Preprint typeset in JHEP style - HYPER VERSION

Chapter 1: Abstract Group Theory

Gregory W. Moore

Abstract: Very Abstract.May 6, 2021

Contents

1. Introduction **5**
- 1.1 Equivalence Relations 5

2. Groups: Basic Definitions And Examples **6**

3. Homomorphism and Isomorphism **19**

4. Group Actions On Sets **25**
- 4.1 Group Actions On Sets 25
- 4.2 Group Actions On Sets Induce Group Actions On Associated Function Spaces 29

5. The Symmetric Group. **29**
- 5.1 Cayley's Theorem 32
- 5.2 Cyclic Permutations And Cycle Decomposition 33
- 5.3 Transpositions 35
- 5.4 Diversion and Example: Card shuffling 40

6. Generators and relations **43**
- 6.1 Example Of Generators And Relations: Fundamental Groups In Topology 50
  - 6.1.1 The Fundamental Group Of A Topological Space 51
  - 6.1.2 Surface Groups: Compact Two-Dimensional Surfaces 57
  - 6.1.3 Braid Groups And Anyons 60
  - 6.1.4 Fundamental Groups Of Three-Dimensional Manifolds 65
  - 6.1.5 Fundamental Groups Of Four-Dimensional Manifolds 67

7. Cosets and conjugacy **68**
- 7.1 Lagrange Theorem 68
- 7.2 Conjugacy 71
- 7.3 Normal Subgroups And Quotient Groups 75
  - 7.3.1 A Very Interesting Quotient Group: Elliptic Curves 84
- 7.4 Conjugacy Classes In S_n 88
  - 7.4.1 Conjugacy Classes In S_n And Harmonic Oscillators 90
  - 7.4.2 Conjugacy Classes In S_n And Partitions 98

8. More About Group Actions And Orbits **100**
- 8.1 Some Definitions And Terminology Associated With Group Actions 101
- 8.2 The Stabilizer-Orbit Theorem 105
- 8.3 Examples Of Orbits 107
  - 8.3.1 Extended Example: The Case Of $1 + 1$ Dimensions 114
  - 8.3.2 Higher Dimensional Light Cones 117

8.3.3 Torsors And Principal Bundles 120
8.4 More About Induced Group Actions On Function Spaces 126

9. Centralizer Subgroups And Counting Conjugacy Classes 128
9.1 $0+1$-Dimensional Gauge Theory 129
9.2 Three Mathematical Applications Of The Counting Principle 130

10. Kernel, Image, And Exact Sequence 134
10.1 The Relation Of $SU(2)$ And $SO(3)$ 141

11. Some Representation Theory 145
11.1 Some Basic Definitions 146
11.2 Characters 150
11.3 Unitary Representations 151
11.4 Haar Measure, a.k.a. Invariant Integration 152
11.5 The Regular Representation 157
11.6 Reducible And Irreducible Representations 160
11.6.1 Definitions 160
11.6.2 Reducible vs. Completely reducible representations 165
11.7 Schur's Lemmas 167
11.8 Pontryagin Duality 170
11.8.1 An Application Of Pontryagin Duality: Bloch's Theorem And Band Structure 173
11.9 Orthogonality Relations Of Matrix Elements And The Peter-Weyl Theorem 175
11.10 Orthogonality Relations For Characters And Character Tables 180
11.10.1 Finite Groups And The Character Table 181
11.10.2 Orthogonality Relations And Pontryagin Duality 187
11.10.3 The Poisson Summation Formula 188
11.11 The Finite Heisenberg Group And The Quantum Mechanics Of A Particle On A Discrete Approximation To A Circle 190
11.12 Decomposition Of Tensor Products Of Representations And Fusion Coefficients 195
11.13 Induced Representations 198
11.13.1 The Geometrical Interpretation 199
11.13.2 Frobenius Reciprocity 200
11.14 Representations Of $SU(2)$ 205
11.14.1 Homogeneous Polynomials 205
11.14.2 Chararacters Of The Representations V_j 210
11.14.3 Unitarization 212
11.14.4 Inhomogeneous Polynomials And Mobius Transformations On $\mathbb{CP}^1$ 212
11.14.5 The Geometrical Interpretation Of $\mathcal{P}_{2j}$ 216
11.15 The Clebsch-Gordon Decomposition For $SU(2)$ 219
11.16 Lie Groups And Lie Algebras And Lie Algebra Representations 221

11.16.1 Some Useful Formulae For Working With Exponentials Of Operators 221
11.16.2 Lie Algebras 228
11.16.3 The Classical Matrix Groups Are Lie Groups 231
11.16.4 Representations Of Lie Algebras 239
11.16.5 Finite Dimensional Irreducible Representations Of $\mathfrak{sl}(2,\mathbb{C})$,$\mathfrak{sl}(2,\mathbb{R})$, and $\mathfrak{su}(2)$ 241
11.16.6 Casimirs 246
11.16.7 Lie Algebra Operators In The Induced Representations Of $SU(2)$ 248

12. Group Theory And Elementary Number Theory 250
12.1 Reminder On gcd And The Euclidean Algorithm 250
12.2 Application: Expressing elements of $SL(2,\mathbb{Z})$ as words in S and T 255
12.3 Products Of Cyclic Groups And The Chinese Remainder Theorem 257
12.3.1 The Chinese Remainder Theorem 262

13. The Group Of Automorphisms 264
13.1 The group of units in $\mathbb{Z}_N$ 270
13.2 Group theory and cryptography 276
13.2.1 How To Break RSA: Period Finding 279
13.2.2 Period Finding With Quantum Mechanics 279

14. Semidirect Products 284

15. Group Extensions and Group Cohomology 299
15.1 Group Extensions 299
15.2 Projective Representations 307
15.2.1 How projective representations arise in quantum mechanics 309
15.3 How To Classify Central Extensions 315
15.4 Extended Example: Charged Particle On A Circle Surrounding A Solenoid 332
15.4.1 Hamiltonian Analysis 332
15.4.2 Remarks About The Quantum Statistical Mechanics Of The Particle On The Ring 344
15.4.3 Gauging The Global $SO(2)$ Symmetry, Chern-Simons Terms, And Anomalies 350
15.5 Heisenberg Extensions 360
15.5.1 Heisenberg Groups: The Basic Motivating Example 360
15.5.2 Example: The Magnetic Translation Group For Two-Dimensional Electrons 363
15.5.3 The Commutator Function And The Definition Of A General Heisenberg Group 364
15.5.4 Classification Of $U(1)$ Central Extensions Using The Commutator Function 366
15.5.5 Pontryagin Duality And The Stone-von Neumann-Mackey Theorem 367
15.5.6 Some More Examples Of Heisenberg Extensions 372

15.5.7 Lagrangian Subgroups And Induced Representations 380
15.5.8 Automorphisms Of Heisenberg Extensions 382
15.5.9 Coherent State Representations Of Heisenberg Groups: The Bargmann Representation 394
15.5.10 Some Remarks On Chern-Simons Theory 394
15.6 Non-Central Extensions Of A General Group G By An Abelian Group A: Twisted Cohomology 394
15.6.1 Crystallographic Groups 399
15.6.2 Time Reversal 402
15.6.3 $T^2 = (-1)^{2j}$ and the Clebsch-Gordon Decomposition 408
15.7 General Extensions 409
15.8 Group cohomology in other degrees 413
15.8.1 Definition 414
15.8.2 Interpreting the meaning of $H^{0+\omega}$ 418
15.8.3 Interpreting the meaning of $H^{1+\omega}$ 418
15.8.4 Interpreting the meaning of $H^{2+\omega}$ 419
15.8.5 Interpreting the meaning of H^3 419
15.9 Some references 421

16. Overview of general classification theorems for finite groups 421
16.1 Brute force 422
16.2 Finite Abelian Groups 427
16.3 Finitely Generated Abelian Groups 430
16.4 The Classification Of Finite Simple Groups 432

17. Categories: Groups and Groupoids 439
17.1 Groupoids 449
17.2 The topology behind group cohomology 451

18. Lattice Gauge Theory 455
18.1 Some Simple Preliminary Computations 455
18.2 Gauge Group And Gauge Field 457
18.3 Defining A Partition Function 460
18.4 Hamiltonian Formulation 465
18.5 Topological Gauge Theory 465

19. Example: Symmetry Protected Phases Of Matter In $1+1$ Dimensions 470

1. Introduction

Historically, group theory began in the early 19th century. In part it grew out of the problem of finding explicit formulae for roots of polynomials. [1]. Later it was realized that groups were crucial in transformation laws of tensors and in describing and constructing geometries with symmetries. This became a major theme in mathematics near the end of the 19th century. In part this was due to Felix Klein's very influential Erlangen program.

In the 20th century group theory came to play a major role in physics. Einstein's 1905 theory of special relativity is based on the symmetries of Maxwell's equations. The general theory of relativity is deeply involved with the groups of diffeomorphism symmetries of manifolds. With the advent of quantum mechanics the representation theory of linear groups, particularly $SU(2)$ and $SO(3)$ came to play an important role in atomic physics, despite Niels Bohr's complaints about "die Gruppenpest." One basic reason for this is the connection between group theory and symmetry, discussed in chapter ****. The theory of symmetry in quantum mechanics is closely related to group representation theory.

Since the 1950's group theory has played an extremely important role in particle theory. Groups help organize the zoo of subatomic particles and, more deeply, are needed in the very formulation of gauge theories. In order to formulate the Hamiltonian that governs interactions of elementary particles one must have some understanding of the theory of Lie algebras, Lie groups, and their representations.

In the late 20th and early 21st century group theory has been essential in many areas of physics including atomic, nuclear, particle, and condensed matter physics. However, the beautiful and deep relation between group theory and geometry is manifested perhaps most magnificently in the areas of mathematical physics concerned with gauge theories (especially supersymmetric gauge theories), quantum gravity, and string theory. It is with that in the background that I decided to cover the topics in the following chapters.

1.1 Equivalence Relations

A very elementary, but very basic idea that we will use repeatedly is that of an *equivalence relation*. A good reference for this elementary material is I.N. Herstein, *Topics in Algebra*, sec. 1.1.

Definition 1.1.1 . Let X be any set. A binary relation $\sim$ is an *equivalence relation* if $\forall a, b, c \in X$

1. $a \sim a$
2. $a \sim b \Rightarrow b \sim a$
3. $a \sim b$ and $b \sim c \Rightarrow a \sim c$

Example 1.1.1 : The notion of equality satisfies these axioms of an equivalence relation. So $a \sim b$ iff $a = b$ is an equivalence relation. The main point, however, is that an equivalence relation is a more flexible notion than equality, and yet captures many of the important aspects of equality.

Example 1.1.2 : $X = \mathbb{Z}$, $a \sim b$ if $a - b$ is even.

[1] For a romantic description, see the chapter "Genius and Stupidity" in E.T. Bell's *Men of Mathematics*. For what is likely a more realistic account see chapter 6 of T. Rothman's *Science à la Mode*.

Example 1.1.3 : More generally, let $X = \mathbb{Z}$, and choose a positive integer N. We can define an equivalence relation by saying that $a \sim b$ iff $a - b$ is divisible by N.

Example 1.1.4 : At the other extreme from equality we could say that every element of the set X is equivalent to every other element. This would be the coarsest possible equivalence relation on the set.

Definition 1.1.2: Let $\sim$ be an equivalence relation on X. The *equivalence class* of an element a is

$$[a] \equiv \{x \in X : x \sim a\} \tag{1.1}$$

In the above two examples we have

Example 1.1.1' : If our equivalence relation is just equality then the equivalence class of every element has only one element: $[a] = \{a\}$. The set of equivalence classes is in bijective correspondence with the original set.

Example 1.1.2' : $[n]$ is the set of all integers with the same parity as n. For example,
$[1] = \{n : n \text{ is an odd integer}\}$
$[4] = \{n : n \text{ is an even integer}\}$.

Example 1.1.3' : Consider the equivalence relation $a \sim b$ iff $a - b$ is divisible by N. Recall that if n is an integer then we can write $n = r + Nq$ in a unique way where the quotient q is integral and the *remainder* or *residue modulo* N is the integer $r \in \{0, 1, \dots, N-1\}$. Thus there is a bijective correspondence between the set of equivalence classes and the set $\{0, 1, \dots, N-1\}$. The equivalence class of an integer n will sometimes be written as $\bar{n}$. One way to write it is

$$\bar{n} := n + N\mathbb{Z} := \{\dots n - 2N, n - N, n, n + N, n + 2N, \dots\} \tag{1.2}$$

Example 1.1.4' : If we take the coarsest possible equivalence relation then there is only one equivalence class, namely the full set X itself.

Here is a simple, but basic, principle:

The distinct equivalence classes of an equivalence relation on X decompose X into a union of mutually disjoint subsets. Conversely, given a *disjoint* decomposition $X = \amalg X_i$ we can define an equivalence relation by saying $a \sim b$ if $a, b \in X_i$.

For example, the integers are the disjoint union of the even and odd integers, and the corresponding equivalence relation is the one mentioned above: $a \sim b$ iff $a - b$ is even.

2. Groups: Basic Definitions And Examples

We begin with the abstract definition of a group.

Definition 2.1: A *group* is a quartet $(G, \mathbf{m}, \mathbf{I}, e)$ where

1. G is a set.

2. $\mathbf{m} : G \times G \to G$ is a map, called the *group multiplication map.*

3. $\mathbf{I} : G \to G$ is a map, called the *inverse map*

4. $e \in G$ is a distinguished element of G called the *identity element.*

These data $(G, \mathbf{m}, \mathbf{I}, e)$ are required to satisfy the following conditions:

1. $\mathbf{m}$ is *associative*: For all $g_1, g_2, g_3 \in G$ we have

$$\mathbf{m}(\mathbf{m}(g_1, g_2), g_3) = \mathbf{m}(g_1, \mathbf{m}(g_2, g_3)) \tag{2.1}$$

2.

$$\forall g \in G \qquad \mathbf{m}(g, e) = \mathbf{m}(e, g) = g \tag{2.2}$$

3.

$$\forall g \in G \qquad \mathbf{m}(\mathbf{I}(g), g) = \mathbf{m}(g, \mathbf{I}(g)) = e \tag{2.3}$$

The above notation is unduly heavy, and we will not use it. Thus, we give the definition again, but more informally:

$\forall a, b \in G$ *there exists a unique element in* G*, called the product, and denoted* $a \cdot b \in G$
in other words, we streamline notation by writing $a \cdot b := \mathbf{m}(a, b)$.
The product is required to satisfy 3 axioms:

1. Associativity: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$

2. Existence of an identity element: $\exists e \in G$ such that:

$$\forall a \in G \qquad a \cdot e = e \cdot a = a \tag{2.4}$$

3. Existence of inverses: Again, we streamline notation by writing $a^{-1} := \mathbf{I}(a)$. so that $a \cdot a^{-1} = a^{-1} \cdot a = e$

Remarks

1. We will often denote e by 1, or, when discussing more than one group at a time, we denote the identity in a particular group G by 1_G. The identity element is also often called the *unit element*, although the term "unit" can have other meanings when dealing with more general mathematical structures such as rings.

2. Also, we sometimes denote the product of a and b simply by ab.

3. We can drop some axioms and still have objects of mathematical interest. For example, a *monoid* is a set M with a multiplication map $\mathbf{m} : M \times M \to M$ which is associative. And that's all. If there is an identity element $e \in M$ which functions as the identity for this multiplication then we speak of a *unital monoid.* The further assumption of inverses turns the monoid into a group. The definition of a group seems to be in the Goldilocks region of having just enough data and conditions to allow a deep theory, but not having too many constraints to allow only a few examples. It is just right to have a deep and rich mathematical theory.

4. We can also put further mathematical structures on the data $(G, \mathbf{m}, \mathbf{I}, e)$. For example, if G is a topological space and $\mathbf{m}$ and $\mathbf{I}$ are both continuous maps, then we have a *topological group*. If G is furthermore a manifold and $\mathbf{m}$ and $\mathbf{I}$ are real analytic in local coordinates, then we have a *Lie group*.

Exercise

a.) Show that e unique. [2]

b.) Given a is a^{-1} unique?

c.) Show that axioms 2,3 above are slightly redundant: For example, just assuming $a \cdot e = a$ and $a \cdot a^{-1} = e$ show that $e \cdot a = a$ follows as a consequence.

Example 2.1: As a set, $G = \mathbb{Z}, \mathbb{R}$, or $\mathbb{C}$. The group operation is ordinary addition:

$$\mathbf{m}(a, b) := a + b \tag{2.5}$$

The reader should check all the axioms.

Example 2.2: A simple generalization is to take n-tuples for a positive integer n: $G = \mathbb{Z}^n, \mathbb{R}^n, \mathbb{C}^n$, with the operation being vector addition, so if $\vec{x} = (x_1, \ldots, x_n)$ and $\vec{y} = (y_1, \ldots, y_n)$ then

$$\mathbf{m}(\vec{x}, \vec{y}) := (x_1 + y_1, \ldots, x_n + y_n) \tag{2.6}$$

Example 2.3: $G = \mathbb{R}^* := \mathbb{R} - \{0\}$ or $G = \mathbb{C}^* := \mathbb{C} - \{0\}$ Now if $x, y \in G$ then $\mathbf{m}(x, y) := xy$ is ordinary multiplication of complex numbers. Check the axioms.

Definition 2.2: Suppose $(G, \mathbf{m}, \mathbf{I}, e)$ is a group and $H \subset G$ is a subset so that $\mathbf{m}$ and $\mathbf{I}$ preserve H, that is, the restriction of $\mathbf{m}$ takes $H \times H \to H$ and the restriction of $\mathbf{I}$ maps $H \to H$. (It then follows that $e \in H$.) In this case we say that $(H, \mathbf{m}, \mathbf{I}, e)$ is a *subgroup of* $(G, \boldsymbol{m}, \boldsymbol{I}, e)$.

Exercise *Subgroups*

a.) $\mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$ with operation $+$, define subgroups.

b.) Is $\mathbb{Z} - \{0\}$ a monoid (with $\mathbf{m}$ given by standard multiplication) ?

c.) Is $\mathbb{Z} - \{0\} \subset \mathbb{R}^*$ a subgroup?

d.) Let $\mathbb{R}^*_{>0}$ and $\mathbb{R}^*_{<0}$ denote the positive and negative real numbers, respectively. Using ordinary multiplication of real numbers, which of these are subgroups of $\mathbb{R}^*$?

e.) Consider the negative real numbers $\mathbb{R}_{<0}$ with the multiplication rule:

$$\mathbf{m}(x, y) = -xy \tag{2.7}$$

[2] *Answer*: Suppose that two elements $e_1, e_2 \in G$ behave as units. Consider the product $e_1 \cdot e_2$. Using e_1 as a unit we can say this is e_2. On the other hand, using e_2 as a unit we can say this is e_1. Therefore $e_1 = e_2$.

Show that this defines a group law on $\mathbb{R}_{<0}$, but that $(\mathbb{R}_{<0}, \mathbf{m}, ...)$ is not a subgroup of $\mathbb{R}^*$. [3]

Definition 2.3: The *order* of a group G, denoted $|G|$, is the cardinality of G as a set. Roughly speaking this is the same as the "number of elements in G." A group G is called a *finite group* if $|G| < \infty$, and is called an *infinite group* otherwise.

Already, with the simple concepts we have just introduced, we can ask nontrivial questions. For example:

Does every infinite group necessarily have proper subgroups of infinite order?

This is of course true of the examples we have just discussed. It is actually not easy to think of counterexamples, but in fact there are infinite groups all of whose proper subgroups are finite. [4]

Let us continue with an overview of examples of groups:

The groups in Examples 1,2,3 above are of infinite order. Here are examples of finite groups:

Example 2.4: The group of Nth roots of unity. Choose a natural number N. [5] We let μ_N be the set of complex numbers z such that $z^N = 1$. Thus we could write

$$\mu_N = \{1, \omega, \ldots, \omega^{N-1}\} \tag{2.8}$$

where $\omega = \exp[2\pi \mathrm{i}/N]$. This is a finite group with N elements, as is easily checked.

Exercise

Does μ_{137} have any nontrivial subgroups? [6]

Exercise

[3] *Answer*: The point is that the multiplication of (2.7) is not the restriction of the multiplication on $\mathbb{R}^*$ to the negative reals.

[4] One example are the *Prüfer groups*. These are subgroups of the group of roots of unity. They are defined by choosing a prime number p and taking the subgroup of roots of unity of order p^n for some natural number n. Even wilder examples are the "Tarski Monster groups" (not to be confused with the Monster group, which we will discuss later). These are infinite groups all of whose subgroups are isomorphic to the cyclic group of order p.

[5] The *natural numbers* are the same as the positive integers.

[6] *Answer*: We will give an elegant answer below.

In example 4 show that if N is even then the subset of classes of even integers forms a proper subgroup of μ_N. What happens if N is odd?

Example 2.5: **The residue classes modulo N, also called "The cyclic group of order N."** Let N be a positive integer. Recall that we can put an equivalence relation on $\mathbb{Z}$ defined by $a \sim b$ iff $a - b$ is divisible by N, and we denoted the class of an integer n by $\bar{n}$. (One could identify the set of equivalence classes with the set $\{0, 1, \ldots, N-1\}$.) We take G to be the set of equivalence classes of integers modulo N, using the notation of equivalence relation of §7.1 above. We need to define $\mathbf{m}(\bar{r}_1, \bar{r}_2)$. To do this we choose a representative r_1, r_2 from each equivalence class and take

$$\mathbf{m}(\bar{r}_1, \bar{r}_2) := \overline{(r_1 + r_2)} \tag{2.9}$$

The main thing to check here is that the equation is well-defined, since we chose representatives for each equivalence class. This group, which appears frequently in the following, will be denoted as $\mathbb{Z}/N\mathbb{Z}$ or $\mathbb{Z}_N$. For example, telling time in hours is arithmetic in $\mathbb{Z}_{12}$, or in $\mathbb{Z}_{24}$ in railroad/military time. The reader should note that it "resembles" closely the group μ_N. We will make that precise in the next section.

So far, all our examples had the property that for any two elements a, b

$$a \cdot b = b \cdot a \tag{2.10}$$

Definition 2.5: When equation (2.10) holds for two elements $a, b \in G$ we say *"a and b commute."* If a and b commute for every pair $(a, b) \in G \times G$ then we say that G is an *Abelian group*:

If a, b commute for all $a, b \in G$ we say "*G is Abelian.*"

Note: Note that our abbreviated notation $a \cdot b$ for the group multiplication $\mathbf{m}(a, b)$ would actually be quite confusing when working with $\mathbb{Z}_N$. The reason is that it is also possible to define a *ring structure* (see Chapter 2) where one multiplies r_1 and r_2 as integers and then takes the residue. This is *NOT* the same as $\mathbf{m}(r_1, r_2)$!! For example, if we take $N = 5$ then $\mathbf{m}(2, 3) = 0$ in $\mathbb{Z}_5$ because $2 + 3 = 5$ is congruent to 0 modulo 5. Of course, multiplying as integers $2 \times 3 = 6$ and 6 is congruent to 1 mod 5. When considering Abelian groups we often prefer to use the abbreviated notation

$$a + b := \mathbf{m}(a, b) \tag{2.11}$$

When we use this additive notation for Abelian groups we will write the identity element as 0 so that $a + 0 = 0 + a = a$. (Writing "$a + 1 = a$" would look extremely weird.) Note that we will not always use additive notation for Abelian groups! For example, for μ_N the multiplicative notation is quite natural.

Since we defined a notion of "Abelian group" we are implicitly suggesting there are examples of groups which are not Abelian. If one tries to use the group axioms to prove that $\mathbf{m}(g_1, g_2) = \mathbf{m}(g_2, g_1)$ one will fail. The only way we can know conclusively that

one will fail is to provide a counterexample. The next example gives a set of examples of nonabelian groups:

Example 2.6: *The General Linear Group*

Let $\kappa = \mathbb{R}$ or $\kappa = \mathbb{C}$. Define $M_n(\kappa)$ to be the set of all $n \times n$ matrices whose matrix elements lie in κ. Note that this is a unital monoid under matrix multiplication. But it is not a group, because some matrices are not invertible. Therefore we define:

♣κ will be our official symbol for a general field. This needs to be changed from k in many places below. ♣

$$GL(n,\kappa) := \{A | A = n \times n \text{ invertible matix over } \kappa\} \subset M_n(\kappa) \tag{2.12}$$

When $\kappa = \mathbb{R}$ or $\kappa = \mathbb{C}$ $GL(n,\kappa)$ is a group of infinite order. It is Abelian if $n = 1$ and nonabelian if $n > 1$. There are some important generalizations of this example: [7] We could let κ be any field. If κ is a finite field then $GL(n,\kappa)$ is a finite group. More generally, if R is a *ring* $GL(n,R)$ is the subset of $n \times n$ matrices with entries in R with an inverse in $M_n(R)$. This set forms a group. For example, $GL(2,\mathbb{Z})$ is the set of 2×2 matrices of integers such that the inverse matrix is also a 2×2 matrix of integers. This set of matrices forms an infinite nonabelian group under matrix multiplication.

Definition 2.5: The center $Z(G)$ of a group G is the set of elements $z \in G$ that commute with all elements of G:

$$Z(G) := \{z \in G | zg = gz \qquad \forall g \in G\} \tag{2.13}$$

Exercise *Due Diligence: The Center* :

a.) Show that for any group G, $Z(G)$ is an Abelian subgroup of G.

b.) Show that the center of $GL(n,\kappa)$ is the subgroup of matrices proportional to the unit matrix with scalar factor in κ^*. [8]

Example 2.7: *The Classical Matrix Groups*

[7] See Chapter 2 for some discussion of the mathematical notions of fields and rings used in this paragraph.

[8] *Answer*: It is obvious that matrices of the form $z1_{n\times n}$ with $z \in \kappa^*$ are in the center. What is not immediately obvious is that there are no <u>other</u> elements of the center. Here is a careful proof that this is indeed the case: Consider the *matrix units*: e_{ij}. The matrix e_{ij} has a 1 in the ith row and j^{th} column and zeroes elsewhere. Note that for any matrix A we have $e_{ii}Ae_{jj} = A_{ij}e_{ij}$ with no sum on i,j here. On the RHS A_{ij} is a matrix element, not a matrix. Now let z be in the center. Check that for any pair ij the matrix $1 + e_{ij}$ is invertible. Therefore, if z is in the center then z must commute with $1 + e_{ij}$ and hence z must commute with e_{ij} for all i,j. Now, as we observed above, $e_{ii}ze_{jj} = z_{ij}e_{ij}$ holds for any matrix, but since z is also central $e_{ii}ze_{jj} = e_{ii}e_{jj}z = \delta_{ij}e_{jj}z$ So z is diagonal. But for any diagonal matrix $z = \sum_k z_k e_{kk}$ we have $(zA)_{ij} = z_i A_{ij}$ and $(Az)_{ij} = A_{ij}z_j$. As long as there are matrices with $A_{ij} \neq 0$ and invertible we can conclude that $z_i = z_j$.

A *matrix group* is a subgroup of $GL(n,\kappa)$. There are several interesting examples which we will study in great detail later. Some examples include:

The special linear group:

$$SL(n,\kappa) \equiv \{A \in GL(n,\kappa) : \det A = 1\} \tag{2.14}$$

The orthogonal groups:

$$\begin{aligned} O(n,\kappa) &:= \{A \in GL(n,\kappa) : AA^{tr} = 1\} \\ SO(n,\kappa) &:= \{A \in O(n,\kappa) : \det A = 1\} \end{aligned} \tag{2.15}$$

Another natural class are the unitary and special unitary groups:

$$U(n) := \{A \in GL(n,\mathbb{C}) : AA^{\dagger} = 1\} \tag{2.16}$$

$$SU(n) := \{A \in U(n) : \det A = 1\} \tag{2.17}$$

Finally, to complete the standard list of classical matrix groups we consider the standard symplectic form on $\mathbb{R}^{2n}$:

$$J = \begin{pmatrix} 0 & 1_{n\times n} \\ -1_{n\times n} & 0 \end{pmatrix} \in M_{2n}(\mathbb{R}) \tag{2.18}$$

Note that the matrix J satisfies the properties:

$$J = J^{*} = -J^{tr} = -J^{-1} \tag{2.19}$$

Definition A *symplectic matrix* is a matrix A such that

$$A^{tr} J A = J \tag{2.20}$$

We define the symplectic groups:

$$Sp(2n,\kappa) := \{A \in GL(2n,\kappa) | A^{tr} J A = J\} \tag{2.21}$$

Remarks:

1. As an exercise you should show from the definition above that the most general element of $SO(2,\mathbb{R})$ must be of the form

$$\begin{pmatrix} x & y \\ -y & x \end{pmatrix} \qquad x^2 + y^2 = 1 \tag{2.22}$$

where the matrix elements x, y are real. Thus we recognize that group elements in $SO(2,\mathbb{R})$ are in 1-1 correspondence with points on the unit circle in the plane. We

can even go further and parametrize $x = \cos\phi$ and $y = \sin\phi$ and ϕ is a coordinate provided we identify $\phi \sim \phi + 2\pi$ so the general element of $SO(2,\mathbb{R})$ is of the form:

$$R(\phi) := \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \tag{2.23}$$

This is familiar from the implementation of rotations of the Euclidean plane in Cartesian coordinates. Note that the group multiplication law is

$$R(\phi_1)R(\phi_2) = R(\phi_1 + \phi_2) \tag{2.24}$$

so, in ϕ "coordinates" the group multiplication law is continuous, differentiable, even (real) analytic. Similarly, taking an inverse is $\phi \to -\phi$.

2. Let us consider the group $U(1)$: This is simply the group of 1×1 unitary matrices. They are not hard to diagonalize. The general matrix can be written as $z(\phi) = e^{\mathrm{i}\phi}$ with multiplication $z(\phi_1)z(\phi_2) = z(\phi_1 + \phi_2)$ where $\phi \sim \phi + 2\pi$ yield identical group elements. Again, as with μ_N and $\mathbb{Z}_N$ the groups look like they are "the same" although strictly speaking they are different sets and therefore have different $\mathbf{m}$'s. We will make this idea precise in the next section.

3. One of the most important groups in both mathematics and physics is $SU(2)$. We claim that the general element of $SU(2)$ looks like

$$g = \begin{pmatrix} z & -w^* \\ w & z^* \end{pmatrix} \tag{2.25}$$

for a pair of complex numbers $(z,w) \in \mathbb{C}^2$ such that

$$|z|^2 + |w|^2 = 1 \tag{2.26}$$

One can prove this by studying the 4 equations for the matrix elements in the identity $gg^\dagger = 1$. Another way to proceed makes use of some concepts from the linear algebra chapter below and goes as follows: Since g is unitary it follows that the basis

$$g\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad g\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.27}$$

Should be orthonormal. Therefore, if we write

$$g = \begin{pmatrix} z & u \\ w & v \end{pmatrix} \tag{2.28}$$

it must be that $\begin{pmatrix} z \\ w \end{pmatrix}$ is orthogonal to $\begin{pmatrix} u \\ v \end{pmatrix}$ and hence $u = -\lambda w^*$ and $v = \lambda z^*$. Moreover, since norms are preserved we know that $|z|^2 + |w|^2 = 1$. So the general unitary matrix must be of the form

$$g = \begin{pmatrix} z & -\lambda w^* \\ w & \lambda z^* \end{pmatrix} \tag{2.29}$$

for some phase λ. But now if we impose $\det g = 1$ we discover that $\lambda = 1$. Note that, writing out the real and imaginary parts of z, w we see that the equation (2.26) is just the equation for a 3-dimensional sphere: Just the way $SO(2)$ and $U(1)$ are, as manifolds, S^1, we see that $SU(2)$ also has the structure of a manifold, namely the three-dimensional sphere.

4. Some of what we have just said can be generalized to all the classical matrix groups: They can be identified with manifolds (although no other matrix groups are spheres - not an obvious fact.) There are parametrizations of these manifolds, so the group multiplication and inverse are smooth operations. Moreover, the groups "act" naturally on various linear spaces. (See Section 4.1 for the notion of "group action.") This is part of the theory of Lie groups. Lie groups have vast applications in physics. For example, $G = SU(3)$ is the gauge group of a Yang-Mills theory that describes the interactions of quarks and gluons, while $G = SU(3) \times SU(2) \times U(1)$ is related to the standard model that describes all known elementary particles and their interactions. The general theory of Lie groups will be discussed in Chapter 8(?) below, although we will meet many many examples before then.

Example 2.8 *Function spaces as groups.*

Suppose G is a group. Suppose X is any set. Consider the set of all functions from X to G:

$$\mathcal{F} = \{f : f \text{ is a function from } X \to G\} \tag{2.30}$$

If we want to stress the role of X and/or G we write $\mathcal{F}[X \to G]$ for $\mathcal{F}$. We claim that $\mathcal{F}$ is also a group. The main step to show this is simply giving a definition of the group multiplication and the inversion operation. The product $\mathbf{m}_{\mathcal{F}}(f_1, f_2)$ of two functions $f_1, f_2 \in \mathcal{F}$ must be another function in $\mathcal{F}$. We define this function by giving a formula for the values at all values of $x \in X$:

$$\mathbf{m}_{\mathcal{F}}(f_1, f_2)(x) := \mathbf{m}_G(f_1(x), f_2(x)) \tag{2.31}$$

It is the only sensible thing we could write given the data at hand. In less cumbersome notation:

$$(f_1 \cdot f_2)(x) := f_1(x) \cdot f_2(x) \tag{2.32}$$

Similarly inverse of f is the function that maps $x \to f(x)^{-1}$, where $f(x)^{-1} \in G$ is the group element in G inverse to $f(x) \in G$.

If both X and G have finite cardinality then $\mathcal{F}[X \to G]$ is a finite group. If X or G has an infinite set of points then this is an infinite order group. If X is a positive dimensional manifold and G is a Lie group (notions defined below) this is an infinite-*dimensional* space.

In the special case of the space of maps from the circle into the group:

$$LG = \mathcal{F}[S^1 \to G] \tag{2.33}$$

we have the famous "loop group" whose representation theory has many wonderful properties, closely related to the subjects of 2d conformal field theory and string theory. In

some cases if X is a manifold and G is a classical matrix group then, taking a subgroup defined by suitable continuity and differerentiability properties, we get the *group of gauge transformations of Yang-Mills theory.* As a simple example, you are probably familiar with the gauge transformation in Maxwell theory:

$$A_\mu \to A_\mu + \partial_\mu \epsilon \tag{2.34}$$

where A_μ is the vector potential so that $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the fieldstrength tensor. Here $\epsilon : \mathbb{M}^{1,3} \to \mathbb{R}$ is a function on 1+3 dimensional Minkowski space. (In a more careful account one would put restrictions on the allowed functions - they should be differentiable and satisfy suitable boundary conditions - etc.) The more canonical object is

$$f : x \mapsto e^{\mathrm{i}\epsilon(x)} \tag{2.35}$$

and this is a function from spacetime, $\mathbb{M}^{1,3}$ to $U(1)$, so $f \in \mathcal{F}[\mathbb{M}^{1,3} \to U(1)]$. This is a better point of view because it generalizes in interesting ways to other spacetimes. Note that the gauge transformation law can be written as:

$$(-\mathrm{i}\partial_\mu + A'_\mu) = f^{-1}(-\mathrm{i}\partial_\mu + A_\mu)f \tag{2.36}$$

For many reasons this is a conceptually superior way to write it.

Example 2.9: *Permutation Groups.*

Let X be any set. A *permutation* of X is a one-one invertible transformation $\phi : X \to X$. The composition $\phi_1 \circ \phi_2$ of two permutations is a permutation. The identity permutation leaves every element unchanged. The inverse of a permutation is a permutation. Thus, composition defines a group operation on the permutations of any set. This group is designated S_X. It is an extremely important group and we will be studying it a lot. In the case where $X = M$ is a manifold we can also ask that our permutations $\phi : M \to M$ be continuous or even differentiable. If ϕ and ϕ^{-1} are differentiable then ϕ is a *diffeomorphism.* The composition of diffeomorphisms is a diffeomorphism by the chain rule, so the set of diffeomorphisms $\mathrm{Diff}(M)$ is a subgroup of the set of all permutations of M. The group $\mathrm{Diff}(M)$ is the group of gauge symmetries in General Relativity. Except in the case where $M = S^1$ is the circle, remarkably little is known about the diffeomorphism groups of manifolds. One can ask simple questions about them whose answers are unknown.

Example 2.10: *Power Sets As Groups.*

Let X be any set and let $\mathcal{P}(X)$ be the power set of X. It is, by definition, the set of all subsets of X. If $Y_1, Y_2 \in \mathcal{P}(X)$ are two subsets of X then define the *symmetric difference*:

$$Y_1 + Y_2 := (Y_1 - Y_2) \cup (Y_2 - Y_1) \tag{2.37}$$

This defines an abelian group structure on $\mathcal{P}(X)$. The identity element 0 is the empty set $\emptyset$ and the inverse of Y is Y itself: That is, in this group

$$2Y := Y + Y = \emptyset = 0 \tag{2.38}$$

Definition 2.4 Let G_1, G_2 be two groups. The *direct product* of G_1, G_2 is the set $G_1 \times G_2$ with product:

$$\mathbf{m}_{G_1 \times G_2}\left((g_1, g_2), (g_1', g_2')\right) = (\mathbf{m}_{G_1}(g_1, g_1'), \mathbf{m}_{G_2}(g_2, g_2')) \tag{2.39}$$

Exercise *Due Diligence: Direct Product Of Groups*

a.) Check the group axioms.

b.) Generalize this to arbitrary products: Given a map $\mathcal{G}$ from a set I to the set of all groups define the product over I as a group.

c.) Show that the direct product of two finite groups is finite.

Remark: It turns out that there can be several interesting ways to put a group structure $\mathbf{m}$ on the set $G_1 \times G_2$. We will explore this in great detail in sections **** below.

Exercise *Classical Matrix Groups: Due diligence*

Check that each of the above sets (2.14),(2.15),(2.16), (2.21), are indeed subgroups of the general linear group.

Exercise *Apparent Asymmetry In The Definitions*

In (2.15) we used $AA^{tr} = 1$ but we could have used $A^{tr}A = 1$. Similarly, in (2.16) we used $AA^\dagger = 1$ rather than $A^\dagger A = 1$. Finally, in (2.21) we could, instead, have defined $Sp(2n, \kappa)$ to be matrices in $M_{2n}(\kappa)$ such that $AJA^{tr} = J$. In all three cases, writing things the other way defines the same group: Why?

(Careful: Just taking the transpose or hermitian conjugate of these equations does not help.) [9]

Exercise $O(2, \mathbb{R})$ *vs.* $SO(2, \mathbb{R})$

a.) Show from the definition above of $O(2, \mathbb{R})$ that the most general element of this group is the form of (2.22) above, OR, of the form

$$\begin{pmatrix} x & y \\ y & -x \end{pmatrix} \qquad x^2 + y^2 = 1 \tag{2.40}$$

[9] *Answer*: Hint: Remember that in a group the inverse matrix is in the group. Consider replacing $g \to g^{-1}$ in the definition.

btw: Note that

$$\begin{pmatrix} x & y \\ y & -x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x & y \\ -y & x \end{pmatrix} \tag{2.41}$$

b.) Show that no matrix in $O(2,\mathbb{R})$ is simultaneously of the form (2.22) and (2.40). Conclude that, as a manifold, $O(2,\mathbb{R})$ is a disjoint union of two circles.

Exercise *Symplectic groups and canonical transformations*

Let q^i, p_i $i = 1, \dots n$ be coordinates and momenta for a classical mechanical system.

The **Poisson bracket** of two functions $f(q^1, \dots q^n, p_1, \dots p_n)$, $g(q^1, \dots q^n, p_1, \dots p_n)$ is defined to be

$$\{f, g\} = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} \right) \tag{2.42}$$

a.) Show that

$$\{q^i, q^j\} = \{p_i, p_j\} = 0 \qquad \{q^i, p_j\} = \delta^i{}_j \tag{2.43}$$

Suppose we define new coordinates and momenta Q^i, P_i to be linear combinations of the old:

$$\begin{pmatrix} Q^1 \\ \vdots \\ Q^n \\ P_1 \\ \vdots \\ P_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1,2n} \\ \vdots & \ddots & \vdots \\ a_{2n,1} & \cdots & a_{2n,2n} \end{pmatrix} \cdot \begin{pmatrix} q^1 \\ \vdots \\ q^n \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \tag{2.44}$$

where $A = (a_{ij})$ is a constant $2n \times 2n$ matrix.

b.) Show that

$$\{Q^i, Q^j\} = \{P_i, P_j\} = 0 \qquad \{Q^i, P_j\} = \delta^i_j \tag{2.45}$$

if and only if A is a symplectic matrix.

c.) Show that $J \in Sp(2n, \mathbb{R})$. Note that it exchanges momenta and coordinates.

d.) What are the conditions on the $n \times n$ matrix B so that

$$\{ \begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} \} \tag{2.46}$$

is a subgroup. [10]

e.) What are the conditions on the $n \times n$ matrix C so that

$$\{ \begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix} \} \tag{2.47}$$

[10] *Answer*: B must be a symmetric matrix

is a subgroup. [11]

f.) Show that

$$\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix} = J \begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} J^{-1} \tag{2.48}$$

for $C = -B$.

Exercise *The Quaternion Group And The Pauli Group*

When working with spin-1/2 particles it is very convenient to introduce the standard Pauli matrices:

$$\sigma^1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.49}$$

$$\sigma^2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \tag{2.50}$$

$$\sigma^3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.51}$$

a.) Show that they satisfy the identity, valid for all $1 \leq i, j \leq 3$:

$$\boxed{\boxed{\sigma^i \sigma^j = \delta^{ij} + \mathrm{i}\epsilon^{ijk}\sigma^k}} \tag{2.52}$$

b.) Show that the set of matrices

$$Q = \{\pm 1, \pm \mathrm{i}\sigma^1, \pm \mathrm{i}\sigma^2, \pm \mathrm{i}\sigma^3\} \tag{2.53}$$

forms a subgroup of order 8 of $SU(2) \subset GL(2, \mathbb{C})$. It is known as the *quaternion group.*

c.) Show that the set of matrices

$$P = \{\pm 1, \pm \mathrm{i}, \pm\sigma^1, \pm\sigma^2, \pm\sigma^3, \pm \mathrm{i}\sigma^1, \pm \mathrm{i}\sigma^2, \pm \mathrm{i}\sigma^3\} \tag{2.54}$$

forms a subgroup of $U(2) \subset GL(2, \mathbb{C})$ of order 16. It is known as the *Pauli group.*

Remark: [12] The Pauli group is often used in quantum information theory. If we think of the quantum Hilbert space of a spin 1/2 particle (isomorphic to $\mathbb{C}^2$ with standard inner product) then there is a natural basis of up and down spins: $v_1 = |\uparrow\rangle$ and $v_2 = |\downarrow\rangle$. Thinking of these as quantum analogs $|0\rangle$ and $|1\rangle$ of classical information bits $0, 1$ we see that $X = \sigma^1$ acts as a "bit flip," while $Z = \sigma^3$ acts as a "phase-flip." $Y = \mathrm{i}\sigma^2$ flips both

[11] *Answer*: C must be a symmetric matrix

[12] Many terms used here will be more fully explained in Chapter 2.

bits and phases. These are then quantum error operators. Note that if we have a chain of N spin 1/2 particles then the Nth direct product

$$P^N = \underbrace{P \times \cdots \times P}_{N \text{ times}} \tag{2.55}$$

acts naturally on this chain of particles. [13] This group is useful in quantum information theory. For example if $H \subset P^N$ is a subgroup such that $(-1, ..., -1)$ is not in H then we can study the subspace of Hilbert space $\{\psi | g\psi = \psi, \quad \forall g \in H\}$. For astutely chosen subgroups these are useful quantum code subspaces, known as *stabilizer codes*.

Exercise *Function Groups*

Interpret the direct product G^n of a group with itself n times as a group of the form $\mathcal{F}[X \to G]$ for some X.

---

3. Homomorphism and Isomorphism

Definition 3.1: Let $(G, \mathbf{m}, \mathbf{I}, e)$ and $(G', \mathbf{m}', \mathbf{I}', e')$ be two groups,

1.) A *homomorphism* from $(G, \mathbf{m}, \mathbf{I}, e)$ to $(G', \mathbf{m}', \mathbf{I}', e')$ is a mapping that preserves the group law. That is, it is a map of sets $\varphi : G \to G'$ such that, for all $g_1, g_2 \in G$ we have:

$$\varphi(\mathbf{m}(g_1, g_2)) = \mathbf{m}'(\varphi(g_1), \varphi(g_2)) \tag{3.1}$$

2.) If φ is 1-1 and onto it is called an *isomorphism*.

3.) One often uses the term *automorphism* of G when φ is an isomorphism and $G = G'$, that is G and G' are literally the same set with the same multiplication law.

Remarks

1. We will henceforward be more informal and simply say that $\varphi : G \to G'$ is a homomorphism of groups if, for all $g_1, g_2 \in G$:

$$\varphi(\underbrace{g_1 g_2}_{\text{product in G}}) = \overbrace{\varphi(g_1)\varphi(g_2)}^{\text{product in G}'} \tag{3.2}$$

2. A common slogan is: "isomorphic groups are the same."

[13] See 4.1 for the formal definition of a group action on a space.

Example 1: μ_N is isomorphic to $\mathbb{Z}_N$: Let N be a positive integer. Then we can define a homomorphism

$$\varphi : \mathbb{Z}_N \to \mu_N \tag{3.3}$$

as follows. We want to define $\varphi(\bar{r})$. Recall that $\bar{r} = r + N\mathbb{Z}$ is an equivalence class. We choose any representative $r' \in r + N\mathbb{Z}$. Then we set:

$$\varphi(\bar{r}) := \exp\left(2\pi \mathrm{i} \frac{r'}{N}\right) \tag{3.4}$$

There is a crucial thing to check here: We need to check that the map is actually well-defined. We know that any two representatives r_1' and r_2' for $\bar{r}$ must have the property that $r_1' - r_2' = 0 \ \mathrm{mod} N$, that is $r_1' - r_2' = \ell N$ for some integer N and now by standard properties of complex numbers we see that indeed $\exp\left(2\pi \mathrm{i} \frac{r_1'}{N}\right) = \exp\left(2\pi \mathrm{i} \frac{r_2'}{N}\right)$.

Next we check that

$$\varphi(\bar{r}_1 + \bar{r}_2) = \varphi(\bar{r}_1)\varphi(\bar{r}_2) \tag{3.5}$$

If you unwind the definitions you should find this follows from a standard property of the exponential map.

Equation (3.5) implies that (3.3) is a homomorphism. In fact one easily checks:

a.) If $\varphi(\bar{r}) = 1$ then $\bar{r} = \bar{0}$. Thus φ is 1-1 (a.k.a. "injective").

b.) Every element of μ_N is of the form $\varphi(\bar{r})$ for some $\bar{r}$. Thus, φ is onto (a.k.a. "surjective"). Note that this is equivalent to saying that every element in μ_N is of the form ω^j where $\omega = e^{2\pi \mathrm{i}/N}$.

Thus, φ is in fact an isomorphism. As we mentioned above, the two groups "seemed to be the same." We have now given precise meaning to that idea.

Example 2: A family of homomorphisms $\mu_N \to \mu_N$: For each integer k we can define the kth power map

$$p_k : \mu_N \to \mu_N \tag{3.6}$$

by

$$p_k(z) = z^k \tag{3.7}$$

where z is any N^{th} root of unity. Note that z^k is also an Nth root of unity. Moreover $(z_1 z_2)^k = z_1^k z_2^k$ by elementary properties of complex numbers, so p_k is a homomorphism. Note that it is not always injective or surjective. For example, if k is a multiple of N it is the stupid homomorphism. In fact $p_{k+N} = p_k$.

Example 3: A family of homomorphisms $\mathbb{Z}_N \to \mathbb{Z}_N$:

For any integer k we can define the "k^{th} multiplication map"

$$m_k : \mathbb{Z}_N \to \mathbb{Z}_N \tag{3.8}$$

by the equation:

$$m_k(\bar{r}) := \overline{kr} \tag{3.9}$$

where on the right hand side $\overline{kr}$ is defined by choosing a representative r for the class $\bar{r}$ and then using ordinary multiplication of integers $k \times r$ (e.g. $2 \times 3 = 6$) and then reducing modulo N. Again, one needs to check the equation is well-defined. Note that $m_{k+N} = m_k$.

Example 4: Relating the homomorphisms in the previous three examples: Since $\mathbb{Z}_N$ and μ_N are isomorphic, one should expect that homomorphisms $\mathbb{Z}_N \to \mathbb{Z}_N$ and $\mu_N \to \mu_N$ should be related. Moreover, one should have the intuition that p_k and m_k somehow have the "same effect." Indeed, note that

$$p_k(\omega^j) = (\omega^j)^k = \omega^{jk}. \tag{3.10}$$

is the essential identity. More formally, one easily checks that

$$\varphi \circ m_k = p_k \circ \varphi \tag{3.11}$$

Or, since φ is invertible,

$$p_k = \varphi \circ m_k \circ \varphi^{-1} \ . \tag{3.12}$$

In mathematics one often uses *commutative diagrams* to express identities such as (3.11). In this case the diagram looks like

$$\begin{array}{ccc} \mathbb{Z}_N & \xrightarrow{m_k} & \mathbb{Z}_N \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \varphi} \\ \mu_N & \xrightarrow{p_k} & \mu_N \end{array} \tag{3.13}$$

We say *a diagram commutes* if the following condition holds: The diagram describes a graph with sets associated to vertices and maps associated with oriented edges. Consider following the arrows around any two paths on the graph with the same beginning and final points. We compose the maps associated with those arrows to get two maps from the initial set to the final set. The diagram commutes iff any pair of maps obtained this way are equal.

Remark We will discuss in detail later on that when k is an integer relatively prime to N the map p_k is an automorphism of μ_N and m_k is an automorphism of $\mathbb{Z}_N$. For example in $\mathbb{Z}/3\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}\}$ if we take $k = 2$, or any even integer not divisible by 3, then μ exchanges $\bar{1}$ and $\bar{2}$. (Check that such an exchange is indeed a homomorphism!) We will discuss this kind of example in greater detail in Section §13 below.

One kind of homomomorphism is especially important:

Definition 3.2: A *matrix representation* of a group G is a homomorphism

$$T : G \to GL(n, \kappa) \tag{3.14}$$

for some positive integer n and field κ. (One can also have matrix representations in $GL(n, R)$ where R is a ring.)

More generally, if V is a vector space over a field κ let $GL(V)$ denote the group of all invertible linear transformations from $V \to V$. (The group multiplication is composition.) If G is a group then a homomorphism $T : G \to GL(V)$ defines a *representation of* G. Sometimes V is referred to as the *carrier space*.

****

PUT EXAMPLE OF HOMOMORPHISM $\pi : SU(2) \to SO(3)$ HERE. IT GIVES A 3-dimensional MATRIX REP OF $SU(2)$.

CURRENTLY IT IS IN SECTION 10.1

BUT SOME OF THAT SHOULD BE PUT HERE

****

Exercise *Preservation Of Structure*

Show that, for any group homomorphism μ we always have:

$$\mu(1_G) = 1_{G'} \tag{3.15}$$

$$\mu(g^{-1}) = \mu(g)^{-1} \tag{3.16}$$

Exercise *The Stupid Homomorphism*

Consider the map $\mu : G \to G'$ defined by $\mu(g) = 1_{G'}$. Show that this is a homomorphism.

Exercise *Some Simple Isomorphisms*

a.) Show that the exponential map $x \to e^x$ defines an isomorphism between the additive group $(\mathbb{R}, > 0)$ and the multiplicative group $(\mathbb{R}^*_+, \times)$.

b.) Show that $SO(2)$ and $U(1)$ are isomorphic groups.

c.) Show that $z \to z^{-1}$ is an automorphism of $U(1) \to U(1)$. What is the corresponding automorphism of $SO(2)$?

Exercise *A group "with one free generator"*

Consider a group with a nontrivial element g_0 such that every element in the group if a power of g_0 or g_0^{-1} and $g_0^n = g_0^m$ iff $n = m$ in the integers.

Show that this group is isomorphic to $\mathbb{Z}$.

Remark: This is an example of what we will call below a group freely generated by one element.

Exercise *Sometimes diagrams don't commute*

Show that the diagram

$$\begin{array}{ccc} \mathbb{Z}_N & \xrightarrow{m_{k_1}} & \mathbb{Z}_N \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \varphi} \\ \mu_N & \xrightarrow{p_{k_2}} & \mu_N \end{array} \tag{3.17}$$

commutes iff $k_1 = k_2 \bmod N$.

Exercise *The Quaternion Group*

Construct a homomorphism

$$\mu : Q \to \mathbb{Z}_2 \times \mathbb{Z}_2 \tag{3.18}$$

where Q is the Quaternion group (2.53).

Exercise *Subgroups of $\mathbb{Z}_N$*

a.) Show that the subgroups of $\mathbb{Z}_N$ are isomorphic to the groups $\mathbb{Z}_M$ for $M|N$.

b.) For $N = 8, M = 4$ write out H.

Exercise

Let S_2 be any set with two elements

a.) Show that there are exactly two possible group structures on S_2, and in each case construct an isomorphism of S_2 with $\mu_2 \cong \mathbb{Z}_2$.

b.) Consider the matrix group of two elements:

$$\hat{S}_2 = \{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \} \tag{3.19}$$

with multiplication being matrix multiplication. Construct an isomorphism with S_2. [14]

Exercise *Some Simple Representations Of* μ_N

Let $\omega = e^{2\pi i/N}$.

a.) Show that for any integer k the kth power map

$$p_k(\omega^j) = \omega^{jk} \tag{3.21}$$

defines a representation of μ_N by 1×1 matrices.

b.) Show that

$$\mu : \omega^j \mapsto R(\frac{2\pi j}{N}) := \begin{pmatrix} \cos(\frac{2\pi j}{N}) & \sin(\frac{2\pi j}{N}) \\ -\sin(\frac{2\pi j}{N}) & \cos(\frac{2\pi j}{N}) \end{pmatrix} \tag{3.22}$$

defines a two-dimensional matrix representation of $\mathbb{Z}_N$.

c.) Let P be the $N \times N$ "shift matrix" all of whose matrix elements are zero except for 1's just below the diagonal and $P_{1,N} = 1$. See equation (10.19) below. Show that

$$\mu(\omega^j) = P^j \tag{3.23}$$

is an $N \times N$ dimensional representation of μ_N.

Exercise *Two Characterizations Of Abelian Groups*

Let G be a group.

a.) Consider the map: $\mu : G \to G$ given by squaring: $\mu(g) = g^2$. Show that μ is a group homomorphism iff G is Abelian.

b.) Consider the map:

$$G \times G \to G \tag{3.24}$$

defined by group multiplication: $\mu(g_1, g_2) = \mathbf{m}(g_1, g_2) = g_1 g_2$. Show that μ is a group homomorphism iff G is Abelian.

---

[14] *Answer*: Write $S_2 = \{e, \sigma\}$ with e the identity and $\sigma^2 = e$. Define $\mu : S_2 \to \hat{S}_2$

$$\begin{aligned} \mu(e) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \mu(\sigma) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{3.20}$$

Exercise *Isomorphisms And Preservation Of Structure*

a.) Suppose $\varphi : G_1 \to G_2$ is an isomorphism. Show that φ^{-1} is an isomorphism.

b.) Suppose that $\varphi : G_1 \to G_2$ is an isomorphism, and $\varphi' : G_1' \to G_2'$ is an isomorphism. Suppose also that $\nu_1 : G_1 \to G_1'$ is a homomorphism (not necessarily an isomorphism). Show that there is a unique homomorphism $\nu_2 : G_2 \to G_2'$ so that we have the commmutative diagram:

$$\begin{array}{ccc} G_1 & \xrightarrow{\nu_1} & G_1' \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \varphi'} \\ G_2 & \xrightarrow{\nu_2} & G_2' \end{array} \tag{3.25}$$

Exercise *Fiber Products*

Given groups G_1 and G_2, and homomorphisms $\psi_1 : G_1 \to H$ and $\psi_2 : G_2 \to H$ one can define a subset of $G_1 \times G_2$ known as a *fiber product*:

$$G_1 \times_{\psi_1,\psi_2} G_2 := \{(g_1, g_2) | \psi_1(g_1) = \psi_2(g_2)\} \ . \tag{3.26}$$

Show that the fiber product is in fact a subgroup of $G_1 \times G_2$, where $G_1 \times G_2$ has the direct product group structure.

---

4. Group Actions On Sets

4.1 Group Actions On Sets

Recall that we said that if X is any set then a *permutation* of X is a 1-1 and onto mapping $X \to X$. The set S_X of all permutations forms a group under composition.

We now define the notion of an action of a group on a set. This is a very important notion, and we will return to it extensively when discussing examples. If the following discussion seems too abstract the reader should consult section 8 for a number of concrete examples beyond the ones we are about to give. There are three ways to think about a group action on a set:

First Way: A *transformation group* on X is a subgroup of S_X.

Second Way: We define a *left G-action on a set X* to be a map $\phi : G \times X \to X$ compatible with the group multiplication law as follows:

$$\phi(g_1, \phi(g_2, x)) = \phi(g_1 g_2, x) \tag{4.1}$$

We would also like $x \mapsto \phi(1_G, x)$ to be the identity map. Now, equation (4.1) implies that

$$\phi(1_G, \phi(1_G, x)) = \phi(1_G, x) \tag{4.2}$$

which is compatible with, but does not quite imply that $\phi(1_G, x) = x$. Thus in defining a group action we must also impose the condition:

$$\phi(1_G, x) = x \qquad \forall x \in X. \tag{4.3}$$

Exercise
Give an example of a map $\phi : G \times X \to X$ that satisfies (4.1) but not (4.3). [15]

Third Way: Yet another way to say this is the following: Define the map $\Phi : G \to S_X$ that takes $g \mapsto \phi(g, \cdot)$. That is, for each $g \in G$, $\Phi(g)$ is the function $X \to X$ taking $x \mapsto \phi(g, x)$. Clearly $\Phi(g_1) \circ \Phi(g_2) = \Phi(g_1 g_2)$ because of (4.1). In order to make sure it is a permutation we need to know that $\Phi(g)$ is invertible and therefore we need to impose that $\Phi(1_G)$ is the identity transformation. This follows from (4.3). Then $\Phi(g) \in S_X$. So, to say we have a group action of G on X is to say that Φ is a homomorphism of G into the permutation group S_X. We will discuss G-actions on sets and their properties extensively in Chapter 3.

Definition: If X has a group action by a group G we say that X is a G-set.

Notation: Again, our notation is overly cumbersome because we want to stress the concept. Usually one writes a left G-action as

$$g \cdot x := \phi(g, x) \tag{4.4}$$

The key axioms become

$$\begin{aligned} g_1 \cdot (g_2 \cdot x) &= (g_1 g_2) \cdot x \\ 1_G \cdot x &= x \end{aligned} \tag{4.5}$$

We can think of $\Phi(g)$ as the map that sends $x \to g \cdot x$.

Definition/Discussion: Orbits: If G acts on a set X then we can define an equivalence relation on X by saying that two elements $x_1, x_2 \in X$ are equivalent, $x_1 \sim x_2$ if there is some $g \in G$ with $\phi(g, x_1) = x_2$. The reader should check that this is indeed an equivalence relation. The equivalence class $[x]$ with this equivalence relation is known as the *orbit of G through a point x*. So, concretely it is the set of points $y \in X$ which can be reached by the action of G:

$$O_G(x) = \{y : \exists g \quad \text{such that} \quad y = g \cdot x\} \tag{4.6}$$

[15] *Answer*: As the simplest example, choose any element $x_0 \in X$ and define $\phi(g, x) = x_0$ for all g, x. For a slightly less trivial example consider $G = S_2$ and let $\phi(e, x) = \phi(\sigma, x) = f(x)$. Then if $f \circ f(x) = f(x)$ the condition (4.1) will be satisfied, but there certainly exist functions with $f \circ f = f$ which are not the identity map.

The notion of orbits is very important in geometry, gauge theory and many other subjects. The *set of orbits* is denoted X/G. We will discuss many examples below.

Examples

1. Consider rotations around the origin of $\mathbb{R}^2$. They act on the points of $\mathbb{R}^2$ as a G-action. The distinct orbits are circles centered on the origin. The origin is also an orbit by itself.

2. Consider rotations around the origin by multiples of $2\pi/3$. Check that this group is isomorphic to $\mathbb{Z}_3$. Consider an equilateral triangle centered on the origin. Then the rotations act on the triangle preserving it. Thus, $\mathbb{Z}_3$ acts as a group of symmetries of the equilateral triangle. Intuitively, group theory is the theory of symmetry. We have just illustrated how that idea can be formalized through the notation of group action on a set. ♣FIGURE HERE!♣

3. Let $G = GL(n, \kappa)$ and $X = \kappa^n$, the n-dimensional vector space over κ. Then the usual linear action on vectors defines a group action of G on X. One can check that there are only two orbits: The zero vector gives one orbit.

4. If $G = \mathbb{Z}_2$ acts linearly on $\mathbb{R}^{n+1}$ (i.e. $V = \mathbb{R}^{n+1}$ is a representation of $\mathbb{Z}_2$) then we can choose coordinates so that the nontrivial element $\sigma \in G$ acts by

$$\sigma \cdot (x^1, \dots, x^{n+1}) = (x^1, \dots, x^p, -x^{p+1}, \cdots, -x^{p+q}) \tag{4.7}$$

   where $p + q = n + 1$. Note that this action preserves the equation of the sphere $\sum_i (x^i)^2 - 1 = 0$ and hence descends to a $\mathbb{Z}_2$-action on the sphere S^n. The case $p = 0, q = n+1$ is the antipodal map. The set of orbits is known as $\mathbb{RP}^n \cong S^n/\langle -1 \rangle$. There are many other natural actions of $\mathbb{Z}_2$ on S^n.

5. Let the group be $G = \mathbb{C}^*$. This acts on $X = \mathbb{C}^n$ by scaling all the coordinates. Note that scaling a nonzero vector by a nonzero scalar gives a nonzero vector so $G = \mathbb{C}^*$ also restricts to act on $\tilde{X} = \mathbb{C}^n - \{0\}$. The set of orbits in these two cases are very different. One can put a natural topology on the set of orbits. One finds that $\mathbb{C}^n/\mathbb{C}^*$ is not a Hausdorff space while $(\mathbb{C}^n - \{0\})/\mathbb{C}^*$ is a nice manifold. This important manifold is often denoted $\mathbb{CP}^{n-1}$. See Chapter 3 for more discussion. We will often denote elements of $\mathbb{CP}^{n-1}$ by $[X^1 : \cdots : X^n]$. This stands for the equivalence class of a vector $(X^1, \dots, X^n) \in \mathbb{C}^n - \{0\}$.

6. Now consider a set of integers $(q_1, \dots, q_n) \in \mathbb{Z}^n$. Then for each such set of integers there is a $\mathbb{C}^*$-action on $\mathbb{CP}^{n-1}$ defined by

$$\mu \cdot [X^1 : \cdots : X^n] := [\mu^{q_1} X^1 : \cdots : \mu^{q_n} X^n] \tag{4.8}$$

   for $\mu \in \mathbb{C}^*$. (Check it is well-defined!)

7. The group $G = SL(2,\mathbb{R})$ acts on the complex upper half plane:

$$\mathcal{H} = \{\tau | \mathrm{Im}\tau > 0\} \tag{4.9}$$

via

$$g \cdot \tau := \frac{a\tau + b}{c\tau + d} \tag{4.10}$$

where

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{4.11}$$

The reader should check that indeed:

$$g_1 \cdot (g_2 \cdot \tau) = (g_1 g_2) \cdot \tau \tag{4.12}$$

8. Consider the action of $\mathbb{Z}$ on $\mathbb{R}$ where $n : x \mapsto x + n$. The orbit of a real number r is $r + \mathbb{Z}$. Note that the value of the function $p(x) := e^{2\pi \mathrm{i} x}$ uniquely determines an orbit. So we can identify the space of orbits $X/G = \mathbb{R}/\mathbb{Z}$ with this action with the circle.

9. *Actions of $\mathbb{Z}$ on <u>any</u> set X.* Let us consider $\mathbb{Z}$ to be the free group with one generator g_0. Then, given <u>any</u> invertible map $f : X \to X$ we can define a group action of $\mathbb{Z}$ on X by

$$g_0^n \cdot x = \begin{cases} \underbrace{f \circ \cdots \circ f}_{n \quad \text{times}}(x) & n > 0 \\ x & n = 0 \\ \underbrace{f^{-1} \circ \cdots \circ f^{-1}}_{|n| \quad \text{times}}(x) & n < 0 \end{cases} \tag{4.13}$$

Conversely, any $\mathbb{Z}$-action must be of this form since we can define $f(x) := g_0 \cdot x$. Thus the orbit of a point $x \in X$ is the set of all images of successive actions of f and f^{-1}.

10. Let $\mathbb{M}^{1,d}$ be $(d+1)$-dimensional Minkowski space and consider an electromagnetic gauge potential A_μ to be a map $A_\mu : \mathbb{M}^{1,d} \to \mathbb{R}^{d+1}$. (This is geometrically inaccurate because A_μ is an example of what is called a connection, but it will suffice for our present purposes.) Then the group $\mathcal{G} = Map[\mathbb{M}^{1,d} \to U(1)]$ is a group which acts on the set $\mathcal{A}$ of all gauge potentials by $A_\mu \to A_\mu - \mathrm{i} f^{-1} \partial_\mu f$. The space of orbits $\mathcal{A}/\mathcal{G}$ parametrizes gauge-inequivalent field configurations.

****

SHOULD DEFINE A MAP, OR MORPHISM, OF G-SPACES HERE. INTRODUCE THE GENERAL IDEA OF AN EQUIVARIANT MAP.

GIVE SOME EXAMPLES.

****

4.2 Group Actions On Sets Induce Group Actions On Associated Function Spaces

The following general abstract idea is of great importance in both mathematics and physics: Suppose X and Y are any two sets and $\mathcal{F}[X \to Y]$ is the set of functions from X to Y. Now suppose that there is a left G-action on X defined by $\phi : G \times X \to X$. Then, automatically, there is also a G action $\tilde{\phi}$ on $\mathcal{F}[X \to Y]$. To define it, suppose $F \in \mathcal{F}[X \to Y]$ and $g \in G$. Then we need to define $\tilde{\phi}(g, F) \in \mathcal{F}[X \to Y]$. We do this by setting $\tilde{\phi}(g, F)$ to be that specific function whose values are defined by:

$$\tilde{\phi}(g, F)(x) := F(\phi(g^{-1}, x)). \tag{4.14}$$

Note the inverse of g on the RHS. It is there so that the group law works out:

$$\begin{aligned} \tilde{\phi}(g_1, \tilde{\phi}(g_2, F))(x) &= \tilde{\phi}(g_2, F)(\phi(g_1^{-1}, x)) \\ &= F(\phi(g_2^{-1}, \phi(g_1^{-1}, x)) \\ &= F(\phi(g_2^{-1} g_1^{-1}, x)) \\ &= F(\phi((g_1 g_2)^{-1}, x)) \\ &= \tilde{\phi}(g_1 g_2, F)(x) \end{aligned} \tag{4.15}$$

and hence $\tilde{\phi}(g_1, \tilde{\phi}(g_2, F)) = \tilde{\phi}(g_1 g_2, F)$ as required for a group action. It should also be clear that $\tilde{\phi}(1_G, F) = F$.

In less cumbersome notation we would simply write

$$(g \cdot F)(x) := F(g^{-1} \cdot x) \tag{4.16}$$

In the above discussion we could impose various conditions, on the functions in $\mathcal{F}[X \to Y]$. For example, if X and Y are manifolds we could ask our maps to be continuous, differentiable, etc. The above discussion would be unchanged.

As just one (important) example of this general idea: In field theory if we have fields on a spacetime, and a group of symmetries acting on that spacetime then that group also acts on the space of fields.

Exercise *When Y Is A G-Set*

Suppose there is a left G-action on a set Y and X is any set. Show that there is a natural left G-action on $\mathcal{F}[X \to Y]$.

---

5. The Symmetric Group.

The symmetric group is an important example of a finite group. As we shall soon see, all finite groups are isomorphic to subgroups of the symmetric group.

Recall from section 2 above that for any set X we can define a group S_X of all permutations of the set X. If n is a positive integer the symmetric group on n elements, denoted S_n, is defined as the group of permutations of the set $X = \{1, 2, \ldots, n\}$.

In group theory, as in politics, there are leftists and rightists and we can actually define *two* group operations:

$$\begin{aligned}(\phi_1 \cdot_L \phi_2)(i) &:= \phi_2(\phi_1(i)) \\ (\phi_1 \cdot_R \phi_2)(i) &:= \phi_1(\phi_2(i))\end{aligned} \tag{5.1}$$

That is, with $\cdot_L$ we read the operations from left to right and first apply the left permutation, and then the right permutation. Etc. Each convention has its own advantages and both are frequently used.

In these notes we will adopt the $\cdot_R$ convention and henceforth simply write $\phi_1\phi_2$ for the product.

We can write a permutation symbolically as

$$\phi = \begin{pmatrix} 1 & 2 & \cdots & n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix} \tag{5.2}$$

meaning: $\phi(1) = p_1, \phi(2) = p_2, \ldots, \phi(n) = p_n$. Note that we could equally well write the same permutation as:

$$\phi = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p_{a_1} & p_{a_2} & \cdots & p_{a_n} \end{pmatrix} \tag{5.3}$$

where $a_1, \ldots, a_n$ is any permutation of $1, \ldots, n$. With this understood, suppose we want to compute $\phi_1 \cdot_L \phi_2$. We should first see what ϕ_1 does to the ordered elements $1, \ldots, n$, and then see what ϕ_2 does to the ordered output from ϕ_1. So, if we write:

$$\begin{aligned}\phi_1 &= \begin{pmatrix} 1 & \cdots & n \\ q_1 & \cdots & q_n \end{pmatrix} \\ \phi_2 &= \begin{pmatrix} q_1 & \cdots & q_n \\ p_1 & \cdots & p_n \end{pmatrix}\end{aligned} \tag{5.4}$$

Then

$$\phi_1 \cdot_L \phi_2 = \begin{pmatrix} 1 & \cdots & n \\ p_1 & \cdots & p_n \end{pmatrix} \tag{5.5}$$

On the other hand, to compute $\phi_1 \cdot_R \phi_2$ we should first see what ϕ_2 does to $1, \ldots, n$ and then see what ϕ_1 does to that output. We could write represent this as:

$$\begin{aligned}\phi_2 &= \begin{pmatrix} 1 & \cdots & n \\ q'_1 & \cdots & q'_n \end{pmatrix} \\ \phi_1 &= \begin{pmatrix} q'_1 & \cdots & q'_n \\ p'_1 & \cdots & p'_n \end{pmatrix}\end{aligned} \tag{5.6}$$

and then

$$\phi_1 \cdot_R \phi_2 = \begin{pmatrix} 1 & \cdots & n \\ p'_1 & \cdots & p'_n \end{pmatrix} \tag{5.7}$$

Exercise

a.) Show that the order of the group is $|S_n| = n!$.

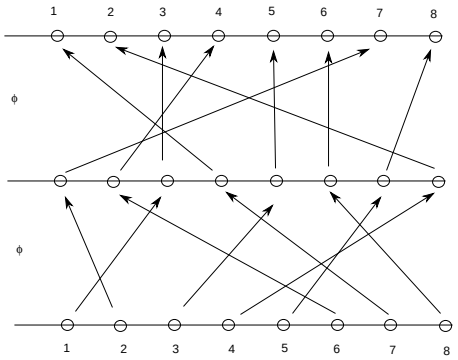
b.) Show that if $n_1 \leq n_2$ then we can consider S_{n_1} as a subgroup of S_{n_2}.

c.) In how many ways can you consider S_2 to be a subgroup of S_3? [16]

d.) In how many ways can you consider S_{n_1} to be a subgroup of S_{n_2} when $n_1 \leq n_2$? [17]

Exercise Show that the inverse of (5.2) is the permutation:

$$\phi = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ 1 & 2 & \cdots & n \end{pmatrix} \tag{5.8}$$



1

Figure 1: A pictorial view of the composition of two permutations ϕ_1, ϕ_2 in S_8. Thus $1 \to 3, 2 \to 7$ etc. for the group product $\phi_2 \cdot \phi_1$.

It is often useful to visualize a permutation in terms of "time evolution" (going up) as shown in 1.

[16] *Answer*: There are three subgroups of S_3 isomorphic to S_2. They are the subgroups that fix $1, 2, 3$ respectively.

[17] *Answer*: for any subset $T \subset \{1, \ldots, n_2\}$ of cardinality $n_2 - n_1$ we can consider the subset of permutations that fix all elements of T. This subset of permutations will be a subgroup isomorphic to S_{n_1}. So there are $\binom{n_2}{n_1}$ distinct subgroups isomorphic to S_{n_1}.

Exercise *Left versus right*

a.) Show that in the pictorial interpretation the inverse is obtained by running arrows backwards in time.

b.) Show that the left- and right- group operation conventions are related by

$$\phi_1 \cdot_L \phi_2 = (\phi_1^{-1} \cdot_R \phi_2^{-1})^{-1} \tag{5.9}$$

c.) Interpret (5.9) as the simple statement that $\phi_1 \cdot_R \phi_2$ puts ϕ_2 in the past while $\phi_1 \cdot_L \phi_2$ puts ϕ_1 in the past.

5.1 Cayley's Theorem

As a nice illustration of some of the concepts we have introduced we now prove Cayley's theorem. This theorem states that *any* finite group is isomorphic to a subgroup of a permutation group S_N for some N.

To prove this we begin with an elementary, but important, observation known as the

The rearrangement lemma: Consider a totally ordered group, that is, we can list the group elements in order

$$G = \{g_1, g_2, \dots, \} \tag{5.10}$$

Consider this as an ordered set, with all the g_i distinct. The set can be finite or infinite. Then, for any $h \in G$ consider the ordered set:

$$h \cdot G = \{h \cdot g_1, h \cdot g_2, \dots, \}. \tag{5.11}$$

With a little thought you can (and should) convince yourself that (5.11) is a a permutation of (5.10): No two elements coincide (since the $g_i \neq g_j$ for $i \neq j$) and every element of G must appear in the list $h \cdot G$.

To put this differently, there is a left-G-action of G on itself: For $h \in G$, define the map $L(h) : G \to G$ by the rule:

$$L(h) : g \mapsto h \cdot g \qquad\qquad \forall g \in G. \tag{5.12}$$

This map is one-one and invertible so $L(h) \in S_G$, the group of permutations of the set G. (In fact, there is no need to assume G is totally ordered.) Now note that

$$L(h_1) \circ L(h_2) = L(h_1 \cdot h_2) \tag{5.13}$$

so the map $\mathcal{L}$ defined by $\mathcal{L} : h \mapsto L(h)$ is a homomorphism

$$\mathcal{L} : G \to S_G \tag{5.14}$$

This is an example of a group action on a set. In this case $X = G$ and G is acting on itself by left-multiplication and $\mathcal{L}$ is the quantity denoted by Φ above. Furthermore, if $L(h_1) = L(h_2)$ then $h_1 = h_2$. Therefore $\mathcal{L}$ is an isomorphism of G with its image in S_G.

The above remarks apply to any group. However, now consider any finite group G with $N = |G|$ then S_G is isomorphic to S_N. Therefore, any finite group is isomorphic to a subgroup of a symmetric group S_N for some N. This is Cayley's theorem. Note that which subgroup of S_N we obtain depends on how we choose to order G, that is, it depends on the choice of isomorphism $S_G \cong S_N$.

Exercise *Concrete Example*

By Cayley's theorem the cyclic group $\mathbb{Z}_n$ of order n is isomorphic to a subgroup of a permutation group. Exhibit such an isomorphic subgoup. [18]

Exercise *Right Action*

♣This is redundant with some material on group actions below. ♣

There are other ways G can act on itself. For example we can define

$$R(a) : g \mapsto g \cdot a \tag{5.15}$$

a.) Show that $R(a)$ permutes the elements of G.

b.) Show that $R(a_1) \circ R(a_2) = R(a_2 a_1)$. Thus, $a \mapsto R(a)$ is not a homomorphism of G into the group S_G of permutations of G.

c.) Show that $a \mapsto R(a^{-1})$ is a homomorphism of G into S_G.

5.2 Cyclic Permutations And Cycle Decomposition

A very important class of permutations are the *cyclic permutations of length* ℓ. Choose ℓ distinct numbers, $a_1, \ldots, a_\ell$ between 1 and n and permute:

$$a_1 \to a_2 \to \cdots \to a_\ell \to a_1 \tag{5.16}$$

holding all other $n - \ell$ elements fixed. Such a permutation is called *a cycle of length* ℓ. We will denote such permutations as:

$$\phi = (a_1 a_2 \ldots a_\ell). \tag{5.17}$$

Bear in mind that with this notation the same permutation can be written in ℓ different ways:

$$(a_1 a_2 \ldots a_\ell) = (a_2 a_3 \ldots a_\ell a_1) = (a_3 \ldots a_\ell a_1 a_2) = \cdots = (a_\ell a_1 a_2 \ldots a_{\ell-1}) \tag{5.18}$$

[18] *Answer*: Choose any cyclic permutation of length n (Cyclic permutations are defined in Section 5.2 below.) Then it generates a subgroup of S_N of length n for any $N \geq n$.

Let us write out the elements of the first few symmetric groups in this notation:

$$S_2 = \{1, (12)\} \tag{5.19}$$

$$S_3 = \{1, (12), (13), (23), (123), (132)\} \tag{5.20}$$

Remarks

1. S_2 is abelian.
2. S_3 is NOT ABELIAN [19]

$$\begin{aligned} (12) \cdot (13) &= (132) \\ (13) \cdot (12) &= (123) \end{aligned} \tag{5.21}$$

and therefore so is S_n for $n > 2$.

It is not true that all permutations are just cyclic permutations, as we first see by considering S_4:

$$\begin{aligned} S_4 = \{1, (12), (13), (14), &(23), (24), (34), (12)(34), (13)(24), (14)(23), \\ (123), (132), (124), (142), &(134), (143), (234), (243) \\ &(1234), (1243), (1324), (1342), (1423), (1432)\} \end{aligned} \tag{5.22}$$

Now a key observation is:

Any permutation $\sigma \in S_n$ can be uniquely written as a product of disjoint cycles. This is called the cycle decomposition of σ.

For example

$$\sigma = (12)(34)(10, 11)(56789) \tag{5.23}$$

is a cycle decomposition in S_{11}. There are 3 cycles of length 2 and 1 of length 5.

The decomposition into products of disjoint cycles is known as the *cycle decomposition.*

Exercise *Decomposition as a product of disjoint cyclic permutations*

Prove the above claim: every permutation above is a product of cyclic permutations on disjoint sets of integers. [20]

---

[19] Note that $(12) \cdot_L (13) = (123)$. But we use the $\cdot_R$ convention.

[20] *Answer*: Use induction: Consider any element, say $x \in \{1, \ldots, n\}$ and let ϕ be a permutation. Consider the elements $x, \phi(x), \phi(\phi(x)), \ldots$. This must be a finite set C, so we get a cyclic permutation of the elements in C. Then ϕ must permute all the elements in $\{1, \ldots, n\} - C$. But this has cardinality strictly smaller than n. So, use the inductive hypothesis.

Exercise

a.) Let ϕ be a cyclic permutation of order ℓ. Suppose we compose ϕ with itself N times. Show that the result is the identity transformation iff ℓ divides N.

b.) Suppose ϕ has a cycle decomposition with cycles of length $k_1, \ldots, k_s$. What is the smallest number N so that if we compose ϕ with itself $\phi \circ \cdots \circ \phi$ for N times that we get the identity transformation?

5.3 Transpositions

A *transposition* is a permutation of the form: (ij). These satisfy some nice properties: Suppose i, j, k are distinct. You can check as an exercise that transpositions obey the following identities:

$$\begin{aligned} (ij)\cdot(jk)\cdot(ij) = (ik) = (jk)\cdot(ij)\cdot(jk) \\ (ij)^2 = 1 \\ (ij)\cdot(kl) = (kl)\cdot(ij) \qquad \{i,j\}\cap\{k,l\} = \emptyset \end{aligned} \tag{5.24}$$

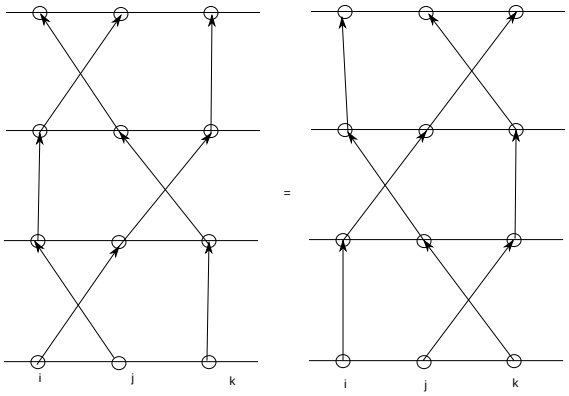


Figure 2: Pictorial illustration of equation (4.21) line one for transpositions where $i < j < k$. Note that the identity is suggested by "moving the time lines" holding the endpoints fixed. Reading time from bottom to top corresponds to reading the composition from left to right in the $\cdot_R$ convention.

The first identity is illustrated in Figure 2. Draw the other two.

We observed above that there is a cycle decomposition of permutations. Now note that

Any cycle $(a_1, \cdots, a_k)$ *can be written as a product of transpositions.* To prove this note that

$$(1,k)(1,k-1)\cdots(1,4)(1,3)(1,2) = (1,2,3,4,\ldots,k) \tag{5.25}$$

Now, consider a permutation that takes

$$1 \to a_1, \quad 2 \to a_2, \quad 3 \to a_3, \cdots, k \to a_k \tag{5.26}$$

For our purposes, it won't really matter what it does to the other integers greater than k. Choose any such permutation and call it ϕ. Note that

$$\phi \circ (1\ 2\ \cdots\ k) \circ \phi^{-1} = (a_1\ a_2\ \cdots\ a_k) \tag{5.27}$$

so now multiply the above identity by ϕ on the left and ϕ^{-1} on the right to get:

$$\phi(1,k)\phi^{-1}\phi(1,k-1)\phi^{-1}\cdots\phi(1,4)\phi^{-1}\phi(1,3)\phi^{-1}\phi(1,2)\phi^{-1} = (a_1, a_2, \ldots, a_k) \tag{5.28}$$

but $\phi(1,j)\phi^{-1} = (a_1, a_j)$. So we get a decomposition of $(a_1\ a_2\ \cdots\ a_k)$ as a product of transpositions.

In general, a group element of the form ghg^{-1} is called a *conjugate of* h. See Section 7.2 below.

Therefore, *every element of* S_n *can be written as a product of transpositions, generalizing* (5.21). We say that the transpositions *generate* the permutation group. Taking products of various transpositions – what we might call a "word" whose "letters" are the transpositions – we can produce any element of the symmetric group. We will return to this notion in §6 below.

Of course, a given permutation can be written as a product of transpositions in many ways. This clearly follows because of the identities (5.24). A nontrivial fact is that the transpositions together with the above relations generate precisely the symmetric group. [21] It therefore follows that all possible nontrivial identities made out of transpositions follow from repeated use of these identities.

Although permutations can be written as products of transpositions in different ways, the number of transpositions in a word *modulo 2* is always the same, because the identities (5.24) have the same number of transpositions, modulo two, on the LHS and RHS. Thus we can define *even, resp. odd, permutations* to be products of even, resp. odd numbers of transpositions.

Definition: The *alternating group* $A_n \subset S_n$ is the subgroup of S_n of even permutations.

Exercise

a.) What is the order of A_n ? [22]

b.) Write out A_2, A_3, and A_4. Show that A_3 is isomorphic to $\mathbb{Z}_3$. [23]

[21] This follows once one has shown that the Coxeter presentation given below gives precisely the symmetric group, and not some larger group (requiring the imposition of further relations) since the above relations all follow from the Coxeter relations.

[22] *Answer*: $\frac{1}{2}n!$ for $n > 1$. To prove this note that the transformation $\phi \to \phi \circ (12)$ is an invertible transformation $S_n \to S_n$ that squares to the identity. On the other hand, it exchanges even and odd permutations.

[23] *Answer*: $A_2 = \{1\}$. $A_3 = \{1, (123), (132)\}$.

$$A_4 = \{1, (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (14)(23)\}.$$

c.) A_3 is Abelian. Is A_4 Abelian? [24]

Exercise

When do two cyclic permutations commute? Illustrate the answer with pictures, as above.

Exercise *A Smaller Set Of Generators*

Show that from the transpositions $\sigma_i := (i, i+1)$, $1 \leq i \leq n-1$ we can generate all other transpositions in S_n. These are sometimes called the elementary generators.

Exercise *An Even Smaller Set Of Generators*

Show that, in fact, S_n can be generated by just two elements: (12) and $(1\ 2\ \cdots\ n)$. [25]

Exercise *Center of S_n*

What is the center of S_n? [26]

Exercise *Decomposing the reverse shuffle*

Consider the permutation which takes $1, 2, \ldots, n$ to $n, n-1, \ldots, 1$.

a.) Write the cycle decomposition.

b.) Write a decomposition of this permutation in terms of the *elementary generators* σ_i. [27]

[24] *Answer*: No. Just multiply a few elements to find a counterexample. For example $(123)(134) = (234)$ but $(134)(123) = (124)$.

[25] *Answer*: Conjugate (12) by the n-cycle to get (23). Then conjugate again to get (34) and so forth. Now we have the set of generators of the previous exercise.

[26] *Answer*: If $n = 2$ then S_n is Abelian and the center is all of S_2. If $n > 2$ then the center is the trivial group. To prove this suppose $z \in Z(S_n)$. If z is not the trivial element then it moves some i to some j. WLOG we can say it moves 1 to $i \neq 1$. Then $z(i) \neq i$. If $z(i) = 1$ then z is the transposition $(1, i)$. If $n > 2$ there will be some other $j \neq 1, i$ and z will not commute with $(1, j)$. If $z(i) = j$ with $j \neq 1, i$ then $\phi = (1, i)$ does not commute with z because $z\phi$ takes $1 \to j$ and ϕz takes $1 \to 1$.

[27] *Hint*: Use the pictorial interpretation mentioned above.

Example 3.2 *The sign homomorphism.*

This is a very important example of a homomorphism:

$$\epsilon : S_n \to \mathbb{Z}_2 \tag{5.29}$$

where we identify $\mathbb{Z}_2$ as the multiplicative group $\{\pm 1\}$ of square roots of 1. The rule is:

$\epsilon : \sigma \to +1$ if σ is a product of an *even* number of transpositions.

$\epsilon : \sigma \to -1$ if σ is a product of an *odd* number of transpositions.

Put differently, we could define $\epsilon(ij) = -1$ for any transposition. This is compatible with the words defining the relations on transpositions. Since the transpositions generate the group the homomorphism is well-defined and completely determined.

In physics one often encounters the sign homomorphism in the guise of the "epsilon tensor" denoted:

$$\epsilon_{i_1 \cdots i_n} \tag{5.30}$$

Its value is:

1. $\epsilon_{i_1 \cdots i_n} = +1$ if
$$\begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \tag{5.31}$$
is an even permutation.

2. $\epsilon_{i_1 \cdots i_n} = -1$ if
$$\begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \tag{5.32}$$
is an odd permutation.

3. $\epsilon_{i_1 \cdots i_n} = 0$ if two indices are repeated. (This goes a bit beyond what we said above since in that case we are not discussing a permutation.)

So, e.g. among the 27 entries of ϵ_{ijk}, $1 \le i, j, k \le 3$ we have

$$\begin{aligned} \epsilon_{123} &= 1 \\ \epsilon_{132} &= -1 \\ \epsilon_{231} &= +1 \\ \epsilon_{221} &= 0 \end{aligned} \tag{5.33}$$

and so forth.

Exercise

Show that

$$\epsilon_{i_1 i_2 \cdots i_n} \epsilon_{j_1 j_2 \cdots j_n} = \sum_{\sigma \in S_n} \epsilon(\sigma) \delta_{i_1 j_{\sigma(1)}} \delta_{i_2 j_{\sigma(2)}} \cdots \delta_{i_n j_{\sigma(n)}} \tag{5.34}$$

This formula is often useful when proving identities involving determinants. An important special case occurs for $n = 3$ where it is equivalent to the rule for the cross-product of 3 vectors in $\mathbb{R}^3$:

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B}) \tag{5.35}$$

The next two exercises assume some familiarity with concepts from linear algebra. See Chapter 2 below if they are not familiar.

Exercise *The Canonical Permutation Representation Of S_n*

Consider the standard Euclidean vector space $\mathbb{R}^n$ (or $\mathbb{C}^n$ or κ^n) with basis vectors $\vec{e}_1, \ldots, \vec{e}_n$ where $\vec{e}_i$ has component 1 in the ith position and zero else. Note that the symmetric group permutes these vectors in an obvious way:

$$T(\phi) : \vec{e}_i \to \vec{e}_{\phi(i)} \ , \tag{5.36}$$

and now extend by linearity so that

$$T(\phi) : \sum_{i=1}^{n} x_i e_i \mapsto \sum_{i=1}^{n} x_i e_{\phi(i)} = \sum_{i=1}^{n} x_{\phi^{-1}(i)} e_i \tag{5.37}$$

Thus to any permutation $\phi \in S_n$ we can associate a linear transformation $T(\phi)$. The matrix $A(\phi)$ of $T(\phi)$ defined by $T(\phi)$ and the ordered basis $\{e_1, \ldots, e_n\}$ is defined by:

$$T(\phi)\vec{e}_i = \sum_{j=1}^{n} A(\phi)_{ji} \vec{e}_j \tag{5.38}$$

a.) Write out $A(\phi)$ for small values of n and some simple permutations ϕ.

b.) Write a general formula for the matrix elements of $A(\phi)$. [28]

c.) Show that $\phi \to A(\phi)$ is a matrix representation of S_n.

d.) The matrices $A(\phi)$ are called *permutation matrices*. In each row and column there is only one nonzero matrix element, and that nonzero element is 1. If B is any other $n \times n$ matrix show that

$$\left(A(\phi) B A(\phi)^{-1}\right)_{i,j} = B_{\phi(i),\phi(j)} \tag{5.39}$$

In general, if we have a representation of a group $T : G \to GL(V)$ and a nontrivial subspace $W \subset V$ such that $T(g)$ takes vectors in W to vectors in W for all $g \in G$, we say that the representation is *reducible*.

[28] *Answer*: $A(\phi)_{i,j} = \delta_{i,\phi(j)} = \delta_{\phi^{-1}(i),j}$.

SEE SECTION ***** BELOW FOR MORE

e.) Show that the natural permutation representation of S_n on $\mathbb{R}^n$ is reducible. [29]

Exercise *Signed Permutation Matrices*

Define *signed permutation matrices* to be invertible matrices such that in each row and column there is only one nonzero matrix element, and the nonzero matrix element can be either $+1$ or -1. Finally, require the matrix to be invertible.

a.) Show that the set of $n \times n$ signed permutation matrices form a group. We will call it $W(B_n)$ for reasons that will not be obvious for a while.

b.) Define a group homomorphism $W(B_n) \to S_n$.

5.4 Diversion and Example: Card shuffling

One way we commonly encounter permutation groups is in shuffling a deck of cards.

A deck of cards is equivalent to an ordered set of 52 elements. Some aspects of card shuffling and card tricks can be understood nicely in terms of group theory.

Mathematicians often use the *perfect shuffle* or the *Faro shuffle*. Suppose we have a deck of $2n$ cards, so $n = 26$ is the usual case. There are actually two kinds of perfect shuffles: the In-shuffle and the Out-shuffle.

In either case we begin by splitting the deck into two equal parts, and then we interleave the two parts perfectly.

Let us call the top half of the deck the left half-deck and the bottom half of the deck the right half-deck. Then, to define the *Out-shuffle* we put the top card of the left deck on top, followed by the top card of the right deck underneath, and then proceed to interleave them perfectly. The bottom and top cards stay the same.

If we number the cards $0, 1, \ldots, 2n-1$ from top to bottom then the top (i.e. left) half-deck consists of the cards numbered $0, 1, \ldots, n-1$ while the bottom (i.e. right) half-deck consists of the cards $n, n+1, \ldots, 2n-1$. Then the Out-shuffle gives the cards in the new order

$$0, n, 1, n+1, 2, n+2, \ldots, n+2, 2n-2, n-1, 2n-1 \tag{5.40}$$

Another way to express this is that the Out-shuffle defines a permutation of $\{0, 1, \ldots, 2n-1\}$. If we let C_x, $0 \le x \le 2n-1$ denote the cards in the original order then the new ordered set of cards C'_x are related to the old ones by:

$$C'_{\mathcal{O}(x)} = C_x \tag{5.41}$$

where

$$\mathcal{O}(x) = \begin{cases} 2x & x \le n-1 \\ 2x-(2n-1) & n \le x \le 2n-1 \end{cases} \tag{5.42}$$

[29] *Answer*: Show that the linear subspace spanned by the "all ones vector" $v_0 = e_1 + \cdots + e_n$ is preserved under the action of $T(\phi)$ for all $\phi \in S_n$: $T(\phi)(\lambda v_0) = \lambda v_0$, for all $\lambda \in \mathbb{R}$.

Note that this already leads to a card trick: Modulo $(2n-1)$ the operation is just $x \to 2x$, so if k is the smallest number with $2^k = 1 \mathrm{mod}(2n-1)$ then k Out-shuffles will restore the deck perfectly.

For example: For a standard deck of 52 cards, $2^8 = 5 \times 51 + 1$ so 8 perfect Out-shuffles restores the deck!

We can also see this by working out the cycle presentation of the Out-shuffle:

$$\begin{aligned}\mathcal{O} = &(0)(1,2,4,8,16,32,13,26)(3,6,12,24,48,45,39,27)\\ &(5,10,20,40,29,7,14,28)(9,18,36,21,42,33,15,30)\\ &(11,22,44,37,23,46,41,31)(17,34)(19,38,25,50,49,47,43,35)(51)\end{aligned} \tag{5.43}$$

Clearly, the 8th power gives the identity permutation.

Now, to define the *In-shuffle* we put the top card of the right half-deck on top, then the top card of the left half-deck underneath, and then proceed to interleave them.

Now observe that if we have a deck with $2n$ cards $\mathcal{D}(2n) := \{0, 1, \ldots, 2n-1\}$ and we embed it in a Deck with $2n+2$ cards

$$\mathcal{D}(2n) \to \mathcal{D}(2n+2) \tag{5.44}$$

by the map $x \to x+1$ then *the Out-shuffle on the deck* $\mathcal{D}(2n+2)$ *permutes the cards* $1, \ldots, 2n$ *amongst themselves and acts as an In-shuffle on these cards!*

♣Explain this some more, e.g. by illustrating with a pack of 6 cards. ♣

Therefore, applying our formula for the Out-shuffle we find that the In-shuffle is given by the formula

$$\mathcal{I}(x) = \begin{cases} 2(x+1) - 1 & x+1 \leq n \\ 2(x+1) - (2n+1) - 1 & n \leq x \leq 2n-1 \end{cases} \tag{5.45}$$

One can check that this is given by the uniform formula

$$\mathcal{I}(x) = (2x+1) \,\mathrm{mod}(2n+1) \tag{5.46}$$

for $x \in \mathcal{D}(2n)$.

For $2n = 52$ this turns out to be one big cycle!

$$\begin{aligned}&(0,1,3,7,15,31,10,21,43,34,16,33,14,29,6,13,27,2,5,\\ &11,23,47,42,32,12,25,51,50,48,44,36,20,41,30,8,17,\\ &35,18,37,22,45,38,24,49,46,40,28,4,9,19,26)\end{aligned} \tag{5.47}$$

so it takes 52 consecutive perfect In-shuffles to restore the deck.

One can do further magic tricks with In- and Out-shuffles. As one example there is a simple prescription for bringing the top card to any desired position, say, position ℓ by doing In- and Out-shuffles.

To do this we write ℓ in its binary expansion:

$$\ell = 2^k + a_{k-1}2^{k-1} + \cdots + a_1 2^1 + a_0 \tag{5.48}$$

where $a_j \in \{0,1\}$. Interpret the coefficients 1 as In-shuffles and the coefficients 0 as Out-shuffles. Then, reading from left to right, perform the sequence of shuffles given by the binary expression: $1a_{k-1}a_{k-2}\cdots a_1a_0$.

To see why this is true consider iterating the functions $o(x) = 2x$ and $i(x) = 2x+1$. Notice that the sequence of operations given by the binary expansion of ℓ are

$$\begin{aligned} 0 &\to 1 \\ &\to 2\cdot 1 + a_{k-1} \\ &\to 2\cdot(2\cdot 1 + a_{k-1}) + a_{k-2} = 2^2 + 2a_{k-1} + a_{k-2} \\ &\to 2\cdot(2^2 + 2a_{k-1} + a_{k-2}) + a_{k-3} = 2^3 + 2^2 a_{k-1} + 2a_{k-2} + a_{k-3} \\ \vdots &\quad \vdots \\ &\to 2^k + a_{k-1}2^{k-1} + \cdots + a_1 2^1 + a_0 = \ell \end{aligned} \tag{5.49}$$

For an even ordered set we can define a notion of permutations preserving *central symmetry*. For $x \in D_{2n}$ let $\bar{x} = 2n-1-x$. Then we define the group $W(B_n) \subset S_{2n}$ to be the subgroup of permutations which permutes the pairs $\{x, \bar{x}\}$ amongst themselves.

Note that there is clearly a homomorphism

$$\phi : W(B_n) \to S_n \tag{5.50}$$

Moreover, both $\mathcal{O}$ and $\mathcal{I}$ are elements of $W(B_n)$. Therefore the *shuffle group*, the group generated by these is a subgroup of $W(B_n)$. Using this one can say some nice things about the structure of the group generated by the in-shuffle and the out-shuffle. It was completely determined in a beautiful paper (the source of the above material):

"The mathematics of perfect shuffles," P. Diaconis, R.L. Graham, W.M. Kantor, Adv. Appl. Math. **4** pp. 175-193 (1983)

It turns out that shuffles of decks of 12 and 24 cards have some special properties. In particular, special shuffles of a deck of 12 cards can be used to generate a very interesting group known as the Mathieu group M_{12}. It was, historically, the first "sporadic" finite simple group. See section §16.4 below.

To describe M_{12} we need to introduce a *Mongean shuffle*. Here we take the deck of cards put the top card on the right. Then from the deck on the left alternatively put cards on the top or the bottom. So the second card from of the deck on the left goes on top of the first card, the third card from the deck on the left goes under the first card, and so on. If we label our deck as cards $1, 2, \ldots, 2n$ then the Mongean shuffle is:

$$m : \{1,2,\ldots,2n\} \to \{2n, 2n-2, \ldots, 4, 2, 1, 3, 5, \ldots, 2n-3, 2n-1\} \tag{5.51}$$

In formulae, acting on $\mathcal{D}(2n)$

$$m(x) = \mathrm{Min}[2x, 2n+1-2x] \tag{5.52}$$

In particular for $2n = 12$ we have

$$\{1,2,3,4,5,6,7,8,9,10,11,12\} \to \{12,10,8,6,4,2,1,3,5,7,9,11\} \tag{5.53}$$

which has cycle decomposition (check!)

$$(3\ 8)\cdot(1\ 12\ 11\ 9\ 5\ 4\ 6\ 2\ 10\ 7) \tag{5.54}$$

Now consider the *reverse shuffle* that simply orders the cards backwards. In general for a deck $\mathcal{D}(2n)$ with $n = 2\text{mod}4$ Diaconis et. al. show that r and m generate the entire symmetric group. However, for a pack of 12 cards r and m generate the Mathieu group M_{12}. It turns out to have order

$$|M_{12}| = 2^6 \cdot 3^3 \cdot 5 \cdot 11 = 95040 \tag{5.55}$$

Compare this with the order of S_{12}:

$$12! = 2^{10} \cdot 3^5 \cdot 5^2 \cdot 7 \cdot 11 = 479001600 \tag{5.56}$$

So with the uniform probability distribution on S_{12}, the probability of finding a Mathieu permutation is $\frac{1}{5040} \sim 2 \times 10^{-4}$.

We mention some final loosely related facts:

1. There are indications that the Mathieu groups have some intriguing relations to string theory, conformal field theory, and K3 surfaces.

2. In the theory of L_∞ algebras and associated topics, which are closely related to string field theory one encounters the concept of the k-shuffle...

   FILL IN.

Exercise *Cycle structure for the Mongean shuffle*

Write the cycle structure for the Mongean shuffle of a deck with 52 cards. How many Mongean shuffles of such a deck will restore the original order?

---

6. Generators and relations

The presentation (5.24) of the symmetric group is an example of presenting a group by *generators and relations.*

Definition 6.1 A subset $\mathcal{S} \subset G$ is a *generating set* for a group if every element $g \in G$ can be written as a "word" or product of elements of $\mathcal{S}$. That is any element $g \in G$ can be written in the form

$$g = s_{i_1} \cdots s_{i_r} \tag{6.1}$$

where, for each $1 \leq k \leq r$ we have $s_{i_k} \in \mathcal{S}$.

Finitely generated means that the generating set $\mathcal{S}$ is finite, that is, there is a finite list of elements $\{s_1, \ldots s_n\}$ so that all elements of the group can be obtained by taking

products – "words" – in the "letters" drawn from $\mathcal{S}$. For example, the symmetric group is finitely generated by the transpositions. Typical Lie groups such as $SU(n)$ or $SO(n,\kappa)$ (over $\kappa = \mathbb{R}, \mathbb{C}$) are not finitely generated.

The *relations* are then equalities between different words such that any two equivalent words in G can be obtained by successively applying the relations. [30]

In general if we have a finitely generated group we write

$$G = \langle g_1, \ldots, g_n | R_1, \cdots R_r \rangle \tag{6.2}$$

where R_i are words in the letters of $\mathcal{S}$ which will be set to 1. **ALL** other relations, that is, all other identities of the form $W = 1$ are supposed to be consequences of these relations.

Remark: It is convenient to exclude the unit 1 from $\mathcal{S}$. When we write our words it is understood that we can raise a generator s to any integer power s^n where, $s^0 = 1$ and, if $n < 0$, this means $(s^{-1})^{|n|}$. Alternatively, we can, for each generator s introduce another generator t, which will play the role of s^{-1} and then impose another relation $st = ts = 1_G$. A generating set that contains s^{-1} for every generator s is said to be *symmetric*.

Example 2.1: If $\mathcal{S}$ consists of one element a then $F(\mathcal{S}) \cong \mathbb{Z}$. The isomorphism is given by mapping $n \in \mathbb{Z}$ to the word a^n.

Example 2.2: The most general group with one generator and one relation must be of the form:

$$\langle a | a^N = 1 \rangle \tag{6.3}$$

where N is an integer and by replacing $a \to a^{-1}$ we can assume it is a positive integer. You will recognize this as the cyclic group, isomorphic to $\mathbb{Z}_N$ and μ_N.

Example 2.3: *Free groups.* If we impose no relations on the generating set $\mathcal{S}$ then we obtain what is known as the *free group on* $\mathcal{S}$, denoted $F(\mathcal{S})$. If $\mathcal{S}$ consists of one element then we just get $\mathbb{Z}$, as above. However, things are completely different if $\mathcal{S}$ consists of two elements a, b. Then $F(\mathcal{S})$ is very complicated. A typical element looks like one of

$$\begin{aligned} &a^{n_1} b^{m_1} \cdots a^{n_k} \\ &a^{n_1} b^{m_1} \cdots b^{m_k} \\ &b^{n_1} a^{m_1} \cdots a^{n_k} \\ &b^{n_1} a^{m_1} \cdots b^{m_k} \end{aligned} \tag{6.4}$$

where n_i, m_i are nonzero integers (positive or negative). Three nice general results on free groups are:

1. Two free groups $F(\mathcal{S}_1)$ and $F(\mathcal{S}_2)$ are isomorphic iff they have the same cardinality.

[30] See Jacobsen, *Basic Algebra I*, sec. 1.11 for a more precise definition.

2. *Nielsen-Shreier theorem*: Any subgroup of a free group is free.

3. Every group has a presentation in terms of generators and relations. For the group G we can consider the free group $F(G)$ with $\mathcal{S} = G$ as a set. There is then a natural homomorphism $\varphi : F(G) \to G$ where we take a word in elements of G and map concatenation of letters to group multiplication in G. As we will see in Section *** below, the kernel of the homomorphism is a normal subgroup $K(G)$ so that $G \cong F(G)/K(G)$ and $K(G)$ are the relations in this presentation. This presentation can be incredibly inefficient and useless. (Think, for example, of Lie groups.)

Combinatorial group theorists use the notion of a *Cayley graph* to illustrate groups presented by generators and relations. Assuming that $1 \notin \mathcal{S}$ the Cayley graph is a graph whose vertices correspond to all group elements in G and the oriented edges are drawn between g_1 and g_2 if there is an $s \in \mathcal{S}$ with $g_2 = g_1 s$. We label the edge by s. (If $\mathcal{S}$ is symmetric we can identify this edge with the edge from g_2 to g_1 labeled by s^{-1}.) For the free group on two elements this generates the graph shown in Figure 3.

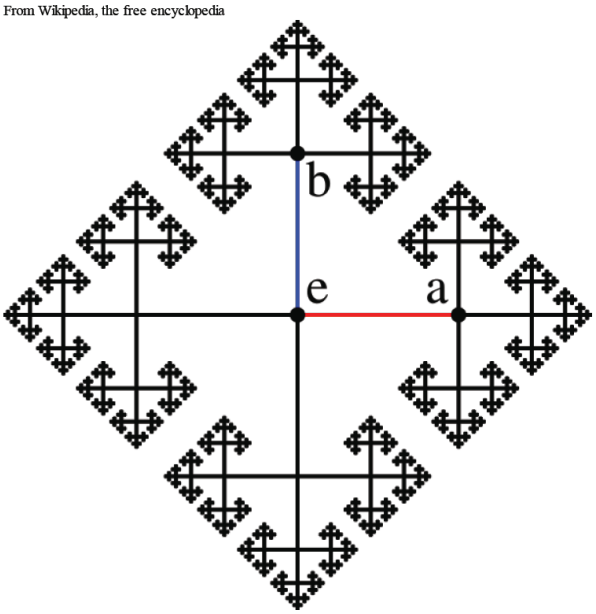


Figure 3: The Cayley graph for the free group on 2 generators a and b.

Example 2.4: *Coxeter groups*: Let m_{ij} by an $n \times n$ symmetric matrix whose entries are positive integers or ∞, such that $m_{ii} = 1$, $1 \leq i \leq n$, and $m_{ij} \geq 2$ or $m_{ij} = \infty$ for $i \neq j$. Then a *Coxeter group* is the group with generators and relations:

$$\langle s_1, \ldots, s_n | \forall i, j : (s_i s_j)^{m_{ij}} = 1 \rangle \tag{6.5}$$

where, if $m_{ij} = \infty$ we interpret this to mean there is no relation.

Note that since $m_{ii} = 1$ we have

$$s_i^2 = 1 \tag{6.6}$$

Quite generally, a group element that squares to 1 is called an *involution*. So all the generators of a Coxeter group are involutions. It then follows that if $m_{ij} = 2$ then s_i and s_j commute. If $m_{ij} = 3$ then the relation can also be written:

$$s_i s_j s_i = s_j s_i s_j \tag{6.7}$$

A theorem of Coxeter's from the 1930's gives a classification of the finite Coxeter groups. [31] Coxeter found it useful to describe these groups by a diagrammatic notation: We draw a graph whose vertices correspond to the generators s_i. We draw an edge between vertices i and j if $m_{ij} \geq 3$. By convention the edges are labeled by m_{ij} and if $m_{ij} = 3$ then the standard convention is to omit the label.

It turns out that the *finite* Coxeter groups can be classified. The corresponding Coxeter diagrams are shown in Figure 4.

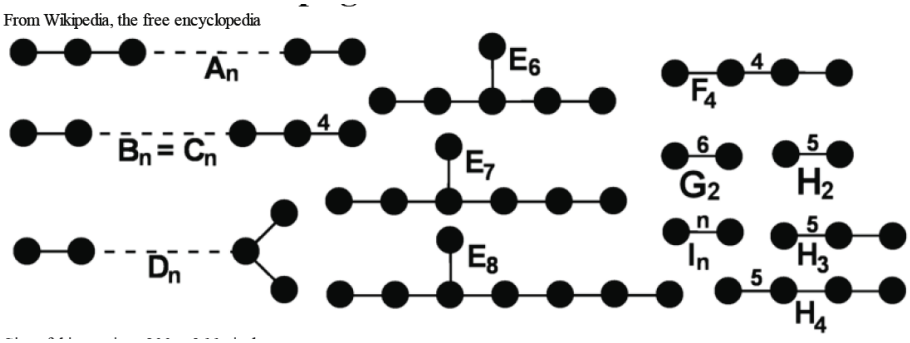


Figure 4: Coxeter's list of finite Coxeter groups. They are finite groups of reflections in some Euclidean space.

The finite Coxeter groups turn out to be isomorphic to concrete groups of *reflections* in some Euclidean space. That is, finite subgroups of $O(N)$ for some N. That is, there is some vector space $\mathbb{R}^N$ and collection of vectors $v_i \in \mathbb{R}^N$ with inner products

$$v_i \cdot v_j = -2\cos(\frac{\pi}{m_{i,j}}) \tag{6.8}$$

so that the group generated by reflections in the plane orthogonal to the vectors v_i:

$$P_{v_i} : v \mapsto v - \frac{2v \cdot v_i}{v_i \cdot v_i} v_i \tag{6.9}$$

is a finite group isomorphic to the Coxeter group with matrix $m_{i,j}$. (Note that since $m_{i,i} = 1$ we have $v_i^2 = 2$ and $P_{v_i}(v) = v - (v \cdot v_i)v_i$.)

Note that, if P_v is the Euclidean reflection in the plane orthogonal to v then $P_{v_1} \circ P_{v_2}$ is just rotation in the plane spanned by v_1, v_2 by an angle 2θ where the angle between v_1 and v_2 is θ. To prove this, note that $P_{v_1} \circ P_{v_2}$ clearly leaves all vectors in the plane orthogonal to v_1, v_2 fixed. Now represent vectors in a 2-dimensional Euclidean plane by

[31] For a quick summary see the expository note by D. Allcock at https://web.ma.utexas.edu/users/allcock/expos/reflec-classification.pdf.

complex numbers, but view $\mathbb{C}$ as a real vector space. WLOG take $v = e^{\mathrm{i}\theta}$. Then P_v is the transformation:

$$P_v : z \mapsto -e^{2\mathrm{i}\theta}\bar{z} \tag{6.10}$$

Note that this is a linear transformation of real vector spaces: $P_v(a_1 z_1 + a_2 z_2) = a_1 P_v(z_1) + a_2 P_v(z_2)$ if $a_1, a_2 \in \mathbb{R}$. To check this formula note that if $z = e^{\mathrm{i}\theta}$ then $P_v(z) = -z$ and if $z = \mathrm{i}e^{\mathrm{i}\theta}$ is in the orthogonal hyperplane to v then $P_v(z) = z$.

Now if $v_a = e^{\mathrm{i}\theta_a}$, $a = 1, 2$, it is an easy matter to compute:

$$\begin{aligned} P_{v_1} \circ P_{v_2} : z &\mapsto -e^{2\mathrm{i}\theta_2}\bar{z} \\ &\mapsto -e^{2\mathrm{i}\theta_1}\overline{-e^{2\mathrm{i}\theta_2}\bar{z}} \\ &= e^{2\mathrm{i}(\theta_1 - \theta_2)} z \end{aligned} \tag{6.11}$$

So: The product of reflections in the hyper-planes orthogonal to two vectors at an angle θ is a rotation by an angle 2θ in the plane spanned by the two vectors.

We will meet some of these groups again later as Weyl groups of simple Lie groups. We have, in fact, already met two of these groups! The case A_n turns out to be isomorphic to the symmetric group S_{n+1}. [32] In this case we have seen that the elementary generators $\sigma_i = (i, i+1)$, $1 \leq i \leq n$ indeed satisfy the Coxeter relations:

$$\begin{aligned} \sigma_i^2 &= 1 \\ (\sigma_i \sigma_{i+1})^3 &= 1 \qquad 1 \leq i \leq n-1 \\ (\sigma_i \sigma_j)^2 &= 1 \qquad |i-j| > 1 \end{aligned} \tag{6.12}$$

Now consider the standard basis e_i for $\mathbb{R}^{n+1}$, $1 \leq i \leq n+1$ and consider the vectors:

$$\alpha_i = e_i - e_{i+1} \tag{6.13}$$

which have the inner products: $\alpha_i^2 = 2$ and $\alpha_i \cdot \alpha_{i\pm 1} = -1$ (so they are at angle $2\pi/3$) and all other inner products vanish. This is summarized in the matrix:

$$\alpha_i \cdot \alpha_j = C_{ij} = 2\delta_{i,j} - \delta_{i,j+1} - \delta_{i,j-1} \tag{6.14}$$

Then the map $s_i \to P_{\alpha_i}$ is an isomorphism of the Coxeter group A_n with a subgroup of $O(n+1)$. Moreover, one computes that

$$P_{\alpha_i}(e_j) = \begin{cases} e_j & j \neq i, i+1 \\ e_{i+1} & j = i \\ e_i & j = i+1 \end{cases} \tag{6.15}$$

So, referring to equation (5.36) we see that this is just the permutation action of σ_i on the standard basis of $\mathbb{R}^{n+1}$. This makes clear that the Coxeter group is isomorphic to the symmetric group S_{n+1}.

♣A presentation of the Monster in terms of generators and relations is known.(Atlas) Give it here? ♣

Remarks

[32] The notation here is standard but exceedingly unfortunate and confusing!!! Here A_n does <u>*NOT*</u> refer to the alternating group! It refers to Cartan's classification of simple Lie groups and the Coxeter group with this label is in fact isomorphic to S_{n+1}.

1. One very practical use of having a group presented in terms of generators and relations is in the construction of homomorphisms. If one is constructing a homomorphism $\phi : G_1 \to G_2$, then it suffices to say what elements the generators map to. That is, if g_i are generators of G_1 we can fully specify a homomorphism by choosing elements $g_i' \in G_2$ (not necessarily generators) and declaring

$$\phi(g_i) = g_i' \quad . \tag{6.16}$$

   However, we cannot choose the g_i' arbitrarily. Rather, the g_i' must satisfy the same relations as the g_i. This puts useful constraints on what homomorphisms you can write down. For example, using this idea you can prove that there is no nontrivial homomorphism $\phi : \mathbb{Z}_N \to \mathbb{Z}$.

2. In general it is hard to say much about a group given a presentation in terms of generators and relations. For example, it is not even obvious, in general, if the group is the trivial group! This is part of the famous "word problem for groups." There are finitely presented groups where the problem of saying whether two words represent the same element is undecidable! [33] However, for many important finitely presented groups the word problem can be solved. Indeed, the word problem was first formulated by Max Dehn in 1911 and solved by him for the surface groups discussed below.

3. Nevertheless, there are four Tietze transformations (adding/removing a relation, adding/removing a generator) which can transform one presentation of a group to a different presentation of an isomorphic group. It is a theorem [REF!] that any two presentations can be related by a finite sequence of Tietze transformations. How is this compatible with the previous remark? The point is that the number $f(n)$ of such transformations needed to transform a presentation of the trivial group with n relations into the trivial presentation grows faster than any recursive function of n.

4. It turns out that the case of Coxeter groups $B_n = C_n$ are isomorphic to the group of symmetric permutations $\mathcal{W}B_n \subset S_{2n}$ discussed in card-shuffling. The Coxeter diagrams are very similar to the *Dynkin diagrams* that are used to label finite dimensional simple Lie algebras over the complex numbers except that H_n and I_n do not correspond to Lie algebras.

♣It would be more effective here to give an example of a set of generators and relations that is actually isomorphic to the trivial group - but not obviously so. ♣

Exercise *Homomorphisms involving $\mathbb{Z}_N$ and $\mathbb{Z}$*

a.) Write a nontrivial homomorphism $\mu : \mathbb{Z} \to \mathbb{Z}_N$.

b.) Show that there is no nontrivial homomorphism $\mu : \mathbb{Z}_N \to \mathbb{Z}$. [34]

[33] The Wikipedia article on "Word problem for groups," is useful.

[34] *Answer*: Since $\mathbb{Z}_N$ can be generated by one element, say $\bar{1}$, it suffices to say what the value of $\phi(\bar{1})$ is. The trivial homomorphism takes the generator to zero: $\phi(\bar{1}) = 0 \in \mathbb{Z}$ and hence takes every element to zero. On the other hand, if $\phi(\bar{1}) = k$ is a nonzero integer, then $Nk = N\phi(\bar{1}) = \phi(N\bar{1}) = \phi(\bar{0}) = 0$, a contradiction. So there is no nontrivial homomorphism.

c.) Find the most general homomorphism $\mu : \mathbb{Z} \to \mathbb{Z}$.

d.) Find the most general homomorphism $\mu : \mathbb{Z}_N \to \mathbb{Z}_N$.

Exercise *One generator and many relations*

Suppose a group has a single generator g but many relations. Describe this group. [35]

Exercise *Simple Roots Of $SU(n+1)$*

a.) Verify equation (6.14). The matrix C_{ij} is known as a *Cartan matrix* of $SU(n+1)$.

b.) Show that the vectors α_i are all orthogonal to the all-one vector: $v = (1, \ldots, 1)$ and that they span the orthogonal complement of v.

c.) Show that the permutation representation of S_{n+1} separately preserves v and the orthogonal complement of v. Thus, $\mathbb{R}^{n+1}$ gives what is known as a *reducible representation* of S_{n+1}.

d.) Compute the action of P_{α_i} on α_j. Give the matrix representation relative to the ordered basis $\{\alpha_1, \ldots, \alpha_n\}$. [36]

Exercise Show that

$$\langle a, b | a^3 = 1, b^2 = 1, abab = 1 \rangle \tag{6.17}$$

is a presentation of S_3

Exercise

Consider the group with presentation:

$$\langle T, S | (ST)^3 = 1, S^2 = 1 \rangle \tag{6.18}$$

[35] *Answer*: Let the generator be g. The relations must be of the form $g^{k_1} = 1, \ldots, g^{k_n} = 1$ for some integers $k_1, \ldots, k_n$, and WLOG we can assume they are positive. Then it is not hard to see, using the Chinese remainder theorem (see below) that the group is isomorphic to $\mathbb{Z}_N$ where $N = gcd(k_1, ..., k_n)$.

[36] *Answer* The matrix is a diagonal matrix of 1's except on the 3×3 block for rows and columns $i-1, i, i+1$ where it looks like

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} .$$

Is this group finite or infinite?

This group plays a very important role in string theory.

Exercise *Bounds on the minimal number of generators of a finite group*

Suppose we have a set of finite groups $G_1, G_2, G_3, \ldots$ with a minimal set of generators $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots$ of cardinality $|\mathcal{S}_k| = k$. Show that $2|G_k| \leq |G_{k+1}|$ and hence as $k \to \infty$ the order $|G_k|$ must grow at least as fast as 2^k.

Remark: Denote the smallest cardinality of a set of generators of G by $d(G)$. If G is a finite and transitive permutation subgroup of S_n (meaning it acts transitively on some set X) then there is a constant C such that

$$d(G) \leq C \frac{n}{\sqrt{\log n}} \tag{6.19}$$

and if G is a primitive permutation group, meaning that it acts on a set X such that it does not preserve any nontrivial disjoint decomposition of X, then there is a constant C so that if $n \geq 3$:

$$d(G) \leq C \frac{\log n}{\sqrt{\log\log n}} \tag{6.20}$$

Moreover, these results are asymptotically the best possible. For a review of such results see. [37]

Exercise *Generators And Relations For Products Of Groups*

Suppose you are given groups G_1 and G_2 in terms of generators and relations. Write a set of generators and relations for the product group $G_1 \times G_2$. [38]

6.1 Example Of Generators And Relations: Fundamental Groups In Topology

Presentations in terms of generators and relations is very common when discussing the *fundamental group* of a topological space X.

This subsection assumes some knowledge of topological spaces and the idea of a homotopy. Without trying to be too precise we choose a basepoint $x_0 \in X$ and let $\pi_1(X, x_0)$ be the set of closed paths in X, beginning and ending at x_0 where we identify two paths if they can be continuously deformed into each other. We can define a group multiplication by concatenation of paths. Inverses exist since we can run paths backwards. The following subsubsection contains more precise definitions.

♣Need to fix pictures and change p_0 to x_0. ♣

[37] F. Menegazzo, "The Number of Generators of a Finite Group," Irish Math. Soc. Bulletin 50 (2003), 117128.

[38] *Answer*: If $G_1 = \langle g_i | R_i \rangle$ and $G_2 = \langle h_a | S_a \rangle$ then $G_1 \times G_2 = \langle g_i, h_a | R_i, S_a, g_i h_a g_i^{-1} h_a^{-1} = 1 \rangle$.

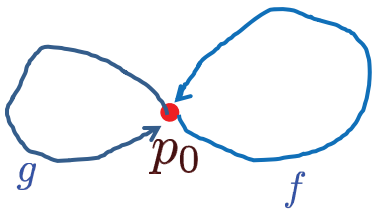


Figure 5: Two loops f, g with basepoint at x_0.

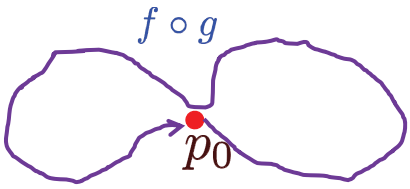


Figure 6: The concatenation of the looops $f \star g$. Note that the "later" loop is written on the right. This is generally a more convenient convention when working with homotopy and monodromy. In order for $f \star g$ to be a map from $[0,1]$ into X we should run each of the individual loops at "twice the speed" so that at time $t = 1/2$ the loop returns to x_0. However, in homotoping $f \star g$ there is no reason why the point at $t = 1/2$ has to stay at x_0.

6.1.1 The Fundamental Group Of A Topological Space

Choose a point $x_0 \in X$. The fundamental group $\pi_1(X, x_0)$ based at x_0 is, as a set, the set of homotopy classes of closed curves.

That is we consider continuous maps:

$$f : ([0,1], \{0,1\}) \to (X, \{x_0\}) \tag{6.21}$$

These define paths in X with beginning and ending point fixed at x_0. The path must be traveled in time 1.

We say that two such paths f_0, f_1 are *homotopic* if there is a continuous map

$$F : [0,1] \times [0,1] \to X \tag{6.22}$$

such that

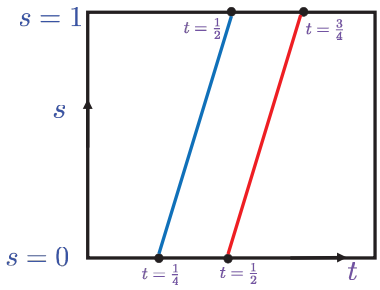


Figure 7: The homotopy demonstrating that loop concatenation is an associative multiplication on homotopy equivalence classes of closed loops. The blue line is $s = 4t - 1$ and the red line is $s = 4t - 2$.

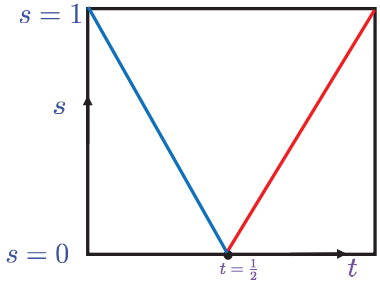


Figure 8: The homotopy demonstrating that the loop $g(t) = f(1-t)$ provides a representative for the inverse of $[f(t)]$.

1. $F(0,t) = f_0(t)$ and $F(1,t) = f_1(t)$
2. $F(s,0) = F(s,1) = x_0$

If we define $f_s(t) := F(s,t)$ and consider $f_s(t)$ as a path in t at fixed s then, as we vary s we are describing a path of paths.

Now, homotopy of paths in X is an equivalence relation. [39] We denote by $[f]$ the equivalence class of a path f and we denote the set of such equivalence class by $\pi_1(X, x_0)$. We will see that this set has a natural and beautiful group structure.

We can define a group structure on $\pi_1(X, x_0)$ by concatenating curves as in Figure 6 and rescaling the time variable so that it runs from 0 to 1. In equations we have

$$f_1 \star f_2(t) := \begin{cases} f_1(2t) & 0 \le t \le \frac{1}{2} \\ f_2(2t-1) & \frac{1}{2} \le t \le 1 \end{cases} \tag{6.23}$$

[39] See section 1.1 above for this notion.

Remarks

1. Note that we are composing successive paths *on the right*. This is slightly nonstandard but a nice convention when working with monodromy and path ordered expontentials of gauge fields - one of the main physical applications.

2. Note well that $(f_1 \star f_2) \star f_3$ is *NOT* the same path as $f_1 \star (f_2 \star f_3)$. This observation ultimately leads to the notion of A_∞ spaces.

3. For the moment we simply notice that if we mod out by homotopy then we have a well-defined product on homotopy classes in $\pi_1(X, x_0)$

$$[f_1] \cdot [f_2] := [f_1 \star f_2] \tag{6.24}$$

and the virtue of passing to homotopy classes is that now the product (6.24) is in fact associative. The proof is in Figure 7. Written out in excruciating detail the homotopy is

$$F(s,t) = \begin{cases} f_1(\frac{4}{s+1}t) & 0 \leq t \leq \frac{s+1}{4} \\ f_2(4t - (s+1)) & \frac{s+1}{4} \leq t \leq \frac{s+2}{4} \\ f_3(\frac{4}{2-s}(t - \frac{s+2}{4})) & \frac{s+2}{4} \leq t \leq 1 \end{cases} \tag{6.25}$$

4. Since we have an associative product on $\pi_1(X, x_0)$ we are now ready to define a group structure. The identity element is clearly given by the (homotopy class of the) constant loop: $f(t) = x_0$. If a homotopy class is represented by a loop $f(t)$ then the inverse is represented by running the loop backwards: $g(t) := f(1-t)$. The two are joined at $t = 1/2$, and since this is in the open interval $(0, 1)$ the image can be deformed away from x_0. See Figure 8. In equations, there is a homotopy of $f \star g$ with the constant loop given by

$$F(s,t) = \begin{cases} f(2t) & t \leq \frac{1-s}{2} \\ f(1-s) & \frac{1-s}{2} \leq t \leq \frac{1+s}{2} \\ f(2-2t) & \frac{1+s}{2} \leq t \leq 1 \end{cases} \tag{6.26}$$

Thus, with the group operation defined by concatenation in the sense of (6.24) the set of homotopy classes $\pi_1(X, x_0)$ is a group. It is known as the fundamental group based at x_0.

5. A connected space such that $\pi_1(X, x_0)$ is the trivial group is called *simply connected*.

A Basic Example: The Fundamental Group Of The Circle:

The first and most basic example of a nontrivial fundamental group is the fundamental group of the circle. It should be intuitively clear that

$$\pi_1(S^1, x_0) \cong \mathbb{Z} \tag{6.27}$$

which just measures the number of times the path winds around the circle. The sign of the integer takes into account winding clockwise vs. counterclockwise.

Let us amplify a little on how one proves this basic fact. We will just give the main idea. For a thorough and careful proof see A. Hatcher's book on Algebraic Topology, or Section 13.2 of

http://www.physics.rutgers.edu/∼gmoore/511Fall2014/Physics511-2014-Ch2-Topology.pdf

We have a standard map

$$p : \mathbb{R} \to S^1 \tag{6.28}$$

given by $p(x) = e^{2\pi i x}$. Note that the inverse image of any phase $e^{2\pi i s}$ with $s \in \mathbb{R}$ is the set of real numbers $s + \mathbb{Z}$. So p is the map that identifies the orbits of the $\mathbb{Z}$-action by translation on $\mathbb{R}$ with S^1. That is $S^1 \cong \mathbb{R}/\mathbb{Z}$. Now, suppose we have a map $\bar{f} : [0,1] \to S^1$. We claim that there is a map $f : [0,1] \to \mathbb{R}$ so that $\bar{f}$ "factors through p," meaning that there is a map $f : [0,1] \to \mathbb{R}$ such that:

$$\bar{f} = p \circ f \tag{6.29}$$

The problem of finding such a map f is nicely expressed in terms of diagrams. One is trying to complete the following diagram to make it a commutative diagram by finding a suitable map f to use on the dashed line in:

$$\begin{array}{ccc} & & \mathbb{R} \\ & \overset{f}{\nearrow} & \big\downarrow p \\ [0,1] & \xrightarrow{\bar{f}} & S^1 \end{array} \tag{6.30}$$

In other words, given $\bar{f}$ one is trying to find a map $f : [0,1] \to \mathbb{R}$ so that

$$e^{2\pi i f(x)} = \bar{f}(x) \tag{6.31}$$

So, $f(x)$ is a logarithm, but a logarithm is not single-valued. Nevertheless, in sufficiently small open sets of $[0,1]$ (small enough that the image of $\bar{f}(x)$ does not "wrap" around the circle) one can choose an unambiguous logarithm. Let us fix (WLOG) $\bar{f}(0) = 1$. Then $f(0)$ can be any integer, n_0. Once we choose that integer the branch of the logarithm is fixed in some small open set $[0, \epsilon)$. Now we continue choosing open sets along the interval so that we can choose an unambiguous logarithm and we fix branches successively as we move from one open set to the next along the positive direction. The net result is an unambiguous map $f : [0,1] \to \mathbb{R}$ satisfying (6.31). Now if $\bar{f}(1) = 1$, then $f(1) = n_1$, with $n_1 \in \mathbb{Z}$. The integer $n_1 - n_0$ only depends on $\bar{f}$, and not on the particular choice n_0, and is the winding number of the map. Moroever, the winding number is continuous as a function of $\bar{f}$ and hence only depends on the homotopy class.

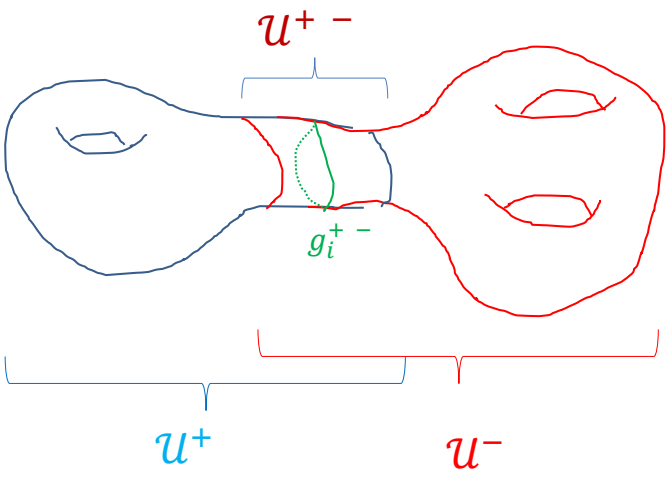


Figure 9: Illustrating the Seifert-VanKampen theorem. The green curve has a homotopy class in $\mathcal{U}^{+-}$ that is one of the generators of $\pi_1(\mathcal{U}^{+-})$. Now it must separately be a word W_i^+ in the generators of $\pi_1(\mathcal{U}^+)$ and W_i^- in the generators of $\pi_1(\mathcal{U}^-)$ so in $\pi_1(X)$ there must be a relation of the form $W_i^+ = W_i^-$.

1. It is not hard to prove that

$$\pi_1(X \times Y, (x_0, y_0)) \cong \pi_1(X, x_0) \times \pi_1(Y, y_0) \tag{6.32}$$

   So the fundamental group of the n-dimensional torus is isomorphic to the n-dimensional lattice $\mathbb{Z}^n$.

2. If $F : X \to Z$ is a continuous map of topological spaces and takes $x_0 \in X$ to $z_0 \in Z$ then we can define $F_* : \pi_1(X, x_0) \to \pi_1(Z, z_0)$ simply by $F_*[f] := [F \circ f]$. This can be shown to be a group homomorphism. In particular, if F is a homotopy equivalence, then it is a group isomorphism.

3. In algebraic topology books a major result which is proved is the *Seifert-van Kampen theorem*. This is an excellent illustration of defining groups by generators and relations. The theorem can be useful because it allows one to compute $\pi_1(X, x_0)$ by breaking up X into simpler pieces. Specifically, suppose that $X = U^+ \cup U^-$ is a union of two open path-connected subsets and that $U^{+-} := U^+ \cap U^-$ is also path-connected and contains x_0. See Figure 9. Now suppose we know presentations of the fundamental groups of the pieces U^+, U^-, U^{+-} in terms of generators and relations:

$$\begin{aligned} \pi_1(U^+, x_0) &\cong \langle g_1^+, \ldots, g_{n^+}^+ | R_1^+, \ldots, R_{m^+}^+ \rangle \\ \pi_1(U^-, x_0) &\cong \langle g_1^-, \ldots, g_{n^-}^- | R_1^-, \ldots, R_{m^-}^- \rangle \\ \pi_1(U^{+-}, x_0) &\cong \langle g_1^{+-}, \ldots, g_{n^{+-}}^{+-} | R_1^{+-}, \ldots, R_{m^{+-}}^{+-} \rangle \end{aligned} \tag{6.33}$$

   Then the recipe for computing $\pi_1(X, x_0)$ is this: Denote the injection $\iota^+ : U^{+-} \to U^+$ and $\iota^- : U^{+-} \to U^-$. Then the generators of $\pi_1(U^{+-}, x_0)$ push forward to words in

g_i^+ or g_i^-, respectively:

$$\begin{aligned} \iota_*^+(g_i^{+-}) &:= W_i^+ \qquad i = 1, \ldots, n^{+-} \\ \iota_*^-(g_i^{+-}) &:= W_i^- \qquad i = 1, \ldots, n^{+-} \end{aligned} \tag{6.34}$$

Finally, we have the presentation:

$$\pi_1(X, x_0) \cong \langle g_1^+, \ldots, g_{n^+}^+, g_1^-, \ldots, g_{n^-}^- | R_\alpha \rangle \tag{6.35}$$

where the relations R_α include the *old relations*

$$R_1^+, \ldots, R_{m^+}^+, R_1^-, \ldots, R_{m^-}^- \tag{6.36}$$

and a set of *new relations*:

$$W_1^+(W_1^-)^{-1}, \ldots, W_{n^{+-}}^+(W_{n^{+-}}^-)^{-1} \tag{6.37}$$

It is obvious that these are relations on the generators. What is not obvious is that these are the only ones. Note that in the final presentation the generators g_i^{+-} and the relations R_i^{+-} have dropped out of the description.

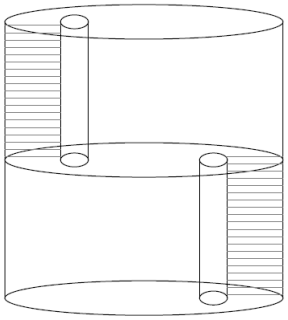


Figure 10: The house that Bing built. Taken from M. Freedman and T. Tam Nguyen-Pham, "Non-Separating Immersions Of Spheres and Bing Houses," which describes nice mathematical properties of this house.

Exercise
Show that if $X = S^n$ with $n > 1$ then $\pi_1(X, x_0)$ is the trivial group.

Exercise
Does the fundamental group depend on a choice of basepoint x_0 ?

Exercise
What is the fundamental group of Serin Physics Laboratory?

Exercise *The house that Bing built*
Show that the house in Figure 10 can be shrink-wrapped with a single balloon so that the complement of the balloon in $\mathbb{R}^3$ is connected and simply connected.

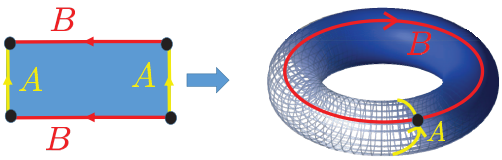


Figure 11: Right: Cutting a torus along the A and B cycles the surface falls apart into a rectangle, shown on the left. Conversely, gluing the sides of the rectangle together produces a torus with distinguished closed curves A, B.

6.1.2 Surface Groups: Compact Two-Dimensional Surfaces

The fundamental groups of two-dimensional surfaces, known as *surface groups* and braid groups turn out to provide a very rich set of examples of groups defined by generators and relations.

The simplest example of a nontrivial surface group is the torus. Let a, b denote the homotopy classes of the cycles A, B shown in Figure 11. One can convince oneself that these generate the fundamental group: Every closed curve based at x_0 can be homotoped to a word in $a^{\pm 1}$ and $b^{\pm 1}$. Now, if we cut the torus along the cycles the surface falls apart

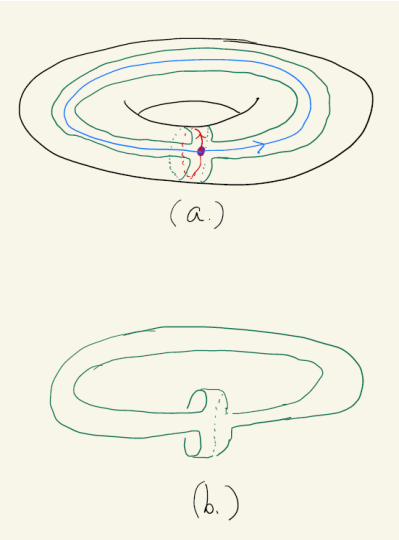


Figure 12: Illustrating the Seifert-VanKampen theorem for the torus: The "tubular neighborhood" - the green region - of the cutting curves, shown in (b) is homotopy equivalent to a one-point union of two circles. The latter space has a π_1 which is a free group on two generators. The boundary of the green region contracts into the remainder of the surface - which can be deformed to a disk. Therefore the group commutator $[a, b] = 1$.

into a rectangle as shown in Figure 11. The edge of the rectangle represents the class $aba^{-1}b^{-1}$. A slightly different way of thinking about this is described in Figure 12.

Definition: In general, in group theory an expression of the form $g_1 g_2 g_1^{-1} g_2^{-1}$ is known as a *group commutator* and is sometimes denoted $[g_1, g_2]$. It should not be confused with the commutator of matrices $[A_1, A_2] = A_1 A_2 - A_2 A_1$.

Returning to the fundamental group of the torus, the group commutator $[a, b]$ it can be contracted inside the rectangle to a point. Therefore, the generators a, b satisfy the relation:

$$aba^{-1}b^{-1} = 1 \tag{6.38}$$

so this means

$$ab = ba \tag{6.39}$$

In fact, this is the only relation and therefore:

$$\pi_1(T^2, x_0) \cong \mathbb{Z} \oplus \mathbb{Z}. \tag{6.40}$$

The above ideas generalize nicely. Let us consider the case of a genus 2, or 2-handled, surface shown in Figure 14. The fundamental group can be presented as a group with four generators a_1, b_1, a_2, b_2 and one relation:

$$[a_1, b_1][a_2, b_2] = 1 \tag{6.41}$$

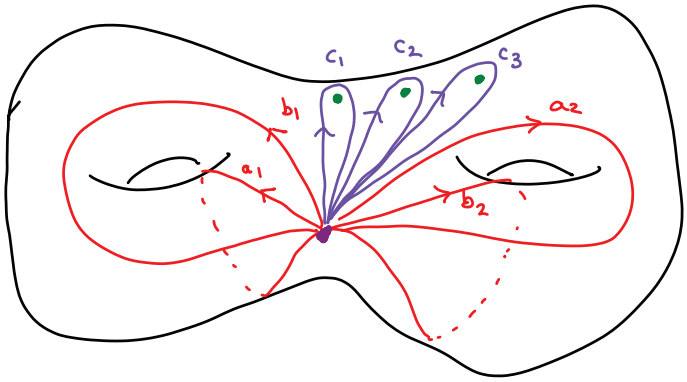


Figure 13: A collection of closed paths at x_0 which generate the fundamental group of a two-dimensional surface with two handles and three (green) holes.

Now let us consider a more complicated surface, perhaps with punctures as shown in Figure 15. By cutting along the paths shown there the surface unfolds to a presentation by gluing as in Figure 16:

From these kinds of constructions one can prove [40] that the fundamental group of an orientable surface with g handles and p punctures will be

$$\pi_1(S, x_0) = \langle a_i, b_i, c_s | \prod_{i=1}^{g} [a_i, b_i] \prod_{s=1}^{p} c_s = 1 \rangle \tag{6.42}$$

There is only one relation so this is very close to a free group! In fact, for $p \geq 1$ we can solve for one generator c_s in terms of the rest so the group is just a free group on $2g+p-1$ generators. When there are no punctures the group is not a free group. Groups of the form (6.42) are sometimes called *surface groups*.

As mentioned above, a flat connection amounts to a representation of this group - so one is searching for matrices A_i, B_i, C_s such that

$$\prod_i (A_i B_i A_i^{-1} B_i^{-1}) \prod_s C_s = 1 \tag{6.43}$$

[40] See, for example, W. Massey, *Introduction to Algebraic Topology*, Springer GTM

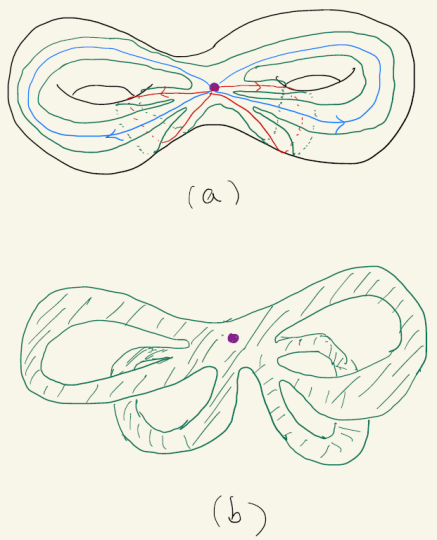


Figure 14: Illustrating the Seifert-VanKampen theorem for a genus two surface: The "tubular neighborhood" - the green region - of the cutting curves, shown in (b) is homotopy equivalent to a one-point union of 4 circles. The latter space has a π_1 which is a free group on four generators which we can call a_1, b_1, a_2, b_2. The boundary of the green region is a single circle homotopic to $[a_1, b_1][a_2, b_2]$. But it contracts into the remainder of the surface - which can be deformed to a disk. Therefore by the Seifert-van Kampen theorem the presentation of π_1 of the genus two surface has a single relation $[a_1, b_1][a_2, b_2] = 1$.

Exercise *Fundamental group of the Klein bottle*

A very interesting unorientable surface is the Klein bottle. Its fundamental group has two natural presentations in terms of generators and relations. One is

$$\langle a, b | a^2 = b^2 \rangle \tag{6.44}$$

and the other is

$$\langle g_1, g_2 | g_1 g_2 g_1 g_2^{-1} = 1 \rangle \tag{6.45}$$

Show that these two presentations are equivalent.

Exercise

Use the Seifert-van Kampen theorem to relate the fundamental group of a torus to that of a torus with a disk cut out.

6.1.3 Braid Groups And Anyons

Let us modify Figure 2 and Figure 1 to include an under-crossing and overcrossing of the strands. So now we are including more information - the topological configuration of the

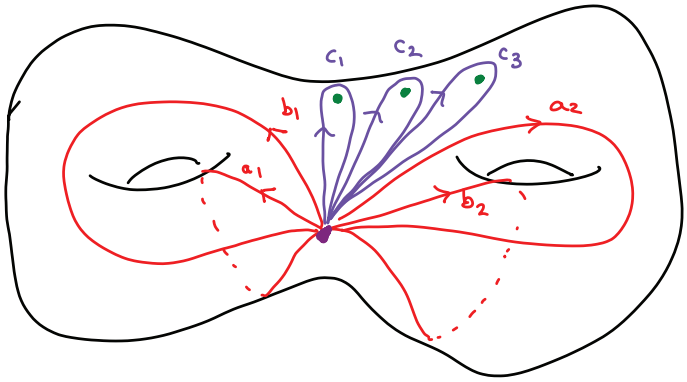


Figure 15: A collection of closed paths at x_0 which generate the fundamental group of a two-dimensional surface with two handles and three (green) holes.

strands in three dimensions. In an intuitive sense, which we will not make precise here we obtain a group called the nth *braid group*. It is generated by the overcrossing $\tilde{\sigma}_i$ of strings $(i, i+1)$, for $1 \leq i \leq n-1$ and may be pictured as in Figure 17. Note that $\tilde{\sigma}_i^{-1}$ is the undercrossing.

Now one verifies the relations

$$\tilde{\sigma}_i \tilde{\sigma}_j = \tilde{\sigma}_j \tilde{\sigma}_i \qquad |i-j| \geq 2 \tag{6.46}$$

and

$$\tilde{\sigma}_i \tilde{\sigma}_{i+1} \tilde{\sigma}_i = \tilde{\sigma}_{i+1} \tilde{\sigma}_i \tilde{\sigma}_{i+1} \tag{6.47}$$

where the relation (6.47) is illustrated in Figure 18.

The braid group $\mathcal{B}_n$ may be defined as the group generated by $\tilde{\sigma}_i$ subject to the relations (6.46)(6.47):

$$\mathcal{B}_n := \langle \tilde{\sigma}_1, \ldots, \tilde{\sigma}_{n-1} | \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\sigma}_i^{-1} \tilde{\sigma}_j^{-1} = 1, |i-j| \geq 2; \tilde{\sigma}_i \tilde{\sigma}_{i+1} \tilde{\sigma}_i = \tilde{\sigma}_{i+1} \tilde{\sigma}_i \tilde{\sigma}_{i+1} \rangle \tag{6.48}$$

The braid group $\mathcal{B}_n$ may also be defined as the fundamental group of the space of configurations of n unordered points on the disk D. We first consider the set:

$$\{(x_1, \ldots, x_n) | x_i \in D \qquad x_i \neq x_j \quad i \neq j\} \tag{6.49}$$

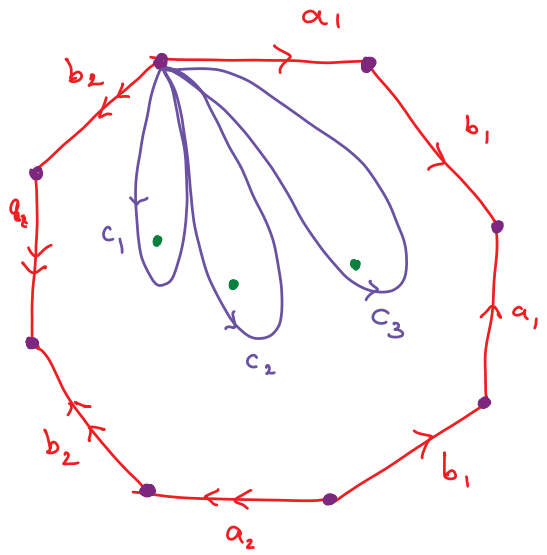


Figure 16: When the directed edges are identified according to their labels the above surface reproduces the genus two surface with three punctures. Since the disk is simply connected we derive one relation on the curves shown here.

Then we observe that there is a group action of S_n on this set. Note that this set is not simply connected: For example if we let x_1 loop around x_2 holding all other x_i fixed it should be intuitively clear that the loop cannot be deformed to the trivial loop. That is even more clear if you view the looping process as taking place in time on particles in a plane.

Now we consider the space of orbits under this group action:

$$\mathcal{C}_n := \{(x_1, \ldots, x_n) | x_i \in D \qquad x_i \neq x_j \quad i \neq j\}/S_n \tag{6.50}$$

There are new nontrivial loops here where, for example, x_i and x_j exchange places, all other x_k staying fixed.

♣Since we must quotient by S_n this needs to be moved to the section on group actions on spaces. ♣

Note that the "only" difference from the presentation of the symmetric group is that we do *not* put any relation like $(\tilde{\sigma}_i)^2 = 1$. Indeed, $\mathcal{B}_n$ is of infinite order because $\tilde{\sigma}_i^n$ keeps getting more and more twisted as $n \to \infty$.

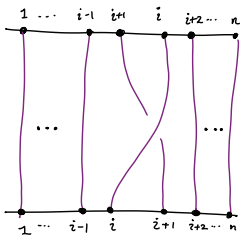


Figure 17: Pictorial illustration of the generator σ_i of the braid group B_n.

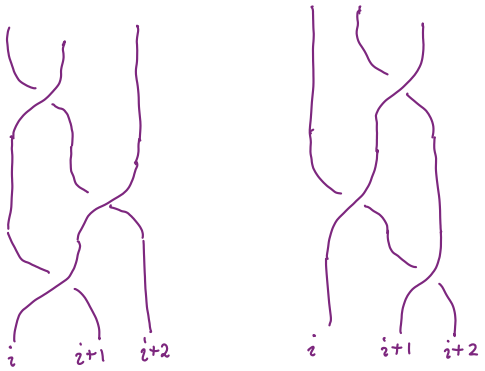


Figure 18: Pictorial illustration of the Yang-Baxter relation.

Exercise *Homomorphisms Between Braid And Symmetric Groups*

a.) Define a homomorphism $\mu : \mathcal{B}_n \to S_n$.

b.) Can you define a homomorphism $s : S_n \to \mathcal{B}_n$ so that $\mu \circ s$ is the identity transformation?

Remarks

1. In the theory of integrable systems the relation (6.47) is closely related to the "Yang-Baxter relation." It plays a fundamental role in integrable models of 2D statistical mechanics and field theory.

2. One interesting application of permutation groups to physics is in the quantum theory of identical particles. It was a major step in the development of quantum theory when Einstein and Bose realized that a system of n identical kinds of particles (photons, for example, or atomic Nuclei of the same isotope) are in fact *indistinguishable*. [41] In mathematical terms, there is a group action of S_n on a set of n indistinguishable particles leaving the physical system "the same." In quantum mechanics this translates into the statement that the Hilbert space of a system of n indistinguishable particles should be a representation of (a central extension of) S_n. There are many different representations of S_n (we have already encountered three different ones). Most of them are higher dimensional. Particles transforming in higher-dimensional representations are said to satisfy "parastatistics." (This idea goes back to B. Green in the 1950's and Messiah and Greenberg in the 1960's.) However, remarkably, in relativistically invariant theories in spacetimes of dimension larger than 3 particles are either bosons or fermions. This is related to the classification of the *projective representations* of $SO(d,1)$, where d is the number of spatial dimensions, for relativistic systems and to representations of $SO(d)$ for nonrelativistic systems. (We will discuss projective representations in section **** below.) Now, when discussing projective representations the fundamental group of $SO(d,1)$ and $SO(d)$ becomes important. In fact $\pi_1(SO(d,1)) \cong \pi_1(SO(d))$ for $d \geq 2$. However, there is a fundamental difference between $d \leq 2$ and $d > 2$. The essential point is that the fundamental group $\pi_1(SO(2)) \cong \mathbb{Z}$ is infinite while $\pi_1(SO(d)) \cong \mathbb{Z}_2$ for $d \geq 3$. A consequence of this, and other principles of physics is that in $2+1$ and $1+1$ dimensions, particles with "anyonic" statistics can exist. [42] Anyons are defined by the property that, if we just consider the wavefunction of two identical such particles, $\Psi(z_1, z_2)$ where z_1, z_2 are points in the plane and then we adiabatically switch their positions using the kind of braiding that defines $\tilde{\sigma}$ then

$$\tilde{\sigma} \cdot \Psi(z_2, z_1) = e^{\mathrm{i}\theta}\Psi(z_1, z_2) \tag{6.51}$$

Unlike bosons and fermions where $\theta = 0, \pi \mathrm{mod} 2\pi$, respectively, for "anyons" the phase can be anything - hence the name. There are even physical realizations of this theoretical prediction in the fractional quantum Hall effect. Moreover, quantum wavefunctions should transform in representations of the braid group. The law (6.51)

[41] See Chapters 24 and 25 in the book *Einstein and the Quantum* by A.D. Stone for a nice historical account of the importance of this discovery in the development of quantum mechanics.

[42] The possible existence of anyons was pointed out by Leinaas and Myrheim in 1977. Another early reference are the papers of Goldin, Melnikof and Sharp. The term "anyon" was invented in F. Wilczek, "Quantum Mechanics of Fractional-Spin Particles". Physical Review Letters 49 (14): 957959. For comments on the early history of these ideas see "The Ancestry of the 'Anyon' " in Physics Today, August 1990, page 90.

leads to the 1-dimensional representation $\tilde{\sigma} \to e^{i\theta}$ but there can also be more interesting "nonabelian representations." That is, there can be interesting irreducible representations of dimension greater than one, and if wavefunctions transform in such representations there can be *nonabelian statistics*. The particles should be called *nonabelions*. There are some theoretical models of fractional quantum Hall states in which this takes place. [43] Nonabelions are of potentially great importance because of their possible use in quantum computation, an observation first made by A. Kitaev.

Here are some sources for more material about anyons:

♣Need to keep updating and refining this reference list. ♣

1. There are some nice lecture notes by John Preskill, which discuss the potential relation to quantum computation and quantum information theory: http://www.theory.caltech.edu/~preski
2. A. Stern, "Anyons and the quantum Hall effectA pedagogical review". Annals of Physics 323: 204; arXiv:0711.4697v1.
3. A. Lerda, *Anyons: Quantum mechanics of particles with fractional statistics* Lect.Notes Phys. M14 (1992) 1-138
4. A. Khare, *Fractional Statistics and Quantum Theory*,
5. G. Dunne, *Self-Dual Chern-Simons Theories.*
6. David Tong, "Lectures on the Quantum Hall Effect," e-Print: arXiv:1606.06687
7. S. Burton, "A Short Guide To Anyons and Modular Functors," 1610.05384
8. J. Pachos, *Introduction To Topological Quantum Computation*
9. K. Beer, "From Categories To Anyons: A Travelogue," 1811.06670
10. E.C. Rowell and Z. Wang, "Mathematics of Topological Quantum Computation," Bulletin American Math. Soc., 55, 183
11. Z. Wang, "Topological Quantum Computation," http://web.math.ucsb.edu/∼zhenghwa/data/course/c

6.1.4 Fundamental Groups Of Three-Dimensional Manifolds

The fundamental groups of surfaces are easy to write down, and surfaces can be classified. The situation is completely different in dimensions three and larger where the problem is much harder. Of course, we could take products like $S^1 \times \Sigma$, where Σ is a two-dimensional manifold. But many more possibilities arise.

[43] Important work on the compatibility of nonabelions with spin-statistics theorems was done by Jurg Fröhlich. Perhaps the first concrete proposal of a physical system with nonabelions is that of G.W. Moore and N. Read, "Nonabelions in the fractional quantum hall effect." Nuclear Physics B360, 1991. It was inspired, in part, by the work of G. Moore and N. Seiberg, "Polynomial Equations For RCFT" and "Classical and Quantum Conformal Field Theory," Commun.Math.Phys. 123 (1989) 177 which gave the first description of a modular tensor category.

Let us start by considering a circle $S^1 \subset S^3$. It is probably helpful to think of S^3 as $\mathbb{R}^3$ with the boundary at infinity identified to a point. Now the tubular neighborhood of the circle is diffeomorphic to $S^1 \times D^2$ and has boundary $S^1 \times S^1$. Imagine cutting out this tubular neighborhood - what remains? It is some three-manifold with a single boundary which is also a torus $S^1 \times S^1$. If you draw a solid torus you will notice that one of the homotopically nontrivial loops on the torus becomes homotopically trivial.

FIGURE HERE OF SOLID TORUS AND A-CYCLE.

The other cycle is nontrivial. Thus, it must be that S^3 can be obtained by gluing together two solid tori (bagels). However, S^3 is simply connected, so the cycle which is nontrivial in one bagel must become trivial in the other, and vice versa. This suggests that there are topologically interesting diffeomorphisms of the torus that are used when gluing together the two bagels: Note that if you glue with the identity transformation you get $S^2 \times S^1$, not S^3.

Indeed a vast number of new constructions comes of three-manifolds comes from ideas of "surgery." An important point is that there are typically many nontrivial diffeomorphisms of a surface. Let us see this with the case of a torus.

We consider the torus to be the space of orbits $\mathbb{R}^2/\mathbb{Z}x\mathbb{Z}$ with coordinates (σ^1, σ^2) with $\sigma^i \sim \sigma^i + 1$. Then consider the transformation

$$\begin{aligned} \sigma^1 &\to a\sigma^1 + b\sigma^2 \\ \sigma^2 &\to c\sigma^1 + d\sigma^2 \end{aligned} \tag{6.52}$$

This will be consistent with the identifications $\sigma^i \sim \sigma^i + 1$ iff $a, b, c, d \in \mathbb{Z}$, and will be invertible if

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{Z}) \tag{6.53}$$

Finally, it will preserve orientation if it is in $SL(2, \mathbb{Z})$, that is, if $ad - bc = 1$. We claim this diffeomorphism cannot be smoothly deformed to the identity if the matrix is not the identity. To see this consider its action on the homotopically nontrivial loops on the torus.

So, we can generalize the above construction by considering an arbitrary knot $K \subset S^3$. Again, its tubular neighborhood is topologically just a solid torus, as is the complement - but the two bagels must be glued together by a nontrivial diffeomorphism.

If we consider a higher genus surface Σ there is similarly an infinite number of nontrivial self-diffeomorphisms.

The idea of a solid torus can be generalized to the notion of a *handlebody*: A connected three-manifold with a surface Σ as boundary.

FIGURE

Note that in a genus g handlebody, g of the generators of π_1 become topologically trivial and the π_1 is just the free group on g generators.

We can now generalize the above construction by taking two genus g handlebodies, together with a diffeomorphism $\phi \in \mathrm{Diff}(\Sigma)$ and glue the handlebodies together. The resulting three-manifold is said to have a Heegaard decomposition. This means we can find an embedded closed oriented surface $\Sigma \subset Y$ so that Y is the gluing of two handlebodies

(bordisms of Σ to the emptyset) using a diffeomorphism of Σ. Such a Heegard decomposition strongly constraints the fundamental group, thanks to the Seifert-van-Kampen theorem: A set of generators is given by a set of generators of the two handlebodies. The fundamental group of the handlebody for Σ of genus g is just a free group on g generators. (The "A-cycles" contract to a point, leaving the "B-cycles" with no relation.) After gluing to the other handlebody there will be g relations expressing the contractibility of the new "A-cycles" of the second handlebody.

In fact, any three-manifold has a Heegaard decomposition: By a theorem of Moise, every 3-manifold can be triangulated. Take the 1-skeleton and thicken it to get a handlebody. This shows that every three-manifold admits a Heegaard decomposition. The trouble is, there are a lot of nontrivial diffeomorphisms of a surface to itself, and it can be hard to recognize two equivalent 3-folds constructed from different Heegaard splittings.

Thus, any three-manifold Y, $\pi_1(Y, y_0)$ admits a presentation where the number of generators is equal to the number of relations. The existence of such a presentation is not possible for a general finitely generated group. So, at least in three dimensions it is not true that any finitely generated group is the fundamental group of some three-dimensional manifold.

6.1.5 Fundamental Groups Of Four-Dimensional Manifolds

A basic fact is the following theorem of Markov:

Theorem Any finitely generated group G is the fundamental group of some four-manifold.

Proof: Suppose the group G has presentation:

$$G \cong \langle g_1, \ldots, g_n | R_1, \ldots, R_m \rangle \tag{6.54}$$

We aim to produce a four-manifold M_4 with fundamental group G. First, consider the free group on one generator $\langle g \rangle \cong \mathbb{Z}$. A good manifold that has this as a fundamental group is $X_4 = S^1 \times S^3$. Now let us consider $\tilde{M}_4 := X_4 \# \cdots \# X_4$. Then

$$\pi_1(\tilde{M}_4) \cong \langle g_1, \ldots, g_n \rangle \tag{6.55}$$

is the free group with generators g_i corresponding to the simple loops around the S^1 factor in each summand (extended to some common basepoint). Now, each relation R_α is a word in the g_i and thus can be represented by some closed based loop $\ell_\alpha \subset \tilde{M}_4$. We can take the ℓ_α to be nonintersecting, by simple codimension arguments. Now, take a tubular neighborhood $N(\ell_\alpha)$ of ℓ_α. By our discussion above of the local picture of submanifolds it is diffeomorphic to $N(\ell_\alpha) \cong S^1 \times D^3$, where D^3 is the 3-dimensional ball. The boundary is thus $\partial N(\ell_\alpha) \cong S^1 \times S^2$. This is also the boundary of $D^2 \times S^2$. So, glue in a copy of $D^2 \times S^2$ along the boundary of $N(\ell_\alpha)$. This procedure is known as *surgery*. Now the loop S^1 in $S^1 \times D^3$ (which was representing the word R_α) becomes contractible! Thus it is a relation on the generators g_i in the new manifold. We can choose the tubular neighborhoods around the different loops ℓ_α to be nonintersecting, and hence we can perform surgeries on each of these loops without interference. If we do this for all the loops we produce our manifold

♣Need to explain why it is a free group and not a free abelian group! ♣

M_4. By the Seifert-van Kampen theorem it follows that the fundamental group of M_4 is exactly G. ♠

Since finitely presented groups cannot be classified it follows that four-manifolds cannot be classified, even up to homotopy type.

7. Cosets and conjugacy

7.1 Lagrange Theorem

The reader should refresh her/his memory about equivalence relations - see section 1.1.

Definition 7.1.1: Let $H \subseteq G$ be a subgroup. The set

$$gH \equiv \{gh | h \in H\} \subset G \tag{7.1}$$

is called a *left-coset* of H.

Example 1: $G = \mathbb{Z}, H = 2\mathbb{Z}$. There are two cosets: H and $H + 1$.

Example 2: $G = S_3$, $H = \{1, (12)\} \cong S_2$. Cosets:

$$\begin{aligned} 1 \cdot H &= \{1, (12)\} \\ (12) \cdot H &= \{(12), 1\} = \{1, (12)\} = H \\ (13) \cdot H &= \{(13), (123)\} \\ (23) \cdot H &= \{(23), (132)\} \\ (123) \cdot H &= \{(123), (13)\} = \{(13), (123)\} = (13) \cdot H \\ (132) \cdot H &= \{(132), (23)\} = \{(23), (132)\} = (23) \cdot H \end{aligned} \tag{7.2}$$

Claim: Two left cosets are either *identical* or *disjoint*. Moreover, every element $g \in G$ lies in some coset. That is, the cosets define an equivalence relation by saying $g_1 \sim g_2$ if there is an $h \in H$ such that $g_1 = g_2 h$. Here's a proof written out in excruciating detail. [44]

First, g is in gH, so every element is in *some* coset. Second, suppose $g \in g_1 H \cap g_2 H$. Then $g = g_1 h_1$ and $g = g_2 h_2$ for some $h_1, h_2 \in H$. This implies $g_1 = g_2 (h_2 h_1^{-1})$ so $g_1 = g_2 h$ for an element $h \in H$. (Indeed $h = h_2 h_1^{-1}$, but the detailed form is not important.) By the rearrangement lemma $hH = H$, and hence $g_1 H = g_2 H$.

The basic principle above leads to a fundamental theorem:

Theorem 7.1.1 (Lagrange) If H is a subgroup of a finite group G then the order of H divides the order of G:

$$|G|/|H| \in \mathbb{Z}_+ \tag{7.3}$$

Proof : If G is finite $G = \amalg_1^m g_i H$ for some set of g_i, leading to *distinct* cosets. Now note that the order of any coset is the order of H:

$$|g_i H| = |H| \tag{7.4}$$

[44] In general, the reader should provide these kinds steps for herself or himself and we will not spell out proofs in such detail.

So $|G|/|H| = m$, where m is the number of distinct cosets. ♠

This theorem is simple, but powerful: For example we have the following

Corollary: Any finite group of prime order p is isomorphic to $\mu_p \cong \mathbb{Z}_p$. Moreover, such groups have exactly two subgroups: The trivial group and itself.

Proof: Choose a nonidentity element $g \in G$ and consider the subgroup generated by g i.e,

$$\{1, g, g^2, g^3, \dots\} \tag{7.5}$$

The order of this group must divide $|G|$ so if $|G| = p$ is prime it must be the entire group. ♠

Definition 7.1.2: If G is any group and H any subgroup then the *set of left cosets of H in G* is denoted G/H. It is the set of orbits under the right H action on G. A set of the form G/H is also referred to as a *homogeneous space*. The order of this set is the *index of H in G*, and denoted $[G : H]$.

Example 1: If $G = S_3, H = \{1, (12)\} \cong S_2$, then $G/H = \{H, (13) \cdot H, (23) \cdot H\}$, and $[G : H] = 3$.

Example 2: Let $G = \{1, \omega, \omega^2, \dots, \omega^{2N-1}\} = \mu_{2N}$ where ω is a primitive $(2N)^{th}$ root of 1. Let $H = \{1, \omega^2, \omega^4, \dots, \omega^{2N-2}\} = \mu_N$. Then $[G : H] = 2$ and $G/H = \{H, \omega H\}$.

Example 3: Let $G = A_4$ and $H = \{1, (12)(34)\} \cong \mathbb{Z}_2$. Then $[G : H] = 6$ and

$$G/H = \{H, (13)(24) \cdot H, (123) \cdot H, (132) \cdot H, (124) \cdot H, (142) \cdot H\} \tag{7.6}$$

Remark Note well! If $H \subset G$ is a subgroup and $g_1 H = g_2 H$ it does <u>not</u> follow that $g_1 = g_2$. All you can conclude is that there is some $h \in H$ with $g_1 = g_2 h$.

Exercise *Is there a converse to Lagrange's theorem?*

Suppose $n||G|$, does there then exist a subgroup of G of order n? Not necessarily! Find a counterexample. That is, find a group G and an n such that n divides $|G|$, but G has no subgroup of order n. [45]

[45] *Answer:* One possible example is A_4, which has order 12, but no subgroup of order 6. By examining the table of groups below we can see that this is the example with the smallest value of $|G|$. Sylow's theorem (discussed below) states that if a prime power p^k divides $|G|$ then there is in fact a subgroup of order p^k. This fails for composite numbers - products of more than one prime. Indeed, the smallest composite number is $6 = 2 \cdot 3$. Thus, in regard to a hypothetical converse to Lagrange's theorem, as soon as things can go wrong, they do go wrong.

Nevertheless, there is a very powerful theorem in group theory known as

Theorem 7.1.2: (Sylow's (first) theorem). Suppose p is prime and p^k divides $|G|$ for a nonnegative integer k. Then there is a subgroup $H \subset G$ of order p^k.

Herstein's book, sec. 2.12, waxes poetic on the Sylow theorems and gives three proofs. We'll give a proof as an application of the class equation in section 9 below. Actually, Sylow has a bit more to say. We will explain some more about this in the next section.

Definition: Thus far we have repeatedly spoken of the "order of a group G" and of various subsets of G, meaning simply the cardinality of the various sets. In addition a common terminology is to say that an element $g \in G$ *has order* n if n is the smallest natural number such that $g^n = 1$.

Note carefully that if g has order n and k is a natural number then $(g^n)^k = g^{nk} = 1$ and hence if $g^m = 1$ for some natural number m it does not necessarily follow that g has order m. However, as an application of Lagrange's theorem we can say the following: *If G is a finite group then the order of g must divide $|G|$, and in particular $g^{|G|} = 1$.* The proof is simple: Consider the subgroup generated by g, i.e. $\{1, g, g^2, \dots\}$. The order of this subgroup is the same as the order of g.

Exercise *Subgroups of A_4*

Write down all the subgroups of A_4. Draw a diagram indicating how these are subgroups of each other.

Exercise *Orders of group elements in infinite groups*

a.) Give an example of an infinite group in which all elements, other than the identity, have infinite order. (This should be quite easy for you.) [46]

b.) Give an example of an infinite group where some group elements have finite order and some have infinite order. [47]

c.) Give an example of an infinite group where all elements have finite order. [48]

[46] *One possible answer*: Take $\mathbb{Z}$ or $\mathbb{Z}^n$ or ....

[47] *One possible answer*: $\mathbb{Z} \times \mathbb{Z}_N$. Another possible answer is $G = U(1)$.

[48] *One possible answer*: Regard $U(1)$ as the group of complex numbers of modulus one. Let G be the subgroup of complex numbers so that $z^N = 1$ for some integer N. This is the group of all roots of unity of any order. It is clearly an infinite group, and by its very definition every element has finite order. Using the notation of the next section, this group is isomorphic to $\mathbb{Q}/\mathbb{Z}$.

Exercise

Suppose a finite group G has subgroups H_i, $i = 1, \ldots, s$ of order h_i where the h_i are all mutually relatively prime integers. Show that $\prod_i h_i$ divides $|G|$.

7.2 Conjugacy

Now introduce a notion generalizing the idea of similarity of matrices:

Definition 7.2.1 :

a.) A group element h is *conjugate* to h' if $\exists g \in G \qquad h' = ghg^{-1}$.

b.) Conjugacy defines an equivalence relation and the *conjugacy class of* h is the equivalence class under this relation:

$$C(h) := \{ghg^{-1} : g \in G\} \tag{7.7}$$

c.) Let $H \subseteq G, K \subseteq G$ be two subgroups. We say "H is conjugate to K" if $\exists g \in G$ such that

$$K = gHg^{-1} := \{ghg^{-1} : h \in H\} \tag{7.8}$$

Example 7.2.1 : We showed above that all cyclic permutations in S_n are conjugate.

Example 7.2.2 : Consider $G = U(N)$. Then conjugacy is the same notion as similarity of matrices. It is a simple consequence of an important theorem - the Spectral theorm that if $u \in U(N)$ there is a $g \in U(N)$ with $gug^{-1} = Diag\{z_1, \ldots, z_N\}$ where $|z_i| = 1$. One proves this by induction on N. See section 17 of Chapter 2.

This does not mean that the set of conjugacy classes can be identified with $U(1)^N$. Consider, for example, the permutation matrix $A(\phi)$ for $\phi \in S_N$. This is unitary and

$$A(\phi)Diag\{z_1, \ldots, z_N\}A(\phi)^{-1} = Diag\{z_{\phi(1)}, \ldots, z_{\phi(N)}\} \tag{7.9}$$

However, once we have taken this into account we are done: *The set of conjugacy classes in* $U(N)$ *is the set of unordered* N*-tuples of phases.*. We can make this assertion plausible by noting that all the traces $\mathrm{Tr}(u^k)$ are invariant under conjugation, so two conjugate diagonal matrices must provide simultaneous solutions to

$$\sum_{i=1}^{N} z_i^k = w_k \tag{7.10}$$

for $k \in \mathbb{Z}$. This can only be the case if they are related by permutation.

The set of conjugacy classes is therefore a space of orbits $U(1)^N/S_N$.

Now suppose we have two commuting unitary matrices. Again, basic linear algebra (explained in Chapter two) shows that they can be simultaneously diagonalized. That is, if $u_1, u_2 \in U(N)$ and $[u_1, u_2] = 1$ (group commutator) then there is a single $g \in U(N)$ with

$$gu_i g^{-1} = D_i \tag{7.11}$$

with D_i diagonal. Now by induction if we have a maximal Abelian subgroup of $U(N)$ then they can all be simultaneously diagonalized. So: Every maximal Abelian subgroup of $U(N)$ is conjugate to the subgroup of diagonal unitary matrices, and this group is isomorphic to $U(1)^N$.

In general, for any compact Lie group G the *maximal torus* is the largest commuting subgroup of G, and it is an important theorem, generalizing the spectral theorem, that they are all conjugate subgroups.

Example 7.2.3 : Now consider $G = GL(n, \mathbb{C})$. We must stress that not all matrices are diagonalizable, so that the full description of conjugacy classes is more complicated. For any matrix $A \in M_n(\mathbb{C})$ we can define its *characteristic polynomial*

$$p_A(x) := \det(x1 - A) \tag{7.12}$$

Note that p_A only depends on the conjugacy class of A:

$$p_{gAg^{-1}}(x) = p_A(x) \tag{7.13}$$

If r is a root of this polynomial then the matrix $r1 - A$ has zero determinant, so it has a nontrivial kernel (see Chapter 2) and therefore there is an eigenvector v of A with eigenvalue r:

$$Av = rv \tag{7.14}$$

It is very important to note that the eigenvectors might not form a basis,. Here is a simple and basic example:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{7.15}$$

Then one easily checks $p_A(x) = x^2$. So an eigenvector would satisfy $Av = 0$. But if there were a basis of eigenvectors then $A = 0$, a contradiction. The general statement is that any matrix is conjugate to its *Jordan canonical form*. See section 10.4 of Chapter 2. Briefly, we define the Jordan block with eigenvalue λ;

$$J_\lambda^{(k)} = \lambda 1 + N^{(k)} \tag{7.16}$$

$$N^{(k)} = e_{1,2} + e_{2,3} + \cdots + e_{k-1,k} \tag{7.17}$$

Let A be any complex $N \times N$ matrix and $p_A(x) = \prod_{i=1}^s (x - \lambda_i)^{k_i}$ where λ_i are the distinct roots. Then A is conjugate to a block form $A_{\lambda_1} \oplus \cdots A_{\lambda_s}$, where each A_{λ_i} is a block diagonal matrix of Jordan blocks

$$A_{\lambda_i} = \sum_{a=1}^{k_i} J_{\lambda_i}^{(n_{a,i})} \tag{7.18}$$

Various permutation matrices can act preserving the Jordan decomposition.

Remark/Definitions: We say that two homomorphisms $\varphi_i : H \to G$ are conjugate if there is an element $g \in G$ such that

$$\varphi_2(h) = g\varphi_1(h)g^{-1} \tag{7.19}$$

for all $h \in H$. Recall that a matrix representation of a group G is a homomorphism

$$\varphi : G \to GL(n, \kappa) \tag{7.20}$$

We say two matrix representations are *equivalent representations* if the two homomorphisms are conjugate.

Definition: A *class function* on a group is a function f on G (it can be valued in any set) such that f takes the same values on conjugate group elements:

$$f(gg_0g^{-1}) = f(g_0) \tag{7.21}$$

for all $g_0, g \in G$. Note particularly that if φ is a matrix representation then

$$\chi_\varphi(g) := \mathrm{Tr}\varphi(g) \tag{7.22}$$

is an example of a class function. This function is called the *character of the representation.* Note that two equivalent representations must have the same character.

Exercise *Conjugacy Is An Equivalence Relation*

a.) Show that conjugacy is an equivalence relation

b.) Prove that if H is a subgroup of G then gHg^{-1} is also a subgroup of G using the multiplication structure on G.

Exercise *Characters Of A Permutation Representation*

Consider the n-dimensional representation of S_n given by $T(\sigma) : e_i \mapsto e_{\sigma(i)}$ where e_i is the standard basis of $\mathbb{R}^n$. Show that the character of this representation is

$$\chi(\sigma) = N(\sigma) = |\{i : \sigma(i) = i\}| \tag{7.23}$$

As we will see later, this is the number of fixed points of σ.

Exercise *Rotations In $SO(3)$* . Consider a subgroup of $SO(3)$ defined by rotations around some axis in $\mathbb{R}^3$. Show that all such subgroups are conjugate subgroups.

Exercise *Conjugacy Classes In $SU(2)$*

a.) Using the spectral theorem show that the conjugacy class of a matrix $u \in SU(2)$ is uniquely determined by its trace. [49]

b.) Show that the set of conjugacy classes in $SU(2)$ can be identified with $S^1/\mathbb{Z}_2 = [0, \pi]$.

c.) Show that the most general continuous homomorphism $\varphi : U(1) \to SU(2)$ looks like

$$\varphi : z \to \begin{pmatrix} |\alpha|^2 z + |\beta|^2 z^{-1} & \alpha\beta(z^{-1} - z) \\ \alpha^*\beta^*(z^{-1} - z) & |\alpha|^2 z^{-1} + |\beta|^2 z \end{pmatrix} \tag{7.24}$$

where $(\alpha, \beta) \in \mathbb{C}^2$ satisfy $|\alpha|^2 + |\beta|^2 = 1$ and $z \in U(1)$. (Hint: Show that all continuous homomorphisms $\varphi : U(1) \to SU(2)$ are conjugate.)

d.) Show that the conjugacy class of a matrix $A \in M_2(\mathbb{C})$ is <u>not</u> uniquely determined by the values of $\mathrm{Tr}(A^k)$, but it is if A is diagonalizable. ($k = 1, 2$ will suffice.)

Exercise *The Complex Conjugate Representation*

a.) Consider the two 2-dimensional representations of $SU(2)$ where φ_1 is the identity and $\varphi_2(u) = u^*$. Show that these are equivalent representations of $SU(2)$. [50]

b.) Consider the two 2-dimensional representations of $U(2)$ where φ_1 is the identity and $\varphi_2(u) = u^*$. Show that these are inequivalent representations of $U(2)$. [51]

c.) Consider the N dimensional representation of $SU(N)$ given by $\varphi_1(u) = u$ and $\varphi_2(u) = u^*$. Are these equivalent representations?

---

[49] *Answer*: By the spectral theorem u is conjugate to

$$\begin{pmatrix} e^{\mathrm{i}\theta} & 0 \\ 0 & e^{-\mathrm{i}\theta} \end{pmatrix}$$

Of course $\theta \sim \theta + 2\pi$. However conjugation by $\mathrm{i}\sigma^1$ shows that θ and $\theta + \pi$ define conjugate elements. So, $Tr(u) = 2\cos\theta$ determines θ up to $\theta \sim \theta + \pi$.

[50] *Answer*: They are equivalent: Conjugation by $\mathrm{i}\sigma^2$ is equivalent to complex conjugation in $SU(2)$:

$$(\mathrm{i}\sigma^2)u(\mathrm{i}\sigma^2)^{-1} = u^* \ . \tag{7.25}$$

[51] *Answer*: They are inequivalent. By the spectral theorem $u \in U(2)$ can be conjugated to a diagonal matrix $Diag\{z_1, z_2\}$ with $z_1, z_2 \in U(1)$. The character of φ_1 is $\chi_1(u) = z_1 + z_2$. The character of φ_2 is $\chi_2(u) = z_1^{-1} + z_2^{-1}$. For general elements of $U(2)$ these are different so the character functions are different.

Exercise *An Example Of Inequivalent Representations*

(To do this exercise you need to understand about tensor products. See Chapter 2, section 5.3.)

Consider the following four-dimensional representations of $SU(2)$:

$$\varphi_1(u) = \begin{pmatrix} u & 0 \\ 0 & u \end{pmatrix} \tag{7.26}$$

$$\varphi_2(u) = u \otimes u \tag{7.27}$$

Are these representations equivalent or inequivalent? [52]

7.3 Normal Subgroups And Quotient Groups

Groups which are self-conjugate are very special:

Definition 7.2.2: A subgroup $N \subseteq G$ is called a *normal* subgroup, or an *invariant* subgroup if

$$gNg^{-1} = N \qquad \forall g \in G \tag{7.28}$$

Sometimes this is denoted as $N \triangleleft G$.

Warning! Equation (7.28) does *not* mean that $gng^{-1} = n$ for all $n \in N$!

There is a beautiful theorem associated with normal subgroups. In general the set of cosets of a subgroup H in G, denoted G/H, does not have any natural group structure. [53] However, if H is normal something special happens:

Theorem 7.2.1. If $N \subset G$ is a normal subgroup then the set of left cosets $G/N = \{gN | g \in G\}$ has a natural group structure with group multiplication defined by:

$$(g_1 N) \cdot (g_2 N) := (g_1 \cdot g_2) N \tag{7.29}$$

Proof- left as an important exercise - see below.

Remarks:

[52] *Answer*: They are inequivalent. One can show this by computing characters of the two representations. If u is in the conjugacy class with $Tr(u) = 2\cos\theta$ then $\chi_1(u) = 4\cos\theta$ while $\chi_2(u) = 4(\cos\theta)^2$.

[53] Note that it might have many unnatural group structures. For example, if G/H is a finite set with n elements that we could choosely - arbitrarily!! - some one-one correspondence between the elements of G/H and the elements in any finite group with n elements and use this to define a group multiplication law on the set G/H. We hope the reader can appreciate how incredibly tasteless such a procedure would be. Technically, it is *unnatural* because it makes use of an arbitrary extra choice of one-one correspondence between the elements of G/H and the elements of some group.

1. All subgroups N of Abelian groups A are normal, and moreover the quotient group A/N is Abelian.

2. Groups of the form G/N are known as *quotient groups*. A very common source of error and confusion is to mix up quotient groups and subgroups. They are very different!

3. As an illustration of the previous remark note that if $T : G \to GL(n, \kappa)$ is a matrix representation of G and if $H \subset G$ is a subgroup then we can also restrict T to H to get an n-dimensional representation of H. However, if Q is a quotient of G it is not true in general that a representation of G determines a representation of Q. We can try to define $T(gN) = T(g)$, but this will only make sense if $T(n) = 1$ for every $n \in N$.

Example 7.2.1 Cyclic Groups For example $n\mathbb{Z} \subset \mathbb{Z}$ is normal, and the quotient group is $\mathbb{Z}/n\mathbb{Z}$. This is isomorphic to the cyclic group we have previously denoted as μ_n or $\mathbb{Z}_n$. So $\bar{r}$ is the equivalence class of an integer $r \in \mathbb{Z}$:

$$\bar{r} = r + n\mathbb{Z} \tag{7.30}$$

$$\bar{r} + \bar{s} = (r + s) + n\mathbb{Z} \tag{7.31}$$

Example 7.2.2 Quotients of $\mathbb{Z}^d$. Consider $G = \mathbb{Z}^d$ and let e_i be a standard basis, with 1 in the ith row and zeroes elsewhere. Let A_{ij} be a $d \times d$ matrix of integers and consider the elements:

$$f_i := \sum_{j=1}^{d} A_{ij} e_j \tag{7.32}$$

Consider the subgroup $H \subset G$ of all integral linear combinations of f_i:

$$H := \{\sum_{i=1}^{d} n_i f_i | n_i \in \mathbb{Z}\} \tag{7.33}$$

H is clearly a subgroup of G, so we can form the quotient group G/H. If $\det A \neq 0$ then in fact G/H is a finite group. One way to see this easily is to consider G as a subgroup of $\mathbb{Q}^d$, so that we can write

$$e_i = A^{-1}_{ij} f_j \tag{7.34}$$

with $A^{-1} \in GL(d, \mathbb{Q})$. Recall that $A^{-1} = (\det A)^{-1} Cof(A)$ where the cofactor matrix $Cof(A)$ is a matrix of minors, and therefore is a matrix of integers. Therefore $(\det A)e_i \in H$ and hence $\det A[e_i] = 0$ in the quotient group so every element of G/H has a representative of the form $[\sum_i x_i e_i]$ with $|x_i| < |\det A|$.

Actually, if we invoke a nontrivial theorem we can say much more: The matrix A can be put into *Smith normal form*. This means that there are matrices $S, T \in GL(d, \mathbb{Z})$,

♣Should give a sketch of a proof of SNF in a supplementary section. ♣

representing change of generators (i.e. change of basis of the $\mathbb{Z}$-module) of H and G so that

$$SAT = Diag\{\alpha_1, \ldots, \alpha_d\} \tag{7.35}$$

with $\alpha_i = d_i/d_{i-1}$ where $d_0 = 1$ and d_i for $i > 0$ is the g.c.d. of the $i \times i$ minors. (For a proof see the Wikipedia article.) Then

$$G/H \cong \mathbb{Z}_{\alpha_1} \times \cdots \mathbb{Z}_{\alpha_d} \tag{7.36}$$

Note that it has order $|G/H| = \det A$. Here is a good example of the difference between a quotient group and a subgroup: No nontrivial finite group will be a subgroup of $\mathbb{Z}^d$.

Example 7.2.3 Discriminant Group. Now consider an embedded lattice in $\mathbb{R}^d$ equipped with Euclidean inner product. This is the integral span of a collection $\{v_i\}$ of vectors. For simplicity we will assume it is full rank, that is, the $\{v_i\}$ form a basis for $\mathbb{R}^d$ over $\mathbb{R}$. We denote it by Λ, so

$$\Lambda := \{\sum_{i=1}^{d} n_i v_i | n_i \in \mathbb{Z}\} \subset \mathbb{R}^d \tag{7.37}$$

We define the *dual lattice* (closely related to the "reciprocal lattice" in solid state physics) as the set of vectors $w \in \mathbb{R}^d$ such that $w \cdot v \in \mathbb{Z}$ for all $v \in \Lambda$:

$$\Lambda^\vee := \{w \in \mathbb{R}^d | \forall v \in \Lambda \quad v \cdot w \in \mathbb{Z}\} \tag{7.38}$$

Now assume that Λ is an *integral lattice*. This means that the matrix of inner products $G_{ij} = v_i \cdot v_j$ is a $d \times d$ matrix of integers. (Note it is symmetric and of nonzero determinant.) Then it follows that $\Lambda \subset \Lambda^\vee$ is a sublattice. The *discriminant group* of Λ is the finite group

$$\mathcal{D} := \Lambda^\vee/\Lambda \tag{7.39}$$

Note that $\Lambda^\vee$ has a basis f_i with $v_i = G_{ij} f_j$ so one can work out $\mathcal{D}$ as a product of cyclic groups using the Smith normal form of G_{ij}.

Example 7.2.4. Let us now consider some nonabelian examples.

$$A_3 \equiv \{1, (123), (132)\} \subset S_3 \tag{7.40}$$

is normal. Note that

$$(12)A_3(12)^{-1} = A_3 \tag{7.41}$$

but conjugation by (12) induces a nontrivial permutation of the set A_3. The group S_3/A_3 has order 2 and hence must be isomorphic to $\mathbb{Z}_2$.

Example 7.2.5. Of course, in any group G the subgroup $\{1\}$ and G itself are normal subgroups. These are the trivial normal subgroups. It can happen that these are the only normal subgroups of G:

Definition . A group with no nontrivial normal subgroups is called a *simple group*.

Remarks

1. Note that a simple group cannot have a nontrivial center.

2. The term is a bit of a misnomer: Some simple groups are pretty darn complicated. What it means is that there is no means of simplifying it using something called the Jordan-Holder decomposition - discussed below. Simple groups are extremely important in the structure theory of finite groups. One example of simple groups are the cyclic groups $\mathbb{Z}/p\mathbb{Z}$ for p prime. Can you think of others?

3. *Sylow's theorems again.* Recall that Sylow's first theorem says that if p^k divides $|G|$ then G has a subgroup of order p^k. If we take the largest prime power dividing $|G|$, that is, if $|G| = p^k m$ with m relatively prime to p then a subgroup of order p^k is called a *p-Sylow subgroup*. Sylow's second theorem states that all the p-Sylow subgroups are conjugate. The third Sylow theorem says something about how many p-Sylow subgroups there are.

4. **WARNING!**: In the theory of Lie groups you will find the term "simple Lie group." A simple Lie group is NOT a simple group in the sense we defined above ! For example $SU(2)$ is a simple Lie group. But it has a nontrivial center namely the two diagonal $SU(2)$ matrices $\{\pm 1_{2\times 2}\}$.

Example 7.2.6. Recall that $SL(n,\kappa) \subset GL(n,\kappa)$, $SO(n,\kappa) \subset O(n,\kappa)$, and $SU(n) \subset U(n)$ are all subgroups defined by the condition $\det A = 1$ on a matrix. Note that, since $\det(gAg^{-1}) = \det A$ for any invertible matrix g these are in fact normal subgroups. The quotient groups are

$$\begin{aligned} GL(n,\kappa)/SL(n,\kappa) &\cong \kappa^* \\ O(n,\mathbb{R})/SO(n,\mathbb{R}) &\cong \mathbb{Z}_2 \\ U(n)/SU(n) &\cong U(1) \end{aligned} \tag{7.42}$$

Lines 1 and 3 follow since every element in $GL(n,\kappa)$ can be written as zA with $z \in \kappa^*$ and $A \in SL(n,\kappa)$. For line 2 take P to be any reflection in any hyperplane orthogonal to some vector v, then $O(n,\mathbb{R}) = SO(n,\mathbb{R}) \amalg PSO(n,\mathbb{R})$ because $\det P = -1$. Recall that $P_{v_1} P_{v_2}$ is a rotation in the plane spanned by v_1, v_2, so it doesn't matter which hyperplane we choose.

Example 7.2.7. Let G be a topological group. Let G_0 be the (path-) connected component of the identity element $1_G \in G$. We claim that G_0 is a normal subgroup: If $g_0 \in G_0$ there is a continuous path of group elements $\gamma : [0,1] \to G$ with $\gamma(0) = 1_G$ and $\gamma(1) = g_0$. Then if $g \in G$ is any other group element $g\gamma(t)g^{-1}$ is a continuous path connecting 1_G to gg_0g^{-1}. The quotient group G/G_0 is the *group of components*, sometimes denoted $\pi_0(G)$. In general for a topological space X, the set of connected components is denoted by $\pi_0(X)$, but it carries no natural group structure.

For some examples:

1. $G = \mathbb{R}^*$ under multiplication. The identity element is 1 and clearly $\mathbb{R}^*_{>0}$ is the connected component of the identity. The quotient group is isomorphic to $\mathbb{Z}_2$.

2. The determinant map defines a homomorphism $\det : GL(n,\mathbb{R}) \to \mathbb{R}^*$. Clearly there is no path of $GL(n,\mathbb{R})$ matrices that connects elements with $\det A < 0$ to $\det A > 0$. (Why?) In fact, the subgroup of $GL(n,\mathbb{R})$ of matrices with positive determinant is connected: The sign of the determinant is the only obstruction to deformation to the identity. [54]

3. Very similar considerations hold for $O(n,\mathbb{R})$. One can show that $SO(n,\mathbb{R})$ is the connected component of the identity and $\pi_0 \cong \mathbb{Z}_2$. For example, consider $O(2)$. In an exercise above you showed that as a manifold it has two components, each of which can be identified with a circle. The connected component of the identity is $SO(2)$. We have $O(2) = SO(2) \amalg SO(2)P$ where P is any $O(2)$ matrix of determinant $= -1$. So $\pi_0(O(2)) \cong \mathbb{Z}_2$. Similarly, $\pi_0(O(n)) \cong \mathbb{Z}_2$.

4. One can show that $\pi_0(Diff(T^2)) \cong GL(2,\mathbb{Z})$. Indeed there is a subgroup of $Diff(T^2)$ isomorphic to $GL(2,\mathbb{Z})$ of diffeomorphisms

$$\begin{pmatrix} \sigma^1 \\ \sigma^2 \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \sigma^1 \\ \sigma^2 \end{pmatrix} \tag{7.43}$$

   which projects isomorphically to the quotient group.

5. If G is a finite group then $\pi_0(G) \cong G$.

Example 7.2.8. The center of $U(N)$ consists of matrices proportional to the unit matrix. See the exercise below. Elements in the center of $SU(N)$ must also be diagonal. However, now if $z1_{N\times N}$ is to be in $SU(N)$ then $z^N = 1$ (why?) so $Z(SU(N)) \cong \mu_N \cong \mathbb{Z}_N$. Since this subgroup is normal we can take a quotient and get another group. It is known as

$$PSU(N) := SU(N)/\mathbb{Z}_N \tag{7.44}$$

One can show that $PSU(N) \cong U(N)/Z(U(N)) \cong U(N)/U(1)$. There are representations of $SU(N)$ that are not representations of $PSU(N)$ so here is another example where $PSU(N)$ cannot be considered as a subgroup of $SU(N)$ in any sense.

Exercise *Due Diligence*

[54] The theory of fiber bundles shows that if H is a Lie subgroup of G so that G/H is a topological space then if H is connected the set of components of G can be identified with the set of components of G/H. We will see later from the stabilizer orbit theorem that $SO(n+1)/SO(n) = S^n$, so we can prove $SO(n)$ is connected by induction on n. By Gram-Schmidt procedure $SL(n,\mathbb{R})$ is a product of $SO(n)$ and upper triangular matrices with unit diagonal - and this space is connected. So $SL(n,\mathbb{R})$ is connected. Then the set of components of $GL(n,\mathbb{R})$ is that of $\mathbb{R}^*$, and this is measured by the determinant.

a.) Check the details of the proof of Theorem 7.2.1 ! [55]

b.) Consider the *right cosets*. Show that $N\backslash G$ is a group.

Exercise *Even Permutations*

Example 7.2.2 has a nice generalization. Recall that a permutation is called *even* if it can be written as a product of an even number of transpositions.

a.) Show that the even permutations, A_n, form a normal subgoup of S_n.

b.) What is S_n/A_n?

Exercise *Subgroups Of Index Two*

a.) Suppose that $H \subset G$ is of index two: $[G : H] = 2$. Show that H is normal in G. What is the group G/H in this case? [56]

b.) Using (a) give another proof that $A_n \triangleleft S_n$ is a normal subgroup.

c.) As we will discuss later, the groups A_n for $n \geq 5$ are simple groups. Accepting this for the moment give an infinite set of counterexamples to the converse of Lagrange's theorem. [57]

Exercise

Look at the 3 examples of homogeneous spaces G/H in section 7.1. Decide which of the subgroups H is normal and what the group G/H would be.

---

[55] *Answer*: The main thing to check is that the product law defined by (7.29) is actually well defined. Namely, you must check that if $g_1N = g_1'N$ and $g_2N = g_2'N$ then $g_1g_2N = g_1'g_2'N$. To show this note that $g_1' = g_1n_1$ and $g_2' = g_2n_2$ for some $n_1, n_2 \in N$. Now note that $g_1'g_2' = g_1n_1g_2n_2 = g_1g_2(g_2^{-1}n_1g_2)n_2$. But, since N is normal $(g_2^{-1}n_1g_2) \in N$ and hence $(g_2^{-1}n_1g_2)n_2 \in N$ and hence indeed $g_1g_2N = g_1'g_2'N$. Once we see that (7.29) is well-defined the remaining checks are straightforward. Essentially all the basic axioms are inherited from the group law for multiplying g_1 and g_2. Associativity should be obvious. The identity is $1_GN = N$ and the inverse of gN is $g^{-1}N$. etc. ♠

[56] *Answer*: Suppose $G = H \amalg g_0H$. Then take any $h \in H$. The element $g_0hg_0^{-1}$ must be in H or g_0H. But if it were in g_0H then there would be an $h' \in H$ such that $g_0hg_0^{-1} = g_0h'$ but this would imply g_0 is in H, which is false. Therefore, for all $h \in H$, $g_0hg_0^{-1} \in H$, and hence H is a normal subgroup. Therefore $G/H \cong \mathbb{Z}_2$.

[57] *Answer*: Note that the order of $|A_n|$ is even and hence $\frac{1}{2}|A_n|$ is a divisor of $|A_n|$. However, a subgroup of order $|A_n|/2$ would have to be a normal subgroup, and hence does not exist, since A_n is simple. More generally, a high-powered theorem, known as the Feit-Thompson theorem states that a finite simple nonabelian group has even order. Therefore if G is a finite simple nonabelian group there is no subgroup of order $\frac{1}{2}|G|$, even though this is a divisor.

Exercise
Show that if the center $Z(G)$ is such that $G/Z(G)$ is cyclic then G is Abelian. [58]

Exercise *Sylow subgroups of A_4*
Write down the 2-Sylow and 3-Sylow subgroups of A_4.

Exercise *Commutator Subgroups And Abelianization*

If g_1, g_2 are elements of a group G then the *group commutator* is the element $[g_1, g_2] := g_1 g_2 g_1^{-1} g_2^{-1}$. If G is any group the *commutator subgroup* usually denoted $[G, G]$ (sometimes denoted G') is the subgroup generated by words in all group commutators $g_1 g_2 g_1^{-1} g_2^{-1}$.

a.) Show that $[G, G]$ is a normal subgroup of G. [59]

b.) Show that $G/[G, G]$ is abelian. This is called the *abelianization* of G. [60]

c.) Consider the free group on 2 generators. What is the abelianization?

d.) Consider a surface group of the type given in (6.42). The abelianization of this group is isomorphic to the *first homology group* $H_1(S)$ where S is the punctured surface. Compute this group. [61]

Exercise *A Less Than Perfect Group*

a.) Recall that a *simple* group is a group with no nontrivial normal subgroups. A *perfect* group is a group which is equal to its commutator subgroup. Show that a nonabelian simple group must be perfect.

b.) Show that S_n is not a perfect group. What is the commutator subgroup? [62]

---

[58] *Answer*: Every element of G would be of the form $g_0^n z$ with $z \in Z$. But then it is easy to check: $g_0^n z g_0^m z' = g_0^m z' g_0^n z$ so G is Abelian. So $G = Z(G)$, and in fact the cyclic subgoup must be trivial.

[59] *Answer*: Note that $g_0[g_1, g_2]g_0^{-1} = [g_1', g_2']$ where $g_i' = g_0 g_i g_0^{-1}$.

[60] *Answer*: Let $G' = [G, G]$ then $g_1 G' g_2 G' = (g_1 g_2)G' = g_2 g_1 (g_1^{-1} g_2^{-1} g_1 g_2)G' = g_2 g_1 G'$.

[61] One can define higher homology groups $H_k(X)$ of a topological space X but these in general are not Abelianiations of the higher homotopy groups $\pi_k(X)$, even though both groups are Abelian. Homology and homotopy groups measure different aspects of the topology of a space.

[62] The commutator subgroup is clearly a subgroup of A_n. In fact A_n is generated by products of two transpositions and hence is generated by (abc). But note that $(ab)(ac)(ab)(ac) = (abc)$. Therefore the commutator subgroup of S_n is just A_n.

Exercise *Signed Permutations Again*

Recall our discussion of a natural matrix representation of S_n and the group $W(B_n)$ of signed permutations from *** above.

a.) Show that the subgroup of $W(B_n)$ of diagonal matrices is a normal subgroup isomorphic to $\mathbb{Z}_2^n$.

b.) Show that every signed permutation matrix can be written in the form $D \cdot \Pi$ where D is a diagonal matrix of ± 1's and Π is a permutation matrix.

c.) Conclude that the quotient of $W(B_n)$ by the normal subgroup of diagonal matrices is isomorphic to S_n.

d.) Show that every signed permutation can also be written as $\Pi' \cdot D'$. How is this decomposition related to writing it as $D \cdot \Pi$.

Exercise *Products Of Simple Groups*

Let G_1 and G_2 be simple groups.

a.) What are the subgroups of the Cartesian product $G_1 \times G_2$? [63]

b.) Suppose G_i, $i \in I$ is a set of simple groups. What are the subgroups of $\prod_{i \in I} G_i$?

♣This exercise should go to the first section when we introduce the matrix groups. ♣

Exercise *The Center Of $U(N)$*

Show that the center of $U(N)$ consists of the subgroup of matrices proportional to the unit matrix and is therefore isomorphic to $U(1)$. [64]

Exercise *Subgroups Which Are Not So Normal*

[63] *Answer*: Only $\{1\}$, $G_1 \times \{1_{G_2}\}$, $\{1_{G_1}\} \times G_2$ and $G_1 \times G_2$.

[64] *Answer*: There are many proofs but one nice one is to use induction on N. First establish the result for $U(2)$ - here the matrix multiplication is easy and this can be done by hand. Now suppose that $\zeta \in U(N+1)$ is in the center. Decompose it as follows

$$\zeta = \begin{pmatrix} A & v \\ w & D \end{pmatrix}$$

where $A \in M_2(\mathbb{C})$, $v \in M_{2 \times N-1}(\mathbb{C})$, $w \in M_{N-1 \times 2}(\mathbb{C})$, and $D \in M_{N-1 \times N-1}(\mathbb{C})$. Now insist that it commute with

$$\begin{pmatrix} u & 0 \\ 0 & 1 \end{pmatrix}$$

with $u \in U(2)$ to show that $uAu^{-1} = A$, $uv = v$ and $wu^{-1} = w$ for all $u \in U(2)$. These equations imply A is diagonal and $v, w = 0$.

a.) Consider $O(n,\mathbb{R}) \subset GL(n,\mathbb{R})$. Is this a normal subgroup?

b.) Consider the subgroup of diagonal matrices in $SU(N)$. Is this a normal subgroup.

Exercise *The Normalizer Subgroup*

If $H \subset G$ is a subgroup then we define the *normalizer of H within G* to be the largest subgroup N of G such that H is a normal subgroup of N. Note that H is normal inside itself so such subgroups exist. If $N_1, N_2 \subset G$ are subgroups and H is normal in both then they generate a subgroup in which H is normal. In fact we have:

$$N_G(H) := \{g \in G | gHg^{-1} = H\} \tag{7.45}$$

a.) Show that (7.45) is a subgroup of G and H is a normal subgroup of $N_G(H)$.

b.) Show that $N_G(H)$ is the largest subgroup of G which contains H as a normal subgroup.

Note that there is no claim that $N_G(H)$ is a normal subgroup of G. In general, it is not.

Exercise *An Important Class of Quotient Groups: The Weyl Groups Of Lie Groups*

a.) Let $D \subset SU(2)$ be the subgroup of diagonal matrices. Note that $D \cong U(1)$. Compute

$$N_{SU(2)}(D) \tag{7.46}$$

explicitly. [65]

b.) Compute the quotient group $N_{SU(2)}(D)/D$. [66]

c.) Show that conjugation by elements of the normalizer act by a permutation of the diagonal elements and the permutation only depends on the projection to the quotient.

d.) Show that there is no subgroup of $N_{SU(2)}(D)$ whose conjugation on D induces the permutation action.

---

[65] *Answer*: The normalizer is the subgroup of $SU(2)$ that is the union of matrices of the form

$$\begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}$$

or of the form

$$\begin{pmatrix} 0 & -z^{-1} \\ z & 0 \end{pmatrix}$$

where z is a phase.

[66] *Answer* The quotient is isomorphic to $\mathbb{Z}_2$.

Remark: In general, in a simple Lie group G there is a unique maximal torus, $T \subset G$ up to conjugation. The Weyl group of G is by definition

$$W(G) := N_G(T)/T \tag{7.47}$$

For example, in $SU(n)$ any maximal torus is conjugate to the subgroup $D \subset SU(n)$ of diagonal matrices. In this case, conjugation by $N_{SU(n)}(D)$ acts on D by permutation of the diagonal elements and in fact

$$W(SU(n)) := N_{SU(n)}(D)/D \cong S_n \tag{7.48}$$

Note that the Weyl group is defined as a quotient of a subgroup of G. (Often this is abbreviated to "subquotient of G.") In general there is no subgroup of G whose conjugation action induces the Weyl group action on T. (It is a common mistake to confuse $W(G)$ with a subgroup of G.)

Exercise *Representations Of $SU(N)$ That Are Not Representations Of $PSU(N)$*

Give an example of a representation of $SU(N)$ that is not a representation of the quotient $PSU(N)$. [67]

Exercise *Homomorphic Images And Normal Subgroups*

Suppose $N \triangleleft G$, and suppose that $\varphi : G \to \tilde{G}$ is a homomorphism to some other group $\tilde{G}$. Then $\varphi(N) \subset \tilde{G}$ is a subgroup. Is it a normal subgroup? [68]

♣Your answer here should have a more explicit counterexample. ♣

7.3.1 A Very Interesting Quotient Group: Elliptic Curves

Consider the Abelian group $\mathbb{C}$ of complex numbers with normal addition as the group operation. If τ is a complex number with nonzero imaginary part then $\mathbb{Z} + \tau\mathbb{Z}$ is the subgroup of complex numbers of the form $n_1 + \tau n_2$ where n_1 and n_2 are integers. Since $\mathbb{C}$ is abelian we can form the Abelian group $\mathbb{C}/\mathbb{Z}+\tau\mathbb{Z}$. Note that $\mathbb{Z}+\tau\mathbb{Z}$ is a rank two lattice in the plane so that this quotient space can be thought of as a torus. As an Abelian group this group is isomorphic to $U(1) \times U(1)$. The explicit isomorphism is

$$(\sigma_1 + \tau\sigma_2) + (\mathbb{Z} + \tau\mathbb{Z}) \mapsto (e^{2\pi i\sigma_1}, e^{2\pi i\sigma_2}) \tag{7.49}$$

[67] The defining representation is not, because the center of $SU(N)$ acts nontrivially.

[68] *Answer*: In general it is not a normal subgroup. However, if φ is surjective then it is easy to see that it is a normal subgroup

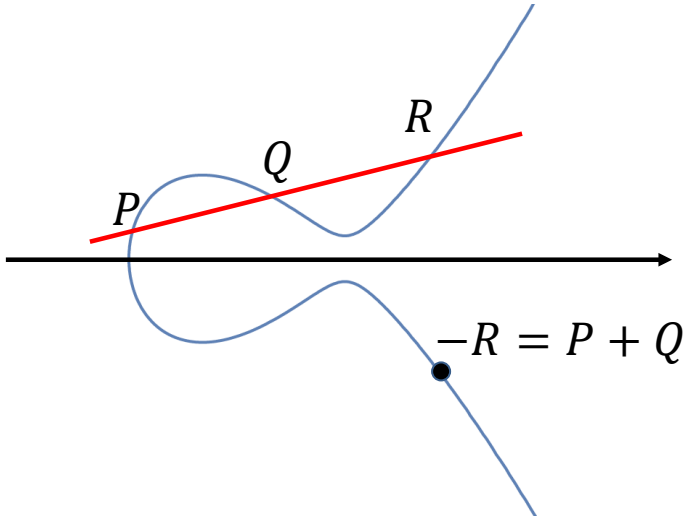


Figure 19: In a suitable range of real values of f, g the real points on the elliptic curve have the above form. Then the elliptic curve group law is easily pictured as shown.

A remarkable fact is that this torus (minus one point) can be thought of as the space of solutions of the algebraic equation

$$y^2 = x^3 + fx + g \tag{7.50}$$

where $(x, y) \in \mathbb{C}^2$ and $f, g \in \mathbb{C}$. [69]

The mapping between $[z] \in \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ and (x, y) and between f, g and the complex number τ involves very interesting functions known as elliptic functions. The solution set is known as an "elliptic curve." It is not difficult to describe the mapping. One introduces a holomorphic function of z known as the *Weierstrass function*:

$$\wp(z|\tau) := \frac{1}{z^2} + \sum_{\omega \in \Lambda - \{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right) \tag{7.51}$$

where $\omega = n_1 + n_2\tau \in \Lambda := \mathbb{Z} + \tau\mathbb{Z}$ and $z \notin \Lambda$ Note that for large values of ω the summand behaves like $\frac{2z}{\omega^3}$, so the series converges absolutely since $\int \frac{dxdy}{r^3}$ is convergent at $r \to \infty$. By general results in complex analysis the function is holomorphic for $z \in \mathbb{C} - \Lambda$. Note that it is also doubly-periodic:

$$\wp(z + m + m'\tau|\tau) = \wp(z|\tau) \tag{7.52}$$

for all $m, m' \in \mathbb{Z}$. So it descends to a function on the quotient to define a meromorphic function on a complex manifold.

[69] We can restore the point at infinity using projective geometry. The equation $ZY^2 = X^3 + fXZ^2 + gZ^3$ makes sense for a point $[X : Y : Z] \in \mathbb{CP}^2$. Indeed note that the equivalence relation says that $[X : Y : Z] = [\lambda X : \lambda Y : \lambda Z]$ and the equation is homogeneous and of degree three. The equation (7.50) is the equation we get in the patch $Z \neq 0$ where we can fix the scaling degree of freedom by choosing λ so that $Z = 1$. We then define x, y by $[x : y : 1] = [X : Y : Z] = [X/Z : Y/Z : 1]$ which makes sense when $Z \neq 0$. The point at infinity has $Z = 0$. Therefore, by the equation $X = 0$, and since we have a point in $\mathbb{CP}^2$ we must have $Y \neq 0$, which can therefore be scaled to $y = 1$. So, the point at infinity is $[0 : 1 : 0] \in \mathbb{CP}^2$.

For $[z] = 0$ the function has a second order pole. Put differently, the Weierstrass function has a double pole at every point $z \in \Lambda$. Indeed we can expand $\wp(z|\tau)$ around $z = 0$:

$$\begin{aligned} \wp(z|\tau) &= \frac{1}{z^2} + \sum_{k=1}^{\infty} (2k+1) G_{2k+2} z^{2k} \\ &= \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + \cdots \end{aligned} \tag{7.53}$$

where

$$G_{2k+2} = \sum_{(n_1,n_2)\in\mathbb{Z}^2-\{(0,0)\}} \frac{1}{(n_1+n_2\tau)^{2k+2}} \tag{7.54}$$

are absolutely convergent and hence holomorphic functions of τ for $k \geq 1$ when $\mathrm{Im}\tau \neq 0$. They are famous functions known as *Eisenstein functions* and are basic examples of a fascinating set of functions known as *modular forms*. In order to produce equation (7.50) we will take $x = \wp(z|\tau)$ and define

$$\begin{aligned} y &:= \frac{\partial}{\partial z} \wp(z|\tau) \\ &= \frac{-2}{z^3} + 6G_4 z + 20 G_6 z^3 + \cdots \end{aligned} \tag{7.55}$$

Now a small amount of algebra shows that we have the series expansion

$$y^2 - 4x^3 + 60 G_4 x = -140 G_6 + \mathcal{O}(z^2) \tag{7.56}$$

So the combination $y^2 - 4x^3 + 60G_4 x$ is entire, i.e., holomorphic in the entire complex plane, and doubly-periodic. But then by Liouville's theorem it must be constant! Thus all the higher terms in the series vanish! Thus we have:

$$y^2 = 4x^3 - 60 G_4 x - 140 G_6 \tag{7.57}$$

which is exactly the form (7.50) up to a simple rescaling.

The Abelian group law expressed in terms of (x, y) is rather nontrivial and closely related to some deep topics in number theory. If one considers f, g to be real and studies the real solutions then the group law can be visualized as in Figure 19. (We are following the Wikipedia article here, which is quite clear.) One first defines the inverse $-P$ of a point P on the curve with coordinates $P = (x, y)$ to be $-P := (x, -y)$. Then, for generic points we can define $P + Q$ by saying that $P + Q = -R$ where R is on the intersection of the straight line through P, Q and the elliptic curve. In formulae we can write the line between P, Q as

$$y = sx + d \tag{7.58}$$

with

$$s = \frac{y_P - y_Q}{x_P - x_Q} \qquad d = y_P - x_P \left(\frac{y_P - y_Q}{x_P - x_Q} \right) = y_Q - x_Q \left(\frac{y_P - y_Q}{x_P - x_Q} \right) \tag{7.59}$$

Now the intersection of this line with the cubic equation has x coordinates given by

$$(sx+d)^2 = x^3 + fx + g \tag{7.60}$$

and by simple rearrangement we can rewrite (7.60) as

$$x^3 - s^2x^2 + (f-2sd)x + (g-d^2) = 0 \tag{7.61}$$

On the other hand, this equation must be of the form

$$(x-x_P)(x-x_Q)(x-x_R) = 0 \tag{7.62}$$

Expanding out (7.61) and equating the coefficient of x^2 we obtain

$$x_R = s^2 - x_P - x_Q \tag{7.63}$$

so we have x_R explicitly as a function of x_P, x_Q, y_P, y_Q. Now the point (x_R, y_R) must lie on the line $y = sx + d$ so we can also say that

$$y_R = y_P + s(x_R - x_P) \tag{7.64}$$

expressing y_R and hence the coordinates of $R = (x_R, -y_R)$ as rational functions of x_P, x_Q, y_P, y_Q. It is not at all obvious that the above group law really satisfies the associativity constraint. When points coincide or the line is tangent to the elliptic curve one must carefully degenerate the above expressions. Indeed, requiring that $P + (-P) = 0$ shows that 0 must correspond to the point at infinity. When f, g are not in the range to give a figure like Figure 19 the algebraic equations above still define a group law. Indeed, these equations make sense over any field, thus allowing one to define an Abelian group law for elliptic curves defined over any field.

Exercise *Modular Transformations*

a.) Show that the transformation law

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : (z,\tau) \mapsto \left(\frac{z}{c\tau+d}, \frac{a\tau+b}{c\tau+d} \right) \tag{7.65}$$

defines a group action of $SL(2,\mathbb{Z})$ on pairs $\mathbb{C} \times \mathbb{H}$.

b.) Show that

$$\wp\left(\frac{z}{c\tau+d} | \frac{a\tau+b}{c\tau+d} \right) = (c\tau+d)^2 \wp(z|\tau) \tag{7.66}$$

c.) Show that

$$G_{2k}(\frac{a\tau+b}{c\tau+d}) = (c\tau+d)^{2k} G_{2k}(\tau) \qquad k = 2,3,4,... \tag{7.67}$$

d.) Using Fourier analysis prove that for $z \notin \mathbb{Z}$ we have

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^2} = \frac{\pi^2}{\sin^2(\pi z)} \tag{7.68}$$

and conclude that for $\mathrm{Im} z > 0$

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^{2k}} = \frac{(2\pi \mathrm{i})^{2k}}{(2k-1)!} \sum_{\ell=1}^{\infty} \ell^{2k-1} e^{2\pi \mathrm{i}\ell z} \tag{7.69}$$

From this derive that

$$G_{2k}(\tau) = 2\zeta(2k) + 2\frac{(2\pi \mathrm{i})^{2k}}{(2k-1)!} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n \tag{7.70}$$

where $\sigma_m(n) = \sum_{d \text{ divides } n} d^m$ and $q := e^{2\pi \mathrm{i}\tau}$. Furthermore show that

$$G_{2k}(\tau) = 2\zeta(2k) E_{2k}(\tau) \tag{7.71}$$

$$E_{2k}(\tau) = 1 - \frac{4k}{B_{2k}} \sum_{n=1}^{\infty} \frac{n^{2k-1} q^n}{1-q^n} \tag{7.72}$$

where B_{2k} is the Bernoulli number (as defined in Wikipedia).

7.4 Conjugacy Classes In S_n

Above we discussed the cycle decomposition of elements of S_n. Now let us study how the cycles change under conjugation.

When showing that transpositions generate S_n we noted the following fact:

If $(i_1 i_2 \cdots i_k)$ is a cycle of length k then $g(i_1 i_2 \cdots i_k) g^{-1}$ is a cycle of length k. It is the cycle where we replace $i_1, i_2, \ldots$ by their images under g. That is, if $g(i_a) = j_a$, $a = 1, \ldots, k$, then $g(i_1 i_2 \cdots i_k) g^{-1} = (j_1 j_2 \cdots j_k)$.

It therefore follows that:

Any two cycles of length k are conjugate.

Example In S_3 there are two cycles of length 3 and they are indeed conjugate:

$$(12)(123)(12)^{-1} = (213) = (132) \tag{7.73}$$

Now recall that any element in S_n can be written as a product of disjoint cycles.

Therefore, the conjugacy classes in S_n are labeled by specifying a nonnegative integer, denoted ℓ_j, where $j = 1, \ldots, n$, and ℓ_j is the number of distinct cycles of length j in the cycle decomposition of any typical element σ of $C(\sigma)$.

Examples:

1. The following two permutations in S_{12} are conjugate:

$$(1,2)(3,4)(5,6)(7,8,9)(10,11,12) \tag{7.74}$$

$$(4,10)(7,8)(9,11)(1,12,6)(2,5,3) \tag{7.75}$$

   This has $\ell_1 = 0$, $\ell_2 = 3$, $\ell_3 = 2$, $\ell_j = 0$ for $j > 3$.

2. In S_4 there are 3 elements with cycle decomposition of type $(ab)(cd)$:

$$(12)(34), \qquad (13)(24), \qquad (14)(23) \tag{7.76}$$

   Note that these can be conjugated into each other by suitable transpositions. So this conjugacy class is determined by

$$\ell_1 = 0 \qquad \ell_2 = 2 \qquad \ell_3 = 0 \qquad \ell_4 = 0 \tag{7.77}$$

In general we can denote a conjugacy class in S_n by:

$$(1)^{\ell_1}(2)^{\ell_2}\cdots(n)^{\ell_n} \tag{7.78}$$

Then, since we must account for all n letters being permuted we must have:

$$n = 1\cdot\ell_1 + 2\cdot\ell_2 + \cdots n\cdot\ell_n = \sum_{j=1}^{n} j\cdot\ell_j \tag{7.79}$$

Definition A decomposition of n into a sum of nonnegative integers is called a *partition of* n.

Therefore:

The conjugacy classes of S_n are in 1-1 correspondence with the partitions of n.

Definition The number of distinct partitions of n is called the partition function of n, and denoted $p(n)$. [70]

Example For $n = 4, 5$ $p(4) = 5$ and $p(5) = 7$ and the conjugacy classes of S_4 and S_5 are:

Partition	Cycle decomposition	Typical g	$\lvert C(g)\rvert$	Order of g
$4 = 1+1+1+1$	$(1)^4$	1	1	1
$4 = 1+1+2$	$(1)^2(2)$	(ab)	$\binom{4}{2} = 6$	2
$4 = 1+3$	$(1)(3)$	(abc)	$2\cdot 4 = 8$	3
$4 = 2+2$	$(2)^2$	$(ab)(cd)$	$\frac{1}{2}\binom{4}{2} = 3$	2
$4 = 4$	(4)	$(abcd)$	6	4

[70] This is a term in number theory. It is not to be confused with the "partition function" of a field theory!

Cycle decomposition	$\lvert C(g)\rvert$	Typical g	Order of g
$(1)^5$	1	1	1
$(1)^3(2)$	$\binom{5}{2} = 10$	(ab)	2
$(1)^2(3)$	$2 \cdot \binom{5}{3} = 20$	(abc)	3
$(1)(4)$	$6 \cdot \binom{5}{4} = 30$	$(abcd)$	4
$(1)(2)^2$	$5 \cdot \frac{1}{2}\binom{4}{2} = 15$	$(ab)(cd)$	2
$(2)(3)$	$2 \cdot \binom{5}{2} = 20$	$(ab)(cde)$	6
(5)	$4! = 24$	$(abcde)$	5

♣Put the tables in uniform format and add one for S_6. ♣

Exercise *Sign of the conjugacy class*

Let $\epsilon : S_n \to \{\pm 1\}$ be the sign homomorphism. Show that $\epsilon(g) = (-1)^{n+\sum_j \ell_j}$ if g is in the conjugacy class (7.78).

Exercise *Order of the conjugacy class*

Given a conjugacy class of type (7.78) compute the order $|C(g)|$. [71]

7.4.1 Conjugacy Classes In S_n And Harmonic Oscillators

There is a beautiful relation of conjugacy classes of the symmetric group with special collections of harmonic oscillators. We'll give a taste of how that happens here.

Let's review briefly some facts about the quantum mechanical harmonic oscillator: The classical harmonic oscillators is described by a phase space with coordinate $q \in \mathbb{R}$ (the displacement of the oscillator) and momentum $p \in \mathbb{R}$ with Hamiltonian

$$H = \frac{1}{2}(p^2 + \omega^2 q^2) \tag{7.80}$$

(we have scaled away the mass). The classical Poisson bracket $\{p, q\} = 1$ is quantized by postulating there are operators $\hat{p}, \hat{q}$ with

$$[\hat{p}, \hat{q}] = -\mathrm{i}\hbar \tag{7.81}$$

[71] *Answer*: $|C(g)| = n!/(\prod_{i=1}^{n} i^{\ell_i} \ell_i!)$.

and there is a Hilbert space representing this operator algebra. One forms the linear combinations:

$$\begin{aligned} a &:= \frac{1}{\sqrt{2\hbar\omega}}(\omega\hat{q} + \mathrm{i}\hat{p}) \\ \bar{a} &:= \frac{1}{\sqrt{2\hbar\omega}}(\omega\hat{q} - \mathrm{i}\hat{p}) \end{aligned} \tag{7.82}$$

so that

$$[a, \bar{a}] = 1 \tag{7.83}$$

The operator algebra is a $*$-algebra so there is a $\mathbb{C}$-antilinear map with $\hat{q}^* = \hat{q}$ and $\hat{p}^* = \hat{p}$. So then $\bar{a} = a^*$.

If we just consider the operator algebra without considering H then there are a number of different ways to represent it. We could postulate there is a vector $|0\rangle$ with $a|0\rangle = 0$. Then the Hilbert space is spanned by $\bar{a}^n|0\rangle$ and $\bar{a} = a^\dagger$. But note that we could make a linear transformation to

$$\begin{aligned} b &= \alpha a + \beta\bar{a} \\ \bar{b} &= \gamma a + \delta\bar{a} \end{aligned} \tag{7.84}$$

and this will preserve the commutation relations: $[b, \bar{b}] = 1$, if $\alpha\delta - \beta\gamma = 1$. If we preserve the $*$ structure so that $\bar{b} = b^*$ then $\delta = \alpha^*$ and $\gamma = \beta^*$ and we find transformations by $SU(1,1)$.

If we represent $\hat{p}, \hat{q}$ on the Hilbert space $L^2(\mathbb{R})$ with

$$\begin{aligned} (\hat{q} \cdot \psi)(x) &= x\psi(x) \\ (\hat{p} \cdot \psi)(x) &= -\mathrm{i}\hbar\frac{\partial}{\partial x}\psi(x) \end{aligned} \tag{7.85}$$

then the groundstate with $a|0\rangle = 0$ is preferred because the corresponding vector in $L^2(\mathbb{R})$ satisfies:

$$(\mathrm{i}\hat{p} + \omega\hat{q})\psi = 0 \qquad \Rightarrow \qquad \psi(x) = Ce^{-\frac{\omega x^2}{2\hbar}} \tag{7.86}$$

This choice of quantization is also preferred when we consider the oscillator Hamiltonian:

The quantum Hamiltonian is $H = \omega(a^\dagger a + \frac{1}{2})$. The states $\bar{a}^n|0\rangle$ are eigenstates of H with eigenvalue $\omega(n + \frac{1}{2})$. Assuming $|0\rangle$ has unit norm the normalized eigenstates $\frac{1}{\sqrt{n!}}\bar{a}^n|0\rangle$ form a complete ON basis for the Hilbert space.

In quantum statistical mechanics a very important quantity is the (physics) *partition function* defined to be

$$\mathrm{Tr}_{\mathcal{H}_{\text{single h.o.}}} e^{-\beta H} = \frac{e^{-\frac{1}{2}\beta\omega}}{1 - e^{-\beta\omega}} = \frac{1}{2\sinh\frac{1}{2}\beta\omega} \tag{7.87}$$

Here β has the physical interpretation of $1/(kT)$ where k is Boltzmann's constant and T is the temperature above absolute zero. Using the (physics) partition function one derives thermodynamic quantities when the oscillator is connected to a heat bath.

Now, suppose we have a system which is described by an infinite collection of harmonic oscillators:

$$[a_j, a_k] = 0 \qquad [a_j^\dagger, a_k^\dagger] = 0 \qquad [a_j, a_k^\dagger] = \delta_{j,k} \qquad j,k = 1,\dots \tag{7.88}$$

Suppose they have frequencies which are all a multiple of a basic harmonic which we'll denote ω, so the frequencies are associated with the oscillators $a_1, a_2, a_3, \dots$ are $\omega, 2\omega, 3\omega, \dots$. The motivation for choosing all frequencies to be multiples of a basic frequency comes from the theory of strings, as we will explain below.

If we write the standard sum of harmonic oscillator Hamiltonians we get, formally,

$$H^{\text{formal}} = \sum_{j=1}^{\infty} j\omega(a_j^\dagger a_j + \frac{1}{2}) \tag{7.89}$$

This is formal, because on the usual lowest weight module of the system defined by saying the vacuum line satisfies:

$$a_j|vac\rangle = 0 \qquad\qquad \forall j \tag{7.90}$$

the groundstate energy is infinite. This is typical of the divergences of quantum field theory [72] An infinite number of degrees of freedom typically leads to divergences in physical quantities. However, there is a very natural way to regularize and renormalize this divergence by using the Riemann zeta function:

$$\sum_{j=1}^{\infty} \frac{j}{2}\omega = \frac{\omega}{2}\sum_{j=1}^{\infty} \frac{1}{j^{-1}} \to \frac{\omega}{2}\zeta(-1) = -\frac{\omega}{24} \tag{7.91}$$

[73] This can be justified much more rigorously and indeed it gives the correct Casimir energy for a massless scalar field on an interval. Multiplying by two we get a similar result for a scalar field on a circle. If we restore proper units so that the radius of the circle so that the length is $2\pi L$ the ground state energy is:

♣We now arranged the string explanation below so this remark is out of place. ♣

$$E_{\text{ground}} = -\frac{1}{24L} = -\frac{\hbar c}{24L} \tag{7.92}$$

where in the second equation we restored $\hbar$ and c which had been set to 1. Of course, unless you couple the system to gravity, the zero of energy is arbitrary. Here the zero of energy is defined by saying the massless scalar field on the real line has zero groundstate energy. Then the above formula for the Casimir energy is meaningful. What is meaningful independent of the choice of zero of energy is the Casimir force $-\frac{\partial E_{\text{ground}}}{\partial L}$.

[72] The quantum field theory in question is that of a massless scalar field in a spacetime of $1+1$ dimensions.

[73] One way to see that $\zeta(-1) = -1/12$ is to use the functional equation for the Riemann zeta function. $\zeta(s)$ is a convergent series for $Re(s) > 1$. Define $\xi(s) = \frac{1}{2}\pi^{-s/2}s(s-1)\Gamma(\frac{s}{2})\zeta(s)$. Then one proves the stunning result: $\xi(s) = \xi(1-s)$ for the analytically continued ζ-function. Now, one can analytically continue the Γ function using $\Gamma(x+1) = x\Gamma(x)$, and then evaluate both sides of the functional equation at $s = 2$ using $\Gamma(1/2) = \sqrt{\pi}$ and $\zeta(2) = \pi^2/6$. The analytic continuation of the ζ-function can be derived from the integral representation $2\pi^{-s/2}\Gamma(s/2)\zeta(s) = \int_0^\infty x^{s/2}(\vartheta(\mathrm{i}x) - 1)dx$ and from this representation one easily derives the functional equation once one knows the modular transformation law of the theta function $\vartheta(\tau)$. See section *** below for a physical derivation of the modular transformation law of the theta function.

In any case, things work out very nicely if we take the Hamiltonian to be:

$$H = \sum_{j=1}^{\infty} j\omega a_j^\dagger a_j - \frac{\omega}{24} \tag{7.93}$$

The dimension of the space of states of energy $n\omega$ above the groundstate is $p(n)$. A natural basis of this space is labeled by partitions of n:

$$(a_1^\dagger)^{\ell_1}(a_2^\dagger)^{\ell_2}\cdots(a_n^\dagger)^{\ell_n}|0\rangle \tag{7.94}$$

and hence the vectors in this basis are in 1-1 correspondence with the conjugacy classes of S_n. This turns out to be significant in the boson-fermion correspondence in 1+1 dimensional quantum field theory.

The quantum statistical mechanical partition function of this collection of oscillators has a truly remarkable property, as we now explain:

Let q be a complex number with $|q| < 1$. Notice that:

$$\begin{aligned}\frac{1}{\prod_{j=1}^{\infty}(1-q^j)} &= (1+q+q^2+\cdots)(1+q^2+q^4+\cdots)(1+q^3+q^6+\cdots)\cdots \\ &= 1+\sum_{n=1}^{\infty} p(n)q^n\end{aligned} \tag{7.95}$$

Indeed, note that this is almost exactly the same as the physical partition function of our system of oscillators!

Therefore, taking into account the full system of harmonic oscillators we have:

$$Z^{osc}(\beta) = \mathrm{Tr}_{\mathcal{H}_{\text{all the h.o.s}}} e^{-\beta H} = \frac{1}{q^{1/24}\prod_{n=1}^{\infty}(1-q^n)}, \tag{7.96}$$

where we trace over the Hilbert space of states of our collection of oscillators. Here we identify $q = e^{-\beta\omega}$. Expanding out (7.95) gives the first few values of $p(n)$:

$$\begin{aligned}&1+q+2q^2+3q^3+5q^4+7q^5+11q^6+15q^7+22q^8+30q^9+ \\ &+42q^{10}+56q^{11}+77q^{12}+101q^{13}+135q^{14}+\cdots\end{aligned} \tag{7.97}$$

and one can easily generate the first few 100 values using Maple or Mathematica or Sage or ...

It turns out the generating series has a remarkable "modular transformation property" relating $Z(\beta)$ to $Z(1/\beta)$:

$$\beta^{-1/4} Z^{osc}(\beta) = \tilde{\beta}^{-1/4} Z^{osc}(\tilde{\beta}) \tag{7.98}$$

$$\beta\tilde{\beta} = \left(\frac{2\pi}{\omega}\right)^2 \tag{7.99}$$

This is a kind of high-low temperature duality.

The property (7.98) and (7.99) is proven in textbooks on analytic number theory. In this context one often uses the variable

$$\tau = \mathrm{i}\frac{\beta\omega}{2\pi} \tag{7.100}$$

so that $q = e^{-\beta\omega} = e^{2\pi i\tau}$. Although the physical motivation starts with $\beta > 0$, one can analytically continue so that τ is in the upper half complex plane. (See below for the physical interpretation of this continuation.) Analytic number theorists define the *Dedekind eta function* :

$$\eta(\tau) = \exp\left(\frac{2\pi i\tau}{24}\right)\prod_{n=1}^{\infty}(1-q^n) \tag{7.101}$$

and prove the crucial identity:

$$\eta(-1/\tau) = (-i\tau)^{1/2}\eta(\tau). \tag{7.102}$$

This equation is equivalent to the identity in equations (7.98) and (7.99).

An Interpretation From Physics

It turns out that (7.98) and (7.99), or equivalently, (7.102) has a beautiful interpretation from physics. The string is a circle with coordinate $\sigma \sim \sigma + 2\pi$ and the displacement of the string $X(\sigma, t)$ is a real number depending on position and time. The action is:

$$S = \frac{1}{4\pi\ell_s^2}\int dt\int_0^{2\pi} d\sigma\left((\partial_t X)^2 - \partial_\sigma X)^2\right) \tag{7.103}$$

where ℓ_s has units of length, or inverse mass, and we have temporarily set $\hbar = c = 1$ by choice of units. So $T = \ell_s^{-2}$ is the tension of the string. Then the general solution of the classical equation of motion (the wave equation) $(\partial_t^2 - \partial_\sigma^2)X = 0$ is:

$$X(t,\sigma) = X_0 + \ell_s^2 pt + i\frac{\ell_s}{\sqrt{2}}\sum_{n\neq 0}\left(\frac{\alpha_n}{n}e^{in(t+\sigma)} + \frac{\tilde{\alpha}_n}{n}e^{in(t-\sigma)}\right) \tag{7.104}$$

with complex numbers $\alpha_n = (\alpha_{-n})^*$ and $\tilde{\alpha}_n = (\tilde{\alpha}_{-n})^*$ and X_0, p are real.

We can think of this as a $1+1$ dimensional field theory. Then spacetime is a cylinder $S^1 \times \mathbb{R}$ with Lorentzian metric. The α_n are the amplitudes of waves moving at the speed of light to the left, while the $\tilde{\alpha}_n$ are the amplitudes of waves moving at the speed of light to the right.

The Hamiltonian computed from the action is

$$H = \frac{1}{4\pi\ell_s^2}\int_0^{2\pi} d\sigma\left((\partial_t X)^2 + \partial_\sigma X)^2\right) \tag{7.105}$$

and evaluating this on the general solution of the equation of motion gives:

$$H = \frac{1}{2}\ell_s p^2 + \frac{1}{2}\ell_s^{-1}\sum_{n\neq 0}(\alpha_{-n}\alpha_n + \tilde{\alpha}_{-n}\tilde{\alpha}_n) \tag{7.106}$$

When quantizing this system we find that $[X_0, p] = i\hbar$ and

$$[\alpha_n, \alpha_m] = n\delta_{n+m,0} \qquad [\tilde{\alpha}_n, \tilde{\alpha}_m] = n\delta_{n+m,0} \qquad [\alpha_n, \tilde{\alpha}_m] = 0 \tag{7.107}$$

If we represent the Heisenberg algebras with vacua so that $\alpha_n|0\rangle = \tilde{\alpha}_n|0\rangle = 0$ for $n > 0$ then we get standard Harmonic oscillators by defining $a_n = \alpha_n/\sqrt{n}$ and $a_n^\dagger = \alpha_{-n}/\sqrt{n}$ for $n > 0$, and similarly for the right-moving modes $\tilde{\alpha}_n$. Therefore, the Hamiltonian becomes:

$$H = \frac{1}{2}\ell_s \hat{p}^2 + \ell_s^{-1}\left[\sum_{n=1}^{\infty} n\left(a_n^\dagger a_n + \tilde{a}_n^\dagger \tilde{a}_n\right) - \frac{2}{24}\right] \tag{7.108}$$

We thus recognize $\omega = \ell_s^{-1}$ as our basic frequency.

Let us now compute the quantum statmech partition function for this string:

$$Z(\beta) := Tre^{-\beta H} = (2\pi\beta\omega)^{-1/2}\,(Z^{osc}((\beta))^2 \tag{7.109}$$

where the prefactor comes from the zeromode degrees of freedom (X_0, p) of the scalar field and is obtained from the Gaussian integral

$$\ell_s \int_{-\infty}^{+\infty} \frac{dp}{2\pi} e^{-\beta\frac{1}{2}\ell_s p^2} = \frac{\ell_s}{\sqrt{2\pi\beta\ell_s}} = \frac{1}{\sqrt{2\pi\beta\omega}} \tag{7.110}$$

We have written this for β real.

Moreover, it is a standard and fundamental result that for real β the partition function $Tre^{-\beta H}$ can be written as a path integral with periodic Euclidean time of period β. The chain of logic is that the partition function is of the form

$$\mathrm{Tr}e^{-\beta H} = \sum_{\psi_n} \langle\psi_n|e^{-\beta H}|\psi_n\rangle \tag{7.111}$$

where ψ_n is a basis of the space of states. But now $\langle\psi_1|e^{-\beta H}|\psi_2\rangle$ is an analytic continuation of the transition amplitude $\langle\psi_1|e^{-\mathrm{i}tH}|\psi_2\rangle$ to imaginary time. On the other hand $\langle\psi_1|e^{-\mathrm{i}tH}|\psi_2\rangle$ can be written as a path integral with initial and final conditions ψ_2, ψ_1, respectively. If we set $\psi_1 = \psi_2$ and sum over a complete basis then the domain of the path integral becomes that of all field configurations on the circle of Euclidean time. For more on this, see the classic book by Feynman and Hibbs.

If we apply the above principle to the case of our $1+1$ dimensional QFT, the quantum field X is already a map from the circle to the real line so altogether we have a path integral on a torus $S^1 \times S^1$ with metric

$$ds^2 = (d\sigma)^2 + (\frac{\beta}{\ell_s})^2 (d\sigma^2)^2 = (2\pi)^2\left[(d\sigma^1)^2 + \left(\frac{\beta}{2\pi\ell_s}\right)^2 (d\sigma^2)^2\right] \tag{7.112}$$

where we have chosen a dimensionless coordinate $\sigma^2 \sim \sigma^2 + 1$ in the Euclidean time direction and rescaled $\sigma = 2\pi\sigma^1$ so that $\sigma^1 \sim \sigma^1 + 1$.

Now, in statistical physics it is natural to consider the analytic continuation of $Z(\beta)$ to the subset of the complex β-plane with positive real part. In this case, to interpret the analytic continuation as a trace we should split the Hamiltonian into contributions of left- and right-moving oscillators. We can do this by introducing separate Hamiltonians for the left- and right-moving degrees of freedom:

$$\begin{aligned} H_L &= \frac{1}{2}(H + P) \\ H_R &= \frac{1}{2}(H - P) \end{aligned} \tag{7.113}$$

where P is the total momentum of the field $\sim \int_0^{2\pi} \partial_t X \partial_\sigma X$. Note that $H = H_L + H_R$. Explicitly, one finds: [74]

$$\begin{aligned} H_L &= \frac{1}{4}\ell_s \hat{p}^2 + \ell_s^{-1}\left[\sum_{n=1}^{\infty} n a_n^\dagger a_n - \frac{1}{24}\right] \\ H_R &= \frac{1}{4}\ell_s \hat{p}^2 + \ell_s^{-1}\left[\sum_{n=1}^{\infty} n \tilde{a}_n^\dagger \tilde{a}_n - \frac{1}{24}\right] \end{aligned} \tag{7.114}$$

We set

$$q = e^{2\pi \mathrm{i}\tau} = e^{-\beta\omega} \qquad \Leftrightarrow \qquad \tau = \mathrm{i}\frac{\beta\omega}{2\pi} \ . \tag{7.115}$$

Then the proper analytic continuation to consider is

$$Z(\beta) = \mathrm{Tr} q^{H_L} \bar{q}^{H_R} = \mathrm{Tr} e^{-Re(\beta\omega)H} e^{-\mathrm{i}Im(\beta\omega)P} \tag{7.116}$$

With this understood the partition function now becomes

$$Z(\beta) = (\mathrm{Im}\tau)^{-1/2} |Z^{osc}(\beta)|^2 \tag{7.117}$$

One of the virtues of (7.116) is that it still has a nice interpretation in terms of a path integral on a torus: After we propagate in Euclidean time by $\mathrm{Re}(\beta\omega)$ we shift in the σ coordinate by $\mathrm{Im}(\beta\omega)$ before gluing, because P is the generator of translations in the σ direction. The net result is that we can identify $Z(\beta)$ with the pathintegral on a torus with metric:

$$ds^2 = (2\pi)^2 |d\sigma^1 + \tau d\sigma^2|^2 = (2\pi)^2 |dz|^2 \tag{7.118}$$

We can identify the torus with our friend: $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ with a flat metric $ds^2 = |dz|^2$ where τ is a complex number in the upper half-plane. One can easily (and rigorously) compute that path integral and show that it is

$$Z(\beta) = (\mathrm{Im}\tau)^{-1/2} |\eta(\tau)|^{-2} \tag{7.119}$$

where

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n), \tag{7.120}$$

confirming the general principle we just enunciated.

On the other hand, the path integral is invariant under "large" diffeomorphisms of the torus, that is, on diffeomorphisms which are not deformable to the identity. Such diffeomorphisms will act nontrivially on $\pi_1(T^2)$. If we take $z = x + \tau y$ with x, y real and x, y identified modulo one, then we can make the diffeomorphism that rotates by 90 degrees in the x, y plane. Note that this exchanges A- and B-cycles: We are exchanging the spatial circle with the (Euclidean) time circle. The transformation also takes the torus to a torus with $\tau \to -1/\tau$ and the flat metric rescaled by a constant factor: If $ds^2 = |dz|^2$ with

[74] The splitting of the zeromode between left- and right-movers is a very subtle point we have elided here. To do this properly one needs to work out the quantization of a self-dual field.

$z = x + \mathrm{i}y$ then the pull-back is $ds^2 = |\tau|^2 |dx' + \tau' dy'|^2$ with $x' = y$ and $y' = -x'$ and $\tau' = -1/\tau$.

What about the overall factor of $|\tau|^2$ in front of the metric? But the massless scalar field above has a beautiful property known as *conformal invariance*. Let us return to the action and restore the $1+1$ dimensional Minkowskian metric:

$$S = -\frac{1}{4\pi \ell_s^2} \int_{S^1 \times \mathbb{R}} d^2\sigma \sqrt{|\det \eta|} \eta^{\alpha\beta} \partial_\alpha X \partial_\beta X \tag{7.121}$$

we choose $\eta_{tt} = -1$. The generalization to an arbitrary metric $h_{\alpha\beta} d\sigma^\alpha d\sigma^\beta$ on the "world-sheet" $S^1 \times \mathbb{R}$ is clear:

$$S = -\frac{1}{4\pi \ell_s^2} \int_{S^1 \times \mathbb{R}} d^2\sigma \sqrt{|\det h|} h^{\alpha\beta} \partial_\alpha X \partial_\beta X \tag{7.122}$$

This is the standard minimal coupling in general relativity. But note that the action is invariant under *conformal transformations*

$$h_{\alpha\beta} \to \Omega^2 h_{\alpha\beta} \tag{7.123}$$

One has to be very careful about the quantum theory: In this case the partition function is not quite invariant but rather scales by an overall functional of Ω. But for a flat metric and constant Ω, the overall scaling factor is just one.

We finally conclude that $(\mathrm{Im}\tau)^{-1/2} |\eta(\tau)|^{-2}$ is invariant under $\tau \to -1/\tau$ and since $\eta(\tau)$ is holomorphic one can deduce the very important result (7.102) above.

As an interesting application of (7.102), when combined with the method of stationary phase, one can derive the Hardy-Ramanujan formula giving an asymptotic formula for large values of n:

$$p(n) \sim \frac{1}{\sqrt{2}} \left(\frac{1}{24}\right)^{3/4} n^{-1} \exp\left(2\pi \sqrt{\frac{n}{6}}\right) \tag{7.124}$$

Note that this grows much more slowly than the order of the group, $n!$. So we conclude that some conjugacy classes must be very large! (See discussion in the next section on the class equation if this is not obvious.)

Analogs of equation (7.124) for a class of functions known as *modular forms* plays an important role in modern discussions of the entropy of supersymmetric (and extreme) black hole solutions of supergravity.

Exercise *Deriving the Hardy-Ramanujan formula*

The function $Z^{osc}(\beta)$ has a nice analytic continuation into the right half complex plane where $\mathrm{Re}(\beta) > 0$. Note that $q^{1/24} Z^{osc}(\beta)$ is periodic under imaginary shifts $\beta \to \beta + \frac{2\pi \mathrm{i}}{\omega}$.

Write

$$p(n) = \int_{\beta_0}^{\beta_0 + \frac{2\pi \mathrm{i}}{\omega}} d\beta e^{-n\beta\omega} q^{1/24} Z^{osc}(\beta) \tag{7.125}$$

and use the above transformation formula, together with the stationary phase method to derive (7.124). [75]

7.4.2 Conjugacy Classes In S_n And Partitions

Another way of thinking about partitions of n uses the general idea of partitions: In general, a *partition* is a sequence of nonnegative integers $\{\lambda_1, \lambda_2, \lambda_3, \dots\}$ so that

a.) λ_i are nonincreasing: $\lambda_i \geq \lambda_{i+1}$.

b.) The λ_i eventually become zero.

Given a partition, we define $|\lambda| = \sum_i \lambda_i$ so that a partition of n can be written:

$$n = \lambda_1 + \lambda_2 + \cdots + \lambda_k \tag{7.126}$$

as a sum of positive integers with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$. The nonzero λ_i are called the *parts of the partition*. The above is a partition of n with k parts.

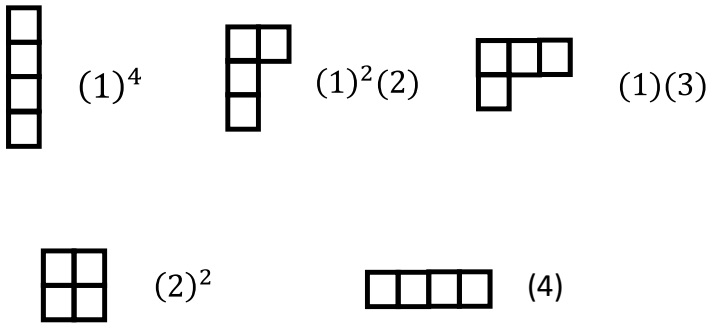


Figure 20: Young diagrams corresponding to the 5 different partitions of 4.

In general, to a partition λ we can associate a *Young diagram*. This is a diagram with λ_1 boxes in the first row, λ_2 boxes in the second row and so forth. The boxes are arranged to make, roughly speaking, an upside-down L-shape. See Figure 20 for some examples. We will talk much more about these when discussing representations of the symmetric group and representations of $SU(n)$.

[75] *Answer*: Write $p(n) = \int_{-1/2+i\beta}^{1/2+i\beta} e^{-2\pi i n\tau} q^{1/24} \eta(\tau)^{-1} d\tau$ where $\tau = x + i\beta$ and the contour is along a horizontal line. One argues that, as $\beta \to 0^+$ the dominant terms in the integral come from the region near $x \cong 0$. (This is a rather subtle step to do correctly.) Then, using the modular transformation law one writes $\eta(\tau)^{-1} = (-i\tau)^{1/2}\eta(-1/\tau)^{-1}$ and for $\mathrm{Im}(-1/\tau) \to \infty$ one approximates $\eta(-1/\tau)^{-1} \cong \exp[2\pi i/(24\tau)]$. Now one applies the standard stationary phase technique. When this procedure is carried out more systematically one is led to the famous Rademacher expansion for coefficients of certain modular functions.

One way to associate a conjugacy class in S_n with a partition is as follows. We let

$$m_i(\lambda) := |\{j|\lambda_j = i\}| \tag{7.127}$$

denote the *multiplicity of* i *in* λ. In terms of the Young diagram it is the number of rows with i boxes. If there are m_i rows with i boxes then that accounts for $i \times m_i$ boxes. Summing over i gives n, the total number of boxes in the diagram. So we can associate the conjugacy class:

$$(1)^{m_1}(2)^{m_2}\cdots \tag{7.128}$$

to a partition/Young diagram.

It is worth noting that there is another associated partition and conjugacy class known as the conjugate partition λ. It corresponds to the partition obtained by flipping on the main diagonal: We exchange rows and columns. Let λ'_i be the number of boxes in the ith column. By the inverted L-shape we see that this is another partition

$$\lambda'_1 \geq \lambda'_2 \geq \cdots \tag{7.129}$$

with $|\lambda'| = n$. Note that $\lambda'' = \lambda$.

Note that

$$\lambda'_i = |\{j|\lambda_j \geq i\}| \tag{7.130}$$

For example, λ'_1 is the number of boxes in the first column. This is clearly the number of rows, and to have a nontrivial row we must have $\lambda_j \geq 1$. Now, to get a box in the second column we must have rows with $\lambda_j \geq 2$, and so on.

Therefore,

$$m_i(\lambda) = \lambda'_i - \lambda'_{i+1} \tag{7.131}$$

So in terms of λ' our conjugacy class above becomes

$$(1)^{\lambda'_1-\lambda'_2}(2)^{\lambda'_2-\lambda'_3}\cdots(n)^{\lambda'_n} \tag{7.132}$$

Note that the identity

$$\lambda'_1+\lambda'_2+\cdots+\lambda'_n = (\lambda'_1-\lambda'_2)+2(\lambda'_2-\lambda'_3)+3(\lambda'_3-\lambda'_4)+\cdots+(n-1)(\lambda'_{n-1}-\lambda'_n)+n\lambda'_n \tag{7.133}$$

assures us the total number of boxes is still n.

Now, when n is large we can ask what the "typical" partition is. That is, what are the "typical" conjugacy classes in S_n when n is large? This is an imprecise, and rather subtle question. To get some sense of an answer it is useful to consider the number $p_k(n)$ of partitions of n into precisely k parts (as in (7.126)). The generating function is

$$\sum_{n=1}^{\infty} p_k(n)x^n = \prod_{j=1}^{k}\frac{x}{1-x^j} \tag{7.134}$$

One natural guess, then, is that the "typical" partition has $k \cong \sqrt{n}$ with "most of the parts" on the order of $\sqrt{n}$. This naive picture can be considerably improved using the

statistical theory of partitions. [76] Without going into a lot of complicated asymptotic formulae, the main upshot is that, for large n, as a function of k, $p_k(n)$ indeed is sharply peaked with a maximum around

$$\bar{k}(n) := \frac{\sqrt{6}}{2\pi}\sqrt{n}\log n \tag{7.135}$$

See Figure 21 for a numerical illustration. Moreover, and again speaking very roughly, the number of terms in the partition λ_j with $\lambda_j \cong \frac{\sqrt{6}}{2\pi}\sqrt{n}$ is order $\sqrt{6n}/\pi$.

Remarks:

1. Recall that in our discussion of a string, or equivalently of a massless scalar field on the circle, there are $p(n)$ states in the energy eigenspace with energy $E = (n - \frac{1}{24})\omega$. Thus we can interpret the above result as a kind of equipartition theorem: The most likely state is the one where the energy is shared equally by the different oscillators.

2. A *Young tableau* is a Young diagram with n boxes where the boxes have been filled in with integers drawn from $\{1, \ldots, n\}$ so that no integer is repeated. Note that the symmetric group S_n acts on Young tableau. For a given tableau T we can define two subgroups of the symmetric group: $R(T)$ are the permutations that only move numbers around within the rows and $C(T)$ are the permutations that only move numbers around within the columns. Young tableau and these subgroups are used in constructing the irreducible representations of the symmetric group.

Exercise *Generating Function For $p_k(n)$*
Prove equation (7.134). [77]

---

8. More About Group Actions And Orbits

♣NOTE BENE! THE MATERIAL IN THIS SECTION IS IDENTICAL TO SECTION 2 OF CHAPTER 3 ♣

In Section 4.1 above we introduced the notion of a group action on a set. In this section we develop this important idea a bit further.

[76] It is a large subject. See P Erdös and J. Lehner, "The distribution of the number of summands in the partition of a positive integer," Duke Math. Journal **8**(1941)335-345 or M. Szalay and P. Turán, "On some problems of the statistical theory of partitions with application to characters of the symmetric group. I," Acta Math. Acad. Scient. Hungaricae, Vol. 29 (1977), pp. 361-379.

[77] *Answer*: A partition of n into exactly k-parts means that $\lambda_k \geq 1$. So now write

$$n - k = (\lambda_1 - \lambda_2) + 2(\lambda_2 - \lambda_3) + \cdots + (k-1)(\lambda_{k-1} - \lambda_k) + k(\lambda_k - 1)$$

This is a partition of $n - k$ as a sum of integers drawn from $\{1, \ldots, k\}$. Enumerating those is clearly given by $\prod_{j=1}^{k}(1 - x^j)^{-1}$.

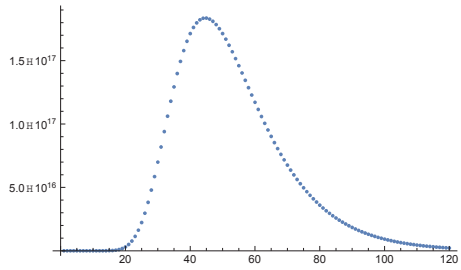


Figure 21: Showing the distribution of $p_k(n)$ as a function of k for $n = 400$ and $1 \leq k \leq 120$. Note that the Erdös-Lehner mean value of k is $\bar{k} = \frac{\sqrt{6}}{2\pi} 20\log(20) \cong 46.7153$ is a very good approximation to where the distribution has its sharp peak. The actual maximum is at $k = 45$.

8.1 Some Definitions And Terminology Associated With Group Actions

Let X be any set (possibly infinite). Recall the definition we gave in Section 4.1.

A *permutation* of X is a 1-1 and onto mapping $X \to X$. The set S_X of all permutations

forms a group under composition. A *transformation group* on X is a subgroup of S_X.

Equivalently, a *G-action on a set* X is a map $\phi : G \times X \to X$ compatible with the group multiplication law as follows:

A *left-action* satisfies:

$$\phi(g_1, \phi(g_2, x)) = \phi(g_1 g_2, x) \tag{8.1}$$

A *right-action* satisfies

$$\phi(g_1, \phi(g_2, x)) = \phi(g_2 g_1, x) \tag{8.2}$$

In addition in both cases we require that

$$\phi(1_G, x) = x \tag{8.3}$$

for all $x \in X$.

Remarks:

1. If ϕ is a left-action then it is natural to write $g \cdot x$ for $\phi(g, x)$. In that case we have

$$g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x. \tag{8.4}$$

   Similarly, if ϕ is a right-action then it is better to use the notation $\phi(g, x) = x \cdot g$ so that

$$(x \cdot g_2) \cdot g_1 = x \cdot (g_2 g_1). \tag{8.5}$$

2. If ϕ is a left-action then $\tilde{\phi}(g, x) := \phi(g^{-1}, x)$ is a right-action, and vice versa. Thus there is no essential difference between a left- and right-action. However, in computations with nonabelian groups it is extremely important to be consistent and careful about which choice one makes. Confusing left- and right- actions is a common source of error.

3. A given set X can admit more than one action by the same group G. If one is working simultaneously with several different G actions on the same set then the notation $g \cdot x$ is ambiguous and one should write, for example, $\phi_g(x) = \phi(g, x)$ or speak of ϕ_g, etc. A good example of a set X with several natural G actions is the case of $X = G$ itself. Then there are the actions of left-multiplication, right-multiplication, and conjugation. The action of g on the group element g' is:

$$\begin{aligned} L(g, g') &= gg' \\ \tilde{L}(g, g') &= g^{-1} g' \\ R(g, g') &= g' g \\ \tilde{R}(g, g') &= g' g^{-1} \\ C(g, g') &= g^{-1} g' g \\ \tilde{C}(g, g') &= g g' g^{-1} \end{aligned} \tag{8.6}$$

   where on the RHS of these equations we use group multiplication. The reader should work out which actions are left actions and which actions are right actions.

There is some important terminology one should master when working with G-actions. First here are some terms used when describing a G-action on a set X:

Definitions:

1. A group action is *effective* or *faithful* if for any $g \neq 1$ there is <u>some</u> x such that $g \cdot x \neq x$. Equivalently, the only $g \in G$ such that ϕ_g is the identity transformation is $g = 1_G$. A group action is *ineffective* if there is <u>some</u> $g \in G$ with $g \neq 1$ so that $g \cdot x = x$ for all $x \in X$. The set of $g \in G$ that act ineffectively is a normal subgroup of G.

2. A group action is *transitive* if for any pair $x, y \in X$ there is some g with $y = g \cdot x$.

3. A group action is *free* if for any $g \neq 1$ then for *every* x, we have $g \cdot x \neq x$.

In summary:

1. *Effective*: $\forall g \neq 1, \exists x$ s.t. $g \cdot x \neq x$.

2. *Ineffective*: $\exists g \neq 1$,s.t. $\forall x\ g \cdot x = x$.

3. *Transitive*: $\forall x, y \in X, \exists g$ s.t. $y = g \cdot x$.

4. *Free*: $\forall g \neq 1, \forall x,\ g \cdot x \neq x$

In addition there are some further important definitions:

1. Given a point $x \in X$ the set of group elements:

$$\mathrm{Stab}_G(x) := \{g \in G : g \cdot x = x\} \tag{8.7}$$

is called the *isotropy group at* x. It is also called the *stabilizer group* of x. It is often denoted G^x. The reader should show that $G^x \subset G$ is in fact a subgroup. Note that a group action is free iff for every $x \in X$ the stabilizer group G^x is the trivial subgroup $\{1_G\}$.

2. A point $x \in X$ is a *fixed point* of the G-action if there exists <u>some</u> element $g \in G$ with $g \neq 1$ such that $g \cdot x = x$. So, a point $x \in X$ is a fixed point of G iff $\mathrm{Stab}_G(x)$ is not the trivial group. Some caution is needed here because if an author says "x is a fixed point of G" the author might mean that $\mathrm{Stab}_G(x) = G$. That would not be implied by our terminology.

3. Given a group element $g \in G$ the *fixed point set* of g is the set

$$\mathrm{Fix}_X(g) := \{x \in X : g \cdot x = x\} \tag{8.8}$$

The fixed point set of g is often denoted by X^g. Note that if the group action is free then for every $g \neq 1$ the set $\mathrm{Fix}_X(g)$ is the empty set.

4. We repeat the definition from section **** above. The *orbit of G through a point x* is the set of points $y \in X$ which can be reached by the action of G:

$$O_G(x) = \{y : \exists g \quad \text{such that} \quad y = g \cdot x\} \tag{8.9}$$

Remarks:

1. If we have a G-action on X then we can define an equivalence relation on X by defining $x \sim y$ if there is a $g \in G$ such that $y = g \cdot x$. (Check this is an equivalence relation!) The orbits of G are then exactly the equivalence classes of under this equivalence relation. Therefore, X is partitioned into a disjoint union of all the G-orbits.
2. The group action restricts to a transitive group action on any orbit.
3. If x, y are in the same orbit then the isotropy groups G^x and G^y are conjugate subgroups in G. Therefore, to a given orbit, we can assign a definite *conjugacy class* of subgroups.

Point 3 above motivates the

Definition If G acts on X a *stratum* is a set of G-orbits such that the conjugacy class of the stabilizer groups is the same. The set of strata is sometimes denoted $X \parallel G$.

Exercise *Group actions of G on G*
Referring to equation (8.6). Which actions are left-actions and which actions are right-actions? [78]

Exercise
Recall that a group action of G on X can be viewed as a homomorphism $\phi : G \to S_X$. Show that the action is effective iff the homomorphism is injective.

Exercise
Suppose X is a G-set.
a.) Show that the subset H of elements which act ineffectively, i.e. the set of $h \in G$ such that $\phi(h, x) = x$ for all $x \in X$ is a normal subgroup of G.

[78] *Answer*: $L, \tilde{R}, \tilde{C}$ are left-actions, while $\tilde{L}, R, C$ are right-actions.

b.) Show that the group G/H acts effectively on X.

Exercise

Let G act on a set X.

a.) Show that the stabilizer group at x, denoted G^x above, is in fact, a subgroup of G.

b.) Show that the G action is free iff the stabilizer group at every $x \in X$ is the trivial subgroup $\{1_G\}$.

c.) Suppose that $y = g \cdot x$. Show that G^y and G^x are conjugate subgroups in G. [79]

Exercise *Derangements*

A permutation in S_n which acts on $\{1, \ldots, n\}$ without fixed points is called a *derangement*. Show that the number of derangements in S_n is given by

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \tag{8.10}$$

8.2 The Stabilizer-Orbit Theorem

There is a beautiful relation between orbits and isotropy groups:

Theorem [Stabilizer-Orbit Theorem]: Each left-coset of G^x in G is in 1-1 correspondence with the points in the G-orbit of x:

$$\psi : Orb_G(x) \to G/G^x \tag{8.11}$$

for a $1-1$ map ψ.

Proof: Suppose y is in a G-orbit of x. Then $\exists g$ such that $y = g \cdot x$. Define $\psi(y) \equiv g \cdot G^x$. You need to check that ψ is actually well-defined.

$$y = g' \cdot x \qquad \to \qquad \exists h \in G^x \qquad g' = g \cdot h \qquad \to \qquad g'G^x = ghG^x = gG^x \tag{8.12}$$

Conversely, given a coset $g \cdot G^x$ we may define

$$\psi^{-1}(gG^x) \equiv g \cdot x \tag{8.13}$$

Again, we must check that this is well-defined. Since it inverts ψ, ψ is 1-1. ♠

[79] *Answer*: If $y = g_0 \cdot x$ and $g \cdot x = x$ then $(g_0 g g_0^{-1}) \cdot y = y$ so $G^y = g_0 G^x g_0^{-1}$.

Corollary: If G acts transitively on a set X then the isotropy groups G^x for all the points $x \in X$ are conjugate subgroups of G, and for any $x \in X$, there is a $1-1$ correspondence between X and the set of cosets G/G^x. If H is any one of these isotropy groups we can therefore identify X with the set of left-cosets G/H.

Remark: Sets of the type G/H are called *homogeneous spaces*. This theorem is the beginning of an important connection between the *algebraic* notions of subgroups and cosets to the *geometric* notions of orbits and fixed points. Below we will show that if G, H are topological groups then, in some cases, G/H are beautifully symmetric topological spaces, and if G, H are Lie groups then, in some cases, G/H are beautifully symmetric manifolds.

Exercise *Orbits Of $\mathbb{Z}_p$ For Prime p*

Let p be a prime and suppose the cycle group $\mathbb{Z}_p$ acts on a space X. Show that any orbit consists of a single point, or of p points.

Exercise *The Lemma that is not Burnside's*

Suppose a finite group G acts on a finite set X as a transformation group. A common notation for the set of points fixed by g is X^g. Show that the number of distinct orbits is the averaged number of fixed points:

$$|\{orbits\}| = \frac{1}{|G|}\sum_g |X^g| \tag{8.14}$$

For the answer see. [80]

Exercise *Jordan's theorem*

Suppose G is finite and acts transitively on a finite set X with more than one point. Show that there is an element $g \in G$ with no fixed points on X. [81]

---

[80] *Answer*: Write

$$\sum_{g\in G} |X^g| = |\{(x,g)|g\cdot x = x\}| = \sum_{x\in X} |G^x| \tag{8.15}$$

Now use the stabilizer-orbit theorem to write $|G^x| = |G|/|\mathcal{O}_G(x)|$. Now in the sum

$$\sum_{x\in X} \frac{1}{|\mathcal{O}_G(x)|} \tag{8.16}$$

the contribution of each distinct orbit is exactly 1.

[81] Hint: Note that $X = G/H$ for some H and apply the Burnside lemma.

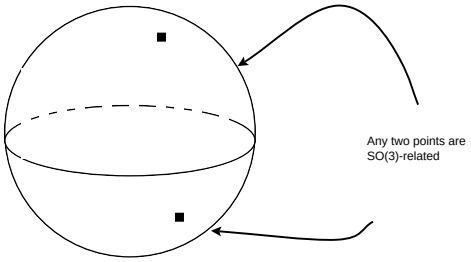


Figure 22: Transitive action of $SO(3, \mathbb{R})$ on the sphere.

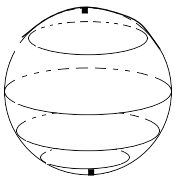


Figure 23: Orbits of $SO(2, \mathbb{R})$ on the two sphere.

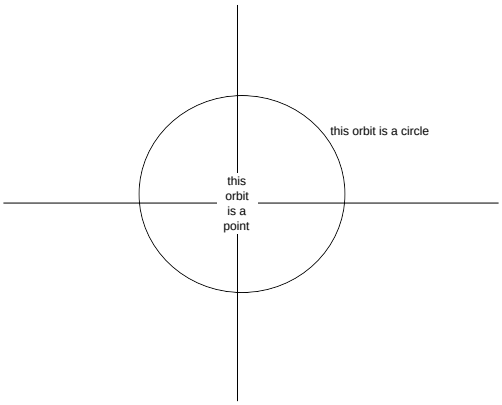


Figure 24: Notice not all orbits have the same dimensionality. There are two qualitatively different kinds of orbits of $SO(2, \mathbb{R})$.

8.3 Examples Of Orbits

The concept of a G-action on a set is an extremely important concept, so let us consider a number of examples:

Examples

1. Let G be any group and consider the group action defined by $\phi(g, x) = x$ for all $g \in G$. This is as ineffective as a group action can be: For every x, the istropy group is all of G, and for all $g \in G$, $\mathrm{Fix}(g) = X$. In particular, this situation will arise if X consists of a single point. This example is not quite as stupid as might at first appear, once one takes the categorical viewpoint, for $pt /\!/ G$ is a very rich category indeed.

2. Let $X = \{1, \cdots n\}$, so $S_X = S_n$ as before. The action is effective and transitive, but not free. Indeed, the fixed point of any $j \in X$ is just the permutations that permute everything else, and hence $S_X^j \cong S_{n-1}$. Note that different j have different stabilizer subgroups isomorphic to S_{n-1}, but they are all conjugate.

3. $GL(n, \mathbb{R})$ acts on $\mathbb{R}^n$ by matrix multiplication. If we act with a matrix on a column vector we get a left action. If we act on a row vector we get a right action. The action is:

   a.) Effective: If $g \neq 1$ some vector $\vec{x}$ is moved.

   b.) Not transitive: If $\vec{x} \neq 0$ it cannot be mapped to 0

   c.) Not free: $\vec{0}$ is a fixed point of the entire group.

   d.) There are two orbits.

   e.) The isotropy group of the vector e_1 is (under the left-action) the subgroup of matrices of the form

   $$\mathrm{Stab}_{GL(n,\mathbb{R})}(e_1) = \{ \begin{pmatrix} 1 & v \\ 0 & B \end{pmatrix} | v \in Mat_{1\times(n-1)}(\mathbb{R}), \qquad B \in GL(n-1, \mathbb{R}) \} \tag{8.17}$$

   The stabilizer group for all other nonzero vectors will be conjugate to this one. The stabilizer group of the origin is the entire group $GL(n, \mathbb{R})$.

4. If we restrict from $GL(n, \mathbb{R})$ to $SO(n, \mathbb{R})$ the picture changes completely. For simplicity consider the case $n = 2$. The left- action is:

   $$R(\phi) : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{8.18}$$

   The group action is effective. It is not free, and it is not transitive. There are now infinitely many orbits of $SO(2)$, and they are all distinguished by the invariant value of x^2+y^2 on the orbit. From the viewpoint of topology, there are two distinct "kinds" of orbits acting on $\mathbb{R}^2$. One has trivial isotropy group and one has isotropy group $SO(2)$. See Figure 24. These give two strata.

5. Orbits of $O(2)$. We have seen that $O(2)$ can be written as a disjoint union:

   $$O(2) = SO(2) \amalg P \cdot SO(2) \tag{8.19}$$

   where P is not canonical and can be taken to be reflection in any line through the origin. The orbits of $SO(2)$ and $O(2)$ are the same. We will find a very different picture when we consider the orbits of the Lorentz group.

6. Now consider a fixed $SO(2, \mathbb{R})$ subgroup of $SO(3, \mathbb{R})$, say, the subgroup defined by rotations around the z-axis, and consider the action of this group on a sphere in $\mathbb{R}^3$ of fixed radius. The action is not <u>not</u> transitive. The G-orbits are shown in Figure 23. It is also not free: The north and south poles are fixed points.

7. Now consider the action of $SO(3,\mathbb{R})$ on a sphere of positive fixed radius in $\mathbb{R}^3$. (WLOG take it to be of radius one.) The action is then transitive on the sphere. Now the isotropy subgroup $\mathrm{Stab}_{SO(3)}(\hat{n}) \subset SO(3)$ of any unit vector $\hat{n} \in S^2$ is isomorphic to $SO(2)$:

$$\mathrm{Stab}_{SO(3)}(\hat{n}) \cong SO(2) \tag{8.20}$$

But, for different choices of $\hat{n}$ we get different subgroups of $SO(3)$. For example, with usual conventions, if $\hat{n} = e_3$ is on the x^3-axis then the subgroup is the subgroup of matrices of the form

$$R_{12}(\phi) = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8.21}$$

but if x is on the x^1-axis the subgroup is the subgroup of matrices of the form

$$R_{23}(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} \tag{8.22}$$

and so on. For any $\hat{n} \in S^2$ let $SO(2)_{\hat{n}} \subset SO(3)$ denote the subgroup, isomorphic to $SO(2)$, which stabilizes $\hat{n}$. According to the stabilizer-orbit theorem there is a natural one-one correspondence

$$S^2 \cong SO(3)/SO(2)_{\hat{n}} \tag{8.23}$$

Therefore, fixing any $\hat{b} \in S^2$ there is a map

$$\pi_{\hat{n}} : SO(3,\mathbb{R}) \to S^2 \tag{8.24}$$

Put simply, $\pi(R)$ rotates $\hat{n} \in S^2$ to $R \cdot \hat{n} \in S^2$:

$$\pi_{\hat{n}}(R) := R \cdot \hat{n} \in S^2 \tag{8.25}$$

Therefore, the inverse image of any other vector $\hat{k} \in S^2$:

$$\pi_{\hat{n}}^{-1}(\hat{k}) := \{R | R\hat{n} = \hat{k}\} \subset SO(3) \tag{8.26}$$

is the set of rotations which can be (noncanonically!) put in 1-1 correspondence with elements of $SO(2)$. That is because if $\hat{k} = R_1\hat{n}$ and $\hat{k} = R_2\hat{n}$ then $R_1^{-1}R_2\hat{n} = \hat{n}$ and therefore $R_2 = R_1R_0$ where $R_0 \in \mathrm{Stab}_{SO(3)}(\hat{n}) \cong SO(2)$.

So, for each point $\hat{k} \in S^2$ we can associate a copy of $SO(2)$ inside $SO(3)$, which is topologically a circle, and clearly every element of $SO(3)$ will be captured this way as $\hat{k}$ ranges over S^2. One might think that this means that, as manifolds, $SO(3)$ is diffeomorphic to $S^2 \times SO(2) = S^2 \times S^1$, but this turns out to be quite false. For example the homotopy groups of $SO(3)$ and $S^2 \times S^1$ are completely different.

Nevertheless, we can try to parametrize the general rotation by using this idea: We choose $\hat{n} = e_3$ to be the basepoint. Then the standard polar angles of a point on the sphere are defined by

$$R_{12}(\phi)R_{23}(\theta)\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}\sin\theta\sin\phi\\ \sin\theta\cos\phi\\ \cos\theta\end{pmatrix} \tag{8.27}$$

But this does NOT mean every rotation matrix is of the form $R_{12}(\phi)R_{23}(\theta)$! It only gives us a parametrization of the cosets $SO(3)/SO(2)_{12}$. The general element can be written as

$$R = R_{12}(\phi)R_{23}(\theta)R_{12}(\psi) \tag{8.28}$$

with range $\phi \sim \phi + 2\pi$, $\psi \sim \psi + 2\pi$ and $0 \leq \theta \leq \pi$. These are the famous *Euler angles* But they are not global coordinates. They can't be: (θ, ϕ) are not global coordinates on S^2 (they go bad at the north and south poles) and $S^2 \times S^1$ is not the same manifold as $SO(3)$.

8. *$GL(2,\mathbb{C})$ And $SU(2)$ Act On $\mathbb{CP}^1$.* Recall that $\mathbb{CP}^1$ can be identified with equivalence classes of points $(z_1, z_2) \in \mathbb{C}^2 - \{0\}$ with equivalence relation $(z_1', z_2') \sim (\lambda z_1, \lambda z_2)$. We denote equivalence classes by $[z_1 : z_2]$.

   Note that $\mathbb{CP}^1$ can also be thought of as the space of states of a single Qbit: $[z_1 : z_2]$ always has a representative with $|z_1|^2 + |z_2|^2 = 1$ and the representative is unique up to multiplication by a phase. We can use such a normalized representative to define a Qbit state:

$$\psi = \begin{pmatrix}z_1\\ z_2\end{pmatrix} \tag{8.29}$$

♣THIS REMARK BELONGS EARLIER, WHEN WE FIRST INTRODUCED $\mathbb{CP}^n$. ♣

   There is a well-defined action of $GL(2;\mathbb{C})$ on $\mathbb{CP}^1$:

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix} : [z_1 : z_2] \mapsto [az_1 + bz_2 : cz_1 + dz_2] \tag{8.30}$$

   (The reader should carefully check that this is a well-defined group action. Since the $GL(2,\mathbb{C})$ action on $\mathbb{C}^2 - \{0\}$ is transitive, the action on $\mathbb{CP}^1$ is transitive. Therefore choosing a point $p \in \mathbb{CP}^1$ we have an identification of $\mathbb{CP}^1$ as a homogeneous space:

$$GL(2,\mathbb{C})/B \cong \mathbb{CP}^1 \tag{8.31}$$

   For example, if we take B to be the stabilizer of $[1 : 0]$ we compute

$$B = \{\begin{pmatrix}a & b\\ 0 & d\end{pmatrix} | a, d \in \mathbb{C}^* \qquad b \in \mathbb{C}\} \tag{8.32}$$

   Note that by the equivalence $[z_1 : z_2] = [\lambda z_1 : \lambda z_2]$ we can always find a representative so that (z_1, z_2) is a unit vector in the Hilbert space $\mathbb{C}^2$. Therefore, the restriction of

the $GL(2,\mathbb{C})$ action on $\mathbb{CP}^1$ to $SU(2)$ is still transitive. Now the stabilizer of $[1:0]$ is the subgroup of diagonal $SU(2)$ matrices and is isomorphic to $U(1)$. Therefore, there is also an identification

$$\mathbb{CP}^1 \cong SU(2)/U(1) \tag{8.33}$$

and hence there is a natural map

$$\pi : SU(2) \to \mathbb{CP}^1 \tag{8.34}$$

defined by this identification. By stereographic projection we can identify $\mathbb{CP}^1$ with the Riemann sphere. and, hence since $S^3 \cong SU(2)$ we have a natural continuous map: ♣Explain more ♣

$$\pi : S^3 \to S^2 \tag{8.35}$$

whose fibers are copies of S^1. This is a famous map in mathematics and physics known as the *Hopf map* and has many beautiful properties. It appears in the physics of magnetic monopoles and in several other related contexts. It is very closely related to the map $\pi : SO(3) \to S^2$ defined above.

Another way of thinking about $\mathbb{CP}^1$ is that it is the space of lines through the origin in $\mathbb{C}^2$. This leads us to our next example:

9. *Grassmannians.* A very nice application of the Stabilizer-Orbit theorem is to the Grassmannian of a vector space. Consider a finite dimensional vector space V, say of dimension n and let $0 < k < n$ be an integer and define $Gr_k(V)$ to be the set of all k-dimensional linear subspaces of V. It is not hard to see that $GL(V)$ acts transitively on this space: If $W \subset V$ is a k-dimensional subspace and $T \in GL(V)$ then $T(W) = \{T(w)|w \in W\}$ is a k-dimensional subspace. It is an easy fact of linear algebra that for any two k-dimensional subspaces $W_1, W_2 \subset V$ there is a T with $T(W_1) = W_2$, i.e. the action is transitive. To compute the stabilizer of some vector space W_0 choose an ordered basis $v_1, \ldots, v_k$ for W_0 and any complementary basis for V so that

$$v_1, \ldots, v_k, u_1, \ldots, u_{n-k} \tag{8.36}$$

is an ordered basis for V. Now, what is the subgroup of T so that $T(W_0) = W_0$? By definition:

$$\begin{aligned} T(v_i) &= A_{ji} v_j + C_{\alpha i} u_\alpha \\ T(u_\alpha) &= B_{j\alpha} v_j + D_{\beta\alpha} u_\beta \end{aligned} \tag{8.37}$$

The condition $T(W_0) = W_0$ is then the condition that $C = 0$ for the matrix of T relative to such a basis. So the stabilizer group is isomorphic to the subgroup P of $GL(n,\kappa)$ of matrices of the form:

$$\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \in GL(n,\kappa) \tag{8.38}$$

In fact, the Grassmannian is a manifold and the representation in terms of homogeneous coordinates helps us to find local coordinates. Suppose that $g \in GL(n,\kappa)$ is "not too far" from the identity matrix then we will try to find a representative g_0 in the coset gP so that if $g_0 \neq g_0'$ then they must be in different cosets. The idea is that we view right-multiplication as a "gauge freedom" and try to "fix the gauge" by choosing g_0 to be of the form:

$$g_0 = \begin{pmatrix} 1_{k\times k} & 0 \\ \gamma_{(n-k)\times k} & 1_{(n-k)\times(n-k)} \end{pmatrix} \tag{8.39}$$

for some matrix $\gamma_{(n-k)\times k} \in Mat_{(n-k)\times k}(\kappa)$. Note that if we right multiply by an element of P and ask that

$$\begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \gamma' & 1 \end{pmatrix} \tag{8.40}$$

Then $B = 0$ and $A = 1$ and $D = 1$ and hence $\gamma' = \gamma$.

"Not to far" means more precisely: that if we write g in block diagonal form

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{8.41}$$

then α and δ are invertible and $\alpha - \beta\delta^{-1}\gamma$ is also invertible. These are the conditions so that we can solve the equation

$$g \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \tilde{\gamma} & 1 \end{pmatrix} \tag{8.42}$$

for some $\tilde{\gamma} \in Mat_{(n-k)\times k}(\kappa)$. Indeed, one finds that we must set: $A = \alpha^{-1}$, $D = \delta^{-1}(1 - \gamma B)$ and and $B = -(\alpha - \beta\delta^{-1}\gamma)^{-1}\beta\delta^{-1}$.

Exercise $\mathbb{Z}_2$ *Actions On The Sphere*
Consider the action of $\mathbb{Z}_2$ on the sphere defined by (4.7):

$$\sigma \cdot (x^1, \ldots, x^{n+1}) = (x^1, \ldots, x^p, -x^{p+1}, \cdots, -x^{p+q}) \tag{8.43}$$

a.) For which values of p, q is the action effective?
b.) For which values of p, q is the action transitive?
c.) Compute the fixed point set of the nontrivial element $\sigma \in \mathbb{Z}_2$.
d.) For which values of p, q is the action free?

Exercise $\mathbb{C}^*$ *Actions On* $\mathbb{CP}^{n-1}$

Consider the action of $G = \mathbb{C}^*$ on $\mathbb{CP}^{n-1}$ defined by (4.8).

a.) For which values of $(q_1, \ldots, q_n)$ is the action effective?

b.) For which values of $(q_1, \ldots, q_n)$ is the action transitive?

c.) What are the fixed points of the $\mathbb{C}^*$ action?

d.) What are the stabilizers at the fixed points of the $\mathbb{C}^*$ action?

Exercise $SL(2,\mathbb{R})$ *Action On The Upper Half-Plane*

a.) Show that (4.10) above defines a left-action of $SL(2,\mathbb{R})$ on the complex upper half-plane. [82]

b.) Is the action effective?

c.) Is the action transitive?

d.) Which group elements have fixed points?

e.) What is the isotropy group of $\tau = \mathrm{i}$? [83]

Conclude that

$$\mathcal{H} \cong SL(2,\mathbb{R})/SO(2) \tag{8.44}$$

Exercise

Using $GL(2,\mathbb{C})/B \cong \mathbb{CP}^1$ show that $GL(2,\mathbb{C})$ has a natural action on the Riemann sphere given by

$$z \mapsto \frac{az+b}{cz+d} \tag{8.45}$$

Exercise

Since there is a left-action of $G \times G$ on $X = G$ there is a left-action of the diagonal subgroup $\Delta \subset G \times G$ where $\Delta = \{(g,g) | g \in G\}$ is a subgroup isomorphic to G.

a.) Show that this action is given by $a \mapsto I(a)$, where $I(a)$ is the conjugation by a.

b.) Show that the orbits of Δ are the conjugacy classes of G.

c.) What is the stabilizer subgroup of an element $g_0 \in G$?

---

[82] *Hint*: Show that $\mathrm{Im}(g \cdot \tau) = \frac{\mathrm{Im}\tau}{|c\tau+d|^2}$.

[83] The isotropy group is the subgroup $SO(2,\mathbb{R}) \subset SL(2,\mathbb{R})$. To see this set $\frac{a\mathrm{i}+b}{c\mathrm{i}+d} = \mathrm{i}$ and conclude that $a = d$ and $b = -c$. Then since $ad - bc = 1$ we have $a^2 + b^2 = 1$ but this implies that the group element is in $SO(2,\mathbb{R})$.

Exercise *Spheres As Homogeneous Spaces*

a.) Show that there is a transitive action of $SO(n+1)$ on S^n, considered as a sphere of fixed radius in $\mathbb{R}^{n+1}$.

b.) Show that $S^n \cong SO(n+1)/SO(n)$.

c.) Give an inductive proof that $SO(n)$ is a connected manifold for $n \geq 2$.

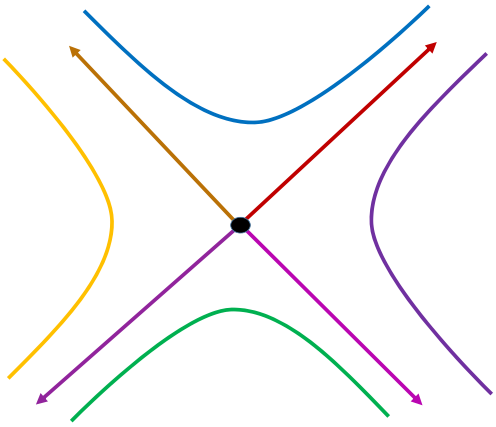


Figure 25: The distinct kinds of orbits of $SO(1,1,\mathbb{R})$ are shown in different colors. If we enlarge the group to include transformations that reverse the orientation of time and/or space then orbits of the larger group will be made out of these orbits by reflection in the space or time axis.

8.3.1 Extended Example: The Case Of $1+1$ Dimensions

Consider $1+1$-dimensional Minkowski space with coordinates $x = (x^0, x^1)$ and metric given by

$$\eta := \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{8.46}$$

i.e. the quadratic form is $(x,x) = -(x^0)^2 + (x^1)^2$. The two-dimensional Lorentz group is defined by

$$O(1,1) = \{A | A^{tr} \eta A = \eta\} \tag{8.47}$$

This group acts on $\mathbb{M}^{1,1}$ preserving the Minkowski metric.

The connected component of the identity is the group of Lorentz boosts of rapidity θ:

$$x^0 \to \cosh\theta \; x^0 + \sinh\theta \; x^1 \tag{8.48}$$

$$x^1 \to \sinh\theta\ x^0 + \cosh\theta\ x^1 \tag{8.49}$$

that is:

$$SO_0(1,1;\mathbb{R}) \equiv \{B(\theta) = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} \mid -\infty < \theta < \infty\} \tag{8.50}$$

In the notation the S indicates we look at the determinant one subgroup and the subscript 0 means we look at the connected component of 1. This is a group since

$$B(\theta_1)B(\theta_2) = B(\theta_1 + \theta_2) \tag{8.51}$$

so $SO_0(1,1) \cong \mathbb{R}$ as groups. Indeed, note that

$$B(\theta) = \exp\left[\theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \tag{8.52}$$

It is often useful to define *light cone coordinates*: [84]

$$x^{\pm} := x^0 \pm x^1 \tag{8.53}$$

and the group action in these coordinates is simply:

$$x^{\pm} \to e^{\pm\theta} x^{\pm} \tag{8.54}$$

so it is obvious that $x^+x^- = -(x,x)$ is invariant.

It follows that the orbits of the Lorentz group are, in general, hyperbolas. They are separated by different values of the Lorentz invariant $x^+x^- = \lambda$, but this is not a complete invariant, since the sign (or vanishing) of x^+ and of x^- is also Lorentz invariant. For a real number r define

$$\text{sign}(r) := \begin{cases} +1 & r > 0 \\ 0 & r = 0 \\ -1 & r < 0 \end{cases} \tag{8.55}$$

Then $(\lambda, \text{sign}(x^+), \text{sign}(x^-))$ is a complete invariant of the orbits. That is, given this triple of data there is a unique orbit with these properties.

It is now easy to see what the different types of orbits there are. They are shown in Figure 25: They are:

1. hyperbolas in the forward/backward lightcone and the left/right of the lightcone
2. 4 disjoint lightrays.
3. the origin: $x^+ = x^- = 0$.

♣Actually, the lightrays and hyperbolas have trivial stabilizer and hence are in the same strata. This is a problem with using strata. ♣

[84] Some authors will define these with a $1/2$ or $1/\sqrt{2}$. One should exercise care with this choice of convention.

It is now interesting to consider the orbits of the full Lorentz group $O(1,1)$ and its relation to the massless wave equations. But there are clearly elements of $O(1,1)$ not continuously connected to the identity such as:

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{8.56}$$

In general we can write (noncanonically),

$$O(1,1) = SO_0(1,1) \amalg P \cdot SO_0(1,1) \amalg T \cdot SO_0(1,1) \amalg PT \cdot SO_0(1,1) \tag{8.57}$$

with

The P and T operations map various orbits of $SO_0(1,1)$ into each other: P is a reflection in the time axis, i.e., a reflection of the spatial coordinate, while T is a reflection in the space axis, i.e. a reflection of the time coordinate. Thus the orbits of the groups $SO(1,1)$, $SO_0(1,1) \amalg PT \cdot SO_0(1,1)$, and $O(1,1)$ all differ slightly from each other.

♣Should give more details here, or form an exercise. ♣

As an example of a physical manifestation of orbits let us consider the energy-momentum dispersion relation of a particle of mass m with energy-momentum $(E,p) \in \mathbb{R}^{1,1}$.

1. Massive particles: $m^2 > 0$ have (E,p) along an orbit in the upper quadrant:

$$\mathcal{O}^+(m) = \{(m\cosh\theta, m\sinh\theta) | \theta \in \mathbb{R}\} \tag{8.58}$$

2. Massless particles move at the speed of light. In 1+1 dimensions there is an interesting refinement of the massless orbits: Left-moving particles with positive energy have support on [85] $p_+ = \frac{1}{2}(E+p) = 0$ and $p_- = \frac{1}{2}(E-p) \neq 0$. Right-moving particles with positive energy have support on $p^- = 0$ and $p^+ \neq 0$. In $d+1$ dimensions with $d > 1$ the orbits of $SO_0(1,d)$ consisting of the forward and backward lightcones (minus the origin) are connected.

3. Tachyons have $E^2 - p^2 = m^2 < 0$ and have their support on the left or right quadrant. If we try to expand a solution to the wave-equation with $e^{\mathrm{i}(k_0 x^0 + k_1 x^1)}$ then $k_0^2 = \sqrt{k_1^2 + m^2}$ and so if the spatial momentum k_1 is sufficiently small then k_0 is pure imaginary and the wave grows exponentially, signaling and instability. This tells us our theory is out of control and some important new physical input is needed.

4. A massless "particle" of zero energy and momentum.

Exercise *Components Of The* $1+1$*-Dimensional Lorentz Group*
a.) Prove equation (8.57). [86]

[85] Note the factors of two, so that $x^0 p_0 + x^1 p_1 = x^+ p_+ x^- p_-$. This is an example of the tricky factors of two one encounters when working with light-cone coordinates.

[86] *Answer*: The general matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{8.59}$$

b.) Show that the group of components of the Lorentz group is $O(1,1)/SO_0(1,1) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$

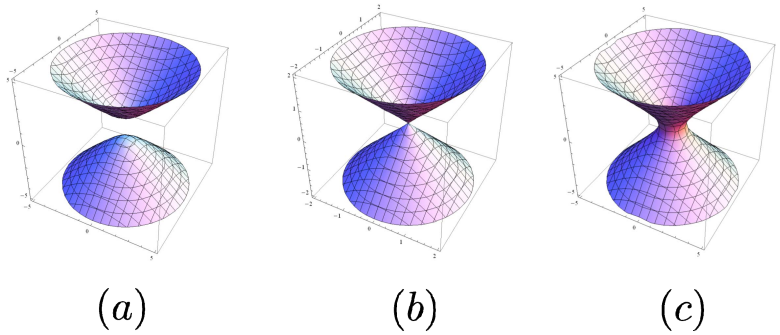


Figure 26: Illustrating orbits of the connected component of the identity in $O(1,3)$. In (a) the top and bottom hyperboloids are separate orbits, and if we include time-reversing transformations the orbits are unions of the two hyperboloids. In (b) there are three orbits shown with $x^0 > 0$ $x^0 < 0$ (the future and past, or forward and backward light cones), and the orbit consisting of the single point. In (c), once x^2 has been specified, there is just one orbit, for $d > 2$.

8.3.2 Higher Dimensional Light Cones

We define d-dimensional Minkowski space $\mathbb{M}^{1,d-1}$ with $d > 2$ to be the vector space $\mathbb{R}^d$

---

is in $O(1,1)$ iff

$$\begin{aligned} a^2 - c^2 &= 1 \\ d^2 - b^2 &= 1 \\ ab &= cd \end{aligned} \tag{8.60}$$

The most general solution of the first two equations is

$$\begin{aligned} a &= \kappa_1 \cosh\theta \\ c &= \sinh\theta \\ d &= \kappa_2 \cosh\theta' \\ b &= \sinh\theta' \end{aligned} \tag{8.61}$$

where $\kappa_i \in \{\pm 1\}$ and $\theta, \theta' \in \mathbb{R}$. Now impose the third equation. The solutions split into two cases: If $\kappa_1/\kappa_2 = 1$ then $\theta = \theta'$. This gives two components. If $\kappa_1/\kappa_2 = -1$ then $\theta = -\theta'$, giving the other two components.

with quadratic form

$$\eta = Diag\{-1, +1_{d-1}\} \tag{8.62}$$

and

$$O(1, d-1) = \{A | A^{tr} \eta A = \eta\} \tag{8.63}$$

The nature of the orbits is slightly different because of the zero-dimensional sphere S^0 is disconnected but the higher dimensional spheres are connected.

1. For $\lambda^2 > 0$ we can define

$$\mathcal{O}^+(\lambda) = \{x | (x^0)^2 - (\vec{x})^2 = \lambda^2 \qquad \& \qquad \text{sign}(x^0) = \text{sign}(\lambda)\} \tag{8.64}$$

   By the stabilizer-orbit theorem we can identify this with

$$SO_0(1, d-1)/SO(d-1) \tag{8.65}$$

   by considering the isotropy group at $(x^0 = \lambda, \vec{x} = 0)$. See Figure 26(a).

2. For $\mu^2 > 0$ we can define

$$\mathcal{O}^-(\lambda^2) = \{x | (x^0)^2 - (\vec{x})^2 = -\mu^2\} \tag{8.66}$$

   By the stabilizer-orbit theorem we can identify this with

$$SO_0(1, d-1)/SO_0(1, d-2) \tag{8.67}$$

   by considering the isotropy group at $x = (x^0 = 0, x^1 = 0, \ldots, x^{d-2} = 0, x^{d-1} = \mu)$. The sign of μ does not distinguish different orbits for $d > 2$ because the sphere S^{d-2} is connected. See Figure 26(c).

3.

$$\mathcal{O}^\pm = \{x | x^2 = 0 \qquad \& \qquad \text{sign}(x^0) = \pm 1\} \tag{8.68}$$

   Vectors in this orbit are of the form $(x^0, |x^0|\hat{n})$ where $\hat{n} \in S^{d-2} \subset \mathbb{R}^{d-1}$ and the sign of x^0 is invariant under the action of the identity component of $O(1, 3)$. (Show this!). Note that, for $d = 2$ the sphere S^0 has two disconnected components, leading to left- and right-movers. But for $d > 2$ there is only one component. We can think of $\hat{n} \in S^{d-2}$ as parametrizing the directions of light-rays. That is, the point where the light ray hits the celestial sphere. In one spatial dimension, a light ray either moves left or right, and this is a Lorentz-invariant concept. In $d - 1 > 1$ spatial dimensions, we can rotate any direction of light ray into any other. See Figure 26(b). One can show that these orbits too are homogeneous spaces: [87]

$$\mathcal{O}^\pm \cong SO_0(1, d-1)/\mathcal{I} \tag{8.69}$$

[87] The isotropy group of a light ray is $\mathcal{I} \cong ISO(d-2)$, where $ISO(d-2)$ is the Euclidean group on $\mathbb{R}^{d-2}$. The easiest way to show this is to use the Lie algebra of $so(1, d-1)$ and work with light-cone coordinates. Choosing a direction of the light ray along the x^{d-1} axis and introducing light-cone coordinates $x^\pm := x^0 \pm x^{d-1}$, and transverse coordinates x^i, $i = 1, \ldots, d-2$ if the lightray satisfies $x^- = 0$ then we have unbroken generators M^{+i} and M^{ij}.

4. The final orbit is of course $\{x = 0\}$.

Remarks

1. *Discrete symmetries in nature.* The higher dimensional Lorentz groups also have four connected components. The Lorentz group in $d+1$ spacetime dimensions is $O(d-1,1)$. In QFT it is clear that if a theory is invariant under infinitesimal Lorentz symmetries, then it is invariant under the connected component of the identity $SO_0(d-1,1)$.

   However, it turns out that such theories can nevertheless not be invariant under the disconnected components. For a long time it was assumed that parity is a symmetry of all the laws of physics. This means that, if you watch a video of a physical process, then you cannot tell whether you are looking at that process in a mirror, or not. However, in 1956 T.D. Lee and C.N. Yang carefully reviewed the evidence for parity conservation in nature and pointed out that there was no careful experimental test for the weak interactions (nuclear beta decay, etc.). They proposed some experimental tests and C.S. Wu and collaborators discovered experimentally in 1957 that parity is indeed violated.

   So, what about time reversal invariance? If you run a movie backwards, is the resulting process physically possible (however unlikely)? There is a famous theorem in QFT stating that the product of parity and time reversal, the PT-component of $O(d-1,1)$ will be a symmetry if the connected component is a symmetry. This is usually called the CPT theorem. In 1964 V. Fitch and J. Cronin discovered that certain very rare processes in nature actually do violate time-reversal invariance. That is a good thing, because, as noted by Sakharov, if there were no time-reversal invariance in the laws of physics it would be impossible to understand why there is matter/anti-matter asymmetry in the context of the big bang theory.

2. An important generalization of the Lorentz groups is the following:

$$O(p,q;\mathbb{R}) := \{A \in GL(p+q;\mathbb{R}) | A^{tr} d_{p,q} A = d_{p,q}\} \tag{8.70}$$

   where

$$d_{p,q} := \begin{pmatrix} 1_{p\times p} & 0 \\ 0 & -1_{q\times q} \end{pmatrix} \tag{8.71}$$

   forms a group, generalizing the Lorentz and rotation-reflection groups and, when both $p > 0$ and $q > 0$ it has four connected components. [88]

   ****

   SHOW CONTRACTION TO $O(p) \times O(q)$ USING THE MATRICES

$$\exp[\begin{pmatrix} 0 & \pi \\ \pi^\dagger & 0 \end{pmatrix}] = \begin{pmatrix} \cosh(\sqrt{\pi\pi^\dagger} & \pi \frac{\sinh(\sqrt{\pi^\dagger\pi})}{\sqrt{\pi^\dagger\pi}} \\ \pi^\dagger \frac{\sinh(\sqrt{\pi\pi^\dagger})}{\sqrt{\pi\pi^\dagger}} & \cosh(\sqrt{\pi^\dagger\pi} \end{pmatrix}$$

[88] The proof proceeds by showing that the group contracts to $O(p) \times O(q)$ and then recall that $O(p)$ has two connected components.

SINCE

$$\begin{aligned} A^{tr}A &= 1 + C^{tr}C \\ D^{tr}D &= 1 + B^{tr}B \end{aligned} \tag{8.72}$$

A, D are positive have polar decompositions $A = O_1P_1$ and $D = O_2P_2$ etc.

Detailed proof in Survey Of Matrix Groups 2009, Chapter 5.

****

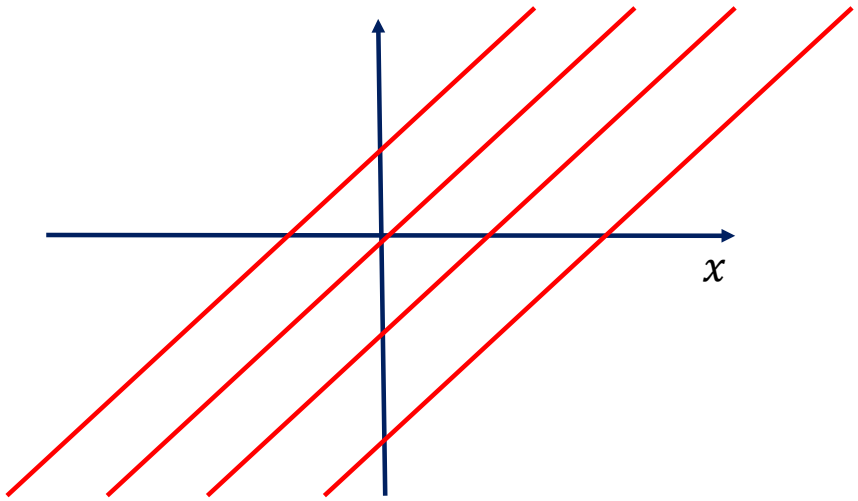


Figure 27: The torsors $\mathbb{Z} + x$, plotted in the y direction are the points above a point x on the $\mathbb{Z}$-axis. Note that the union of the red lines has a symmetry under the action $(x, y) \to (x + n, y)$ for $n \in \mathbb{Z}$. If we quotient by this action then the projection to the x-axis becomes the projection $\mathbb{R} \to S^1$ given by the exponential map. This gives us a principal bundle for the group $\mathbb{Z}$.

8.3.3 Torsors And Principal Bundles

Definition A *torsor* or *principal homogeneous space* for a group G is a G-set X on which the action is transitive and free.

One can set up a 1-1 correspondence between a torsor and elements of G, but in general there is no natural correspondence: A torsor has no distinguished element we can call the identity.

Example 1: Let $x \in \mathbb{R}$ be a real number. Consider the subset $X = \mathbb{Z} + x \subset \mathbb{R}$. This set is a torsor for $\mathbb{Z}$. But there is no natural zero in X. Indeed, let x vary continuously, then any purported natural zero would vary continuously to any other number.

Example 2: Imagine that the surface of the earth is flat and of infinite extent. Is this a copy of $\mathbb{R}^2$? Yes and no. We can identify it with $\mathbb{R}^2$, but not in any natural way: $\mathbb{R}^2$

is a vector space with a distinguished vector $\vec{0}$. Where should we put the origin? Rome? Beijing? Moscow? London? New York? Piscataway? Wuhan? If the UN tried to assign an origin there would be endless disputes. However, there would never be any dispute about the vector in $\mathbb{R}^2$ needed to translate from New York to London. So, the difference of London minus New York is well-defined, but the sum is not. If London and New York represented vectors in a vector space then one could both add and subtract these vectors.

The infinite flat earth is an example of two-dimensional affine Euclidean space $\mathbb{E}^2$. More formally: An *affine space* $\mathbb{E}^d$ *modeled on* $\mathbb{R}^d$ is a space of points with an action of $\mathbb{R}^d$ that translates the points so that nonzero vectors always move points and one can get from one point to any other by the action of a vector. But there is no natural choice of origin. In equations:

1. If $v \in \mathbb{R}^d$ and $p \in \mathbb{E}^d$ then there is a point $p+v \in \mathbb{E}^d$ so that $(p+v)+v' = p+(v+v')$.
2. If $p + v = p$ then $v = 0$.
3. If $p, p' \in \mathbb{E}^d$ there is a (unique) vector $v \in \mathbb{R}^d$ so that $p' = p + v$. We can therefore say $p' - p = v$.

If we do choose an origin (this choice is arbitrary) then we can identify $\mathbb{E}^d \cong \mathbb{R}^d$. Indeed, given the above statements, given $p \in \mathbb{E}^d$, every $p' \in \mathbb{E}^d$ is of the form $p' = p + v$ for a unique $v \in \mathbb{R}^d$. So we map $\Psi_p : \mathbb{E}^d \to \mathbb{R}^d$ by taking $\Psi_p : p' \mapsto v$.

In this language we can say that $\mathbb{E}^d$ is a principal homogeneous space for the Abelian group $\mathbb{R}^d$.

There is a distance between points which we take to be the Euclidean norm of the vector:

$$dist(p, p') := \| v \| \tag{8.73}$$

Now we can study the group of *isometries of* $\mathbb{E}^d$. This is the group of transformations $T : \mathbb{E}^d \to \mathbb{E}^d$ that preserves these distances:

$$dist(T(p), T(p')) = dist(p, p') \tag{8.74}$$

We denote it as $\mathrm{Euc}(d)$ and refer to it as the *Euclidean group.*

In a similar way, we can consider d-dimensional Minkowski space $\mathbb{M}^{1,d-1}$ to be an affine space modeled on $\mathbb{R}^d$, but now with quadratic form on $p' - p = v$ given by

$$v \cdot v = v^\mu \eta_{\mu\nu} v^\nu \tag{8.75}$$

♣Earlier we defined it as a vector space, so need to alter that to avoid confusion. ♣

so the "distance squared" between two points is now $dist(p, p')^2 = v \cdot v$. Considered as an affine space we define the *Poincaré group* as the group of transformations $T : \mathbb{M}^{1,d-1} \to \mathbb{M}^{1,d-1}$ preserving the squared distance.

Example 3: Let V be a finite-dimensional vector space over a field κ. The set $X = \mathcal{B}(V)$ of all ordered bases for V is a $GL(n, \kappa)$ torsor: $g \cdot \{v_1, \ldots, v_n\} := \{\tilde{v}_1, \ldots, \tilde{v}_n\}$ where $\tilde{v}_i = g_{ji} v_j$. Any two such bases are related by some g, but, if we are just given an abstract vector space,

there is no natural basis. In the case $\kappa = \mathbb{R}$ the torsor $\mathcal{B}(V)$ has two connected components. A choice of connected component is known as an *orientation of* V.

Example 4: It is quite interesting to consider continuous families of torsors. For example, let us return to Example 1 and vary x so that we have a family of torsors of the form $\mathbb{Z}+x$. Let

$$\tilde{P} = \cup_{x\in\mathbb{R}}(\mathbb{Z}+x) \tag{8.76}$$

Note that $\mathbb{Z}+x+1 = \mathbb{Z}+x$ so that we can quotient by the group action of $\mathbb{Z}$ on $\mathbb{R}$ to get

$$P = \tilde{P}/(x,y) \sim (x+1,y) \tag{8.77}$$

See Figure 27.

Example 5: Let us return to our example $\pi_{\hat{n}} : SO(3) \to S^2$, or more generally,

$$\pi : G \to G/H \tag{8.78}$$

Note that the fibers of a coset gH are:

$$\pi^{-1}(gH) = gH \tag{8.79}$$

The subset $gH \subset G$ can be put into 1-1 correspondence with H, but not in any natural way. While it is identified with H as a set, it is not identified as a group because there is no natural element in gH to identify with the unit in H.

The above examples are special cases of an extremely important idea in mathematics - that of a *principal fiber bundle*. In the language of this section, a principal G-bundle is a continuous family of G-torsors. Here is the formal definition:

Definition: Let G be a topological group. A *principal G-bundle* is a pair of topological spaces P and X together with a surjective continuous map $\pi : P \to X$ so that there is a continuous free right G action (not necessarily transitive!) on P with

$$\pi(p\cdot g) = \pi(p) \tag{8.80}$$

Technically, we need it to have a local triviality property: For all $x \in X$ there is a neighborhood $\mathcal{U}_x \subset X$ with a smooth map $\phi_x : \pi^{-1}(\mathcal{U}_x) \to \mathcal{U}_x \times G$ such that

$$\begin{array}{ccc} \pi^{-1}(\mathcal{U}_x) & \xrightarrow{\phi_x} & \mathcal{U}_x \times G \\ & \searrow^{\pi} & \downarrow{\scriptstyle \pi} \\ & & \mathcal{U}_x \end{array} \tag{8.81}$$

commutes. Moreover, we require that the right G-action on $\pi^{-1}(\mathcal{U}_x)$ is equivalent to the natural right G-action on $\mathcal{U}_x \times G$, where g_0 acts by $(y,g) \mapsto (y, gg_0)$, i.e. if $p \in \pi^{-1}(\mathcal{U}_x)$ and $\phi_x(p) = (y,g)$ then

$$\phi_x(p\cdot g_0) = (y, g\cdot g_0) \tag{8.82}$$

P is called the *total space* of the principal bundle and X is called the *base space*. The preimage $\pi^{-1}(x)$ is called the *fiber above* x.

A consequence of the local triviality property is that if we cover the base space X with charts $\{\mathcal{U}_\alpha\}$ with $\phi_\alpha : \pi^{-1}(\mathcal{U}_\alpha) \to \mathcal{U}_\alpha \times G$ then

$$\phi_{\alpha\beta} := \phi_\alpha \circ \phi_\beta^{-1} : (x, g) \mapsto (x, g_{\alpha\beta}(x)g) \qquad x \in \mathcal{U}_{\alpha\beta} \tag{8.83}$$

Note that on triple overlaps $\mathcal{U}_{\alpha\beta\gamma}$

$$g_{\alpha\beta}(x)g_{\beta\gamma}(x)g_{\gamma\alpha}(x) = 1 \qquad x \in \mathcal{U}_{\alpha\beta\gamma} \tag{8.84}$$

a condition called the *cocycle condition*.

Examples:

1. Let X be any topological space and G any topological group. Then $P = X \times G$ with $\pi : (x, g) \mapsto x$ is a principal bundle. A bundle of this form is known as the trivial bundle.

2. $\pi : G \to G/H$ is a principal H bundle over the homogeneous space $X = G/H$. (Note the free right H-action that commutes with π is $g \mapsto g \cdot h$ so that $\pi(g) = gH = \pi(gh)$.

3. In particular, $\pi : SO(3) \to S^2$ is a principal $SO(2)$ bundle over S^2

4. $\pi : SU(2) \to \mathbb{CP}^1$ is a principal $U(1)$ bundle over $\mathbb{CP}^1 \cong S^2$.

5. *G-bundles over the circle.* Let G be a discrete group. Then $\mathbb{R} \times G$ is the trivial principal G- bundle over $\mathbb{R}$. We can make a more interesting bundle by considering the left $\mathbb{Z}$-action defined by choosing an element $g_0 \in G$ and defining

   $$\phi_{g_0}((x, g), n) = (x + n, g_0^n g) \tag{8.85}$$

   so equivalence classes satisfy: $[(x, g)] = [(x + n, g_0^n g)]$. The quotient $P = (\mathbb{R} \times G)/\mathbb{Z}$ by this action is a principal G-bundle over S^1. Note that there is still a well-defined free right G-action. Call this bundle $\pi : P_{g_0} \to S^1$. We could also describe it as $([0, 1] \times G)$ where we glue $(0, g)$ to $(1, g_0 g)$. We will return to it below.

6. A nice special case of the bundle P_{g_0} described above is obtained by considering the map of the unit disk in the complex plane $\pi : z \mapsto z^n$. Restricting this map to the boundary of the disk we get a map from $S^1 \to S^1$, but notice that the inverse image of any point is a torsor for multiplication by elements of the group μ_n. So π describes a principle $\mathbb{Z}_n$ bundle over the circle.

A *map of G-torsors*, or, more properly, a *morphism of G-torsors* is a map the preserves the mathematical structure of being a G-torsor. So, if $\mathcal{X}_1$ and $\mathcal{X}_2$ are two G-torsors then a morphism of G-torsors is a map

$$\psi : \mathcal{X}_1 \to \mathcal{X}_2 \tag{8.86}$$

such that

$$\psi(y \cdot g) = \psi(y) \cdot g \tag{8.87}$$

for all $y \in \mathcal{X}_1$. Such maps are also said to be *equivariant*.

A *bundle map* between two principal G-bundles, or, more properly a *morphism of G-bundles* is a continuous map that preserves fibers and restricts on fibers to a be a morphism of G-torsors. In diagrams, if ψ is a bundle map between two principal G-bundles P_1 and P_2 then $\psi : P_1 \to P_2$ is a map so that it preserves fibers, meaning that:

$$\begin{array}{ccc} P_1 & \xrightarrow{\psi} & P_2 \\ & {\scriptstyle \pi_1} \searrow \quad \swarrow {\scriptstyle \pi_2} & \\ & X & \end{array} \tag{8.88}$$

commutes. That is $\pi_2(\psi(p_1)) = \pi_1(p_1)$ for all $p_1 \in P$. Moreover, the map must be G equivariant, or a morphism of torsors on the fibers, so that [89]

$$\psi(p_1 \cdot g) = \psi(p_1) \cdot g \tag{8.89}$$

If there are bundle maps $\psi_1 : P_1 \to P_2$ and $\psi_2 : P_2 \to P_1$ whose composition is the identity then the bundles are said to be *isomorphic bundles*.

Let us return to our example of $\pi : P_{g_0} \to S^1$. Note that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R} \times G & \xrightarrow{\psi_h} & \mathbb{R} \times G \\ {\scriptstyle \phi_{g_0,n}} \downarrow & & \downarrow {\scriptstyle \phi_{hg_0h^{-1},n}} \\ \mathbb{R} \times G & \xrightarrow{\psi_h} & \mathbb{R} \times G \end{array} \tag{8.90}$$

This implies that ψ_h descends to a well-defined bundle map $P_{g_0} \to P_{hg_0h^{-1}}$. Clearly $\psi_{h^{-1}}$ defines the inverse bundle map.

Therefore, the isomorphism classes of principal G bundles over the circle are labeled by conjugacy classes of elements of G.

Finally, let $\pi : P \to X$ be a principal G bundle and let Y be any G-space with a left G-action. Then we can define a G action on $P \times Y$:

$$\phi_g(p, y) \mapsto (p \cdot g^{-1}, g \cdot y) \tag{8.91}$$

Notice this is a left G-action. The quotient space, usually denoted $P \times_G Y$ has a well-defined continuous map

$$\tilde{\pi} : P \times_G Y \to X \tag{8.92}$$

defined by $\tilde{\pi}([p, y]) = \pi(p)$. The fibers of $\tilde{\pi}$ can be identified with the space Y. (See the exercise below.) This is an example of a more general fiber bundle. The map $\tilde{\pi} : P \times_G Y \to$

[89] The reader familiar with the general theory of fiber bundles will note that (8.88) alone serves as the definition of a bundle map for a fiber bundle. But a principal bundle has more structure, and for a morphism of principal bundles we require the additional condition (8.89).

X defines what is called *an associated bundle to the principal bundle* $\pi : P \to X$ *by the* G*-set* Y.

Sections Of Bundles

If $\pi : E \to X$ is a principal bundle, or an associated bundle to a principal bundle (or any fiber bundle, if you know what that means) then we define a *section of* π to be a continuous map

$$s : X \to E \tag{8.93}$$

which is a right-inverse to π. That is

$$\pi \circ s(x) = x \qquad\qquad \forall x \in X \tag{8.94}$$

Note that $s(x)$ is always an element of E which is in the fiber of $\pi : E \to X$ above x.

Because of local triviality, local sections always exist. That is, near any $x \in X$ there will be sections in the bundle $\pi^{-1}(\mathcal{U}) \to \mathcal{U}$ for some neighborhood $\mathcal{U}$ of x. It is less obvious what happens globally. In fact, we have

Theorem A principal G-bundle $\pi : P \to X$ is isomorphic to the trivial bundle $\pi : X \times G \to X$ iff there is a globally defined continuous section.

Proof: Suppose there is a globally defined section $s : X \to P$. Then let

$$\psi : X \to G \to P \tag{8.95}$$

be defined by $\psi(x, g) := s(x)g$. One checks this is a bundle morphism. Conversely, suppose there is a bundle morphism $\psi : P \to G \times X$. Then for each $x \in X$ there is a unique $s(x) \in P$ so that $\psi(s(x)) = (x, 1_G)$. ♠

The situation is rather different for associated bundles. For example if $Y = V$ is a linear space then the space of sections is an infinite dimensional vector space.

♣Explain more and write this out with transition functions ♣

Exercise *Bundle Maps For Trivial Bundles*

Show that the most general bundle map from the trivial bundle $\pi : X \times G \to X$ to itself is of the form:

$$\psi : (x, g) \mapsto (x, h(x)g) \tag{8.96}$$

for some continuous map $h : X \to G$.

Exercise *Morphisms Of Principal G-Bundles Are Isomorphisms*

a.) Show that any morphism of G-torsors is an isomorphism of G-torsors.

b.) Extend this to show that any morphism of principal G-bundles is an isomorphism of principal G-bundles.

Exercise *Fibers Of An Associated Bundle*

Show that the fibers of the associated bundle $\tilde{\pi} : P \times_G Y \to X$ described above are in 1-1 correspondence with the space Y. [90]

8.4 More About Induced Group Actions On Function Spaces

Let us return to the considerations of section 4.2.

Let X be a G-set and let Y be any set. There are natural left- and right- actions on the function space $\mathrm{Map}(X, Y)$. Given $\Psi \in \mathrm{Map}(X, Y)$ and $g \in G$ we need to produce a new function $\phi(g, \Psi) \in \mathrm{Map}(X, Y)$. The rules are as follows:

1. If G is a left-action on X then
$$\phi(g, \Psi)(x) := \Psi(g \cdot x) \qquad \text{right action on } \mathrm{Map}(X, Y) \tag{8.97}$$
2. If G is a left-action on X then
$$\phi(g, \Psi)(x) := \Psi(g^{-1} \cdot x) \qquad \text{left action on } \mathrm{Map}(X, Y) \tag{8.98}$$
3. If G is a right-action on X then
$$\phi(g, \Psi)(x) := \Psi(x \cdot g) \qquad \text{left action on } \mathrm{Map}(X, Y) \tag{8.99}$$
4. If G is a right-action on X then
$$\phi(g, \Psi)(x) := \Psi(x \cdot g^{-1}) \qquad \text{right action on } \mathrm{Map}(X, Y) \tag{8.100}$$

Example: Consider a spacetime $\mathcal{S}$. With suitable analytic restrictions the space of scalar fields on $\mathcal{S}$ is $\mathrm{Map}(\mathcal{S}, \kappa)$, where $\kappa = \mathbb{R}$ or $\mathbb{C}$ for real or complex scalar fields. If a group G acts on the spacetime, there is automatically an induced action on the space of scalar fields. To be even specific, suppose $X = \mathbb{M}^{1,d-1}$ is d-dimensional Minkowski space time, G is the Poincaré group, and $Y = \mathbb{R}$. Given one scalar field Ψ and a Poincaré transformation $g^{-1} \cdot x = \Lambda x + v$ we have $(g \cdot \Psi)(x) = \Psi(\Lambda x + v)$.

[90] *Answer*: Let us define a map $f : \tilde{\pi}^{-1}(x) \to Y$. To do this we must choose an element $p_0 \in \pi^{-1}(x)$. Then if $[p, y] \in \tilde{\pi}^{-1}(x)$ it follows that $\pi(p) = x$ and hence $p = p_0 \cdot g_0$ for some g_0. Then we define $f[p, y] = g_0^{-1} \cdot y$. The reader needs to check that this map is well-defined. Since we can choose any $y \in Y$ the map is clearly surjective. Finally note that $[p_1, y_1] = [p_1, y_2]$ implies that $y_1 = y_2$ since the G-action on P is free. Therefore the map is also injective. Note that the map does depend on a choice of p_0 for each x, so there is no canonical identification of the fibers with Y.

Similarly, suppose that X is any set, but now Y is a G-set. Then again there is a G-action on $\mathrm{Map}(X,Y)$:

$$(g\cdot\Psi)(x) := g\cdot\Psi(x) \qquad \text{or} \qquad \Psi(x)\cdot g \tag{8.101}$$

according to whether the G action on Y is a left- or a right-action, respectively. These are left- or right-actions, respectively.

We can now combine these two observations and get the general statement: We assume that both X is a G_1-set and Y is a G_2-set. We can assume, without loss of generality, that we have left-actions on both X and Y. Then there is a natural $G_1\times G_2$-action on $\mathrm{Map}(X,Y)$ defined by:

$$\phi((g_1,g_2),\Psi)(x) := g_2\cdot(\Psi(g_1^{-1}\cdot x)) \tag{8.102}$$

note that if one writes instead $g_2\cdot(\Psi(g_1\cdot x))$ on the RHS then we do not have a well-defined $G_1\times G_2$-action (if G_1 and G_2 are both nonabelian). In most applications X and Y both have a G action for a single group and we write

$$\phi(g,\Psi)(x) := g\cdot(\Psi(g^{-1}\cdot x)) \tag{8.103}$$

This is a special case of the general action (8.102), with $G_1=G_2=G$ and specialized to the diagonal $\Delta\subset G\times G$.

Example: Again let $X=\mathbb{M}^{1,d-1}$ be a Minkowski space time. Take $G_1=G_2$ and let $G=\Delta\subset G\times G$ be the diagonal subgroup, and take G to be the Poincaré group. Now let $Y=V$ be a finite-dimensional representation of the Poincaré group. Let us denote the action of $g\in G$ on V by $\rho(g)$. Then a field $\Psi\in\mathrm{Map}(X,Y)$ has an action of the Poincaré group defined by

$$g\cdot\Psi(x) := \rho(g)\Psi(g^{-1}x) \tag{8.104}$$

This is the standard way that fields with nonzero "spin" transform under the Poincaré group in field theory. As a very concrete related example, consider the transformation of electron wavefunctions in nonrelativistic quantum mechanics. The electron wavefunction is governed by a two-component function on $\mathbb{R}^3$:

$$\Psi(\vec{x}) = \begin{pmatrix}\psi_+(\vec{x})\\ \psi_-(\vec{x})\end{pmatrix} \tag{8.105}$$

Then, suppose $G=SU(2)$. Recall there is a surjective homomorphism $\pi: G\to SO(3)$ defined by $\pi(u)=R$ where

$$u\vec{x}\cdot\vec{\sigma}u^{-1} = (R\vec{x})\cdot\vec{\sigma} \tag{8.106}$$

Then the (double-cover) of the rotation group acts to define the transformed electron wavefunction $u\cdot\Psi$ by

$$(u\cdot\Psi)(\vec{x}) := u\begin{pmatrix}\psi_+(R^{-1}\vec{x})\\ \psi_-(R^{-1}\vec{x})\end{pmatrix} \tag{8.107}$$

In particular, $u=-1$ acts trivially on $\vec{x}$ but nontrivially on the wavefunction.

♣IT HASN'T BEEN DEFINED YET. IT IS IN SECTION 10.1. SOME OF THAT MATERIAL SHOULD BE MOVED EARLIER SO WE CAN RECALL IT HERE. ♣

9. Centralizer Subgroups And Counting Conjugacy Classes

Definition 9.1: Let $g \in G$, the *centralizer subgroup* of g, (also known as the *normalizer subgroup*), denoted, $Z(g)$, is defined to be:

$$Z(g) := \{h \in G | hg = gh\} = \{h \in G | hgh^{-1} = g\} \tag{9.1}$$

Exercise *Due Diligence*
a.) Check that $Z(g) \subset G$ is a subgroup.
b.) Show that $g^n \in Z(g)$ for any integer n.
c.) If $g_1 = g_0 g_2 g_0^{-1}$ show that $Z(g_1) = g_0 Z(g_2) g_0^{-1}$.
d.) Show that

$$Z(G) = \cap_{g \in G} Z(g) \tag{9.2}$$

Exercise *Is $Z(g)$ always an Abelian group?*
a.) Show that $Z(1) = G$. Answer the above question.
b.) Show that the centralizer of the transposition (12) in S_n for $n \leq 3$ is isomorphic to S_2.
c.) Show that the centralizer of the transposition (12) in S_n for $n \geq 4$ is isomorphic to $S_2 \times S_{n-2}$.

Recall that $C(g)$ denotes the conjugacy class of g. Using the Stabilizer-Orbit theorem we can establish a 1-1 correspondence between $C(g)$ and the cosets of $G/Z(g)$. As in the proof of that theorem we have a map $\psi : G/Z(g) \to C(g)$ by

$$\psi : g_i Z(g) \to g_i g g_i^{-1} \in C(g) \tag{9.3}$$

It is 1-1 and onto.

Since conjugacy is an equivalence relation G decomposes as a disjoint union of the orbits, which in this case are the conjugacy classes. When G is a finite group this decomposition leads to some useful theorems based on simple counting ideas. When $|G|$ is finite we can usefully write:

$$|G| = \sum_{conj.\ classes} |C(g)| \tag{9.4}$$

The sum is over distinct conjugacy classes. What is g in this formula? For each class we may choose any representative element from that class.

Now, if G is finite, then by the above 1-1 correspondence we may write:

$$|C(g)| = \frac{|G|}{|Z(g)|} \tag{9.5}$$

which allows us to write the above decomposition of $|G|$ in a useful form sometimes called the *class equation*:

$$|G| = \sum_{conj.\ classes} \frac{|G|}{|Z(g)|} \tag{9.6}$$

Again, we sum over a complete set of distinct non-conjugate elements g. Which g we choose from each conjugacy class does not matter since if $g_1 = hg_2h^{-1}$ then $Z(g_1) = hZ(g_2)h^{-1}$ are conjugate groups, and hence have the same order. So, for each distinct conjugacy class we just choose any element we like.

9.1 $0+1$-Dimensional Gauge Theory

Recall the definition above of a morphism of bundles, and an isomorphism of bundles. An automorphism of a principal G-bundle $\pi : P \to X$ is an isomorphism of the principal bundle with itself. Since the composition of automorphisms is an automorphism, and automorphisms are invertible, and the identity is an automorphism the set of automorphisms form a group under composition, called the *group of automorphisms of the bundle* $\pi : P \to X$. Recall the bundles $P_{g_0} \to S^1$ determined by the group element g_0. As we showed above, the invertible bundle maps are of the form $\psi_h : P_{g_0} \to P_{hg_0h^{-1}}$. Therefore, the isomorphism classes of G-bundles over the circle are in 1-1 correspondence with the conjugacy classes in G, and the group of automorphisms of P_{g_0} is precisely the centralizer group $Z(g_0)$.

A *gauge theory* is a physical theory where physical quantities are defined by summing over principal G-bundles with connection. We haven't defined the term "connection," but for principal G-bundles with discrete group there is a unique connection so we can discuss those here. [91] In the case of $0+1$ dimensional gauge theory a basic quantity of interest is the "partition function" on a one-dimensional manifold. The only closed connected one-dimensional manifold is the circle. So we have

$$Z(S^1) = \sum F(P_g) \tag{9.7}$$

where the sum is over isomorphism classes of principal G bundles over the circle. and F is a function on the set of such bundles, or equivalently, a function on all the principal G-bundles that only depends on the isomorphism class. We call this a *gauge invariant Boltzman factor.* The standard physical Boltzman factors involve curvature and holonomy. In this setting there is no curvature, so F should be proportional to a character of g in some representation.

In gauge theory we must also divide by the "volume" of the group of automorphisms of the bundle so

$$Z(S^1) = \sum_{cc} \frac{\chi_\rho(g)}{|Z(g)|} \tag{9.8}$$

for some character χ_ρ. We will see later that the sum is zero unless ρ contains some copies of the trivial representation, so we might as well take $\chi_\rho(g) = 1$.

[91] See section 18 below for a discussion of gauge theory that uses minimal prerequisites and is sufficient to understand this remark.

Now notice that we can use the stabilizer-orbit theorem to rewrite this as:

$$Z(S^1) = \sum_{cc} \frac{1}{|Z(g)|} = \frac{1}{|G|} \sum_{P_g} 1 \tag{9.9}$$

In the second equality we have summed over all the G-bundles P_g weighted by 1 and divided by the full "volume" $|G|$ of the gauge group. So

$$Z(S^1) = \frac{1}{|G|} \sum_{g \in G} 1 = 1 \tag{9.10}$$

The Hilbert space is one-dimensional with zero Hamiltonian so indeed this is $\mathrm{Tr} e^{-\beta H} = 1$.

9.2 Three Mathematical Applications Of The Counting Principle

In this section let p be a prime number.

Application 1:

Theorem: If $|G| = p^n$ then the center is nontrivial, i.e., $Z(G) \neq \{1\}$.

Proof: Observe that an element g is central *if and only if* $C(g) = \{g\}$ has order 1. Now let us use the class equation. We can usefully split up the sum over conjugacy classes as a sum over the center and the rest:

$$|G| = |Z(G)| + \sum_i |\mathcal{C}_i| \tag{9.11}$$

where the sum over i is a sum over the distinct conjugacy classes more than one element. As we noted above, by the stabilizer orbit theorem

$$|\mathcal{C}_i| = \frac{|G|}{|Z(g_i)|} \tag{9.12}$$

where g_i is any element of the conjugacy class $\mathcal{C}_i$. But, for these conjugacy classes $|Z(g_i)| < |G|$ and by Lagrange's theorem, and the assumption that p is prime, $|Z(g_i)| = p^{n-n_i}$ for some $n_i < n$. Therefore, the second term on the RHS of (9.11) is divisible by p and hence $p||Z(G)|$. ♠

Application 2: *Cauchy's theorem*:

In a similar style, we can prove the very useful:

Theorem: If p divides $|G|$ then there is an element $g \in G$, $g \neq 1$ with order p.

Proof 1: This is a nice application of the stabilizer-orbit theorem. Consider the set

$$X = \{(g_1, \ldots, g_p) | g_1 \cdots g_p = 1\} \subset G^p \tag{9.13}$$

♣Proof 1 is not really an application of the class equation, rather it is an application of stabilizer-orbit. ♣

Note that the cyclic group $\mathbb{Z}_p$ acts on this set with the standard generator acting by

$$\omega \cdot (g_1, \ldots, g_p) = (g_p, g_1, g_2, \ldots, g_{p-1}) \tag{9.14}$$

A fixed point of the $\mathbb{Z}_p$-action corresponds to an element of the form $(g, \ldots, g)$ such that $g^p = 1$. If $g \neq 1$ then this corresponds to an element of order p. Now, by the stabilizer-orbit theorem, the orbits of any $\mathbb{Z}_p$ action (on any set) have cardinality either 1 or p. Let N_1 be the number of orbits of length one and let N_p be the number of orbits of length p. Note that the order of X is just $|G|^{p-1}$ since one can always solve for g_p in terms of $g_1, \ldots, g_{p-1}$. Then, by the counting principle we have:

$$|G|^{p-1} = N_1 + pN_p \tag{9.15}$$

It follows that p divides N_1. Also $N_1 > 0$ because $(1, ..., 1)$ is a fixed point of the $\mathbb{Z}_p$ action. Therefore $N_1 = kp > 1$ and hence there are other fixed points, i.e. there are group elements of order p. In fact, there must be at least $(p-1)$ of them. ♠

Proof 2: We can also prove Cauchy's theorem using induction on the order of G, dividing the proof into two cases: First we consider the case where G is Abelian and then the case where it is nonabelian.

Case 1: G is Abelian:

If $|G| = p$ then G is cyclic and the statement is obvious: Any generator has order p. More generally, note that if G is a cyclic group $\mathbb{Z}/N\mathbb{Z}$ with $N > p$ and p divides N then $\overline{N/p} \in \mathbb{Z}/N\mathbb{Z}$ has order p. This establishes the result for cyclic groups.

Now suppose our Abelian group has order $|G| > p$. Choose an element $g_0 \neq 1$ and suppose that g_0 does not have order p. Let $H = \langle g_0 \rangle$. If $H = G$ then G would be cyclic but then as we just saw, it would have an element of order p. So now assume H is a proper subgroup of G. If p divides $|H|$ then H (and hence G) has an element of order p by the inductive hypothesis. If p does not divide $|H|$ then we consider the group G/H. But this has order strictly less than $|G|$ and p divides the order of G/H. So there is an element aH of order p meaning $a^p = g_0^x$ for some x. If $g_0^x = 1$ we are done. If not then there is some smallest positive integer y so that $g_0^{xy} = 1$ but then a^y has order p. We have now proved Cauchy's theorem for abelian groups.

Case 2: G is non-Abelian: By the class equation we can write

$$|G| = |Z(G)| + \sum{}' \frac{|G|}{|Z(g_i)|} \tag{9.16}$$

If p divides the order of the centralizer $Z(G)$ then we can apply our previous result about Cauchy's theorem for Abelian groups. If p does not divide $Z(G)$ then there must be some g_i so that p does not divide $\frac{|G|}{|Z(g_i)|}$ but this means p divides $|Z(g_i)|$, but now by the inductive hypothesis $Z(g_i)$, and hence G has an element of order p. This completes the proof. ♠

Application 3: *Sylow's theorem*:

Finally, as a third application we give a simple proof of Sylow's first theorem: If p is prime and p^k divides $|G|$ then G has a subgroup of order p^k.

♣Proof 1 is not really an application of the class equation, rather it is an application of stabilizer-orbit. ♣

Proof 1: The first proof is again an application of the stabilizer-orbit theorem. [92] Suppose $|G| = p^{k+r}u$ with $gcd(u,p) = 1$ and $r \geq 0$ and $k > 0$. We will show that G has a subgroup of order p^k. Consider the power set $\mathcal{P}(G)$, namely the set of all subsets of G, and consider the subset of $\mathcal{P}(G)$ of <u>all subsets</u> (not subgroups!) of G of cardinality p^k. Call this set of subsets $\mathcal{P}(G, p^k)$. The cardinality of $\mathcal{P}(G, p^k)$ is clearly:

$$|\mathcal{P}(G,p^k)| = \binom{p^{k+r}u}{p^k} = p^r u \prod_{j=1}^{p^k-1} \frac{p^{k+r}u/j - 1}{p^k/j - 1} \tag{9.17}$$

In the product we have a ratio of rational numbers of the form $p^{k+r}u/j - 1$ (the denominator is a special case of this form). Any rational number r can be expressed as a product of prime powers $r = \prod_{\tilde{p} \text{ prime}} \tilde{p}^{v_{\tilde{p}}(r)}$ where the $v_{\tilde{p}}(r) \in \mathbb{Z}$ is known as the *valuation of* r *at* $\tilde{p}$ and the product runs over all primes $\tilde{p}$. Now, given a specific prime p, note that if a, b are relatively prime to p then

$$\frac{p^k a}{b} - 1 = \frac{p^k a - b}{b} \tag{9.18}$$

and hence for such rational numbers r the integer $v_p(r) = 0$ for the prime p. It follows that p^r divides $\mathcal{P}(G, p^k)$ and that it is the maximal power which does so.

Now note that G acts on $\mathcal{P}(G, p^k)$ via:

$$\phi_g : S \mapsto g \cdot S := \{gh | h \in S\} \tag{9.19}$$

where we are denoting an element of $\mathcal{P}(G, p^k)$ by S. Consider the stabilizer subgroup G^S of any $S \in \mathcal{P}(G, p^k)$. Note that if $h \in S$ then every element $g \cdot h \in S$ for $g \in G^S$. (Why? Because $g \cdot S = S$ if $g \in G^S$.) But this means that $G^S \cdot h$ is a subset of S. Since the left-G action is free

$$|G^S| = |G^S \cdot \alpha| \leq |S| = p^k \tag{9.20}$$

We now aim to show that <u>some</u> stabilizer group G^S has order exactly p^k. This will be our subgroup predicted by Sylow's theorem. Suppose, on the contrary that no stabilizer group has order p^k. Then every stabilizer group satisfies $|G^S| < p^k$, and therefore it is divisible at most by p^{k-1}. Now, by the stabilizer-orbit theorem

$$|G| = |G^S| \cdot |\mathcal{O}(S)| \tag{9.21}$$

where $\mathcal{O}(S)$ is the G-orbit through S. Now p^{k+r} divides $|G|$ and if G^S is divisible by at most p^{k-1} then p^s divides $|\mathcal{O}(S)|$ for $s > r$. But now

$$|\mathcal{P}(G,p^k)| = \sum_{\text{distinct orbits}} |\mathcal{O}(S_i)| \tag{9.22}$$

If all the orbits on the RHS were divisible by p^s with $s > r$ then $|\mathcal{O}(S_i)|$ would be divisible by p^s with $s > r$. But this is not true. Therefore, some orbit is divisible by p^r and no higher power. Therefore some $|G^S|$ is divisible by p^k, therefore $|G^S| = p^k$. Since G^S is a stabilizer group it is a subgroup of G. ♠

[92] We are following the nice article on Wikipedia here.

Proof 2: The more conventional proof is similar to that of Cauchy's theorem. We work by induction on $|G|$, and divide the proof into two cases:

Case 1: p divides the order of $Z(G)$.: By Cauchy's theorem $Z(G)$ has an element of order p and hence a subgroup $N \subset Z(G)$ of order p. N is clearly a normal subgroup of G (being a subgroup of the center of G) so G/N is a group. It is clearly of order $p^{k-1}m$. So, by the inductive hypothesis there is a subgroup $\bar{H} \subset G/N$ of order p^{k-1}. Now let $H = \{g \in G | gN \in \bar{H}\}$. It is not hard to show that H is a a subgroup of G containing N and in fact $H/N = \bar{H}$. Therefore $|H| = p^k$, so H is a p-Sylow subgroup of G.

Case 2: p does not divide the order of $Z(G)$.: In this case, by the class equation p must not divide $|C(g)| = |G|/|Z(g)|$ for some nontrivial conjugacy class $C(g)$. But that means that for such an element g we must have that p^k divides $|Z(g)| < |G|$. So $Z(g)$ has a p-Sylow subgroup which can serve as a p-Sylow subgroup of G. ♠

♣Again, there is a nice proof using the orbit-stabilizer theorem. See Wikipedia article. Give this in the section on Orbit-Stabilizer below? ♣

Exercise
If p^k divides $|G|$ with $k > 1$ does it follow that there is an element of order p^k? [93]

Exercise *Groups Whose Order Is A Square Of A Prime Number*
If $|G| = p^2$ where p is a prime then show that
1. G is abelian
2. $G \cong \mathbb{Z}_p \times \mathbb{Z}_p$ or $\mathbb{Z}_{p^2}$.

Exercise
Write out the class equation for the groups S_4 and S_5.

Exercise
Find the centralizer $Z(g) \subset S_n$ of $g = (12 \dots n)$ in S_n.

Exercise

[93] *Answer*: NO! $\mathbb{Z}_p^k$ is a counterexample: It has order p^k and every element has order p.

Prove that if $|G| = 15$ then $G = \mathbb{Z}/15\mathbb{Z}$.

Exercise *Groups whose order is a product of two primes*

Suppose that G has order pq where p and q are distinct primes. We assume WLOG that $p < q$. We now also assume that and p does not divide $q - 1$.

a.) Show that G is isomorphic to $\mathbb{Z}_{pq}$.

Warning!! This is hard. [94]

b.) Why is it important to say that p does not divide $q - 1$? [95]

c.) Show that this result implies that if a nonabelian group has odd order then the order must be ≥ 21. (And in fact, there does exist a nonabelian group of order 21.)

---

10. Kernel, Image, And Exact Sequence

Given an arbitrary homomorphism

$$\mu : G \to G' \tag{10.1}$$

there is automatically a "God-given" subgroup of both G and G':

[94] *Answer.* By Cauchy's theorem we know there is an element a of order p and an element b of order q. We can easily reduce to the case the center of G is trivial. In general the subgroup $Z(G)$ must have order pq, p, q, or 1. If $Z(G)$ has order pq then a and b commute and $G \cong \mathbb{Z}_p \times \mathbb{Z}_q \cong \mathbb{Z}_{pq}$. If $|Z(G)|$ has order p or q then $G/Z(G)$ must be cyclic of order q or p, respectively. Hence by an easy exercise above G is cyclic. This leaves us with the hard case where $Z(G) = \{1\}$ is the trivial subgroup. Let us consider the conjugacy classes of the powers of a, $C(a)$, $C(a^2)$,.... Since $Z(a)$ has order at least p and its order must divide pq and it can't be the whole group (since $Z(G) = \{1\}$) it must be that $Z(a) = \{1, a, \ldots, a^{p-1}\}$ and hence $C(a)$ has order q. Indeed, for any element $g \in G$ that is not the identity it must be that $Z(g)$ has order p or q and $C(g)$ has order q or p. Now note that $Z(a) \supset Z(a^2) \supset \cdots$. So, as long as a^x is not one, it must be that $Z(a^x) = Z(a)$ and $C(a^x)$ has order q. Now we claim that the different conjugacy classes $C(a)$, $C(a^2)$,..., $C(a^{p-1})$ are all distinct. The statement that these are distinct can be reduced to the statement that it is not possible to have $bab^{-1} = a^x$ for any x, so now we verify this latter statement. If it were the case that $bab^{-1} = x$ then since the general element of the conjugacy class is $b^j ab^{-j}$ the conjugacy class would have to be $\{a, a^x, a^{2x}, \ldots, a^{(q-1)x}\}$. But that set must be the set $C(a) = \{a, a^2, \cdots, a^{p-1}\}$ of p elements. Since $q > p$ it must be that $b^{j_1} ab^{-j_1} = b^{j_2} ab^{-j_2}$ where $1 \leq j_1, j_2 \leq (q-1)$ and $j_1 \neq j_2$. So we have to have $b^j ab^{-j} = a$ for some $1 \leq j \leq (q-1)$. But then $b^j \neq 1$. But then such an element b^j would be in $Z(a)$. This is impossible. So we can never have $bab^{-1} = a^x$ and hence $C(a)$, $C(a^2)$,..., $C(a^{p-1})$ are all distinct. Now the class equation says that

$$pq = 1 + (p-1)q + X$$

where X accounts for all the other conjugacy classes. As we have remarked these must have order p or q and hence $X = rp + sq$ for nonnegative integers r, s. But now

$$q - 1 = rp + sq$$

But this is impossible: If $s \geq 1$ the RHS is too large. So $s = 0$ but then p would have to divide $q - 1$.

[95] *Answer*: Consider $p = 2$ and $q = 3$ and note that S_3 is not isomorphic to $\mathbb{Z}_6$.

Definition 10.1:

a.) The *kernel* of μ is

$$K = \ker\mu := \{g \in G | \mu(g) = 1_{G'}\} \tag{10.2}$$

b.) The *image* of μ is

$$\mathrm{im}\mu := \mu(G) \subset G' \tag{10.3}$$

Exercise *Due Diligence*

b.) Check that $\mu(G) \subseteq G'$ is indeed a subgroup.

a.) Check that $\ker(\mu) \subset G$ is indeed a subgroup. [96]

In mathematics one often encounters the notation of an *exact sequence*: Suppose we have three groups and two homomorphisms f_1, f_2

$$G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} G_3 \tag{10.4}$$

We say the sequence is *exact at* G_2 if $\mathrm{im}f_1 = \ker f_2$.

This generalizes to sequences of several groups and homomorphisms

$$\cdots G_{i-1} \xrightarrow{f_{i-1}} G_i \xrightarrow{f_i} G_{i+1} \xrightarrow{f_{i+1}} \cdots \tag{10.5}$$

The sequence can be as long as you like. It is said to be *exact at* G_i if $\mathrm{im}(f_{i-1}) = \ker(f_i)$.

A *short exact sequence* is a sequence of the form

$$1 \longrightarrow G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} G_3 \longrightarrow 1 \tag{10.6}$$

which is exact at G_1, G_2, and G_3. Here 1 refers to the trivial group with one element. There is then a unique homomorphism $1 \to G_1$ and $G_3 \to 1$ so we don't need to specify it. Thus, the meaning of saying that (10.6) is a short exact sequence is that

1. Exactness at G_1: The kernel of f_1 is the image of the inclusion $\{1\} \hookrightarrow G_1$, and hence is the trivial group. Therefore f_1 an injection of G_1 into G_2.

2. Exactness at G_2: $\mathrm{im}f_1 = \ker f_2$.

3. Exactness at G_3: $G_3 \to 1$ is the homomorphism which takes every element of G_3 to the identity element in the trivial group. The kernel of this homomorphism is therefore all of G_3. Exactness at G_3 means that this kernel is the image of the homomorphism f_2, and hence f_2 is a surjective homomorphism.

[96] *Answer*: If $k_1, k_2 \in K$ then $\mu(k_1k_2) = \mu(k_1)\mu(k_2) = 1_{G'}$. So K is closed under multiplication. The group properties of K now follow.

In particular, note that if $\mu : G \to G'$ is any group homomorphism then we automatically have a short exact sequence:

$$1 \to K \to G \xrightarrow{\mu} \mathrm{im}(\mu) \to 1 \tag{10.7}$$

where K is the kernel of μ.

When we have a short exact sequence of groups there is an important relation between them, as we now explain.

Theorem 10.1: Let $K \subseteq G$ be the kernel of a homomorphism (10.1). Then K is a normal subgroup of G.

Proof: $\mu(gkg^{-1}) = \mu(g)\mu(k)\mu(g^{-1}) = \mu(g)1_{G'}\mu(g)^{-1} = 1_{G'} \Rightarrow K$ is normal. ♠

Exercise *Is the image of a homomorphism a normal subgroup?*
If $\mu : G \to G'$ is a group homomorphism is $\mu(G)$ a normal subgroup of G'?
Answer the question with a proof or a counterexample. [97]

It follows by Theorem 7.2.1, that G/K has a group structure. Note that $\mu(G)$ is also naturally a group.

These two groups are closely related because

$$\mu(g) = \mu(g') \qquad \leftrightarrow \qquad gK = g'K \tag{10.8}$$

Thus we have

Theorem 10.2:

$$\boxed{\mu(G) \cong G/K} \tag{10.9}$$

Proof: We associate the coset gK to the element $\mu(g)$ in G'.

$$\psi : gK \mapsto \mu(g) \tag{10.10}$$

Claim: ψ is an isomorphism. You have to show three things:

1. ψ is a well defined map:

$$gK = g'K \Rightarrow \exists k \in K, g' = gk \Rightarrow \mu(g') = \mu(gk) = \mu(g)\mu(k) = \mu(g) \tag{10.11}$$

2. ψ is in fact a homomorphism of groups

$$\psi(g_1K \cdot g_2K) = \psi(g_1K) \cdot \psi(g_2K) \tag{10.12}$$

where on the LHS we have the product in the group G/K and on the RHS we have the product in G'. We leave this as an exercise for the reader.

[97] *Answer*: Definitely not! Any subgroup $H \subset G$ is the image of the inclusion homomorphism. In general, subgroups are not normal subgroups.

3. ψ is one-one, i.e. ψ is onto and invertible. The surjectivity should be clear. To prove injectivity note that:

$$\mu(g') = \mu(g) \Rightarrow \exists k \in K, g' = gk \Rightarrow g'K = gK \qquad \spadesuit \tag{10.13}$$

Remarks:

1. If we have a short exact sequence

$$1 \to N \to G \to Q \to 1 \tag{10.14}$$

then it automatically follows that N is isomorphic to a normal subgroup of G (it is the kernel of a homomorphism $G \to Q$) and moreover Q is isomorphic to G/N. For this reason we call Q the *quotient group*. A frequently used terminology is that *"G is an extension of Q by N."* Some authors [98] will use the terminology that "G is an extension of N by Q." So it is best simply to speak of a group extension with kernel N and quotient Q.

2. **VERY IMPORTANT**: In quantum mechanics physical states are actually represented by "rays" in Hilbert space, or better, by one-dimensional subspaces of Hilbert space, or, even better, by orthogonal projection operators of rank one. (This is for "pure states." More generally, "states" are described mathematically by density matrices.) When comparing symmetries of quantum systems with their classical counterparts, group extensions play an important role so we will discuss them rather thoroughly in §15 below. For the moment we quote three important examples:

Example 1: Consider the group of fourth roots of unity, $Res(4)$ and the homomorphism $\pi : Res(4) \to Res(2)$ given by $\pi(g) = g^2$. The kernel is $\{\pm 1\} = Res(2)$ and so we have:

$$1 \to \mathbb{Z}_2 \to \mathbb{Z}_4 \to \mathbb{Z}_2 \to 1 \tag{10.15}$$

As an exercise the reader should also describe this extension thinking of $\mathbb{Z}_4$ additively as $\mathbb{Z}/4\mathbb{Z}$, and generalize it to

$$1 \to \mathbb{Z}_p \to \mathbb{Z}_{p^2} \to \mathbb{Z}_p \to 1 \tag{10.16}$$

where p is prime.

Example 2: Consider the homomorphism

$$r_N : \mathbb{Z} \to \mathbb{Z}/N\mathbb{Z} \tag{10.17}$$

given by reduction modulo N. (Or, if you prefer to think multiplicatively, $r_N(n) = \omega^n$ where ω is a primitive N^{th} root of 1.) The kernel is $K = N\mathbb{Z} \subset \mathbb{Z}$. As a group this kernel is isomorphic to $\mathbb{Z}$ and so we have

$$0 \to \mathbb{Z} \stackrel{\iota_N}{\to} \mathbb{Z} \stackrel{r_N}{\to} \mathbb{Z}/N\mathbb{Z} \to 0 \tag{10.18}$$

[98] notably, S. MacLane, one of the inventors of group cohomology,

where $\iota_N(x) = Nx$.

Example 3: *Finite Heisenberg Groups*: Let P, Q be $N \times N$ "clock" and "shift" matrices. To define these introduce an Nth root of unity, say $\omega = \exp[2\pi i/N]$. Then

$$P_{i,j} = \delta_{i=j+1 \bmod N} \tag{10.19}$$

$$Q_{i,j} = \delta_{i,j}\omega^j \tag{10.20}$$

Note that $P^N = Q^N = 1$ and no smaller power is equal to 1. Further note that [99]

$$QP = \omega PQ \tag{10.21}$$

For $N = 4$ the matrices look like

$$P = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad Q = \begin{pmatrix} \omega & 0 & 0 & 0 \\ 0 & \omega^2 & 0 & 0 \\ 0 & 0 & \omega^3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{10.22}$$

with $\omega = e^{2\pi i/4}$. The group of matrices generated by P, Q and $\omega \mathbf{1}_{N\times N}$ is a finite subgroup of $GL(N, \mathbb{C})$ isomorphic to a *finite Heisenberg group*, denoted Heis_N. It is an extension

$$1 \to \mathbb{Z}_N \to \text{Heis}_N \xrightarrow{\pi} \mathbb{Z}_N \times \mathbb{Z}_N \to 1 \tag{10.23}$$

and has many pretty applications to physics and we will return to this group several times below. See, for example section 11.11 below for a physical interpretation.

Exercise
Give a formula for π in the exact sequence (10.23).

Exercise A_n
Use Theorem 7.1 to show that A_n is a normal subgroup of S_n.

Exercise *Induced maps on quotient groups*

We will use the following result in §12.3: Suppose $\mu : G_1 \to G_2$ is a homomorphism and $H_2 \subset G_2$ is a subgroup.

[99] The fastest way to check that - and thereby to check that you have your conventions under control - is to compute QPQ^{-1} because $(Q^{-1}PQ)_{ij} = Q_{ii}P_{ij}(Q_{jj})^{-1} = \omega P_{ij}$.

a.) Show that $\mu^{-1}(H_2) \subset G_1$ is a subgroup.

b.) If $H_1 \subset \mu^{-1}(H_2)$ is a subgroup show that there is an induced map $\bar{\mu} : G_1/H_1 \to G_2/H_2$.

c.) Show that if H_1 and H_2 are normal subgroups then $\bar{\mu}$ is a homomorphism.

d.) In this case there is an exact sequence

$$1 \to \mu^{-1}(H_2)/H_1 \to G_1/H_1 \to G_2/H_2 \tag{10.24}$$

Exercise

Let A, B be abelian groups and $A_1 \subset A$ and $B_1 \subset B$ subgroups, and suppose $\phi : A \to B$ is a homomorphism such that ϕ takes A_1 into B_1.

a.) Show that ϕ induces a homomorphism

$$\bar{\phi} : A/A_1 \to B/B_1 \tag{10.25}$$

b.) Show that if $\phi : A_1 \to B_1$ is *surjective* then

$$\ker\{\bar{\phi} : A/A_1 \to B/B_1\} \cong \frac{\ker\{\phi : A \to B\}}{\ker\{\phi : A_1 \to B_1\}} \tag{10.26}$$

Exercise

Let n be a natural number and let

$$\psi : \mathbb{Z}/n\mathbb{Z} \to (\mathbb{Z}/n\mathbb{Z})^d \tag{10.27}$$

be given by the diagonal map $\psi(\omega) = (\omega, \cdots, \omega)$.

Find a set of generators and relations for $G/\psi(H)$.

Exercise

Let $G = \mathbb{Z} \times \mathbb{Z}_4$. Let K be the subgroup generated by $(2, \omega^2)$ where we are writing $\mathbb{Z}_4$ as the multiplicative group of 4th roots of 1. Note $(2, \omega^2)$ is of infinite order so that $K \cong \mathbb{Z}$. Show that $G/K \cong \mathbb{Z}_8$.

Exercise *The Finite Heisenberg Groups*

a.) Using the matrices of (10.19) and (10.20) show that the word

$$P^{n_1}Q^{m_1}P^{n_2}Q^{m_2}\cdots P^{n_k}Q^{m_k} \tag{10.28}$$

where $n_i, m_i \in \mathbb{Z}$ can be written as $\xi P^x Q^y$ where $x, y \in \mathbb{Z}$ and ξ is an Nth root of unity. Express x, y, ξ in terms of n_i, m_i.

b.) Show that $P^N = Q^N = 1$.

c.) Find a presentation of Heis_N in terms of generators and relations.

d.) What is the order of Heis_N ?

Exercise

Let $\mathcal{B}_n$ be a braid group. Compute the kernel of the natural homomorphism $\phi : \mathcal{B}_n \to S_n$ and show that there is an exact sequence

$$1 \to \mathbb{Z}^{n-1} \to \mathcal{B}_n \to S_n \to 1 \tag{10.29}$$

Exercise *Centrally symmetric shuffles*

Let us consider again the permutation group of the set $\{0, 1, \ldots, 2n-1\}$. Recall we let $W(B_n)$ denote the subgroup of S_{2n} of centrally symmetric permutations which permutes the pairs $x + \bar{x} = 2n - 1$ amongst themselves.

Show that there is an exact sequence

$$1 \to \mathbb{Z}_2^n \to W(B_n) \to S_n \to 1 \tag{10.30}$$

and therefore $|W(B_n)| = 2^n n!$.

Exercise *Weyl Group Of $SU(N)$*

Every element of $SU(N)$ can be conjugated into the set T Of diagonal matrices.

a.) Show that the normalizer $N(T)$ of T within $SU(N)$ is larger than T by considering permutation matrices $\mathrm{i}(e_{ij} + e_{ji})$.

b.) Show that these act by permuting the ii and jj diagonal elements .

c.) Show that there is a homomorphism $N(T) \to S_N$.

d.) In fact, there is an exact sequence

$$1 \to T \to N(T) \to S_N \to 1 \tag{10.31}$$

10.1 The Relation Of $SU(2)$ And $SO(3)$

There is a standard homomorphism

$$\pi : SU(2) \to SO(3) \quad . \tag{10.32}$$

To define it we note that for any $u \in SU(2)$ there is a unique $R \in SO(3)$ such that, for all $\vec{x} \in \mathbb{R}^3$ we have:

$$u\vec{x} \cdot \vec{\sigma} u^{-1} = (R\vec{x}) \cdot \vec{\sigma} \tag{10.33}$$

where $R \in SO(3)$.

To prove (10.33) we begin by noting that, since $u^{-1} = u^\dagger$ and $\vec{x}$ is real, the 2×2 matrix $u\vec{x} \cdot \vec{\sigma} u^{-1}$ is hermitian, and traceless, and hence has to be of the form $\vec{y} \cdot \vec{\sigma}$, where $\vec{y} \in \mathbb{R}^3$. Moreover, $\vec{y}$ depends linearly on $\vec{x}$. So the transformation $\vec{x} \mapsto \vec{y}$ defined by $u\vec{x} \cdot \vec{\sigma} u^{-1} = \vec{y} \cdot \vec{\sigma}$ is a linear transformation of $\mathbb{R}^3$. In fact it is a norm-preserving transformation. One way to prove this is to note that (see exercise below)

$$(\vec{x} \cdot \vec{\sigma})^2 = \vec{x}^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{10.34}$$

or, alternatively (see exercise below)

$$\det \vec{x} \cdot \vec{\sigma} = -\vec{x}^2 \tag{10.35}$$

From either formula we conclude that (see exercise below) $\vec{x}^2 = \vec{y}^2$. We therefore conclude that $\vec{y} = R\vec{x}$ with $R \in O(3)$.

We define $\pi(u) = R$ by using this equation. To be totally explicit

$$u\sigma^i u^{-1} = R_{ji}\sigma^j \quad . \tag{10.36}$$

It should be clear from the definition that $\pi(u_1 u_2) = \pi(u_1)\pi(u_2)$, that is that π is a homomorphism of groups. Now to show that actually $R \in SO(3) \subset O(3)$ note that

$$\begin{aligned} 2\mathrm{i} &= \mathrm{tr}\left(\sigma^1\sigma^2\sigma^3\right) \\ &= \mathrm{tr}\left(u\sigma^1 u^{-1} u\sigma^2 u^{-1} u\sigma^3 u^{-1}\right) \\ &= R_{j_1,1}R_{j_2,2}R_{j_3,3}\mathrm{tr}\left(\sigma^{j_1}\sigma^{j_2}\sigma^{j_3}\right) \\ &= 2\mathrm{i}\epsilon^{j_1 j_2 j_3}R_{j_1,1}R_{j_2,2}R_{j_3,3} \\ &= 2\mathrm{i}\det R \end{aligned} \tag{10.37}$$

and hence $\det R = 1$. Alternatively, if you know about Lie groups, you can use the fact that π is continuous, and $SU(2)$ is a connected manifold.

We will now prove that:

1. $\ker(\pi) = \{\pm \mathbf{1}_{2\times 2}\} = Z(SU(2))$.

2. Every proper rotation R comes from some $u \in SU(2)$:

Thus we have the extremely important extension:

$$1 \to \mathbb{Z}_2 \quad \stackrel{\iota}{\to} \quad SU(2) \quad \stackrel{\pi}{\to} \quad SO(3) \to 1 \tag{10.38}$$

Thus, $SU(2)$ is a two-fold cover of $SO(3)$ and in fact

$$SO(3,\mathbb{R}) \cong SU(2)/\mathbb{Z}_2 \tag{10.39}$$

where the $\mathbb{Z}_2$ we quotient by is the center $\{\pm \mathbf{1}_{2\times 2}\}$. This is arguably the most important exact sequence in physics.

To prove the above two claims we will need to get to know $SU(2)$ a bit better.

First we claim that, as a manifold, $SU(2)$ can be identified with the unit three-dimensional sphere. One way to see this is to consider the unit sphere in $\mathbb{R}^4$ as the space of unit vectors in a two-dimensional complex Hilbert space (the space of states of "one Qbit"):

$$S^3 \cong \{\vec{z} | \vec{z}^\dagger \vec{z} = 1\} \subset \mathbb{C}^2 \tag{10.40}$$

This is easily seen by writing

$$\vec{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{10.41}$$

and decomposing z_1, z_2 into their real and imaginary parts. Next, we note that $SU(2)$ has a transitive action on the unit sphere:

$$\phi_u : \vec{z} \mapsto u\vec{z} \tag{10.42}$$

The action is transitive because, given any unit vector we can find another orthogonal unit vector. But any two ON bases are related by some unitary transformation. By changing the phase of the second vector we can arrange that they are related by a special unitary transformation.

Therefore, we should invoke the stabilizer-orbit theorem and compute the stabilizer of, say

$$\vec{z}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ . \tag{10.43}$$

These are the upper triangular $SU(2)$ matrices with one on the diagonal: The stabilizer is trivial. So

$$SU(2) \cong S^3 \tag{10.44}$$

In particular, a general $SU(2)$ element must have the form

$$u = \begin{pmatrix} z_1 & * \\ z_2 & * \end{pmatrix} \tag{10.45}$$

with $|z_1|^2 + |z_2|^2 = 1$. Now imposing the condition

$$u^{-1} = u^\dagger \tag{10.46}$$

we solve for the other two matrix elements and conclude that every $SU(2)$ element is of the form

$$u = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \tag{10.47}$$

where

$$|\alpha|^2 + |\beta|^2 = 1 \tag{10.48}$$

This makes the identification of the group as a manifold quite clear.

There are many ways to parametrize S^3. One is to introduce a polar angle and stratify S^3 by two-dimensional spheres. Viewed this way, we can write the general $SU(2)$ element as

$$u = \cos\chi + \mathrm{i}\sin\chi \vec{n}\cdot\vec{\sigma} \tag{10.49}$$

where $0 \leq \chi \leq \pi$ and $\vec{n} \in S^2$. From this form it is easy to check that u only with all the σ^i if $\sin\chi = 0$ so $\cos\chi = \pm 1$. From this we conclude

$$\ker(\pi) = \{\pm \mathbf{1}_{2\times 2}\} = Z(SU(2)) \tag{10.50}$$

Now let us study the restriction of the homomorphism π to some special $U(1)$ subgroups of $SU(2)$. First consider the subgroup of diagonal matrices D of the form:

$$\begin{pmatrix} \xi & 0 \\ 0 & \xi^{-1} \end{pmatrix} \tag{10.51}$$

with $|\xi| = 1$. These act by

$$\begin{pmatrix} \xi & 0 \\ 0 & \xi^{-1} \end{pmatrix} \begin{pmatrix} x_3 & x_1 - \mathrm{i}x_2 \\ x_1 + \mathrm{i}x_2 & -x_3 \end{pmatrix} \begin{pmatrix} \xi^{-1} & 0 \\ 0 & \xi \end{pmatrix} = \begin{pmatrix} x_3 & \xi^2(x_1 - \mathrm{i}x_2) \\ \xi^{-2}(x_1 + \mathrm{i}x_2) & -x_3 \end{pmatrix} \tag{10.52}$$

If we write

$$\xi = e^{-\mathrm{i}\phi/2} \tag{10.53}$$

for some angle ϕ then π maps the diagonal matrix to $R \in SO(3)$ that is a rotation around the x^3 axis. It is a counterclockwise rotation by ϕ in the $x^1 - x^2$ plane with the orientation $dx^1 \wedge dx^2$.

Another obvious subgroup of $SU(2)$ is $SO(2)$, the real unitary matrices. We parametrize the group by

$$R(\theta/2) = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix} e^{-\mathrm{i}\frac{\theta}{2}\sigma^2} \tag{10.54}$$

Of course, the eigenvalues are $e^{\pm \mathrm{i}\theta/2}$ and indeed

$$SR(\theta/2)S^{-1} = e^{-\mathrm{i}\frac{\theta}{2}\sigma^3} \tag{10.55}$$

where

$$S = \frac{1}{\sqrt{2}}(1 - \mathrm{i}\sigma^1) \in SU(2) \tag{10.56}$$

(all you have to check is $S\sigma^2 S^{-1} = \sigma^3$.) We find that

$$\pi(R(\theta/2)) = \begin{pmatrix} \cos(\theta/2) & 0 & \sin(\theta/2) \\ 0 & 1 & 0 \\ -\sin(\theta/2) & 0 & \cos(\theta/2) \end{pmatrix} \tag{10.57}$$

is rotation by θ around the x^2 axis. By the Euler angle parametrization we therefore learn that π is onto. In fact, we can parametrize all $SU(2)$ elements by

$$u = e^{\phi T^3} e^{\theta T^2} e^{\psi T^3} \tag{10.58}$$

where

$$T^i = -\frac{\mathrm{i}}{2}\sigma^i \qquad 1 \le i \le 3 \tag{10.59}$$

The range of Euler angles that covers $SO(3)$ once is $0 \le \theta \le \pi$ with ϕ and ψ identified modulo 2π. Because $SU(2)$ is a double cover we should extend the range of ϕ or ψ by a factor of 2 if we want to cover the group $SU(2)$ once. For example, taking:

$$\begin{aligned} &0 \le \theta \le \pi \\ &\phi \sim \phi + 2\pi \\ &\psi \sim \psi + 4\pi \end{aligned} \tag{10.60}$$

Then for <u>generic</u> $SU(2)$ elements we will have a unique representation

$$u = e^{\phi T^3} e^{\theta T^2} e^{\psi T^3} = \exp[-\frac{\mathrm{i}}{2}\phi\sigma^3]\exp[-\frac{\mathrm{i}}{2}\theta\sigma^2]\exp[-\frac{\mathrm{i}}{2}\psi\sigma^3] = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \alpha \end{pmatrix} \tag{10.61}$$

with

$$\alpha = e^{\mathrm{i}\frac{1}{2}(\phi+\psi)}\cos(\theta/2) \qquad \beta = -e^{\mathrm{i}\frac{1}{2}(\phi-\psi)}\sin(\theta/2) \tag{10.62}$$

The Euler angle coordinates on $SU(2)$ break down at $\theta = 0, \pi$. At $\theta = 0$ the product only depends on $(\phi + \psi)$ even though we have a three-dimensional manifold Similarly at $\theta = \pi$ the product only depends on $(\phi - \psi)$.

Remark: As we will discuss later, a good parametrization near the identity would be

$$u = \exp[\theta^k T^k] \tag{10.63}$$

where we are exponentiating the general element of the Lie algebra $\mathfrak{su}(2)$

Exercise *Simple Identities For* $\vec{x}\cdot\vec{\sigma}$
a.) Prove (10.34)
b.) Prove (10.35)
c.) Show that both these formulae imply that $\vec{x}^2 = \vec{y}^2$.

Exercise *Polar Angle Decomposition Of* $SU(2)$

a.) Prove that every element of $SU(2)$ can be written in the form of (10.49).

b.) Express α, β in terms of χ and $\hat{n}$.

c.) The coordinates $\chi, \hat{n}$ cover a product of an interval and a two-dimensional sphere. Prove that $[0, \pi] \times S^2$ is not topologically the same as $SU(2)$. Where does the $\chi, \hat{n}$ coordinate system go bad?

Exercise *A Basis For The Lie Algebra* $\mathfrak{su}(2)$

a.) Show that every traceless anti-Hermitian 2×2 matrix is a real linear combination of the three matrices T^i defined in (10.59).

b.) Show that the matrix commutators satisfy

$$[T^i, T^j] = \epsilon^{ijk} T^k \tag{10.64}$$

with the convention $\epsilon^{123} = +1$.

Exercise

Show that in the Euler angle parametrization the shift

$$\psi \to \psi + 2\pi \tag{10.65}$$

takes $u \to -u$.

---

11. Some Representation Theory

One of the main motivations from physics for studying representation theory stems from *Wigner's theorem* discussed below. The basic upshot of Wigner's theorem is that, in quantum mechanics, if G is a group of symmetries of a physical system then the Hilbert space of the theory will be a representation space of G and in fact will define a unitary representation: [100] For every symmetry operation $g \in G$ there is a unitary operator $U(g)$ acting on the Hilbert space $\mathcal{H}$ so that

$$U(g_1)U(g_2) = U(g_1 g_2) \tag{11.1}$$

The use of representation can be very powerful and have far-reaching consequences. A few examples:

[100] We will need to amend this in two ways to be completely accurate: First, classical symmetries in general are represented projectively on a quantum Hilbert space. Second, one must allow symmetries to be represented by both unitary and anti-unitary operators in general.

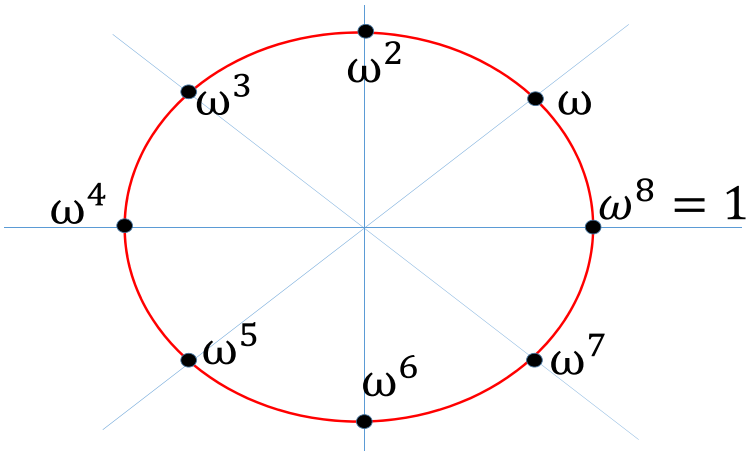


Figure 28: Roots of unity on the unit circle in the complex plane. Here $\omega = e^{2\pi i/8}$ is a primitive eighth root of 1.

1. The use of representation theory greatly aids in the diagonalization of physical observables, such as Hamiltonians.

2. Quantum states can be classified according to their symmetry types. This has important applications to selection rules governing what kind of transition amplitudes can be nonzero.

3. Conservation laws are associated with symmetry operations.

4. The very formulation of Lagrangians and actions makes heavy use of representation theory. For example, in relativistically invariant field theory the fields form a representation of the Poincaré group (induced from the action on spacetime) and one wishes to make a Lorentz invariant density when forming a Lagrangian.

11.1 Some Basic Definitions

Let V be a vector space over a field κ and recall that $GL(V)$ denotes the group of all invertible linear transformations $V \to V$. It is also denoted as $\mathrm{Aut}(V)$ since it is the group of linear automorphisms of V with itself.

Definition 11.1.1 A *representation* of G is a group homomorphism from G to a group of the form $GL(V) = \mathrm{Aut}(V)$ where V is a vector space over a field κ:

$$\begin{aligned} T : g &\mapsto T(g) \\ G &\to GL(V) \end{aligned} \tag{11.2}$$

V is sometimes called the *representation space* or the *carrier space*. We will abbreviate the action of $T(g)$ on a vector $v \in V$ from $T(g)(v)$ to $T(g)v$ for readability.

Put differently, in terms of group actions, a representation of G is a G action on a vector space that respects the linear structure:

$$g \cdot (\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 g \cdot v_1 + \alpha_2 g \cdot v_2 \tag{11.3}$$

where $v_1, v_2 \in V$ are any vectors and $\alpha_1, \alpha_2 \in \kappa$ are any scalars.

If the vector space has an ordered basis then we get a matrix representation. For example, if V is finite-dimensional then we can choose an ordered basis $\{v_1, \ldots, v_n\}$ and the corresponding matrix representation $g \mapsto M(g) \in GL(n, \kappa)$ is defined by:

$$T(g)v_i = \sum_j M(g)_{ji} v_j \tag{11.4}$$

One easily checks that $T(g_1) \circ T(g_2) = T(g_1 g_2)$ implies $M(g_1)M(g_2) = M(g_1 g_2)$.

1. As a simple example, take $V = \kappa$, the standard one-dimensional vector space over κ and let $T(g) = 1_V$ for all $g \in G$. This representation is known as the trivial representation.

2. We will often abbreviate "representation" to "rep." Moreover we sometimes refer to "the rep (V, T)" or simply by V, when we wish to stress the representation space. Or we might refer to "the rep T," when the rep of the data is understood.

3. The "dimension of the representation" is by definition the dimension $\dim V$ of the vector space V. This number can be finite or infinite. Sometimes representations are simply denoted by the dimension of the carrier space. This can be dangerous. For example, there are $p(n)$ inequivalent n-dimensional representations of $SU(2)$, although, as we will see, there is a unique irreducible n-dimensional representation of $SU(2)$ - up to isomorphism. See below.

4. The notion of representation can be generalized by replacing $GL(V)$ by $GL(R)$ where R is a ring. Then one often speaks of a *module for* G.

♣Should it be (V, T) or (T, V)? Depends on which is logically prior. Some would say V is defined by the codomain of T. Others would say you first choose a V and then define a T. Need to be consistent... ♣

Definition 11.1.2. Let (V_1, T_1) and (V_2, T_2) be two representations of a group G. An *intertwiner* between these representations is a linear transformation $A : V_1 \to V_2$ such that, for all $g \in G$ the diagram

$$\begin{array}{ccc} V_1 & \xrightarrow{A} & V_2 \\ {\scriptstyle T_1(g)}\downarrow & & \downarrow{\scriptstyle T_2(g)} \\ V_1 & \xrightarrow{A} & V_2 \end{array} \tag{11.5}$$

commutes. Equivalently,

$$T_2(g)A = AT_1(g) \tag{11.6}$$

for all $g \in G$. Put differently: A is a morphism of G actions, and put yet another way: A is an equivariant linear map of G spaces. We denote the vector space of all intertwiners by $\mathrm{Hom}_G(V_1, V_2)$.

Note that if an intertwiner is invertible then T^{-1} is also an intertwiner. So we have:

Definition 11.1.3. Two representations T_1 and T_2 are *equivalent* $T_1 \cong T_2$ if there is an intertwiner $A : V_1 \to V_2$ which is an isomorphism. That is,

$$T_2(g) = AT_1(g)A^{-1} \tag{11.7}$$

for all $g \in G$.

Examples:

1. The general linear group $GL(n, \kappa)$ with $\kappa = \mathbb{R}, \mathbb{C}$ always has a family of one-dimensional real representations, labeled by $\mu \in \mathbb{C}$ given by

$$T(g) := |\det g|^\mu \tag{11.8}$$

   This is a representation because:

$$T(g_1 g_2) = |\det g_1 g_2|^\mu = |\det g_1|^\mu |\det g_2|^\mu = T(g_1)T(g_2) \tag{11.9}$$

   Note that for different μ these are inequivalent representations.

2. We saw that the connected component of the Lorentz group in $1+1$ dimensions is isomorphic to $\mathbb{R}$, $SO_0(1,1) \cong \mathbb{R}$ with boosts $B(\theta_1)B(\theta_2) = B(\theta_1 + \theta_2)$ with $\theta_i \in IR$. The *"spin-s" representation* is

$$\rho_s(B(\theta)) = e^{s\theta} \tag{11.10}$$

   For different values of s these one-dimensional representations are inequivalent.

Familiar notions of linear algebra generalize to representations:

1. *The direct sum $\oplus$ of representations.* The direct sum of (T_1, V_1) and (T_2, V_2) is the rep $(T_1 \oplus T_2, V_1 \oplus V_2)$ where the representation space is $V_1 \oplus V_2$ *and* the operators are:

$$\big((T_1 \oplus T_2)(g)\big)\big(v_1 \oplus v_2\big) := (T_1(g))(v_1) \oplus (T_2(g))(v_2) \tag{11.11}$$

   If $\{v_1, \ldots, v_n\}$ is an ordered basis for V_1 and $\{w_1, \ldots, w_m\}$ is an ordered basis for V_2 then $\{v_1, \ldots, v_n, w_1, \ldots, w_m\}$ is an ordered basis for $V_1 \oplus V_2$ and relative to this ordered basis the matrix representation is block diagonal:

$$M_{T_1 \oplus T_2}(g) = \begin{pmatrix} M_{T_1}(g) & 0 \\ 0 & M_{T_2}(g) \end{pmatrix} \tag{11.12}$$

2. Similarly, for the tensor product, the carrier space is $V_1 \otimes V_2$. See Chapter 2 for a proper definition of the tensor product. For finite-dimensional vector spaces we can say that if $\{v_1, \ldots, v_n\}$ is a basis for V_1 and $\{w_1, \ldots, w_m\}$ is a basis for V_2 then the set of vectors of the form $v_i \otimes w_a$ form a basis for $V_1 \otimes V_2$ and we impose the rules

$$(\alpha_1 v_1 + \alpha_2 v_2) \otimes w = \alpha_1 v_1 \otimes w + \alpha_2 v_2 \otimes w \tag{11.13}$$

$$v \otimes (\alpha_1 w_1 + \alpha_2 w_2) = \alpha_1 v \otimes w_1 + \alpha_2 v \otimes w_2 \tag{11.14}$$

The tensor product can be defined without reference to a basis (See Chapter 2) and so can the tensor product of representations. We set:

$$((T_1 \otimes T_2)(g)) \left(v \otimes w\right) := (T_1(g)v) \otimes (T_2(g)w) \tag{11.15}$$

for all $v_{\in} V_1$ and $w \in V_2$ and then extend by κ-linearity.

If $\{v_1, \ldots, v_n\}$ is an ordered basis for V_1 and $\{w_1, \ldots, w_m\}$ is an ordered basis for V_2 then the matrix elements of $(T_1 \otimes T_2)(g)$ will be of the form

$$(M_1 \otimes M_2)(g)_{ia,jb} = (M_1(g))_{ij}(M_2(g))_{ab} \tag{11.16}$$

Note that while the set $\{v_i \otimes w_a\}$ forms a basis for $V_1 \otimes V_2$ it does not in any natural way define an ordered basis: One needs to make a further choice of how to order this basis and there are several choices. Once one has done this, there will be an ordering on pairs (i, a) and (11.16) will define the matrix relative to this ordered basis.

3. Given a representation of V we get a dual representation on the dual space $V^\vee := \mathrm{Hom}(V, \kappa)$ of linear maps $V \to \kappa$. We do this by demanding that the natural pairing between V and $V^\vee$ is preserved by the group transformation:

$$\langle T^\vee(g)\ell, T(g)v\rangle = \langle \ell, v\rangle, \tag{11.17}$$

where $\ell \in V^\vee, v \in V$. If V is finite dimensional and $\{v_1, \ldots, v_n\}$ is an ordered basis then there is a natural ordered basis $\{v_1^\vee, \ldots, v_n^\vee\}$ for $V^\vee$ and relative to these bases we have

$$(M_{T^\vee}(g)) = M_T(g)^{tr,-1} \tag{11.18}$$

4. As a corollary of the previous remark note that the vector space of linear transformations from V to W, denoted $Hom(V, W)$, is canonically isomorphic to $V^\vee \otimes W$ and hence naturally becomes a representation of G. In concrete terms, if $\phi \in Hom(V, W)$ is a linear transformation $\phi : V \to W$ then the G-action on ϕ is determined by our general remarks about induced actions on function spaces:

$$(\tilde{T}(g) \cdot \phi)(v) := T_W(g) \cdot \phi \left(T_V(g^{-1})v\right) \tag{11.19}$$

5. If V is a complex vector space then the complex conjugate representation sends $g \to \bar{T}(g) \in GL(\bar{V})$. A *real representation* is one where $(\bar{T}, \bar{V})$ is equivalent to (T, V). If $\{v_i\}$ is an ordered basis for V then there is a canonical ordered basis $\{\bar{v}_i\}$ for $\bar{V}$ and the matrices are related by

$$M_{\bar{T}}(g) = (M_T(g))^* \tag{11.20}$$

Exercise *New Matrix Representations From Old Ones*

Given a matrix representation of a group $g \to M(g)$ show that

a.) $g \to (M(g))^{tr,-1}$ is also a representation.

b.) Check the claim (11.19) above.

c.) If M is a matrix representation in $GL(n,\mathbb{C})$ then $g \mapsto M(g)^*$ is also a representation.

d.) If T is a real representation, then there exists an $S \in GL(n,\mathbb{C})$ such that for all $g \in G$:

$$M^*(g) = SM(g)S^{-1} \tag{11.21}$$

Warning: The matrix elements $M(g)_{ij}$ of a real representation might not be real numbers:

e.) Show that the defining two-dimensional representation of $SU(2)$ acting on $\mathbb{C}^2$ is a real representation, but there is no basis in which the matrix elements are all real.

As we will explain later, real representations can be further distinguished as totally real and quaternionic (a.k.a. pseudoreal).

Exercise *Representation Matrices On* $\mathrm{Hom}(V,W)$

Let $V = \mathbb{C}^n$ and $W = \mathbb{C}^m$ equipped with their standard ordered bases e_i, $i = 1,\ldots,n$ and e_a, $a = 1,\ldots,m$, respectively. Identify $\mathrm{Hom}(V,W) \cong \mathrm{Mat}_{m\times n}(\mathbb{C})$ and consider the basis $e_{a,i}$ for $\mathrm{Mat}_{m\times n}(\mathbb{C})$. Show that if V and W are representation spaces of G then the representation on $\mathrm{Hom}(V,W)$ satisfies:

$$\tilde{T}(g)e_{ai} = M_{ba}(g)(M(g)^{tr,-1})_{ki}e_{bk} \tag{11.22}$$

♣Some duplication in this section with previous material ♣

11.2 Characters

For any finite-dimensional representation $T : G \to \mathrm{Aut}(V)$ of any group G we can define the *character of the representation*, denoted χ_T. It is a function on the group:

$$\chi_T : G \to \kappa \tag{11.23}$$

and it is defined by

$$\chi_T(g) := \mathrm{Tr}_V(T(g)) \tag{11.24}$$

Some useful general remarks about characters:

1. The character is independent of any choice of basis for V.

2. Equivalent representations define precisely the same character function.

3. $\chi_T(h^{-1}gh) = \chi_T(g)$ for all $g, h \in G$. In other words, $\chi_T(g)$ only depends on g via its conjugacy class. In general, a function $F : G \to \mathbb{C}$ that only depends on conjugacy class, that is, that satisfies $F(h^{-1}gh) = F(g)$ for all $g, h \in G$ is known as a *class function*. Such functions "descend" to functions on the set of conjugacy classes of G.

4. $\chi_{T_1 \oplus T_2} = \chi_{T_1} + \chi_{T_2}$

5. $\chi_{T_1 \otimes T_2} = \chi_{T_1}\chi_{T_2}$

Exercise *Complex Conjugate Representation*

Show that the fundamental representation of $SU(2)$ is equivalent to its complex conjugate representation.

Show that the fundamental representation of $SU(3)$ is not equivalent to its complex conjugate representation

---

11.3 Unitary Representations

In physics, unitary representations play a distinguished role. The basic reason for this is that symmetries should preserve probability amplitudes.

Definition 11.3.1. Let V be an inner product space. [101] A *unitary representation* is a representation (V, T) such that $\forall g \in G$, $T(g)$ is a unitary operator on V, i.e.,

$$\langle T(g)v, T(g)v\rangle = \langle v, v\rangle \qquad \forall g \in G, v \in V \tag{11.25}$$

1. The canonical representation of S_n on $\mathbb{R}^n$ or $\mathbb{C}^n$ is unitary.

2. The fundamental representations of $SO(n)$ and $U(n)$ are unitary.

Definition 11.3.2. If a rep (V, T) is equivalent to a unitary rep then such a rep is said to be *unitarizable*.

Example. A simple example of non-unitarizable reps are the $\det^\mu$ reps of $GL(n, \kappa)$ with $\kappa = \mathbb{R}, \mathbb{C}$ and the "spin-s" representations ρ_s of the Lorentz group $SO_0(1, 1)$ described in section *****

Exercise

a.) Show that if $T(g)$ is a rep on an inner product space then $T(g^{-1})^\dagger$ is a rep also.

[101] See Chapter 2, section 12

b.) Suppose $T : G \to GL(V)$ is a unitary rep on an inner product space V. Let $\{v_i\}$ be an ordered orthonormal basis for V. Show that the corresponding matrix rep $M(g)_{ij}$ is a unitary matrix rep. That is:

$$M : G \to U(\dim V) \tag{11.26}$$

is a homomorphism.

c.) Show that for a unitary matrix rep the transpose-inverse and complex conjugate representations are equal.

Exercise *Characters Of Unitarizable Representations*
Show that if $V(\rho)$ is unitarizable (in particular, if it is unitary), then

$$\chi_\rho(g^{-1}) = \chi_\rho(g)^* \ . \tag{11.27}$$

11.4 Haar Measure, a.k.a. Invariant Integration

When proving facts about representations a very important tool is the notion of *invariant integration*. In many situations we would like to consider a group to be a measure space and give the average value of functions on the group.

If the group is finite then the obvious way to do that is:

$$f \to \frac{1}{|G|} \sum_{g \in G} f(g) := \langle f \rangle \tag{11.28}$$

We can write, very suggestively:

$$\frac{1}{|G|} \sum_{g} f(g) := \int_G f(g) dg \tag{11.29}$$

The map $f \mapsto \int_G f(g)dg$ on complex-valued functions is clearly linear in the function f. So, it defines an element of the dual space of the space of functions, i.e. a map from the functions to $\mathbb{C}$. There are lots of other elements of the dual space, such as evaluation on a particular group element g_0,

$$\mathrm{ev}_{g_0} : f \mapsto f(g_0) \tag{11.30}$$

As a measure this is i.e. the Dirac measure supported at g_0 (a.k.a. the evaluation map at g_0). Of course we could take various linear combinations of functionals of the form ev_{g_0} to get others. What is special about the measure (11.29) is that it satisfies the *left invariance* property:

$$\int_G f(hg) dg = \int_G f(g) dg \tag{11.31}$$

for all $h \in G$.

The left-action on G induces an action L_h^* on the functions on G. Let $L_h^*(f)$ denote the function on G defined by:

$$(L_h^* f)(g) := f(hg) \tag{11.32}$$

Then $\langle L_h^*(f) \rangle = \langle f \rangle$.

Note that, in this case of a finite group, the measure is also right-invariant:

$$\int_G f(gh)dg = \int_G f(g)dg \tag{11.33}$$

For a finite group left-invariant and right-invariant measures are unique up to overall scale. Indeed, the most general measure will be of the form

$$\sum_{g \in G} \rho(g) f(g) \tag{11.34}$$

for some weight function $\rho(g)$. Left invariance implies that

$$\begin{aligned} \sum_{g \in G} \rho(g) f(g) &= \sum_{g \in G} \rho(g) f(hg) \\ &= \sum_{g \in G} \rho(h^{-1} g) f(g) \end{aligned} \tag{11.35}$$

Now apply the statement of left-invariance to the "Dirac function" at g_0, a.k.a. the characteristic function at g_0: [102]

$$\delta_{g_0}(g) := \begin{cases} 1 & g = g_0 \\ 0 & else \end{cases} \tag{11.36}$$

Then $\rho(g_0) = \rho(h^{-1} g_0)$ for every g_0 and every h. Therefore $\rho(g) = c$ is just some constant function on the group. It is now easy to check that the measure is also right-invariant.

The idea of a left- or right-invariant measure extends to continuous groups. For example for $G = \mathbb{R}$ the general measure is given by

$$\int_{-\infty}^{+\infty} f(x) \rho(x) dx \tag{11.37}$$

for some measure $\rho(x)dx$. By a similar argument to the above we learn that $\rho(x+y)dx = \rho(x)dx$ for all x, y, and hence $\rho(x)$ is a constant.

Remark: Haar's Theorem: The existence of left and right-invariant measures on topological groups is very general. The topological group G must be "locally compact" and "Hausdorff" - two mild topological conditions on G. [103]

[102] Do not confuse this function with the Dirac measure ev_{g_0}.

[103] A topological space X is *locally compact* if, for every $x \in X$ there is a compact neighborhood of x. That is, there is a compact subspace $K \subset X$ so that $x \in U \subset K$ for some open neighborhood U of x. A topological space X is *Hausdorff* if, for all pairs of distinct points $x_1, x_2 \in X$ there exist neighborhoods $x_1 \in \mathcal{U}_1$ and $x_2 \in \mathcal{U}_2$ such that $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$. In other words, open sets separate points.

Then one can define a set $\mathfrak{B}$ of measurable subsets of G and a measure $\mu : \mathfrak{B} \to \mathbb{R}$ is "left-invariant" if $\mu(gS) = \mu(S)$ for all measurable subsets $S \in \mathfrak{B}$. There is a similar definition of "right-invariant." If one imposes a few more technical conditions then Haar's theorem states that such measures are unique up to multiplication by a positive constant.

Given a Haar measure one can define integrals of some class of functions - known as measureable functions. For example, for $G = \mathbb{R}$ clearly we can only integrate functions so that

$$\int_{-\infty}^{+\infty} f(x)dx \tag{11.38}$$

exists.

In general, the left- and right- invariant measures on a topological group need not coincide, even up to scale. However, in the case of compact groups the left- and right-invariant measures, are unique and coincide up to scale. For compact Lie groups the essential observation is that left-invariance shows the volume form must be proportional to $\langle g^{-1}dg \wedge \cdots g^{-1}dg, v\rangle$ where v is a volume form on $\mathfrak{g}^\vee$. [104]

Examples:

1. $G = \mathbb{R}$: The most general Haar measure is of the form:

$$\int_{G=\mathbb{R}} f(g)dg := c\int_{-\infty}^{+\infty} dx f(x) \tag{11.39}$$

   where c is a constant.

2. $G = \mathbb{Z}$: The most general Haar measure is of the form:

$$\int_{G=\mathbb{Z}} f(g)dg := c\sum_{n\in\mathbb{Z}} f(n) \tag{11.40}$$

   where c is a constant.

3. Now let $G = \mathbb{R}^*_{>0}$ be the multiplicative group of positive real numbers. The most general Haar measure is of the form

$$\int_{G=\mathbb{R}^*_{>0}} f(g)dg := c\int_0^{\infty} f(x)\frac{dx}{x} \tag{11.41}$$

   where c is a constant.

[104] See Chapter **** for a discussion of the Maurer-Cartan form $g^{-1}dg$ and the notion of left- and right-invariant differential forms on Lie groups. Another, more sophisticated, but also more elegant proof was explained to me by Dan Freed: Recall that for a finite dimensional real vector space V of dimension n there is a $GL(n,\mathbb{R})$ torsor $\mathcal{B}(V)$ of the bases. Consider the character $|\det|$ on $GL(n,\mathbb{R})$. There is a one-dimensional line of equivariant functions $\mathcal{B}(V) \to \mathbb{R}$ transforming according to this character. There is a natural orientation on this line given by the positive functions and a nonzero function corresponds to a measure on V. It is unique up to scalar multiplication by a positive constant. Now let V be the vector space of left-invariant vector fields on G. A measure on V corresponds to a left-invariant measure on the group. Now right translation acts by a positive scalar, and this gives a homomorphism $G \to \mathbb{R}_{>0}$. But G is compact so the only possible homomorphism is the trivial one. Therefore the measure is also right-invariant.

4. Similarly, consider $G = GL(n, \mathbb{R})$. We can take the matrix elements g_{ij} to be coordinates on the open domain $\{g|\det g \neq 0\} \subset M_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$. The usual Euclidean measure $\prod_{ij} dg_{ij}$ changes under $g \to g_0 g$ by

$$\prod_{ij} dg_{ij} \to |\det g_0|^n \prod_{ij} dg_{ij} \tag{11.42}$$

and so the most general Haar measure is of the form

$$\int_{G=GL(n,\mathbb{R})} f(g)dg := c \int_{\det g \neq 0} f(g)|\det g|^{-n} \prod_{ij} dg_{ij} \tag{11.43}$$

where c is a constant.

5. $G = U(1)$: Up to scale we have the Haar measure:

$$\int_{G=U(1)} f(g)dg := \frac{1}{2\pi \mathrm{i}} \oint_{|z|=1} f(z)\frac{dz}{z} = \int_0^{2\pi} \frac{d\theta}{2\pi} g(\theta) \tag{11.44}$$

where $g(\theta) = f(e^{\mathrm{i}\theta})$ and the range of the last integral is over any interval of length 2π. Here the scale of the measure has been chosen so that the "volume" of the group is 1.

6. For the important case of $G = SU(2)$ we can write it as follows. First, every element of $SU(2)$ can be written as:

$$g = \begin{pmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{pmatrix} \tag{11.45}$$

♣Definition of β here is backwards from what we use when we describe reps using homogeneous polynomials below. ♣

for 2 complex numbers α, β with

$$|\alpha|^2 + |\beta|^2 = 1. \tag{11.46}$$

In this way we identify the group as a manifold as S^3. That manifold has no globally well-defined coordinate chart. The best we can do is define coordinates that cover "most" of the group but will have singularities are some places. (It is always important to be careful about those singularities when using explicit coordinates!) We can always write:

$$\begin{aligned} \alpha &= \zeta_1 \cos\theta/2 \\ \beta &= \zeta_2 \sin\theta/2 \end{aligned} \tag{11.47}$$

where ζ_1, ζ_2 are phases, and the magnitude is parametrized in a 1-1 fashion by taking $0 \leq \theta \leq \pi$. Next it is standard to parametrize the phases by:

$$\begin{aligned} \alpha &= e^{i\frac{1}{2}(\psi+\phi)} \cos\theta/2 \\ \beta &= ie^{i\frac{1}{2}(\psi-\phi)} \sin\theta/2 \end{aligned} \tag{11.48}$$

The virtue of this definition is that we can then write:

$$\begin{aligned} g &= \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix} \begin{pmatrix} \cos\theta/2 & i\sin\theta/2 \\ i\sin\theta/2 & \cos\theta/2 \end{pmatrix} \begin{pmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2} \end{pmatrix} \\ &= e^{i\frac{1}{2}\phi\sigma^3} e^{i\frac{1}{2}\theta\sigma^1} e^{i\frac{1}{2}\psi\sigma^3} \end{aligned} \tag{11.49}$$

and under the standard homomorphism $\pi : SU(2) \to SO(3)$ the angles θ, ϕ, ψ become the Euler angles.

We need to be a little careful about ϕ and ψ since they are not defined at $\theta = 0, \pi$ and we also need to be careful about their ranges. The above expression is invariant under the transformations:

$$\begin{aligned} &(\phi, \psi) \to (\phi + 4\pi, \psi) \\ &(\phi, \psi) \to (\phi, \psi + 4\pi) \\ &(\phi, \psi) \to (\phi + 2\pi, \psi + 2\pi) \end{aligned} \tag{11.50}$$

If we think of this as generating a group of transformations on $\mathbb{R}^2$ we can choose a fundamental domain in various ways, and then taking (ϕ, ψ) to be in that fundamental domain we will cover the group elements exactly once (away from $\theta = 0, \pi$). One standard fundamental domain is:

$$\begin{aligned} 0 \le \phi < 2\pi \\ 0 \le \psi < 4\pi \end{aligned} \tag{11.51}$$

This is good because if we then act on the unit vector e_3, (ϕ, θ) become the standard angular coordinates on S^2.

The normalized Haar measure for $SU(2)$ in these coordinates is [105]

$$[dg] = \frac{1}{16\pi^2} d\psi \wedge d\phi \wedge \sin\theta d\theta \tag{11.52}$$

For much more about this, see Chapter 5 below.

We give formulae for the Haar measures on the classical compact matrix groups in Chapter 6 below.

A consequence of the existence of invariant integration is that every finite-dimensional representation is equivalent to a unitary representation. This follows because you can use a suitable averaging procedure over the group to make the matrices unitary.

Proposition If (T, V) is a rep of a *compact group* G and V is an inner product space then (T, V) is unitarizable.

Proof We make essential use of the Haar measure. If T is not already unitary with respect to the inner product $\langle \cdot, \cdot \rangle_1$ then we can define a new inner product by:

$$\langle v, w \rangle_2 := \int_G \langle T(g)v, T(g)w \rangle_1 \, dg \tag{11.53}$$

[105] We have chosen an orientation so that, with a positive constant, this is $\mathrm{Tr}_{\mathbf{2}}(g^{-1}dg)^3)$.

For a compact group $\langle T(g)v, T(g)w\rangle_1$ will be a nice continuous function, hence bounded, and therefore integrable. If $v \neq 0$ then $T(g)v \neq 0$ hence $\langle T(g)v, T(g)v\rangle_1 > 0$ and therefore $0 < \langle v, v\rangle_2 < \infty$. Here we used the fact that for a compact group the volume is finite. So $\langle \cdot, \cdot\rangle_2$ will be a good inner product. Then using the properties of the Haar measure it is easily checked that

$$\langle T(g)v, T(g)w\rangle_2 = \langle v, w\rangle_2 \tag{11.54}$$

and hence $T(g)$ is unitary w.r.t. the inner product $\langle \cdot, \cdot\rangle_2$ ♠

Remark: We saw that the representations $\det^\mu$ of $GL(n,\mathbb{R})$, $GL(n,\mathbb{C})$ and ρ_s of $SO_0(1,1)$ are not unitarizable. What fails in the above argument is the infinite volume of these noncompact groups.

Exercise *Due Diligence*
Show that

$$\langle T(g)v, T(g)w\rangle_2 = \langle v, w\rangle_2 \tag{11.55}$$

Exercise *Computation For $SU(2)$*
Using the parametrization by Euler angles show that

$$\int_{SU(2)} g_{\alpha\beta} dg = 0 \tag{11.56}$$

$$\int_{SU(2)} g_{\alpha\beta} g_{\gamma\delta} dg = \frac{1}{2}\epsilon_{\alpha\gamma}\epsilon_{\beta\delta} \tag{11.57}$$

We will later interpret these equations as special cases of the orthogonality relations of matrix elements in irreducible representations. (See section on the Peter-Weyl theorem below.)

11.5 The Regular Representation

Let G be a group. Then there is a left action of $G \times G$ on G: $(g_1, g_2) \mapsto L(g_1)R(g_2^{-1})$:

$$(g_1, g_2) \cdot g_0 = g_1 g_0 g_2^{-1} \tag{11.58}$$

and hence an induced action on $\mathrm{Map}(G, Y)$ for any Y. Now let $Y = \mathbb{C}$. Then $\mathrm{Map}(G, \mathbb{C})$ is a representation of $G \times G$ because the induced left-action:

$$\left((g_1, g_2) \cdot \Psi\right)(h) := \Psi(g_1^{-1} h g_2) \tag{11.59}$$

converts the vector space of functions $\Psi : G \to \mathbb{C}$ into a representation space for $G \times G$.

If we equip G with a Haar measure then we can speak of $L^2(G)$, namely the Hilbert space based on the complex-valued functions such that

$$\langle f, f \rangle := \int_G |f(g)|^2 dg < \infty \tag{11.60}$$

Note that the $G \times G$ action preserves the L^2-property thanks to left- and right- invariance, and in fact the $G \times G$ action is unitary.

Definition The representation $L^2(G)$ is known as the *regular representation of* G.

Note that $L^2(G)$ is a representation of $G \times G$ although by restriction to the natural subgroups $G \times \{1\}$ and $\{1\} \times G$ it becomes a representation of G. So sometimes people speak of "the regular representation of G." More precisely, if we restrict to operations of the form:

$$(L(h) \cdot \Psi)(g) := \Psi(h^{-1}g) \tag{11.61}$$

We have the *left regular representation*, while

$$(R(h) \cdot \Psi)(g) := \Psi(gh) \tag{11.62}$$

defines the *right regular representation*. Both actions are left- actions on the functions space, so the terminology is slightly confusing.

Suppose that (T, V) is a representation of G. As we explained above the vector space of linear transformations $\mathrm{End}(V) := \mathrm{Hom}(V, V)$ of V to itself is then also a representation. In fact, it is a representation of $G \times G$ because if $S \in \mathrm{End}(V)$ then we can define a linear left-action of $G \times G$ on $\mathrm{End}(V)$ by:

$$(g_1, g_2) \cdot S := T(g_1) \circ S \circ T(g_2)^{-1} \tag{11.63}$$

Now, we have two representations of $G \times G$. They are related as follows: If V is finite-dimensional we have a map

$$\iota : \mathrm{End}(V) \to L^2(G) \tag{11.64}$$

The map ι takes a linear transformation $S : V \to V$ to the complex-valued function $\Psi_S : G \to \mathbb{C}$ defined by

$$\Psi_S(g) := \mathrm{Tr}_V(ST(g^{-1})) \tag{11.65}$$

That is

$$\iota(S) := \Psi_S \tag{11.66}$$

We claim that ι is a $G \times G$-equivariant map:

$$(h_1, h_2) \cdot \Psi_S = \Psi_{(h_1, h_2) \cdot S} \tag{11.67}$$

You are asked to prove this in an exercise below. Put differently, denoting by $T_{\mathrm{End}(V)}$ the representation of $G\times G$ on $\mathrm{End}(V)$ and $T_{Reg.Rep.}$ the representation of $G\times G$ on $\mathrm{Map}(G,\mathbb{C})$ we get a commutative diagram:

$$\begin{array}{ccc} \mathrm{End}(V) & \xrightarrow{\iota} & \mathrm{Map}(G,\mathbb{C}) \\ \Big\downarrow{\scriptstyle T_{\mathrm{End}(V)}} & & \Big\downarrow{\scriptstyle T_{Reg.Rep.}} \\ \mathrm{End}(V) & \xrightarrow{\iota} & \mathrm{Map}(G,\mathbb{C}) \end{array} \tag{11.68}$$

If we choose an ordered basis $\{v_i\}$ for V then the operators $T(g)$ are represented by matrices:

$$T(g)\cdot v_i = \sum_j M(g)_{ji} v_j \tag{11.69}$$

If we take $S = e_{ij}$ to be the matrix unit in this basis then Ψ_S is the function on G given by the matrix element $M(g^{-1})_{ji} = M^{tr,-1}(g)_{ij}$. So the Ψ_S's are linear combinations of matrix elements of the representation matrices of G. (Replacing V by its dual $V^\vee$ we will get the representation matrices $M(g)_{ij}$.) The advantage of (11.65) is that it is completely canonical and basis-independent.

See section 11.9 below for more about the regular representation.

Exercise *Due Diligence*
Prove equation (11.67). [106]

Exercise

a.) Let $\delta_0,\delta_1,\delta_2$ be a basis of functions in the regular representation of $\mathbb{Z}_3$ which are 1 on $1,\omega,\omega^2$, respectively, and zero elsewhere. Show that ω is represented as

$$L(\omega) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{11.71}$$

b.) Show that

$$\begin{aligned} L(h)\cdot\delta_g &= \delta_{h\cdot g} \\ R(h)\cdot\delta_g &= \delta_{g\cdot h^{-1}} \end{aligned} \tag{11.72}$$

[106] *Answer*: This is a straightforward computation:

$$\begin{aligned} (h_1,h_2)\cdot\Psi_S(g) &= \Psi_S(h_1^{-1}gh_2) \\ &= \mathrm{Tr}_V(ST(h_2^{-1}g^{-1}h_1)) \\ &= \mathrm{Tr}_V(T(h_1)ST(h_2^{-1})T(g^{-1})) \\ &= \mathrm{Tr}_V((h_1,h_2)\cdot ST(g^{-1})) \\ &= \Psi_{(h_1,h_2)\cdot S}(g) \end{aligned} \tag{11.70}$$

and conclude that for the left, or right, regular representation of a finite group the representation matrices in the δ-function basis are permuation matrices.

11.6 Reducible And Irreducible Representations

11.6.1 Definitions

In general, representations of a group on a vector space of "large dimension" are harder to understand and work with than representations on a vector space of "small dimension." So, presenting a representation as a direct sum of smaller ones is often a very useful simplification. In terms of matrices it corresponds to block diagonalization. We now investigate how that can be done systematically.

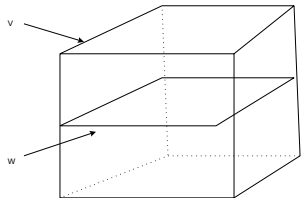


Figure 29: $T(g)$ preserves the subspace W.

Definition. Let $W \subset V$ be a linear subspace of the carrier space V of a group representation $T : G \to GL(V)$. Then W is *invariant* under T, a.k.a. an *invariant subspace* if $\forall g \in G, w \in W$

$$T(g)w \in W \tag{11.73}$$

This may be pictured as in 29.

Examples

1. Both $\{\vec{0}\}$ and V are always invariant subspaces.

2. Consider the three-dimensional representation of $SO(2)$ as rotations around the z-axis. Then the vector subspace of the xy plane at $z = 0$ is an invariant subspace. Note that the parallel planes at $z = z_0 \neq 0$ are invariant under the group action but are not linear subspaces.

3. Consider the canonical representation of S_n on κ^n. The line through the all ones vector is an invariant subspace.

4. We saw above that spaces of functions of the form (11.65) for a fixed V define invariant subspaces of the regular representation with action (11.59). Let us write this out in more detail.

   Let $M : G \to GL(n, \kappa)$ be *any* matrix n-dimensional representation of G. For a fixed i, j consider the matrix element M_{ij} as a κ-valued function on G: So M_{ij} is the

function on G whose value at $g \in G$ is just $M(g)_{ij} \in \kappa$. Now consider the linear span of functions where we fix i:

$$\mathcal{R}_i := Span\{M_{ij}\}_{j=1,\ldots,n} \tag{11.74}$$

We claim this is an invariant subspace of $L^2(G)$ in the right regular representation: To see this, just check:

$$\begin{aligned}(R(g) \cdot M_{ij})(h) &= M_{ij}(hg) \\ &= \sum_{s=1}^{n} M_{is}(h) M_{sj}(g)\end{aligned} \tag{11.75}$$

which is equivalent to the equation on functions:

$$\overbrace{R(g) \cdot M_{ij}}^{\text{function on G}} = \sum_{s=1}^{n_\mu} \underbrace{M_{sj}(g)}_{\text{matrix element for } \mathcal{R}} \overbrace{M_{is}}^{\text{function on G}} \tag{11.76}$$

Similarly, suppose we fix j and consider:

$$\mathcal{L}_j := Span\{M_{ij}\}_{i=1,\ldots,n} \tag{11.77}$$

is an invariant subspace under the left-regular representation of G. Putting these together we see that the space

$$\mathcal{LR} := Span\{M_{ij}\}_{i,j=1,\ldots,n} \tag{11.78}$$

is an invariant subspace under the action of $G \times G$ under the action (11.59) on $L^2(G)$.

Remarks:

1. If (V, T) is a rep and $W \subset V$ is an invariant subspace we can define a smaller group representation (T, W) called *restriction of T to W*. We also say the (T, W) *is a subrepresentation of* (V, T). Strictly speaking we should write $T|_W$ but we will not write that out.
2. If T is unitary on V then it is unitary on W.

Definition. A representation T is called *reducible* if there is an invariant subspace $W \subset V$, under T, which is nontrivial, i.e., such that $W \neq 0, V$, If V is not reducible we say V is *irreducible*. That is, in an irreducible rep, the only invariant subspaces are $\{\vec{0}\}$ and V. We often shorten the term "irreducible representation" to "irrep."

Remarks:

1. Given any nonzero vector $v \in V$, the linear span of $\{T(g)v\}_{g\in G}$ is an invariant subspace. In an irrep this will span all of V. Such a vector is called a *cyclic vector*. Caution!! The existence of a cyclic vector does not imply the representation is irreducible. Consider the vector e_1 in the permutation representation of S_n on $\mathbb{R}^n$.

2. Suppose (T, W) is a subrepresentation of (V, T). Choose an ordered basis

$$\{w_1, \ldots, w_k\} \tag{11.79}$$

for W. Then it can be completed to an ordered basis

$$\{w_1, \ldots, w_k, u_{k+1}, \ldots, u_n\} \tag{11.80}$$

for V. Let us write w_i, $i = 1, \ldots, k$ for the basis vectors for W and u_a, $a = k+1, \ldots, n$ for our choice of a set of complementary ordered basis vectors for V. The matrix representation associated with such a choice of basis is defined by:

$$\begin{aligned} T(g)(w_i) &= (M_{11}(g))_{ji} w_j + (M_{12}(g))_{ai} u_a \\ T(g)(u_a) &= (M_{21}(g))_{ja} w_j + (M_{22}(g))_{ba} u_b \end{aligned} \tag{11.81}$$

where

$$\begin{aligned} &M_{11}(g) \in Mat_{k\times k} \\ &M_{21}(g) \in Mat_{(n-k)\times k} \\ &M_{22}(g) \in Mat_{(n-k)\times(n-k)} \end{aligned} \tag{11.82}$$

Because W is an invariant subspace we have $M_{12} = 0$ so the matrix representation looks like:

$$M(g) = \begin{pmatrix} M_{11}(g) & 0 \\ M_{21}(g) & M_{22}(g) \end{pmatrix} \tag{11.83}$$

Since $M(g_1)M(g_2) = M(g_1 g_2)$ it follows that M_{11} gives a matrix representation on W.

3. If $W \subset V$ is an invariant subspace then the quotient vector space (see Chapter two) V/W is a representation of G in a natural way:

$$T(g)(v + W) := T(g)(v) + W \tag{11.84}$$

The reader can check this is well-defined. The vectors u_a above define a basis for V/W of the form $u_a + W$, and relative to this basis the representation will look like M_{22}.

Definition. A representation T is called *completely reducible* if it is isomorphic to a direct sum of representations:

$$W_1 \oplus \cdots \oplus W_n \tag{11.85}$$

where the W_i are irreducible reps. Thus, there is a basis in which the matrices look like:

$$M(g) = \begin{pmatrix} M_{11}(g) & 0 & 0 & \cdots \\ 0 & M_{22}(g) & 0 & \cdots \\ 0 & 0 & M_{33}(g) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{11.86}$$

Examples

1. Let $G = U(1)$. Then, for any $n \in \mathbb{Z}$ we define the one-dimensional representation ρ_n by
$$\rho_n(z) = z^n \tag{11.87}$$
for $z \in U(1)$. This is clearly an irreducible representation. We will argue below that these are the only irreducible representations.

2. Finite-dimensional representations of Abelian groups are completely reducible. Choosing an ordered ON basis the matrices $M(g)$, $g \in G$ are commuting unitary matrices and, by the spectral theorem, can be simultaneously diagonalized. For example, a fd unitary representation of $U(1)$ will be a family of commuting matrices and we can choose a basis so that
$$M(z) = Diag\{z^{n_1}, \ldots, z^{n_d}\} \tag{11.88}$$
so if $V \cong \mathbb{C}^d$ is the carrier space we would write
$$V \cong \rho_{n_1} \oplus \cdots \oplus \rho_{n_d} \tag{11.89}$$

3. $G = \mathbb{Z}_2$
$$\begin{aligned} 1 &\to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ (12) &\to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{11.90}$$
is a 2-dimensional *reducible* rep on $\mathbb{R}^2$ because $W = \{(x, x)\} \subset \mathbb{R}^2$ is a nontrivial invariant subspace. Indeed, σ^1 is diagonalizable, so this rep is equivalent to
$$\begin{aligned} 1 &\to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ (12) &\to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{11.91}$$
which is a direct sum of two 1×1 reps.

4. Now consider the nonabelian group S_3. There is a natural 3-dimensional permutation representation of S_3 we defined above: $T(\sigma)e_i = e_{\sigma(i)}$ where e_i is the standard basis

for $\mathbb{R}^3$. As we have noted, the all ones vector $u_0 := e_1 + e_2 + e_3$ spans an invariant subspace $L \subset \mathbb{R}^3$. We can choose basis vectors for the orthogonal complement

$$\begin{aligned} u_1 &:= e_1 - e_2 \\ u_2 &:= e_2 - e_3 \end{aligned} \tag{11.92}$$

and the subspace spanned by u_1, u_2 is also an invariant subspace. In fact, one easily computes that relative to this basis:

$$\begin{aligned} M((12)) &= \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \qquad M((23)) = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \qquad M((13)) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \\ M((123)) &= \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \qquad M((132)) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \end{aligned} \tag{11.93}$$

Note that this representation is clearly irreducible: Since it is a representation of a finite group it is completely reducible (see below). The reduction would have to give diagonal matrices. But diagonal matrices commute. The above matrices do not commute since $T(12)T(23) = T(123)$ is not the same as $T(23)T(12) = T(132)$.

Note that the basis u_1, u_2 is <u>not</u> an ON basis and the above matrices are <u>not</u> unitary matrices, even though this is a unitary representation. One could use the averaging procedure above to make a unitary representation, although this would be tedious. A better way to proceed is to note that the reflections and rotations in the plane that preserve an equilateral triangle form the group S_3. (We will discuss this in much more detail later.) If we label the vertices 123 in counter-clockwise order with vertex 3 on the y-axis, as in figure 35 then:

$$M((12)) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{11.94}$$

$$M((13)) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \tag{11.95}$$

$$M((23)) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \tag{11.96}$$

$$M((123)) = R(\frac{2\pi}{3}) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \tag{11.97}$$

$$M((132)) = R(\frac{-2\pi}{3}) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \tag{11.98}$$

generates a unitary representation.

5. Consider the representation of S_n on $\mathbb{R}^n$. Then the one-dimensional subspace $L = \{(x, \cdots, x)\}$ is a subrepresentation. Moreover we can take

$$L^\perp = \{x_1, \cdots x_n | \sum x_i = 0\}. \tag{11.99}$$

Then $L^\perp$ is an $(n-1)$-dimensional representation of S_n and the representation is equivalent to a direct sum $L \oplus L^\perp$. So, if we choose a basis of the all ones vector and an orthogonal basis for $L^\perp$ the matrices will be block diagonal. It is not obvious, but it does follow from the general representation theory of the symmetric group that the representation $L^\perp$ of dimension $(n-1)$ is irreducible.

11.6.2 Reducible vs. Completely reducible representations

Irreps are the "atoms" out of which all reps are made. Thus we are naturally led to study the *irreducible* reps of G. In real life it can and does actually happen that a group G has representations which are reducible but not completely reducible. Reducible, but not completely reducible reps are sometimes called *indecomposable*.

Example 1 An example of an indecomposible rep which is not completely reducible is the rep

$$A \to \begin{pmatrix} 1 & \log|\det A| \\ 0 & 1 \end{pmatrix} \tag{11.100}$$

of $GL(n, \mathbb{R})$.

Example 2 Similarly, we can write an indecomposable representation of the connected component of the identity of the $1+1$ dimensional Lorentz group. If $B(\eta) \in SO_0(1,1)$ is the boost of rapidity η, with $\eta \in \mathbb{R}$, then:

$$T(B(\eta)) = \begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix} \tag{11.101}$$

Example 3 Let $G = GL(n, \kappa)$ and $H = \mathrm{Mat}_n(\kappa)$. We can define a group which, as a set is $H \times G$, but it has a "twisted" group multiplication law:

$$\mathbf{m}((h_1, g_1), (h_2, g_2)) := (h_1 + g_1 h_2 g_1^{tr}, g_1 g_2) \tag{11.102}$$

One can check that this really does define a group structure. It is a special case of the semi-direct product structure that we will study in more detail in Section **** . Now check that the following is a matrix representation:

$$T(h, g) := \begin{pmatrix} g & h g^{tr,-1} \\ 0 & g^{tr,-1} \end{pmatrix} \tag{11.103}$$

Because of the upper-triangular nature of the matrices we cannot simultaneously block-diagonalize all the matrices $T(h, g)$ but there is a nontrivial invariant subspace. This kind

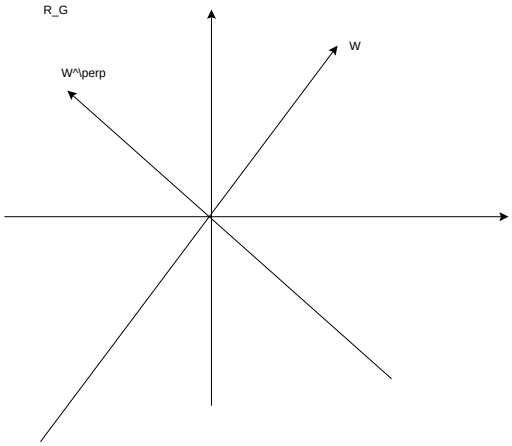


Figure 30: The orthogonal complement of an invariant subspace is an invariant subspace.

of construction gives the finite-dimensional indecomposable representations of the Poincaré group, the Euclidean group, and its crystallographic subgroups.

It is useful to have criteria for when this complication cannot occur:

Proposition 11.6.2.1 Suppose that the representation (V, T) is a unitary representation on an inner product space V and $W \subset V$ is an invariant subspace then $W^\perp$ is an invariant subspace:

Proof: Recall that

$$y \in W^\perp \Leftrightarrow \forall x \in W, \quad \langle y, x \rangle = 0 \tag{11.104}$$

Let $g \in G$, $y \in W^\perp$. Compute

$$\begin{aligned} \langle T(g)y, x \rangle &= \langle y, T(g)^\dagger x \rangle \\ &= \langle y, T(g^{-1})x \rangle \end{aligned} \tag{11.105}$$

But, $T(g^{-1})x \in W$, since $W \subset V$ is an invariant subspace. Therefore: $\forall x \in W$, $\langle T(g)y, x \rangle = 0$. Therefore: $T(g)y \in W^\perp$ Therefore: $W^\perp$ is an invariant subspace. ♠

Therefore:

1. *Finite dimensional unitary representations are always completely reducible.* This follows from the above proposition together with induction on the dimension.

2. *Finite dimensional representations of compact groups are always completely reducible.*

3. *In particular, for a finite group the regular representation $L^2(G)$ is completely reducible as a representation of G or of $G \times G$.*

4. The complete reduction $L^2(G)$ is given by the Peter-Weyl theorem described below. The beautiful aspect of this theorem is that it holds not just for finite groups but for all compact groups.

Isotypical Components: For each isomorphism class of irreducible representation of G choose a representative $(T^{(\mu)}, V^{(\mu)})$ where μ runs over the set of distinct irreducible representations. If V is a completely decomposable representation then we can write

$$V \cong \oplus_\mu \oplus_{i=1}^{a_\mu} V^{(\mu)} \tag{11.106}$$

where $V^{(\mu)}$ is the carrier space of an irreducible representation of G, we are summing over all irreps in (11.106), and a_μ is the number of times that irrep appears in the decomposition. (We understand that if a particular irrep does not appear at all then we take $a_\mu = 0$.) When $a_\mu \neq 0$ we can write

$$\oplus_{i=1}^{a_\mu} V^{(\mu)} \cong \kappa^{a_\mu} \otimes_\kappa V^{(\mu)} \tag{11.107}$$

as vector spaces where κ is the ground field. We can moreover interpret this as an interpretation as representations where $T(g)$ acts as the identity operator on κ^{a_μ}. With this understood, the summand $\kappa^{a_\mu} \otimes_\kappa V^{(\mu)}$ is called the *isotypical component of* V belonging to μ. Note that here If we abbreviate $\kappa^{a_\mu} \otimes V^{(\mu)}$ to $a_\mu V^{(\mu)}$ then the decomposition into isotypical components can be written as:

$$V = \oplus a_\mu V^{(\mu)} \tag{11.108}$$

11.7 Schur's Lemmas

A very important remark about equivariant maps between irreducible representations is known as *Schur's lemma*. It is almost a tautology - but it is a very powerful tautology.

> Schur's Lemma: Let V_1, V_2 be vector spaces over any field κ such that they are carrier spaces of irreducible representations of any group G. If $A : V_1 \to V_2$ is an intertwiner between these two irreps then A is either zero or an isomorphism of representations.

Proof: Note that the kernel and image of A are invariant subspaces of V_1 and V_2, respectively. Recall these are defined by:

$$\ker A := \{v_1 \in V_1 | A(v_1) = 0\} \tag{11.109}$$

$$\mathrm{im} A := \{v_2 \in V_2 | \exists v_1 \in V_1 \qquad v_2 = A(v_1)\} \tag{11.110}$$

The reader should check that these are linear subspaces because A is linear and they are invariant subspaces because A is an intertwiner.

Now, since V_1 is an irrep we conclude that $\ker A$ is either the 0 vector space or all of V_1. Similarly, since V_2 is an irrep $\mathrm{im} A$ is either 0 or the entire space V_2. Now, if $\ker A = V_1$ then $A = 0$. If $A \neq 0$ then $\ker A \neq V_1$ so therefore $\ker A = 0$, therefore A is injective. Moreover if $A \neq 0$ then there is a nonzero vector in $\mathrm{im} A$, and therefore $\mathrm{im} A = V_2$, so A is surjective. Therefore, A is an isomorphism. ♠

Note that if $V_1 = V_2 = V$ is an irrep then the set of intertwiners $V \to V$ is not only a linear space but is also an algebra (see Chapter 2). Namely, $A_1 \circ A_2$ is also an intertwiner. In the case of the algebra of self-intertwiners of a representation one can say more. But now the choice of field κ becomes important.

Theorem: Suppose (V, T) is an irreducible representation of a group G and V is a complex vector space. Suppose $A : V \to V$ is an intertwiner, i.e., A commutes with $T(g)$ for all $g \in G$. Then A is proportional to the identity transformation: There exists a scalar $\lambda \in \mathbb{C}$ such that for all $v \in V$, $A(v) = \lambda v$.

Proof: Since we are working over the complex field A has a nonzero eigenvector $Av = \lambda v$. That follows because the characteristic polynomial $p_A(x) = \det(x1 - A)$ is a polynomial in the complex field and has a root in the complex numbers. The eigenspace $C = \{w : Aw = \lambda w\}$ is therefore not the zero vector space. But it is also an invariant subspace. Therefore, it must be the entire carrier space. ♠.

Remarks:

1. *Degeneracy Spaces As Spaces Of Intertwiners.* Let us return to the isotypical decomposition of a completely reducible representation, (11.106). Let $\mathrm{Hom}^G(V_1, V_2)$ denote the vector space of G-equivariant maps between two G-spaces V_1, V_2.

$$\begin{aligned} \mathrm{Hom}_G(V^{(\mu)}, V) &\cong \oplus_\nu \mathrm{Hom}_G(V^{(\mu)}, \kappa^{a_\nu} \otimes V^{(\nu)}) \\ &\cong \oplus_\nu \kappa^{a_\nu} \otimes \mathrm{Hom}_G(V^{(\mu)}, V^{(\nu)}) \\ &= \kappa^{a_\mu} \otimes \mathrm{Hom}_G(V^{(\mu)}, V^{(\mu)}) \end{aligned} \tag{11.111}$$

   In the second line we used the fact that G acts trivially on κ^{a_ν}. If we work over $\kappa = \mathbb{C}$ then we just showed $\mathrm{Hom}_G(V^{(\mu)}, V^{(\mu)}) \cong \mathbb{C}$ and hence we have a better interpretation of the degeneracy space: It is the linear space of G-invariant maps from $V^{(\mu)} \to V$. Indeed, note that there is a canonical equivariant map:

$$\mathrm{Hom}_G(V^{(\mu)}, V) \otimes V^{(\mu)} \to V \tag{11.112}$$

   given by $A \otimes u \mapsto A(u)$. So we have a canonical G-equivariant map

$$\oplus_\mu \mathrm{Hom}_G(V^{(\mu)}, V) \otimes V^{(\mu)} \to V \tag{11.113}$$

   and complete reducibility is the statement that this is an isomorphism.

2. *Block diagonalization of Hamiltonians.* Suppose $\mathcal{H}$ is a physical Hilbert space which is a representation of a group G so that $\mathcal{H}$ is completely reducible into isotypical components:

$$\mathcal{H} \cong \oplus_\mu \mathcal{H}^{(\mu)} \tag{11.114}$$

That is $\mathcal{H}^{(\mu)} \cong D_\mu \otimes V^{(\mu)}$ where $V^{(\mu)}$ is an irreducible representation and D_μ is the *degeneracy space*. The sum in (11.114) is over the irreps of G. Suppose now that H is a Hermitian operator, such as a Hamiltonian, that commutes with the G-action. That is

$$H : \mathcal{H} \to \mathcal{H} \tag{11.115}$$

is an intertwiner. This is typically what happens when we have a symmetry of some dynamics. By Schur's lemma we therefore have under this isomorphism

$$H \cong \oplus_\mu H^{(\mu)} \otimes 1_{V^{(\mu)}} \tag{11.116}$$

In terms a basis compatible with the isotypical decomposition it means that $H^{(\mu)}$ acts only on the degeneracy space, and hence the Hamiltonian has been partially (or sometimes completely) diagonalized. This also leads to *selection rules*. If ψ_1, ψ_2 are two states in different isotypical components and $\mathcal{O}$ is an operator that commutes with the G-action then the transition amplitude

$$\langle \psi_1, \mathcal{O}\psi_2 \rangle = 0 \tag{11.117}$$

****

ABOVE DISCUSSION NEEDS TO BE SUPPLEMENTED IN MORE CONCRETE TERMS:

$D \otimes V$ with G acting as $1 \otimes T(g)$ commutes with $A \otimes 1$, with A any linear transformation on D. If V is irreducible

$$\mathrm{Hom}_G(D \otimes V, D \otimes V) \cong \mathrm{Hom}(D, D) \tag{11.118}$$

WRITE OUT WHAT THIS MEANS IN TERMS OF BLOCK DIAGONAL MATRICES

****

3. Schur's lemma over other fields can lead to more complicated possibilities. All we can say in general is that $\mathrm{Hom}^G(V^{(\mu)}, V^{(\mu)})$ is a *division algebra* over κ, meaning that it is an algebra in which all nonzero elements are invertible. (This follows immediately from our proof of Schur's lemma above.) When we take $\kappa = \mathbb{R}$ we get a division algebra over $\mathbb{R}$ and there are three possibilities: $\mathbb{R}, \mathbb{C}, \mathbb{H}$. As an example of the case $\mathbb{R}$ just consider the real linear transformations between the irreducible real representations of $\mathbb{Z}_2$. The algebra of intertwiners is clearly just $\mathbb{R}$. If we consider the defining representation $\mathbb{C}^2 \cong \mathbb{R}^4$ as a representation of $SU(2)$ over $\mathbb{R}$, so that $SU(2)$ elements are represented as 4×4 real matrices, then there are more $\mathbb{R}$-linear transformations on $\mathbb{R}^4$ that can commute with the representation matrices $T(g)$. (If we tried to express them as transformations of $\mathbb{C}^2$ they would typically be a linear combination of $\mathbb{C}$-linear and $\mathbb{C}$-antilinear transformations.) This is an example where the algebra of intertwiners is $\mathbb{H}$. For more on this see the section in chapter two on quaternions.

11.8 Pontryagin Duality

In this section we introduce the beautiful idea of the *Pontryagin dual* of an Abelian group. [107] We will use it in section 11.11, and again we will use it to give a very general construction of interesting Heisenberg extensions of Abelian groups.

Definition: Let S be an Abelian group. The *Pontryagin dual* group $\widehat{S}$ is defined to be the group of homomorphisms $\mathrm{Hom}(S, U(1))$. Note that if $\chi_1, \chi_2 \in \mathrm{Hom}(S, U(1))$ then the pointwise product

$$(\chi_1 \cdot \chi_2)(s) := \chi_1(s)\chi_2(s) \tag{11.119}$$

is again a homomorphism $S \to U(1)$, thus making $\widehat{S}$ into an Abelian group.

Remarks:

1. The Pontryagin dual group $\widehat{S}$ can also be thought of as the group of all complex one-dimensional unitary representations of S. It follows from Schur's lemma that all irreducible finite dimensional complex representations of an Abelian group are one-dimensional.

   Note that the adjective complex is essential here. After all the defining representation of the Abelian group $SO(2)$ is $\mathbb{R}^2$ and is irreducible as a representation over $\mathbb{R}$.

2. Elements of the group $\widehat{S}$ are also called characters.

3. It is best to discuss Pontryagin duality in the context of topological groups. In this case we should only consider the continuous characters $\chi : S \to U(1)$. For the duality theorem below we should consider *locally compact Abelian groups*. Examples of locally compact Abelian groups are $\mathbb{R}^n$, tori, lattices, and finite Abelian groups with compact topology. An infinite-dimensional Hilbert space is a topological Abelian group under addition, but it is not locally compact.

Note that, for a fixed $s \in S$ we can define a homomorphism $\mathrm{Hom}(\widehat{S}, U(1))$ by

$$\hat{s} : \chi \mapsto \chi(s) \tag{11.120}$$

Note that $\hat{s}$ is just the evaluation map ev_s discussed previously. The map $s \mapsto \hat{s}$ is a homomorphism $S \to \widehat{\widehat{S}}$. The main theorem is:

Theorem[Pontryagin-van Kampen duality]. If G is a locally compact Abelian group then the canonical homomorphism $S \to \widehat{\widehat{S}}$ is in fact an isomorphism:

$$\widehat{\widehat{S}} \cong S \tag{11.121}$$

For a proof see, for example, the book on representation theory by A.A. Kirillov.

[107] The transliteration from the Cyrillic to the Latin alphabets takes various forms. Another common one is Pontrjagin.

Example 1: Consider $S = \mathbb{Z}/n\mathbb{Z}$, thought of additively. To determine $\chi \in \mathrm{Hom}(S, U(1))$ it suffices to determine $\chi(\bar{1})$, since $\chi(\bar{\ell}) = \chi(\bar{1})^\ell$ for any $\ell \in \mathbb{Z}$. Put $\chi(\bar{1}) = \omega \in U(1)$. But now we need to impose the relation $\chi(\bar{n}) = \chi(\bar{0}) = 1$. This implies $\omega^n = 1$, so ω is an nth root of unity. So the most general element of $\widehat{\mathbb{Z}/n\mathbb{Z}}$ is

$$\chi(\bar{\ell}) = \chi_\omega(\bar{\ell}) := \omega^\ell \tag{11.122}$$

where ω is an nth root of unity. Moreover $\chi_{\omega_1}\chi_{\omega_2} = \chi_{\omega_1\omega_2}$ so $\widehat{\mathbb{Z}/n\mathbb{Z}}$ is identified in this way with the multiplicative group of nth roots of unity μ_n In this way we see that, as abstract groups

$$\widehat{\mathbb{Z}/n\mathbb{Z}} = \mu_n \cong \mathbb{Z}/n\mathbb{Z} \tag{11.123}$$

So a finite cyclic group is self-dual.

Example 2: Consider $\mathbb{R}$, additively. Then if $\chi \in \widehat{\mathbb{R}}$ we have $\chi(x+y) = \chi(x)\chi(y)$ so $\chi(x) = e^{ax}$ for some constant a. For χ to be valued in $U(1)$ we must have $a = \mathrm{i}k$ with $k \in \mathbb{R}$ and hence

$$\chi(x) = \chi_k(x) := e^{\mathrm{i}kx} \tag{11.124}$$

moreover,

$$\chi_k\chi_\ell = \chi_{k+\ell} \tag{11.125}$$

and hence $\widehat{\mathbb{R}} \cong \mathbb{R}$. In an entirely similar way $\widehat{\mathbb{R}^n} \cong \mathbb{R}^n$

Example 3: Consider $S = \mathbb{Z}$. To determine $\chi \in \mathrm{Hom}(S, U(1))$ it suffices to determine $\chi(1)$. Choose any phase $\xi \in U(1)$ and set $\chi(1) = \xi$. Then it must be that, for all $n \in \mathbb{Z}$:

$$\chi(n) = \chi_\xi(n) := \xi^n \tag{11.126}$$

Moreover $\chi_{\xi_1}\chi_{\xi_2} = \chi_{\xi_1\xi_2}$. Thus,

$$\widehat{\mathbb{Z}} \cong U(1) \tag{11.127}$$

Example 4: Consider $S = U(1)$. To determine $\chi \in \mathrm{Hom}(S, U(1))$ it might help to think of $U(1) \cong \mathbb{R}/\mathbb{Z}$. We know from the Pontryagin dual of $\mathbb{R}$ that χ should be of the form

$$\chi(x + \mathbb{Z}) = \exp[\mathrm{i}k(x + \mathbb{Z})] \tag{11.128}$$

for some real number k. However, for this to be well-defined we must have $k = 2\pi n$ with $n \in \mathbb{Z}$. Therefore χ must be of the form

$$\chi_n(x + \mathbb{Z}) = \exp[2\pi\mathrm{i}nx] \tag{11.129}$$

Or, if we think of $S = U(1)$ multiplicatively as complex numbers of modulus one, then we can say that every character on $U(1)$ is of the form:

$$\chi_n(\xi) := \xi^n \tag{11.130}$$

for some $n \in \mathbb{Z}$. Therefore,

$$\widehat{U(1)} \cong \mathbb{Z} \tag{11.131}$$

Comparing (11.127) and (11.131) we verify the general result (11.121).

Example 5: *Tori.* Consider the group $G = \mathbb{Z}^d$. It will be useful to consider a free G action on affine Euclidean space $\mathbb{E}^d$. This defines a subset $\Gamma \subset \mathbb{E}^d$ known as a *lattice* (sometimes called an *embedded lattice*). The quotient space $\mathbb{E}^d/\Gamma$ has a natural basepoint, namely the coset of Γ and, as a group it is isomorphic to $U(1)^d$. By the same arguments as above its Pontryagin dual will be isomorphic to $\mathbb{Z}^d$.

There is a nice way to think about the Pontryagin duality between lattices and tori. Suppose Γ is a lattice in $\mathbb{R}^d$. Using the Euclidean norm we can define another lattice, the *dual lattice*

♣Really we should do this using the dual vector space... ♣

$$\Gamma^\vee = \{g \in \mathbb{R}^d | g \cdot \gamma \in \mathbb{Z} \qquad \forall \gamma \in \Gamma\} \tag{11.132}$$

Note we can identify $\Gamma^\vee \cong \mathrm{Hom}(\Gamma, \mathbb{Z})$ since any $g \in \Gamma^\vee$ defines a homomorphism ϕ_g whose values are: $\phi_g(\gamma) = g \cdot \gamma$. Also, as an abstract Abelian group of course $\Gamma^\vee \cong \mathbb{Z}^d$, but the above definition identifies it as a specific subgroup of $\mathbb{R}^d$.

The unitary irreps of Γ are represented by points in the torus $T^\vee := \mathbb{R}^d/\Gamma^\vee$. Note that $T^\vee$ is a torus as a manifold and is isomorphic to the group $U(1)^d$, as an Abelian group. For any $\bar{k} \in T^\vee$ we can

$$\chi_{\bar{k}}(\gamma) = \exp[2\pi \mathrm{i} k \cdot \gamma] \tag{11.133}$$

where k is any representative of $\bar{k}$, that is $\bar{k} = k + \Gamma^\vee$. Note that the above formula is well-defined because if we choose any two lifts k_1 and k_2 of $\bar{k}$ then $k_1 = k_2 + g$ with $g \in \Gamma^\vee$, and then $g \cdot \gamma \in \mathbb{Z}$ for all $\gamma \in \Gamma$. So

$$\widehat{\Gamma} \cong \mathbb{R}^d/\Gamma^\vee \cong U(1)^d \tag{11.134}$$

Conversely, the Pontryagin dual of the torus $\mathbb{R}^d/\Gamma^\vee$ can naturally be identified with Γ by the same formula:

$$\chi_\gamma(\bar{k}) = \exp[2\pi \mathrm{i} k \cdot \gamma] \tag{11.135}$$

Exercise *Pontryagin Dual Of The Prüfer Groups*

Recall that the Prüfer groups $Pr(p)$ are defined for each prime p as the union over all n of roots of unity of order p^n.

What is the Pontryagin dual of $Pr(p)$? (Give it the discrete topology.) [108]

---

[108] *Answer*: Use the isomorphism $U(1) \cong \mathbb{R}/\mathbb{Z}$. One needs to say what is the image of p^{-n}. So $\chi(\frac{1}{p^n}) = \exp[2\pi \mathrm{i} a_n/p^n]$, for some integer a_n because $\chi(p^{-n})$ must itself be a $(p^n)^{th}$ root of unity. Note we can regard $a_n \in \mathbb{Z}/p^n\mathbb{Z}$. Now note that

$$\exp[2\pi \mathrm{i} \frac{a_{n-1}}{p^{n-1}}] = \chi(\frac{1}{p^{n-1}}) = \chi(\frac{1}{p^n})^p = \exp[2\pi \mathrm{i} \frac{a_n}{p^{n-1}}] \tag{11.136}$$

So the Pontryagin dual is the subgroup of $\prod_n \mathbb{Z}/p^n\mathbb{Z}$ consisting of sequences a_n so that a_n projects to a_{n-1}. This is known as the group of p-adic integers.

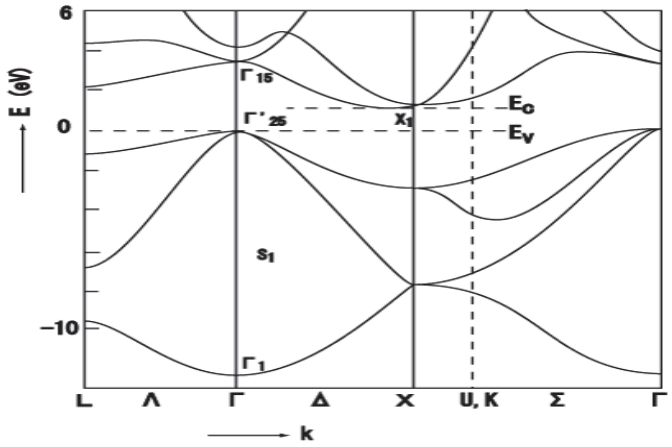


Figure 31: Example of a bandstructure. (For silicon.) On the horizontal axis the structure is plotted as a function of k along lines inside the Brillouin torus. The letters refer to points where the (cubic) crystallographic group has fixed points. Γ denotes the identity element $\bar{k} = 0$ where the full cubic symmetry group is restored.

11.8.1 An Application Of Pontryagin Duality: Bloch's Theorem And Band Structure

The Pontryagin duality between an embedded lattice $\Gamma \subset \mathbb{R}^d$ and the torus $T^\vee = \mathbb{R}^d/\Gamma^\vee$ has a very significant application in condensed matter physics known as *Bloch's theorem*.

In the one-electron approximation to the Schrodinger problem of electrons in a crystal one considers the Schrödinger Hamiltonian on $L^2(\mathbb{R}^d)$ of the form:

$$H = -\frac{\hbar^2}{2m}\nabla^2 + U(x) \tag{11.137}$$

where the potential $U : \mathbb{R}^d \to \mathbb{R}$ is assumed to be invariant under some crystallographic group (see below). In particular, it is invariant under translation by a lattice $\Gamma \subset \mathbb{R}^d$. For exampe, if we take into account the Coulomb interaction between the electron and a collection of ions of charge $Z_i e$ at positions $x_i \in C$, where C is a crystal then

$$U(x) = \sum_i \frac{-Z_i e^2}{|x - x_i|} \tag{11.138}$$

But for the statement we are going to make all we need is that $U(x+\gamma) = U(x)$ for $\gamma \in \Gamma$. Now the group Γ acts on the Hilbert space through unitary operators commuting with H. Explicitly:

$$\rho(\gamma) = \exp[\mathrm{i}\gamma \cdot \hat{p}/\hbar] \tag{11.139}$$

where $\hat{p} = -\mathrm{i}\hbar\vec{\nabla}$ as is standard in quantum mechanics. Note that

$$\rho(\gamma_1)\rho(\gamma_2) = \rho(\gamma_1 + \gamma_2) \tag{11.140}$$

so the Hilbert space will be a representation of Γ. The Hamiltonian cannot make transitions between different irreps in an isotypical decomposition.

We have classified the one-dimensional representations above so if $\psi \in \mathcal{H}$ were to be in a one-dimensional representation then it would have to be quasi-periodic:

$$\psi(x+\gamma) = \chi_{\bar{k}}(\gamma)\psi(x) \tag{11.141}$$

where $\bar{k} \in \mathbb{R}^d/\Gamma^\vee$. In this context the Pontryagin dual torus $T^\vee$ is known as the *Brillouin torus*.

Note that if ψ obeys (11.141) then we can always write (noncanonically!!!) as

$$\psi(x) = e^{2\pi i k\cdot x} u_k(x) \tag{11.142}$$

where we have chosen a specific representative $k \in \mathbb{R}^d$ of the element $\bar{k}$ in the Brillouin torus and $u_k(x)$ is periodic in x, i.e. invariant under shifts of $x \to x+\gamma$ for $\gamma \in \Gamma$. In condensed matter physics the vector $k \in \mathbb{R}^d$ is known as a *reciprocal vector*. We stress that this decomposition is noncanonical. If k_1 and k_2 both represent the same $\bar{k}$ then $k_1 = k_2 + g$ for $g \in \Gamma^\vee$ and $u_{k_2}(x) = e^{2\pi i g\cdot x} u_{k_1}(x)$. Note that both functions u_{k_2} and u_{k_1} are periodic under shifts by Γ. Therfore, they can be considered as functions on the *real space torus* $T := \mathbb{R}^d/\Gamma$.

Note that a quasiperiodic function (11.141) cannot also be L^2. For each point $\bar{k}$ in the Brillouin torus define the Hilbert space

$$\mathcal{H}_{\bar{k}} := \{\psi(x) | \psi(x+\gamma) = \chi_{\bar{k}}(\gamma)\psi(x) \qquad \int_{\mathbb{R}^d/\Gamma} |\psi(x)|^2 dx < \infty\} \tag{11.143}$$

Then there is a measure on the torus $T^\vee$ so that we have an isomorphism

$$\mathcal{H} \cong \oint_{\mathbb{R}^d/\Gamma^\vee} d\bar{k} \mathcal{H}_{\bar{k}} \tag{11.144}$$

Now, the Hamiltonian acts within $\mathcal{H}_{\bar{k}}$. It is useful to write the eigenvalue problem as

$$H_k u_k(x) = E_k u_k(x) \tag{11.145}$$

where

$$H_k = e^{-2\pi i k\cdot x} H e^{2\pi i k\cdot x} \tag{11.146}$$

Note that we had to make a choice of k that projects to $\bar{k}$ to write the Hamiltonian H_k. However, if $k' = k + g$ where $g \in \Gamma^\vee$ then $H_{k'}$ is unitarily equivalent to H_k. Indeed $U = e^{2\pi i \hat{x}\cdot g}$ is a nice unitary operator on the wavefunctions on the torus $\mathbb{R}^d/\Gamma$.

H_k is an Hermitian elliptic operator acting on the functions on a compact manifold. Explicitly, it works out to

$$H_k = -\frac{\hbar^2}{2m}\nabla^2 - 4\pi\frac{\hbar^2}{2m} k\cdot(i\nabla) + (U + \frac{\hbar}{2m} 4\pi^2 k^2) \tag{11.147}$$

and it acts on L^2 functions on the torus $T = \mathbb{R}^d/\Gamma$. This operator should be viewed as a perturbation of a Laplace operator on functions on a compact manifold. The latter has a

discrete spectrum of eigenvalues. The spectrum of $h = -\nabla^2$ is $\{4\pi^2 g^2\}_{g\in\Gamma^\vee}$. The theory of elliptic Hermitian operators on compact manifolds shows that the lower order terms in (11.147) do not change this property and hence the operator (11.147) has a discrete set of eigenvalues $\{E_n(k)\}$. Note that while H_k depends on k, the spectrum itself only depends on $\bar{k}$. This is not obvious from (11.147) but it follows from the unitary equvalence between H_{k_1} and H_{k_2} where k_1 and k_2 are two representatives of $\bar{k}$.

The eigenvalues vary continuously as functions of $\bar{k} \in \mathbb{R}^d/\Gamma^\vee$ to give what is called a *band structure*. See Figure 31.

11.9 Orthogonality Relations Of Matrix Elements And The Peter-Weyl Theorem

The very beautiful *Peter-Weyl theorem* states that, if G is a compact group, then there is an isomorphism of $G \times G$ representations:

$$L^2(G) \cong \oplus_\mu \mathrm{End}(V^{(\mu)}) \tag{11.148}$$

where we sum over the isomorphism class of each irreducible representation exactly once, and for each irrep we choose a representative $(T^{(\mu)}, V^{(\mu)})$.

♣Notation is too heavy: Replace $V^{(\mu)}$ by V^μ etc. below. ♣

The key to proving the Peter-Weyl theorem are the *orthogonality relations for matrix elements of irreps*:

Theorem: Let G be a compact group, and define an Hermitian inner product on $L^2(G)$ by

$$\langle \Psi_1, \Psi_2 \rangle := \int_G \Psi_1^*(g)\Psi_2(g)dg \tag{11.149}$$

where, WLOG we normalize the Haar measure so the volume of G is one. Let $\{V^{(\mu)}\}$ be a set of representatives of the distinct isomorphism classes of irreducible unitary representations for G. For each representation $V^{(\mu)}$ choose an ON basis $w_i^{(\mu)}$, $i = 1, \ldots, n_\mu$ with

$$n_\mu := \dim_{\mathbb{C}} V^{(\mu)} \ . \tag{11.150}$$

Then the matrix elements $M_{ij}^\mu(g)$ defined by

$$T(g)w_i^{(\mu)} = \sum_{j=1}^{n_\mu} M_{ji}^\mu(g) w_j^{(\mu)} \tag{11.151}$$

form a complete orthogonal set of functions on $L^2(G)$ so that

$$(M_{i_1,j_1}^{\mu_1}, M_{i_2,j_2}^{\mu_2}) = \frac{1}{n_\mu}\delta^{\mu_1,\mu_2}\delta_{i_1,i_2}\delta_{j_1,j_2} \tag{11.152}$$

Proof: The proof is based on linear algebra and Schur's lemma. For any linear transformation $A : V^{(\mu)} \to V^{(\nu)}$ we can average using the Haar measure

$$\tilde{A} := \int_G T^{(\nu)}(g) A T^{(\mu)}(g^{-1}) dg \tag{11.153}$$

And then a small computation shows that $\tilde{A}$ is an intertwiner:

$$\begin{aligned} T^{(\nu)}(h)\tilde{A} &= \int_G T^{(\nu)}(hg)AT^{(\mu)}(g^{-1})dg \\ &= \int_G T^{(\nu)}(g)AT^{(\mu)}(g^{-1}h)dg \\ &= \tilde{A}T^{(\mu)}(h) \end{aligned} \tag{11.154}$$

Therefore, by Schur's lemma, $\tilde{A} = \delta_{\mu,\nu}\hat{A}$ where $\hat{A}$ is a multiple of the identity transformation. It is useful at this point to choose an ordered basis and examine our conclusion: $\tilde{A} = \delta_{\mu,\nu}\hat{A}$, with $\hat{A}$ proportional to the identity, in that basis:

For any matrix $A \in \mathrm{Mat}_{n_\nu \times n_\mu}(\mathbb{C})$ we have

$$\int_G [dg] M^{(\nu)}_{ij}(g) A_{ja} M^{(\mu)}_{ab}(g^{-1}) = \delta_{\mu,\nu} c_A \delta_{i,b} \tag{11.155}$$

We can determine the constant c_A by setting $\mu = \nu$ and $b = i$ and summing on i to get

$$\mathrm{Tr}_{V^{(\mu)}}(A) = n_\mu c_A \tag{11.156}$$

where we normalized the volume of the group to 1.

Note that for the matrix unit $A = e_{jk}$ we have $\mathrm{Tr} e_{jk} = \delta_{jk}$ and hence

$$\int_G [dg] M^{(\nu)}_{ij}(g) M^{(\mu)}_{k\ell}(g^{-1}) = \frac{1}{n_\mu} \delta_{\mu,\nu}\delta_{jk}\delta_{i,\ell} \tag{11.157}$$

Equation (**??**) holds for the matrix elements relative to <u>any</u> ordered bases for the $V^{(\mu)}$. If we specialize to ON bases then the matrices $M^{(\mu)}(g)$ are unitary. It is now useful to define functions $\phi^\mu_{ij} \in L^2(G)$ by

$$\phi^{(\mu)}_{ij} : g \mapsto \sqrt{n_\mu} M^{(\mu)}_{ij}(g) \tag{11.158}$$

so that we have

$$\int_G [dg] (\phi^{(\mu)}_{ij}(g))^* \phi^{(\nu)}_{k\ell}(g) = \delta_{\mu,\nu}\delta_{ik}\delta_{j\ell} \ . \tag{11.159}$$

Now recall the definition of the map $\iota : \mathrm{End}(V) \to L^2(G)$ in (11.66). This is easily generalized: If we have any collection of finite-dimensional representations $\{V_\lambda\}$ of G then we have

$$\iota : \oplus_\lambda \mathrm{End}(V_\lambda) \hookrightarrow L^2(G) \tag{11.160}$$

where we just add the functions $\iota(\oplus_i S_i) := \sum_i \Psi_{S_i}$. Thanks to the equivariance, the image of (11.160) is a $G \times G$-invariant subspace, i.e. a subrepresentation of $L^2(G)$. It follows from the orthogonality of matrix elements that if we choose the collection $\{V_\lambda\}$ to be the set of distinct irreps of G, namely $\{V^{(\mu)}\}$ then ι is injective. We claim it is also surjective: The orthogonal complement would have to be a representation of G, and hence would be isomorphic to a direct sum of representations $V^{(\mu)}$. But then the matrix elements are already accounted for by the image of ι. Thus the functions $\phi^{(\mu)}_{ij}$ are in fact an <u>orthonormal</u> basis for $L^2(G)$. ♠

♣Be more explicit and careful about this step. ♣

When working with finite groups then we have the following beautiful:

Corollary: If G is a finite group then

$$|G| = \sum_{\mu} n_\mu^2 \tag{11.161}$$

Proof: On the one hand $L^2(G)$ clearly has a basis of delta-functions δ_g so the dimension is $|G|$ on the other hand $\mathrm{End}(V^{(\mu)})$ has dimension n_μ^2

This fact is extremely useful: If you are trying to list the dimensions of the irreps of a finite group it is always a very useful check to verify (11.161).

♣SHOULD WE CHANGE D TO M FOR MATRIX ELEMENTS BELOW? ♣

Example 1: Let $G = \mathbb{Z}_2 = \{1, \sigma\}$ with $\sigma^2 = 1$. Then the general complex valued-function on G is specified by two complex numbers $(\psi_+, \psi_-) \in \mathbb{C}^2$:

$$\Psi(1) = \psi_+ \qquad \Psi(\sigma) = \psi_- \tag{11.162}$$

This identifies $\mathrm{Map}(G, \mathbb{C}) \cong \mathbb{C}^2$ as a vector space. We found the irreps of any cyclic group above. For $\mathbb{Z}_2$ there are just two irreducible representations $V_\pm \cong \mathbb{C}$ with $\rho_\pm(\sigma) = \pm 1$. The matrix elements give two functions on the group $M^\pm$:

$$M^+(1) = 1 \qquad M^+(\sigma) = 1 \tag{11.163}$$

$$M^-(1) = 1 \qquad M^-(\sigma) = -1 \tag{11.164}$$

(Here and in the next examples when working with 1×1 matrices we drop the $\mu\nu$ subscript!) The reader can check they are orthonormal, and they are complete because any function Ψ can be expressed as:

$$\Psi = \frac{\psi_+ + \psi_-}{2} M^+ + \frac{\psi_+ - \psi_-}{2} M^- \tag{11.165}$$

Note that if V is any representation of $\mathbb{Z}_2$ with nontrivial element represented by $T(\sigma)$ then we can form orthogonal projection operators $P_\pm = \frac{1}{2}(1 \pm \rho(\sigma))$ onto direct sums of isotypical components. Note that we can write these projectors as:

$$P_\pm = \int_G (M^\pm(g))^* T(g) \tag{11.166}$$

Example 2: We can generalize the previous example slightly by taking $G = \mathbb{Z}/n\mathbb{Z} = \langle \omega | \omega^n = 1 \rangle$. Let us identify this group with the group of nth roots of unity and choose a generator $\omega = \exp[2\pi \mathrm{i}/n]$. Since G is abelian all the representation matrices can be simultaneously diagonalized so all the irreps are one-dimensional. They are:

$V = \mathbb{C}$ and $\rho_m(\omega) = \omega^m$ where m is an integer. Note that $m \sim m + n$ so the set of irreps is again labeled by $\mathbb{Z}/n\mathbb{Z}$ and in fact, under tensor product the set of irreps itself forms a group isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

The matrix elements in the irrep (ρ_m, V) are

♣These were called ρ_m previously.... ♣

$$M^{(m)}(\omega^j) = \omega^{mj} = e^{2\pi \mathrm{i} \frac{mj}{n}} \tag{11.167}$$

Now we can check that indeed

$$\frac{1}{|G|} \sum_{g \in G} (M^{(m_1)}(g))^* M^{(m_2)}(g) = \delta_{m_1 - m_2 = 0 \mathrm{mod} n} \tag{11.168}$$

The decomposition of a function Ψ on the group G is known as the discrete Fourier transform: If $\Psi : \mathbb{Z}_n \to \mathbb{C}$ is any function we can write it as

$$\Psi = \sum_m \hat{\Psi}_m M^{(m)} \tag{11.169}$$

$$\hat{\Psi}_m = \int_{\mathbb{Z}_n} (M^{(m)}(g))^* \Psi(g) dg \tag{11.170}$$

Example 3: The theorem applies to all compact Lie groups. For example, when $G = U(1) = \{z || z| = 1\}$ then the invariant measure on the group is just $-\mathrm{i}\frac{dz}{z} = \frac{d\theta}{2\pi}$ where $z = e^{\mathrm{i}\theta}$:

$$\langle \Psi_1, \Psi_2 \rangle = \int_0^{2\pi} \frac{d\theta}{2\pi} (\Psi_1(\theta))^* \Psi_2(\theta) \tag{11.171}$$

Now, again since G is abelian the irreducible representations are 1-dimensional and the unitary representations are (ρ_n, V_n) where $n \in \mathbb{Z}$, $V_n \cong \mathbb{C}$ and

$$\rho_n(z) := z^n \tag{11.172}$$

Now, the orthonormality of the matrix elements is the standard orthonormality of $e^{in\theta}$ and the Peter-Weyl theorem specializes to Fourier analysis: An L^2-function $\Psi(\theta)$ on the circle can be expanded in terms of the matrix elements of the irreps:

$$\Psi = \sum_{\mathrm{Irreps}\ \ \rho_n} \hat{\Psi}_n M^{(n)} \tag{11.173}$$

This is just the standard Fourier decomposition of a periodic function.

We stress that the Peter-Weyl theorem applies to all compact groups, not just Abelian ones. For this reason it is an important defining subject of the subject of "nonabelian Fourier analysis." Here is a simple nonabelian example:

Example 4: So far all our examples have been Abelian groups. Let us consider $G = S_3$. It has order 6 so that $L^2(G)$ is a six-dimensional vector space. So far, we have discussed three different irreps of dimensions $1, 1, 2$. Are there any others? Note that $6 = 1^2 + 1^2 + 2^2$. So we conclude that there are no other irreps. The matrix elements are:

1. *The trivial representation*: $M^+(g) = 1$ for all $g \in S_3$

2. *The sign representation*:

$$\begin{aligned} M^-(1) = 1 \qquad & M^-(123) = M^-(132) = 1 \\ M^-(12) \qquad & = M^-(13) = M^-(23) = -1 \end{aligned} \tag{11.174}$$

3. *The two-dimensional representation.* See equation (11.94) et. seq. So we have

$$\begin{aligned} M^2_{11}(1) = M^2_{11}((12)) &= 1 \\ M^2_{11}((13)) = M^2_{11}((23)) &= -1/2 \\ M^2_{11}((123)) = M^2_{11}((123)) &= -1/2 \end{aligned} \tag{11.175}$$

and so on for the other 3 matrix elements in the 2-dimensional representation. The reader should check that the function M^2_{11} is orthogonal to the functions M^+ and M^- and that $\sqrt{2}M^2_{11}$ has unit norm.

Remark: This leaves the separate question of actually *constructing* the representations of the finite group G. Until recently there has been no general algorithm for doing this. Recently there has been a claim that it can be done. See

Vahid Dabbaghian-Abdoly, "An Algorithm for Constructing Representations of Finite Groups." Journal of Symbolic Computation Volume 39, Issue 6, June 2005, Pages 671-688

Exercise *Due Diligence*
Check the orthogonality relations for the other matrix elements D^2_{ij} on S_3.

Exercise *How Many One Dimensional Represenations Does An Arbitrary Finite Group Have?*

Show that the number of distinct one-dimensional representations of a finite group G is the same as the index of the commutator subgroup $[G, G]$ in G. [109]

Exercise *Projection Operators*

Let (T, V) be any representation of a compact group G. Consider the linear operators on V:

$$P^{(\mu)}_{ij} := \int_G (\phi^{(\mu)}_{ij}(g))^* T(g) dg \in \mathrm{End}(V) \tag{11.176}$$

a.) Show that

$$P^{(\mu)}_{ij} P^{(\nu)}_{kl} = \delta_{\mu\nu}\delta_{j,k} P^{(\nu)}_{i\ell} \tag{11.177}$$

[109] Hint: A one-dimensional representation is trivial on $[G, G]$ and hence descends to a representation of the abelianization of G, namely the quotient group $G/[G, G]$.

b.) Show that, for any vector $\psi \in V$ we have

$$T(h)P^{\mu}_{ij}\psi = \sum_{k=1}^{n_\mu} M^{\mu}_{ki}(h)P^{\mu}_{kj}\psi \tag{11.178}$$

Conclude that, if the $P^{\mu}_{ij}\psi \neq 0$ then, for fixed μ, j the span

$$Span\{P^{\mu}_{ij}\psi | i = 1, \ldots, n_\mu\} \tag{11.179}$$

is a subspace of V transforming in the representation (V^μ, T^μ).

c.) In particular show that the projector onto the isotypical component corresponding to the trivial representation is:

$$P = \int_G T(g)dg \tag{11.180}$$

We will discuss this more later, giving examples of the decomposition of a reducible representation into its irreps.

11.10 Orthogonality Relations For Characters And Character Tables

An important simplification of the orthogonality relations is the corresponding relations for the character functions of the irreps: If we take traces by setting $i = j$ and $k = \ell$ and summing in equation (11.157) we get the orthogonality relations of characters:

$$\int_G (\chi^{(\mu)}(g))^* \chi^{(\nu)}(g)dg = \delta_{\mu,\nu} \tag{11.181}$$

This beautiful fact is extremely powerful. For example, as we have seen, any representation (T, V) of G is completely reducible so

$$V \cong \oplus_\mu a_\mu V^{(\mu)} \tag{11.182}$$

with $a_\mu \in \mathbb{Z}_+$. But then we can determine these degeneracies by

$$a_\mu = \int_G (\chi^{(\mu)}(g))^* \chi_V(g)dg = \langle \chi^{(\mu)}, \chi_V \rangle \tag{11.183}$$

Therefore, we conclude the extremely important fact:

A representation of a compact group is completely determined, up to equivalence, by its character function!

Moreover, by using the characters we can easily determine the decomposition into isotypical components.

Recall that a class function on G is a function $f : G \to \mathbb{C}$ whose value only depends on the conjugacy class: $f(g) = f(hgh^{-1})$ for all $g, h \in G$. The class functions form a subspace $L^2(G)^{class} \subset L^2(G)$. The characters of the irreps provide an ON basis:

Theorem. $\{\chi^\mu\}$ is an ON basis for the vector space of class functions $L^2(G)^{class}$.

Proof: This is a corollary of the orthogonality relations for matrix elements. Any function $f \in L^2(G)$ be expanded

$$f(g) = \sum_{\mu,i,j} \hat{f}^{\mu}_{ij} M^{\mu}_{ij}(g) \tag{11.184}$$

Now, if f is a class function then when we compute the integral:

$$\tilde{f}(g) := \int_G f(hgh^{-1}) dh \tag{11.185}$$

we of course get $\tilde{f} = f$, provided we normalized the group to have volume 1.

On the other hand,

$$\begin{aligned} \tilde{f}(g) &= \int_G f(hgh^{-1}) dh \\ &= \sum_{\mu,i,j} \hat{f}^{\mu}_{ij} \int_G M^{\mu}_{ij}(hgh^{-1}) dh \\ &= \sum_{\mu,i,j} \hat{f}^{\mu}_{ij} M^{\mu}_{kl}(g) \int_G M^{\mu}_{ik}(h) M^{\mu}_{lj}(h^{-1}) dh \\ &= \sum_{\mu,i} \frac{\hat{f}^{\mu}_{ii}}{n_\mu} \chi^{\mu}(g) \end{aligned} \tag{11.186}$$

We already knew that the χ^μ were ON and now we see that they span the subpsace of class functions. ♠

11.10.1 Finite Groups And The Character Table

In the case of a finite group there is another obvious basis of the space of class functions: Denote the distinct conjugacy classes by $C_i, i = 1, \ldots r$. For each C_i we can define a class function δ_{C_i} in $L^2(G)$ to be the characteristic function of C_i. In equations:

$$\delta_{C_i}(g) := \begin{cases} 1 & g \in C_i \\ 0 & else \end{cases} \tag{11.187}$$

Since any class function takes the same value for all group elements g in a fixed conjugacy class it follows that the functions δ_{C_i} form a basis for the space of class functions. Any two bases for a given vector space have the same cardinality so we conclude:

Theorem The number of conjugacy classes of G is the same as the number of irreducible representations of G.

Since the number of representations and conjugacy classes are the same we can define an $r \times r$ array known as a *character table*:

$$\begin{array}{cccccccc} & m_1 C_1 & m_2 C_2 & \cdots & \cdots & m_r C_r \\ \chi_1 & \chi_1(C_1) & \cdots & \cdots & \cdots & \chi_1(C_r) \\ \chi_2 & \chi_2(C_1) & \cdots & \cdots & \cdots & \chi_2(C_r) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \chi_r & \chi_r(C_1) & \cdots & \cdots & \cdots & \chi_r(C_r) \end{array} \tag{11.188}$$

Here m_i denotes the order of C_i.

Generally speaking, most of what you typically want to know about a group is contained in its character table.

For a finite group we can rewrite the orthogonality relations of the characters, equation (11.181), more explicitly as:

$$\frac{1}{|G|}\sum_{C_i\in\mathcal{C}} m_i \chi_\mu(C_i)\chi_\nu(C_i)^* = \delta_{\mu\nu} \tag{11.189}$$

where $m_i = |C_i|$ is the order of the conjugacy class C_i.

We claim that there is a "dual" orthogonality relation where we sum over irreps rather than conjugacy classes

$$\sum_\mu \chi_\mu(C_i)^*\chi_\mu(C_j) = \frac{|G|}{m_i}\delta_{ij} \tag{11.190}$$

The proof of (11.190) is very elegant: Note that equation (11.189) can be interpreted as the statement that the $r\times r$ matrix

$$S_{\mu i} := \sqrt{\frac{m_i}{|G|}}\chi_\mu(C_i) \qquad \mu = 1,\dots,r \quad i=1,\dots,r \tag{11.191}$$

satisfies

$$\sum_{i=1}^{r} S_{\mu i}S^*_{\nu i} = \delta_{\mu\nu} \tag{11.192}$$

Therefore, $S_{\mu i}$ is a unitary matrix. The left-inverse is the same as the right-inverse, and hence we obtain (11.190).

Example 1: For the simplest group there are two conjugacy classes $\mathcal{C}_1 = [1]$ and $\mathcal{C}_2 = [(12)]$. They both have cardinality $m_i = 1$. There are two irreps $\mathbf{1}^+$ and the sign representation $\mathbf{1}^-$. We thus have the character table

	$[1]$	$[(12)]$
1^+	1	1
1^-	1	-1

It is now straightforward to write out the character table for S_3.

	[1]	3[(12)]	2[(123)]
$\mathbf{1}^+$	1	1	1
$\mathbf{1}^-$	1	-1	1
$\mathbf{2}$	2	0	-1

Here is an example of the use of characters to decompose a representation:

Now let us see how we can use the orthogonality relations on characters to find the decomposition of a reducible representation.

Example 1 Consider the 3×3 rep generated by the $\natural$ action of the permutation group S_3 on $\mathbb{R}^3$. We'll compute the characters by choosing one representative from each conjugacy class:

$$1 \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (12) \to \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (132) \to \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{11.193}$$

From these representatives the character of $V = \mathbb{R}^3$ is easily calculated:

$$\chi_V(1) = 3 \qquad \chi_V([(12)]) = 1 \qquad \chi_V([(132)] = 0 \tag{11.194}$$

Using the orthogonality relations we compute

$$a_{\mathbf{1}_+} = (\chi_{\mathbf{1}_+}, \chi) = \frac{1}{6}3 + \frac{3}{6}1 + \frac{2}{6}0 = 1 \tag{11.195}$$

$$a_{\mathbf{1}_-} = (\chi_{\mathbf{1}_-}, \chi) = \frac{1}{6}3 + \frac{3}{6}(-1) \cdot 1 + \frac{2}{6}0 = 0 \tag{11.196}$$

$$a_{\mathbf{2}} = (\chi_{\mathbf{2}}, \chi) = \frac{1}{6}3 \cdot 2 + \frac{3}{6}0 \cdot 1 + \frac{2}{6}(-1) \cdot 0 = 1 \tag{11.197}$$

Therefore:

$$\chi_V = \chi_{\mathbf{1}_+} + \chi_{\mathbf{2}} \tag{11.198}$$

showing the decomposition of $\mathbb{R}^3$ into irreps, and confirming what we showed above.

Example 2: Let V be any vector space. Consider natural permutation $V \otimes V$ by S_2. where V is a finite dimensional vector space (over any field) of dimension d. Let us write out the isotypical decomposition under the S_2 action. The character is easily computed:

$$\begin{aligned} \chi_{V^{\otimes 2}}(1) &= d^2 \\ \chi_{V^{\otimes 2}}((12)) &= d \end{aligned} \tag{11.199}$$

To check the second line we choose a basis $\{v_i\}$ for V so that $\{v_i \otimes v_j\}$ is a basis for $V \otimes V$. In this basis we have $(12) \cdot v_i \otimes v_j = v_j \otimes v_i$ and hence only the basis elements $v_i \otimes v_i$ contribute to the trace. Therefore, we compute the degeneracies of the trivial and sign representation from:

$$\begin{aligned} a_{1^+} &= \langle \chi^+, \chi_{V^{\otimes 2}} \rangle = \frac{1}{2}(\chi_{V^{\otimes 2}}(1) + \chi_{V^{\otimes 2}}((12))) \\ &= \frac{1}{2}d(d+1) \\ a_{1^-} &= \langle \chi^-, \chi_{V^{\otimes 2}} \rangle = \frac{1}{2}(\chi_{V^{\otimes 2}}(1) - \chi_{V^{\otimes 2}}((12))) \\ &= \frac{1}{2}d(d-1) \end{aligned} \tag{11.200}$$

so we have isotypical decomposition:

$$V^{\otimes 2} = \frac{1}{2}d(d+1)\mathbf{1}^+ \oplus \frac{1}{2}d(d-1)\mathbf{1}^- \tag{11.201}$$

General elements of $V^{\otimes 2}$ are of the form $T_{ij}v_i \otimes v_j$ and the components are often referred to as 2-index covariant tensors. The above decomposition is a decomposition into *symmetry types* of tensors: We can choose a basis of symmetric tensors:

$$\frac{1}{2}(e_i \otimes e_j + e_j \otimes e_i) \qquad\qquad 1 \leq i \leq j \leq d \tag{11.202}$$

and anti-symmetric tensors

$$\frac{1}{2}(e_i \otimes e_j - e_j \otimes e_i) \qquad\qquad 1 \leq i < j \leq d \tag{11.203}$$

Example 3: The previous example is the beginning of a very beautiful story called Schur-Weyl duality. Let us consider the next case: Consider S_3 acting by permuting the various factors in the tensor space $V \otimes V \otimes V$ for any vector space V. Now, if $\dim V = d$ then we have

$$\begin{aligned} \chi([1]) &= d^3 \\ \chi([(ab)]) &= d^2 \\ \chi[(abc)] &= d \end{aligned} \tag{11.204}$$

as is easily computed by considering the action on the basis $\{v_i \otimes v_j \otimes v_k\}$.

So we can compute

$$a_{\mathbf{1}_+} = (\chi_{\mathbf{1}_+}, \chi) = \frac{1}{6}d^3 + \frac{3}{6}d^2 + \frac{2}{6}d = \frac{1}{6}d(d+1)(d+2) \tag{11.205}$$

$$a_{\mathbf{1}_-} = (\chi_{\mathbf{1}_-}, \chi) = \frac{1}{6}d^3 + \frac{3}{6}(-1) \cdot d^2 + \frac{2}{6}d = \frac{1}{6}d(d-1)(d-2) \tag{11.206}$$

$$a_{\mathbf{2}} = (\chi_{\mathbf{2}}, \chi) = \frac{1}{6}2d^3 + \frac{3}{6}0 \cdot d^2 + \frac{2}{6}(-1) \cdot d = \frac{1}{3}d(d^2-1) \tag{11.207}$$

Thus, *as a representation of* S_3, we have

$$V^{\otimes 3} \cong \frac{d(d+1)(d+2)}{6}\mathbf{1}_+ \oplus \frac{d(d-1)(d-2)}{6}\mathbf{1}_- \oplus \frac{d(d+1)(d-1)}{3}\mathbf{2} \tag{11.208}$$

Note that the first two dimensions are those of S^3V and $\Lambda^3 V$, respectively, and that the dimensions add up correctly. It is not supposed to be obvious, but it turns out that the last summand corresponds to tensor of mixed symmetry type that satisfy

$$T_{ijk} + T_{jki} + T_{kij} = 0 \tag{11.209}$$

together with

$$T_{ijk} = -T_{kji} \tag{11.210}$$

Why this is so, and the generalization to the S_n action on $V^{\otimes n}$ is best discussed in the context of Young diagrams, representation theory of S_n and Schur-Weyl duality. See section *****

Exercise

Show that right-multiplication of the character table by a diagonal matrix produces a unitary matrix.

Exercise

Using the orthogonality relations on matrix elements, derive the more general relation on characters:

$$\frac{1}{|G|}\sum_{g\in G}\chi_\mu(g)\chi_\nu(g^{-1}h) = \frac{\delta_{\mu\nu}}{n_\mu}\chi^{(\nu)}(h) \tag{11.211}$$

We will interpret this more conceptually later.

Exercise *Another proof of* (11.190)

Consider the operator $L(g_1)\otimes R(g_2)$ acting on the regular representation. We will compute

$$\mathrm{Tr}_{R_G}[L(g_1)\otimes R(g_2)] \tag{11.212}$$

in two bases.

First consider the basis ϕ^μ_{ij} of matrix elements. We have:

$$L(g_1)\otimes R(g_2)\cdot\phi^\mu_{ij} = \sum_{i',j'} T^\mu_{j'j}(g_2)T^\mu_{i'i}(g_1^{-1})\phi^\mu_{i'j'} \tag{11.213}$$

so the trace in this basis is just:

$$\mathrm{Tr}_{R_G}[L(g_1) \otimes R(g_2)] = \sum_\mu \chi_\mu(g_1)^* \chi_\mu(g_2) \tag{11.214}$$

On the other hand, we can use the delta-function basis: δ_g. Note that

$$L(g_1) \otimes R(g_2) \cdot \delta_g = \delta_{g_1 g g_2^{-1}} \tag{11.215}$$

So in the delta function basis we get a contribution of $+1$ to the trace iff $g = g_1 g g_2^{-1}$, that is iff $g_2 = g^{-1} g_1 g$, that is iff g_1 and g_2 are conjugate, otherwise we get zero.

Now when g_1 and g_2 are conjugate we might as well take them to be equal. Then the functions δ_g contributing to the trace are precisely those with $g \in Z(g_1)$. But then

$$|Z(g_1)| = \frac{|G|}{|C(g_1)|} = \frac{|G|}{m_1} \tag{11.216}$$

completing the proof of (11.190).

Exercise *Average Number Of Fixed Points - Again*

Recall that in a previous exercise you showed that if a finite group G acts on a finite set X then

$$\frac{1}{|G|} \sum_g |X^g| = |\{orbits\}| \tag{11.217}$$

Prove this again by viewing $L^2(X)$ as a G-representation and using the orthgonality relations on characters. [110]

Exercise

Write out the unitary matrix $S_{\mu i}$ for $G = S_3$. [111]

[110] *Answer*: Note that one basis for $L^2(X)$ is given by δ_x. Here $g \cdot \delta_x = \delta_{g \cdot x}$ so $|X^g|$ is just the character of g in this representation. Therefore the average number of fixed points is the degeneracy of the trivial representation in the isotypical decomposition. On the other hand, if $f : X \to \mathbb{C}$ is invariant under the g action then f takes the same value on any two x_1, x_2 in the same G-orbit. Therefore, the subspace corresponding to the trivial representation in the isotypical decomposition has a basis given by the characteristic function for each orbit.

[111] *Answer*: The unitary matrix $S_{\mu i}$ for S_3 is:

$$S_{\mu i} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{pmatrix} \tag{11.218}$$

Exercise

a.) Suppose we tried to define a representation of S_3 by taking $(12) \to 1$ and $(23) \to -1$. What goes wrong?

b.) Show that for any n there are only two one-dimensional representations of S_n

♣this is really about generators and relations and should be moved up. ♣

Exercise *Universal Properties Of Character Tables*

Show that if we partially order the conjugacy classes and the representations so that the class of the identity and the trivial representation come first then the first row of the character table is all 1's and the first column of the character table gives the dimensions of the irreducible representations.

11.10.2 Orthogonality Relations And Pontryagin Duality

It is interesting to think about the orthogonality of characters in the context of Pontryagin duality. Let S be a locally compact Abelian group and $\widehat{S}$ its Pontryagin dual.

Compact groups have discrete sets of irreducible representations, but the representations of noncompact groups typically come in continuous families. Thus, for example, the Pontryagin dual of $\mathbb{Z}$ is the continuous group $U(1)$. The orthogonality relations for locally compact Abelian groups extend to this case, but the $\delta_{\mu,\nu}$ delta function on the representations becomes a delta function $\delta_{\hat{S}}(\chi_1 - \chi_2)$ on the characters, where the Dirac measure is relative to the Haar measure on $\hat{S}$. Thus we have:

$$\int_S \chi_1(s)^* \chi_2(s) ds = \delta_{\widehat{S}}(\chi_1 - \chi_2) \tag{11.219}$$

Now, by Pontryagin duality, we have $\widehat{\widehat{S}} \cong S$ so we also have the relation

$$\langle \psi_1, \psi_2 \rangle_{L^2(S)} = \langle \hat{\psi}_1, \hat{\psi}_2 \rangle_{L^2(\widehat{S})} \tag{11.220}$$

Therefore, given a function $\psi \in L^2(S)$, we can define its Fourier transform, quite generally, as

$$\hat{\psi}(\chi) := \int_S \chi(s)^* \psi(s) ds \tag{11.221}$$

and the orthogonality relations above show that

$$\int_{\widehat{S}} \chi(s_1)^* \chi(s_2) d\chi = \delta_S(s_1 - s_2) \tag{11.222}$$

This is a result known as either the *Plancherel theorem* or the *Parseval theorem.* [112]

An important special case of the above is the case $S = \Gamma \subset \mathbb{R}^n$, an embedded lattice. Then we have:

$$\sum_{\gamma\in\Gamma} \chi_{\bar{k}_1}(\gamma)^* \chi_{\bar{k}_2}(\gamma) = \delta_{\hat{\Gamma}}(\bar{k}_1 - \bar{k}_2) \tag{11.223}$$

Here $\delta_{\hat{\Gamma}}(\bar{k})$ is the delta measure on the dual group

$$\hat{\Gamma} = \mathbb{R}^n/\Gamma^\vee \tag{11.224}$$

We can lift $\bar{k}$ to $k \in \mathbb{R}^n$, and then

$$\delta_{\hat{\Gamma}}(\bar{k}_1 - \bar{k}_2) = \sum_{\gamma^\vee\in\mathbb{R}^n} \delta_{\mathbb{R}^n}(k_1 - k_2 - \gamma^\vee) \tag{11.225}$$

On the other hand, using the explicit formula:

$$\chi_{\bar{k}}(\gamma) = e^{2\pi \mathrm{i} k\cdot\gamma} \tag{11.226}$$

we get the relation

$$\sum_{\gamma\in\Gamma} e^{2\pi\mathrm{i}(k_2-k_1)\cdot\gamma} = \sum_{\gamma^\vee\in\Gamma^\vee} \delta_{\mathbb{R}^n}(k_2 - k_1 - \gamma^\vee) \tag{11.227}$$

This is one version of the *Poisson summation formula.* Since the PSF is an important result we will unpack this a bit in the next section.

11.10.3 The Poisson Summation Formula

Let us begin with a standard derivation of the very useful Poisson summation formula:

Let $f : \mathbb{R} \to \mathbb{C}$ that decays fast enough that

$$F(x) := \sum_{n\in\mathbb{Z}} f(x+n) \tag{11.228}$$

exists. Then $F(x)$ is clearly periodic of period one and therefore defines a function $F : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$. Since $\mathbb{R}/\mathbb{Z} \cong U(1)$ we can decompose in terms of irreducible representations:

$$F(x) = \sum_{\rho_m} \hat{F}(\rho_m) e^{2\pi\mathrm{i}mx} \tag{11.229}$$

$$\hat{F}(\rho_m) = \int_0^1 e^{-2\pi\mathrm{i}mt} F(t)dt \tag{11.230}$$

Now note that

$$\begin{aligned} \int_0^1 e^{-2\pi\mathrm{i}mt} F(t)dt &= \int_0^1 e^{-2\pi\mathrm{i}mt} \sum_m f(m+t)dt \\ &= \int_{-\infty}^{+\infty} e^{-2\pi\mathrm{i}mt} f(t)dt \end{aligned} \tag{11.231}$$

[112] The original statements by Plancherel and Parseval concerned special cases and the two terms are not consistently used in the literature.

Putting $x = 0$ we learn that for suitably rapidly decaying functions:

$$\sum_{m\in\mathbb{Z}} f(m) = \sum_{w\in\mathbb{Z}} \widehat{f}(w) \tag{11.232}$$

where $\widehat{f}$ is the Fourier transform:

$$\widehat{f}(w) = \int_{\mathbb{R}} e^{-2\pi i t w} f(t) \tag{11.233}$$

This is valid for functions such that

1. f decays rapidly enough so that the sum on the LHS converges.
2. The Fourier transform $\widehat{f}$ exists.
3. The Fourier transform $\widehat{f}$ decays rapidly enough so that the sum on the RHS converges.

Note that this can also be understood as a special case of the orthogonality relations for characters on $\mathbb{Z}$:

$$\boxed{\sum_{n\epsilon\mathbb{Z}} e^{2\pi i n t} e^{-2\pi i n t'} = \delta_{\mathbb{R}/\mathbb{Z}}(t-t') = \sum_{k\in\mathbb{Z}} \delta_{\mathbb{R}}(t-t'-k)} \tag{11.234}$$

where $\delta_{\mathbb{Z}}$ means the delta function on the Pontryagin dual group $\mathbb{Z}$.

The generalization of this statement is our result (11.227) above.

Put differently, we have:

$$\sum_{\vec{n}\in\mathbb{Z}^d} f(\vec{n}) = \sum_{\vec{m}\in\mathbb{Z}^d} \int_{\mathbb{R}^d} e^{-2\pi i \vec{m}\cdot\vec{t}} f(t) dt = \sum_{\vec{m}\in\mathbb{Z}^d} \int_{R^d} e^{+2\pi i \vec{m}\cdot\vec{t}} f(t) dt \tag{11.235}$$

That is:

$$\sum_{\vec{v}\in\Gamma} f(\vec{v}) = \sum_{\vec{l}\in\Gamma^\vee} \hat{f}(\vec{l}) \ . \tag{11.236}$$

Remarks

1. Since people have different conventions for the factors of 2π in Fourier transforms it is hard to remember the factors of 2π in the PSF. The equation (11.234) has no factors of 2π. One easy way to see this is to integrate both sides from $t = -1/2$ to $t = +1/2$.

2. One application of this is the x-ray crystallography: The LHS is the sum of scattered waves. The RHS constitutes the bright peaks measured on a photographic plate.

3. Another application is to analytic number theory. If τ is in the upper half complex plane, and θ, ϕ, z are complex numbers define the *Riemann theta function with characteristics* θ, ϕ

$$\vartheta[\begin{matrix}\theta\\ \phi\end{matrix}](z|\tau) := \sum_{n\in\mathbb{Z}} e^{i\pi\tau(n+\theta)^2+2\pi i(n+\theta)(z+\phi)} \tag{11.237}$$

(usually, θ, ϕ are taken to be real numbers, but z is complex). This converges to an entire function of z and is also holomorphic for $\mathrm{Im}\tau > 0$.

Using the Poisson summation formula one can show that it obeys the modular transformation law:

$$\vartheta[\begin{matrix}\theta\\ \phi\end{matrix}](\frac{-z}{\tau}|\frac{-1}{\tau}) = (-i\tau)^{1/2} e^{2\pi i\theta\phi} e^{i\pi z^2/\tau} \vartheta[\begin{matrix}-\phi\\ \theta\end{matrix}](z|\tau) \tag{11.238}$$

Exercise

a.) Show that

$$\sum_{n\in Z} e^{-\pi a n^2+2\pi i b n} = \sqrt{\frac{1}{a}} \sum_{m\in Z} e^{-\frac{\pi(m-b)^2}{a}} \tag{11.239}$$

b.) Check equation (11.238).

11.11 The Finite Heisenberg Group And The Quantum Mechanics Of A Particle On A Discrete Approximation To A Circle

****

SOME OF THE MATERIAL IN THIS SUBSECTION IS NOW REDUNDANT WITH MATERIAL ABOVE AND NEEDS TO BE REMOVED

****

It is very illuminating to interpret the group Heis_N in terms of the quantum mechanics of a particle on a discrete approximation to a circle.

♣This sub-section assumes some knowledge of linear algebra and quantum mechanics explained in chapter 2. ♣

Recall that if G acts on a set X then it acts on the functions $\mathcal{F}[X \to Y]$ for any Y. Moreover, G always acts on itself by left-translation. Let us apply this general idea to $G = \mathbb{Z}_N$, thought of as the Nth roots of unity and $Y = \mathbb{C}$, the complex numbers. So we are studying complex-valued functions on the group G. We can picture the group as a discrete set of points on the unit circle so we can think, physically, of $\mathcal{F}[X \to Y]$ as the space of wavefunctions of a particle moving on a discrete approximation to a circle.

Now, as a vector space it is clear that $\mathcal{F}[\mathbb{Z}_N \to \mathbb{C}]$ is isomorphic to $\mathbb{C}^N$. To specify a function is to specify the N different complex values $\Psi(\omega^k)$ where ω is a primitive N^{th} root of one, say, $\omega = \exp[2\pi\mathrm{i}/N]$ for definiteness, and $k = 0, \ldots, N-1$. (We will not try to normalize our wavefunctions Ψ, but we could. It would make no difference to the present considerations.)

Another way to see we have an isomorphism is to choose a natural basis, the delta-function basis:

$$\delta_j(\omega^k) = \delta_{\bar{j},\bar{k}} \tag{11.240}$$

where $\bar{j}, \bar{k} \in \mathbb{Z}/N\mathbb{Z}$, viewed additively. So our isomorphism is $\delta_j \mapsto \vec{e}_j$. Put differently, every wavefunction can be uniquely expressed as

$$\Psi = \sum_{j=0}^{N-1} z_j \delta_j \tag{11.241}$$

where $z_j \in \mathbb{C}$. Indeed $z_j = \Psi(\omega^j)$.

In fact, we can make $\mathcal{H} = \mathcal{F}[\mathbb{Z}_N \to \mathbb{C}]$ into a Hilbert space in a natural way by declaring that the inner product is:

$$\langle \Psi_1, \Psi_2 \rangle := \frac{1}{|G|} \sum_{g \in G} \Psi_1^*(g) \Psi_2(g) \tag{11.242}$$

Note that with this inner product the basis δ_j, $j = 0, \ldots, N-1$, to be an orthonormal basis of $\mathcal{H}$. Note that

Note that the sum on the RHS of (11.242) defines a measure on the group: If $F : G \to \mathbb{C}$ we define its integral

$$\int_G F d\mu := \frac{1}{|G|} \sum_{g \in G} F(g) \tag{11.243}$$

and we have normalized the measure so that the group has "volume 1." This is an example of an important idea called a *Haar measure* that we will discuss more later.

Now, recall the general definition from section 4.1. G acts naturally on X (which happens to be G itself) by left-multiplication. The induced action of G on the complex-valued functions on G in this case is such that the generator ω of $\mathbb{Z}_N$ acts on the space of functions via:

$$\begin{aligned} \tilde{\phi}(\omega, \Psi)(\omega^k) &:= \Psi(\phi(\omega^{-1}, \omega^k)) \\ &= \Psi(\omega^{k-1}) \end{aligned} \tag{11.244}$$

So the generator ω of the group $\mathbb{Z}_N$ acts linearly on the functions $\mathcal{F}[\mathbb{Z}_N \to \mathbb{C}]$. We call this linear operator P. We can therefore rewrite (11.244) as

$$(P \cdot \Psi)(\omega^k) := \Psi(\omega^{k-1}) \tag{11.245}$$

Note that with respect to the inner product (11.242) P is clearly a unitary operator.

The operator P can be viewed as translation operator around the discrete circle by one step in the clockwise direction. Recall that in the quantum mechanics of a particle on the line translation by a distance a is

$$(T(a) \cdot \Psi)(x) = \Psi(x - a) \tag{11.246}$$

This equation makes sense also for a particle on the circle, that is, with x, a considered periodically. So our P is $T(a)$ for translation by $2\pi/N$ times around the circle clockwise.

Remark: In the quantum mechanics of a particle on a line or circle we could also write

$$(T(a)\cdot\Psi)(x) = \Psi(x-a) = (\exp[\mathrm{i}a\hat{p}])\Psi(x) = (\exp[-a\frac{d}{dx}]\cdot\Psi)(x) \tag{11.247}$$

so the momentum operator $\hat{p}$ generates translations. In the finite Heisenberg group there is no analog of the infinitesimal translations generated by $\hat{p}$, but only of a finite set of discrete translations.

Now let Q be the position operator:

$$(Q\cdot\Psi)(\omega^k) := \omega^k\Psi(\omega^k) \tag{11.248}$$

Q is likewise a unitary operator.

Now note that

$$\begin{aligned}(P\circ Q\cdot\Psi)(\omega^k) &= (Q\cdot\Psi)(\omega^{k-1})\\ &= \omega^{k-1}\Psi(\omega^{k-1})\end{aligned} \tag{11.249}$$

while

$$\begin{aligned}(Q\circ P\cdot\Psi)(\omega^k) &= \omega^k(P\cdot\Psi)(\omega^k)\\ &= \omega^k\Psi(\omega^{k-1})\end{aligned} \tag{11.250}$$

and therefore we conclude that we have the operator equation:

$$Q\circ P = \omega P\circ Q \tag{11.251}$$

Given a linear transformation and ordered bases for domain and range we can associate a matrix. (See Chapter two for details if you do not know this.) Now, let us choose the ordered basis $\delta_1,\ldots,\delta_N$. Then we easily compute

$$P\cdot\delta_j = \delta_{j+1} \tag{11.252}$$

and therefore the matrix for P relative to the basis $\{\delta_j\}$, is the matrix with matrix elements

$$P_{i,j} = \delta_{i,j+1} \tag{11.253}$$

so, for $N=3$ it is

$$P = \begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix} \tag{11.254}$$

Similarly, in the basis δ_j we have

$$Q_{i,j} = \omega^j\delta_{i,j} \tag{11.255}$$

and since $j=0,1,\ldots,N-1$ we have for $N=3$:

$$Q = \begin{pmatrix}1&0&0\\0&\omega&0\\0&0&\omega^2\end{pmatrix} \tag{11.256}$$

Thus, we have recovered the $N \times N$ clock and shift matrices we discussed above. The group of unitary operators generated by discrete position and translation operators is the finite Heisenberg group.

It is interesting to study the operators P and Q in a different basis. We can introduce a "plane wave basis" of functions $\Psi_j \in \mathcal{H}$, with $j = 0, \ldots, N-1$ defined by

$$\Psi_j(\omega^k) = \omega^{jk} \tag{11.257}$$

It is now easy to compute the action of P and Q on this basis:

$$\begin{aligned} P \cdot \Psi_j &= \omega^{-j} \Psi_j \\ Q \cdot \Psi_j &= \Psi_{j+1} \end{aligned} \tag{11.258}$$

The roles of P and Q have been exchanged! P (as is P^{-1}) is now representated by "clock matrix" and Q is represented by a "shift matrix". Indeed, what we have done is perform a transformation from a position representation to a momentum representation in the language of quantum mechanics.

The basis Ψ_j represents a very general and beautiful fact: The Ψ_j are actually group homomorphisms $\Psi_j : \mathbb{Z}_N \to U(1) \subset GL(1, \mathbb{C})$ because

$$\Psi_j(\omega^{k_1}\omega^{k_2}) = \Psi_j(\omega^{k_1}) \cdot \Psi_j(\omega^{k_2}) \tag{11.259}$$

as is easily checked. The one-dimensional vector spaces spanned by the individual Ψ_j decompose the Hilbert space $L^2(\mathbb{Z}_N)$, which is an N-dimensional representation of $\mathbb{Z}_N$ into a direct sum of 1-dimensional (irreducible) representations. [113]

♣The next several paragraphs are now redundant with material in section 11.8 et. seq. You might want to leave this out and bring back the example in those later sections. ♣

In general, for compact Lie groups there is a left action of G on $L^2(G)$ and, as a representation of G

$$L^2(G) \cong \oplus_j (\dim V_j) V_j \tag{11.260}$$

where V_j runs over the distinct irreducible representations. This is part of an important general theorem about compact Lie groups known as the *Peter-Weyl theorem*.

Moreover, the exchange of clock and shift matrices by passing from the δ_j basis to the basis of characters is again an instance of a general fact:

Let $\widehat{G}$ be the set of homomorphisms $\chi : G \to U(1)$. For an Abelian group G the set $\widehat{G}$ is also a group, using the product law:

$$(\chi_1 \cdot \chi_2)(g) := \chi_1(g)\chi_2(g) \tag{11.261}$$

The reader should be able to check that this defines a group law on $\widehat{G}$. The group $\widehat{G}$ is known as the *Pontryagin dual group*

For reasonable [114] Abelian groups we have

$$\widehat{\widehat{G}} \cong G \tag{11.262}$$

[113] See Chapter four for a detailed discussion of the idea of reducible and irreducible representations. In brief, a representation $\rho : G \to GL(V)$ is *reducible* if there is a nonzero linear subspace $W \subset V$ which is preserved by all the operators $\rho(g)$. Note that one-dimensional representations are trivially irreducible.

[114] e.g. locally compact Abelian groups. See the book by Kirillov on representation theory.

If χ is a homomorphism we can define the Fourier transform

$$\widehat{\Psi}(\chi) := \frac{1}{\sqrt{|G|}} \sum_{g \in G} \chi(g) \Psi(g) \tag{11.263}$$

giving an isometry of $L^2(\widehat{G})$ with $L^2(G)$. Being an isometry means that

$$\langle \widehat{\Psi}_1, \widehat{\Psi}_2 \rangle_{L^2(\widehat{G})} = \langle \Psi_1, \Psi_2 \rangle_{L^2(G)} \tag{11.264}$$

and this in turn implies that

$$\begin{aligned} \frac{1}{|\widehat{G}|} \sum_{\chi \in \widehat{G}} \chi^*(g_1) \chi(g_2) &= \delta_{g_1, g_2} \\ \frac{1}{|G|} \sum_{g \in G} \chi_1^*(g) \chi_2(g) &= \delta_{\chi_1, \chi_2} \end{aligned} \tag{11.265}$$

These equations are special cases of the famous orthogonality relations for the matrix elements of irreducible representations of compact groups.

Now, for $G = \mathbb{Z}_N$ we have $\widehat{G} \cong \mathbb{Z}_N$. The passage from the δ_j basis to the Ψ_j basis, which diagonalizes P is just the *finite Fourier transform*.

In more concrete terms: The trace of all the powers of P less than N is also obviously zero and $P^N = 1$ and no smaller power of P is the identity. So P must be unitarily equivalent to Q. Now we can easily check that

$$SPS^{-1} = Q \tag{11.266}$$

where S is the finite Fourier transform matrix

$$S_{j,k} = \frac{1}{\sqrt{N}} e^{2\pi \mathrm{i} \frac{jk}{N}} \tag{11.267}$$

One easy way to check this is to multiply the matrices SP and QS in the δ_j basis. The reader should check that S is in fact a unitary matrix and that the matrix elements only depend on the projections $\bar{j}, \bar{k} \in \mathbb{Z}/N\mathbb{Z}$.

In any case, $S_{j,k}$ takes us from a position basis δ_j to a "momentum basis" where P is diagonal, in beautiful analogy to how the Fourier transform converts a position basis to a momentum basis for a particle on the line.

We will return to these ideas and discuss Pontryagin duality and Fourier transforms in Chapter 4 below.

Exercise *The Pontryagin Dual Of $\mathbb{Z}_N$ Is Isomorphic To $\mathbb{Z}_N$*

Let $\chi : \mathbb{Z}_N \to U(1)$ be a homomorphism. Let g be a generator of $\mathbb{Z}_N$. Show that $\chi(g)$ must be an Nth root of unity, and choosing any N^{th} root of unity defines the homomorphism. Conclude that $\widehat{\mathbb{Z}_N} \cong \mathbb{Z}_N$.

Exercise *Orthogonality For $\mathbb{Z}_N$*
Write out the equations (11.265) for the case $G = \mathbb{Z}_N$

11.12 Decomposition Of Tensor Products Of Representations And Fusion Coefficients

A frequently asked question is the following: Suppose we know the reduction of two representations of G, say, V_1, V_2 into irreps. What is the decomposition of the tensor product $V_1 \otimes V_2$ into irreps? For compact groups this is nicely answered with characters.

Suppose we have two representations (T_1, V_1) and (T_2, V_2) and we know the isotypical decompositions:

$$V_1 = \oplus a_\mu V^\mu \qquad V_2 = \oplus b_\nu V^\nu \tag{11.268}$$

then

$$V_1 \otimes V_2 = \oplus_{\mu,\nu} a_\mu b_\nu V^\mu \otimes V^\nu \tag{11.269}$$

so the isotypical decomposition of $V_1 \otimes V_2$ follows immediately if we can find the isotypical decomposition of the tensor product of two irreps.

On general grounds we know that ♣$\kappa = \mathbb{C}$ here? ♣

$$V^\mu \otimes V^\nu \cong \oplus_\lambda \mathrm{Hom}_G(V^\lambda, V^\mu \otimes V^\nu) \otimes V^\lambda \tag{11.270}$$

The dimensions

$$N^\lambda_{\mu\nu} := \dim_\kappa \mathrm{Hom}_G(V^\lambda, V^\mu \otimes V^\nu) \tag{11.271}$$

are known as the *fusion coefficients* and they give the isotypical decomposition:

$$\begin{aligned} V^{(\mu)} \otimes V^{(\nu)} &= \oplus_\lambda \Big(\overbrace{V^{(\lambda)} \oplus \cdots \oplus V^{(\lambda)}}^{N^\lambda_{\mu\nu}\ times} \Big) \\ &= \oplus_\lambda N^\lambda_{\mu\nu} V^{(\lambda)} \end{aligned} \tag{11.272}$$

Recall that

$$\chi_{V_1 \otimes V_2}(g) = \chi_{V_1}(g)\chi_{V_2}(g) \tag{11.273}$$

Therefore, taking the trace of (11.272) we get a formula for the fusion coefficients:

$$\chi_\mu(g)\chi_\nu(g) = \sum_\lambda N^\lambda_{\mu\nu}\chi_\lambda(g) \tag{11.274}$$

Now, taking the inner product we get:

$$N^\lambda_{\mu\nu} = \langle \chi_\lambda, \chi_\mu \chi_\nu \rangle \tag{11.275}$$

There is a very beautiful explicit formula for the fusion coefficients in the case of finite groups. In this case we can write:

$$N^\lambda_{\mu\nu} = \frac{1}{|G|} \sum_{g \in G} \chi_\mu(g)\chi_\nu(g)\chi_\lambda(g^{-1}) \tag{11.276}$$

This can be written in a different way. For simplicity choose unitary irreps, and recall that the orthogonality relations on characters is equivalent to the statement that

$$S_{\mu i} := \sqrt{\frac{m_i}{|G|}}\chi_\mu(C_i) \qquad \mu = 1, \ldots, r \quad i = 1, \ldots, r \tag{11.277}$$

is a unitary matrix. Let 1 denote the trivial representation or the trivial conjugacy class. Then $S_{1i} = \sqrt{\frac{m_i}{|G|}}$. Thus we can write

$$N_{\mu\nu}^\lambda = \sum_i \frac{S_{\mu i} S_{\nu i} S_{\lambda i}^*}{S_{1i}} \tag{11.278}$$

This is a prototype of a celebrated result in conformal field theory known as the "Verlinde formula."

Equations (11.275)(11.276)(11.278) give a very handy way to get the numbers $N_{\mu\nu}^\lambda$. Note that, by their very definition the coefficients $N_{\mu\nu}^\lambda$ are nonnegative integers, although this is hardly obvious from, say, (11.278).

Example 1: Consider the irreps ρ_m of $\mathbb{Z}/N\mathbb{Z}$. We then have the fusion rules:

$$\rho_m \otimes \rho_n \cong \rho_{m+n} \tag{11.279}$$

Recall that ρ_m only depends on the integer $m \mathrm{mod} N$. The S-matrix above is the finite Fourier transform matrix:

$$S_{\mu i} = \sqrt{\frac{1}{|G|}} e^{2\pi \mathrm{i} \mu i / N} \tag{11.280}$$

from which one easily verifies (11.278).

Remarks

1. Denote the intertwiners from V^λ to $V^\mu \otimes V^\nu$ by:

   $$\mathcal{V}_{\mu\nu}^\lambda := \mathrm{Hom}_G(V^\lambda, V^\mu \otimes V^\nu) \tag{11.281}$$

   Note that we can decompose the triple product $V^\mu \otimes V^\nu \otimes V^\lambda$ in two natural ways leading to the isomorphism

   $$\oplus_\rho \mathcal{V}_{\mu\nu}^\rho \otimes \mathcal{V}_{\rho\lambda}^\kappa \cong \oplus_\rho \mathcal{V}_{\mu\rho}^\kappa \otimes \mathcal{V}_{\nu\lambda}^\rho \tag{11.282}$$

   If we choose bases for the spaces of intertwiners then the components of the isomorphism relative to these bases are called *fusion matrices* and they satisfy some nice identities.

   ♣Explain pentagon and hexagon? ♣

2. The considerations of this section lead rather naturally to the beautiful subject of Frobenius algebras and 2d topological quantum field theory. See Chapter 4 section **** for more about this.

3. The various identities satisfied by the fusion matrix and the S-matrix have remarkable analogues in the subject of 2d rational conformal field theory. For more about this see:

   a.) G. Moore and N. Seiberg, "Classical and Quantum Conformal Field Theory," Commun. Math. Phys. **123**(1989)177

   b.) G. Moore and N. Seiberg, "Lectures on Rational Conformal Field Theory," in *Strings '89*,Proceedings of the Trieste Spring School on Superstrings, 3-14 April 1989, M. Green, et. al. Eds. World Scientific, 1990. Available on G. Moore's home page.

   c.) Philippe Di Francesco, Pierre Mathieu, David Senechal, *Conformal Field Theory*, Springer

   d.) Jurgen Fuchs, *Affine Lie Algebras And Quantum Groups*

♣Need to add other references ♣

Exercise *Fusion Algebra For S_3*
Let V^+, V^-, V^2 denote the irreps of S_3 of dimensions $1, 1, 2$. Show that

$$\begin{aligned} V^+ \otimes V^\mu &\cong V^\mu \\ V^- \otimes V^- &\cong V^+ \\ V^- \otimes V^2 &\cong V^2 \\ V^2 \otimes V^2 &\cong V^+ \oplus V^- \oplus V^2 \end{aligned} \tag{11.283}$$

Exercise *Identities For Fusion Coefficients*
a.) Show that $N_{\mu\nu}^\lambda = N_{\nu\mu}^\lambda$
b.) Show that $N_{\mu 1}^\lambda = \delta_\mu^\lambda$
c.) For fixed μ regard $N_{\mu\nu}^\lambda$ as a matrix. It gives the matrix for tensor product with V^μ. Show that this matrix is diagonalized by $S_{\mu i}$:

$$\sum_\lambda N_{\mu\nu}^\lambda S_{\lambda j} = \frac{S_{\mu j} S_{\nu j}}{S_{1j}} \tag{11.284}$$

$$\sum_{\nu,\lambda} S_{\nu i}^* N_{\mu\nu}^\lambda S_{\lambda j} = \delta_{ij} \left(\frac{S_{\mu j}}{S_{1j}} \right) \tag{11.285}$$

d.) Show that

$$\frac{S_{\mu 1}}{S_{11}} = \dim V^\mu \tag{11.286}$$

11.13 Induced Representations

Let G be a group and H a subgroup. Suppose that $\rho : H \to \mathrm{Aut}(V)$ is a representation of the *subgroup* H. Using this data we are going to produce, canonically, a representation of G known as an *induced representation*. Note that, in general, there is no way to extend $\rho : H \to \mathrm{Aut}(V)$ to a homomorphism $\rho : G \to \mathrm{Aut}(V)$. As a counterexample, we will see later that the only one-dimensional representation of $SU(2)$ is the trivial representation. However, there are nontrivial representations of the subgroup of diagonal matrices. So, in general, a representation $\rho : H \to \mathrm{Aut}(V)$ of a subgroup $H \subset G$ is not just the restriction of a representation of G on V.

Then, as we have seen, $\mathrm{Map}(G, V)$ is canonically a $G \times H$-space. The left-action of $G \times H$ is defined by declaring that for $(g, h) \in G \times H$ and $\Psi \in \mathrm{Map}(G, V)$ the new function $\phi((g, h), \Psi) \in \mathrm{Map}(G, V)$ is the function $G \to V$ defined by:

$$\phi((g, h), \Psi)(g_0) := \rho(h) \cdot \Psi(g^{-1} g_0 h) \tag{11.287}$$

for all $g_0 \in G$. Or, in slightly lighter notation:

$$(g, h) \cdot \Psi(g_0) := \rho(h)\Psi(g^{-1} g_0 h) \tag{11.288}$$

Now, we can consider the subspace of functions *fixed by the action of* $1 \times H$. That is, we consider functions which satisfy

$$\boxed{\Psi(gh^{-1}) = \rho(h)\Psi(g)} \tag{11.289}$$

for every $g \in G$ and $h \in H$. Put differently: There are two natural left-actions by H on $\mathrm{Map}(G, V)$ and we consider the subspace where they are equal. Such functions are said to be the *H-equivariant*. See the exercise below for the justification of this terminology. Note that the space of H-equivariant functions $G \to V$ is a linear subspace of $\mathrm{Map}(G, V)$. We will denote it by $\mathrm{Ind}_H^G(V)$:

$$\mathrm{Ind}_H^G(V) := \{\Psi : G \to V \;\; | \;\; \Psi(gh^{-1}) = \rho(h)\Psi(g) \qquad \forall g \in G, h \in H\} \tag{11.290}$$

Note that since we are taking the fixed points of the subgroup $\{1_G\} \times H$ of the $G \times H$ action on $\mathrm{Map}(G, V)$ there is still a G action on the fixed point set. More explicitly, if Ψ is an H-equivariant function satisfying (11.289) then $g \cdot \Psi$ with values

$$(g \cdot \Psi)(g_0) := \Psi(g^{-1} g_0) \tag{11.291}$$

is also an H-equivariant function. (Check this!) Thus $\mathrm{Ind}_H^G(V)$ is a representation space of G.

The subspace $\mathrm{Ind}_H^G(V) \subset \mathrm{Map}(G, V)$ of H-equivariant functions, i.e. functions satisfying (11.289) is called the *induced representation of G, induced by the representation V of the subgroup H*. This is an important construction with a beautiful underlying geometrical interpretation. In a sense we will explain below all the representations of compact groups follow from this construction. One can also use it to construct representations of many important noncompact and infinite-dimensional groups. For this reason it appears in many places in physics.

Two examples of important applications in physics are:

1. The irreducible unitary representations of space groups in condensed matter physics.

2. The irreducible unitary representations of the Poincaré group in QFT.

Example: Let us take $V = \mathbb{C}$ with the trivial representation of H, i.e. $\rho(h) = 1$. Then the induced representation is the vector space of functions on G which are invariant under right-multiplication by H. This is precisely the vector space of $\mathbb{C}$-valued functions on the homogeneous space G/H. Recall that for $G = SU(2)$ and $H = U(1)$ we have seen that $G/H = \mathbb{CP}^1 \cong S^2$ so we can get a nice basis of orthogonal functions on S^2 from the functions on $SU(2)$. These are called *spherical harmonics* and we will discuss them more below.

Exercise *Functoriality Of Induction*

a.) Show that

$$\mathrm{Ind}_H^G(V_1 \oplus V_2) \cong \mathrm{Ind}_H^G(V_1) \oplus \mathrm{Ind}_H^G(V_2) \tag{11.292}$$

b.) If $A \in \mathrm{Hom}_H(V_1, V_2)$ is an intertwiner between H-reps then there is an induced map $F(A) : \mathrm{Ind}_H^G(V_1) \to \mathrm{Ind}_H^G(V_2)$ which is an intertwiner of G-representations. [115]

Exercise *H-Equivariance*

Show that a function $\Psi \to V$ satisfying (11.289) fits in a commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{\Psi} & V \\ {\scriptstyle R(h)}\downarrow & & \downarrow{\scriptstyle \rho(h^{-1})} \\ G & \xrightarrow{\Psi} & V \end{array} \tag{11.293}$$

where $R(h) : g \mapsto gh$ is the right action of H on G. Thus a function satisfying (11.289) is a morphism of H-spaces, justifying the term "equivariant."

11.13.1 The Geometrical Interpretation

Note that there is a right H-action on the set $G \times V$:

$$\phi_h : (g, v) \mapsto (gh, \rho(h^{-1})v) \tag{11.294}$$

we can therefore form the quotient space of orbits. In this case it is usually denoted $G \times_H V$, but it is just the set of equivalence classes under the above right H-action. There is a natural map

$$\pi : G \times_H V \to G/H \tag{11.295}$$

[115] *Answer* $(F(A)(\Psi))(g) = A(\Psi(g))$.

given by $\pi : [(g, v)] \mapsto gH$. Referring back to the discussion of equation (8.92) we see that this is the *associated bundle to the principal H bundle $\pi : G \to G/H$.*

When G, H are Lie groups and ρ is a continous representation the map π is continuous. Moreover, the fiber above any coset gH is the vector space V. We therefore have an example of a vector bundle over G/H with fiber V. The sections of the vector bundle are, by definition, continuous maps

♣WE SHOULD EXPLAIN THE NOTION OF SECTIONS IN THE DISCUSSION OF BUNDLES ABOVE. ♣

$$s : G/H \to G \times_H V \tag{11.296}$$

that are a right-inverse to π, that is $\pi \circ s = Id_{G/H}$. To construct such a section we have to identify, for each coset gH an equivalence class in $G \times V$ which projects back down to gH. If we represent gH by g then it must be an equivalence class of the form

$$s(gH) = [(g, v(g))] \tag{11.297}$$

for some vector $v(g)$ associated to g. But now $gH = \tilde{g}H$ when $\tilde{g} = gh$ for $h \in H$. So it must be that

$$[(g, v(g))] = s(gH) = s(\tilde{g}H) = [(\tilde{g}, v(\tilde{g})] = [(gh, v(\tilde{g})] = [(g, \rho(h)v(\tilde{g}))] \tag{11.298}$$

and hence

$$v(g) = \rho(h)v(\tilde{g}) = \rho(h)v(gh) \qquad \Rightarrow \qquad v(gh) = \rho(h^{-1})v(g) \tag{11.299}$$

This must hold for all $g \in G$, and hence $g \mapsto v(g)$ is an equivariant function: Thus,

the space of sections of the homogeneous vector bundle $\pi : G \times_H V \to G/H$ is canonically identified with the space of H-equivariant functions $G \to V$ satisfying (11.289).

11.13.2 Frobenius Reciprocity

The theory of induced representations is already interesting and nontrivial for G, H finite groups. In this case $G \to G/H$ is a finite cover (by H) of a discrete set of points. Nevertheless, the general geometrical ideas apply.

Let $\mathrm{Rep}(G)$ denote the category of finite-dimensional representations of G. Morphisms between $W_1, W_2 \in \mathrm{Rep}(G)$ are linear transformations commuting with G, i.e. G-intertwiners, and the vector space of all morphisms is denoted $\mathrm{Hom}_G(W_1, W_2)$. The induced representation construction defines a functor

$$\mathrm{Ind} : \mathrm{Rep}(H) \to \mathrm{Rep}(G). \tag{11.300}$$

(We denoted this by Ind_H^G before but H, G will be fixed in what follows so we simplify the notation.) On the other hand, there is an obvious functor going the other way, since any G-rep W is *a foriori* an H-rep, by restriction. Let us denote this "restriction functor"

$$R : \mathrm{Rep}(G) \to \mathrm{Rep}(H) \tag{11.301}$$

How are these two maps related? The answer is that they are "adjoints" of each other! This is the statement of Frobenius reciprocity:

$$\mathrm{Hom}_G(W, \mathrm{Ind}(V)) = \mathrm{Hom}_H(R(W), V) \tag{11.302}$$

We can restate the result in another way which is illuminating because it helps to answer the question: How is $\mathrm{Ind}_H^G(V)$ decomposed in terms of irreducible representations of G? Let W_α denote the distinct irreps of G. Then Schur's lemma tells us that

$$\mathrm{Ind}_H^G(V) \cong \oplus_\alpha W_\alpha \otimes \mathrm{Hom}_G(W_\alpha, \mathrm{Ind}_H^G(V)) \tag{11.303}$$

But now Frobenius reciprocity (11.302) allows us to rewrite this as

$$\mathrm{Ind}_H^G(V) \cong \oplus_\alpha W_\alpha \otimes \mathrm{Hom}_H(R(W_\alpha), V) \tag{11.304}$$

where the sum runs over the unitary irreps W_α of G, with multiplicity one.

The statement (11.304) can be a very useful simplification of (11.303) if H is "much smaller" than G. For example, G could be nonabelian, while H is abelian. But the representation theory for abelian groups is much easier! Similarly, G could be noncompact, while H is compact. etc.

Proof of Frobenius reciprocity:

In order to prove (11.304) we note that it is equivalent (see the exercise below) to the statement that the character of $\mathrm{Ind}_H^G(V)$ is given by

$$\chi(g) = \sum_{x \in G/H} \hat{\chi}(x^{-1} g x) \tag{11.305}$$

where x runs over a set of representatives and $\hat{\chi}$ is the character χ_V for H when the argument is in H and zero otherwise.

On the other hand, (11.305) can be understood in a very geometrical way. Think of the homogeneous vector bundle $G \times_H V$ as a collection of points $g_j H$, $j = 1, \ldots, n$ with a copy of V sitting over each point. Now, choose a representative $g_j \in G$ for each coset. Having chosen representatives g_j for the distinct cosets, we may write:

$$g \cdot g_j = g_{g \cdot j} h(g, j) \tag{11.306}$$

where $j \mapsto g \cdot j$ is just a permutation of the integers $1, \ldots, n$, or more invariantly, a permutation of the points in G/H.

Now let us define a basis for the induced representation by introducing a basis v_a for the H-rep V and the equivariant functions determined by:

$$\psi_{i,a}(g_j) := v_a \delta_{i,j} \tag{11.307}$$

Geometrically, this is a section whose support is located at the point $g_i H$. The equivariant function is then given by

$$\psi_{i,a}(g_j h) := \rho(h^{-1}) v_a \delta_{i,j} \tag{11.308}$$

Now let us compute the action of $g \in G$ in this basis:

$$\begin{aligned}(g \cdot \psi_{i,a})(g_j) &= \psi_{i,a}(g^{-1} g_j) \\ &= \psi_{i,a}(g_{g^{-1}\cdot j} h(g^{-1}, j)) \\ &= \delta_{i,g^{-1}\cdot j} \rho(h(g^{-1}, j)^{-1}) \cdot v_a \end{aligned} \tag{11.309}$$

Fortunately, we are only interested in the trace of this G-action. The first key point is that *only the fixed points of the g-action on G/H contribute*. Note that the RHS above is supported at $j = g \cdot i$, but if we are taking the trace we must have $i = j$. But in this case $g g_i = g_i h(g, i)$ and hence $g^{-1} g_i = g_i h(g, i)^{-1}$ so for fixed points we can simplify $h(g^{-1}, i) = h(g, i)^{-1}$, and hence when we take the trace the contribution of a fixed point $g g_i H = g_i H$ is the trace in the H-rep of $h(g, i) = g_i^{-1} g g_i$, as was to be shown ♠

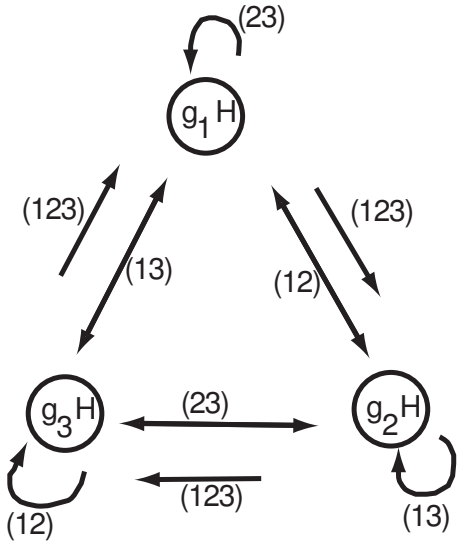


Figure 32: The left action of $G = S_3$ on G/H. In fact, this picture should be considered as a picture of a category, in this case, a groupoid.

Remark: The *Ind* map does not extend to a ring homomorphism of representation rings.

Exercise

♣Physics 619, ch. 5, 2002 for more ♣

Let G be the symmetric group on $\{1, 2, 3\}$ and let $H = \{1, (12)\}$. Choose a representation of H with $V \cong \mathbb{C}$ and $\rho(\sigma) = +1$ or $\rho(\sigma) = -1$.

a.) Show that in either case, the induced representation $\mathrm{Ind}_H^G(V)$ is a three-dimensional vector space.

b.) Choose a basis for $\mathrm{Ind}_H^G(V)$ and compute the representation matrices of the elements of S_3 explicitly.

Example A simple example from finite group theory nicely illustrates the general idea. Let $G = S_3$ be the permutation group. Let $H = \{1, (12)\} \cong \mathbb{Z}_2$ be a $\mathbb{Z}_2$ subgroup. G/H consists of 3 points. The left action of G on this space is illustrated in (**??**).

There are two irreducible representations of H, the trivial and the sign representation. These are both 1-dimensional. Call them $V(\epsilon)$, with $\epsilon = \pm$. Accordingly, we are looking at a line bundle over G/H and the vector space of sections of $G \times_H V(\epsilon)$ is 3-dimensional. A natural basis for the space of sections is given by the functions which are "δ-functions supported at each of the three points":

$$\begin{aligned} s^i(g_j H) &= \delta_{ij} \\ g_1 H = (13)H &= \{(13), (123)\} \\ g_2 H = (23)H &= \{(23), (132)\} \\ g_3 H = (12)H &= \{1, (12)\} \end{aligned} \tag{11.310}$$

These sections correspond to equivariant functions on the total space. The space of all functions $F : G \to \mathbb{R}$ is a six-dimensional vector space. The equivariance condition:

$$\begin{aligned} F(12) &= \epsilon F(1) \\ F(123) &= \epsilon F(13) \\ F(132) &= \epsilon F(23) \end{aligned} \tag{11.311}$$

cuts this six-dimensional space down to a three-dimensional space.

We can choose a basis of equivariant functions by *choosing* 3 representatives g_1, g_2, g_3 for the cosets in G/H and setting $F^i(g_j) = \delta_{ij}$. Using such a basis the representation of the group is easily expressed as a permutation representation.

In our example of $G = S_3$ it is prudent to choose $g_1 = (13), g_2 = (23), g_3 = (12)$ so that

$$\begin{aligned} (12)g_1 &= g_2(12) \\ (12)g_2 &= g_1(12) \\ (12)g_3 &= g_3(12) \\ (13)g_1 &= g_3(12) \\ (13)g_2 &= g_2(12) \\ (13)g_3 &= g_1(12) \\ (23)g_1 &= g_1(12) \\ (23)g_2 &= g_3(12) \\ (23)g_3 &= g_2(12) \end{aligned} \tag{11.312}$$

From this one easily gets the induced representation

$$\rho_{ind}(12) = \begin{pmatrix} 0 & \epsilon & 0 \\ \epsilon & 0 & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \tag{11.313}$$

$$\rho_{ind}(132) = \begin{pmatrix} 0 & 0 & \epsilon \\ \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \end{pmatrix} \tag{11.314}$$

and so forth.

Now let us look at Frobenius reciprocity. The irreducible representations of G are $W(\epsilon)$ defined by $(ij) \to \epsilon$, and W_2 defined by the symmetries of the equilateral triangle, embedded into $O(2)$:

$$\begin{aligned} \rho_2(12) &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ \rho_2(123) &= \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \end{aligned} \tag{11.315}$$

As $H = \mathbb{Z}_2$ representations, we have $W(\epsilon) \cong V(\epsilon)$ and

$$W_2 \cong V(+1) \oplus V(-1) \tag{11.316}$$

Therefore

$$\begin{aligned} \mathrm{Hom}_H(W(\epsilon), V(\epsilon')) &= \delta_{\epsilon,\epsilon'}\mathbb{R} \\ \mathrm{Hom}_H(W_2, V(\epsilon)) &= \mathbb{R} \end{aligned} \tag{11.317}$$

By Frobenius reciprocity we have

$$\mathrm{Ind}_{S_2}^{S_3}(V(\epsilon)) = W(\epsilon) \oplus W_2 \tag{11.318}$$

Let us check this by computing characters from the geometric perspective.

The induced representation consists of functions on G/H valued in $V(\epsilon)$. The action of g acts as a permutation representation on the support of the functions.

Therefore, we can compute the character of g in the induced representation by looking at its action on fixed points:

$$\chi_{\rho_{ind}}(g) = \sum_{Fix(g)} \chi_{V(\epsilon)}(h(g,i)) \tag{11.319}$$

To find out how to decompose the representation $\mathrm{Ind}_{S_2}^{S_3}(V(\epsilon))$ in terms of $G = S_3$ irreps it suffices to compute the character for $g = (12)$ and $g = (123)$. Now, $g = (12)$ has exactly one fixed point, namely $g_3 H$ and $h(g, 3) = (12)$ for this element. Therefore,

$$\chi_{\rho_{ind}}(12) = \chi_V(12) = \epsilon \tag{11.320}$$

On the other hand, $g = (123)$ clearly has no fixed points, and therefore the character is zero. It follows immediately that we have the decomposition (11.318).

****

MORE MATERIAL IN GTLect4-IntroRepTheory-2020

****

****

11.14 Representations Of $SU(2)$

11.14.1 Homogeneous Polynomials

We now use the idea of induced representations to construct representations of $SU(2)$. In fact, we will construct all the irreducible representations.

♣Using ρ rather than T for representations here. Uniformize? ♣

We take $G = SU(2)$ and we take the subgroup $H \subset SU(2)$ to be a "maximal torus" namely the subgroup of diagonal $SU(2)$ matrices. We will denote it by T. As a group $T \cong U(1)$. For V we choose an irreducible representation ρ_k of $U(1)$ where $k \in \mathbb{Z}$ and the carrier space is $V \cong \mathbb{C}$. Thus we choose the one-dimensional representation of T defined by:

♣This conflicts with our notation T for the homomorphism in a rep. So we use ρ here. ♣

$$\rho_k\left(\begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}\right) := z^k \tag{11.321}$$

where $z \in U(1)$ and $k \in \mathbb{Z}$. Then the induced representation

$$\mathrm{Ind}_{U(1)}^{SU(2)}(\rho_k) \tag{11.322}$$

is, by definition the space of functions $F : SU(2) \to V = \mathbb{C}$ satisfying the equivariance condition

$$F(\begin{pmatrix} u & -\bar{v} \\ v & \bar{u} \end{pmatrix} \begin{pmatrix} e^{\mathrm{i}\theta} & \\ & e^{-\mathrm{i}\theta} \end{pmatrix}) = e^{-\mathrm{i}k\theta} F(\begin{pmatrix} u & -\bar{v} \\ v & \bar{u} \end{pmatrix}) \tag{11.323}$$

We can abbreviate

$$F(\begin{pmatrix} u & -\bar{v} \\ v & \bar{u} \end{pmatrix}) \Rightarrow F(u, v) \tag{11.324}$$

so we just have

$$F(ue^{\mathrm{i}\theta}, ve^{\mathrm{i}\theta}) = e^{-\mathrm{i}k\theta} F(u, v) \tag{11.325}$$

Of course $g \in SU(2)$ implies $|u|^2 + |v|^2 = 1$. So we can also view this as the space of equivariant functions $S^3 \to \mathbb{C}$ for the right-action of $U(1)$.

Actually, we don't want <u>all</u> functions - that space is too big. One useful subspace to restrict to is $L^2(S^3)$.

The induced representation is infinite dimensional and very far from being irreducible. For example, the functions $u^k|u|^\ell$ for ℓ any positive integer $\geq -k$ are all smooth functions on $SU(2)$ and are independent vectors in $L^2(SU(2))$.

It turns out we can cut down the ρ_k-equivariant functions $L^2(SU(2))$ down to a finite dimensional irrep of $SU(2)$ by imposing one more condition. It is a condition of *holomorphy* as we now explain.

One way to view this restriction is first to view $S^3 \subset \mathbb{C}^2$ and consider instead the space of all functions of two variables

$$\begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{C}^2 \tag{11.326}$$

which are equivariant in the sense of equation (11.325). This larger space of functions is also a representation of $SU(2)$ and also we can restrict functions to give equivariant

functions on $SU(2)$. But now we can look at the subspace of equivariant functions which are also holomorphic functions on $\mathbb{C}^2$. This will be a finite-dimensional space of functions.

Holomorphy requires that $-k \geq 0$. Such homogeneous holomorphic functions must be polynomials in u, v. It is convenient to set $-k = 2j \in \mathbb{Z}_+$. The restriction to the holomorphic equivariant functions gives the finite dimensional space $\mathcal{H}_{2j}$ of homogeneous polynomials in u, v of degree $2j$. In physics this is called the *spin-j representation* for reasons explained below. In general we will denote the isomorphism class of the spin-j representation by V_j.

There is a second, very useful, way to think about the restriction to holomorphic functions. Recall that:

$$SU(2)/U(1) \cong SL(2,\mathbb{C})/B \tag{11.327}$$

where B is the subgroup of upper triangular matrices.

$$B = \{ \begin{pmatrix} z & w \\ 0 & z^{-1} \end{pmatrix} | z \in \mathbb{C}^*, w \in \mathbb{C} \} \tag{11.328}$$

Both quotients are $\mathbb{CP}^1$. But the latter description is useful because we can introduce the idea of *holomorphy*. Note that the representation ρ_k of $U(1)$ extends uniquely to a holomorphic representation of B:

$$\rho_k(\begin{pmatrix} z & w \\ 0 & z^{-1} \end{pmatrix}) = z^k \tag{11.329}$$

Now consider

$$\mathrm{Ind}_B^{SL(2,\mathbb{C})}(\rho_k) \tag{11.330}$$

Once again we can consider the equivariant functions:

$$F(gb) = \rho_k(b^{-1})F(g) \qquad\qquad b \in B \tag{11.331}$$

Now $g \in SL(2,\mathbb{C})$ looks like

$$g = \begin{pmatrix} u & s \\ v & t \end{pmatrix} \qquad\qquad ut - vs = 1 \tag{11.332}$$

and any three matrix elements will serve as holomorphic coordinates as long as we can solve for the fourth. However we have matrices

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \in B \tag{11.333}$$

and equivariance under these matrices implies that

$$F\left(\begin{pmatrix} u & s \\ v & t \end{pmatrix} \right) = F\left(\begin{pmatrix} u & s+ux \\ v & t+vx \end{pmatrix} \right) \tag{11.334}$$

and hence B-equivariant functions on $SL(2,\mathbb{C})$ are again functions of just the vector

$$\begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{C}^2 - \{0\} \ . \tag{11.335}$$

such that

$$F(\begin{pmatrix} u \\ v \end{pmatrix} z) = z^{-k} F(\begin{pmatrix} u \\ v \end{pmatrix}) \tag{11.336}$$

for $z \in \mathbb{C}^*$. Such functions restrict to the functions on the 3-sphere $|u|^2 + |v|^2 = 1$ and a function equivariant for B by ρ_k restricts to a function on $SU(2)$, equivariant for T by ρ_k since

$$B \cap SU(2) = T \tag{11.337}$$

Conversely we can use the equivariance to extend the functions in $\mathrm{Ind}_{U(1)}^{SU(2)}(\rho_k)$ to functions in $\mathrm{Ind}_{B}^{SL(2,\mathbb{C})}(\rho_k)$ so they are really the same representation:

$$\mathrm{Ind}_{B}^{SL(2,\mathbb{C})}(\rho_k) \cong \mathrm{Ind}_{U(1)}^{SU(2)}(\rho_k) \tag{11.338}$$

The great advantage of the description $\mathrm{Ind}_{B}^{SL(2,\mathbb{C})}(\rho_k)$ is that in this description it is manifest that there is a subspace of holomorphic equivariant functions: Such functions extend uniquely to holomorphic functions on all of $\mathbb{C}^2$ (by a theorem known as Hartog's theorem) so we are simply considering holomorphic functions on $\mathbb{C}^2$. In this way we have found a nice finite dimensional subspace of $\mathrm{Ind}_{B}^{SL(2,\mathbb{C})}(\rho_k)$ of equivariant holomorphic functions.

Since the $SU(2)$ action is holomorphic, the restriction to the finite-dimensional subspace of holomorphic functions in fact forms a representation of $SU(2)$. It is the space of homogeneous polynomials of degree $2j = -k$ in u, v. (The space is empty if $k > 0$.)

A basis for $\mathcal{H}_{2j}$ is

$$\tilde{f}_{j,m}(u,v) := u^{j+m} v^{j-m} \tag{11.339}$$

for $m = -j, -j+1, -j+2, \cdots, j-1, j$. Note that m increases in steps of $+1$ and hence $j \pm m$ is always an integer even though j, m might be half-integer. Thus $V_j \cong \mathcal{H}_{2j}$ is a complex vector space of dimension $2j+1$. Now, for $g \in SU(2)$,

$$g = \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix} \qquad |\alpha|^2 + |\beta|^2 = 1 \tag{11.340}$$

we compute the matrix elements for this representation of $SU(2)$ relative to this basis via:

$$\begin{aligned} (g \cdot \tilde{f}_{j,m})(u,v) &:= \tilde{f}_{j,m}(\bar{\alpha}u + \bar{\beta}v, -\beta u + \alpha v) \\ &= (\bar{\alpha}u + \bar{\beta}v)^{j+m}(-\beta u + \alpha v)^{j-m} \\ &:= \sum_{m'} \tilde{D}^j_{m'm}(g) \tilde{f}_{j,m'} \end{aligned} \tag{11.341}$$

Note that the basis $\tilde{f}_{j,m}$ diagonalizes the action of the diagonal $SU(2)$ matrices with $\beta = 0$:

$$g \cdot \tilde{f}_{j,m} = \bar{\alpha}^{j+m} \alpha^{j-m} \tilde{f}_{j,m} = \alpha^{-2m} \tilde{f}_{j,m} \tag{11.342}$$

That is,

$$\tilde{D}^j_{m_L,m_R}(g) = \delta_{m_L,m_R} \alpha^{-2m_R} \tag{11.343}$$

when $\beta = 0$.

More generally, we can derive an explicit formula for the matrix elements $\tilde{D}^j_{m'm}(g)$ as functions on $SU(2)$ by expanding out the two factors in (11.341) using the binomial theorem and collecting terms:

$$\tilde{D}^j_{m'm}(g) = \sum_{s+t=j+m'} \binom{j+m}{s} \binom{j-m}{t} \bar{\alpha}^s \alpha^{j-m-t} \bar{\beta}^{j+m-s} (-\beta)^t \tag{11.344}$$

Here the sum is over integers $s \geq 0$ and $t \geq 0$ and we recall that $\binom{a}{b} = 0$ when $b > a$ so this is a finite sum. We recover the functions $\tilde{f}_m$, up to scale from

$$\tilde{D}^j_{-m,-j}(g) = (-1)^{j-m} \binom{2j}{j-m} \alpha^{j+m} \beta^{j-m} \tag{11.345}$$

We claim that the representations $\mathcal{H}_{2j}$ are irreducible, and moreover give a full set of irreducible representations of $SU(2)$. We will prove this using characters in the next section.

Remarks

1. *Representations Of $SO(3)$ That Lift To Representations Of $SU(2)$*

   Since $SO(3) \cong SU(2)/Z$ where $Z \cong \mathbb{Z}_2$ is the central subgroup consisting of $SU(2)$ matrices proportional to the unit matrix: $Z = \{\pm 1_{2\times 2}\}$, we can easily determine the irreducible representations of $SO(3)$. On the spin j representation the central element acts as $\rho_j(-1) = (-1)^{2j} 1_{V_j}$. Therefore: *the irreducible representations of $SO(3)$ are given by V_j for $j \in \mathbb{Z}$, that is, the representations where $\dim_{\mathbb{C}} V_j$ is odd. The other irreducible representations of $SU(2)$ with $\dim_{\mathbb{C}} V_j$ even are <u>not</u> representations of $SO(3)$, although they are projective representations - see section **** below.*

2. Returning to the induced representation $\mathrm{Ind}_{U(1)}^{SU(2)}(\rho_k)$ of smooth (not holomorphic) equivariant functions on $SU(2)$ we see that the functions $\tilde{D}$ satisfy the equivariance condition

   $$\tilde{D}^j_{m_L,m_R}(g \begin{pmatrix} e^{\mathrm{i}\theta} & \\ & e^{-\mathrm{i}\theta} \end{pmatrix}) = (e^{\mathrm{i}\theta})^{-2m_R} \tilde{D}^j_{m_L,m_R}(g) \tag{11.346}$$

   and hence we get the equivariance condition for $2m_R = k$. Such values of m_R can only appear for $j \geq |k|/2$ with $j = |k|/2 \mathrm{mod} 1$. So the induced representation is spanned by the functions $\tilde{D}^j_{m_L,k/2}$ with $j \geq |k|/2$ and $j = |k|/2 \mathrm{mod} 1$ and $m \in \{-j, -j+1, \ldots, j-1, j\}$. Thus as a representation of $SU(2)$ we have:

   $$\mathrm{Ind}_{U(1)}^{SU(2)}(\rho_k) \cong V_{|k|/2} \oplus V_{|k|/2+1} \oplus \cdots \tag{11.347}$$

When k is nonpositive we can set $k = -2j_0$ and the span of the set of functions $D^{j_0}_{m_L,-j_0}$ where m ranges over $-j_0, -j_0+1, \ldots, j_0-1, j_0$ transforms under the left regular representation in a representation isomorphic to V_j. These are the functions with a holomorphic interpretation. When k is positive there is no holomorphic interpretation.

3. It is instructive to evaluate the functions $\tilde{D}^j_{m'm}(g)$ where g is parametrized by Euler angles. We find:

$$\tilde{D}^j_{m_L,m_R}(g) = e^{-\mathrm{i}\phi m_L}\tilde{P}^j_{m_L,m_R}(\theta)e^{-\mathrm{i}\psi m_R} \tag{11.348}$$

where $\tilde{P}^j_{m_L,m_R}(\theta)$ is a polynomial in $\cos(\theta/2)$ and $\sin(\theta/2)$:

$$\tilde{P}^j_{m_L,m_R}(\theta) = (-\mathrm{i})^{2j+m_L+m_R}\sum_{s+t=j+m_L}(-1)^s\binom{j+m_R}{s}\binom{j-m_R}{t}(\cos\theta/2)^{j-m_R+s-t}((\sin\theta/2)^{j+m_R+t-s} \tag{11.349}$$

It is closely related to an *associated Legendre polynomial* and the functions $\tilde{D}^j_{m'm}$ are closely related to *Wigner functions*. We will explain more about this below.

4. Looking ahead to the relation to Lie algebras, diagonal $SU(2)$ matrices can be written as $g = \exp[-\mathrm{i}\sigma^3\phi]$ and then

$$g\cdot\tilde{f}_m = e^{\mathrm{i}2m\phi}\tilde{f}_m \tag{11.350}$$

so the basis $\tilde{f}_m$ is proportional to the physicist's basis $|j,m\rangle$. Note that

$$\begin{pmatrix} e^{\mathrm{i}\phi} & \\ & e^{-\mathrm{i}\phi}\end{pmatrix} = \exp[-2\phi\left(-\frac{\mathrm{i}}{2}\sigma^3\right)] \tag{11.351}$$

so this basis diagonalizes the generator $T^3 = -\frac{\mathrm{i}}{2}\sigma^3$, that is $\rho(T^3)\cdot\tilde{f}_{j,m} = -m\tilde{f}_{j,m}$.

Exercise *Explicit Matrices For Small j*

a.) Show that for $j = 1/2$, relative to the ordered basis $\{\tilde{f}_{+1/2}, \tilde{f}_{-1/2}\}$ we have

$$\tilde{D}^{j=1/2}(g) = \begin{pmatrix}\bar{\alpha} & \bar{\beta}\\ -\beta & \alpha\end{pmatrix} = g^* \tag{11.352}$$

b.) Show that for $j = 1$, relative to the ordered basis $\{\tilde{f}_1, \tilde{f}_0, \tilde{f}_{-1}\}$ we have

$$\tilde{D}^{j=1}(g) = \begin{pmatrix}\bar{\alpha}^2 & -\bar{\alpha}\beta & \beta^2\\ 2\bar{\alpha}\bar{\beta} & |\alpha|^2-|\beta|^2 & -2\alpha\beta\\ \bar{\beta}^2 & \alpha\bar{\beta} & \alpha^2\end{pmatrix} \tag{11.353}$$

11.14.2 Chararacters Of The Representations V_j

Every $g \in SU(2)$ is diagonalizable so we can say

$$g \sim d(z) := \begin{pmatrix} z & \\ & z^{-1} \end{pmatrix} \tag{11.354}$$

where $|z| = 1$ and both z and z^{-1} define the same conjugacy class [116]

For the spin j representation denote the character by χ_j. In the basis described above $\tilde{D}^j(d(z))$ is diagonal and given by

$$\tilde{D}_j(d(z)) = \mathrm{Diag}\{z^{-2j}, z^{-2j+2}, \ldots, z^{2j-2}, z^{2j}\} \tag{11.355}$$

Therefore

$$\chi_j(g) = z^{-2j} + z^{-2j+2} + \cdots + z^{2j-2} + z^{2j} = \frac{z^{2j+1} - z^{-2j-1}}{z - z^{-1}} \tag{11.356}$$

Now we write the orthogonality relations for the characters. We have written the Haar measure above in Euler angles, but this is not the most convenient form for integrating class functions. Given the Euler angles (ϕ, θ, ψ) of $u \in SU(2)$ it is not so evident how to find the value of the angle ξ such that

$$g = \cos\xi + \mathrm{i}\sin\xi \hat{n}\cdot\vec{\sigma} = \exp[\mathrm{i}\xi\hat{n}\cdot\vec{\sigma}] \tag{11.357}$$

The conjugacy class with angle ξ is a two-dimensional sphere in S^3 with radius $\sin\xi$. Using this we see that the properly normalized measure on $SU(2)$ is such that on a class function F we have:

$$\int_{SU(2)} F(u)[du] = \frac{1}{\pi}\int_0^{2\pi} f(\theta)\sin^2\theta d\theta = -\frac{1}{4\pi\mathrm{i}}\oint g(z)(z - z^{-1})^2\frac{dz}{z} \tag{11.358}$$

where

$$f(\theta) = F(\begin{pmatrix} e^{\mathrm{i}\theta} & \\ & e^{-\mathrm{i}\theta} \end{pmatrix}) = g(z) \tag{11.359}$$

with $z = e^{\mathrm{i}\theta}$.

Now, the space of class functions $L^2(SU(2))^{class}$ can be identified with the (completion of) the space of Laurent polynomials in z which is even under the involution $z \to 1/z$. This space is clearly spanned by $z^{2j} + z^{-2j}$, and that basis is related to the χ_j by an upper triangular matrix. Moreover, one easily confirms the general relation

$$\langle \chi_j, \chi_{j'} \rangle = \delta_{j,j'} \tag{11.360}$$

From this we learn that we have a complete set of representations and moreover these are all irreducible representations.

Remarks

[116] $d(z) = vd(z^{-1})v^{-1}$ where, for example, we can take $v = \mathrm{i}\sigma^1$.

1. If we write $n = 2j$ and $z = e^{\mathrm{i}\theta}$ then

$$\chi_j(u) = \frac{\sin((n+1)\theta)}{\sin\theta} = U_{n+1}(\cos\theta) \tag{11.361}$$

where $U_{n+1}(x)$ are known as the *Chebyshev polynomials of the second kind.*

2. *Weyl Density Formula.* For more details about the following see Chapter 5, Survey of Matrix Groups. The formula for integrating class functions with the Haar measure has a nice analog for all the classical groups. For example, if $F : SU(n) \to \mathbb{C}$ is a class function then $F(u)$ only depends on the eigenvalues of u. Write the unordered set of eigenvalues as $\{e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_n}\}$. Then define

$$F(u) = f(\theta_1, \ldots, \theta_n) \tag{11.362}$$

and then, we have

$$\int_{SU(n)} F(u)du = \frac{1}{n!}\int_0^{2\pi} f(\theta_1, \ldots, \theta_n) \prod_{1\le i<j\le n} |e^{\mathrm{i}\theta_i} - e^{\mathrm{i}\theta_j}|^2 \prod_i \frac{d\theta_i}{2\pi} \tag{11.363}$$

and there are similar formulae for $SO(n)$ and $USp(2n)$. In $SO(n)$ we can conjugate any matrix to the form

$$Diag\{R(\theta_1), \ldots, R(\theta_r)\} \tag{11.364}$$

when $n = 2r$ is even and

$$Diag\{R(\theta_1), \ldots, R(\theta_r), 1\} \tag{11.365}$$

when $n = 2r + 1$ is odd. We then have

$$\int_{SO(2n)} f[dg] = \frac{2^{(n-1)^2}}{\pi^n n!}\int_{-\pi}^{\pi} \prod_{1\le j<k\le n} (\cos\theta_j - \cos\theta_k)^2 f(\theta_1, \ldots, \theta_n) \prod d\theta_i \tag{11.366}$$

$$\int_{SO(2n+1)} f[dg] = \frac{2^{n^2}}{\pi^n n!}\int_{-\pi}^{\pi} \prod_{1\le j<k\le n} (\cos\theta_j - \cos\theta_k)^2 \prod_{j=1}^{n} \sin^2(\theta_j/2) f(\theta_1, \ldots, \theta_n) \prod d\theta_i \tag{11.367}$$

Where we have normalized the Haar measure to have volume one. Finally, for the unitary symplectic group $USp(2r)$, any matrix can be conjugated to

$$Diag\{e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_r}, e^{-\mathrm{i}\theta_1}, \ldots, e^{-\mathrm{i}\theta_r}\} \tag{11.368}$$

in terms of which the formula for integrating class functions is:

$$\int_{USp(2n)} f[dg] = \frac{2^{n^2}}{\pi^n n!}\int_{-\pi}^{\pi} \prod_{1\le j<k\le n} (\cos\theta_j - \cos\theta_k)^2 \prod_{j=1}^{n} \sin^2(\theta_j) f(\theta_1, \ldots, \theta_n) \prod d\theta_i \tag{11.369}$$

These kinds of formulae appear very frequently in the subject of random matrix theory.

♣Need to give a reference here. ♣

3. *Weyl-Kac Character Formula* The character formula for $SU(2)$ can be written as

$$\chi_j = e^{-\rho} \frac{\sum_{w \in W} \epsilon(w) e^{w(\lambda+\rho)}}{\prod_{\alpha>0}(1-e^{-\alpha})} \tag{11.370}$$

where e^{λ}, e^{α} and e^{ρ} are functions on the maximal torus such that

$$e^{\lambda}(d(z)) = z^{2j} \tag{11.371}$$

$$e^{\alpha}(d(z)) = z^2 \qquad\qquad e^{\rho} = e^{\alpha/2} \tag{11.372}$$

and the sum on w is a sum over the Weyl group, with $\epsilon(w) \in \{\pm 1\}$ a homomorphism to μ_2.

11.14.3 Unitarization

As we have seen, the finite-dimensional representations of $SU(2)$ are unitarizable, and unitarity is very important in physics and mathematics.

One might first be tempted to define a unitary structure on $\mathcal{H}_{2j}$ by declaring the inner product of two homogeneous polynomials to be the integral of $\psi_1^* \psi_2$. But this will not converge. Instead we take

$$\langle \psi_1, \psi_2 \rangle_{\mathcal{H}_{2j}} := \frac{1}{\pi(2j+1)!} \int_{\mathbb{C}^2} \psi_1(u,v)^* \psi_2(u,v) e^{-|u|^2-|v|^2} d^2u d^2v \tag{11.373}$$

This will give finite overlaps on $\mathcal{H}_{2j}$. Note that since $du \wedge dv$ and $|u|^2 + |v|^2$ are $SU(2)$ invariant the inner product is $SU(2)$-invariant, i.e. it is unitary:

$$\langle g \cdot \psi_1, g \cdot \psi_2 \rangle_{\mathcal{H}_{2j}} = \langle \psi_1, \psi_2 \rangle_{\mathcal{H}_{2j}} \tag{11.374}$$

The reason for the funny normalization constant will be clearer below.

Exercise *ON Basis*
Show that an ON basis is given by

$$f_{j,m} = \frac{1}{\sqrt{\pi}} \sqrt{\frac{(2j+1)!}{(j+m)!(j-m)!}} u^{j+m} v^{j-m} \tag{11.375}$$

11.14.4 Inhomogeneous Polynomials And Mobius Transformations On $\mathbb{CP}^1$

We saw in section **** above that there is a close relation between $SU(2)/U(1)$ and $SL(2,\mathbb{C})/B$ and $\mathbb{CP}^1$. Indeed, $SL(2,\mathbb{C})$ acts naturally on $\mathbb{CP}^1$:

$$\begin{pmatrix} \alpha & s \\ \beta & t \end{pmatrix} \cdot [z_1 : z_2] \to [\alpha z_1 + s z_2 : \beta z_1 + t z_2] \tag{11.376}$$

On the other hand we can identify $\mathbb{CP}^1$ with S^2 and then we can stereographically project $\mathbb{CP}^1$ to $\mathbb{C} \cup \{\infty\}$. See the exercise below for the precise formulae. The projection from $\mathbb{CP}^1$ to the Riemann sphere is just:

$$[z_1 : z_2] \mapsto \begin{cases} z := z_1/z_2 & z_2 \neq 0 \\ \infty & z_2 = 0 \end{cases} \tag{11.377}$$

It follows that $SL(2, \mathbb{C})$ acts on the Riemann sphere by Mobius transformations:

$$\begin{pmatrix} \alpha & s \\ \beta & t \end{pmatrix} \cdot z \to \frac{\alpha z + s}{\beta z + t} \tag{11.378}$$

and restricted to $SU(2)$ this becomes:

$$\begin{pmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{pmatrix} \cdot z \to \frac{\alpha z - \beta^*}{\beta z + \alpha^*} \tag{11.379}$$

Since the $SU(2)$ action on S^2 factors through the usual $SO(3)$ rotations it is not surprising that the round volume form is invariant under $SU(2)$. We can verify this with the volume form

$$\omega := \frac{1}{2\pi} \frac{\mathrm{i} dz \wedge d\bar{z}}{(1 + |z|^2)^2} \tag{11.380}$$

directly since if $\tilde{z} = g \cdot z$ then one easily checks:

$$d\tilde{z} = \frac{dz}{(\beta z + \alpha^*)^2} \tag{11.381}$$

$$\frac{1}{(1 + |\tilde{z}|^2)^2} = \frac{|\beta z + \alpha^*|^2}{(1 + |z|^2)^2} \tag{11.382}$$

and therefore the measure ω is invariant. Indeed, as you show in an exercise below, it is just the usual round measure on S^2 evaluated in stereographic coordinates.

Now let us return to the space $\mathcal{H}_{2j}$ of homogeneous polynomials in z_1, z_2. This vector space is canonically isomorphic to the vector space of polynomials in $z = z_1/z_2$ of degree $\leq 2j$ since we can identify a homogenous polynomial F with a polynomial p by

$$p(z) := z_2^{-2j} F(z_1, z_2) \tag{11.383}$$

The map $F \to p$ defines a map $\mathcal{H}_{2j} \to \mathcal{P}_{2j}$ and we can make this an $SU(2)$-equivariant map if we declare the action of $SU(2)$ on polynomials p to be [117]

$$(g \cdot p)(z) := (-\beta z + \alpha)^{2j} p\left(\frac{\alpha^* z + \beta^*}{-\beta z + \alpha}\right) \tag{11.384}$$

The model for the spin j representation V_j as the space of polynomials of degree $\leq 2j$ can be turned into a Hilbert space by defining

$$h(\psi_1, \psi_2) := (1 + |z|^2)^{-2j} \psi_1^*(z) \psi_2(z) \tag{11.385}$$

[117] One needs to be careful to remember that the left g-action on functions takes the argument z of the function to $g^{-1}z$. Note it is g^{-1} and not g.

One checks that this is invariant under the $SU(2)$ action so we integrate:

$$\langle \psi_1, \psi_2 \rangle_{\mathcal{P}_{2j}} = \int_{\mathbb{C}} h(\psi_1, \psi_2)\omega \tag{11.386}$$

and then

$$\langle g \cdot \psi_1, g \cdot \psi_2 \rangle = \langle \psi_1, \psi_2 \rangle \tag{11.387}$$

so g acts as a unitary operator on $\mathcal{P}_{2j}$ in this inner product.

Using the integrals

$$\int_0^\infty \frac{x^{n_1}}{(1+x)^{n_2}} dx = \frac{\Gamma(n_1+1)\Gamma(n_2-n_1-1)}{\Gamma(n_2)} \qquad -1 < n_1 \quad \& \quad 1 < n_2 - n_1 \tag{11.388}$$

we check that an ON basis for $\mathcal{P}_{2j}$ is given by:

$$\psi_\ell := N_\ell z^\ell \qquad \ell = 0, \ldots, 2j \tag{11.389}$$

$$N_\ell := \sqrt{(2j+1)\binom{2j}{\ell}} \tag{11.390}$$

In physics this is usually written as:

$$\psi_{j,m} = \sqrt{\frac{(2j+1)!}{(j+m)!(j-m)!}} z^{j+m} = N_{j,m} z^{j+m} \tag{11.391}$$

Now the Wigner functions are defined by the matrix elements relative to this ON basis:

$$g \cdot \psi_{j,m_R} = \sum_{m_L} D^j_{m_L,m_R}(g)\psi_{j,m_L} \tag{11.392}$$

In physics notation $\psi_{j,m}$ would usually be written as the ket vector $|j,m\rangle$, the spin j representation might be written as operators $T^j(g)$ and we would have:

$$D^j_{m'm}(g) := \langle j, m' | T^j(g) | j, m \rangle \tag{11.393}$$

Remarks:

1. There is an isometry of Hilbert spaces between $\mathcal{H}_{2j}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_{2j}}$ defined in equation (11.373) above and $\langle \cdot, \cdot \rangle_{\mathcal{P}_{2j}}$. Indeed, in the integral (11.373) we can change coordinates to $z = u/v$ and v. Then $du \wedge dv = v dz \wedge dv$ and one can do the v integral explicitly to show the isometry.

2. To relate the Wigner function $D^j_{m'm}$ to our functions $\tilde{D}^j_{m'm}$ defined above we compute

$$\begin{aligned} D^j_{m',m}(g) &:= \langle \psi_{j,m'}, g \cdot \psi_{j,m} \rangle \\ &= N_{j,m'} N_{j,m} \int_{\mathbb{C}} (z^{j+m'})^* (-\beta z + \alpha)^{j-m} (\alpha^* z + \beta^*)^{j+m} \frac{d\phi}{2\pi} \frac{d(r^2)}{(1+r^2)^{2j+2}} \\ &= N_{j,m'} N_{j,m} \sum_{s+t=j+m'} \binom{j+m}{s} \binom{j-m}{t} \bar{\alpha}^s \alpha^{j-m-t} \bar{\beta}^{j+m-s} (-\beta)^t \int_0^\infty r^{2j+2m'} \frac{d(r^2)}{(1+r^2)^{2j+2}} \\ &= \sqrt{\frac{(j+m')!(j-m')!}{(j+m)!(j-m)!}} \tilde{D}^j_{m',m}(g) \end{aligned} \tag{11.394}$$

3. As we noted before, if we induce from the trivial representation then we get functions on G/H which in this case is $G/H = S^2$. The trivial representation of $U(1)$ is $m_R = 0$. In this case $j = \ell$ must be an integer. The functions

$$D^{\ell}_{m_L,0}(g) = e^{-\mathrm{i}m_L\phi}P^{\ell}_{m_L}(\theta) = Y_{\ell,m}(\theta,\phi) \tag{11.395}$$

are known as *spherical harmonics*. From what we have said they form a complete orthonormal set of functions on S^2 and are widely used in electromagnetism and quantum mechanics.

4. If we specialize further to $m_L = 0$ we get the famous Legendre polynomials

$$P_\ell(\cos\theta) = 2^{-\ell}\sum_{s=0}^{\ell}\binom{\ell}{s}^2(-1)^{\ell-s}(1+\cos\theta)^s(1-\cos\theta)^{\ell-s} \tag{11.396}$$

whereas the $P_{\ell,m}(\theta)$ are known as associated Legendre functions. Now some addition theorems become transparent from group theory. For example, if we consider $g_1 = e^{\mathrm{i}\theta_1\sigma^1/2}$ and $g_2 = e^{\mathrm{i}\theta_2\sigma^1/2}$ then

$$D^{\ell}_{00}(g_1g_2^{-1}) = \sum_m D^{\ell}_{0m}(g_1)D^{\ell}_{m0}(g_2^{-1}) \tag{11.397}$$

becomes one of the identities for Legendre polynomials known as the "addition theorem." See, for example, J.D. Jackson, *Classical Electrodynamics*, section 3.6. [118]

5. The Wigner functions have many applications in physics. As we will see, from group theory, they satisfy some nice differential equations and hence appear in the wavefunctions of atoms. They also appear in the study of the quantum Hall effect on S^2 and in the study of wavefunctions of electrons in the field of a magnetic monopole of magnetic charge $2j$.

Exercise *Relations Between Coordinates On S^2 And Stereographic Projection*

a.) Let $\hat{n} = (\hat{x}^1,\hat{x}^2,\hat{x}^3) = (\sin\theta\cos\phi, \sin\theta, \sin\phi, \cos\theta)$ parametrize the unit sphere S^2. Verify the following relations: Now we have the relation between stereographic and angular coordinates where $\hat{x}^3 = 1$ corresponds to $z = \infty$ and $\hat{x}^3 = -1$ corresponds to $z = 0$:

$$\begin{aligned}
\frac{z}{1+|z|^2} &= \frac{1}{2}(\hat{x}^1 + \mathrm{i}\hat{x}^2) = \frac{1}{2}e^{\mathrm{i}\phi}\sin\theta \\
\frac{|z|^2}{1+|z|^2} &= \frac{1}{2}(1+\hat{x}^3) = \frac{1+\cos\theta}{2} \\
\frac{1}{1+|z|^2} &= \frac{1}{2}(1-\hat{x}^3) = \frac{1-\cos\theta}{2} \\
\frac{|z|^2-1}{|z|^2+1} &= \hat{x}^3 \\
z &= \frac{x_1+\mathrm{i}x_2}{1-x_3} = e^{\mathrm{i}\phi}\cot(\frac{\theta}{2})
\end{aligned} \tag{11.398}$$

[118] We have specialized Jackson's equation 3.62. For the more general statement let $g_1 = e^{\mathrm{i}\theta_1\sigma^1/2}e^{\mathrm{i}\phi_1\sigma^3/2}$ and similarly for g_2.

b.) Also show that the standard round metric of the unit sphere can be written as

$$ds^2 = \frac{1}{\pi}\frac{|dz|^2}{(1+|z|^2)^2} \tag{11.399}$$

and the volume form is:

$$\begin{aligned}\omega :=& \frac{1}{2\pi}\frac{\mathrm{i}dz\wedge d\bar{z}}{(1+|z|^2)^2} \\ =& \frac{1}{(1+|z|^2)^2}\frac{d\phi}{2\pi}d(|z|^2) \\ =& \frac{1}{4\pi}\sin\theta d\phi\wedge d\theta\end{aligned} \tag{11.400}$$

is the spherically symmetric unit volume measure on the sphere.

c.) Show that

$$\frac{1}{2}(1+\hat{x}\cdot\vec{\sigma}) = \frac{1}{1+|z|^2}\begin{pmatrix}|z|^2 & \bar{z}\\ z & 1\end{pmatrix} \tag{11.401}$$

is a 2×2 projection matrix that depends smoothly on $\hat{x}\in S^2$. It plays many important roles and is known as the *Bott projector*.

Exercise *Spherical Harmonics And Polynomials*

Show that the (rescaled) spherical harmonics can be written as

$$\tilde{D}^{\ell}_{m_L,0} = |\beta|^{2\ell}\sum_{s+t=\ell+m_L}\binom{\ell}{s}\binom{\ell}{t}(-1)^t\bar{z}^s z^{\ell-t} \tag{11.402}$$

where $z=\alpha/\beta$.

11.14.5 The Geometrical Interpretation Of $\mathcal{P}_{2j}$

We discussed in section 11.13.1 above that equivariant functions can be viewed as sections of an associated bundle. In this section we unpack that idea a bit for the case of $SU(2)$.

In the case of $SU(2)/U(1)$ we have the line bundle

$$\mathcal{L}_k = (SU(2)\times V_k)/U(1) \tag{11.403}$$

where V_k is the irreducible representation of $U(1)$.

Abstractly, a section is a map $s: SU(2)/U(1)\to\mathcal{L}_k$ so that $\pi\circ s = Id$. This can be made much more explicit using *local trivializations*.

To motivate the description of sections in terms of local trivializations consider the transition we made above from homogenous functions of the (z_1,z_2) to polynomials in $z=z_1/z_2$.

If we think of $\mathbb{CP}^1$ as the space of equivalence classes $[z_1 : z_2]$ then on the "patch" $\mathcal{U}_N$ where $z_2 \neq 0$ we can define a map

$$\phi_N : \mathcal{U}_N \to \mathbb{C} \qquad \phi_N([z_1 : z_2]) := z_N := z_1/z_2 \tag{11.404}$$

Note that the point $[1 : 0]$ corresponds to a "point at infinity" in this mapping. In terms of the identification $S^2 \cong \mathbb{CP}^1$ this corresponds to stereographic projection from the north pole:

$$z_N = \frac{\hat{x}^1 + \mathrm{i}\hat{x}^2}{1 - \hat{x}^3} \tag{11.405}$$

In this language we identified a homogeneous polynomial of z_1, z_2 of degree $2j$ with a polynomial $p_N(z_N)$ by

$$p_N(z_N) = z_2^{-2j} F(z_1, z_2) \tag{11.406}$$

We could clearly have given an analogous discussion based on stereographic projection from the south pole:

$$z_S = \frac{\hat{x}^1 - \mathrm{i}\hat{x}^2}{1 + \hat{x}^3} \tag{11.407}$$

Note that

$$z_S z_N = 1 \tag{11.408}$$

In terms of $\mathbb{CP}^1$, we could introduce a 'patch" $\mathcal{U}_N$ where $z_1 \neq 0$ we can define a map

$$\phi_S : \mathcal{U}_S \to \mathbb{C} \qquad \phi_S([z_1 : z_2]) := z_S := z_2/z_1 \tag{11.409}$$

Note that the point $[0 : 1]$ corresponds to a "point at infinity" in this mapping. Again it is manifest that $z_N z_S = 1$ on the patch overlap $\mathcal{U}_S \cap \mathcal{U}_N$ where both functions are defined.

Now, the same homogeneous function F could be used to define a different polynomial

$$p_S(z_S) = z_1^{-2j} F(z_1, z_2) \tag{11.410}$$

Comparing the equations we see that on patch overlaps we can say:

$$p_S(z_S) = z_N^{-2j} p_N(z_N) \tag{11.411}$$

or equally well

$$p_N(z_N) = z_S^{-2j} p_S(z_S) \tag{11.412}$$

Equations (11.411) and (11.412) are known as gluing conditions and z_S^{-2j} and z_N^{-2j} are known as transition functions. We can glue together the total space of the line bundle $\mathcal{L}_k$ by taking two trivial line bundles

$$\mathcal{U}_S \times \mathbb{C} \tag{11.413}$$

$$\mathcal{U}_N \times \mathbb{C} \tag{11.414}$$

and identifying (z_S, v_S) with $(1/z_N, z_N^{-2j} v_N)$ on the patch overlaps.

Note the the function h is invariant under this description:

$$h(p_S, p_S) = h(p_N, p_N) \tag{11.415}$$

because, on patch overlaps we have:

$$(1+|z_S|^2)^{-2j}|p_S(z_S)|^2 = (1+|z_N|^2)^{-2j}|p_N(z_N)|^2 \tag{11.416}$$

as the reader should carefully check. Thus this is a globally defined notion of the length-square of a section. It is called a *Hermitian metric on the line bundle.*

Remarks:

1. Recalling that $2j = -k$ the transition rules for all sections of the line bundle $\mathcal{L}_k$ can be written

$$\psi_N(z_N, \bar{z}_N) = z_S^k \psi_S(z_S, \bar{z}_S) \tag{11.417}$$

where there is now no condition of holomorphy. The space of all sections can be identified with the space of all functions in $\mathrm{Ind}_{U(1)}^{SU(2)}(\rho_k)$. This formulation makes sense for any integer k. When $k > 0$ there are no holomorphic sections, but there are plenty of $\mathcal{C}^\infty$ sections.

2. *The Borel-Weil-Bott Theorem*: Here is a rough statement: If G is a compact group and T a maximal subgroup then one can give G/T a complex structure so that $G/T \cong G_{\mathbb{C}}/B$. Irreps of T induce reps of B and then the holomorphic sections of a line bundle of $G_{\mathbb{C}}/B$, corresponding to the equivalent holomorphic induced representation $Ind_B^{G_{\mathbb{C}}}(V)$ is an irreducible representation of G, and all such arise in this way. G_c/B for other finite dimensional simple Lie groups. ♣Improve this statement. ♣

3. We will see later that $SL(2,\mathbb{C})$ is isomorphic to the Lorentz group in 3+1 dimensions. The Lorentz group acts on the set of light rays through the origin of Minkowski space, and we can identify a light ray with its point on the celestial sphere. Under this identification, the action of the Lorentz group on the set of light rays is just the Mobius action on the sphere.

4. *Representations Of Lorentz, Poincaré, and Affine Euclidean Groups.* Also, the construction of the irreducible unitary representations of Lorentz groups, and affine Euclidean and Poincaré groups proceeds using this method. (That observation goes back to Wigner and Bargmann.) Briefly, for a representation of the Poincaré group one induces from a representation of the translation group (or the translation group semidirect product with a compact rotation group). For a representation of the Lorentz group one considers the homogeneous spaces from orbits of $SO(1,d)$ in momentum space. For example, the mass shell $p^2 = m^2$ with $p^0 > 0$ can be identified with $SO(1,d)/SO(d)$. Then one induces from a representation of $SO(d)$ to produce a unitary representation of $SO(1,d)$. One source that approaches the subject from the present viewpoint is M. Carmeli, *Group Theory And General Relativity.*

5. Loop Groups: See the book of Pressley and Segal

6. Diffeomorphism Groups: Nice paper of A. Alekseev and S. Shatashvili

7. *Geometric Quantization.* EXPLAIN A LITTLE

11.15 The Clebsch-Gordon Decomposition For $SU(2)$

In physics the spin j representation V_j shows up almost universally. Among other applications it appears in the quantum mechanical theory of spin. (See the relation to Lie algebras below for an explanation of this.)

The combination of two systems of spins naturally leads to the question of how to give an isotypical decomposition of $V_{j_1} \otimes V_{j_2}$. This is known as the Clebsch-Gordon decomposition. The general formula is:

$$V(j_1) \otimes V(j_2) \cong V(|j_1 - j_2|) \oplus V(|j_1 - j_2| + 1) \oplus \cdots\cdots \oplus V(j_1 + j_2) \tag{11.418}$$

Note that every representation on the RHS has the same parity of $(-1)^{2j}$.

Let us give a proof of (15.610) using characters. Because a representation is uniquely determined by its character we can consider the character of $V(j_1) \otimes V(j_2)$. If we can write this as a linear combination of characters χ_j with nonnegative integer coefficients we can uniquely determine the decomposition into irreps.

The easiest thing to do is prove that

$$V_{\frac{1}{2}} \otimes V_j \cong V_{j+1/2} \oplus V_{j-1/2} \tag{11.419}$$

If $j = 0$ we interpret $V_{-1/2}$ as the zero vector space. Since a representation is uniquely determined by its character we can prove this by simply computing

$$\begin{aligned} \chi_{1/2}(z)\chi_j(z) &= (z + z^{-1}) \frac{z^{2j+1} - z^{-2j-1}}{z - z^{-1}} \\ &= \frac{z^{2j+2} - z^{-2j-2}}{z - z^{-1}} + \frac{z^{2j} - z^{-2j}}{z - z^{-1}} \\ &= \chi_{j+1/2} + \chi_{j-1/2} \end{aligned} \tag{11.420}$$

The general result (15.610) now follows by induction.

Alternatively, we can write:

$$\begin{aligned} \chi_{j_1}(z)\chi_{j_2}(z) &= \left(\frac{z^{2j_1+1} - z^{-2j_1-1}}{z - z^{-1}} \right) \cdot \left(\frac{z^{2j_2+1} - z^{-2j_2-1}}{z - z^{-1}} \right) \\ &= \frac{1}{z - z^{-1}} \left(\frac{z^{2j_1+2j_2+2} - z^{2(j_1-j_2)}}{z - z^{-1}} + \frac{z^{-2j_1-2j_2-2} - z^{-2(j_1-j_2)}}{z - z^{-1}} \right) \end{aligned} \tag{11.421}$$

Now, WLOG assume that $j_1 \geq j_2$. Then we use the identity:

$$\frac{z^{a+2} - z^b}{z - z^{-1}} = z^{b+1} \frac{z^{a-b+2} - 1}{z^2 - 1} = z^{b+1} + z^{b+3} + \cdots + z^{a+1} \tag{11.422}$$

for each of the two terms in the sum above, then realize that the two terms are related by $z \to 1/z$ and we directly obtain:

$$\chi_{j_1}\chi_{j_2} = \chi_{j_1+j_2} + \chi_{j_1+j_2-1} + \cdots + \chi_{|j_1-j_2|} \tag{11.423}$$

It is also instructive to give a proof using the orthogonality of characters. One can check directly by contour integration that

$$-\frac{1}{4\pi \mathrm{i}}\oint \chi_{j_1}(z)\chi_{j_2}(z)\chi_j(z)(z-z^{-1})^2\frac{dz}{z} = \begin{cases} +1 & |j_1-j_2|\leq j\leq j_1+j_2 \qquad \& \qquad 2j = 2j_1+2j_2 \mathrm{mod} 2 \\ 0 & \text{else} \end{cases} \tag{11.424}$$

♣Explain this better ♣

Definition Choose an ON basis $\psi_{j,m}$ of V_j. Let P_j the the orthogonal projector onto the subspace of $V_{j_1}\otimes V_{j_2}$ transforming in the representation V_j. Then

$$\langle \psi_{j,m}, P_j(\psi_{j_1,m_1}\otimes\psi_{j_2,m_2})\rangle \tag{11.425}$$

is known as a *Clebsch-Gordon coefficient*. In physics they are denoted as: **********

Example: Let us consider in detail the important case of

$$V_{1/2}\otimes V_{1/2}\cong V_0\oplus V_1 \tag{11.426}$$

The orthogonal projectors are obtained from

$$P^j_{m_L,m_R} = \int_{SU(2)} (D^j_{m_L,m_R}(g))^* T(g) dg \tag{11.427}$$

In particular, as noted above in equation (11.180) the projector to the isotypical component of the trivial representation is always:

$$P = \int_G T(g)dg \tag{11.428}$$

In our example,

$$T(g)|\beta\rangle\otimes|\delta\rangle = g_{\alpha\beta}g_{\gamma\delta}|\alpha\rangle\otimes|\gamma\rangle \tag{11.429}$$

Tradition demands that an ON basis for the fundamental representation be denoted $\{|+\rangle,|-\rangle\}$ with

$$d(z)\cdot|+\rangle = z|+\rangle \qquad\qquad d(z)\cdot = z^{-1}|-\rangle \tag{11.430}$$

We denote the basis vectors $|\alpha\rangle$ with $\alpha,.... \in \{+,-\}$. If we look at the projector to the trivial representation then we need to evaluate

$$\int_{SU(2)} g_{\alpha\beta}g_{\gamma\delta}dg \tag{11.431}$$

Now recall that $g_{\alpha\beta} = \epsilon_{\alpha\alpha'}\epsilon_{\beta\beta'}(g^*)^{\alpha'\beta'}$ so we can use the orthogonality relations to say

$$\int_{SU(2)} g_{\alpha\beta}g_{\gamma\delta}dg = \frac{1}{2}\epsilon_{\alpha\gamma}\epsilon_{\beta\delta} \tag{11.432}$$

confirming the direct computation from the exercise (11.57).

It follows that an ON basis for the spin 0 ("singlet") isotypical component of $V_{1/2} \otimes V_{1/2}$ is

$$\frac{1}{\sqrt{2}} \left(|+\rangle \otimes |-\rangle - |-\rangle \otimes |+\rangle \right) \tag{11.433}$$

This state is also called a *Bell pair* in quantum information theory. It is the simplest example of a nontrivially entangled state and was used in the famous EPR paper - the first paper on quantum information theory.

The orthogonal complement must be the spin one states, and a basis that diagonalizes the action of the diagonal unitary matrices is

$$\begin{aligned} |1,1\rangle &= |+\rangle \otimes |+\rangle \\ |1,0\rangle &= \frac{1}{\sqrt{2}} \left(|+\rangle \otimes |-\rangle + |-\rangle \otimes |+\rangle \right) \\ |1,-1\rangle &= |-\rangle \otimes |-\rangle \end{aligned} \tag{11.434}$$

****

In physics a standard example of a general selection rule is the *Wigner-Eckart theorem*: Suppose $\mathcal{O}_{j,m}$ a collection of operators transforming in the spin j rep

$$\langle \psi_{j,m} | \mathcal{O}_{j_1,m_1} | \psi_{j_2,m_2} \rangle = CLEBSCHx" GroupIndependent" \tag{11.435}$$

EXPLAIN.

****

Exercise *Recovering Dimension*

As a nice check of the expression on the far RHS in (11.356) show that the limit $u \to 1$ reproduces the dimension of V_j.

11.16 Lie Groups And Lie Algebras And Lie Algebra Representations

11.16.1 Some Useful Formulae For Working With Exponentials Of Operators

In this section we will discuss some formulae that are very useful for working with exponentials of matrices (and linear operators). In particular we will derive the Baker-Campbell-Hausdorff formula.

Let us recall that if A is a matrix or an operator then e^A is the matrix, or operator, defined by the exponential series. The following three identities are easily shown by direct use of the exponential series:

1.
$$e^{\alpha A} e^{\beta A} = e^{(\alpha+\beta)A} \tag{11.436}$$

2.
$$\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A \tag{11.437}$$

3.

$$e^A e^B e^{-A} = e^{e^A B e^{-A}} \tag{11.438}$$

Now we prove some identities that are not directly obvious from the exponential series:

Definition: For $A \in M_n(\kappa)$ we denote by $\mathrm{Ad}(A)$ the linear transformation $M_n(\kappa) \to M_n(\kappa)$ defined by

$$\mathrm{Ad}(A) : B \mapsto [A, B] \tag{11.439}$$

We also denote:

$$(\mathrm{Ad}(A))^m B = \overbrace{[A, [A, \cdots [A, B] \cdots]}^{m \text{ times}} \tag{11.440}$$

where there are m commutators on the RHS.

First we prove

$$e^A B e^{-A} = e^{\mathrm{Ad}(A)} B \tag{11.441}$$

in other words:

$$\begin{aligned} e^A B e^{-A} &= e^{\mathrm{Ad}(A)} B \\ &= B + \mathrm{Ad}(A)B + \frac{1}{2!}(\mathrm{Ad}(A))^2 B + \cdots \\ &= B + [A, B] + \frac{1}{2!}[A, [A, B]] + \cdots \end{aligned} \tag{11.442}$$

To prove this define $B(t) := e^{tA} B e^{-tA}$. So $B(0) = B$ and $B(1) = e^A B e^{-A}$ is the quantity we want. Now it is easy to derive the differential equation:

$$\frac{d}{dt} B(t) = \mathrm{Ad}(A) B(t) \tag{11.443}$$

so

$$B(t) = e^{t\mathrm{Ad}(A)} B(0) \tag{11.444}$$

Now set $t = 1$.

Combining with (11.438) we now have the somewhat less trivial identity:

$$e^A e^B e^{-A} = e^{e^{\mathrm{Ad}(A)} B} \tag{11.445}$$

All these identities follow from a much more nontrivial formula, known as the Baker-Campbell-Hausdorff formula that expresses the operator C defined by

$$e^A e^B = e^C \tag{11.446}$$

as a power series in A, B. We have

$$C = A + B + s(A, B) \tag{11.447}$$

where $s(A, B)$ is an infinite series and every term involves nested commutators.

We will give a complete statement and proof of the BCH formula below. In order to do that we first state the extremely useful

Lemma : Let

$$f(z) = \frac{e^z - 1}{z} = 1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots \tag{11.448}$$

Then

$$\big(\frac{d}{dt}e^{A(t)}\big)e^{-A(t)} = -e^{A(t)}\frac{d}{dt}e^{-A(t)} = f(\mathrm{Ad}(A(t))) \cdot \dot{A}(t) \tag{11.449}$$

where $A(t)$ is any differentiable matrix function of t.

Note that this is nontrivial because $\dot{A}(t)$ does not commute with $A(t)$ in general! Indeed, from the exponential series you can easily show that

$$\begin{aligned} \frac{d}{dt}e^{A(t)} - \dot{A}(t)e^{A(t)} &= \frac{1}{2}[A(t), \dot{A}(t)] + \cdots \\ \frac{d}{dt}e^{A(t)} - e^{A(t)}\dot{A}(t) &= -\frac{1}{2}[A(t), \dot{A}(t)] + \cdots \end{aligned} \tag{11.450}$$

Proof: Introduce a matrix function of two variables and take derivatives wrt s:

$$\begin{aligned} B(s,t) &:= e^{sA(t)}\frac{d}{dt}e^{-sA(t)} \\ \frac{\partial B}{\partial s} &= A(t)e^{sA(t)}\frac{d}{dt}e^{-sA(t)} - e^{sA(t)}\frac{d}{dt}\big[e^{-sA(t)}A(t)\big] \\ &= \mathrm{Ad}(A(t))B(s,t) - \dot{A}(t) \\ \frac{\partial^j B}{\partial s_j} &= (\mathrm{Ad}(A(t)))^j B(s,t) - (\mathrm{Ad}A(t))^{j-1}\dot{A}(t) \end{aligned} \tag{11.451}$$

$B(0,t) = 0$ therefore again by Taylor:

$$\frac{1}{j!}\frac{\partial^j}{\partial s^j}B(s,t) \mid_{s=0} = -\mathrm{Ad}(A(t))^{j-1}\dot{A}(t) \qquad j \geq 1 \tag{11.452}$$

So

$$e^{sA(t)}\frac{d}{dt}(e^{-sA(t)}) = -\sum_{j=1}^{\infty}\frac{s^j(\mathrm{Ad}(A(t)))^{j-1}}{j!}\dot{A}(t) \tag{11.453}$$

Now set s=1. ♠

Note: you can rewrite this lemma as the statement:

$$\frac{d}{dt}e^{A(t)} = \int_0^1 e^{sA(t)}\dot{A}(t)e^{(1-s)A(t)}ds \tag{11.454}$$

because:

$$\begin{aligned}\frac{d}{dt}e^{A(t)} &= \int_0^1 e^{sA(t)}\dot{A}(t)e^{(1-s)A(t)}ds \\ &= \int_0^1 e^{sAd(A(t))}ds\dot{A}(t)e^{A(t)} \\ &= \left[\left(\frac{e^{Ad(A(t))}-1}{Ad(A(t))}\right)\dot{A}(t)\right]e^{A(t)}\end{aligned} \tag{11.455}$$

Remark: Equation (11.454) is an intuitively appealing formula. For a finite product we have:

$$\begin{aligned}\frac{d}{dt}\Big[M_1(t)M_2(t)M_3(t)\cdots M_n(t)\Big] &= (\frac{d}{dt}M_1(t))M_2(t)M_3(t)\cdots \\ &+ (M_1(t))(\frac{d}{dt}M_2(t))M_3(t)\cdots \\ &+ (M_1(t))(M_2(t))(\frac{d}{dt}M_3(t))\cdots \\ &+ \cdots + M_1(t)M_2(t)\cdots(\frac{d}{dt}M_n(t))\end{aligned} \tag{11.456}$$

Now write

$$e^{A(t)} = \prod_{i=1}^{N}[e^{A(t)\Delta s}] \tag{11.457}$$

where $\Delta s = 1/N$. By equation (11.450), we can replace $\frac{d}{dt}e^{\Delta sA(t)}$ by $e^{\Delta sA(t)}\Delta s\dot{A}(t)$ up to order $(\Delta s)^2$. Then write the general term in the sum (11.456) as

$$\left(\prod_{i<s}e^{\Delta sA(t)}\right)e^{\Delta sA(t)}\dot{A}(t)\left(\prod_{s<i}e^{\Delta sA(t)}\right)\Delta s \tag{11.458}$$

up to terms of order $(\Delta s)^2$. Next we sum over these terms and take $N\to\infty$ to get (11.454).

Now we are finally ready to state the main theorem:

Theorem: (Baker-Campbell-Hausdorff formula)
Let:

$$g(w) = \frac{\log w}{w-1} = \sum_{j=0}^{\infty}\frac{(1-w)^j}{j+1} = 1 + \frac{1-w}{2} + \frac{(1-w)^2}{3} + \cdots \tag{11.459}$$

be a power series in w about 1. Then when A, B are $n\times n$ matrices with $\| A \|$, $\| B \|$ sufficiently small, the matrix C given by the expansion:

$$\boxed{C = B + \int_0^1 g(e^{t\mathrm{Ad}A}e^{\mathrm{Ad}B})(A)dt} \tag{11.460}$$

satisfies $C = \log(e^A e^B)$.

Proof:

Introduce the matrix-valued function $C(t)$ via:

$$e^{C(t)} = e^{tA} e^B \tag{11.461}$$

and note that $C(0) = B$, and $C(1)$ is the matrix we want. We derive a differential equation for $C(t)$. By our lemma we have:

$$e^{C(t)} \frac{d}{dt} e^{-C(t)} = -f(\mathrm{Ad}C(t))\dot{C}(t) \tag{11.462}$$

with

$$f(z) = \frac{e^z - 1}{z} \tag{11.463}$$

On the other hand, plugging in the definition (11.461) we compute directly the simple result

$$e^{C(t)} \frac{d}{dt} e^{-C(t)} = e^{tA} \frac{d}{dt} e^{-tA} = -A \tag{11.464}$$

Therefore we get a differential equation:

$$f(\mathrm{Ad}C(t))\dot{C}(t) = A \tag{11.465}$$

Now, f is a power series about 1 so it immediately follows that

$$\dot{C}(t) = f(Ad(C(t)))^{-1} A \tag{11.466}$$

Let us make this more explicit: Using the power series $g(w)$ above with $w = e^z$ note that

$$f(z)g(e^z) = \frac{e^z - 1}{z} \cdot \frac{z}{e^z - 1} = 1 \tag{11.467}$$

regarded as an identity of power series in z. Now we can substitute for z any operator $\mathcal{O}$, and use

$$g(e^{\mathcal{O}}) = f(\mathcal{O})^{-1}, \tag{11.468}$$

and therefore we can solve for $\dot{C}$:

$$\begin{aligned} \dot{C}(t) &= f(\mathrm{Ad}(C(t)))^{-1} \cdot A \\ &= g(\exp(\mathrm{Ad}(C(t))) \cdot A \end{aligned} \tag{11.469}$$

where we applied (11.468) with $\mathcal{O} = \mathrm{Ad}(C(t)$. This hardly seems useful, since we still don't know $C(t)$, but now since we have power series we can say

$$e^{\mathcal{O}} = e^{Ad(C(t))} = e^{\mathrm{Ad}(tA)} e^{\mathrm{Ad}(B)} \tag{11.470}$$

To prove (11.470) note that for all H we have:

$$\begin{aligned} e^{\mathrm{Ad}C(t)}H &= e^{C(t)}He^{-C(t)} \\ &= e^{tA}e^{B}He^{-B}e^{-tA} \\ &= e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)}H \\ \Rightarrow e^{\mathrm{Ad}(C(t))} &= e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)} \end{aligned} \tag{11.471}$$

Therefore:

$$\dot{C}(t) = g(e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)}) \cdot A \tag{11.472}$$

Now we integrate equation (11.472)

$$C(t) = C(0) + \int_0^t g(e^{\mathrm{Ad}(sA)}e^{\mathrm{Ad}(B)})A ds \tag{11.473}$$

but $C(0) = B$, so

$$C = C(1) = \log(e^A e^B) = B + \int_0^1 g(e^{\mathrm{Ad}(s\ A)}e^{\mathrm{Ad}(B)})A\ ds \tag{11.474}$$

which is what we wanted to show. ♠.

Remarks:

1. To evaluate $g(e^{\mathcal{O}})$ for an operator $\mathcal{O}$ we expand $e^{\mathcal{O}}$ around 0 so $e^{\mathcal{O}} = 1 + \mathcal{O} + \cdots$ and then we expand around 1 to get an expansion of $g(e^{\mathcal{O}})$ around $\mathcal{O} = 0$. A similar remark applies to $g(e^{\mathcal{O}_1}e^{\mathcal{O}_2})$.

2. Explicitly the first few terms are: [119]

$$\boxed{C = A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \frac{1}{12}[B,[B,A]] + \frac{1}{24}[A,[B,[A,B]]] + \cdots} \tag{11.475}$$

   where the next terms are order ϵ^5 if we scale A, B by ϵ.[120]

3. For suitable operators A, B on Hilbert space the BCH formula continues to hold. But the series has a finite radius of convergence: See the exercises below.

Exercise
Work out the BCH series to order 5 in A, B.

---

[119] It is useful to note that $[A,[B,[A,B]]] = -[B,[A,[B,A]]] = B^2A^2 - A^2B^2$

[120] One can find an algorithm for generating the higher order terms in Varadarajan's book on group theory.

Exercise

Show that we can also write:

$$C = \log(e^A e^B) = A + \int_0^1 g(e^{-\mathrm{Ad}(s\ B)} e^{-\mathrm{Ad}(A)}) B\ ds \tag{11.476}$$

Exercise *All Orders In B, First Order in A*

Write $A = \epsilon$ and consider it to be small. Show that the formula for C given by BCH to all orders in B and first order in ϵ is

$$\begin{aligned} C &= B + \frac{\mathrm{Ad}B}{e^{\mathrm{Ad}B} - 1}(\epsilon) \\ &= B + \epsilon - \frac{1}{2}[B,\epsilon] + \frac{1}{12}[B,[B,\epsilon]] - \frac{1}{720}[B,[B,[B,[B,\epsilon]]]] + \cdots \end{aligned} \tag{11.477}$$

Note:

$$\begin{aligned} \frac{x}{e^x - 1} &= 1 - \frac{1}{2}x + \frac{1}{12}x^2 - \frac{x^4}{720} + \frac{x^6}{30240} - \frac{x^8}{1209600} + \frac{x^{10}}{47900160} \\ &\quad - \frac{691}{1307674368000}x^{12} + \cdots \\ &\equiv \sum_{n=0}^{\infty} \frac{B_n x^n}{n!} \end{aligned} \tag{11.478}$$

is an important expansion in classical function theory -the numbers B_n are known as the Bernoulli numbers

There are many applications of this formula. One in particle physics is to spontaneous symmetry breaking where the formula above gives the chiral transformation law of the pion field. Here $B = \pi(x)$ is the pion field and ϵ is the chiral transformation parameter.

Exercise *Eigenvalues of* $\mathrm{Ad}(A)$ *For Diagonalizable A*

a.) Show that if $A \in M_n(\kappa)$ is diagonalizable $A \sim \mathrm{Diag}\{\lambda_1, \ldots, \lambda_n\}$ then the eigenvalues of $\mathrm{Ad}(A)$ acting on $M_n(\kappa)$ are $\lambda_i - \lambda_j$.

b.) Using (a) and the previous exercise conclude that the BCH formula has finite radius of convergence.

11.16.2 Lie Algebras

We recall two basic definitions - the definition of an *algebra* and of a *Lie algebra*. See Chapter 2, Linear Algebra User's Manual for more discussion and examples. The relation between Lie groups and Lie algebras is covered in much more detail in Chapter 8. But here is a preview.

Definition An *algebra* over a field κ is a vector space A over κ with a notion of multiplication of two vectors

$$A \times A \to A \tag{11.479}$$

denoted:

$$a_1, a_2 \in A \to a_1 \odot a_2 \in A \tag{11.480}$$

which has a ring structure compatible with the scalar multiplication by the field. Concretely, this means we have axioms:

i.) $(a_1 + a_2) \odot a_3 = a_1 \odot a_3 + a_2 \odot a_3$

ii.) $a_1 \odot (a_2 + a_3) = a_1 \odot a_2 + a_1 \odot a_3$

iii.) $\alpha(a_1 \odot a_2) = (\alpha a_1) \odot a_2 = a_1 \odot (\alpha a_2), \qquad \forall \alpha \in \kappa.$

The product $\odot$ might, or might not, be associative. When it is associative, it is called an *associative algebra*.

Example: A basic example of an algebra is the vector space of $n \times n$ matrices over a field κ. The vector addition is simply addition of matrices. The algebra product $\odot$ is matrix multiplication.

Definition A *Lie algebra* over a field κ is an algebra A over κ where the multiplication of vectors $a_1, a_2 \in A$, satisfies in addition the two conditions:

1. $\forall a_1, a_2 \in A$:

$$a_2 \odot a_1 = -a_1 \odot a_2 \tag{11.481}$$

2. $\forall a_1, a_2, a_3 \in A$:

$$((a_1 \odot a_2) \odot a_3) + ((a_3 \odot a_1) \odot a_2) + ((a_2 \odot a_3) \odot a_1) = 0 \tag{11.482}$$

This is known as the *Jacobi relation*.

Now, tradition demands that the product on a Lie algebra be denoted not as $a_1 \odot a_2$ but rather as $[a_1, a_2]$ where it is usually referred to as the *bracket*. So then the two defining conditions (11.481) and (11.482) are written as:

1. $\forall a_1, a_2 \in A$:

$$[a_2, a_1] = -[a_1, a_2] \tag{11.483}$$

2. $\forall a_1, a_2, a_3 \in A$:

$$[[a_1, a_2], a_3] + [[a_3, a_1], a_2] + [[a_2, a_3], a_1] = 0 \tag{11.484}$$

Remark: Note that if we consider the Lie algebra product as the algebra product on a vector space, then the algebra is non-associative. If we have a Lie algebra product and use it to define an algebra product: $a_1 \odot a_2 := [a_1, a_2]$ then

$$\begin{aligned}(a_1 \odot a_2) \odot a_3 - a_1 \odot (a_2 \odot a_3) &= [[a_1, a_2], a_3] - [a_1, [a_2, a_3]] \\ &= -[[a_3, a_1], a_2]]\end{aligned} \tag{11.485}$$

by the Jacobi identity.

Example 1: Whenever we have an associative algebra A, we can automatically turn it into a Lie algebra by defining the Lie product as a commutator:

$$[a_1, a_2] := a_1 \odot a_2 - a_2 \odot a_1 \tag{11.486}$$

However, it should be stressed that not all Lie algebras arise in this way. A good example would be first order smooth differential operators on, say, $\mathbb{R}^n$, (or more generally, smooth vector fields on a manifold). If

$$V = v^\mu(x) \frac{\partial}{\partial x^\mu} \tag{11.487}$$

then the product of the differential operators - which makes sense as a second order operator - is not a vector field, but we can define

$$[V_1, V_2] = \left(v_1^\mu \frac{\partial}{\partial x^\mu} v_2^\nu - v_2^\mu \frac{\partial}{\partial x^\mu} v_1^\nu \right) \frac{\partial}{\partial x^\nu} \tag{11.488}$$

Example 2: An important example arises by applying the general remark above to the algebra $M_n(\kappa)$. In this case the Lie algebra is often denoted $\mathfrak{gl}(n, \kappa)$. Now $M_n(\kappa)$ and $\mathfrak{gl}(n, \kappa)$ are precisely the same as sets, and are precisely the same as vector spaces. The notation simply puts a different stress on the algebraic structure which the author is considering.

Example 3: If $\mathfrak{g} \subset \mathfrak{gl}(n, \kappa)$ is a vector subspace of matrices that is closed under matrix commutation, then $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{gl}(n, \kappa)$. For example, there is a vector subspace of $\mathfrak{gl}(n, \kappa)$ consisting of traceless matrices. The property of being traceless is preserved by commutator. Therefore this subspace is a sub-Lie algebra. It is denoted as $\mathfrak{sl}(n, \kappa)$. In the classification of semi-simple Lie algebras over $\mathbb{C}$ these Lie algebras have special names: $A_n := \mathfrak{sl}(n + 1, \kappa)$.

Example 4: Consider $\mathfrak{so}(n, \kappa) \subset M_n(\kappa)$ defined to be the vector subspace of $n \times n$ antisymmetric matrices. The matrix product of anti-symmetric matrices is not anti-symmetric, but the matrix commutator is. In the classification of semi-simple Lie algebras

over $\mathbb{C}$ these Lie algebras have special names, and for good reasons the even and odd cases are considered as separate families. They are denoted $B_n = \mathfrak{so}(2n+1)$ and $D_n = \mathfrak{so}(2n)$.

Example 5: $\mathfrak{u}(n) \subset M_n(\mathbb{C})$ is defined to be the vector subspace of $n \times n$ antihermitian matrices. The matrix product of anti-hermitian matrices is not anti-hermitian, but the matrix commutator is. Note carefully that $\mathfrak{u}(n)$ is a real vector space and is not a complex vector space! After all, if $A^\dagger = -A$ then $(\mathrm{i}A)^\dagger = \mathrm{i}A$ is not anti-hermitian. We can further define a sub-Lie algebra $\mathfrak{su}(n) \subset \mathfrak{u}(n)$ of anti-hermitian matrices that are, in a addition, traceless.

Example 6: Finally $\mathfrak{sp}(2n, \kappa) \subset M_n(\kappa)$ is the Lie subalgebra such that $(Ja)^{tr} = Ja$ where

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{11.489}$$

Example 7: The intersection of two Lie algebras is a Lie algebra and a particularly important example is $\mathfrak{usp}(2n) = \mathfrak{su}(2n) \cap \mathfrak{sp}(2n, \mathbb{C}) \subset M_{2n}(\mathbb{C})$. ♣say why ♣

Let $\mathfrak{g} \subset \mathfrak{gl}(n, \kappa)$ be a sub-Lie algebra with $\kappa = \mathbb{Q}, \mathbb{R}, \mathbb{C}$. The BCH formula gives one way to understand a relation between Lie algebras and Lie groups: Provided the BCH series converges it follows that if $A, B \in \mathfrak{g}$ then

$$e^A e^B = e^C \tag{11.490}$$

with $C \in \mathfrak{g}$. Thus, up to convergence issues, the invertible operators e^A with $A \in \mathfrak{g}$ will close to form a group. The series will converge for A, B in an open neighborhood of the origin. So, taking the closure (under group multiplication) of the set of matrices $\exp[A]$ with $A \in \mathfrak{g}$ will give the Lie group.

One can show that for compact connected Lie groups the exponential map

$$\exp : \mathfrak{g} \to G \tag{11.491}$$

is indeed surjective, although it will not be injective. A good example is the case $U(1)$. Here the Lie algebra is $\mathfrak{g} = \mathrm{i}\mathbb{R}$ (and the commutator is zero). Then all of $2\pi\mathrm{i}\mathbb{Z}$ is in the kernel of the exponential map.

Conversely, from a matrix Lie group one can recover the Lie algebra by considering the general one-parameter subgroups $g(t)$ with $g(0) = 1$ and computing $g^{-1}(t)\frac{d}{dt}g(t)$ at $t = 0$. We will elaborate on this idea by first closing a gap in our discussion at the very beginning of the course and prove that the classical Lie groups are manifolds as follows:

Exercise *Structure Constants*

If $\mathfrak{g}$ is a Lie algebra over κ then we can choose a basis T^i for $\mathfrak{g}$ and necessarily we have an expansion

$$[T^i, T^j] = \sum_k f_k^{ij} T^k \qquad\qquad f_k^{ij} \in \kappa \tag{11.492}$$

from which one can construct all commutators. The constants f_k^{ij} are known as *structure constants*.

a.) Show that

$$f_k^{ij} = -f_k^{ji} \tag{11.493}$$

$$f_\ell^{ij} f_m^{\ell k} + f_\ell^{jk} f_m^{\ell i} + f_\ell^{ki} f_m^{\ell j} = 0 \tag{11.494}$$

b.) Conversely, show that given tensors $f_k^{ij} \in \kappa$ satisfying equations (11.493) and (11.494) one can define a Lie algebra over κ.

Remark: Sometimes Lie algebras are presented by giving a list of structure constants. If someone tries to sell you a Lie algebra by giving you a list of structure constants don't buy it until you have checked equations (11.493) and (11.494).

11.16.3 The Classical Matrix Groups Are Lie Groups

This section assumes some knowledge of differential geometry. Some readers might wish to skip it.

First we recall the pre-image theorem. A map $f : M_1 \to M_2$ between two manifolds M_1 and M_2 is said to be a *submersion* if at every $p \in M_1$ the map of tangent spaces:

$$df : T_p M_1 \to T_{f(p)} M_2 \tag{11.495}$$

is surjective. In this case $f(p)$ is said to be a *regular value*. Using results from the calculus of many variables one can show that if f is a submersion at p then there are local coordinates so that in a neighborhood of p it has the form:

$$f : (x^1, \ldots, x^{n_1}) \mapsto (x^1, \ldots, x^{n_2}) \tag{11.496}$$

That is, locally, in suitable coordinate systems f is literally the map $f : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ keeping only the first n_2 coordinates. Then, every point in the target $\mathbb{R}^{n_2}$ is a regular value of f and the inverse image of a regular value is

$$f^{-1}(c^1, \ldots, c^{n_2}) = \{x \in \mathbb{R}^{n_1} | x^1 = c^1, \ldots, x^{n_2} = c^{n_2}\} \cong \mathbb{R}^{n_1 - n_2} \tag{11.497}$$

is a submanifold of dimension $n_1 - n_2$.

Therefore we have

Theorem: [Preimage Theorem] If $f : M_1 \to M_2$ and $q \in M_2$ is a regular value in the image of f then the preimage $f^{-1}(q)$ is a submanifold of M_1 of dimension $\dim M_1 - \dim M_2$. That

is, the preimage $f^{-1}(q)$ is a submanifold of M_1 of codimension $\dim M_2$. The tangent space to $f^{-1}(q)$ at any point p is $\ker(df_p)$.

Now we can apply this idea to describe subsets defined by equations. If $f : M \to \mathbb{R}^\ell$ and $\dim M \geq \ell$ then for $\vec{c} \in \mathbb{R}^\ell$ the sets

$$M_{\vec{c}} := f^{-1}(\vec{c}) = \cap_{i=1}^{\ell} \{p \in M | f^i(p) = c^i\} \tag{11.498}$$

are called level sets. If $\vec{c}$ is a regular value then the level set is a submanifold of M of codimension ℓ. Note that for each i, $f^i : M \to \mathbb{R}$ so $df^i : T_pM \to T_{c^i}\mathbb{R} \cong \mathbb{R}$, and hence df^i is a linear functional on T_pM. The regularity condition is the condition that these linear functionals are all linearly independent. We say that the functions f^i are *independent*.

We can apply these ideas to give easy proofs that the classical matrix groups are in fact Lie groups. First we prove they are all manifolds:

Example 1: $GL(n, \mathbb{R}) \subset \mathbb{R}^{n^2}$ and $GL(n, \mathbb{C}) \subset \mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$ are both manifolds. The coordinates are the matrix elements. The inverse image $\det^{-1}(0)$ is a closed subset of $M_n(\kappa)$, with $\kappa = \mathbb{R}, \mathbb{C}$ and hence any invertible matrix has an open neighborhood of invertible matrices. The matrix elements thus serve as global coordinates. Unless we insist that our coordinate patches are diffeomorphic to $\mathbb{R}^{n^2}$ there is no need to use more than one coordinate patch. The tangent space at any point $A \in GL(n, \kappa)$ is isomorphic to the vector space $M_n(\kappa)$ of $n \times n$ matrices over κ.

Example 2: Now, for $SL(n, \kappa)$ consider $f : M_n(\kappa) \to \kappa$ defined by $f(A) := \det A - 1$. We claim that $0 \in \kappa$ is a regular value of f. Indeed, if A is invertible then for any

$$M \in T_A M_n(\kappa) \cong M_n(\kappa) \tag{11.499}$$

we have

$$df_A(M) = \det A \mathrm{Tr}(A^{-1}M) \tag{11.500}$$

This is usually written as the (very useful) identity [121]

$$\delta \mathrm{logdet} A = \mathrm{Tr}\left(A^{-1}\delta A\right) \tag{11.501}$$

for A invertible. When A is invertible the kernel of df_A is the linear subspace of $n \times n$ matrices M such that $A^{-1}M$ is traceless, which is linearly equivalent to the linear subspace of traceless matrices, and therefore has dimension (over κ) equal to $n - 1$. Therefore the rank of df_A is 1, and $f = \det$ is a submersion. So the inverse image is a manifold.

Example 3: For $O(n; \kappa)$ we define $f : M_n(\kappa) \to S_n(\kappa)$ where $S_n(\kappa) \cong \kappa^{\frac{1}{2}n(n+1)}$ is the vector space over κ of $n \times n$ symmetric matrices. We take f to be

$$f(A) = AA^{tr} - 1 \tag{11.502}$$

[121] For a proof see the Linear Algebra User's Manual, ch. 3.

Then $O(n) = f^{-1}(0)$. We aim to show it is a manifold. Note that df_A is a linear operator $M_n(\kappa) \to S_n(\kappa)$. It is just

$$df_A(M) = MA^{tr} + AM^{tr} \tag{11.503}$$

Therefore $\ker(df_A)$ is the linear subspace of $M_n(\kappa)$ of matrices such that MA^{tr} is anti-symmetric. When A is invertible this subspace is isomorphic to the linear subspace of anti-symmetric matrices and hence has dimension $\frac{1}{2}n(n-1)$. It follows that 0 is a regular value of f and $O(n,\kappa)$ is a manifold.

Example 4: For $Sp(2n;\kappa)$ we define $f : M_{2n}(\kappa) \to A_{2n}(\kappa)$ where $A_{2n}(\kappa)$ is the set of $(2n) \times (2n)$ matrices over κ such that (Jm) is antisymmetric. This is isomorphic to the vector space over κ of dimension $\frac{1}{2}(2n)(2n-1) = n(2n-1)$. Now we take f to be

$$f(A) = AJA^{tr}J^{tr} - 1 \tag{11.504}$$

so that $Sp(2n;\kappa) = f^{-1}(0)$. Again we claim that 0 is a regular value of f. Now df_A is the linear operator $M_n(\kappa) \to A_{2n}(\kappa)$. It is just

$$df_A(M) = MJA^{tr}J^{tr} + AJM^{tr}J^{tr} \tag{11.505}$$

Therefore $\ker(df_A)$ is the linear subspace of $M_n(\kappa)$ of matrices such that MJA^{tr} is symmetric. When A is invertible this subspace is isomorphic to the linear subspace of symmetric matrices and hence has dimension $\frac{1}{2}2n(2n+1)$, which is complementary to the dimension of the image $\frac{1}{2}(2n)(2n-1)$ and hence df_A is surjective. It follows that 0 is a regular value of f and $Sp(2n;\kappa)$ is a manifold.

Example 5: Finally, for $U(n)$ consider $f : M_n(\mathbb{C}) \to \mathcal{H}_n$ where $\mathcal{H}_n$ is the *real* vector space of $n \times n$ Hermitian matrices in $M_n(\mathbb{C})$. This has real dimension $n + 2 \times \frac{1}{2}n(n-1) = n^2$. We now take $f(A) = AA^\dagger - 1$. Then

$$df_A(M) = MA^\dagger + AM^\dagger \tag{11.506}$$

When A is invertible the kernel is the subspace of $M_n(\mathbb{C})$ of matrices such that $MA^\dagger$ is anti-hermitian. This is again a real vector space of real dimension $n + 2 \times \frac{1}{2}n(n-1) = n^2$. Since $M_n(\mathbb{C})$ is a real vector space of real dimension $2n^2$ it follows that df_A is surjective and hence 0 is a regular value of f. Therefore $U(n)$ is a manifold.

Example 6: It is useful to combine the previous two examples and define the Lie group:

$$USp(2n) := U(2n) \cap Sp(2n,\mathbb{C}) \tag{11.507}$$

Now we have a map: $f : M_{2n}(\mathbb{C}) \to A_{2n}(\mathbb{C}) \oplus \mathcal{H}_{2n}$ defined by taking the direct sum. Again, one must check that the *real linear* map df_A at an invertible matrix in the preimage of 0 has a kernel of the correct dimension. [122]

[122] We have not covered quaternions yet, but a superior viewpoint is to view $USp(2n)$ as the group of $n \times n$ unitary matrices over the quaternions. In this viewpoint we should define $f : M_n(\mathbb{H}) \to \mathcal{H}_n(\mathbb{H})$

All the examples above are submanifolds of $GL(n,\kappa)$. The group operation of multiplication is polynomial in the matrix elements and hence certainly a $\mathcal{C}^\infty$ function. The group operation of inversion is a rational function of the matrix elements and is also $\mathcal{C}^\infty$ on $GL(n,\kappa)$. It follows from the exercise of Section §?? that the group operations of multiplication and inversion are $\mathcal{C}^\infty$ maps in all the cases of the matrix subgroups. This concludes the argument that the above examples are all Lie groups.

In general, the *Lie algebra* of a Lie group G is defined, as a vector space, to be the tangent space at the identity:

$$Lie(G) := T_1 G \tag{11.508}$$

As we will demonstrate, it is in fact a Lie algebra.

Our proof above that the classical matrix groups are manifolds also leads nicely to an immediate computation of the Lie algebras of these groups.

♣For consistency, use small gothic font for Lie algebras ♣

1. $GL(n,\kappa)$:

$$T_1 GL(n,\kappa) \cong M_n(\kappa) = \mathfrak{gl}(n,\kappa) \tag{11.509}$$

2. $SL(n,\kappa)$:

$$T_1 SL(n,\kappa) \cong \{M \in M_n(\kappa) | \mathrm{Tr}(M) = 0\} = \mathfrak{sl}(n,\kappa) \tag{11.510}$$

3. $O(n,\kappa)$:

$$\mathrm{so}(n;\kappa) := T_1 O(n,\kappa) \cong \{M \in M_n(\kappa) | M^{tr} = -M\} = \mathfrak{o}(n,\kappa) = \mathfrak{so}(n,\kappa) \tag{11.511}$$

4. $Sp(2n,\kappa)$:

$$T_1 Sp(2n,\kappa) \cong \{M \in M_{2n}(\kappa) | (MJ)^{tr} = +MJ\} = \mathfrak{sp}(2n,\kappa) \tag{11.512}$$

5. $U(n)$:

$$T_1 U(n) \cong \{M \in M_n(\mathbb{C}) | M^\dagger = -M\} = \mathfrak{u}(n) \tag{11.513}$$

6. $SU(n)$:

$$T_1 SU(n) \cong \{M \in M_n(\mathbb{C}) | M^\dagger = -M \quad \& \quad \mathrm{Tr}(M) = 0\} = \mathfrak{su}(n) \tag{11.514}$$

7. $USp(2n)$:

$$\begin{aligned} T_1 USp(2n) &\cong \{M \in M_{2n}(\mathbb{C}) | M^\dagger = -M \quad \& \quad (MJ)^{tr} = MJ\} \\ &\cong \{M \in M_n(\mathbb{H}) | M^\dagger = -M\} := \mathfrak{usp}(2n) \end{aligned} \tag{11.515}$$

---

where $\mathcal{H}_n(\mathbb{H})$ is the space of $n \times n$ quaternionic Hermitian matrices. The above arguments work in the same way: $\dim_{\mathbb{R}} M_n(\mathbb{H}) = 4n^2$, while $\dim_{\mathbb{R}} \mathcal{H}_n(\mathbb{H}) = n + 4\frac{1}{2}n(n-1) = 2n^2 - n$. As before, the kernel of df_A, for A invertible is the space of $n \times n$ quaternionic-antihermitian matrices. This has real dimension $3n + 4\frac{1}{2}n(n-1) = 2n^2 + n$. (The $3n$ is there because one can have an arbitrary imaginary quaternion on the diagonal.) Therefore 0 is a regular value, and $USp(2n)$ is a manifold. Many authors denote this group simply as $Sp(n)$.

We now recognize that the above vector spaces of matrices are in fact the Lie algebras we introduced earlier. We can get back the groups by exponentiation. It is a good exercise to check, from the defining relations of T_1G above that the exponentiated matrix indeed satisfies the defining relations of the group. Thus, for example, one should check that if M is anti-Hermitian, i.e. if $M^\dagger = -M$ then $\exp(M)$ is unitary.

We can argue more generally that they must be Lie algebras as follows: Given a matrix $V \in T_1G \subset M_N(\kappa)$ we can form the family of group elements

$$g_V(t) = \exp[tV] := \sum_{n=0}^{\infty} \frac{(tV)^n}{n!} \tag{11.516}$$

Note that for $t \in \mathbb{R}$ these elements form a subgroup of G:

$$g_V(t_1)g_V(t_2) = g_V(t_1 + t_2) \tag{11.517}$$

In general, in differential geometry, there is an exponential map from the tangent space T_pM of a manifold M to a neighborhood of p in M. It is not as explicit as an exponential series of a matrix, but involves using the vector to define a differential equation. An important property of the tangent spaces T_1G for the various groups above is that:

If $V_1, V_2 \in T_1G$ then the matrix commutator $[V_1, V_2]$ is also in T_1G.

This can be verified by directly checking each case. For example, in the case of $so(n, \kappa)$, if V_1, V_2 are antisymmetric matrices over κ then neither V_1V_2, nor V_2V_1 is antisymmetric, but $[V_1, V_2]$ is. The reader should check the other cases in this way. Nevertheless, this fact also follows from more general principles, and that is important because as we will see not every Lie group is a classical matrix group. In fact, it is not true that every finite-dimensional Lie group is a subgroup of $GL(N, \mathbb{R})$ for some N. [123] Given V_1, V_2 we can consider the path through $g = 1$ at $t = 0$ given by the group commutator:

$$\lambda(t) = [g_{V_1}(\sqrt{t}), g_{V_2}(\sqrt{t})] \tag{11.518}$$

Now, usin the BCH formula one can show that for t_1, t_2 small we have

$$g_{V_1}(t_1)g_{V_2}(t_2) = \exp[t_1V_1 + t_2V_2 + \frac{1}{2}t_1t_2[V_1, V_2] + \mathcal{O}(t_1^a t_2^b)] \tag{11.519}$$

where the higher order terms have $a + b > 2$, and therefore the tangent vector to the path through $\lambda(t)$ is the matrix commutator.

Therefore, just based on group theory and manifold theory, one can deduce that the tangent space at the identity $\mathfrak{g} = T_1G$ is indeed a Lie algebra.

One of the main theorems about the relation of Lie groups and Lie algebras is the following:

Theorem:

[123] A counterexample is the metaplectic group, a group which arises as a central extension of the symplectic group when one tries to implement symplectic transformations on a the quantum mechanics of a system of free particles.

a.) Every finite dimensional Lie algebra $\mathfrak{g}$ over $\kappa = \mathbb{R}$ arises from a unique (up to isomorphism) connected and simply connected Lie group G.

b.) Under this correspondence, Lie group homomorphisms $f : G_1 \to G_2$ are in $1-1$ correspondence with Lie algebra homomorphisms $\mu : T_1G_1 \to T_1G_2$.

For more about this see Chapter 8. The best statement makes use of the language of categories. What is described here is an equivalence of categories. See Section **** below.

Examples:

1. $\mathfrak{su}(2)$ has a standard basis
$$T^a := -\frac{\mathrm{i}}{2}\sigma^a \tag{11.520}$$
with structure constants
$$[T^a, T^b] = \epsilon^{abc}T^c \tag{11.521}$$
Note that any traceless anti-Hermitian matrix can be diagonalized: For any $A \in \mathfrak{su}(2)$ there is a $u \in SU(2)$ with
$$u^{-1}Au = \mathrm{i}\lambda\sigma^3 \tag{11.522}$$
for some $\lambda \in \mathbb{R}$. Note that it follows that the exponentiation of any one-parameter subgroup is subgroup of $SU(2)$ isomorphic to $U(1)$. In particular, it is compact. Finally, as we have seen, every $SU(2)$ group element can be written as
$$g = \cos\chi + \mathrm{i}\sin\chi\hat{n}\cdot\vec{\sigma} = \exp[\mathrm{i}\chi\hat{n}\cdot\vec{\sigma}] \tag{11.523}$$
where $\hat{n} \in S^2 \subset \mathbb{R}^3$. So the exponential map is surjective.

2. $\mathfrak{sl}(2,\mathbb{R})$ has a standard basis:
$$e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad h = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad f = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \tag{11.524}$$
You should check the structure constants, taking careful note of signs:
$$[h,e] = -2e \qquad [e,f] = h \qquad [h,f] = +2f \tag{11.525}$$
Note that $\mathfrak{sl}(2,\mathbb{R})$ is qualitatively different: The elements $e, f \in \mathfrak{sl}(2,\mathbb{R})$ have non-trivial Jordan form and cannot be diagonalized (even within $GL(2,\mathbb{C})$) So this Lie algebra is inequivalent to $\mathfrak{su}(2)$. It is not hard to show that if $A \in \mathfrak{sl}(2,\mathbb{R})$ then it is conjugate (via $SL(2,\mathbb{R})$ conjugation) to one of three distinct forms
$$SAS^{-1} = xe = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \qquad x \in \mathbb{R} \tag{11.526}$$
$$SAS^{-1} = xh = \begin{pmatrix} -x & 0 \\ 0 & x \end{pmatrix} \qquad x \in \mathbb{R} \tag{11.527}$$

$$SAS^{-1} = x(e-f) = \begin{pmatrix} 0 & x \\ -x & 0 \end{pmatrix} \qquad x \in \mathbb{R} \tag{11.528}$$

and hence there are three distinct maximal Abelian subgroups (up to conjugation):

$$\exp[xe] = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \qquad x \in \mathbb{R} \tag{11.529}$$

$$\exp[xh] = \begin{pmatrix} e^{-x} & 0 \\ 0 & e^{x} \end{pmatrix} \qquad x \in \mathbb{R} \tag{11.530}$$

$$\exp[x(e+f)] = \begin{pmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{pmatrix} \qquad x \in \mathbb{R} \tag{11.531}$$

Note that the last subgroup is compact and is just a copy of $SO(2,\mathbb{R})$ and $x \sim x + 2\pi$ parametrize the same group element.

It is easy to show these subgroups are not conjugate by considering the trace - see the exercise below. It is now easy to see that the exponential map cannot be surjective onto $SL(2,\mathbb{R})$. Consider the group elements

$$\begin{pmatrix} -1 & x \\ 0 & -1 \end{pmatrix} \qquad x \neq 0 \tag{11.532}$$

If this were of the form $\exp[A]$ then it would be in the one-parameter group $\exp[tA]$. But the trace is -2 and the only way one of the above one-parameter groups can have trace $= -2$ is to take (11.531) with $x = \pi(2n+1)$ with $n \in \mathbb{Z}$. Nevertheless, via the Gram-Schmidt procedure (see chapter 2) one can prove the so-called *KAN decomposition*: Every $g \in SL(2,\mathbb{R})$ can be uniquely written in the form

$$g = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \cdot \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \tag{11.533}$$

with $\lambda > 0$ and $x \in \mathbb{R}$. So the products of the one-parameter groups generate all group elements.

Remark: *The group manifold $SL(2,\mathbb{R})$ as anti-deSitter space:* Every 2×2 real matrix can plainly be written as:

$$g = \begin{pmatrix} T_1 + X_1 & X_2 + T_2 \\ X_2 - T_2 & T_1 - X_1 \end{pmatrix} \tag{11.534}$$

for real variables T_1, T_2, X_1, X_2. Restricting to matrices with $\det g = 1$ gives the hyperboloid in $\mathbb{R}^{2,2}$:

$$-T_1^2 - T_2^2 + X_1^2 + X_2^2 = -1 \tag{11.535}$$

Thus one picture of $SL(2,\mathbb{R})$ can be given as a hyperboloid in $\mathbb{R}^{2,2}$. The induced metric has signature $(-^1,+^2)$. Indeed, a global set of coordinates is:

$$\begin{aligned} T_1 &= \cosh\rho\cos t \\ T_2 &= \cosh\rho\sin t \\ X_1 &= \sinh\rho\cos\phi \\ X_2 &= \sinh\rho\sin\phi \end{aligned} \tag{11.536}$$

These coordinates smoothly cover the manifold once for $0 \le \rho < \infty$, $t \sim t + 2\pi$, $\phi \sim \phi + 2\pi$. Substituting into

$$ds^2 = -(dT_1)^2 - (dT_2)^2 + (dX_1)^2 + (dX_2)^2 \tag{11.537}$$

the induced metric becomes

$$ds^2 = -\cosh^2\rho dt^2 + \sinh^2\rho d\phi^2 + d\rho^2. \tag{11.538}$$

This is one form of the *anti-deSitter metric* of constant curvature -1. Note that the time coordinate t is periodic. What is usually meant by anti-deSitter space is the universal cover of $SL(2,\mathbb{R})$.

Exercise *Due Diligence*
Show that if matrices a_1, a_2 satisfy $(Ja)^{tr} = Ja$ then $[a_1, a_2]$ has the same property.

Exercise *Group Commutators And Lie Algebra Commutators*
a.) Use the BCH theorem to show that if

$$g_1 = e^{t_1 A_1}, \quad g_2 = e^{t_2 A_2} \tag{11.539}$$

the *group commutator*, $g_1 g_2 g_1^{-1} g_2^{-1}$ corresponds to the Lie algebra commutator:

$$g_1 \cdot g_2 \cdot g_1^{-1} \cdot g_2^{-1} = 1 + t_1 t_2 [A_1, A_2] + \mathcal{O}(t_1^2, t_2^2) \tag{11.540}$$

b.) We say a Lie algebra is "Abelian" if $[A_1, A_2] = 0$ for all $A_1, A_2 \in \mathfrak{g}$. That that such a Lie algebra exponentiates to form an Abelian group.
c.) If $A_i = \frac{d}{dt}|_0 g_i(t)$ then $[A_1, A_2]$ is the Lie algebra element associated to the curve

$$g_{12}(t) = g_1(\sqrt{t}) \cdot g_2(\sqrt{t}) \cdot g_1^{-1}(\sqrt{t}) \cdot g_2^{-1}(\sqrt{t}) \tag{11.541}$$

Exercise *Inequivalent One-Parameter Subgroups Of* $SL(2,\mathbb{R})$

a.) Show that the three maximal Abelian subgroups (11.529), (11.530), (11.531) are non-conjugate. [124]

b.) Consider the one-parameter subgroup $\exp[xf]$. Is it conjugate to one of the above three?

11.16.4 Representations Of Lie Algebras

A *representation of a Lie algebra* is a linear map

$$\rho : \mathfrak{g} \to \mathrm{End}(V) \tag{11.542}$$

for some vector space V (again sometimes called the *carrier space*) such that

$$[\rho(x), \rho(y)] = \rho([x,y]) \tag{11.543}$$

for all $x, y \in \mathfrak{g}$.

Note that any Lie algebra $\mathfrak{g}$ has a canonical representation with $V = \mathfrak{g}$ and

$$\dot{\rho}(x) : y \mapsto \mathrm{Ad}(x)(y) = [x,y] \tag{11.544}$$

This follows because, as one easily checks, the equation $[\dot{\rho}(x), \dot{\rho}(y)] = \dot{\rho}([x,y])$ is equivalent to the Jacobi identity. This representation is known as the *adjoint representation.*

Given a representation $\dot{\rho}$ of a Lie algebra $\mathfrak{g}$ we get a representation of a corresponding (connected) Lie group G so that $\mathfrak{g} = T_1 G$. Recall a *representation of a Lie group* is a group homomorphism

$$\rho : G \to GL(V) \tag{11.545}$$

We do this by setting [125]

$$\rho(e^x) := e^{\dot{\rho}(x)} \tag{11.546}$$

For $\mathfrak{g} = \mathfrak{gl}(n,\kappa)$ the group representation associated to the adjoint representation has carrier space $\mathfrak{g}$ and the group action is:

$$\rho(g)(x) = gxg^{-1} \tag{11.547}$$

Conversely, given a representation ρ of a Lie group we can define a representation of the Lie algebra by

$$\dot{\rho}(X) = \frac{d}{dt}|_0 \rho(e^{tX}) \tag{11.548}$$

Remarks:

[124] *Answer*: Consider the traces of these matrices. They are $= 2$, ≥ 2, and ≤ 2, respectively.

[125] We are assuming there is a corresponding Lie group. Some Lie algebras can, in fact, not be exponentiated to form Lie groups.

1. *Adjoint Representation Of* $\mathfrak{su}(2)$ *And Rotations.* We have already seen this in a slightly different form when we defined the natural double-cover homomorphism $\pi : SU(2) \to SO(3)$. We can identify $\mathfrak{su}(2) \cong \mathbb{R}^3$ as a vector space by identifying $\vec{x} \in \mathbb{R}^3$ with

$$M_{\vec{x}} := \mathrm{i}\vec{x} \cdot \vec{\sigma} \tag{11.549}$$

The adjoint representation is $\rho_{adj}(u) \cdot M = uMu^{-1}$, so ρ_{adj} acts linearly on $\vec{x} \in \mathbb{R}^3$. Note that $M_{\vec{x}}^2 = -\vec{x}^2 1_{2\times 2}$, so this linear action preserves the norm and therefore $\rho_{adj}(u) \in O(3)$. Since

$$\mathrm{tr}\,(M_{\vec{x}_1} M_{\vec{x}_2} M_{\vec{x}_3}) = 2\mathrm{i}\vec{x}_1 \cdot (\vec{x}_2 \times \vec{x}_3) \tag{11.550}$$

is also preserved $\rho_{adj}(u) \in SO(3)$.

2. *Adjoint Representation Of* $\mathfrak{sl}(2,\mathbb{R})$ *And Lorentz Transformations* Every element of $\mathfrak{sl}(2,\mathbb{R})$ can be written as

$$M_{x^\mu} = \begin{pmatrix} x & y-t \\ y+t & -x \end{pmatrix} = (y-t)e + xh + (y+t)f \tag{11.551}$$

Note $g \in SL(2,\mathbb{R})$ acts by $\rho_{adj}(g) \cdot M = gMg^{-1}$. But now

$$M_{x^\mu}^2 = (-t^2 + x^2 + y^2) 1_{2\times 2} \tag{11.552}$$

So, in a manner similar to the previous example $SL(2,\mathbb{R})$ double covers the connected component of the $2+1$ dimensional Lorentz group $SO_0(1,2)$.

3. $\mathfrak{sl}(2,\mathbb{C})$ *And Lorentz Transformations* For completeness we note that 2×2 Hermitian matrices can be identified with $3+1$ dimensional Minkowski space $\mathbb{M}^{1,3}$ via

$$M_{x^\mu} = x^0 1_{2\times 2} + \vec{x} \cdot \vec{\sigma} \tag{11.553}$$

Note $g \in SL(2,\mathbb{C})$ acts linearly via $\rho(g) \cdot M = gMg^\dagger$ since this preserves Hermiticity. However

$$\det M_{x^\mu} = (x^0)^2 - \vec{x}^2 \tag{11.554}$$

and the determinant is preserved by this actions so $g \mapsto \rho(g)$ describes $SL(2,\mathbb{C})$ as a double cover of the connected component $SO_0(1,3)$ of the $3+1$ dimensional Lorentz group.

4. $\mathfrak{su}(2) \oplus \mathfrak{su}(2)$ *And Rotations In* $\mathbb{R}^4$ Finally, a very similar construction gives the double cover $\pi : SU(2) \times SU(2) \to SO(4)$. But this is best discussed in the context of the *quaternions*. See the section on quaternions in Chapter 2.

Exercise *Another Proof That* $T_1 GL(n,\kappa) = \mathfrak{gl}(n,\kappa)$

Show that for $G = GL(n, \kappa)$ the corresponding Lie algebra is $\mathfrak{gl}(n, \kappa)$ by considering the 1-parameter subgroups $\exp[te_{ij}]$ where e_{ij} are the matrix units.

Exercise

Interpret equation (11.441) as a special case of (11.546) for the adjoint representation. Use this to derive (11.470) from the group homomorphism property of ρ.

Exercise *Tensor Products Of Representations*

We have noted that if $\rho_1 : G \to \mathrm{Aut}(V_1)$ and $\rho_2 : G \to \mathrm{Aut}(V_2)$ are two representations of any group G then there is a tensor product representation

$$\rho_{12}(g) := \rho_1(g) \otimes \rho_2(g) \tag{11.555}$$

Show that if G is a Lie gorup then the corresponding Lie algebra representation on $V_1 \otimes V_2$ representations $X \in \mathfrak{g} = T_1 G$ by

$$\dot{\rho}_{12}(X) = \dot{\rho}_1(X) \otimes 1_{V_2} + 1_{V_1} \otimes \dot{\rho}_2(X) \tag{11.556}$$

11.16.5 Finite Dimensional Irreducible Representations Of $\mathfrak{sl}(2,\mathbb{C})$,$\mathfrak{sl}(2,\mathbb{R})$, and $\mathfrak{su}(2)$

Relation of $\mathfrak{sl}(2,\mathbb{R})$, $\mathfrak{su}(2)$ *and* $\mathfrak{sl}(2,\mathbb{C})$:

If we regard $\mathfrak{sl}(2,\mathbb{C})$ as a Lie algebra over $\kappa = \mathbb{R}$ there are two inequivalent real Lie subalgebras: $\mathfrak{sl}(2,\mathbb{R})$ and $\mathfrak{su}(2)$, as we have described in detail above. However, if we allow ourselves to multiply by complex numbers, that is, if we consider $\mathfrak{sl}(2,\mathbb{R}) \otimes \mathbb{C}$ and $\mathfrak{su}(2) \otimes \mathbb{C}$ then we obtain a single Lie algebra $\mathfrak{sl}(2,\mathbb{C})$. Indeed, in the complexification, we have

$$\begin{aligned} e &= \mathrm{i}T^1 - T^2 \\ h &= -2\mathrm{i}T^3 \\ f &= -\mathrm{i}T^1 - T^2 \end{aligned} \tag{11.557}$$

Therefore, there is no distinction between the finite-dimensional representations of $\mathfrak{su}(2)$, $\mathfrak{sl}(2,\mathbb{R})$, and $\mathfrak{sl}(2,\mathbb{C})$ on complex vector spaces. It is actually easiest to construct the finite dimensional representation of $\mathfrak{sl}(2,\mathbb{R})$ or $\mathfrak{sl}(2,\mathbb{C})$ on a complex vector space. But by the above formulae these will immediately give the finite dimensional representations of $\mathfrak{su}(2)$.

Suppose we have a finite-dimensional complex vector space V and linear operators $\rho(e), \rho(f), \rho(h)$ on V satisfying the commutation relations:

$$\begin{aligned} [\rho(h), \rho(f)] &= 2\rho(f) \\ [\rho(h), \rho(e)] &= -2\rho(e) \\ [\rho(e), \rho(f)] &= \rho(h) \end{aligned} \tag{11.558}$$

As shown in Chapter two, any linear operator on a complex vector space has at least one eigenvector. (It might have only one eigenvector.)

Suppose we choose an eigenvector v of $\rho(h)$ and suppose the eigenvalue is λ. Then, we claim that, so long as $\rho(e)^n v \neq 0$ the vector $\rho(e)^n v$ has eigenvalue $\lambda - 2n$. To prove this apply the general identity

$$[A, B^n] = \sum_{i=0}^{n-1} B^i [A, B] B^{n-1-i} \tag{11.559}$$

to conclude:

$$[\rho(h), \rho(e)^n] = -2n\rho(e)^n \tag{11.560}$$

and the result follows. Now, it is general fact of linear algebra that if we have nonzero vectors $v_1, \ldots, v_n$ with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ for some operator then the $v_1, \ldots, v_n$ are linearly independent. (Prove this as an exercise.) Therefore, if $\rho(e)^n v \neq 0$ then the vectors $v, \rho(e)v, \ldots, \rho(e)^n v$ are linearly independent. Therefore, since we have a finite-dimensional representation there must be a nonnegative integer n so that $\rho(e)^n v \neq 0$ but $\rho(e)^{n+1} v = 0$. Let us denote $v_0 := \rho(e)^n v$. So $\rho(e)v_0 = 0$ with $\rho(h)v_0 = \lambda_0 v_0$ and $v_0 \neq 0$. Now, using (11.559) again we get:

$$[\rho(h), \rho(f)^k] = 2k\rho(f)^k \tag{11.561}$$

and therefore

$$\rho(h)(\rho(f)^k v_0) = (\lambda_0 + 2k)\rho(f)^k v_0 \tag{11.562}$$

By a similar argument to that above, we know that if $\rho(f)^n v_0$ is nonzero then the vectors $v_0, \rho(f)v_0, \ldots, \rho(f)^n v_0$ are linearly independent, and therefore there must exist an integer N so that $\rho(f)^N v_0 \neq 0$ but $\rho(f)^{N+1} v_0 = 0$. Therefore

$$[\rho(e), \rho(f)^{N+1}] v_0 = 0 \tag{11.563}$$

Now, using

$$[\rho(e), \rho(f)^{N+1}] = \sum_{i=0}^{N} \rho(f)^i [\rho(e), \rho(f)] \rho(f)^{N-i} = \sum_{i=0}^{N} \rho(f)^i \rho(h) \rho(f)^{N-i} \tag{11.564}$$

and applying this identity to v_0 we get:

$$0 = \left(\sum_{i=0}^{N} (\lambda_0 + 2(N-i)) \right) \rho(f)^N v_0 \tag{11.565}$$

But $\rho(f)^N v_0 \neq 0$, by the definition of N so

$$\left(\sum_{i=0}^{N}(\lambda_0 + 2(N-i))\right) = 0 \tag{11.566}$$

which implies $\lambda_0 = -N$. Thus, we have produced a set of vectors

$$v_0, \rho(f)v_0, \cdots, \rho(f)^N v_0 = v_0, v_1, \ldots, v_N \tag{11.567}$$

spanning an $(N+1)$-dimensional representation of $\mathfrak{sl}(2,\mathbb{R})$ inside any finite-dimensional representation. Indeed, if we define:

$$\tilde{v}_s := \rho(f)^s v_0 \tag{11.568}$$

understanding that $v_{N+1} = v_{N+2} = \cdots = 0$ then

$$\begin{aligned} \rho(f)\tilde{v}_s &= \tilde{v}_{s+1} \\ \rho(h)\tilde{v}_s &= (2s - N)\tilde{v}_s \\ \rho(e)\tilde{v}_s &= -s(N+1-s)\tilde{v}_{s-1} \end{aligned} \tag{11.569}$$

So the span, which we will denote $W_N \subset V$ is a subrepresentation. Note that in this ordered basis for W_N the representation matrices are: (See chapter 2 for the proper way to associate a matrix to a linear transformation on a basis with an ordered basis.)

$$\rho(f) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{11.570}$$

$$\rho(h) = \begin{pmatrix} -N & 0 & 0 & \cdots & 0 & 0 \\ 0 & -N+2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -N+4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & N-2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & N \end{pmatrix} \tag{11.571}$$

$$\rho(e) = \begin{pmatrix} 0 & -N & 0 & \cdots & 0 & 0 \\ 0 & 0 & -2(N-1) & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -N \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{11.572}$$

Note that the Jordan form for $\rho(f)$ shows that the representation W_N is irreducible.

As mentioned above, we automatically get a representation of $\mathfrak{su}(2)$, and hence of the group $SU(2)$. Therefore V is completely reducible. Therefore, we have recovered - this time at the Lie algebra level - that there is one irreducible representation of $SU(2)$ of dimension n for each positive integer n. In fact, identifying $N = 2j$ the representation W_N gives another model for the spin j representation V_j.

Remarks

1. In physics the *angular momentum* operators in quantum mechanics generate the Lie algebra $\mathfrak{su}(2)$. See the exercise below. In physics observable quantities (like angular momentum) are represented by Hermitian operators. The observable angular momentum operators are related to the anti-Hermitian generators of $\mathfrak{su}(2)$ by a factor of $\sqrt{-1}$. Physicists usually define

$$J^a := \mathrm{i}T^a = \frac{1}{2}\sigma^a \tag{11.573}$$

and they satisfy the commutation relations

$$[J^a, J^b] = \mathrm{i}\epsilon^{abc}J^c \tag{11.574}$$

which, unfortunately, obscures the fact that we are working with a real Lie algebra. Group elements are of the form $\exp[i\vec{\theta}\cdot\vec{J}]$ where $\vec{\theta}$ is real. When working with representations physicists generally define

$$\begin{aligned} J^+ &= J^1 + \mathrm{i}J^2 = e \\ J^- &= J^1 - \mathrm{i}J^2 = -f \end{aligned} \tag{11.575}$$

so that

$$\begin{aligned} [J^3, J^\pm] &= \pm J^\pm \\ [J^+, J^-] &= 2J^3 \end{aligned} \tag{11.576}$$

More generally, in a physical system with $SU(2)$ symmetry the J^a are the conserved "Noether charges" for that symmetry. Recall that we called the irreducible representations of $SU(2)$ of dimension $n = 2j+1$ the "spin j representations." This is the origin of that terminology.

2. V_j is a unitary representation of $\mathfrak{su}(2)$ and $SU(2)$ but is definitely <u>not</u> unitary as a representation of $\mathfrak{sl}(2,\mathbb{R})$. (There are unitary irreps of $\mathfrak{sl}(2,\mathbb{R})$, but they are infinite-dimensional.) The standard physics notation for an ON basis diagonalizing $\rho(J^3)$ is $|j,m\rangle$ with

$$\begin{aligned} \rho(J^+)|j,m\rangle &= \sqrt{(j-m)(j+m+1)}|j,m+1\rangle \\ \rho(J^3)|j,m\rangle &= m|j,m\rangle \\ \rho(J^-)|j,m\rangle &= \sqrt{(j+m)(j-m+1)}|j,m-1\rangle \end{aligned} \tag{11.577}$$

with $m = -j, -j+1, \ldots, j-1, j$. We can relate this to our basis by identifying $N = 2j$ and

$$\rho(f)^s v_0 = C_{j,s}|j, j-s\rangle \tag{11.578}$$

for a suitable normalization factor $C_{j,s}$.

Exercise *Justifying The Relation To Angular Momentum*

In the classical mechanics of a particle moving in $\mathbb{R}^3$ the angular momentum of a particle around the origin is the function on phase space given by

$$L_a = \epsilon_{abc} x_b p_c \qquad\qquad a, b, c \in \{1, 2, 3\} \tag{11.579}$$

L_a is a (pseudo)-vector. In quantum mechanics we adopt the same expression and there is no issue of operator ordering because the epsilon symbol prevents $b = c$:

$$\hat{L}_a = \epsilon_{abc} \hat{x}_b \hat{p}_c \qquad\qquad a, b, c \in \{1, 2, 3\} \tag{11.580}$$

Using $[\hat{p}_a, \hat{x}_b] = -\mathrm{i}\hbar\delta_{ab}$ show that

$$[\hat{L}_a, \hat{L}_b] = \mathrm{i}\epsilon_{abc}\hbar\hat{L}_c \tag{11.581}$$

so that $\hat{T}^a = -\frac{\mathrm{i}}{\hbar}\hat{L}_a$ are anti-Hermitian operators generating a copy of $\mathfrak{su}(2)$.

Remark: In spin j representations if j is order 1 the physical angular momentum is of order $\hbar$ and hence intrinsically quantum mechanical. On the other hand, for fixed $\hat{L}$, the semiclassical $\hbar \to 0$ limit is the large spin limit.

Exercise *Another basis*

Show that if we use the basis

$$v_s = \frac{1}{s!}\rho(f)^s v_0 \tag{11.582}$$

then

$$\begin{aligned} \rho(f)v_s &= (s+1)v_{s+1} \\ \rho(h)v_s &= (2s - N)v_s \\ \rho(e)v_s &= (s - N - 1)v_{s-1} \end{aligned} \tag{11.583}$$

Exercise *Unitarizing V_j*

Unitarizing V_j means equipping it with an inner product so that the $\rho(J^a)$ are Hermitian operators. Equivalently, $\rho(h)$ is Hermitian and

$$\rho(J^+)^\dagger = \rho(J^-) \qquad \leftrightarrow \qquad \rho(e)^\dagger = -\rho(f) \tag{11.584}$$

Show that by rescaling $u_s = c_s \tilde{v}_s$ and declaring u_s to be ON we get a unitary structure provided that

$$\left(\frac{c_{s-1}}{c_s}\right)^2 = s(2j+1-s) \tag{11.585}$$

so that

♣Fix this equation: sign issue ♣

$$\begin{aligned} \rho(e)u_s &= -\sqrt{s(2j+1-s)}u_{s-1} \\ \rho(f)u_s &= \sqrt{s(2j+1-s)}u_{s+1} \end{aligned} \tag{11.586}$$

11.16.6 Casimirs

We have stressed that if two elements a, b in a Lie subalgebra of a matrix Lie algebra are multiplied as matrices ab then in general the result is not in the Lie algebra. Nevertheless, if we have a representation $\rho(a), \rho(b) \in \mathrm{End}(V)$ nothing stops us form multiplying the operators $\rho(a)\rho(b)$. Certain relations among the algebra of the operators $\rho(a)$ for $a \in \mathfrak{g}$ are universal and independent of the representation. They can be expressed in terms of tensor algebra using the *universal enveloping algebra*. See Chapter two for a description. Here we just look at one important aspect of such universal relations.

In any representation of $\mathfrak{su}(2)$ the operator

$$C_2(\rho) := \sum_{a=1}^{3} \rho(T^a)^2 \tag{11.587}$$

commutes with all the operators $\rho(T^a)$. This operator is known as a quadratic Casimir. It is a theorem in the theory of universal enveloping algebras that any representation, any operator that commutes with all operators made by multiplying and adding the $\rho(T^a)$ is a polynomial in the operator $C_2(\rho)$. This fact generalizes to all simple Lie algebras: The center of the universal enveloping algebra is a polynomial in the Casimirs and there are r independent Casimirs where r is the rank. For $SU(N)$ there are $N-1$ independent Casimirs.

Returning to $SU(2)$, we can express $C_2(\rho)$ in terms of the representations of the basis for $\mathfrak{sl}(2,\mathbb{R})$ using

$$C_2(\rho) = \frac{1}{4}\left(2(\rho(e)\rho(f) + \rho(f)\rho(e)) - \rho(h)^2\right) \tag{11.588}$$

In any irreducible representation the Casimir operator must be a multiple of the identity operator. We can easily compute the value by acting on any convenient vector. For example,

$$\begin{aligned} C_2(\rho)v_0 &= \frac{1}{4}\left(2(\rho(e)\rho(f) + \rho(f)\rho(e)) - \rho(h)^2\right)v_0 \\ &= -\frac{N(N+2)}{4}v_0 \end{aligned} \tag{11.589}$$

which the physicists will prefer to write as

$$\rho_j(\vec{J})^2 = j(j+1)\mathbf{1}_{(2j+1)\times(2j+1)} \tag{11.590}$$

where ρ_j is the representation on V_j and we have merely substituted $N = 2j$ above.

Examples:

1. Consider a quantum system of two spin 1/2 particles with Hamiltonian:

$$H = J_1^a \otimes J_2^a \tag{11.591}$$

We can easily use representation theory to find the spectrum of this Hamiltonian. We note that

$$H = \frac{1}{2}\left((J_1^a \otimes 1 + 1 \otimes J_2^a)^2 - \vec{J}_1^2 \otimes 1 - 1 \otimes \vec{J}_2^2\right) \tag{11.592}$$

But the Hilbert space is $V_{1/2} \otimes V_{1/2} \cong V_0 \oplus V_1$. On the one-dimensional subspace $\cong V_0$ we use (11.590) to compute

$$H|_{V_0} = \frac{1}{2}(0 - \frac{3}{4} - \frac{3}{4}) = -\frac{3}{4} \tag{11.593}$$

But on the three-dimensional subspace $\cong V_1$ we have

$$H|_{V_1} = \frac{1}{2}(2 - \frac{3}{4} - \frac{3}{4}) = +\frac{1}{4} \tag{11.594}$$

Exercise *Three Qbits On A Ring*

Consider a quantum system of three spin 1/2 particles on a ring with $\vec{J}_i \cdot J_{i+1}$ interaction between neighboring spins. Compute the spectrum of the Hamiltonian.

$$H = \sum_i J_i^a \otimes J_{i+1}^a \tag{11.595}$$

where i is understood as an index modulo 3. [126]

---

[126] *Answer*: We can write:

$$H = \frac{1}{2}\left((J_1^a \otimes 1 \otimes 1 + 1 \otimes J_2^a \otimes 1 + 1 \otimes 1 \otimes J_3^a)^2 - \vec{J}_1^2 \otimes 1 \otimes 1 - 1 \otimes \vec{J}_2^2 \otimes 1 - 1 \otimes 1 \otimes \vec{J}_3^2\right) \tag{11.596}$$

But we have the isotypical decomposition:

$$V_{1/2} \otimes V_{1/2} \otimes V_{1/2} \cong 2V_{1/2} \oplus V_{3/2} \tag{11.597}$$

On the 4 dimensional space $2V_{1/2}$ we have

$$H|_{2V_{1/2}} = -\frac{3}{4} \tag{11.598}$$

and on the 4-dimensional space $V_{3/2}$ we have

$$H|_{V_{3/2}} = +\frac{3}{4} \tag{11.599}$$

11.16.7 Lie Algebra Operators In The Induced Representations Of $SU(2)$

Using the above models for the irreducible representations of $SU(2)$ and the relation between elements of the Lie algebra and infinitesimal group elements we get representations of the Lie algebra in terms of differential operators.

In terms of homogeneous polynomials $\psi(u,v)$ of two variables in the space $\mathcal{H}_{2j}$ recall that

$$(\rho(g)\cdot\psi)(\begin{pmatrix}u\\v\end{pmatrix}) := \psi\left(g^{-1}\begin{pmatrix}u\\v\end{pmatrix}\right) \tag{11.600}$$

Now suppose g is infinitesimally close to the identity so that

$$(\rho(1+\epsilon X)\cdot\psi)(\begin{pmatrix}u\\v\end{pmatrix}) := \psi\left((1-\epsilon X)\begin{pmatrix}u\\v\end{pmatrix}\right) \tag{11.601}$$

up to order ϵ^2. Since $\rho(X)$ is linear in X we can form complex linear combinations, and therefore the same formula will apply to any $X\in\mathfrak{sl}(2,\mathbb{C})$. In particular, we deduce:

$$\begin{aligned}\rho(e)\cdot\psi &= -v\frac{\partial}{\partial u}\psi\\ \rho(h)\cdot\psi &= \left(u\frac{\partial}{\partial u}-v\frac{\partial}{\partial v}\right)\psi\\ \rho(f)\cdot\psi &= u\frac{\partial}{\partial v}\psi\end{aligned} \tag{11.602}$$

The reader should check the operators really satisfy $\rho([X,Y])=[\rho(X),\rho(Y)]$. Note that the Casimir is:

$$C_2(\rho) = -\frac{1}{4}\left(u\frac{\partial}{\partial u}+v\frac{\partial}{\partial v}\right)^2 - \frac{1}{2}\left(u\frac{\partial}{\partial u}+v\frac{\partial}{\partial v}\right) \tag{11.603}$$

so acting on $\mathcal{H}_{2j}$ we immediately get that $C_2(\rho_j)$ acts as the scalar operator of multiplication by $j(j+1)$.

It is also interesting to consider the action in the inhomogeneous representation. As above, we extend the action to all of $SL(2,\mathbb{C})$ so that if

$$g = \begin{pmatrix}a & b\\ c & d\end{pmatrix} \in SL(2,\mathbb{C}) \tag{11.604}$$

then

$$\rho(g)\cdot p(z) = (-cz+a)^{2j}p\left(\frac{dz-b}{-cz+a}\right) \tag{11.605}$$

and if $g=1+\epsilon X+\mathcal{O}(\epsilon^2)$ then $\rho(g)=1+\epsilon\rho(X)+\mathcal{O}(\epsilon^2)$, and in this way we derive:

$$\begin{aligned}\rho(e)\cdot p &= -\frac{\partial}{\partial z}p\\ \rho(h)\cdot p &= -(2z\frac{\partial}{\partial z}+2j)p\\ \rho(f)\cdot p &= -(z^2\frac{\partial}{\partial z}-2jz)p\end{aligned} \tag{11.606}$$

For example:

$$\begin{aligned}(\rho(1+\epsilon f)\cdot p)(z) &:= (\epsilon z+1)^{2j}p\left(\frac{z}{\epsilon z+1}\right)\\ &= (1+2j\epsilon z)p(z-\epsilon z^2)+\mathcal{O}(\epsilon^2)\\ &= p(z)-\epsilon z^2\frac{\partial}{\partial z}p(z)+\epsilon 2jzp(z)+\mathcal{O}(\epsilon^2)\end{aligned} \tag{11.607}$$

Similarly, acting on the Wigner functions themselves we learn that, after complexification $\epsilon\rho(e)$ translates the group element in the argument of $D^j_{m_L,m_R}$ by

$$g=\begin{pmatrix}\alpha & -\bar\beta\\ \beta & \bar\alpha\end{pmatrix}\to g-\begin{pmatrix}\epsilon\beta & \epsilon\bar\alpha\\ 0 & 0\end{pmatrix} \tag{11.608}$$

and similarly $\epsilon\rho(f)$ translates it by

$$g=\begin{pmatrix}\alpha & -\bar\beta\\ \beta & \bar\alpha\end{pmatrix}\to g+\begin{pmatrix}0 & 0\\ \epsilon\alpha & -\epsilon\bar\beta\end{pmatrix} \tag{11.609}$$

So that, acting on the Wigner functions considered as polynomials in $\alpha,\bar\alpha,\beta,\bar\beta$ if we treat $\alpha,\bar\alpha,\beta,\bar\beta$ as independent then

$$\begin{aligned}\rho(e) &= \bar\alpha\frac{\partial}{\partial\bar\beta}-\beta\frac{\partial}{\partial\alpha}\\ \rho(h) &= \bar\alpha\frac{\partial}{\partial\bar\alpha}-\alpha\frac{\partial}{\partial\alpha}-\bar\beta\frac{\partial}{\partial\bar\beta}+\beta\frac{\partial}{\partial\beta}\\ \rho(f) &= \alpha\frac{\partial}{\partial\beta}-\bar\beta\frac{\partial}{\partial\bar\alpha}\end{aligned} \tag{11.610}$$

One can check explicitly that

$$\begin{aligned}\rho(e)\tilde D^j_{m_L,m_R} &= const.\tilde D^j_{m_L+1,m_R}\\ \rho(h)\tilde D^j_{m_L,m_R} &= \tilde D^j_{m_L,m_R}\\ \rho(f)\tilde D^j_{m_L,m_R} &= const.\tilde D^j_{m_L-1,m_R}\end{aligned} \tag{11.611}$$

We can specialize this to get the standard identities on spherical harmonics which are frequently used in mathematical physics: [127]

♣NEED TO GIVE MORE DETAILS ON THIS LAST REMARK ♣

$$\begin{aligned}L_+ &= e^{\mathrm{i}\phi}\left(\frac{\partial}{\partial\theta}+\mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\right)\\ L_- &= e^{-\mathrm{i}\phi}\left(-\frac{\partial}{\partial\theta}+\mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\right)\\ L_3 &= -\mathrm{i}\frac{\partial}{\partial\phi}\end{aligned} \tag{11.612}$$

[127] See, for example, J.D. Jackson, *Classical Electrodynamics*, page 743.

$$
\begin{aligned}
L_+ Y_{\ell,m} &= \sqrt{(\ell-m)(\ell+m+1)}\, Y_{\ell,m+1} \\
L_- Y_{\ell,m} &= \sqrt{(\ell+m)(\ell-m+1)}\, Y_{\ell,m-1} \\
L_3 Y_{\ell,m} &= m Y_{\ell,m} \\
\vec{L}^2 Y_{\ell,m} &= \ell(\ell+1) Y_{\ell,m}
\end{aligned}
\tag{11.613}
$$

In particular $\vec{L}^2$, the Casimir, as a differential shows up in the expression for the Laplacian expressed in terms of spherical coordinates. The equation $\vec{L}^2 Y_{\ell,m} = \ell(\ell+1) Y_{\ell,m}$ becomes the standard differential equation satisfied by associated Legendre functions and Legendre polynomials, whereas the first two lines of (11.613) become standard identities relating associated Legendre functions.

Remark: We have seen above that group theory gives a nice perspective on many identities satisfied by the family of special functions associated with Legendre polynomials and spherical harmonics. This viewpoint extends very nicely to many other special functions in mathematical physics. Two references that explain this in some detail are:

1. J.D. Talman, *Special Functions: A Group Theoretic Approach Based on Lectures by Eugene P. Wigner*

2. N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*

♣The following section on Kernel, Image, Exact Sequence should be moved to be just before the representation theory section, but the quantum mechanics on the circle should be moved either to the Heisenberg section or to the Pontryagin duality section. ♣

12. Group Theory And Elementary Number Theory

In this chapter we review some very elementary number theory that has a strong connection to group theory. The facts here can be very useful in thinking about many physics problems.

Two general references are

Hardy and Wright, *An Introduction To The Theory Of Numbers*

Ireland and Rosen, *A Classical Introduction to Modern Number Theory*

12.1 Reminder On gcd And The Euclidean Algorithm

Let us recall some basic facts from grade school arithmetic:

First, if $A > B$ are two positive integers then we can write

$$
A = qB + r \qquad 0 \leq r < B \tag{12.1}
$$

for unique nonnegative integers q and r known as the *quotient* and the *residue*, respectively.

Next, let $(A, B) = (\pm A, \pm B) = (\pm B, \pm A)$ denote the greatest common divisor of A, B. Then we can find it using the *Euclidean algorithm* by looking at successive quotients. If

$A = qB$ with $r = 0$ we are done! Then $(A, B) = B$. If $r > 0$ then we proceed as follows:

$$\begin{aligned} A &= q_1 B + r_1 \qquad 0 < r_1 < B \\ B &= q_2 r_1 + r_2 \qquad 0 < r_2 < r_1 \\ r_1 &= q_3 r_2 + r_3 \qquad 0 < r_3 < r_2 \\ r_2 &= q_4 r_3 + r_4 \qquad 0 < r_4 < r_3 \\ &\vdots \quad \vdots \\ r_{j-2} &= q_j r_{j-1} + r_j \qquad 0 < r_j < r_{j-1} \\ r_{j-1} &= q_{j+1} r_j \end{aligned} \tag{12.2}$$

Note that $B > r_1 > r_2 > \cdots \geq 0$ is a strictly decreasing sequence of nonnegative integers and hence must terminate at $r_* = 0$ after a finite number of steps.

Examples

$A = 96$ and $B = 17$:

$$\begin{aligned} 96 &= 5 \cdot 17 + 11 \\ 17 &= 1 \cdot 11 + 6 \\ 11 &= 1 \cdot 6 + 5 \\ 6 &= 1 \cdot 5 + 1 \\ 5 &= 5 \cdot 1 \end{aligned} \tag{12.3}$$

$A = 96$ and $B = 27$:

$$\begin{aligned} 96 &= 3 \cdot 27 + 15 \\ 27 &= 1 \cdot 15 + 12 \\ 15 &= 1 \cdot 12 + 3 \\ 12 &= 4 \cdot 3 \end{aligned} \tag{12.4}$$

Note well: In (12.1) the remainder might be zero but in the first j lines of the Euclidean algorithm the remainder is positive, unless B divides A, in which case rather trivially $(A, B) = B$. The last positive remainder r_j is the gcd (A, B). Indeed if m_1, m_2 are integers then the gcd satisfies:

$$(m_1, m_2) = (m_2, m_1) = (m_2, m_1 - x m_2) \tag{12.5}$$

for any integer x. Applying this to the Euclidean algorithm above we get:

$$(A, B) = (B, r_1) = (r_1, r_2) = \cdots = (r_{j-1}, r_j) = (r_j, 0) = r_j. \tag{12.6}$$

A corollary of this algorithm is that if $g = (A, B)$ is the greatest common divisor then there exist integers (x, y) so that

$$Ax + By = g \tag{12.7}$$

In particular, two integers m_1, m_2 are *relatively prime*, that is, have no common integral divisors other than ± 1, if and only if there exist integers x, y such that

$$m_1 x + m_2 y = 1. \tag{12.8}$$

Of course x, y are not unique. Equation (12.8) is sometimes known as "Bezout's theorem." We can prove these statements from the Euclidean algorithm as follows.

♣Putting this discussion here makes part of the section on $SL(2,\mathbb{Z})$ and continued fractions a little redundant. ♣

For an integer n define

$$T(n) := \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = T^n \tag{12.9}$$

where

$$T := T(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} . \tag{12.10}$$

Now let us write the first line of the Euclidean algorithm as a matrix identity as

$$T(-q_1) \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} r_1 \\ B \end{pmatrix} \tag{12.11}$$

and better, we write this as:

$$\sigma^1 T(-q_1) \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} B \\ r_1 \end{pmatrix} \tag{12.12}$$

Then the second line of the Euclidean algorithm becomes:

$$\sigma^1 T(-q_2)\sigma^1 T(-q_1) \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \tag{12.13}$$

Thus we have

$$\sigma^1 T(-q_j) \cdots \sigma^1 T(-q_2)\sigma^1 T(-q_1) \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} r_{j-1} \\ r_j \end{pmatrix} \tag{12.14}$$

and in the final step:

$$\sigma^1 T(-q_{j+1})\sigma^1 T(-q_j) \cdots \sigma^1 T(-q_2)\sigma^1 T(-q_1) \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} r_j \\ 0 \end{pmatrix} \tag{12.15}$$

Multiplying out the matrices on the LHS gives an expression:

$$\begin{pmatrix} x & y \\ u & v \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} r_j \\ 0 \end{pmatrix} \tag{12.16}$$

(Note that x, y, u, v are polynomials in the q_i. See comments on continued fractions below.)

Remarks:

1. *The Euclidean algorithm is fast*: A theorem of Lamé asserts that the Euclidean algorithm is very efficient. It should be completely obvious to you that the number of steps cannot exceed B. (Recall that $A > B$.) However, Lamé asserts that in fact the number of steps never exceeds $5\log_{10} B$. This is important for RSA (see below).

2. *Relation to continued fractions*: Note that from equation (12.15) we can also write

$$\begin{pmatrix} A \\ B \end{pmatrix} = T(q_1)\sigma^1 T(q_2)\sigma^1 \cdots T(q_j)\sigma^1 T(q_{j+1})\sigma^1 \begin{pmatrix} r_j \\ 0 \end{pmatrix} \tag{12.17}$$

Let us write:

$$M(q) := T(q)\sigma^1 = \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \tag{12.18}$$

We now define two sequences of polynomials in n variables that we call $N_n(q_1, \ldots, q_n)$ and $D_n(q_1, \ldots, q_n)$ for all $n \geq 1$. It is convenient to define $N_0 = 1$ and $D_0 = 1$ and then we can write:

$$M(q_1) \cdots M(q_n) := \begin{pmatrix} N_n(q_1, \ldots, q_n) & N_{n-1}(q_1, \ldots, q_{n-1}) \\ D_n(q_1, \ldots, q_n) & D_{n-1}(q_1, \ldots, q_{n-1}) \end{pmatrix} \tag{12.19}$$

(The reader should check that this is a consistent definition for all n.) One easily generates:

$$\begin{aligned} N_1(q_1) &= q_1 \\ N_2(q_1, q_2) &= 1 + q_1 q_2 \\ N_3(q_1, q_2, q_3) &= q_1 + q_3 + q_1 q_2 q_3 \end{aligned} \tag{12.20}$$

$$\begin{aligned} D_1(q_1) &= 1 \\ D_2(q_1, q_2) &= q_2 \\ D_3(q_1, q_2, q_3) &= 1 + q_2 q_3 \end{aligned} \tag{12.21}$$

These polynomials are closely related to continued fractions, defined as:

$$[q_1, q_2, q_3, \cdots, q_j] := q_1 + \frac{1}{q_2 + \frac{1}{q_3 + \cdots + \frac{1}{q_j}}} \tag{12.22}$$

Indeed, now that

$$M(q_1) \cdot (M(q_2) \cdots M(q_{n+1})) = M(q_1) \begin{pmatrix} N_n(q_2, \ldots, q_{n+1}) & N_{n-1}(q_2, \ldots, q_n) \\ D_n(q_2, \ldots, q_{n+1}) & D_{n-1}(q_2, \ldots, q_n) \end{pmatrix} \tag{12.23}$$

from which one deduces the recursion relations:

$$\begin{aligned} N_{n+1}(q_1, \ldots, q_{n+1}) &= q_1 N_n(q_2, \ldots, q_{n+1}) + D_n(q_2, \ldots, q_{n+1}) \\ D_{n+1}(q_1, \ldots, q_{n+1}) &= N_n(q_2, \ldots, q_{n+1}) \end{aligned} \tag{12.24}$$

On the other hand, writing

$$[q_1, q_2, q_3, \cdots, q_n] := \frac{P_n(q_1, \ldots, q_n)}{Q_n(q_1, \ldots, q_n)} \tag{12.25}$$

we see that

$$\begin{aligned}[q_1, q_2, q_3, \cdots, q_{n+1}] &= q_1 + \frac{1}{[q_2, \ldots, q_{n+1}]} \\ &= \frac{q_1 P_n(q_2, \ldots, q_{n+1}) + Q_n(q_2, \ldots, q_{n+1})}{P_n(q_2, \ldots, q_{n+1})}\end{aligned} \tag{12.26}$$

So, P_n, Q_n satisfy the same recursion relations as N_n, D_n, respectively, and since the initial values are also the same we conclude that $P_n = N_n$ and $Q_n = D_n$.

Exercise
Check the Lamé bound for the two examples above.

Exercise
Given one solution for (12.7), find all the others.

Exercise *Continued fractions and the Euclidean algorithm*

a.) Show that the quotients q_i in the Euclidean algorithm define a continued fraction expansion for A/B:

$$\frac{A}{B} = q_1 + \frac{1}{q_2 + \frac{1}{q_3 + \cdots + \frac{1}{q_j}}} := [q_1, q_2, q_3, \cdots, q_j] \tag{12.27}$$

The fractions $[q_1], [q_1, q_2], [q_1, q_2, q_3], \ldots$ are known as the *convergents* of the continued fraction.

b.) Show that [128]

$$\begin{aligned}N_{n+1}(q_1, \ldots, q_{n+1}) &= q_{n+1} N_n(q_1, \ldots, q_n) + N_{n-1}(q_1, \ldots, q_{n-1}) \\ D_{n+1}(q_1, \ldots, q_{n+1}) &= q_{n+1} D_n(q_1, \ldots, q_n) + D_{n-1}(q_1, \ldots, q_{n-1})\end{aligned} \tag{12.28}$$

c.) Show that

$$N_n D_{n-1} - D_n N_{n-1} = (-1)^n \tag{12.29}$$

---

[128] *Answer*: Write $M(q_1) \cdots M(q_{n+1}) = (M(q_1) \cdots M(q_n)) \cdot M(q_{n+1})$.

12.2 Application: Expressing elements of $SL(2,\mathbb{Z})$ as words in S and T

The group $SL(2,\mathbb{Z})$ is generated by

$$S := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \& \qquad T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{12.30}$$

Here is an algorithm for decomposing an arbitrary element

$$h = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in SL(2,\mathbb{Z}) \tag{12.31}$$

as a word in S and T.

First, note the following simple

Lemma Suppose $h \in SL(2,\mathbb{Z})$ as in (12.31). Suppose moreover that $g \in SL(2,\mathbb{Z})$ satisfies:

$$g \cdot \begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{12.32}$$

Then

$$gh = T^n \tag{12.33}$$

for some integer $n \in \mathbb{Z}$.

The proof is almost immediate by combining the criterion that $gh \in SL(2,\mathbb{Z})$ has determinant one and yet must have the first column $(1,0)$.

Now, suppose h is such that $A > C > 0$. Then $(A,C) = 1$ and hence we have the Euclidean algorithm to define integers q_ℓ, $\ell = 1, \dots N+1$, where $N \geq 1$, such that

$$\begin{aligned} A &= q_1 C + r_1 \qquad 0 < r_1 < C \\ C &= q_2 r_1 + r_2 \qquad 0 < r_2 < r_1 \\ r_1 &= q_3 r_2 + r_3 \qquad 0 < r_3 < r_2 \\ &\vdots \quad \vdots \\ r_{N-2} &= q_N r_{N-1} + r_N \qquad 0 < r_N < r_{N-1} \\ r_{N-1} &= q_{N+1} r_N \end{aligned} \tag{12.34}$$

with $r_N = (A,C) = 1$. (Note you can interpret $r_0 = C$, as is necessary if $N = 1$.) Now, write the first line in the Euclidean algorithm in matrix form as:

♣$N = 0$ here? ♣

$$\begin{pmatrix} 1 & -q_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} r_1 \\ C \end{pmatrix} \tag{12.35}$$

We would like to have the equation in a form that we can iterate the algorithm, so we need the larger integer on top. Therefore, rewrite the identity as:

$$\sigma^1 \begin{pmatrix} 1 & -q_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} C \\ r_1 \end{pmatrix} \tag{12.36}$$

We can now iterate the procedure. So the Euclidean algorithm implies the matrix identity:

$$\tilde{g}\begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{12.37}$$

$$\tilde{g} = (\sigma^1 T^{-q_{N+1}}) \cdots (\sigma^1 T^{-q_1}) \tag{12.38}$$

Now, to apply the Lemma we need g to be in $SL(2,\mathbb{Z})$, but

$$\det \tilde{g} = (-1)^{N+1} \tag{12.39}$$

We can easily modify the equation to obtained a desired element g. We divide the argument into two cases:

1. Suppose first that $N+1 = 2s$ is even. Then we group the factors of $\tilde{g}$ in pairs and write

$$\begin{aligned}(\sigma^1 T^{-q_{2\ell}})(\sigma^1 T^{-q_{2\ell-1}}) &= (\sigma^1\sigma^3)(\sigma^3 T^{-q_{2\ell}}\sigma^3)(\sigma^3\sigma^1)T^{-q_{2\ell-1}} \\ &= -ST^{q_{2\ell}}ST^{-q_{2\ell-1}}\end{aligned} \tag{12.40}$$

   where we used that $\sigma^1\sigma^3 = -\mathrm{i}\sigma^2 = S$. Therefore, we can write

$$\tilde{g} = g = (-1)^s \prod_{\ell=1}^{s} (ST^{q_{2\ell}}ST^{-q_{2\ell-1}}) \tag{12.41}$$

2. Now suppose that $N+1 = 2s+1$ is odd. Then we rewrite the identity (12.37) as:

$$\sigma^1\tilde{g}\begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{12.42}$$

   so now we simply take

$$g = \sigma^1\tilde{g} = (-1)^{s+1}(ST^{-q_{2s+1}}) \prod_{\ell=1}^{s} (ST^{q_{2\ell}}ST^{-q_{2\ell-1}}) \tag{12.43}$$

Thus we can summarize both cases by saying that

$$g = (-1)^{\lfloor \frac{N+1}{2} \rfloor} \prod_{\ell=1}^{N+1} (ST^{(-1)^\ell q_\ell}) \tag{12.44}$$

Then we can finally write

$$h = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = g^{-1}T^n \tag{12.45}$$

as a word in S and T for a suitable integer n. (Note that $S^2 = -1$.)

♣It would be good to give an algorithm for determining n. ♣

Now we need to show how to bring the general element $h \in SL(2,\mathbb{Z})$ to the form with $A > C > 0$ so we can apply the above formula. Note that

$$\begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ C+mA & D+mB \end{pmatrix} \tag{12.46}$$

while

$$\begin{pmatrix} 1 & 0 \\ -m & 1 \end{pmatrix} = ST^m S^{-1} \tag{12.47}$$

Thus, if $A > 0$ we can use this operation to shift C so that $0 \leq C < A$. In case $A < 0$ we can multiply by $S^2 = -1$ to reduce to the case $A > 0$. Finally, if $A = 0$ then

$$h = \begin{pmatrix} 0 & \pm 1 \\ \mp 1 & n \end{pmatrix} \tag{12.48}$$

and we write

$$ST^n = \begin{pmatrix} 0 & -1 \\ 1 & n \end{pmatrix} \tag{12.49}$$

♣Need to summarize the result in a useful way ♣

12.3 Products Of Cyclic Groups And The Chinese Remainder Theorem

Recall the elementary definition we met in the last exercise of section 2.

Definition Let H, G be two groups. The *direct product* of H and G, denoted $H \times G$, is the set $H \times G$ with product:

$$(h_1, g_1) \cdot (h_2, h_2) = (h_1 \cdot h_2, g_1 \cdot g_2) \tag{12.50}$$

We will consider the direct product of cyclic groups. According to our general notation we would write this as $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$. However, since $\mathbb{Z}_m$ is also a ring the notation $\mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2}$ also often used, and we will use it below, especially when we write our Abelian groups additively.

Let us begin with the question: Is it true that

$$\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \stackrel{?}{\cong} \mathbb{Z}_{m_1 m_2}. \tag{12.51}$$

In general (12.51) is *false*!

Exercise

a.) Show that $\mathbb{Z}_4$ is not isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. (There is a one-line proof.) [129]

b.) Is p is prime is $\mathbb{Z}_p \oplus \mathbb{Z}_p$ isomorphic to $\mathbb{Z}_{p^2}$?

c.) Is $\mathbb{Z}_3 \oplus \mathbb{Z}_5$ isomorphic to $\mathbb{Z}_{15}$?

Write $g = \gcd(m_1, m_2)$ and $\ell = \mathrm{lcm}(m_1, m_2)$. Then there are two natural exact sequences:

$$1 \to \mathbb{Z}_g \to \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \to \mathbb{Z}_\ell \to 1 \tag{12.52}$$

$$0 \to \mathbb{Z}/\ell\mathbb{Z} \to \mathbb{Z}/m_1\mathbb{Z} \oplus \mathbb{Z}/m_2\mathbb{Z} \to \mathbb{Z}/g\mathbb{Z} \to 0 \tag{12.53}$$

[129] *Answer*: Every element in $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ is of order two. But some elements of $\mathbb{Z}_4$ have order four. The but the order of a group element is preserved under isomorphism.

In fact, we will show below that

$$\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \cong \mathbb{Z}_g \times \mathbb{Z}_\ell \tag{12.54}$$

Remarks:

1. The sequence (12.52) is easier to write down multiplicatively, while (12.53) is easier to write down additively. See the discussion below. (Of course, both are true in either formulation!)

2. If $g = 1$ since $\mathbb{Z}_1 = \mathbb{Z}/\mathbb{Z}$ is the trivial group we can indeed conclude that $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \cong \mathbb{Z}_{m_1 m_2}$ but otherwise this is false. We will return to this point.

Now, let us prove (12.52) and (12.53).
Recall that

$$m_1 m_2 = g\ell \tag{12.55}$$

a fact that will be useful momentarily. (If you do not know this we will prove it below.) It will also be useful to write $m_1 = \mu_1 g$ and $m_2 = \mu_2 g$ where μ_1, μ_2 are relatively prime. Thus there are integers ν_1, ν_2 with

$$\mu_1 \nu_1 + \mu_2 \nu_2 = 1 \tag{12.56}$$

and hence

$$m_1 \nu_1 + m_2 \nu_2 = g. \tag{12.57}$$

To prove (12.52) think of $\mathbb{Z}_m$ as the multiplicative group of mth roots of 1, so they are all subgroups of $U(1)$. Now define a group homomorphism:

$$\pi : \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \to \mathbb{Z}_\ell \tag{12.58}$$

by:

$$\pi : (\xi_1, \xi_2) \to \xi_1 \xi_2 \tag{12.59}$$

That is, we merely multiply the two entries. (This makes it clear that it is a group homomorphism since the group law is multiplication of complex numbers and that multiplication is commutative.) Here ξ_1 is an m_1^{th} root of unity and ξ_2 is an m_2^{th} root of unity. The only thing you need to check is that indeed then $\xi_1 \xi_2$ is an ℓth root of unity, so π indeed maps into $\mathbb{Z}_\ell$.

Now we prove that π is surjective: Let $\omega_1 = e^{\frac{2\pi i}{m_1}}$ and $\omega_2 = e^{\frac{2\pi i}{m_2}}$. These are generators of $\mathbb{Z}_{m_1}$ and $\mathbb{Z}_{m_2}$. Choose integers ν_1, ν_2 so that $\nu_1 m_1 + \nu_2 m_2 = g$ then π maps

$$\begin{aligned}
\pi : (\omega_1^{\nu_2}, \omega_2^{\nu_1}) &\mapsto \omega_1^{\nu_2} \omega_2^{\nu_1} \\
&= \exp\left[2\pi \mathrm{i}\left(\frac{\nu_2}{m_1} + \frac{\nu_1}{m_2}\right)\right] \\
&= \exp\left[2\pi \mathrm{i}\left(\frac{m_2\nu_2 + m_1\nu_1}{m_1 m_2}\right)\right] \\
&= \exp\left[2\pi \mathrm{i}\frac{g}{m_1 m_2}\right] \\
&= \exp\left[2\pi \mathrm{i}\frac{1}{\ell}\right]
\end{aligned} \tag{12.60}$$

But $\exp\left[2\pi i\frac{1}{\ell}\right]$ is a generator of the multiplicative group of ℓth roots of unity, isomorphic to $\mathbb{Z}_\ell$, and hence the homomorphism π is onto. Thus, we have checked exactness of the sequence at $\mathbb{Z}_\ell$.

On the other hand the injection map

$$\iota : \mathbb{Z}_g \to \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \tag{12.61}$$

is defined by identifying $\mathbb{Z}_g$ with the multiplicative group of gth roots of unity and just sending:

$$\iota(\xi) = (\xi, \xi^{-1}) \tag{12.62}$$

Note that a gth root of unity ξ has the property that $\xi^{\pm 1}$ is also both an m_1^{th} and an m_2^{th} root of unity. So this makes sense. It is now easy to check that indeed the kernel of π is the image of ι. Since π takes the product of the two entries it is immediate from the definition (12.62) that $im(\iota) \subset \ker(\pi)$. On the other hand, if $\pi(\xi_1, \xi_2) = \xi_1\xi_2 = 1$ then clearly $\xi_2 = \xi_1^{-1}$, so this must be in the image of ι. Now we have checked exactness at the middle of the sequence. Exactness at $\mathbb{Z}_g$ is trivial. This concludes the proof of (12.52) ♠

It is worth noting that we can write "additive" version of the maps ι and π as:

$$\begin{aligned} \iota(x) &= \mu_1 x \oplus (-\mu_2 x) \\ \pi(x_1 \oplus x_2) &= \mu_2 x_1 + \mu_1 x_2 \end{aligned} \tag{12.63}$$

You should check that written this way it is well defined, and the sequence is exact.

Exercise

a.) Show that there is an exact sequence

$$0 \to \mathbb{Z}/\ell\mathbb{Z} \xrightarrow{\iota} \mathbb{Z}/m_1\mathbb{Z} \oplus \mathbb{Z}/m_2\mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/g\mathbb{Z} \to 0 \tag{12.64}$$

where

$$\pi : x_1 \oplus x_2 \mapsto (x_1 - x_2) \mathrm{mod} g \ . \tag{12.65}$$

$$\iota : x \mapsto (x \mathrm{mod} m_1 \oplus x \mathrm{mod} m_2) \tag{12.66}$$

b.) Show that if we think of these groups as groups of roots of unity then we have $\pi(\xi_1, \xi_2) = \xi_1^{\mu_1}\xi_2^{-\mu_2}$ and $\iota(\omega) = (\omega^{\mu_2}, \omega^{\mu_1})$ with $m_1 = \mu_1 g$ and $m_2 = \mu_2 g$.

Now we prove that in general

$$\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \cong \mathbb{Z}_g \times \mathbb{Z}_\ell \tag{12.67}$$

First, it follows from either of the two exact sequences we proved above that if $(m_1, m_2) = 1$ then indeed

$$\mathbb{Z}_{m_1 m_2} \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \tag{12.68}$$

Next, recall that any integer can be decomposed into its prime factors:

$$m = \prod_p p^{v_p(m)} \tag{12.69}$$

where $v_p(m) \in \mathbb{Z}_+$, known as the *valuation of m at p* is zero for all but finitely many primes. (So we have an infinite product of 1's on the RHS of the above equation.)

Now in terms of the prime factorizations of m_1, m_2 we can write:

$$\begin{aligned} g &= \gcd(m_1, m_2) = \prod_p p^{\min[v_p(m_1), v_p(m_2)]} \\ \ell &= \mathrm{lcm}(m_1, m_2) = \prod_p p^{\max[v_p(m_1), v_p(m_2)]} \end{aligned} \tag{12.70}$$

Now, from the above we know that $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \cong \mathbb{Z}_{m_1 m_2}$ if m_1 and m_2 are relatively prime. Therefore we can write

$$\mathbb{Z}/m\mathbb{Z} \cong \prod_p (\mathbb{Z}/p^{v_p(m)}\mathbb{Z}) \tag{12.71}$$

Applying this to each of the two factors in $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$ and using $G_1 \times G_2 \cong G_2 \times G_1$ to arrange the factors so the minimum power is on the left and maximum on the right and regrouping gives (12.67). In equations:

$$\begin{aligned} \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} &\cong \prod_p \mathbb{Z}_{p^{\nu_p(m_1)}} \times \prod_p \mathbb{Z}_{p^{\nu_p(m_2)}} \\ &\cong \prod_p \mathbb{Z}_{p^{\min[\nu_p(m_1), \nu_p(m_2)]}} \times \prod_p \mathbb{Z}_{p^{\max[\nu_p(m_1), \nu_p(m_2)]}} \\ &\cong \mathbb{Z}_g \times \mathbb{Z}_\ell \end{aligned} \tag{12.72}$$

A second proof gives some additional insight by providing an interesting visual picture of what is going on, as well as relating this fact to lattices. It is related to the first by "taking a logarithm" and involves exact sequences of infinite groups which induce sequences on finite quotients.

Consider the sublattice of $\mathbb{Z} \oplus \mathbb{Z}$ given by

$$\Lambda = m_1\mathbb{Z} \oplus m_2\mathbb{Z} = \{ \begin{pmatrix} m_1\alpha \\ m_2\beta \end{pmatrix} | \alpha, \beta \in \mathbb{Z} \} \tag{12.73}$$

Then $\Lambda \subset \mathbb{Z} \oplus \mathbb{Z}$ is a sublattice and it should be pretty clear that

$$\mathbb{Z}^2/\Lambda = \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \tag{12.74}$$

Now, write $m_1 = \mu_1 g, m_2 = \mu_2 g$ as above. Choose integers ν_1, ν_2 so that $\mu_1\nu_1 + \mu_2\nu_2 = 1$ and consider the matrix

$$\begin{pmatrix} \mu_2 & \mu_1 \\ -\nu_1 & \nu_2 \end{pmatrix} \in SL(2, \mathbb{Z}) \tag{12.75}$$

This is an invertible matrix over the integers, so we can change coordinates on the lattice from $x = m_1\alpha, y = m_2\beta$ to

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \mu_2 & \mu_1 \\ -\nu_1 & \nu_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{12.76}$$

that is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \nu_2 & -\mu_1 \\ \nu_1 & \mu_2 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \tag{12.77}$$

which we prefer to write as:

$$\begin{pmatrix} x \\ y \end{pmatrix} = x' \begin{pmatrix} \nu_2 \\ \nu_1 \end{pmatrix} + y' \begin{pmatrix} -\mu_1 \\ \mu_2 \end{pmatrix} \tag{12.78}$$

We interpret this as saying that x', y' are the coordinates of the vector $(x, y) \in \mathbb{Z}^2$ relative to the new basis vectors for $\mathbb{Z}^2$.

$$v_1 = \begin{pmatrix} \nu_2 \\ \nu_1 \end{pmatrix} \qquad v_2 = \begin{pmatrix} -\mu_1 \\ \mu_2 \end{pmatrix} \tag{12.79}$$

The good property of this basis is that the smallest multiple of v_1 that sits in Λ is ℓv_1 (prove this) [130] Similarly, the smallest multiple of v_2 in Λ is $g v_2$. Thus, we have a way of writing $\mathbb{Z}^2$ as $\mathbb{Z}v_1 \oplus \mathbb{Z}v_2$ such that the projection of Λ to the v_1 axis is the group $\ell\mathbb{Z}$ while the kernel is the subgroup of $\mathbb{Z}v_2$ that maps into Λ, and that is just $\cong g\mathbb{Z}$.

Put differently, there is a homomorphism $\psi : \mathbb{Z}^2 \to \mathbb{Z}$ that takes

$$\psi : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x' \ . \tag{12.80}$$

This is the projection on the v_1 axis. This defines a surjective homomorphism onto $\mathbb{Z}$. (Explain why.) On the other hand, using (12.76) and $\mu_1\mu_2 g = \ell$ we see that the image of Λ under ψ is $\ell\mathbb{Z}$. Therefore, using the exercise result (10.24) ψ descends to a map

$$\bar{\psi} : \mathbb{Z}^2/\Lambda \to \mathbb{Z}/\ell\mathbb{Z} \tag{12.81}$$

Now note from (12.78) that

$$\begin{pmatrix} -\mu_1 \\ \mu_2 \end{pmatrix} \mathrm{mod}\Lambda \tag{12.82}$$

is in the kernel of $\bar{\psi}$, and moreover it generates a cyclic subgroup of order g in $\mathbb{Z}^2/\Lambda$. By counting, this cyclic subgroup must be the entire kernel of $\bar{\psi}$. Therefore we have an exact sequence

$$0 \to \mathbb{Z}_g \to \mathbb{Z}^2/\Lambda \to \mathbb{Z}_\ell \to 0 \tag{12.83}$$

****

AND THERE IS A MAP TO y' AND TOGETHER THESE GIVE ISOMORPHISM TO $\mathbb{Z}/\ell\mathbb{Z} \oplus \mathbb{Z}/g\mathbb{Z}$.

****

This concludes our second proof. ♠

♣Should really add a figure to illustrate this. Unfortunately the first really nontrivial case is $m_1 = 2 \cdot 3$ and $m_2 = 2 \cdot 5$ so $\ell = 30$. ♣

Exercise

Using the Kronecker theorem show that if a finite Abelian group G is not a cyclic group then there is a nontrivial divisor n of $|G|$ so that $g^n = 1$ for all $g \in G$.

12.3.1 The Chinese Remainder Theorem

In fact, there is an important generalization of this statement known as the *Chinese remainder theorem*:

Theorem Suppose $m_1, \ldots, m_r$ are pairwise relatively prime positive integers, (i.e. $(m_i, m_j) = 1$ for all $i \neq j$) then

$$(\mathbb{Z}/m_1\mathbb{Z}) \oplus (\mathbb{Z}/m_2\mathbb{Z}) \cdots \oplus (\mathbb{Z}/m_r\mathbb{Z}) \cong \mathbb{Z}/M\mathbb{Z} \tag{12.84}$$

where $M = m_1 m_2 \cdots m_r$.

Proof:

The fastest proof makes use of the previous result and induction on r.

A second proof offers some additional insight into solving simultaneous congruences: We construct a homomorphism

$$\psi : \mathbb{Z} \to (\mathbb{Z}/m_1\mathbb{Z}) \oplus (\mathbb{Z}/m_2\mathbb{Z}) \cdots \oplus (\mathbb{Z}/m_r\mathbb{Z}) \tag{12.85}$$

by

$$\psi(x) = (x \mathrm{mod} m_1, x \mathrm{mod} m_2, \ldots, x \mathrm{mod} m_r) \tag{12.86}$$

We first claim that $\psi(x)$ is *onto*. That is, for any values $a_1, \ldots, a_r$ we can solve the simultaneous congruences:

$$\begin{aligned} x &= a_1 \mathrm{mod} m_1 \\ x &= a_2 \mathrm{mod} m_2 \\ \vdots &\quad \vdots \\ x &= a_r \mathrm{mod} m_r \end{aligned} \tag{12.87}$$

for some common value $x \in \mathbb{Z}$.

[130] *Answer*: We have $x v_1 \in \Lambda$ iff $x\nu_2 = 0 \mathrm{mod}(g\mu_1)$ and $x\nu_1 = 0 \mathrm{mod}(g\mu_2)$. Multiply these equations by μ_2 and μ_1, respectively, and add them. Find that $x = 0 \mathrm{mod} \ell$.

To prove this note that $\hat{m}_i := M/m_i = \prod_{j\neq i} m_j$ is relatively prime to m_i (by the hypothesis of the theorem). Therefore there are integers x_i, y_i such that

$$x_i m_i + y_i \hat{m}_i = 1 \tag{12.88}$$

Let $g_i = y_i \hat{m}_i$. Note that

$$g_i = \delta_{i,j} \mathrm{mod} m_j \qquad \forall 1 \leq i,j \leq r \tag{12.89}$$

Therefore if we set

$$x = \sum_{i=1}^{r} a_i g_i \tag{12.90}$$

then x is a desired solution to (12.87) and hence is a preimage under ψ.

On the other hand, the kernel of ψ is clearly $M\mathbb{Z}$. Therefore:

$$0 \to M\mathbb{Z} \to \mathbb{Z} \to (\mathbb{Z}/m_1\mathbb{Z}) \oplus (\mathbb{Z}/m_2\mathbb{Z}) \cdots \oplus (\mathbb{Z}/m_r\mathbb{Z}) \to 0 \tag{12.91}$$

and hence the desired isomorphism follows. ♠

Remarks

1. Equation (12.67) is used implicitly all the time in physics, whenever we have two degrees of freedom with different but commensurable frequencies. Indeed, it is used all the time in everyday life. As a simple example, suppose you do X every other day. You will then do X on Mondays every other week, i.e., every 14 days, because 2 and 7 are relatively prime. More generally, consider a system with a discrete configuration space $\mathbb{Z}/p\mathbb{Z}$ thought of as the multiplicative group of pth roots of 1. Suppose the time evolution for $\Delta t = 1$ is $\omega_p^r \to \omega_p^{r+1}$ where ω_p is a primitive p^{th} root of 1. The basic period is $T = p$. Now, if we have *two* oscillators of periods p, q, the configuration space is $\mathbb{Z}_p \times \mathbb{Z}_q$. The basic period of this system is - obviously - the least common multiple of p and q. That is the essential content of (**??**).

2. Our second proof shows that in fact equation (12.84) is a statement of an isomorphism of rings.

3. One might wonder how the theorem got this strange name. (Why don't we refer to the "Swiss-German theory of relativity?") The theorem is attributed (see, e.g. Wikipedia) to Sun-tzu Suan-ching in the 3rd century A.D. (He should not be confused with Sun Tzu who lived in the earlier Spring and Autumn period and wrote *The Art of War*.) For an interesting historical commentary see [131] which documents the historical development in India and China up to the definitive treatments by Euler, Lagrange, and Gauss who were probably unaware of previous developments hundreds of years earlier. The original motivation was apparently related to construction of calendars, and this is certainly mentioned by Gauss in his renowned book *Disquisitiones Arithmeticae*. The Chinese calendar is based on *both* the lunar and solar cycles. Roughly speaking, one starts the new year based on both the winter solstice

and the new moon. Thus, to find periods of time in this calendar one needs to solve simultaneous congruences. I suspect the name "Chinese Remainder Theorem" is an invention of 19th century mathematicians. Hardy & Wright (1938) do not call it that, but do recognize Sun Tzu.

♣Some students with Chinese background say this is wrong. Check it out. ♣

Exercise *Counting your troops*

Suppose that you are a general and you need to know how many troops you have from a cohort of several hundred. Time is too short to take attendance.

So, you have your troops line up in rows of 5. You observe that there are 3 left over. Then you have your troops line up in rows of 11. Now there are 2 left over. Finally, you have your troops line up in rows of 13, and there is only one left over.

How many troops are there? [132]

Exercise

a.) Show that the Chinese Remainder theorem is false if the m_i are not pairwise relatively prime.

b.) Show that the obstruction to finding a solution x to $x = a_i \mathrm{mod} m_i$ is given by the reductions $(a_i - a_j)\mathrm{mod}(m_i, m_j)$ over all pairs $i \neq j$. That is, a solution exists iff all of these vanish.

---

13. The Group Of Automorphisms

Recall that an *automorphism* of a group G is an isomorphism $\mu : G \to G$, i.e. an isomorphism of G onto itself.

One easily checks that the composition of two automorphisms μ_1, μ_2 is an automorphism. The identity map is an automorphism, and every automorphism is invertible. In this way, the set of automorphisms, $\mathrm{Aut}(G)$, is *itself a group* with group law given by composition.

Given a group G there are God-given automorphisms given by conjugation. That is, if $a \in G$ then

$$I(a) : g \to aga^{-1} \tag{13.1}$$

[131] Kang Sheng Shen, "Historical development of the Chinese remainder theorem," Arch. Hist. Exact Sci. 38 (1988), no. 4, 285305.

[132] Apply the Chinese remainder theorem with $m_1 = 5, m_2 = 11, m_3 = 13$. Then $M = 715$, $\hat{m}_1 = 143$, $\hat{m}_2 = 65$ and $\hat{m}_3 = 55$. Using the Euclidean algorithm you find convenient lifts to the integers $g_1 = 286$, $g_2 = -65$ and $g_3 = -220$. Then the number of troops is $3 \times 286 - 2 \times 65 - 1 \times 220 = 508 \mathrm{mod} 715$. Therefore there are 508 soldiers.

defines an automorphism of G. Indeed $I(a) \circ I(b) = I(ab)$ and hence $I : G \to \mathrm{Aut}(G)$ is a homomorphism. The subgroup $\mathrm{Inn}(G)$ of such automorphisms is called the group of *inner automorphisms*. Note that if $a \in Z(G)$ then $I(a)$ is trivial, and conversely. Thus we have:

$$\mathrm{Inn}(G) \cong G/Z(G). \tag{13.2}$$

Moreover, $\mathrm{Inn}(G)$ is a normal subgroup of $\mathrm{Aut}(G)$, since for any automorphism $\phi \in \mathrm{Aut}(G)$:

$$\phi \circ I(a) \circ \phi^{-1} = I(\phi(a)). \tag{13.3}$$

Therefore we have another group

$$\mathrm{Out}(G) := \mathrm{Aut}(G)/\mathrm{Inn}(G) \tag{13.4}$$

known as the group of "outer automorphisms." Thus

$$1 \to \mathrm{Inn}(G) \to \mathrm{Aut}(G) \to \mathrm{Out}(G) \to 1 \tag{13.5}$$

Note we can also write and exact sequence of length four:

$$1 \to Z(G) \to G \to \mathrm{Aut}(G) \to \mathrm{Out}(G) \to 1 \tag{13.6}$$

Remarks

1. In practice one often reads or hears the statement that an element $\varphi \in \mathrm{Aut}(G)$ is an "outer automorphism." What this means is that it projects to a nontrivial element of $\mathrm{Out}(G)$. However, strictly speaking this is an abuse of terminology and an outer automorphism is in the quotient group (13.4). These notes might sometimes perpetrate this abuse of terminology.

2. Note that for any abelian group G all nontrivial automorphisms are outer automorphisms.

Example 13.1: Consider $\mathrm{Aut}(\mathbb{Z}_3)$. This group is Abelian so all automorphisms are outer. Thinking of it multiplicatively, the only nontrivial choice is $\omega \to \omega^{-1}$. If we think of $A_3 \cong \mathrm{Aut}(\mathbb{Z}_3)$ then we are taking

$$(123) \to (132) \tag{13.7}$$

So: $Aut(\mathbb{Z}_3) \cong \mathbb{Z}_2$.

Example 13.2: Consider $Aut(\mathbb{Z}_4)$. Think of $\mathbb{Z}_4$ as the group of fourth roots of unity, generated by $\omega = \exp[i\pi/2] = i$. A generator must go to a generator, so there is only one possible nontrivial automorphism: $\phi : \omega \to \omega^3$. Note that $\omega \to \omega^2$ is a nontrivial homomorphism of $\mathbb{Z}_4 \to \mathbb{Z}_4$, but it is not an automorphism. Thus $Aut(\mathbb{Z}_4) \cong \mathbb{Z}_2$.

Example 13.3: Consider $Aut(\mathbb{Z}_5)$. Think of $\mathbb{Z}_5$ as the group of fifth roots of unity, generated by $\omega = \exp[2\pi i/5]$. Now there are several automorphisms: ϕ_2 defined by its action on the generator $\omega \to \omega^2$. Similarly, we can define ϕ_3, by $\omega \to \omega^3$ and ϕ_4, by $\omega \to \omega^4$. Letting ϕ_1 denote the identity we have

$$\phi_2^2 = \phi_4 \qquad \phi_2^3 = \phi_3 \qquad \phi_2^4 = \phi_4^2 = \phi_1 = 1 \tag{13.8}$$

So $Aut(\mathbb{Z}_5) \cong \mathbb{Z}_4$. The explicit isomorphism is

$$\begin{aligned} \phi_2 &\to \bar{1} \\ \phi_4 &\to \bar{2} \\ \phi_3 &\to \bar{3} \end{aligned} \tag{13.9}$$

Example 13.4: Consider $\mathrm{Aut}(\mathbb{Z}_N)$, and let us think of $\mathbb{Z}_N$ multiplicatively as the group of Nth roots of 1. An automorphism ϕ of $\mathbb{Z}_N$ must send $\omega \mapsto \omega^r$ for some r. On the other hand, ω^r must also be a generator of $\mathbb{Z}_N$. Automorphisms must take generators to generators. Hence r is relatively prime to N. This is true iff there is an s with

$$rs = 1 \mathrm{mod} N \tag{13.10}$$

Thus, $Aut(\mathbb{Z}_N)$ is the group of transformations $\omega \to \omega^r$ where r admits a solution to $rs = 1\mathrm{mod}N$. We will examine this interesting group in a little more detail in §13.1 below.

Example 13.5: *Automorphisms Of The Symmetric Group S_n*: There are no outer automorphisms of S_n so

$$Aut(S_n) \cong Inn(S_n) \cong S_n, \qquad n \neq 2, 6 \tag{13.11}$$

Note the exception: $n = 2, 6$. Note the striking contrast from an abelian group, all of whose automorphisms are outer.

This is not difficult to prove: Note that an automorphism ϕ of S_n must take conjugacy classes to conjugacy classes. Therefore we focus on how it acts on transpositions. These are involutions, and involutions must map to involutions so the conjugacy class of transpositions must map to a conjugacy class of the form $(1)^k(2)^\ell$ with $k + 2\ell = n$. We will show below that, just based on the order of the conjugacy class, ϕ must map transpositions to transpositions. We claim that any automorphism that maps transpositions to transpositions must be inner. Let us say that

$$\phi((ab)) = (xy) \qquad \phi((ac)) = (zw) \tag{13.12}$$

where a, b, c are all distinct. We claim that x, y, z, w must comprise precisely three distinct letters. We surely can't have $(xy) = (zw)$ because ϕ is 1-1, and we also can't have (xy) and (zw) commuting because the group commutator of (ab) and (ac) is (abc). Therefore we can write

$$\phi((ab)) = (xy) \qquad \phi((ac)) = (xz) \tag{13.13}$$

Therefore, we have defined a permutation $a \to x$ and ϕ is the inner automorphism associated with this permutation.

Now let us consider the size of the conjugacy classes. This was computed in exercise *** above. The size of the conjugacy class of transpositions is of course

$$\binom{n}{2} = \frac{n!}{(n-2)!2!} \tag{13.14}$$

The size of a conjugacy class of the form $(1)^k(2)^\ell$ with $k + 2\ell = n$ is

$$\frac{n!}{(n-2\ell)!\ell!2^\ell} \tag{13.15}$$

Setting these equal results in the identity

$$\frac{(n-2)!}{(n-2\ell)!} = \ell!2^{\ell-1} \qquad n \geq 2\ell \tag{13.16}$$

For a fixed ℓ the LHS is a polynomial in n which is growing for $n \geq 2\ell$ and therefore bounded below by $(2\ell-2)!$. Therefore we consider whether there can be a solution with $n = 2\ell$:

$$(2\ell-2)! = \ell!2^{\ell-1} \tag{13.17}$$

For $\ell = 3$, corresponding to $n = 6$, there is a solution, but for $\ell > 3$ we have $(2\ell-2)! > \ell!2^{\ell-1}$. The peculiar exception $n = 6$ is related to the symmetries of the icosahedron. For more information see

1.http://en.wikipedia.org/wiki/Automorphisms of the symmetric and alternating groups

2. http://www.jstor.org/pss/2321657

3. I.E. Segal, "The automorphisms of the symmetric group," *Bulletin of the American Mathematical Society* **46**(1940) 565.

Example 13.6: *Automorphisms Of Alternating Groups.* For the group $A_n \subset S_n$ there is an automorphism which is not obviously inner: Conjugation by any odd permutation. Recall that $Out(G) = Aut(G)/Inn(G)$ is a quotient group so conjugation by any odd permutation represents the same element in $Out(G)$. If we consider $A_3 \subset S_3$ then

$$(12)(123)(12)^{-1} = (132) \tag{13.18}$$

is indeed a nontrivial automorphism of A_3 and since A_3 is abelian this automorphism must be an outer automorphism. In general conjugation by an odd permutation defines an outer automorphism of A_n. For example suppose conjugation by (12) were inner. Then there would be an even permutation a so that conjugation by $a \cdot (12)$ centralizes every $h \in A_n$. But $a \cdot (12)$ together with A_n generates all of S_n and then $a \cdot (12)$ would have to be in the center of S_n, a contradiction. Thus, the outer automorphism group of A_n contains a nontrivial involution. Again for $n = 6$ there is an exceptional outer automorphism.

The above example nicely illustrates a general idea: If $N \triangleleft G$ is a normal subgroup of G and $g \notin H$ then conjugation by g defines an automorphism $H \to H$ which is, in general, not an inner automorphism.

Example 13.7: Consider $G = GL(n, \mathbb{C})$. Then $A \to A^*$ is an outer automorphism: That is, there is no invertible complex matrix $S \in GL(n, \mathbb{C})$ such that, for every invertible matrix $A \in GL(n, \mathbb{C})$ we have

$$A^* = SAS^{-1} \tag{13.19}$$

Exercise *Outer Automorphisms Of Some Matrix Groups*

a.) Prove (13.19). [133]

b.) Consider maps of $GL(n,\mathbb{C})$ given by $A \to A^{tr}$, $A \to A^{-1}$ and $A \to A^{tr,-1}$. Which of these are automorphisms? Which of these are outer automorphisms?

c.) Consider $G = SU(2)$. Is $A \to A^*$ an outer automorphism? [134]

d.) Consider the automorphism of $G = SO(2)$

$$R(\phi) \to R(-\phi) \tag{13.20}$$

Is this inner or outer?

Exercise *Automorphisms of* $\mathbb{Z}$

Show that $\mathrm{Aut}(\mathbb{Z}) \cong \mathbb{Z}_2$. [135]

Exercise

Although $\mathbb{Z}_2$ does not have any automorphisms the product group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ certainly does.

a.) Show that an automorphism of $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ must be of the form

$$\phi(x_1, x_2) = (a_1 x_1 + a_2 x_2, a_3 x_3 + a_4 x_4) \tag{13.21}$$

where we are writing the group additively, and

$$\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \in GL(2,\mathbb{Z}_2) \tag{13.22}$$

b.) Show that $GL(2,\mathbb{Z}_2) \cong S_3$. [136]

---

[133] *Hint*: Consider the invertible matrix $A = \mathrm{i}1_{n\times n}$.

[134] *Answer*: No! Note that $(\mathrm{i}\sigma^k)^* = -\mathrm{i}(\sigma^k)^* = (\mathrm{i}\sigma^2)(\mathrm{i}\sigma^k)(\mathrm{i}\sigma^2)^{-1}$. But $\mathrm{i}\sigma^2 \in SU(2)$ and every $SU(2)$ matrix is a real linear combination of 1 and $\mathrm{i}\sigma^k$. This has an important implication for the representation theory of $SU(2)$: Every irreducible representation is either real or "pseudoreal" (quaternionic).

[135] *Answer*: The most general homomorphism $\mathbb{Z} \to \mathbb{Z}$ is the map $n \mapsto an$ for some integer a. But for an automorphism a must be mutliplicatively invertible in the integers. Therefore a is $+1$ or -1.

[136] *Hint*: Consider what the group does to the three nontrivial elements $(1,0)$, $(0,1)$, and $(1,1)$. The three transpositions correspond to $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and the two elements of order 3 are $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$.

c.) Now describe $\mathrm{Aut}(\mathbb{Z}_4 \times \mathbb{Z}_4)$. [137]

Exercise *Automorphisms of* $\mathbb{Z}_p^N$

(Warning: This is hard and uses some other ideas from algebra.)

Let p be prime. Describe the automorphisms of $\mathbb{Z}_p^N$, and show that the group has order [138]

$$|\mathrm{Aut}(\mathbb{Z}_p^N)| = (p^N - 1)(p^N - p)(p^N - p^2) \cdots (p^N - p^{N-1}) \tag{13.23}$$

Exercise *Automorphisms Of The Quaternion Group*

♣This exercise should go later, perhaps in the section on extensions. Perhaps in the chapter on symmetries of regular objects. ♣

Show that the group of automorphisms of the quaternion group $Q = \{\pm 1, \pm \mathfrak{i}, \pm \mathfrak{j}, \pm \mathfrak{k}\}$ is isomorphic to S_4. [139]

(This assumes you know what the quaternions are. See below for various descriptions of the quaternion group Q.)

Exercise *Isomorphisms between two different groups*

Let G_1, G_2 be two groups which are isomorphic, but not presented as the same set with the same multiplication table. Let $\mathrm{Isom}(G_1, G_2)$ be the set of all isomorphisms from $G_1 \to G_2$.

Show that

a.) Any two isomorphisms $\Psi, \Psi' \in \mathrm{Isom}(G_1, G_2)$ are related by $\Psi' = \Psi \circ \phi$ where $\phi \in \mathrm{Aut}(G_1)$.

---

[137] *Answer*: This group has a homomorphism onto $\mathrm{Aut}(\mathbb{Z}_2 \times \mathbb{Z}_2)$ with a kernel isomorphic to $\mathbb{Z}_2^4$.

[138] *Answer*: The group is the group of $N \times N$ invertible matrices over the ring $\mathbb{Z}_p$. These are in one-one correspondence with the possible bases for the vector space $\mathbb{Z}_p^N$. How many ordered bases are there? Note that any nonzero vector can serve as the first basis vector, and there are $p^N - 1$ nonzero vectors. Choose one and call it e_1. Now, e_2 can be any vector not in the linear span of e_1. But the linear span of e_1 is a one-dimensional subspace of p elements. These are all excluded so e_2 must be chosen from a set of $p^N - p$ vectors. Make a choice of e_2. Then e_3 must be chosen from a vector not in the span of e_1, e_2. The span of e_1, e_2 consists of p^2 vectors so there are $p^N - p^2$ choices for e_3, and so on.

[139] *Answer*: First, the group of inner automorphisms is $Q/Z(Q) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. The three nontrivial elements are given by conjugation by $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}$. Now, any automorphism must permute the three normal subgroups generated by $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}$, and automorphisms leading to nontrivial permutations of normal subgroups must be outer. So the outer automorphism group must be a subgroup of S_3. Now, in fact, one can construct such outer automorphisms. In fact, it suffices to say what the image of $\mathfrak{i}$ and $\mathfrak{j}$ are since these generate the whole group. Thus, the automorphism group is an extension of S_3 by $\mathbb{Z}_2 \times \mathbb{Z}_2$ and one can then map this isomorphically to S_4.

b.) Any two isomorphisms $\Psi, \Psi' \in \mathrm{Isom}(G_1, G_2)$ are related by $\Psi' = \phi \circ \Psi$ where $\phi \in \mathrm{Aut}(G_2)$.

The set $Isom(G_1, G_2)$ with G_1, G_2 not equal but isomorphic is a good example of what is called a *torsor*. A *torsor* for a group G is a set X with a free transitive action.

13.1 The group of units in $\mathbb{Z}_N$

We have seen that $\mathbb{Z}/N\mathbb{Z}$ is a group inherited from the *additive* law on $\mathbb{Z}$. For an integer $n \in \mathbb{Z}$ denote its image in $\mathbb{Z}/N\mathbb{Z}$ by $\bar{n}$. With this notation the group law on $\mathbb{Z}/N\mathbb{Z}$ is

$$\bar{n}_1 + \bar{n}_2 = \overline{n_1 + n_2}, \tag{13.24}$$

and $\bar{0}$ is the unit element.

However, note that since

$$(n_1 + N\ell_1)(n_2 + N\ell_2) = n_1 n_2 + N\ell'' \tag{13.25}$$

we do have a well-defined operation on $\mathbb{Z}/N\mathbb{Z}$ inherited from *multiplcation* in $\mathbb{Z}$:

$$\bar{n}_1 \cdot \bar{n}_2 := \overline{n_1 \cdot n_2}. \tag{13.26}$$

In general, even if we omit $\bar{0}$, $\mathbb{Z}/N\mathbb{Z}$ is *not* a group with respect to the multiplication law (find a counterexample). Nevertheless, $\mathbb{Z}/N\mathbb{Z}$ with $+, \times$ is an interesting object which is an example of something called a *ring*. See the next chapter for a general definition of a ring.

Let us define *the group of units in the ring* $\mathbb{Z}/N\mathbb{Z}$:

$$(\mathbb{Z}/N\mathbb{Z})^* := \{\bar{m} : 1 \leq m \leq N-1, \gcd(m, N) = 1\} \tag{13.27}$$

where (m, N) is the *greatest common divisor* of m and N. We will also denote this group as $\mathbb{Z}_N^*$.

Then, $(\mathbb{Z}/N\mathbb{Z})^*$ *is* a group with the law (13.26) ! Clearly the multiplication is closed and $\bar{1}$ is the unit. The existence of multiplicative inverses follows from (12.8).

Moreover, as we have seen above, we can identify

$$\mathrm{Aut}(\mathbb{Z}/N\mathbb{Z}) \cong (\mathbb{Z}/N\mathbb{Z})^* \tag{13.28}$$

The isomorphism is that $a \in (\mathbb{Z}/N\mathbb{Z})^*$ is mapped to the transformation

$$\psi_a : n \bmod N \to an \bmod N \tag{13.29}$$

if we think of $\mathbb{Z}/N\mathbb{Z}$ additively or

$$\psi_a : \omega \to \omega^a \tag{13.30}$$

if we think of it multiplicatively. Note that $\psi_{a_1} \circ \psi_{a_2} = \psi_{a_1 a_2}$.

The order of the group $(\mathbb{Z}/N\mathbb{Z})^*$ is denoted $\phi(N)$ and is called the Euler ϕ-function or *Euler's totient function.* [140] One can check that

$$\begin{aligned}\phi(2) &= 1\\ \phi(3) &= 2\\ \phi(4) &= 2\end{aligned} \tag{13.31}$$

What can we say about the structure of $\mathbb{Z}_N^*$? Now, in general it is <u>not</u> true that $\mathrm{Aut}(G_1 \times G_2)$ and $\mathrm{Aut}(G_1) \times \mathrm{Aut}(G_2)$ are isomorphic. Counterexamples abound. For example $\mathrm{Aut}(\mathbb{Z}) \cong \mathbb{Z}_2$ but $\mathrm{Aut}(\mathbb{Z} \oplus \mathbb{Z}) \cong GL(2,\mathbb{Z})$. Nevertheless, it actually is true that $\mathrm{Aut}(\mathbb{Z}_n \times \mathbb{Z}_m) \cong \mathrm{Aut}(\mathbb{Z}_n) \times \mathrm{Aut}(\mathbb{Z}_m)$ when n and m are relatively prime. To prove this, let v_1 be a generator of $\mathbb{Z}_n$ and v_2 a generator of $\mathbb{Z}_m$ and let us write our Abelian group additively. The general endomorphism of $\mathbb{Z}_n \oplus \mathbb{Z}_m$ is of the form

$$\begin{aligned}v_1 &\to \alpha v_1 + \beta v_2\\ v_2 &\to \gamma v_1 + \delta v_2\end{aligned} \tag{13.32}$$

with $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$. Now impose the conditions $nv_1 = 0$ and $mv_2 = 0$ and the fact that $\bar{n}$ is multiplicatively invertible in $\mathbb{Z}_m$ and $\bar{m}$ is multiplicatively invertible in $\mathbb{Z}_n$ to learn that in fact an endomorphism must have $\beta = 0 \mathrm{mod}\ m$ and $\gamma = 0 \mathrm{mod}\ n$. Therefore $\beta v_2 = 0$ and $\gamma v_1 = 0$. Therefore, an automorphism of $\mathbb{Z}_n \oplus \mathbb{Z}_m$ is determined by $v_1 \to \alpha v_1$ with $\bar{\alpha} \in \mathbb{Z}_n^*$ and $v_2 \to \delta v_2$ with $\bar{\delta} \in \mathbb{Z}_m^*$ and hence $\mathrm{Aut}(\mathbb{Z}_n \oplus \mathbb{Z}_m) \cong \mathrm{Aut}(\mathbb{Z}_n) \times \mathrm{Aut}(\mathbb{Z}_m)$ when n and m are relatively prime. (The corresponding statement is absolutely false when they are not relatively prime.) So we have:

$$\mathbb{Z}_{nm}^* \cong \mathbb{Z}_n^* \times \mathbb{Z}_m^* \tag{13.33}$$

In particular, ϕ is a multiplicative function: $\phi(nm) = \phi(n)\phi(m)$ if $(n,m) = 1$. Therefore, if $N = p_1^{e_1} \cdots p_r^{e_r}$ is the decomposition of N into distinct prime powers then

$$(\mathbb{Z}/N\mathbb{Z})^* \cong (\mathbb{Z}/p_1^{e_1}\mathbb{Z})^* \times \cdots (\mathbb{Z}/p_r^{e_r}\mathbb{Z})^* \tag{13.34}$$

Moreover, $(\mathbb{Z}/p^e\mathbb{Z})^*$ is of order $\phi(p^e) = p^e - p^{e-1}$, as is easily shown [141] and hence

$$\phi(N) = \prod_i (p_i^{e_i} - p_i^{e_i - 1}) = N \prod_{p|N} (1 - \frac{1}{p}) \tag{13.35}$$

Remark: For later reference in our discussion of cryptography note one consequence of this: If we choose, randomly - i.e. with uniform probability density - a number between 1 and N the probability that it will be relatively prime to N is

$$\frac{\phi(N)}{N} = \prod_{p|N} (1 - \frac{1}{p}) \tag{13.36}$$

[140] Do not confuse $\phi(N)$ with the ϕ_a above!

[141] Proof: The numbers between 1 and p^e which have gcd larger than one must be of the form px where $1 \le x \le p^{e-1}$. So the rest are relatively prime.

This means that, if N is huge and a product of just a few primes, then a randomly chosen number will almost certainly be relatively prime to N.

In elementary number theory textbooks it is shown that if p is an odd prime then $(\mathbb{Z}/p^e\mathbb{Z})^*$ is a cyclic group.

♣Finish proof for $e > 1$. ♣

To prove this let us begin with $(\mathbb{Z}/p\mathbb{Z})^*$. (This proof uses some ideas from the algebra of fields.) Suppose this group were not cyclic. Then there would be some n which is a nontrivial divisor of the order, $\phi(p) = p - 1$ such that $x^n = 1$ for all $x \in (\mathbb{Z}/p\mathbb{Z})^*$. That would imply that in the field $\mathbb{F}_p$ the equation $x^n - 1$ would have $p - 1$ distinct roots. On the other hand, the equation $x^n - 1$ can have at most n roots, and that is a contradiction. We conclude that, in fact,$(\mathbb{Z}/p\mathbb{Z})^*$ must be cyclic.

Of course, all primes are odd, and two is the oddest prime of all. If $p = 2$ the result is a little different and we have:

$$(\mathbb{Z}/4\mathbb{Z})^* \cong \{\pm 1\} \tag{13.37}$$

is cyclic but

$$(\mathbb{Z}/2^e\mathbb{Z})^* = \{(-1)^a 5^b | a = 0, 1, 0 \leq b < 2^{e-2}\} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2^{e-2}\mathbb{Z}) \tag{13.38}$$

when $e \geq 3$.

In fact, more generally, it turns out that $(\mathbb{Z}/n\mathbb{Z})^*$ is cyclic iff $n \in \{1, 2, 4, p^k, 2p^k\}$ where p runs over odd primes and $k > 0$.

Note that if we take a product of two distinct odd prime powers then

$$(\mathbb{Z}/(p_1^{k_1} p_2^{k_2}\mathbb{Z})^* \cong (\mathbb{Z}/p_1^{k_1}\mathbb{Z})^* \times (\mathbb{Z}/p_2^{k_2}\mathbb{Z})^* \tag{13.39}$$

But $\phi(p_1^{k_1})$ and $\phi(p_2^{k_2})$ are both even, being divisible by $p_1 - 1$ and $p_2 - 1$, respectively, and hence are not relatively prime, and hence $(\mathbb{Z}/(p_1^{k_1} p_2^{k_2}\mathbb{Z})^*$ is not cyclic.

Examples

1. $(\mathbb{Z}/7\mathbb{Z})^* = \{1, 2, 3, 4, 5, 6\}\text{mod}7 \cong \mathbb{Z}_6$. Note that 3 and 5 are generators:

$$3^1 = 3, \quad 3^2 = 2, \quad 3^3 = 6, \quad 3^4 = 4, \quad 3^5 = 5, \quad 3^6 = 1 \qquad \text{mod}7 \tag{13.40}$$

$$5^1 = 5, \quad 5^2 = 4, \quad 5^3 = 6, \quad 5^4 = 2, \quad 5^5 = 3, \quad 5^6 = 1 \qquad \text{mod}7 \tag{13.41}$$

   However, $2 = 3^2\text{mod}7$ is *not* a generator, even though it is prime. Rather, it generates an index 2 subgroup $\cong \mathbb{Z}_3$, as does 4, while 6 generates an index 3 subgroup $\cong \mathbb{Z}_2$. Do not confuse this isomorphic copy of $\mathbb{Z}_6$ with the additive presentation $\mathbb{Z}_6 \cong \mathbb{Z}/6\mathbb{Z} = \{0, 1, 2, 3, 4, 5\}$ with the *additive* law. Then 1 and 5 are generators, but not $2, 3, 4$.

2. $(\mathbb{Z}/9\mathbb{Z})^* = \{1, 2, 4, 5, 7, 8\}\text{mod}9 \cong \mathbb{Z}_6$. It is a cyclic group generated by 2 and $2^5 = 5\text{mod}9$, but it is not generated by $2^2 = 4$, $2^3 = 8$ or $2^4 = 7\text{mod}9$, because $2, 3, 4$ are not relatively prime to 6.

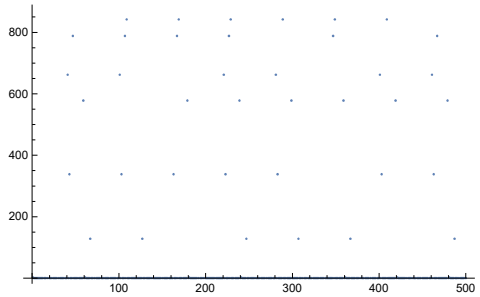


Figure 33: A plot of the residues of $2^x \delta_x$ modulo $N = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 2310$, for $1 \leq x \leq 500$. Here $\delta_x = 0$ if $gcd(x, N) > 1$ so that we only see the values in $(\mathbb{Z}/N\mathbb{Z})^*$. Notice the apparently random way in which the value jumps as we increase x.

3. $(\mathbb{Z}/8\mathbb{Z})^* = \{1, 3, 5, 7\} \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. Note that $3^2 = 5^2 = 7^2 = 1 \text{mod} 8$ and $3 \cdot 5 = 7 \text{mod} 8$, so we can take 3 and 5 to be the generators of the two $\mathbb{Z}_2$ subgroups.

4. $(\mathbb{Z}/15\mathbb{Z})^* \cong \mathbb{Z}_2 \times \mathbb{Z}_4$ is not cyclic.

Remarks

1. When $(\mathbb{Z}/n\mathbb{Z})^*$ is cyclic a generator is called a *primitive root modulo n*, and is not to be confused with a primitive n^{th} root of one. It is trivial to find examples of the latter and highly nontrivial to find examples of the former.

2. The values of $f(x) = a^x \mathrm{mod} N$ for $(a, N) = 1$ appear to jump about randomly as a function of x, as shown in Figure 33. Therefore, finding the period of this function, that is, the smallest positive integer r so that $f(x + r) = f(x)$ is not easy. This is significant because of the next remark.

3. *Factoring Integers.* Suppose N is a positive integer and a is a positive integer so that $(a, N) = 1$, and the order, denoted r, of $\bar{a} \in (\mathbb{Z}/N\mathbb{Z})^*$ is <u>even</u> and finally suppose that $b := a^{r/2} \neq \pm 1 \mathrm{mod} N$. Note that $b^2 = 1 \mathrm{mod} N$ so $\bar{b}$ is a nontrivial squareroot of $\bar{1} \in (\mathbb{Z}/N\mathbb{Z})^*$. Then we claim that $d_\pm := gcd(b \pm 1, N)$ are in fact <u>nontrivial</u> factors of N. To see this we need to rule out $d_\pm = 1$ and $d_\pm = N$, the trivial factors of N. If we had $d_\pm = N$ then N would divide $b \pm 1$ but that would imply $b = \mp 1 \mathrm{mod} N$, contrary to assumption. Now, suppose $d_\pm = 1$, then by Bezout's theorem there would be integers $\alpha_\pm, \beta_\pm$ so that

$$(b \pm 1)\alpha_\pm + N\beta_\pm = 1 \tag{13.42}$$

But then multiply the equation by $b \mp 1$ to get

$$(b^2 - 1)\alpha_\pm + N\beta_\pm(b \mp 1) = b \mp 1 \tag{13.43}$$

But now, N divides the LHS so $b \mp 1 = 0 \mathrm{mod} N$ which implies $b = \pm 1 \mathrm{mod} N$, again contrary to assumption. Thus, $d_\pm$ are <u>nontrivial</u> divisors of N.

To give a concrete example, take $N = 3 \cdot 5 \cdot 7 = 105$, so $\phi(N) = 48$. Then the period of $f(x) = 2^x$ is $r = 12$, and $b = 2^{12/2} = 64$. Well $gcd(64 + 1, 105) = 5$ and $gcd(64 - 1, 105) = 21$ are both divisors of 105. In fact $105 = 5 \cdot 21$.

4. *Artin's Conjecture*: Finding a generator is not always easy, and it is related to some deep conjectures in number theory. For example, the Artin conjecture on primitive roots states that for any positive integer a which is not a perfect square there are an infinite number of primes so that $\bar{a}$ is a generator of the cyclic group $(\mathbb{Z}/p\mathbb{Z})^*$. In fact, if a is not a power of another integer, and the square-free part of a is not $1 \mathrm{mod} 4$ then Artin predicts the density of primes for which a is a generator to be

$$\prod_{Artin\ primes} \left(1 - \frac{1}{p(p-1)}\right) = 0.37.... \tag{13.44}$$

According to the Wikipedia page, there is not a single number a for which the conjecture is known to be true. For example, the primes $p < 500$ for which $a = 2$ is a

generator of $(\mathbb{Z}/p\mathbb{Z})^*$ is

$$\{3,5,11,13,19,29,37,53,59,61,67,83,101,107,131,139,149,163,173,179,181,197,211,227,269,293,317,347,349,373,379,389,419,421,443,461,467,491\} \tag{13.45}$$

5. A good reference for this material is Ireland and Rosen, *A Classical Introduction to Modern Number Theory* Springer GTM

Exercise *Euler's theorem and Fermat's little theorem*

a.) Let G be a finite group of order n. Show that if $g \in G$ then $g^n = e$ where e is the identity element.

b.) Prove *Euler's theorem*: For all integers a relatively prime to N, $g.c.d(a,N) = 1$,

$$a^{\phi(N)} = 1 \mathrm{mod} N \tag{13.46}$$

Note that a special case of this is Fermat's little theorem: If a is an integer and p is prime then

$$a^p = a \mathrm{mod} p \tag{13.47}$$

Remark: This theorem has important practical applications in *prime testing*. If we want to test whether an odd integer n is prime we can compute $2^n \mathrm{mod} n$. If the result is $\neq 2 \mathrm{mod} n$ then we can be sure that n is not prime. Now $2^n \mathrm{mod} n$ can be computed *much* more quickly with a computer than the traditional test of seeing whether the primes up to $\sqrt{n}$ divide n. If $2^n \mathrm{mod} n$ is indeed $= 2 \mathrm{mod} n$ then we can suspect that n is prime. Unfortunately, there are composite numbers which will masquerade as primes in this test. They are called "base 2 pseudoprimes." In fact, there are numbers n, known as *Carmichael numbers* which satisfy $a^n = a \mathrm{mod} n$ for all integers a. The good news is that they are rare. The bad news is that there are infinitely many of them. According to Wikipedia the first few Carmichael numbers are

$$561, 1105, 1729, 2465, 2821, 6601, 8911, 10585, \ldots, \tag{13.48}$$

The first Carmichael number is $561 = 3 \cdot 11 \cdot 17$ and Erdös proved that the number $C(X)$ of Carmichael numbers smaller than X is bounded by

$$C(X) < X \exp\left(-\frac{\kappa \log X \log\log\log X}{\log\log X} \right) \tag{13.49}$$

where κ is a positive real number.

Exercise *Periodic Functions*

a.) Consider the function

$$f(x) = 2^x \text{mod } N \tag{13.50}$$

for an odd integer N. Show that this function is periodic $f(x+r) = f(x)$ for a minimal period r which divides $\phi(N)$.

b.) Compute the period for $N = 15, 21, 105$. [142]

c.) More generally, if $(a, N) = 1$ show that $f(x) = a^x \text{ mod } N$ is a periodic function.

Exercise *How Many Primitive Roots Of n Are There?*

Show that n has either zero or $\phi(\phi(n))$ different primitive roots.

13.2 Group theory and cryptography

Any invertible map $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ can be used to define a code. For example, if $N = 26$ we may identify the elements in $\mathbb{Z}/26\mathbb{Z}$ with the letters in the Latin alphabet:

$$a \leftrightarrow \bar{0}, b \leftrightarrow \bar{1}, c \leftrightarrow \bar{2}, \ldots \tag{13.51}$$

Exercise *Caesar Shift*

a.) Show that $f(m) = (m-3)\text{mod}26$ defines a code. In fact, the above remark, and this example in particular, is attributed to Julius Caesar. Using this decode the message:

$$ZOLPPQEBORYFZLK! \tag{13.52}$$

b.) Is $f(m) = (3m)\text{mod}26$ a valid code? By adding symbols or changing the alphabet we can change the value of N above. Is $f(m) = (3m)\text{mod}27$ a valid code?

The RSA public key encryption system is a beautiful application of Euler's theorem and works as follows. The basic idea is that with numbers with thousands of digits it is relatively easy to compute powers $a^n \text{mod} m$ and greatest common divisors, but it is very difficult to factorize such numbers into their prime parts. For example, for a 1000 digit number the brute force method of factorization requires that we sample up to

$$\sqrt{10^{1000}} = 10^{500} \tag{13.53}$$

[142] *Answer*: $r = 4, 6, 12$ divides $\phi(N) = 8, 12, 48$.

divisors. Bear in mind that our universe is about $\pi \times 10^7 \times 13.79 \times 10^9 \cong 4 \times 10^{17}$ seconds old. [143] There are of course more efficient algorithms, but all the publicly known ones are still far too slow.

Now, Alice wishes to receive and decode secret messages sent by any member of the public. She chooses two large primes (thousands of digits long) p_A, q_A and computes $n_A := p_A q_A$. These primes are to be kept secret. How does she find her secret thousand-digit primes? She chooses a random thousand digit number and applies the Fermat primality test. By the prime number theorem she need only make a few thousand attempts, and she will find a prime. [144]

Next, Alice computes $\phi(n_A) = (p_A - 1)(q_A - 1)$, and then she chooses a random thousand-digit number d_A such that $gcd(d_A, \phi(n_A)) = 1$ and computes an inverse $d_A e_A = 1 \mathrm{mod} \phi(n_A)$. All these steps are relatively fast and easy, because Euclid's algorithm is very fast. Thus there is some integer f so that

$$d_A e_A - f\phi(n_A) = 1 \tag{13.54}$$

That is, she solves the congruence $x = 1 \mathrm{mod} \phi(n_A)$ and $x = 0 \mathrm{mod} d_A$, for the smallest positive x and then computes $e_A = x/d_A$.

Finally, she publishes for the world to see the encoding key: $\{n_A, e_A\}$, but she keeps the numbers $p_A, q_A, \phi(n_A), d_A$ secret. This means that if anybody, say Bob, wants to send Alice a secret message then he can do the following:

Bob converts his plaintext message into a number less than n_A by writing $a \leftrightarrow 01$, $b \leftrightarrow 02$, ..., $z \leftrightarrow 26$. (Thus, when reading a message with an odd number of digits we should add a 0 in front. If the message is long then it should be broken into pieces of length smaller than n_A.) Let Bob's plaintext message thus converted be denoted m. It is a positive integer smaller than n_A.

Now to compute the ciphertext Bob looks up Alice's numbers $\{n_A, e_A\}$ on the public site and uses these to compute the ciphertext:

$$c := m^{e_A} \mathrm{mod} n_A \tag{13.55}$$

Bob sends the ciphertext c to Alice over the internet. Anyone can read it.

Then Alice can decode the message by computing

$$\begin{aligned} c^{d_A} \mathrm{mod} n_A &= m^{e_A d_A} \mathrm{mod} n_A \\ &= m^{1 + f\phi(n_A)} \mathrm{mod} n_A \\ &= m \mathrm{mod} n_A \end{aligned} \tag{13.56}$$

Thus, to decode the message Alice just needs one piece of private information, namely the integer d_A.

[143] There are $\pi \times 10^7$ seconds in a year, to 0.3% accuracy.

[144] The prime number theorem says that if $\pi(x)$ is the number of primes between 1 and x then as $x \to \infty$ we have $\pi(x) \sim \frac{x}{\log x}$. Equivalently, the nth prime is asymptotically like $p_n \sim n \log n$. This means that the density of primes for large x is $\sim 1/\log x$, so if $x \sim 10^n$ then the density is $1/n$ so if we work with thousand-digit primes then after about one thousand random choices we will find a prime.

Now Eve, who has a reputation for making trouble, cannot decode the message without knowing d_A. Just knowing n_A and e_A but not the prime factorization $n_A = p_A q_A$ there is no obvious way to find d_A. The reason is that even though the number n_A is public it is hard to compute $\phi(n_A)$ without knowing the prime factorization of n_A. Of course, if Eve finds out about the prime factorization of n_A then she can compute $\phi(n_A)$ immediately and then quickly (using the Euclidean algorithm) invert e_A to get d_A. Thus, the security of the method hinges on the inability of Eve to factor n_A into primes.

In summary,

1. The intended receiver of the message, namely Alice in our discussion, knows
$$(p_A, q_A, n_A = p_A q_A, \phi(n_A) = (p_A - 1)(q_A - 1), e_A, d_A). \tag{13.57}$$
2. Alice publishes (n_A, e_A). Anybody can look these up.
3. The sender of the message, namely Bob in our discussion, takes a secret message m_B and computes the ciphertext $c = m_B^{e_A} \mathrm{mod} n_A$.
4. Alice can decode Bob's message by computing $m_B = c^{d_A} \mathrm{mod} n_A$ using her secret knowledge of d_A.
5. The attacker, namely Eve in our discussion, knows (n_A, e_A, c) but will have to work to find d_A or some other way of decoding the ciphertext.

Remarks

1. Note that the decoding will *fail* if m and n_A have a common factor. However, $n_A = p_A q_A$ and p_A, q_A are primes with thousands of digits. The probability that Bob's message is one of these is around 1 in 10^{1000}.

Exercise *Your turn to play Eve*
Alice has published the key
$$(n = 661643, e = 325993) \tag{13.58}$$
Bob sends her the ciphertext in four batches:
$$c_1 = 541907 \quad c_2 = 153890 \quad c_3 = 59747 \quad c_4 = 640956 \tag{13.59}$$
What is Bob's message? [145]

---

[145] Factor the integer $n = 541 * 1223$. Then you know p, q and hence $\phi(n) = 659880$. Now take e and compute d by using the Chinese Remainder theorem to compute $x = 1 \mathrm{mod} \phi$ and $x = 0 \mathrm{mod} e$. This gives $x = 735766201 = de$ and hence $d = 2257$. Now you can compute the message from the ciphertext $m = c^d \mathrm{mod} n$.

13.2.1 How To Break RSA: Period Finding

The attacker, Eve, can read the ciphertext $c \mathrm{mod} n_A$. That means the attacker can try to compute the period of the function

$$f(x) := c^x \mathrm{mod} n_A \tag{13.60}$$

Suppose (as is extremely likely when n_A is a product of two large primes) that c is relatively prime to n_A. Then the cyclic group $\langle c \rangle \in (\mathbb{Z}/n_A\mathbb{Z})^*$ generated by c must coincide with the cyclic group generated by the message m_B and in particular they both have the same period r, which divides $\phi(n_A)$. Suppose Eve figures out the period r. Since the published value e_A is relatively prime to $\phi(n_A)$ it will be relatively prime to r and therefore there exists a new decoding method: Compute d_E such that

$$e_A d_E = 1 \bmod r \tag{13.61}$$

Then

$$c^{d_E} = m^{e_A d_E} \mathrm{mod} n_A = m_B^{1+\ell r} \mathrm{mod} n_A = m_B \mathrm{mod} n_A \tag{13.62}$$

decodes the message.

Thus, if the attacker can find the period of $f(x)$ the message can be decoded.

Another way in which finding the period leads to rapid decoding is through explicit factoring:

We saw in our discussion of $\mathbb{Z}_N^*$ that, if one has an element $\bar{a} \in \mathbb{Z}_N^*$ with even period r and $\bar{b} = \bar{a}^{r/2} \neq \overline{\pm 1}$ then $d_\pm = gcd(b \pm 1, N)$ are nontrivial factors of N. Suppose there were a quick method to find the period r. Then we could quickly factor N as follows:

1. Choose a random integer a and using Euclid check that $(a, N) = 1$. If N is a product of two large primes you will only need to make a few choices of a before succeeding.

2. Compute the period r of the function $f(x) = a^x \mathrm{mod} N$.

3. If r is odd go back and choose another a until you get one with r even.

4. Then check that $b = a^{r/2} \neq -1 \mathrm{mod} N$. Again this can be done quickly, thanks to Euclid. If you get $b = -1 \mathrm{mod} N$ go back and choose another a, until you find one that works. The point is that, with high probability, if you pick a at random you will succeed. So you might have a try a few times, but not many.

So, the only real bottleneck in factoring N is computing the order r of $\bar{a}$ in $\mathbb{Z}_N^*$. Equivalently, this is computing the period of the function $f(x) = a^x \bmod N$ where $(a, N) = 1$. This is where the "quantum Fourier transform" and "phase estimation" come in. Quantum computers give a way to compute this period in polynomial time in N, as opposed to classical computers which take exponential time in N. We will come back to this.

13.2.2 Period Finding With Quantum Mechanics

♣This section is out of place. Goes later in the course ♣

Here we sketch how quantum computation allows one to find the period of the function $f(x) = a^x \mathrm{mod} N$ where $(a, N) = 1$. This is just a sketch. A nice and clear and elementary account (which we used heavily) can be found in D. Mermin's book *Quantum Computer Science* and more details and a more leisurely discussion can be found there.

Quantum computation is based on the action of certain unitary operators on a system of n Qbits, that is, on a Hilbert space

$$\mathcal{H}_n = (\mathbb{C}^2)^{\otimes n} \tag{13.63}$$

equipped with the standard inner product. For each factor $\mathbb{C}^2$ one chooses a basis $\{|0\rangle, |1\rangle\}$, which one should think of as, for example spin up/down eigenstates of an electron or photon helicity polarization states. Then there is a natural basis for $\mathcal{H}_n$:

$$|\vec{x}\rangle := |x_{n-1}\rangle \otimes \cdots \otimes |x_1\rangle \otimes |x_0\rangle \tag{13.64}$$

Here, for each i, $x_i \in \{0, 1\}$. One can identify the vector $\vec{x} \in \mathbb{F}_2^n$, the n-dimensional vector space over the field $\mathbb{F}_2$. In our discussion we will only use its Abelian group structure, so one can also think of it as $\mathbb{Z}_2 \oplus \cdots \oplus \mathbb{Z}_2$ with n summands. The basis of states (13.64) is known as the *computational basis* or, the *Classical basis*. Now to each computational basis vector we can assign an integer by its binary expansion:

$$N(\vec{x}) := 2^{n-1}x_{n-1} + \cdots + 2^2 x_2 + 2^1 x_1 + x_0 \tag{13.65}$$

Now, let $N := 2^n$. We can also define the Hilbert space $L^2(\mathbb{Z}_N)$ of functions on the group $\mathbb{Z}_N$ with the natural Haar measure. Of course $\mathcal{H}_n$ is isomorphic to $L^2(\mathbb{Z}_N)$ and the isomorphism we choose to use is the one which identifies the computational basis vector $|\vec{x}\rangle$ with the delta function supported at $N(\vec{x})\mathrm{mod}\ N$. We will denote the latter states as $|N(\vec{x})\rangle\rangle_N$, where the subscript indicates which Abelian group $\mathbb{Z}_N$ we are working with.

In quantum computation one works with a Hilbert space decomposed as

$$\mathcal{H} = \mathcal{H}_{input} \otimes \mathcal{H}_{output} \tag{13.66}$$

The two factors have dimension $N_{in} = 2^{n_{in}}$ and $N_{out} = 2^{n_{out}}$, respectively. The quantum gates are unitary operators and, moreover, under identification of $\mathcal{H}$ as a tensor product of Qbits there should be a notion of "locality" in the sense that they only act nontrivially on "a few" adjacent factors. The locality reflects the spatial locality in some realization in the lab in terms of, say, spin systems. Moreover, we should only have to apply " a few" quantum gates in a useful circuit. With an arbitrary number of gates we can construct any unitary out of products of local ones to arbitrary accuracy. The above notions can be made precise, but that is beyond the scope of this section.

♣Still, it is essential to explain more about the notion of "local quantum gate" and quantum circuit and illustrate a few examples of simple gates. ♣

Now suppose we have a function

$$f : (\mathbb{Z}_N)^* \to (\mathbb{Z}_N)^* \tag{13.67}$$

(such as $f(x) = a^x \mathrm{mod} N$ for $(a, N) = 1$). We would like to convert this to a map

$$\check{f} : \mathbb{F}_2^{n_{in}} \to \mathbb{F}_2^{n_{out}} \tag{13.68}$$

We now choose a fundamental domain which is a subset of $\{1, 2, \ldots, N-1\}$ for $(\mathbb{Z}_N)^*$ with $N < N_{out}$ and $N < N_{in}$ (in fact we will eventually assume $N \ll N_{out}$ and $N \ll N_{in}$) so that we can view elements of $(\mathbb{Z}/N\mathbb{Z})^*$ as elements of the set $\{1, 2, \ldots, N-1\}$ which is, in

turn, a subset of $\mathbb{Z}/N_{in}\mathbb{Z}$ and $\mathbb{Z}/N_{out}\mathbb{Z}$. We use the function $N(\vec{x})$ above to define $\check{f}$ such that

$$f(N(\vec{x})) = N(\check{f}(\vec{x}))\mathrm{mod}2^{n_{out}} \tag{13.69}$$

This does not uniquely specify $\check{f}$ but the ambiguity will not affect the discussion. To read this equation, suppose you want to compute $\check{f}(\vec{x})$ for some $\vec{x} \in \mathbb{F}_2^{n_{in}}$. Then you compute $N(\vec{x})$ which is a nonnegative integer between 0 and N_{in}. Then you reduce it modulo N. If it is relatively prime to N you can compute $f(N(\vec{x}))$ and considerate the result as a number between 1 and $2^{n_{out}} - 1$. The above equation then pins down $\check{f}(\vec{x})$. Using $\check{f}$ we can define a unitary operator U_f by its action on the computational basis:

$$U_f : |\vec{x}\rangle \otimes |\vec{y}\rangle \to |\vec{x}\rangle \otimes |\vec{y} + \check{f}(\vec{x})\rangle \tag{13.70}$$

where on the right-hand side addition is in the Abelian group $(\mathbb{Z}_2)^{n_{out}}$. We will say that the function f *is nice* if U_f can be implemented with a "reasonable" number of local unitary gates. (Of course, one could make this notion much more precise.)

♣And what if $N(\vec{x})$ is not relatively prime to N? Of course, we are thinking this is rare, but it can happen. What is the best way to extend the function? ♣

A good example of a local unitary operator on a Qbit is the *Hadamard gate* that acts by

$$\begin{aligned} H : |0\rangle &\to \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ H : |1\rangle &\to \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{aligned} \tag{13.71}$$

This can be summarized by the formula

$$H|y\rangle = \frac{1}{\sqrt{2}} \sum_x (-1)^{xy} |x\rangle \tag{13.72}$$

So,

$$H^{\otimes n}|\vec{y}\rangle = \left(\frac{1}{\sqrt{2}}\right)^n \sum_{\vec{x}\in\mathbb{F}_2^n} (-1)^{\vec{x}\cdot\vec{y}} |\vec{x}\rangle \tag{13.73}$$

and in particular

$$H^{\otimes n}|\vec{0}\rangle = \left(\frac{1}{\sqrt{2}}\right)^n \sum_{\vec{x}\in\mathbb{F}_2^n} |\vec{x}\rangle \tag{13.74}$$

Therefore, (recall that double brackets $|j\rangle\rangle$ refer to the position basis in $L^2(\mathbb{Z}_N)$) we have

$$U_f \left(H^{\otimes n_{in}} \otimes 1\right) : |0\rangle_{in} \times |0\rangle_{out} \to \left(\frac{1}{\sqrt{2}}\right)^{n_{in}} \sum_{j\in\mathbb{Z}_{N_{in}}} |j\rangle\rangle_{N_{in}} \otimes |f(j)\rangle\rangle_{N_{out}} \tag{13.75}$$

Now we wish to apply this to the function $f(x) = a^x \mod N$ with $(a, N) = 1$ and N is the number we would like to factorize. So, in particular we don't want $N = 2^n$ for some n. Nobody will be impressed if you can factor a power of two! Rather, we identify the group $\mathbb{Z}_N^*$, as a set, with the integers $\{1, \ldots, N-1\}$ so that it can be considered to be a subset of the natural fundamental domain $\{0, \ldots, N_{in} - 1\}$ for $\mathbb{Z}_{N_{in}}$, and similarly for

$\mathbb{Z}_{N_{out}}$. Then to compute f we compute a^x, take the residue modulo N to get an integer in the fundamental domain, and then consider that number modulo N_{out}. Hence, we should choose $N_{out} = 2^{n_{out}}$ to be some integer larger than N. A key claim, explained in textbooks on quantum information theory is that such a function f is nice. That is, it makes sense to compute it with a quantum circuit.

So, we conclude that a suitable quantum circuit can implement:

$$\begin{aligned} U_f\left(H^{\otimes n}\otimes 1\right) : |0\rangle_{N_{in}} \times |0\rangle_{N_{out}} &\to \left(\frac{1}{\sqrt{2}}\right)^{n_{in}} \sum_{k\in\mathbb{Z}_{N_{in}}} |k\rangle\rangle_{N_{in}} \otimes |a^k\rangle\rangle_{N_{out}} \\ &= \left(\frac{1}{\sqrt{2}}\right)^{n_{in}} {\sum_{f_0\in\mathbb{Z}_N^*}}' (|j_0\rangle\rangle_{N_{in}} + |j_0+r\rangle\rangle_{N_{in}} + |j_0+2r\rangle\rangle_{N_{in}} + \cdots) \otimes |f_0\rangle\rangle_{N_{out}} \end{aligned} \tag{13.76}$$

In the second line we are considering $\mathbb{Z}_N^*$ as a subset of $\mathbb{Z}_{N_{out}}$ as explained above and the prime on the sum means that we are just summing over the values that are in the image of $f(x) = a^x \mathrm{mod}\ N$. This will be all the values in $\mathbb{Z}_N^*$ if a is a generator of $\mathbb{Z}_N^*$ but in general might be smaller. Also, j_0 is some solution of $f_0 = a^{j_0} \bmod N$, and r is the period of $f(x)$, that is, the smallest positive integer so that $f(x+r) = f(x)$ for all x. We can choose j_0 so that $0 \le j_0 < r$ and write the RHS of (13.76) as

$$\Psi = \left(\frac{1}{\sqrt{2}}\right)^{n_{in}} \sum_{0\le j_0<r} \left(\sum_{s=0}^{\mathcal{O}-1} |j_0+sr\rangle\rangle_{N_{in}}\right) \otimes |a^{j_0}\rangle\rangle_{N_{out}} \tag{13.77}$$

Here $\mathcal{O}$ is the smallest integer such that $j_0 + \mathcal{O}r \ge N_{in}$. So

$$\mathcal{O} = \lfloor \frac{N_{in} - j_0}{r} \rfloor \tag{13.78}$$

In the applications we have in mind N_{in} and r are typically very large numbers so that this is a (very weak) function of j_0. At the end of this section we will use the observation that in this case, $r\mathcal{O}/N_{in}$ is, to very good accuracy, just equal to 1.

Now we measure the output system and get some result, say, $f_0 = a^{k_0} \mathrm{mod} N$. Applying the usual Born rule we get the state for the input system

$$P_{f_0}(\Psi) = \frac{1}{\sqrt{\mathcal{O}}} \sum_{s=0}^{\mathcal{O}-1} |k_0+sr\rangle\rangle_{N_{in}} \tag{13.79}$$

It is some kind of plane wave state in $L^2(\mathbb{Z}_N)$, so measuring position will give no useful information on r. Of course, we should therefore go to the Fourier dual basis to learn about the period. In terms of the position basis of $L^2(\mathbb{Z}_N)$ we can apply V_{FT} to get:

$$V_{FT}P_{f_0}\Psi = \frac{1}{\sqrt{N_{in}\mathcal{O}}} \sum_{p\in\mathbb{Z}_{N_{in}}} \sum_{s=0}^{\mathcal{O}-1} e^{2\pi \mathrm{i}\frac{(k_0+sr)p}{N_{in}}} |p\rangle\rangle_{N_{in}} \tag{13.80}$$

Note that this Fourier transform is quite nontrivial and nontransparent in the computational basis because of the nontrivial isomorphism between $\mathcal{H}_n$ and $L^2(\mathbb{Z}_N)$ with $N = 2^n$.

Nevertheless, and this is nontrivial and part of the magic of Shor's algorithm, the Fourier transform operator V_{FT} can be implemented nicely with quantum gates in the computational basis. Again, the textbooks on quantum information theory give explicit construction of V_{FT} as a quantum circuit in the computational basis. It is exactly at this point that the exponential speed-up of the period finding takes place:

1. Classical Fourier Transform: $N^2 = 2^{2n}$ operations. We learn every Fourier coefficient.

2. Fast Fourier Transform: $nN = n2^n$ operations. We learn every Fourier coefficient.

3. Quantum Fourier Transform: n^2 quantum gates. We only learn about correlations of the output state.

Now to find the period we make a measurement of the amplitudes for the various Fourier components $|p\rangle\rangle_{N_{in}}$. (Here we are using the isomorphism between $\mathbb{Z}_{N_{in}}$ and its unitary dual.) The probability to measure p is

$$\begin{aligned} Prob(p) &= \frac{1}{N_{in}\mathcal{O}} \Big| \sum_{s=0}^{\mathcal{O}-1} (e^{2\pi \mathrm{i} \frac{p}{N_{in}/r}})^s \Big|^2 \\ &= \frac{1}{N_{in}\mathcal{O}} \frac{\sin^2\left(\pi \cdot p \cdot \frac{\mathcal{O}r}{N_{in}}\right)}{\sin^2\left(\pi \cdot p \cdot \frac{r}{N_{in}}\right)} \end{aligned} \tag{13.81}$$

The basic idea is that the probability, as a function of p, will be peaked near values of p from which we can deduce the crucial number r, but we need to be a bit careful at this point.

Let us ask what is the probability that we will measure a value p of the form

$$p_j = j\frac{N_{in}}{r} + \delta_j \qquad\qquad |\delta_j| \leq 1/2 \tag{13.82}$$

If p_j is of this form with any value $j = 1, \ldots, r-1$ then we can extract r. Thus, substituting such a value for p_j into the formula for the probability we have

$$\frac{1}{N_{in}\mathcal{O}} \frac{\sin^2\left(\pi \cdot \delta_j \cdot \frac{\mathcal{O}r}{N_{in}}\right)}{\sin^2\left(\pi \cdot \delta_j \cdot \frac{r}{N_{in}}\right)} \tag{13.83}$$

Now recall that $\mathcal{O}r/N_{in}$ is equal to 1 to excellent accuracy. Suppose we also choose a number of Qbits so that

$$N_{in} \gg N > r \tag{13.84}$$

Then the argument of the sign in the denominator is extremely small and we can replace $\sin(x)$ by x. So we get:

$$\begin{aligned} Prob(p_j) &\cong \frac{1}{N_{in}\mathcal{O}} \frac{\sin^2(\pi\delta_j)}{\left(\pi\cdot\delta_j\cdot\frac{r}{N_{in}}\right)^2} \\ &= \frac{N_{in}}{r\mathcal{O}}\cdot\frac{1}{r}\left(\frac{\sin\pi\delta_j}{\pi\delta_j}\right)^2 \\ &\cong \frac{1}{r}\left(\frac{\sin\pi\delta_j}{\pi\delta_j}\right)^2 \end{aligned} \tag{13.85}$$

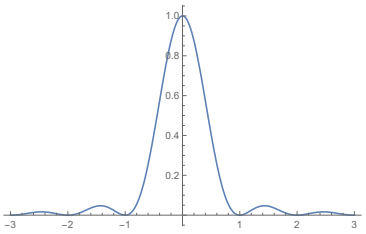


Figure 34: A plot of the function $\sin^2(\pi x)/(\pi x)^2$ as a function of x. This function, very familiar from the theory of diffraction, is symmetric in $x \to -x$ and monotonically decreasing in the interval $0 \le x \le 1/2$. It therefore takes its minimal value in this interval at $x = 1/2$ where it is about $\cong 0.405$.

Now for $0 \le \delta \le \pi/2$ we have $\frac{\sin x}{x} \ge \frac{2}{\pi}$ so

$$\sum_j Prob(p_j) \ge \frac{r-1}{r}\frac{4}{\pi^2} \cong 0.4 \tag{13.86}$$

Now, using various tricks one can raise this 40% value to near 100%. For these tricks consult Mermin's book. Two other useful textbooks on quantum information theory and quantum computing where one can look these things up include:

1. Nielsen and Chuang, Quantum Information Theory
2. A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*, Grad Studies in Math 47, AMS
3. J. Preskill, lecture notes at http://www.theory.caltech.edu/∼preskill/ph219/ph219-2019-20

14. Semidirect Products

We have seen a few examples of direct products of groups above. We now study a more subtle notion, the semidirect product. The semidirect product is a twisted version of the direct product of groups H and G which can be defined once we are given one new piece of extra data. The new piece of data we need is a homomorphism

$$\alpha : G \to \mathrm{Aut}(H). \tag{14.1}$$

For an element $g \in G$ we will denote the corresponding automorphism by α_g. The value of α_g on an element $h \in H$ is denoted $\alpha_g(h)$. Thus $\alpha_g(h_1 h_2) = \alpha_g(h_1)\alpha_g(h_2)$ because α_g is a homomorphism of H to itself while we also have $\alpha_{g_1 g_2}(h) = \alpha_{g_1}(\alpha_{g_2}(h))$ because α is a homomorphism of G into the group of automorphisms $\mathrm{Aut}(H)$. We also have that α_1 is the identity automorphism. (Prove this!)

Using the extra data given by α we can form a more subtle kind of product called the **semidirect product** $H \rtimes G$, or $H \rtimes_\alpha G$ when we wish to stress the role of α. In the math literature on group theory the notation $H : G$ is also used. This group is the Cartesian product $H \times G$ as a *set* but has the "twisted" multiplication law:

$$(h_1, g_1) \cdot (h_2, g_2) := (h_1 \alpha_{g_1}(h_2), g_1 g_2) \tag{14.2}$$

A good intuition to have is that "as g_1 moves from left to right across the h_2 they interact via the action of g_1 on h_2."

Exercise *Due diligence*

a.) Show that (14.2) defines an associative group law.

b.) Show that $(1_H, 1_G)$ defines the unit and

$$(h, g)^{-1} = \left(\alpha_{g^{-1}}(h^{-1}), g^{-1}\right) \tag{14.3}$$

c.) Compute the group commutator $[(h_1, g_1), (h_2, g_2)]$ for a semidirect product. [146]

d.) Let $\mathrm{End}(H)$ be the set of all homomorphisms $H \to H$. Note that this set is closed under the operation of composition, and this operation is associative, but $\mathrm{End}(H)$ is <u>not</u> a group because some homomorphisms will not be invertible. Nevertheless, it is a monoid. Show that if $\alpha_g : G \to \mathrm{End}(H)$ is a homomorphism of monoids then (14.2) still defines a monoid. When is it a group?

Example 14.1: *Infinite dihedral group.* Let $G = \{e, \sigma\} \cong \mathbb{Z}_2$ with generator σ, and let $H = \mathbb{Z}$, written additively. Then define a nontrivial $\alpha : G \to Aut(H)$ by letting α_σ act on $x \in H$ as $\alpha_\sigma(x) = -x$. Then $\mathbb{Z} \rtimes \mathbb{Z}_2$ is a group with elements (x, e) and (x, σ), for $x \in \mathbb{Z}$. Note the multiplication laws:

$$\begin{aligned}
(x_1, e)(x_2, e) &= (x_1 + x_2, e) \\
(x_1, e)(x_2, \sigma) &= (x_1 + x_2, \sigma) \\
(x_2, \sigma)(x_1, e) &= (x_2 - x_1, \sigma) \\
(x_1, \sigma)(x_2, \sigma) &= (x_1 - x_2, e)
\end{aligned} \tag{14.4}$$

and hence the resulting group is nonabelian with this twisted multiplication law. Since $Aut(\mathbb{Z}) \cong \mathbb{Z}_2$ this is the only nontrivial semidirect product we can form. This group is known as the *infinite dihedral group* sometimes denoted D_∞. It has a presentation:

$$\mathbb{Z} \rtimes \mathbb{Z}_2 \cong \langle r, s | s^2 = 1 \qquad srs = r^{-1} \rangle \tag{14.5}$$

[146] *Answer*: $[(h_1, g_1), (h_2, g_2)] = \left(h_1 \alpha_{g_1}(h_2) \alpha_{g_1 g_2 g_1^{-1}}(h_1^{-1}) \alpha_{g_1 g_2 g_1^{-1} g_2^{-1}}(h_2^{-1}), g_1 g_2 g_1^{-1} g_2^{-1}\right)$.

(e.g. take $s = (0, \sigma)$ and $r = (1, e)$)

Remark: Taking $x = s$ and $y = rs$ we see that D_∞ also has a presentation as a Coxeter group:

$$\mathbb{Z} \rtimes \mathbb{Z}_2 \cong \langle x, y | x^2 = 1 \qquad y^2 = 1 \rangle \tag{14.6}$$

Indeed it is the Weyl group for the affine Lie group $LSU(2)$.

Example 14.2: *Finite dihedral group.* We can use the same formulae as in Example 1, retaining $G = \{e, \sigma\} \cong \mathbb{Z}_2$ but now we take $H = \mathbb{Z}/N\mathbb{Z}$. We still have

$$\alpha_\sigma : \bar{n} \to -\bar{n} \tag{14.7}$$

where we are writing $\mathbb{Z}/N\mathbb{Z}$ additively. The semi-direct product of $\mathbb{Z}/N\mathbb{Z}$ with $\mathbb{Z}_2$ using this automorphism gives one definition of an important group, the *finite dihedral group* D_N. Observe that, when we write $\mathbb{Z}_N = \mu_N$ multiplicatively, the automorphism is $\alpha_\sigma(\omega^j) = \omega^{-j}$. In this way we can obtain a presentation of D_N of the form:

$$\mathbb{Z}_N \rtimes_\alpha \mathbb{Z}_2 \cong \langle r, s | s^2 = 1, \qquad srs = r^{-1}, \qquad r^N = 1 \rangle \tag{14.8}$$

Note that here we have switched to a multiplicative model for the group $\mathbb{Z}_N$. The group has the order given by the cardinality of the Cartesian product $\mathbb{Z}_N \times \mathbb{Z}_2$ so it has order $2N$. Note that using the relations, every word in the r, s can be converted to the form r^x or $r^x s$ with $0 \leq x \leq N - 1$, thus accounting for all $2N$ elements.

$$|D_N| = 2N. \tag{14.9}$$

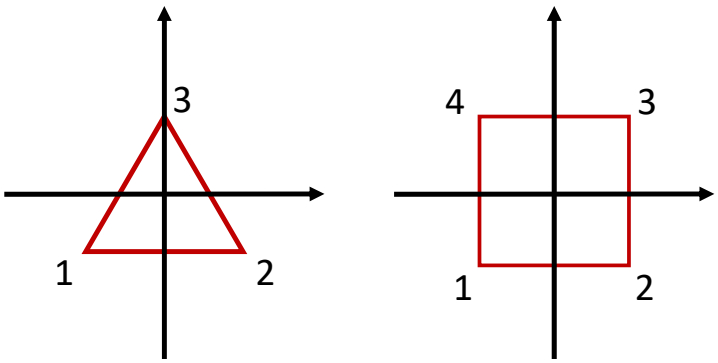


Figure 35: A regular 3-gon and 4-gon in the plane, centered at the origin. The subgroup of $O(2)$ that preserves these is D_3 and D_4, respectively.

Important Remark: *The Dihedral Groups And Symmetries Of Regular Polygons.* The group D_N has a natural action on the vector space $\mathbb{R}^2$. The generator r acts by a

rotation around the origin: $\phi_r = R(2\pi/N)$. This generates a group isomorphic to $\mathbb{Z}_N$ and in this context it is usually denoted C_N. If P is any reflection through a line through the origin then $\phi_s = P$ will satisfy all the relations. The resulting group of transformations of the plane generated by ϕ_r and ϕ_s is isomorphic to D_N. Moreover, if we consider the regular N-gon centered at the origin of the plane $\mathbb{R}^2$ then the subgroup of $O(2)$ that maps it to itself is isomorphic to D_N, although to preserve the polygon we must choose P carefully so that it is a reflection through an axis of symmetry. For example, if we consider the regular triangle illustrated in 35 then reflection in the y-axis is a symmetry, as is rotation by integral amounts of $2\pi/3$. So we have a two-dimensional matrix representation of D_3:

$$\begin{aligned} s \to P &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ r \to R(2\pi/3) &= \begin{pmatrix} \cos(2\pi/3) & \sin(2\pi/3) \\ -\sin(2\pi/3) & \cos(2\pi/3) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix} \end{aligned} \tag{14.10}$$

Note that, if we label the vertices of the triangle $1, 2, 3$ as shown in the figure then the various symmetries are in 1-1 correspondence with permutations of $\{1, 2, 3\}$. So in fact, we have an isomorphism

$$D_3 \cong S_3 \tag{14.11}$$

with $P \to (12)$ and $R(2\pi/3) \to (123)$. Similarly, D_4 is the group of symmetries of the square. We can take $s \to P$ as before and now $r \to R(2\pi/4)$. Again we can label the vertices of the square $1, 2, 3, 4$. Again transformations are uniquely determined by permutations of $\{1, 2, 3, 4\}$. However, the group of permutations we get this way is only a subgroup of the permutation group S_4.

Clearly there is something quite different about the groups D_N when N is even and odd. This is nicely seen in the set of conjugacy classes. As you show in the exercise below the conjugacy classes in D_N are of the form

$$C(r^x) = \{r^x, r^{-x}\} \tag{14.12}$$

and

$$C(r^x s) = \{r^{x+2y} s | y \in \mathbb{Z}\} \tag{14.13}$$

Now, thinking in terms of symmetry actions on the plane, r^x correspond to rotations by $2\pi x/N$, whereas $r^x s$ correspond to reflections. Now note that for N odd, since $(2, N) = 1$ the conjugacy class $C(r^x s)$ will contain all the elements of the form $r^z s$, in other words all the reflections. However, if N is even then there are two distinct conjugacy classes: $C(r^x s)$ for x even and odd are distinct. This is nicely in accord with symmetries of the N-gon: For N odd it is clear that all the symmetry axes can be mapped to each other by using the symmetries of the N-gon. Whereas for N even there are two distinct kinds of reflection axes: Those that go through vertices and those that go through edges.

Exercise *D_4 and permutations*

Show that under the map to S_4 described above the dihedral group D_4 maps to the subgroup containing just the identity, the cyclic permtuations (the rotations)

$$(1234), (1234)^2 = (13)(24), (1234)^3 = (1432) \tag{14.14}$$

and four reflections:

$$(12)(34), \quad (23)(14), \quad (13), \quad , (24) \tag{14.15}$$

Exercise

Show that D_N is a quotient of the infinite dihedral group. [147]

Exercise *Conjugacy Classes In* D_N

a.) Using the presentation of D_N in terms of generators r, s and relations $s^2 = 1$ and $srs = r^{-1}$, and $r^N = 1$ show that we have conjugacy classes given by (14.12) and (14.13).

b.) List the distinct conjugacy classes in D_N for N even and N odd.

♣Should provide answer in a footnote here. ♣

Example 14.3: In equations (2.22) and (2.40) we found the most general form of a matrix in $O(2)$. It is a disjoint union of two circles, each circle being the elements of determinant $\det(A) = \pm 1$. As a group we have an isomorphism

$$O(2) \cong SO(2) \rtimes \mathbb{Z}_2 \tag{14.17}$$

In fact, there are many such isomorphisms.

We can write an explicit isomorphism as follows: Let $\mathbb{Z}_2 = \{1, \sigma\}$. Then

$$\alpha : \mathbb{Z}_2 \to \mathrm{Aut}(SO(2)) \tag{14.18}$$

is given by

$$\alpha(\sigma) : R(\phi) \to R(-\phi) \tag{14.19}$$

Now choose any nonzero vector $v \in \mathbb{R}^2$ and define an isomorphism

$$\Psi_v : SO(2) \rtimes \mathbb{Z}_2 \to O(2) \tag{14.20}$$

by

$$\begin{aligned} \Psi_v &: (1, \sigma) \to P_v \\ \Psi_v &: (R(\phi), 1) \to R(\phi) \end{aligned} \tag{14.21}$$

[147] *Answer*: Note that $\mathcal{N} = \{(x, e) | x = 0 \mathrm{mod} N\} \subset \mathbb{Z} \rtimes \mathbb{Z}_2$ is a normal subgroup and

$$(\mathbb{Z}/N\mathbb{Z}) \rtimes \mathbb{Z}_2 \cong (\mathbb{Z} \rtimes \mathbb{Z}_2)/\mathcal{N}. \tag{14.16}$$

Here P_v is the reflection in the line orthogonal to v. We need to check that this is a well-defined homomorphism by checking that the images we have specified are indeed compatible with the relations in the semidirect product. This amounts to checking that $P_v^2 = 1$, which is obvious, and

$$P_v R(\phi) P_v^{-1} = R(-\phi) \tag{14.22}$$

Thanks to (14.22) the relations are indeed satisfied and now it is an easy matter to check that Ψ_v is injective and surjective, so it is an isomorphism.

Note that there are many different such isomorphisms, Ψ_v, depending on the choice of v. If v' is another nonzero vector in the plane then recall from (6.11) that $P_v P_{v'} = R(2\theta)$ where θ is the angle between v and v'. Then

$$\begin{aligned} \Psi_{v'} : (1, \sigma) &\mapsto P_{v'} \\ &= (P'_v P_v) P_v \\ &= R(2\theta) P_v \\ &= R(\theta) P_v R(-\theta) \end{aligned} \tag{14.23}$$

So

$$\Psi_{v'} = I(R(\theta)) \circ \Psi_{v'} \tag{14.24}$$

and changing $v \to v'$ changes Ψ_v by composition with an inner automorphism of $O(2)$.

More generally, it is true that:

1. When d is odd
$$O(d) \cong SO(d) \times \mathbb{Z}_2 \tag{14.25}$$

2. When d is even
$$O(d) \cong SO(d) \rtimes \mathbb{Z}_2 \tag{14.26}$$

In the case when d is odd the element $-\mathbf{1}_{d\times d}$ has determinant -1, so it is in the nontrivial component of $O(d)$, and yet it is also central: so conjugating elements $R \in SO(d)$ acts trivially. The semi-direct product is isomorphic to a direct product in this case. On the other hand, when d is even $-\mathbf{1}_{d\times d}$ has determinant $+1$ and is an element of $SO(d)$. However, it is still true that if Ψ_v is reflection in the hyperplane orthogonal to a nonzero vector v then

$$\alpha_v(R) := P_v R P_v \tag{14.27}$$

is a nontrivial automorphism of $SO(d)$ and we have a family of isomorphisms:

$$\Psi_v : SO(d) \rtimes_{\alpha_v} \mathbb{Z}_2 \to O(d) \tag{14.28}$$

The same discussion as above shows that the dependence on v is through composition with an inner automorphism of $SO(d)$.

Example 14.4: *Affine Euclidean Space.* Recall the discussion of Affine Euclidean space in section ******** above. We defined there the Euclidean group.

Some natural examples of isometries are the following: Given any vector $v \in \mathbb{R}^d$ we have the translation operator

$$T_v : p \to p' \tag{14.29}$$

Moreover, if $R \in O(d)$ then, if we choose a point p, we can define an operation:

$$R_p : p + v \to p + Rv \tag{14.30}$$

It turns out (this is a nontrivial theorem) that all isometries are obtained by composing such transforamtions. A simple way to express the general transformation, then, is to choose a point p as the "origin" of the affine space thus giving an identification $\mathbb{E}^d \cong \mathbb{R}^d$. Then, to a pair $R \in O(d)$ and $v \in \mathbb{R}^d$ we can associate the isometry: [148]

$$\{v|R\} : x \mapsto v + Rx \qquad\qquad \forall x \in \mathbb{R}^d \tag{14.31}$$

In this notation the group multiplication law is

$$\{v_1|R_1\}\{v_2|R_2\} = \{v_1 + R_1 v_2|R_1 R_2\} \tag{14.32}$$

which makes clear that there is a nontrivial automorphism used to construct the semidirect product of the group of translations, isomorphic to $\mathbb{R}^d$ with the rotation-inversion group $O(d)$.

Put differently: $O(d)$ acts as an automorphism group of $\mathbb{R}^d$:

$$\alpha_R : v \mapsto Rv \tag{14.33}$$

so we can form the abstract group $\mathbb{R}^d \rtimes_\alpha O(d)$. Then, once we choose an origin $p \in \mathbb{E}^d$ we can write an isomorphism

$$\Psi_p : \mathbb{R}^d \rtimes_\alpha O(d) \to \mathrm{Euc}(d) \tag{14.34}$$

To be concrete:

$$\Psi_p(v, R) : p + x \mapsto p + (v + Rx) \tag{14.35}$$

As in our example of $O(d)$ we now have a family of isomorphisms. If $\Psi_{p'}$ is another such based on a different origin $p' = p + v_0$ then the two isomorphisms are related by a translation. See the exercise below.

Exercise *Internal Definition Of Semidirect Products*

Suppose there is a homomorphism $\alpha : G \to \mathrm{Aut}(H)$ so that we can form the semidirect product $H \rtimes_\alpha G$.

a.) Show that elements of the form $(1, g)$, $g \in G$ form a subgroup $Q \subset H \rtimes G$ isomorphic to G, while elements of the form $(h, 1)$, $h \in H$ constitute another subgroup, call it N, which is isomorphic to H.

[148] Our notation is logically superior to the standard notation in the condensed matter physics literature where it is known as the Seitz notation. In the cond-matt literature we have $\{R|v\} : x \mapsto Rx + v$.

b.) Show that $N = \{(h,1)|h \in H\}$ is a *normal* subgroup of $H \rtimes G$, while $Q = \{(1,g)|g \in G\}$ in general is not a normal subgroup. [149] This explains the funny product symbol $\rtimes$ that looks like a fish: it is a combination of $\times$ with the normal subgroup symbol $\triangleleft$.

c.) Show that we have a short exact sequence:

$$1 \to N \to H \rtimes_\alpha G \to Q \to 1 \tag{14.37}$$

d.) Show that $H \rtimes G = NQ = QN$ and show that $Q \cap N = \{1\}$. Here NQ means the sent of elements nq where $n \in N$ and $q \in Q$. [150]

e.) Conversely, show that if $\tilde{G} = NQ$ where N is a normal subgroup of $\tilde{G}$ and Q is a subgroup of $\tilde{G}$, (that is, every element of $\tilde{G}$ can be written in the form $g = nq$ with $n \in N$ and $q \in Q$ and $N \cap Q = \{1\}$) then $\tilde{G}$ is a semidirect product of N and Q. Show how to recover the action of Q as a group of automorphisms of N by defining $\alpha_q(n) := qnq^{-1}$. Note that α_q in general is *NOT* an inner automorphism of N.

Exercise *When is a semidirect product actually a direct product?*

Show that if $G = NQ$ is a semidirect product and Q is *also* a normal subgroup of G, then G is the direct product of N and Q. [151]

It is useful to think about the Euclidean group in terms of the "internal" characterization of semi-direct products explained in the exercise above. Here we have a normal subgroup $N := \{\{v|1\}|v \in \mathbb{R}^d\}$ and a subgroup Q given by the set of elements of the form $\{0|R\}$. To check that N is normal a short computation using the group law reveals

$$\{v|R\}\{w|1\} = \{Rw|1\}\{v|R\} \tag{14.38}$$

and hence:

$$\{v|R\}\{w|1\}\{v|R\}^{-1} = \{Rw|1\} \tag{14.39}$$

Note that, again, thanks to the group law, $\pi : \{v|R\} \to R$ is a surjective homomorphism $\mathrm{Euc}(d) \to O(d)$. Thus there is an exact sequence:

$$0 \to \mathbb{R}^d \to Euc(d) \to O(d) \to 1 \tag{14.40}$$

[149] Answer to (b): Compute $(h_1,g_1)(h,1)(h_1,g_1)^{-1} = (h_1\alpha_{g_1}(h)h_1^{-1},1)$ and

$$(h_1,g_1)(1,g)(h_1,g_1)^{-1} = (h_1\alpha_{g_1gg_1^{-1}}(h_1^{-1}), g_1gg_1^{-1}). \tag{14.36}$$

[150] The notation is slightly dangerous here: We are considering the group Q both as a subgroup of G and, in equation (14.37), as a quotient. In general, as we will see below in the chapter on exact sequences, there is no way to view a quotient of a group G as a subgroup of G. Failure to appreciate this point has led to many, many, many errors in the physics literature.

[151] *Answer:* Note that $n_1q_1n_2q_2 = n_1n_2(n_2^{-1}q_1n_2q_1^{-1})q_1q_2$. However, if both N and Q are normal subgroups then $(n_2^{-1}q_1n_2q_1^{-1}) \in N \cap Q = \{1\}$. Therefore $n_1q_1n_2q_2 = n_1n_2q_1q_2$ is the direct product structure.

Almost identical considerations show that the Poincaré group is isomorphic to the semidirect product of the translation and Lorentz groups:

$$\text{Poincare}(\mathbb{M}^{1,d-1})) \cong \mathbb{M}^{1,d-1} \rtimes O(1,d-1) \tag{14.41}$$

where, once again, the choice of isomorphism depends on a choice of origin.

Example 14.5: *Wreath Products.* If X and Y are sets then let $\mathcal{F}[X \to Y]$ be the set of functions from X to Y. Recall that

1. If $Y = G_1$ is a group then $\mathcal{F}[X \to G_1]$ is itself a group.
2. If a group G_2 acts on X and Y is any set then G_2 actions on the function space $\mathcal{F}[X \to Y]$ in a natural way.

We can combine these two ideas as follows: Suppose that G_2 acts on a set X and $Y = G_1$ is itself a group. Then let

$$\alpha : G_2 \to \text{Aut}(\mathcal{F}[X \to G_1]) \tag{14.42}$$

be the canonical G_2 action on the function space: so if $\phi : G_2 \times X \to X$ is the action on X (part of our given data) then the induced action on the function space is

$$\alpha_g : F \mapsto \alpha_g(F) \in \mathcal{F}[X \to G_1] \tag{14.43}$$

where we define

$$\alpha_g(F)(x) = F(\phi(g^{-1}, x)) \qquad \forall g \in G_2, x \in X \tag{14.44}$$

Then we can form the semidirect product

$$\mathcal{F}[X \to G_1] \rtimes G_2 \tag{14.45}$$

This is a *generalized wreath product.* The traditional wreath product is a special case where $G_2 = S_n$ for some n and S_n acts on $X = \{1, \ldots, n\}$ by permutations in the standard way. Note that the group $\mathcal{F}[X \to G_1] \cong G_1^n$. The traditional wreath product $G_1 \text{wr} S_n$, also denoted $G_1 \wr S_n$, is then $\mathcal{F}[X \to G_1] \rtimes S_n$. To be quite explicit, the group elements in $G_1 \wr S_n$ are

$$(h_1, \ldots, h_n; \phi) \tag{14.46}$$

with $h_i \in G_1$ and $\phi \in S_n$ and the product is

$$(h_1, \ldots, h_n; \phi)(h'_1, \ldots, h'_n; \phi') = (h_1 h'_{\phi^{-1}(1)}, h_2 h'_{\phi^{-1}(2)}, \ldots, h_n h'_{\phi^{-1}(n)}, \phi \circ \phi') \tag{14.47}$$

Example 14.6: *Kaluza-Klein theory.* The basic idea of Kaluza-Klein theory is that we study physics on a product manifold $\mathcal{X} \times \mathcal{Y}$ and partially rigidify the situation by putting some structure on $\mathcal{Y}$. We then regard $\mathcal{Y}$ as "small" and study the physics as "effectively" taking place on $\mathcal{X}$.

The idea is intuitively understood by imagining a $2+1$ dimensional world where space is a cylinder of radius R. If we imagine beings in this flatland of a fixed lengthscale,

and shrink $R \to 0$ then the beings will end up perceiving themselves as living in a $1+1$ dimensional world.

♣SUITABLE FIGURE NEEDED HERE ♣

Historically, Kaluza-Klein theory arose from attempts to unify the field theories of general relativity with Maxwell's theory of electromagnetism. The basic idea is that pure general relativity on $\mathcal{X} \times \mathcal{Y}$ appears, when $\mathcal{Y}$ is "small" to be a theory of several fields, including electro-magnetism, in $\mathcal{X}$. As originally conceived the idea is very beautiful, but now regarded as too naive and simplistic. Nevertheless, the idea that there might be extra dimensions of spacetime in a compact manifold survives to this day and models that make use of it come astonishingly close to describing the standard model of particle physics and gravity, in the context of "string compactification."

The canonical example of Kaluza-Klein theory is the case where $\mathcal{X} = \mathbb{M}^{1,d-1}$ is d-dimensional Minkowski space and $\mathcal{Y} = S^1$ is the circle. We rigidify the situation by putting a metric on the circle S^1 so that the metric on space-time is

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu + R^2 (d\theta)^2 \tag{14.48}$$

where R is the radius of the circle, $\theta \sim \theta + 2\pi$ and $0 \leq \mu \leq d-1$. Our signature is mostly plus. Now we consider a massless scalar field in $(d+1)$ dimensions on this spacetime. A massless scalar field would satisfy the wave equation:

$$\left[\eta^{\mu\nu} \frac{\partial}{\partial x^\mu} \frac{\partial}{\partial x^\nu} + \frac{1}{R^2} \left(\frac{\partial}{\partial \theta} \right)^2 \right] \phi = 0 \tag{14.49}$$

In Quantum Field Theory one makes a huge leap: The quantization of the field leads to quantum states which are interpreted as the states of a system of particles. An essential step in this feat of magic is that one makes a Fourier-decomposition of the field. The Fourier modes of the field are interpreted as creation/annihilation operators of particle states. For the massless scalar field the Fourier modes are

$$e^{\mathrm{i} p_M x^M} = e^{\mathrm{i} p_\mu x^\mu} e^{\mathrm{i} p_\theta \theta} \tag{14.50}$$

corresponding to single-particle creating and annihilation operators of definite energy-momentum. But since $\theta \sim \theta + 2\pi$ single-valuedness of the field implies that $p_\theta = n \in \mathbb{Z}$ is an integer. But now the wave-equation implies that we have a dispersion relation:

$$E^2 - \vec{p}^2 = \frac{n^2}{R^2} \tag{14.51}$$

where $p_\mu = (E, \vec{p})$. From the viewpoint of a d-dimensional field theory, Fourier modes with $n \neq 0$ describe massive particles with $m^2 = n^2/R^2$.

Now consider the case that R is very small compared to the scale of any observer. Then that observer will perceive only a d-dimensional spacetime. If R is very small the single massless particle in $d+1$-dimensions is percieved as an infinite set of different massive particles with mass $|n|/R$ in d dimensions. As $R \to 0$ the masses of the particles $\sim |n|/R$ goes to infinity. So, except for the $n = 0$ modes, the particles are very massive and therefore will not be created by low energy processes, and are hence in general unobservable. For

example, if R is on the order of the Planck scale then the nonzero Fourier modes are fields that represent particles of Planck-scale mass.

In a similar spirit, one finds that the Einstein-Hilbert action in $(d+1)$ dimensions describing gravity in $(d+1)$ dimensions is equivalent, upon keeping only the $n=0$ Fourier modes, to the action of d-dimensional general relativity together with the Maxwell action and the action for a scalar field. In a little more detail, suppose that $\mathcal{Y}=S^1$ and we use coordinates X^M, $M=0,\ldots,d+1$ on $\mathcal{X}\times S^1$ and coordinates x^μ with $\mu=0,\ldots,d$ on $\mathcal{X}$. So that $X^M=(x^\mu,\theta)$ where θ is an angular coordinate around S^1. Then we consider the metric:

$$ds^2 = g_{MN}dX^M dX^N = g_{\mu\nu}(x)dx^\mu dx^\nu + \Omega^2(x)(d\theta + A_\mu(x)dx^\mu)^2 \tag{14.52}$$

where the metric $g_{\mu\nu}$, the "warp factor" Ω^2 and the "gauge connection" A_μ are only functions of x^μ (that is, we make the restriction to zero Fourier modes).

Note that this means the metric tensor looks like

$$g_{MN}(x^\mu,\theta) = \begin{pmatrix} g_{\mu\nu}(x)+\Omega^2(x)A_\mu(x)A_\nu(x) & \Omega^2(x)A_\mu(x) \\ \Omega^2(x)A_\nu(x) & \Omega^2(x) \end{pmatrix} \tag{14.53}$$

The general symmetric $(d+1)\times(d+1)$ matrix has

$$\frac{1}{2}(d+1)(d+2) = \frac{1}{2}d(d+1)+d+1 \tag{14.54}$$

independent components, so we have not lost any generality in the form of the matrix, but writing it this way makes the connection to physical quantities (and to connections on a principal $U(1)$ bundle) clearer. By writing the fields on the RHS as functions of x^μ and not (x^μ,θ) we have made a severe restriction - limiting attention to the massless modes in d dimensions, as explained above.

Under these conditions one computes that the Riemann scalar for the $(d+1)$-dimensional metric is:

$$\mathcal{R}[g_{MN}] = \mathcal{R}[g_{\mu\nu}] - \frac{\Omega^2}{4}F_{\mu\nu}F^{\mu\nu} - 2(\nabla\log\Omega)^2 - 2\nabla^2\log\Omega \tag{14.55}$$

so the Einstein-Hilbert action for GR in $(d+1)$ dimensions reduces to that of Einstein-Hilbert-Maxwell-Scalar in d dimensions. This is a truly remarkable equation: Pure gravity in $(d+1)$ dimensions leads to both gravity and electricity and magnetism in d dimensions!

Remarks:

1. The KK ansatz also leads to a scalar field $\log\Omega^2(x)$, known as the "dilaton" because it can dilate, in a space-time dependent way, the size of the "internal dimensions" $\mathcal{Y}$. Note that in electricity and magnetism the coupling constant enters via

$$S_{\text{Maxwell}} = \int \sqrt{g}\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} \tag{14.56}$$

so the presence of the dilaton can lead to space-time variation of the fine structure constant. In physically viable models one must explain why the dilaton does not

fluctuate wildly. The discovery of the naturally occurring nuclear reactor in Oklo, Africa, has led to the bound

$$|\frac{\dot{\alpha}}{\alpha}| < few \times 10^{-17} yr^{-1} \tag{14.57}$$

2. By considering internal spaces $\mathcal{Y}$ equipped with metric with isometry group H similar considerations lead gauge theory with gauge group H in $\mathcal{X}$.

It is interesting to understand how gauge symmetries in theories on $\mathcal{X}$ arise in this point of view. Suppose $\mathcal{D} \cong Diff(\mathcal{X})$ is a subgroup of diffeomorphisms of $\mathcal{X} \times \mathcal{Y}$ of the form

$$\psi_f : (x, y) \to (f(x), y) \qquad\qquad f \in Diff(\mathcal{X}). \tag{14.58}$$

We also consider a subgroup $\mathcal{G}$ of $Diff(\mathcal{X} \times \mathcal{Y})$ where $\mathcal{G}$ is isomorphic to a subgroup of $Map(\mathcal{X}, Diff(\mathcal{Y}))$. For the moment just take $\mathcal{G} = Map(\mathcal{X}, Diff(\mathcal{Y}))$, so an element $g \in \mathcal{G}$ is a family of diffeomorphisms of $\mathcal{Y}$ parametrized by $\mathcal{X}$: For each x we have a diffeomorphism of $\mathcal{Y}$: $g_x : y \to g(y; x)$. Then we take $\mathcal{G}$ to be the subgroup of diffeomorphisms of $Diff(\mathcal{X} \times \mathcal{Y})$ of the form

$$\psi_g : (x, y) \to (x, g(y; x)) \qquad\qquad g \in \mathcal{G} \tag{14.59}$$

Note that within $Diff(\mathcal{X} \times \mathcal{Y})$ we can write the subgroup

$$\mathcal{G}\mathcal{D} \tag{14.60}$$

and $\mathcal{D}$ acts as a group of automorphisms of $\mathcal{G}$ via

$$\begin{aligned} \psi_f \psi_g \psi_f^{-1} : (x, y) &\to (f^{-1}(x), y) \\ &\to (f^{-1}(x), g(y; f^{-1}(x))) \\ &\to (x, g(y; f^{-1}(x))) \end{aligned} \tag{14.61}$$

so if $g \in \mathcal{G}$ and $f \in \mathcal{D}$ then $\psi_f \psi_g \psi_f^{-1} = \psi_{g'}$ with $g' \in \mathcal{G}$ and hence $\mathcal{G}\mathcal{D}$ is a semidirect product. In fact, it is an example of the generalized wreath product of the previous example.

This is a model for the group of gauge transformations in Kaluza-Klein theory. So $\mathcal{X}$ is the "large", possibly noncompact, spacetime where we have general relativity, while $\mathcal{Y}$ is the "small," possibly compact space giving rise to gauge symmetry. $\mathcal{D}$ is the diffeomorphism group of the large spacetime and is the gauge symmetry of general relativity on $\mathcal{X}$. Typically, $\mathcal{Y}$ is endowed with a fixed metric $ds^2_{\mathcal{Y}}$ and the diffeomorphism symmetry of $\mathcal{Y}$ is (spontaneously) broken down to the group of isometries of $\mathcal{Y}$, $Isom(\mathcal{Y}, ds^2_{\mathcal{Y}})$. So in the above construction we take $\mathcal{G}$ to be the unbroken subgroup $Map(\mathcal{X}, Isom(\mathcal{Y}, ds^2_{\mathcal{Y}})) \subset Map(\mathcal{X}, Diff(\mathcal{Y}))$. This subgroup $Map(\mathcal{X}, Isom(\mathcal{Y}, ds^2_{\mathcal{Y}}))$ is interpreted as a group of gauge transformations of a gauge theory on $\mathcal{X}$ coupled to general relativity on $\mathcal{X}$.

As a simple example of the remarks of the previous paragraph let us suppose that $\mathcal{Y} = S^1$ with coordinate θ and round metric $(d\theta)^2$. The isometries of the circle are just constant translations, $\theta \to \theta + \epsilon$. So if

$$g \in \mathrm{Map}(\mathcal{X}, Isom(S^1)) \tag{14.62}$$

then $g(x)$ will take $\theta \to \theta + \epsilon(x)$, so

$$\psi_g : (x^\mu, \theta) \to (x^\mu, \theta + \epsilon(x)) \tag{14.63}$$

so the metric in (14.52) transforms as

$$\psi_g^*(ds^2) = g_{\mu\nu}(x)dx^\mu dx^\nu + \Omega^2(x)(d\theta + dx^\mu \partial_\mu \epsilon + A_\mu(x)dx^\mu)^2 \tag{14.64}$$

meaning that the fields $g_{\mu\nu}$ and Ω are invariant, but

$$A_\mu \to A_\mu + \partial_\mu \epsilon \tag{14.65}$$

and thus, these special diffeomorphisms appear as gauge transformations of the Maxwell field!

Exercise

Show that $S_3 \cong \mathbb{Z}_3 \rtimes \mathbb{Z}_2$ where the generator of $\mathbb{Z}_2$ acts as the nontrivial outer automorphism of $\mathbb{Z}_3$.

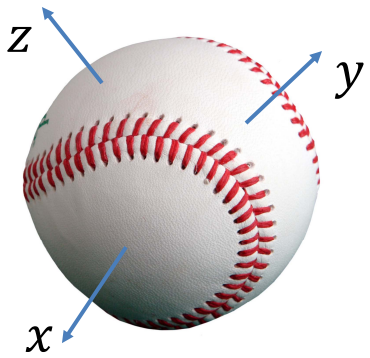


Figure 36: A baseball.

Exercise *Symmetries Of The Baseball*

Ignoring any writing, but taking into account the seams, find the symmetry group of a baseball. (See figure 36.) [152]

---

[152] *Answer* D_4.

Exercise *Symmetries Of The Cube*

a.) Consider a perfect cube. By considering the action of proper rotations on the four diagonal axes through vertices show that the symmetry group is isomorphic to S_4.

b.) Centering the cube on the origin with vertices $(\pm 1, \pm 1, \pm 1)$ show that the symmetry group is $(\mathbb{Z}_2 \times \mathbb{Z}_2) \rtimes S_3$. Deduce that

$$S_4 \cong (\mathbb{Z}_2 \times \mathbb{Z}_2) \rtimes S_3 \tag{14.66}$$

Exercise *Centralizers in the symmetric group*

a.) Suppose that $g \in S_n$ has a conjugacy class given by $\prod_{i=1}^{n} (i)^{\ell_i}$. Show that the centralizer $Z(g)$ is isomorphic to

$$Z(g) \cong \prod_{i=1}^{n} \left(\mathbb{Z}_i^{\ell_i} \rtimes S_{\ell_i} \right) \tag{14.67}$$

where $\prod_i$ is a direct product.

b.) Use this to compute the order of a conjugacy class in the symmetric group.

Exercise *The Lorentz Group As A Semidirect Product*

Let $\eta = Diag\{-1, \mathbf{1}_d\}$. The Lorentz group in $d+1$ dimensions, denoted $O(1,d)$ is the matrix group

$$O(1,d) = \{A | A\eta A^{tr} = \eta\} \tag{14.68}$$

a.) Considering the case $d = 1$ find the general solution and show that there are four connected components. [153]

b.) Show that, group-theoretically, we have

$$O(1,1) = SO_0(1,1) \rtimes (\mathbb{Z}_2 \times \mathbb{Z}_2) \tag{14.70}$$

where $SO_0(1,1)$ is the connected component of the identity.

Remark: In fact, more generally $O(1,d)$ has four connected components for $d \geq 1$ and

$$O(1,d) = SO_0(1,d) \rtimes (\mathbb{Z}_2 \times \mathbb{Z}_2) \tag{14.71}$$

[153] *Answer*: Writing out the four matrix elements of the defining equation easily arrives at the general solution:

$$A = \begin{pmatrix} \xi_1 \cosh\theta & \xi_2 \sinh\theta \\ \xi_4 \sinh\theta & \xi_3 \cosh\theta \end{pmatrix} \tag{14.69}$$

where $\theta \in \mathbb{R}$, $\xi_i \in \{\pm 1\}$ and $\xi_4 = \xi_1\xi_2\xi_3$. By changing the sign of θ we can set $\xi_{=}2$. The sign of ξ_1, ξ_2 is meaningful, giving us four components.

You can easily see that there are at least four components since $\det A \in \{\pm 1\}$ and moreover $A_{00}^2 = 1 + \sum_{i=1}^{d} A_{0i}^2$ so that we can independently have $A_{00} \geq 1$ or $A_{00} \leq -1$.

Exercise *Holomorph*

Given a finite group G a canonical semidirect product group is $G \rtimes \mathrm{Aut}(G)$ known as the holomorph of G.

a.) Show that this is the normalizer of the copy of G in the symmetric group $S_{|G|}$ given by Cayley's theorem.

b.) Show that the affine Euclidean group $Euc(d)$ is the holomorph of the Abelian group $\mathbb{R}^d$.

Exercise *Equivalence of semidirect products*

A nontrivial homomorphism $\alpha : G \to \mathrm{Aut}(H)$ can lead to a semidirect product which is in fact isomorphic to a direct product. Show this as follows: Suppose $\phi : G \to H$ is a homomorphism. Define $\alpha : G \to \mathrm{Aut}(H)$ by $\alpha_g = I(\phi(g))$. Construct an isomorphism [154]

$$\Psi : H \rtimes_\alpha G \to H \times G \tag{14.72}$$

Exercise *Manipulating the Seitz notation*

a.) Show that:

$$\begin{aligned}
\{v|R\}^{-1} &= \{-R^{-1}v|R^{-1}\} \\
\{0|R\}\{v|1\} &= \{Rv|R\} \\
\{v|1\}\{0|R\} &= \{v|R\} \\
\{w|1\}\{v|R\} &= \{w+v|R\} \\
\{v_1|R_1\}\{v_2|R_2\}\{v_1|R_1\}^{-1} &= \{R_1v_2 + (1 - R_1R_2R_1^{-1})v_1|R_1R_2R_1^{-1}\} \\
[\{v_1|R_1\},\{v_2|R_2\}] &= \{(1 - R_1R_2R_1^{-1})v_1 - R_1R_2R_1^{-1}R_2^{-1}(1 - R_2R_1R_2^{-1})v_2|R_1R_2R_1^{-1}R_2^{-1}\}
\end{aligned} \tag{14.73}$$

b.) Show that the subgroup of pure translations, that is, the subgroup of elements of the form $\{v|1\}$ with $v \in \mathbb{R}^d$ is a normal subgroup of $\mathrm{Euc}(d)$.

[154] *Answer*: $\Psi(h,g) = (h\phi(g), g)$.

c.) Can you construct a homomorphism $O(d) \to \mathrm{Euc}(d)$?

d.) Consider the group $G = L \rtimes \mathbb{Z}_2$ where L is a lattice and the nontrivial element in $\mathbb{Z}_2$ acts on L by $v \to -v$. Compute the conjugacy classes in G. [155]

Exercise *Dependence on basepoint for isomorphism of Euclidean group*
Show that if $p' = p + v_0$ then

$$\Psi_{p'}(v', R) = \Psi_p(v, R) \in \mathrm{Euc}(d) \tag{14.74}$$

for

$$v' = v + (1 - R)v_0 \tag{14.75}$$

---

15. Group Extensions and Group Cohomology

15.1 Group Extensions

♣Add: Pushforward extensions ♣

Recall that an extension of Q by a group N is an exact sequence of the form:

$$1 \to N \quad \stackrel{\iota}{\to} \quad G \quad \stackrel{\pi}{\to} \quad Q \to 1 \tag{15.1}$$

There is a notion of *homomorphism of two group extensions*

$$1 \to N \quad \stackrel{\iota_1}{\to} \quad G_1 \quad \stackrel{\pi_1}{\to} \quad Q \to 1 \tag{15.2}$$

$$1 \to N \quad \stackrel{\iota_2}{\to} \quad G_2 \quad \stackrel{\pi_2}{\to} \quad Q \to 1 \tag{15.3}$$

This means that there is a group homomorphism $\varphi : G_1 \to G_2$ so that the following diagram commutes:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & N & \xrightarrow{\iota_1} & G_1 & \xrightarrow{\pi_1} & Q & \longrightarrow & 1 \\ & & \uparrow{\scriptstyle \mathrm{Id}} & & \downarrow{\scriptstyle \varphi} & & \uparrow{\scriptstyle \mathrm{Id}} & & \\ 1 & \longrightarrow & N & \xrightarrow{\iota_2} & G_2 & \xrightarrow{\pi_2} & Q & \longrightarrow & 1 \end{array} \tag{15.4}$$

To say that a "diagram commutes" means that if one follows the maps around two paths with the same beginning and ending points then the compositions of the maps are the same. Thus (15.4) is completely equivalent to the pair of equations:

$$\begin{aligned} \pi_1 &= \pi_2 \circ \varphi \\ \iota_2 &= \varphi \circ \iota_1 \end{aligned} \tag{15.5}$$

[155] *Answer*: $C(\{v|1\}) = \{\{\pm v|1\}\}$ has two elements while $C(\{v|-1\}) = \{\{\pm v + 2v_1|-1\}|v_1 \in L\}$ has infinitely many elements.

However, drawing a diagram makes the relations between maps, domains and codomains much more transparent. Sometimes a picture is worth a thousand equations. This is why mathematicians like commutative diagrams.

When there is a homomorphism of group extensions based on $\psi : G_2 \to G_1$ such that $\varphi \circ \psi$ and $\psi \circ \varphi$ are the identity then the group extensions are said to be isomorphic.

It can certainly happen that there is more than one nonisomorphic extension of Q by N. Classifying all extensions of Q by N is a difficult problem. We will discuss it more in section 15.7 below.

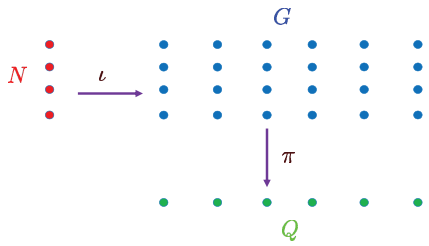


Figure 37: Illustration of a group extension $1 \to N \to G \to Q \to 1$ as an N-bundle over Q.

We would encourage the reader to think geometrically about this problem, even in the case when Q and N are finite groups, as in Figure 37. In particular we will use the important notion of a *section*, that is, a right-inverse to π: It is a map $s : Q \to G$ such that $\pi(s(q)) = q$ for all $q \in Q$. Such sections always exist.[156] Note that in general $s(\pi(g)) \neq g$. This is obvious from Figure 37. The set of pre-images, $\pi^{-1}(q)$, is called *the fiber of* π *over* q. The map π projects the entire fiber over q to the single element q. A choice of section s is a choice, for each and every $q \in Q$, of just one single point in the fiber above q.

In order to justify the picture of Figure 37 let us prove that, as a set, G is just the product $N \times Q$. Note that for any $g \in G$ and any section s:

$$g(s(\pi(g)))^{-1} \tag{15.6}$$

maps to 1 under π (check this). Therefore, since the sequence is exact

$$g(s(\pi(g)))^{-1} = \iota(n) \tag{15.7}$$

for some $n \in N$. That is, every $g \in G$ can be written as

$$g = \iota(n)s(q) \tag{15.8}$$

for some $n \in N$ and some $q \in Q$. In fact, this decomposition is *unique*: Suppose that:

$$\iota(n_1)s(q_1) = \iota(n_2)s(q_2) \tag{15.9}$$

[156] By the axiom of choice. For continuous groups such as Lie groups there might or might not be continuous sections.

Then we rewrite this as

$$\iota(n_2^{-1}n_1) = s(q_2)s(q_1)^{-1} \tag{15.10}$$

Now, applying π we learn that $1 = q_2\pi(s(q_1)^{-1}) = q_2\left(\pi(s(q_1))\right)^{-1} = q_2q_1^{-1}$, so $q_1 = q_2$. But that implies $n_1 = n_2$. Therefore, as a set, G can be identified with $N \times Q$.

Remark: As a nice corollary of the decomposition (15.8) note that if φ defines a morphism of group extensions then φ is in fact an isomorphism of G_1 to G_2. It is a homomorphism by definition. Now note that if $s_1 : Q \to G_1$ is a section of π_1 then $s_2 := \varphi \circ s_1 : Q \to G_2$ is a section of π_2 so

$$\begin{aligned} \varphi(g) &= \varphi(\iota_1(n)s_1(q)) \\ &= \varphi(\iota_1(n))\varphi(s_1(q)) \\ &= \iota_2(n)s_2(q) \end{aligned} \tag{15.11}$$

and since the decomposition is unique (given a choice of section) the map φ is $1-1$.

Now, given an extension and a choice of section s we define a map

$$\omega : Q \to \mathrm{Aut}(N) \tag{15.12}$$

denoted by

$$q \mapsto \omega_q \tag{15.13}$$

where the definition of ω_q is given by

$$\iota(\omega_q(n)) = s(q)\iota(n)s(q)^{-1} \tag{15.14}$$

Because $\iota(N)$ is normal the RHS is again in $\iota(N)$. Because ι is injective $\omega_q(n)$ is well-defined. Moreover, for each q the reader should check that indeed $\omega_q(n_1n_2) = \omega_q(n_1)\omega_q(n_2)$, and ω_q is one-one on N. Therefore we really have a map of sets (15.12). Note carefully that we are not saying that $q \mapsto \omega_q$ is a group homomorphism. In general, it is not.

Remark: Clearly the ι is a bit of a nuisance and leads to clutter and it can be safely dropped if we consider N simply to be a subgroup of G, for then ι is simply the inclusion map. The confident reader is encouraged to do this. The formulae will be a little cleaner. However, we will be pedantic and retain the ι in most of our formulae.

Let us stress that the map $\omega : Q \to \mathrm{Aut(N)}$ *in general is not a homomorphism* and *in general depends on the choice of section* s. We will discuss the dependence on the choice of section s below when we have some more machinery and context. For now let us see how close ω comes to being a group homomorphism:

$$\begin{aligned} \iota\left(\omega_{q_1} \circ \omega_{q_2}(n)\right) &= s(q_1)\iota(\omega_{q_2}(n))s(q_1)^{-1} \\ &= s(q_1)s(q_2)\iota(n)(s(q_1)s(q_2))^{-1} \end{aligned} \tag{15.15}$$

We want to compare this to $\iota(\omega_{q_1q_2}(n))$. In general they will be different unless $s(q_1q_2) = s(q_1)s(q_2)$, that is, unless $s : Q \to G$ is a homomorphism. In general the section is not a homomorphism, but clearly something nice happens when it is:

Definition: We say an extension *splits* if there exists a section $s : Q \to G$ which is *also* a *group homomorphism*. A choice of a section which is a group homomorphism is called a (choice of) *splitting*.

Theorem: An extension is isomorphic to a semidirect product iff it is a split extension.

Proof:

First suppose it splits. Choose a splitting s. Then from (15.15) we know that

$$\omega_{q_1} \circ \omega_{q_2} = \omega_{q_1q_2} \tag{15.16}$$

and hence $q \mapsto \omega_q$ defines a homomorphism $\omega : Q \to \mathrm{Aut}(N)$. Therefore, we can aim to prove that there is an isomorphism of G with $N \rtimes_\omega Q$.

In general if s is just a section the image $s(Q) \subset G$ is not a subgroup. But if the sequence splits, then it is a subgroup. The equation (15.8) implies that $G = \iota(N)s(Q)$ where $s(Q)$ is a subgroup, and by the internal characterization of semidirect products that means we have a semidirect product.

To give a more concrete proof, let us write the group law in the parametrization (15.8). Write

$$\iota(n)s(q)\iota(n')s(q') = \iota(n)\left(s(q)\iota(n')s(q)^{-1}\right)s(qq') \tag{15.17}$$

Note that

$$s(q)\iota(n')s(q)^{-1} = \iota(\omega_q(n')) \tag{15.18}$$

so

$$\iota(n_1)s(q_1)\iota(n_2)s(q_2) = \iota\left(n_1\omega_{q_1}(n_2)\right)s(q_1q_2) \tag{15.19}$$

But this just means that

$$\Psi(n,q) = \iota(n)s(q) \tag{15.20}$$

is in fact an isomorphism $\Psi : N \rtimes_\omega Q \to G$. Indeed equation (15.19) just says that:

$$\Psi(n_1,q_1)\Psi(n_2,q_2) = \Psi((n_1,q_1)\cdot_\omega(n_2,q_2)) \tag{15.21}$$

where $\cdot_\omega$ stresses that we are multiplying with the semidirect product rule.

Thus, we have shown that a split extension is isomorphic to a semidirect product $G \cong N \rtimes Q$. The converse is straightforward. ♠

In §15.7 below we will continue the general line of reasoning begun here. However, in order to appreciate the formulae better it is a good idea first to step back and consider a simple but important special case of extensions, namely, the *central extensions*. These are extensions such that $\iota(N)$ is a subgroup of the center of G. Here is the official definition: (Note the change of notation from the general situation above):

♣Do the general case first and then specialize? ♣

Let A be an abelian group and G any group.

Definition A *central extension* of G by A, [157] is a group $\tilde{G}$ and an extension such that

$$1 \to A \quad \stackrel{\iota}{\to} \quad \tilde{G} \quad \stackrel{\pi}{\to} \quad G \to 1 \tag{15.22}$$

such that $\iota(A) \subset Z(\tilde{G})$.

We stress again that what we called G in the previous discussion is here called $\tilde{G}$, and what we called Q in the previous discussion is here called G.

Example And Remark:*Sections of group extensions vs. continuous sections of principal fiber bundles.* Let us return to the very important exact sequence (10.38):

$$1 \to \mathbb{Z}_2 \quad \stackrel{\iota}{\to} \quad SU(2) \quad \stackrel{\pi}{\to} \quad SO(3) \to 1 \tag{15.23}$$

The $\mathbb{Z}_2$ is embedded as the subgroup $\{\pm 1\} \subset SU(2)$, so this is a central extension. We said above that there is always a section, but when we said that we did not impose any properties of continuity in the case where G and Q are continuous groups. In this example while there is a section of π there is, in fact, no continuous section. Such a continuous section $\pi s = Id$ would imply that $\pi_* s_* = 1$ on the first homotopy group of $SO(3)$. But that is impossible since it would have factor through $\pi_1(SU(2)) = 1$.

We are using a few facts here:

1. Every $SU(2)$ matrix can be written as

$$\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \tag{15.24}$$

   where α, β are complex numbers with $|\alpha|^2 + |\beta|^2 = 1$. Writing this equation in terms of the real and imaginary parts of α, β we recognize the equation of the unit three dimensional sphere. Now recall that all the spheres of dimension ≥ 2 are simply connected. Therefore $\pi_1(SU(2)) = 1$ is simply connected.

2. But $SO(3)$ is not simply connected! In fact, using a coffee cup you can informally demonstrate that $\pi_1(SO(3)) \cong \mathbb{Z}_2$. [Demonstrate].

3. If there were a continuous section then $s_* : \pi_1(SO(3)) \to \pi_1(SU(2))$ would be a well-defined group homomorphism and $s \circ \pi = Id$ would imply that on the fundamental groups $Id_* = s_* \pi_*$ in

$$\pi_1(SO(3)) \to \pi_1(SU(2)) = 1 \to \pi_1(SO(3)) \tag{15.25}$$

   But Id_* takes the nontrivial element of $\mathbb{Z}_2$ to the nontrivial element, not to the trivial element. This is impossible if you factor through the trivial group.

In algebraic topology one introduces another kind of topological invariant known as homology. The homology groups of a manifold are Abelian groups that encode many important properties of the manifold. The homology group $H_1(M; \mathbb{Z}_2)$ tells us what are

[157] Some authors say an extension of A by G.

the possible 2-fold covers of the manifold M. It turns out that $H_1(SO(3);\mathbb{Z}_2) \cong \mathbb{Z}_2$, (this is closely related to $\pi_1(SO(3)) \cong \mathbb{Z}_2$. so there are two double covers of $SO(3)$. One is $O(3) = \mathbb{Z}_2 \times SO(3)$ and the other is $SU(2)$, the nontrivial double cover.

The extension (15.1) generalizes to

$$1 \to \mathbb{Z}_2 \quad \stackrel{\iota}{\to} \quad \mathrm{Spin}(d) \quad \stackrel{\pi}{\to} \quad SO(d) \to 1 \tag{15.26}$$

as well as the two Pin groups which extend $O(d)$:

$$1 \to \mathbb{Z}_2 \quad \stackrel{\iota}{\to} \quad \mathrm{Pin}^{\pm}(d) \quad \stackrel{\pi}{\to} \quad O(d) \to 1 \tag{15.27}$$

we discuss these in Section *** below. Again, in these cases there is no continuous section. Thus, these examples are nontrivial as fiber bundles. Moreover, even if we allow ourselves to choose a discontinuous section, we cannot do so and make it a group homomorphism. In other words these examples are also nontrivial as group extensions.

Exercise

If $s : Q \to G$ is any section of π show that for all $q \in Q$,

$$s(q^{-1}) = s(q)^{-1}n = n's(q)^{-1} \tag{15.28}$$

for some $n, n' \in N$.

Exercise *The pullback construction*

There is one general construction with extensions which is useful when discussing symmetries in quantum mechanics. This is the notion of *pullback extension*. Suppose we are given both an extension

$$1 \longrightarrow H' \xrightarrow{\iota} H \xrightarrow{\pi} H'' \longrightarrow 1 \tag{15.29}$$

and a homomorphism

$$\rho : G \to H'' \tag{15.30}$$

one can define another extension of G by H' known as a *pullback extension*. We are trying to fill in the diagram:

$$\begin{array}{ccccccccc} & & & & & & G & & \\ & & & & & & \downarrow{\scriptstyle \rho} & & \\ 1 & \longrightarrow & H' & \xrightarrow{\iota} & H & \xrightarrow{\pi} & H'' & \longrightarrow & 1 \end{array} \tag{15.31}$$

with an extension on the first row of G by H'.

We do this by defining a subgroup of the Cartesian product $\tilde{G} \subset H \times G$:

$$\tilde{G} := \{(h, g) | \pi(h) = \rho(g)\} \subset H \times G \tag{15.32}$$

We have an extension of the form

$$1 \longrightarrow H' \xrightarrow{\iota} \tilde{G} \xrightarrow{\tilde{\pi}} G \longrightarrow 1 \tag{15.33}$$

where $\tilde{\pi}(h, g) := g$. Show that this extension fits in the commutative diagram

$$\begin{array}{ccccccccc} 1 & \longrightarrow & H' & \longrightarrow & \tilde{G} & \xrightarrow{\tilde{\pi}} & G'' & \longrightarrow & 1 \\ & & \| & & \downarrow{\scriptstyle \tilde{\rho}} & & \downarrow{\scriptstyle \rho} & & \\ 1 & \longrightarrow & H' & \longrightarrow & H & \xrightarrow{\pi} & H'' & \longrightarrow & 1 \end{array} \tag{15.34}$$

(N.B. This is <u>not</u> a morphism of extensions.)

Exercise *The pushforward extension*

Under some circumstances one can complete the diagram

$$\begin{array}{ccccccccc} 1 & \longrightarrow & H' & \xrightarrow{\iota} & H & \xrightarrow{\pi} & H'' & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle \rho} & & & & & & \\ & & \tilde{H} & & & & & & \end{array} \tag{15.35}$$

to get an extension of H'' by $\tilde{H}$. This is not as universal as the pullback. But one can construct it if $\rho : H' \to \tilde{H}$ is surjective and $\iota(\ker(\rho)) \triangleleft H$. Give the construction. [158]

Exercise *Choice of splitting and the Euclidean group* $\mathrm{Euc}(d)$

As we noted, the Euclidean group $\mathrm{Euc}(d)$ is isomorphic to the semidirect product $\mathbb{R}^d \rtimes O(d)$, but to exhibit that we needed to choose an origin about which to define rotation-inversions. See equation (14.35) above.

a.) Show that a change of origin corresponds to a change of splitting.

b.) Using the Seitz notation show that another choice of origin leads to the splitting $R \mapsto \{(1 - R)v|R\}$, and verify that this is also a splitting.

♣This is almost redundant with another exercise above. ♣

[158] *Answer*: Define the group $\tilde{G} := H/\iota(\ker(\rho))$. Then note that we can define $\tilde{\iota} : \tilde{H} \to \tilde{G}$ via $\tilde{\iota}(\rho(h)) = \iota(h) + \iota(\ker(\rho))$. Note that if $\rho(h) = \rho(h')$ then the RHS is the same so this does give a well-defined map $\tilde{\iota}$ on the image of ρ, but if ρ is surjective that is enough. Now define $\tilde{\pi}(\tilde{e}) = \pi(e)$ where $\tilde{e} = e + \iota(\ker\rho)$. Then we have a commutative diagram:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & H' & \xrightarrow{\iota} & H & \xrightarrow{\pi} & H'' & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle \rho} & & \downarrow{\scriptstyle \varphi} & & \updownarrow{\scriptstyle Id} & & \\ 1 & \longrightarrow & \tilde{H} & \xrightarrow{\tilde{\iota}} & \tilde{G} & \xrightarrow{\tilde{\pi}} & H'' & \longrightarrow & 1 \end{array} \tag{15.36}$$

where $\varphi(h) = h + \iota(\ker(\rho))$.

Exercise *Another form of splitting*

Show that an equivalent definition of a split exact sequence for a central extension is that there is a homomorphism $t : \tilde{G} \to A$ which is a left-inverse to ι, $t(\iota(a)) = a$.

(Hint: Define $s(\pi(\tilde{g})) = \iota t(\tilde{g}^{-1}))\tilde{g}$.)

Exercise *The exact sequence for a product of two cyclic groups*

Revisit the exact sequence discussed in equation (12.52). Show that this sequence splits. [159]

Exercise *Is A Restriction Of A Split Extension Split?*

Suppose

$$1 \longrightarrow N \xrightarrow{\iota} G \xrightarrow{\pi} Q \longrightarrow 1 \tag{15.37}$$

is a split extension and $N_1 \subset N$ and $Q_1 \subset Q$, so that there is an extension

$$1 \longrightarrow N_1 \xrightarrow{\iota} G_1 \xrightarrow{\pi} Q_1 \longrightarrow 1 \tag{15.38}$$

given by restriction to $G_1 \subset G$. Does it follow that this extension is a split extension?

Exercise *A Split Central Extension Is A Direct Product*

Suppose

$$1 \longrightarrow N \xrightarrow{\iota} G \xrightarrow{\pi} Q \longrightarrow 1 \tag{15.39}$$

is a split central extension. Show that $G \cong N \times Q$ and the extension is isomorphic to the trivial extension with ι inclusion into the first factor and π projection onto the second factor. [160]

[159] *Answer*: Let ω_ℓ generate $\mathbb{Z}_\ell$ then, using the notation above let $s : \omega_\ell \mapsto (\omega_\ell^{\mu_2\nu_2}, \omega_\ell^{\mu_1\nu_1})$.

[160] *Answer*: Choose a splitting s. Then use the parametrization $g = \iota(n)s(q)$. The group multiplication can be written

$$\begin{aligned} g_1 g_2 &= \iota(n_1)s(q_1)\iota(n_2)s(q_2) \\ &= \iota(n_1)\iota(n_2)s(q_1)s(q_2) \\ &= \iota(n_1 n_2)s(q_1 q_2) \end{aligned} \tag{15.40}$$

In the going from the first to second line we used that $\iota(n_2)$ is in the center. In going from the second to the third line we used that ι and s are group homomorphisms. (Recall s is a group homomorphism because it is a splitting.)

15.2 Projective Representations

We have already encountered the notion of a matrix representation of a group G. This is a homomorphism from G into $GL(d,\kappa)$ for some field κ. In many contexts in mathematics and physics (especially in quantum physics) one encounters a generalization of the notion of matrix representation known as a *projective representation*. The theory of projective representations is closely related to the theory of central extensions.

Recall that a matrix representation of a group G is a group homomorphism

$$\rho : G \to GL(d,\kappa) \tag{15.41}$$

A *projective representation* is a map

$$\rho : G \to GL(d,\kappa) \tag{15.42}$$

which is "almost a homomorphism" in the sense that

$$\rho(g_1)\rho(g_2) = f(g_1,g_2)\rho(g_1,g_2) \tag{15.43}$$

for some function $f : G \times G \to \kappa^*$. Of course $f(g_1,g_2)$ is "just a c-number" so you might think it is an unimportant nuisance. You might try to get rid of it by redefining

$$\tilde{\rho}(g) = b(g)\rho(g) \tag{15.44}$$

where $b(g) \in \kappa^*$ is a c-number. Then if there exists a function $b : G \to \kappa^*$ such that

$$f(g_1,g_2) \stackrel{?}{=} \frac{b(g_1g_2)}{b(g_1)b(g_2)} \tag{15.45}$$

then $\tilde{\rho}$ would be an honest representation.

The trouble is, in many important contexts, there is no function b so that (15.45) holds. So we need to deal with it.

A simple example is the "spin representation of the rotation group $SO(3)$" where one attempts to define a map:

$$\rho : SO(3) \to SU(2) \subset GL(2,\mathbb{C}) \tag{15.46}$$

that attempts to describe the effects of a rotation on - say - a spin 1/2 particle. In fact, there is no such thing as the "spin representation of the rotation group $SO(3)$." There is a spin projective representation of $SO(3)$. [161]

We saw above that there is a very natural group homomorphism from $SU(2)$ to $SO(3)$, but there is no group homomorphism back from $SO(3)$ to $SU(2)$: There is no splitting. The so-called "spin representation of $SO(3)$" is usually presented by attempting to construct a splitting $\rho : SO(3) \to SU(2)$ using Euler angles. Indeed, under the standard

[161] However, there is a perfectly well-defined spin representation of the Lie algebra $\mathfrak{so}(3)$.

homomorphism $\pi : SU(2) \to SO(3)$ one recognizes that $\exp[\mathrm{i}\theta\sigma^i]$ maps to a rotation by angle 2θ around the the ith axis. For example,

$$u = \exp[\mathrm{i}\theta\sigma^3] = \begin{pmatrix} e^{\mathrm{i}\theta} & 0 \\ 0 & e^{-\mathrm{i}\theta} \end{pmatrix} = \cos\theta + \mathrm{i}\sin\theta\sigma^3 \tag{15.47}$$

acts by

$$\begin{aligned} u\vec{x}\cdot\vec{\sigma}u^{-1} &= u\begin{pmatrix} z & x-\mathrm{i}y \\ x+\mathrm{i}y & -z \end{pmatrix}u^{-1} \\ &= \begin{pmatrix} z & e^{2\mathrm{i}\theta}(x-\mathrm{i}y) \\ e^{-2\mathrm{i}\theta}(x+\mathrm{i}y) & -z \end{pmatrix} \end{aligned} \tag{15.48}$$

One can represent any rotation in $SO(3)$ by a rotation around the z-axis, then around the x-axis, then around the z axis. Call this $R(\phi,\theta,\psi)$. So one attempts to define ρ by assigning

$$\rho : R(\phi,\theta,\psi) \to e^{\mathrm{i}\frac{\phi}{2}\sigma^3}e^{\mathrm{i}\frac{\theta}{2}\sigma^1}e^{\mathrm{i}\frac{\psi}{2}\sigma^3}. \tag{15.49}$$

Clearly, we are going to have problems making this mapping well-defined. For example, $R(\phi,0,0)$ would map to $e^{\mathrm{i}\frac{\phi}{2}\sigma^3}$, but this is not well-defined for all ϕ because $R(2\pi,0,0) = 1$ and $e^{\mathrm{i}\frac{2\pi}{2}\sigma^3} = -1$. The problem is that the Euler angle coordinates on $SO(3)$ are sometimes singular. So, we need to restrict the domain of ϕ,θ,ψ so that (15.49) is well-defined for every $R \in SO(3)$. However, when we make such a restriction we will spoil the group homomorphism property, but only up to a phase. So, in this way, we get a two-dimensional projective representation of $SO(3)$.

As an exercise you can try the following: Every $SU(2)$ matrix can be written as $u = \cos(\chi) + \mathrm{i}\sin(\chi)\hat{n}\cdot\vec{\sigma}$ and this maps under π to a rotation by 2χ around the $\hat{n}$ axis. But again, you cannot smoothly identify every $SO(3)$ rotation by describing it as a rotation by 2χ around an axis.

Exercise *Projective representations of $\mathbb{Z}_2 \times \mathbb{Z}_2$ and the Pauli group*
Consider the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ multiplicatively:

$$\mathbb{Z}_2 \times \mathbb{Z}_2 = \{1, g_1, g_2, g_1g_2\} \tag{15.50}$$

with relations $g_1^2 = g_2^2 = 1$ and $g_1g_2 = g_2g_1$. Consider the map

$$\rho : \mathbb{Z}_2 \times \mathbb{Z}_2 \to GL(2,\mathbb{C}) \tag{15.51}$$

defined by

$$\begin{aligned} \rho(1) &= 1 \\ \rho(g_1) &= \sigma^1 \\ \rho(g_2) &= \sigma^2 \\ \rho(g_1g_2) &= \sigma^3 \end{aligned} \tag{15.52}$$

a.) Show that this defines a projective representation of $\mathbb{Z}_2 \times \mathbb{Z}_2$.

b.) Try to remove the phase to get a true representation.

c.) Show that ρ defines a section of an exact sequence with G given by the Pauli group.

15.2.1 How projective representations arise in quantum mechanics

The following material, while very important, assumes knowledge of some of the linear algebra from Chapter 2 and some familiarity with quantum mechanics. For further details see Chapter **** below. The reader should also consult Section 2 of [162]

Let us review, very briefly the most essential points of quantum mechanics: The Dirac-von Neumann axioms of quantum mechanics posit that to a physical system we associate a complex Hilbert space $\mathcal{H}$ such that

1. Physical states are identified with traceclass positive operators ρ of trace one. They are usually called *density matrices*. We denote the space of physical states by $\mathcal{S}$. [163]

2. Physical observables are identified with self-adjoint operators. We denote the set of (bounded) self-adjoint operators by $\mathcal{O}$.

3. The *Born rule* states that when measuring the observable O in a state ρ the probability of measuring value $e \in E \subset \mathbb{R}$, where E is a Borel-measurable subset of $\mathbb{R}$, is

$$P_{\rho,O}(E) = \mathrm{Tr} P_O(E)\rho. \tag{15.53}$$

Here P_O is the projection-valued-measure associated to the self-adjoint operator O by the spectral theorem. For example, if $\mathcal{O}$ has a complete discrete spectrum $\{\lambda_i\}$ of eigenvalues so that

$$\mathcal{O} = \sum_i \lambda_i P(\lambda_i) \tag{15.54}$$

where $P(\lambda_i)$ is the projection operator onto the eigenspace with eigenvalue λ_i then

$$P_O(E) = \sum_{\lambda_i \in E} P(\lambda_i) \tag{15.55}$$

When the spectrum of O is more complicated, e.g. if there is a continuous spectrum, one can still define $P_O(E)$, but the story is more involved. See Chapter 2, the Linear Algebra User's Manual.

---

[162] http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf

[163] As explained in the Linear Algebra User's Manual, positive operators can be defined as operators A such that $(\psi, A\psi) \geq 0$ for every $\psi \in \mathcal{H}$. Such operators are always self-adjoint. Indeed, any operator A such that $(\psi, A\psi) \in \mathbb{R}$ for all $\psi \in \mathcal{H}$ is self-adjoint. To see this note that $(\psi, A\psi)^* = (\psi, A\psi)$ and hence $(\psi, A\psi) = (\psi, A^\dagger\psi)$. Now apply this equation to $\psi_1 + z\psi_2$ for $z = 1$ and $z = \sqrt{-1}$ and add the resulting equations to deduce that $(\psi_1, A\psi_2) = (\psi_1, A^\dagger\psi_2)$ for all pairs $\psi_1, \psi_2 \in \mathcal{H}$. Choose an ON basis for $\mathcal{H}$ to deduce that $A = A^\dagger$.

4. There are further axioms regarding time-development, and so on, but the above is all we need for the present discussion.

Given this setup up the natural notion of a general "symmetry" in quantum mechanics is the following:

Definition An *automorphism* of a quantum system is a pair of bijective maps $s_1 : \mathcal{S} \to \mathcal{S}$ and $s_2 : \mathcal{O} \to \mathcal{O}$ and where s_2 is real linear on $\mathcal{O}$ such that (s_1, s_2) preserves probability measures:

$$P_{s_1(\rho),s_2(O)} = P_{\rho,O} \tag{15.56}$$

This set of mappings forms a group which we will call the group of *quantum automorphisms*.

♣Need to state some appropriate continuity properties. ♣

While this is conceptually straightforward, it is an unwieldy notion of symmetry. We will now simplify it considerably, ending up with the crucial theorem known as *Wigner's theorem*.

We begin by noting that the space of density matrices is a convex set. The convexity means that if ρ_1, ρ_2 are density matrices then for all $0 \leq t \leq 1$

$$t\rho_1 + (1-t)\rho_2 \tag{15.57}$$

is a density matrix. Given a convex set on defines an *extremal point* to be a point in the set which cannot be written in the above form with $0 < t < 1$. By definition, the *pure states* are the extremal points of $\mathcal{S}$. The pure states are just the dimension one projection operators.

Pure states are often referred to in the physics literature as "rays in Hilbert space" for the following reason:

If $\psi \in \mathcal{H}$ is a nonzero vector then it determines a line

$$\ell_\psi := \{z\psi | z \in \mathbb{C}\} := \psi\mathbb{C} \tag{15.58}$$

Note that the line does not depend on the normalization or phase of ψ, that is, $\ell_\psi = \ell_{z\psi}$ for any nonzero complex number z. Put differently, the space of such lines is projective Hilbert space

$$\mathbb{P}\mathcal{H} := (\mathcal{H} - \{0\})/\mathbb{C}^* \tag{15.59}$$

Equivalently, this can be identified with the space of rank one projection operators. Indeed, given any line $\ell \subset \mathcal{H}$ we can write, in Dirac's bra-ket notation: [164]

$$P_\ell = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} \tag{15.60}$$

[164]We generally denote inner products in Hilbert space by $(x_1, x_2) \in \mathbb{C}$ where $x_1, x_2 \in \mathcal{H}$. Our convention is that it is complex-linear in the second argument. However, we sometimes write equations in Dirac's bra-ket notation because it is very popular. In this case, identify x with $|x\rangle$. Using the Hermitian structure there is a unique anti-linear isomorphism of $\mathcal{H}$ with $\mathcal{H}^*$ which we denote $x \mapsto \langle x|$. Sometimes we denote vectors by Greek letters $\psi, \chi, \ldots$, and scalars by Latin letters $z, w, \ldots$. But sometimes we denote vectors by Latin letters, $x, w, \ldots$ and scalars by Greek letters, $\alpha, \beta, \ldots$.

where ψ is any nonzero vector in the line ℓ.

It is possible to argue (see the above reference) that such a symmetry maps pure states to pure states, and is completely determined by this map. So we can view the group of quantum automorphisms as the group of transformations of one-dimensional projection operators (or rays) that preserves overlaps

$$\mathrm{Tr}(P_1 P_2) = \frac{|\langle\psi_1|\psi_2\rangle|^2}{\| \psi_1 \|^2 \| \psi_2 \|^2} \tag{15.61}$$

The group of such transformations is denoted $\mathrm{Aut}(QM)$. Any symmetry of a quantum mechanical system will define a subgroup of this group.

Example: Let us consider this case of a single Qbit, namely $\mathcal{H} = \mathbb{C}^2$. First we write the most general general density matrix. Any 2×2 Hermitian matrix is of the form $a + \vec{b} \cdot \vec{\sigma}$ where $\vec{\sigma}$ is the vector of "Pauli matrices":

$$\begin{aligned} \sigma^1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \sigma^2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ \sigma^3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{15.62}$$

$a \in \mathbb{R}$ and $\vec{b} \in \mathbb{R}^3$. Now a density matrix ρ must have trace one, and therefore $a = \frac{1}{2}$. Then the eigenvalues are $\frac{1}{2} \pm |\vec{b}|$ so positivity means it must have the form

$$\rho = \frac{1}{2}(1 + \vec{x} \cdot \vec{\sigma}) \tag{15.63}$$

where $\vec{x} \in \mathbb{R}^3$ with $\vec{x}^2 \leq 1$.

The extremal states, corresponding to the rank one projection operators are therefore of the form

$$P_{\vec{n}} = \frac{1}{2}(1 + \vec{n} \cdot \vec{\sigma}) \tag{15.64}$$

where $\vec{n}$ is a unit vector. This gives the explicit identification of the pure states with elements of S^2. Moreover, we can easily compute:

$$\mathrm{Tr} P_{\vec{n}_1} P_{\vec{n}_2} = \frac{1}{2}(1 + \vec{n}_1 \cdot \vec{n}_2) \tag{15.65}$$

and $\vec{n}_1 \cdot \vec{n}_2 = \cos(\theta_1 - \theta_2)$ where $|\theta_1 - \theta_2|$ (with θ's chosen so this is between 0 and π) is the geodesic distance between the two points on the unit sphere.

There is another viewpoint which is useful. Nonzero vectors in $\mathbb{C}^2$ can be normalized to be in the unit sphere S^3. Then the association of projector to state given by

$$|\psi\rangle \to |\psi\rangle\langle\psi| = \frac{1}{2}(1 + \vec{n} \cdot \vec{\sigma}) \tag{15.66}$$

defines a map $\pi : S^3 \to S^2$ known as the *Hopf fibration*.

The unit sphere is a principal homogeneous space for $SU(2)$ and we may coordinatize $SU(2)$ by the Euler angles:

$$u = e^{-i\frac{\phi}{2}\sigma^3} e^{-i\frac{\theta}{2}\sigma^2} e^{-i\frac{\psi}{2}\sigma^3} \tag{15.67}$$

with range $0 \leq \theta \leq \pi$ and identifications:

$$(\phi, \psi) \sim (\phi + 4\pi, \psi) \sim (\phi, \psi + 4\pi) \sim (\phi + 2\pi, \psi + 2\pi) \tag{15.68}$$

We can make an identification with the unit sphere in $\mathbb{C}^2$ by viewing it as a homogeneous space:

$$\psi = \begin{pmatrix} e^{-i\frac{\psi+\phi}{2}} \cos\theta/2 \\ e^{-i\frac{\psi-\phi}{2}} \sin\theta/2 \end{pmatrix} = u \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{15.69}$$

The projector onto the line through this space is

$$P_{\ell_\psi} = |\psi\rangle\langle\psi| = \frac{1}{2}(1 + \vec{n} \cdot \vec{\sigma}) \tag{15.70}$$

with $\vec{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ as usual. Alternatively, we could map $\pi : S^3 \to S^2$ by $\pi(\psi) = [\psi_1 : \psi_2] \cong \mathbb{C}P^1$, and this will correspond to the point in S^2 by the usual stereographic projection.

♣from which pole? ♣

In any case, for the case $N = 2$ we see that $\mathrm{Aut}_{\mathrm{qtm}}(\mathbb{P}\mathcal{H})$ is just the group of isometries of S^2 with its round metric. This group is well known to be the orthogonal group $O(3)$.

In general for $\mathcal{H} = \mathbb{C}^{N+1}$ the space of pure states is $\mathbb{CP}^N$. This space has a natural homogeneous metric known as the Fubini-Study metric. When it is suitably normalized the overlap function $\mathfrak{o}$ is nicely related to the Fubini-Study distance d by

$$\mathfrak{o}(\ell_1, \ell_2) = \left(\cos\frac{d(\ell_1, \ell_2)}{2}\right)^2 \tag{15.71}$$

Now, it is hard to work with the space of one-dimensional projection operators since it is not a linear space: The sum of two one-dimensional projection operators is typically not even proportional to a projection operator. It would be much nicer to work with operators acting on Hilbert space. A fundamental theorem of quantum mechanics known as *Wigner's theorem* states that $\mathrm{Aut}(QM)$ is indeed a quotient of a certain group of operators on a Hilbert space. This group, denoted $\mathrm{Aut}(\mathcal{H})$, is the group of the norm-preserving unitary and anti-unitary operators on Hilbert space. We now explain a bit more about $\mathrm{Aut}(\mathcal{H})$.

A unitary operator on $\mathcal{H}$ is a $\mathbb{C}$-linear operator $u : \mathcal{H} \to \mathcal{H}$ that preserves norms:

$$\| u\psi \| = \| \psi \| \tag{15.72}$$

An anti-unitary operator on $\mathcal{H}$ is an $\mathbb{C}$-anti-linear operator [165] $a : \mathcal{H} \to \mathcal{H}$ that preserves norms:

$$\| a\psi \| = \| \psi \| \tag{15.73}$$

[165] See Chapter 2, the Linear Algebra User's Manual, for more on this. Briefly this means $a(\psi_1 + \psi_2) = a(\psi_1) + a(\psi_2)$ for any two vectors but $a(z\psi) = z^* a(\psi)$ for any scalar $z \in \mathbb{C}$. Note that we must then define the adjoint by $\langle a^\dagger \psi_1, \psi_2\rangle := \langle \psi_1, a\psi_2\rangle^*$ in a convention where the sesquilinear form is $\mathbb{C}$-antilinear in the first argument and linear in the second.

The composition of unitary operators is clearly unitary. The composition of unitary and antiunitary is antiunitary, and the composition of antiunitaries is unitary so we have an exact sequence

$$1 \to U(\mathcal{H}) \to \mathrm{Aut}(\mathcal{H}) \xrightarrow{\phi} \mathbb{Z}_2 \to 1 \tag{15.74}$$

where

$$\phi(g) = \begin{cases} +1 & g \text{ unitary} \\ -1 & g \text{ anti} - \text{unitary} \end{cases} \tag{15.75}$$

There is a homomorphism $\pi : \mathrm{Aut}(\mathcal{H}) \to \mathrm{Aut}(QM)$ defined by

$$\pi(u) : P \mapsto uPu^\dagger \tag{15.76}$$

for $u \in U(\mathcal{H})$ and similarly

$$\pi(a) : P \mapsto aPa^\dagger \tag{15.77}$$

for anti-unitary operators a. See the footnote above for the definition of the adjoint of an anti-unitary operator. In both cases $u^{-1} = u^\dagger$ and $a^{-1} = a^\dagger$ and these operations preserve the overlap function.

The fiber of the map π can be thought of as possible c-number phases which can multiply the operator on Hilbert space representing a symmetry operation:

$$1 \ \to \ U(1) \ \to \ \mathrm{Aut}(\mathcal{H}) \xrightarrow{\pi} \mathrm{Aut}(QM) \ \to 1 \tag{15.78}$$

Here the $U(1)$ is the group of phases acting on quantum states: $\psi \to z\psi$ for $z \in U(1)$.

The upshot is that, given a classical symmetry group G of a physical system, for each $g \in G$ we can associate a unitary, or antiunitary, operator $U(g)$ acting on a Hilbert space. Quantum mechanics only guarantees that

$$U(g_1)U(g_2) = c(g_1, g_2)U(g_1 g_2) \tag{15.79}$$

for some phase factor $c(g_1, g_2)$, which, in general, cannot be removed by a redefinition of $U(g)$ by a phase $\tilde{U}(g) = b(g)U(g)$. So we have a projective representation of the classical symmetry G.

Put differently: In a given physical system we will not consider the full group $\mathrm{Aut}(QM)$ as a group of symmetries because. For example, the time-dynamics of a nonrelativistic quantum system is governed by the Schrödinger equation:

$$\mathrm{i}\hbar \frac{\partial \psi}{\partial t} = H\psi \tag{15.80}$$

where H is a self-adjoint operator, the Hamiltonian. For time-independent Hamiltonians the unitary time evolution is governed by

$$U(t) = \exp[-\mathrm{i}\frac{t}{\hbar}H] \tag{15.81}$$

Only a subgroup of $Aut(QM)$ will commute with the resultant flows on the space of states. If we think of G as embedded in $\mathrm{Aut}(QM)$ then we have

$$\begin{array}{ccccccccc} & & & & & & G & & \\ & & & & & & \downarrow \rho & & \\ 1 & \longrightarrow & U(1) & \xrightarrow{\iota} & \mathrm{Aut}(\mathcal{H}) & \xrightarrow{\pi} & \mathrm{Aut}(QM) & \longrightarrow & 1 \end{array} \tag{15.82}$$

The group which acts on the Hilbert space will be a central extension of G given by the pullback construction.

One thing we can note is that finite dimensional matrices are always associative! [166] So for all $g_1, g_2, g_3 \in G$

$$(U(g_1)U(g_2))U(g_3) = U(g_1)(U(g_2)U(g_3)) \tag{15.83}$$

and hence

$$c(g_1, g_2)c(g_1g_2, g_3) = c(g_1, g_2g_3)c(g_2, g_3) \tag{15.84}$$

We note that projective representations are quite pervasive in modern physics:

1. Projective representations appear naturally in quantization of bosons and fermions. The Heisenberg group is an extension of a translation group (on phase space). In addition, the symplectic group of linear canonical transformations gets quantum mechanically modified by a central extension to define something called the metaplectic group.

2. Projective representations are important in the theory of anomalies in quantum field theory.

3. Projective representations are very important in conformal field theory. The Virasoro group, and the Kac-Moody groups are both nontrivial central extensions of simpler objects.

Now, as we will explain near the end of section 15.3 below projective representations are very closely connected to central extensions. So in the next section we turn to a deeper investigation into the structure of central extensions.

Remark: The fact that the symmetry operators $U(g)$ should commute with the Hamiltonian has an important implication. Suppose that H has a complete set of eigenvectors Ψ_λ where $\lambda \in \mathrm{Spec}(H)$ is a discrete set of eigenvalues. Then we can restrict $U(g)$ to the different eigenspaces $\mathcal{H}_\lambda$ of H, for if

$$H\psi_\lambda = E_\lambda \psi_\lambda \tag{15.85}$$

and $U(g)$ commutes with H then $U(g)\psi_\lambda$ also is an eigenvector of eigenvalue E_λ. This means that the eigenspaces $\mathcal{H}_\lambda$ are each projective representations of G. This can be extremely useful in diagonalizing H and simplifying computations.

♣Need to work in some examples ♣

[166] And the same holds for linear operators on infinite-dimensional Hilbert spaces provided the domains are such that the composition of the three operators is well-defined.

Exercise *Kernel Of π In The Wigner Sequence*

Show that the kernel of π in the exact sequence (15.78) is precisely the group of phases times the identity operator. [167]

15.3 How To Classify Central Extensions

There is an interesting way to classify central extensions of G by A.

As before let $s : G \to \tilde{G}$ be a "section" of π. That is, a map such that

$$\pi(s(g)) = g \qquad \forall g \in G \tag{15.86}$$

As we have stressed, in general s is not a homomorphism. In the case when the sequence splits, that is, when there exists a section which is a homomorphism, then we can say $\tilde{G}$ is isomorphic to a direct product $\tilde{G} \cong A \times G$ via

$$\iota(a)s(g) \to (a, g) \tag{15.87}$$

When the sequence splits the semidirect product of the previous section is a direct product because A is central, so $\omega_g(a) = a$.

Now, let us allow that (15.22) does not necessarily split. Let us choose any section s and measure by how much s differs from being a homomorphism by considering the combination:

$$s(g_1)s(g_2)\left(s(g_1g_2)\right)^{-1}. \tag{15.88}$$

Now the quantity (15.88) is in the kernel of π and hence in the image of ι. Since ι is injective we can *define* a function $f_s : G \times G \to A$ by the equation

$$\iota(f_s(g_1, g_2)) := s(g_1)s(g_2)\left(s(g_1g_2)\right)^{-1}. \tag{15.89}$$

That is, we can write:

$$s(g_1)s(g_2) = \iota(f_s(g_1, g_2))s(g_1g_2) \tag{15.90}$$

The function f_s satisfies the important *cocycle identity*

$$\boxed{\boxed{f(g_2, g_3)f(g_1, g_2g_3) = f(g_1, g_2)f(g_1g_2, g_3)}} \tag{15.91}$$

Exercise *Derivation Of The Cocycle Identity*

Verify (15.91) by using (15.89) to compute $s(g_1g_2g_3)$ in two different ways.

[167] *Answer*: Suppose $uP = Pu$ for every rank one projection operator. Consider the projection operator for the line through $\psi_1 + z\psi_2$ for any two nonzero vectors $\psi_1, \psi_2 \in \mathcal{H}$. Applying the condition for the cases $z = 1$ and $z = \sqrt{-1}$ deduce that u commutes with $|\psi_1\rangle\langle\psi_2|$. Thus, choosing an ON basis e_i it commutes with $|e_i\rangle\langle e_j|$ and therefore must be proportional to the identity matrix. On the other hand, it also must preserve norms, so it is the group of multiplication by a phase.

(Note that simply substituting (15.89) into (15.91) is not obviously going to work because $\tilde{G}$ need not be abelian.)

Exercise *Simple consequences of the cocycle identity*
a.) By putting $g_1 = 1$ and then $g_3 = 1$ show that any cocycle f must satisfy:

$$f(g,1) = f(1,g) = f(1,1) \qquad \forall g \in G \tag{15.92}$$

b.) Show that [168]

$$f(g,g^{-1}) = f(g^{-1},g). \tag{15.93}$$

Now we introduce some fancy terminology:

Definition: In general
1. A *2-cochain on G with values in A* is a function

$$f : G \times G \to A \tag{15.94}$$

We denote the set of all such 2-cochains by $C^2(G,A)$.
2. A *2-cocycle* is a 2-cochain $f : G \times G \to A$ satisfying (15.91). We denote the set of all such 2-cocycles by $Z^2(G,A)$.

Remarks:

1. The fancy terminology is introduced for a good reason because there is a topological space and a cohomology theory underlying this discussion. See Section §15.8 and Section §17.2 for further discussion.
2. Note that $C^2(G,A)$ is naturally an abelian group because A is an abelian group. (Recall example 2.7 of Section §2.) $Z^2(G,A)$ inherits an abelian group structure from $C^2(G,A)$.

So, in this language, given a central extension of G by A and a section s we naturally obtain a two-cocycle $f_s \in Z^2(G,A)$ via (15.89).
Now, if we choose a different section $\hat{s}$ then [169]

$$\hat{s}(g) = \iota(t(g))s(g) \tag{15.95}$$

[168] *Answer*: Consider the triple $g \cdot g^{-1} \cdot g$ and apply part (a).
[169] Since we are working with central extensions we could put the $\iota(t(g))$ on either side of the $s(g)$. However, when we discuss non-central extensions later the order will matter.

for some function $t : G \to A$. It is easy to check that

$$f_{\hat{s}}(g_1, g_2) = f_s(g_1, g_2)t(g_1)t(g_2)t(g_1g_2)^{-1} \tag{15.96}$$

where we have used that $\iota(A)$ is central in $\tilde{G}$.

Definition: In general two 2-cochains f and $\hat{f}$ are said to *differ by a coboundary* if there exists a function $t : G \to A$ such that

$$\hat{f}(g_1, g_2) = f(g_1, g_2)t(g_1)t(g_2)t(g_1g_2)^{-1} \tag{15.97}$$

for all $g_1, g_2 \in G$.

One can readily check, using the condition that A is Abelian, that if f is a cocycle then any other $\hat{f}$ differing by a coboundary is also a cocycle. Moreover, being related by a cocycle defines an equivalence relation on the set of cocycles $f \sim \hat{f}$. Thus, we may define:

Definition: The *group cohomology* $H^2(G, A)$ is the set of equivalence classes of 2-cocycles modulo equivalence by coboundaries.

Now, the beautiful theorem states that group cohomology classifies central extensions:
170

Theorem: Isomorphism classes of central extensions of G by an abelian group A are in 1-1 correspondence with the second cohomology set $H^2(G, A)$.

Proof: Let $\mathcal{E}(G, A)$ denote the set of all extensions of G by A, and let $\overline{\mathcal{E}}(G, A)$ denote the set of all isomorphism classes of extensions of G by A.

We first construct a map:

$$\Psi_{\mathcal{E}\to H} : \mathcal{E}(G, A) \to H^2(G, A) \tag{15.98}$$

To do this, we choose a section, then from (15.89)(15.91)(15.96) we learn that we get a cocycle whose cohomology class does not depend on the section. So

$$\Psi_{\mathcal{E}\to H}\left(1 \to A \to \tilde{G} \to G \to 1\right) = [f_s] \tag{15.99}$$

is well-defined, because the RHS does not depend on the choice of section s.

Now we claim that this map descends to a map $\overline{\Psi}_{\mathcal{E}\to H} : \overline{\mathcal{E}}(G, A) \to H^2(G, A)$. Indeed, if we have an isomorphism of central extensions:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & A & \xrightarrow{\iota} & \widetilde{G} & \xrightarrow{\pi} & G & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle \mathrm{Id}} & & \downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \mathrm{Id}} & & \\ 1 & \longrightarrow & A & \xrightarrow{\iota'} & \widetilde{G}' & \xrightarrow{\pi'} & G & \longrightarrow & 1 \end{array} \tag{15.100}$$

where $\psi : \tilde{G} \to \tilde{G}'$ is an isomorphism such that the inverse also leads to a commutative diagram, then ψ can be used to map sections of $\pi : \tilde{G} \to G$ to sections of $\pi' : \tilde{G}' \to G$ by $s \mapsto s'$ where $s'(g) = \psi(s(g))$. Then

♣where do we use this condition in the proof? ♣

$$\begin{aligned} s'(g_1)s'(g_2) &= \psi(s(g_1))\psi(s(g_2)) \\ &= \psi(s(g_1)s(g_2)) \\ &= \psi(\iota(f_s(g_1,g_2))s(g_1g_2)) \\ &= \psi(\iota(f_s(g_1,g_2)))\psi(s(g_1g_2)) \\ &= \iota'(f_s(g_1,g_2))s'(g_1g_2) \end{aligned} \tag{15.101}$$

and hence we assign precisely the same 2-cocycle $f(g_1, g_2)$ to the section s'. Hence the map $\Psi_{\mathcal{E}\to H}$ only depends on the isomorphism class of the extension. This defines the map $\overline{\Psi}_{\mathcal{E}\to H}$.

Conversely, we can define a map $\Psi_{H\to\mathcal{E}} : Z^2(G, A) \to \mathcal{E}(G, A)$ as follows: Given a cocycle $f \in Z^2(G, A)$ we may define $\tilde{G} = A \times G$ as a set and we use f to *define* the multiplication law:

$$(a_1, g_1)(a_2, g_2) := (a_1a_2f(g_1, g_2), g_1g_2) \tag{15.102}$$

You should check that this does define a valid group multiplication: The associativity follows from the cocycle relation. Note that if we use the *trivial cocycle*: $f(g_1, g_2) = 1$ for all $g_1, g_2 \in G$ then we just get the direct product of groups.

Now suppose that we use two 2-cocycles f and f' which are related by a coboundary as in (15.97) above. Then we claim that the map $\psi : \tilde{G} \to \tilde{G}'$ defined by

$$\psi : (a, g) \to (at(g)^{-1}, g) \tag{15.103}$$

is an isomorphism of central extensions as in (15.100). This means that the map $\Psi_{H\to\mathcal{E}} : Z^2(G, A) \to \mathcal{E}(G, A)$ actually descends to a well-defined map

$$\overline{\Psi}_{H\to\mathcal{E}} : H^2(G, A) \to \overline{\mathcal{E}}(G, A) \tag{15.104}$$

We leave it to the reader to check that $\overline{\Psi}_{H\to\mathcal{E}}$ and $\overline{\Psi}_{\mathcal{E}\to H}$ are inverse maps. ♠

Remarks:

1. *Central extensions and projective representations.* A very important consequence of the construction (15.102) is that, if we are given a projective representation of G then we can associate a centrally extended group $\widetilde{G}$:

   $$1 \to U(1) \to \widetilde{G} \to G \to 1 \tag{15.105}$$

   and a <u>true</u> representation $\widetilde{\rho}$ of $\widetilde{G}$:

   $$\widetilde{\rho}(z, g) := z\rho(g) \tag{15.106}$$

[170] In fact, both the set of isomorphism classes of extensions and $H^2(G, A)$ are Abelian groups and a stronger statement is that the 1-1 correspondence described here is an isomorphism of Abelian groups.

The evil failure of $\rho(g)$ to be a true representation of G now becomes a virtuous fact that allows $\widetilde{\rho}$ to be a true representation of $\widetilde{G}$. This is the typical situation in quantum mechanics, where G is a group of classical symmetries and $\widetilde{G}$ is the group that is implemented quantum-mechanically. A good example is the spin-1/2 system where $G = SO(3)$ is the classical group of rotations but for a quantum rotor the proper symmetry group is $\widetilde{G} = SU(2)$. There are many many other examples.

2. *Group Structure on* $\overline{\mathcal{E}}(G, A)$. The set $H^2(G, A)$ carries a natural structure of an Abelian group. Indeed, as we remarked above $C^2(G, A)$, being a set of maps with target space a group, A, is naturally a group. Then, because A is Abelian, we can define a group structure on $Z^2(G, A)$ by the rule:

$$(f_1 \cdot f_2)(g, g') = f_1(g, g') \cdot f_2(g, g') \tag{15.107}$$

where we are writing the product in A multiplicatively. Again using the fact that A is abelian this descends to a well-defined multiplication on cohomology classes: $[f_1] \cdot [f_2] := [f_1 \cdot f_2]$. Therefore $H^2(G, A)$ itself is an abelian group. The identity element corresponds to the cohomology class of the trivializable cocycles, which in turn corresponds to the split extension $A \times G$.

♣It might be clearer to write A additively... ♣

It is natural to ask whether one can give a more canonical description of the abelian group structure on the set of equivalence classes of central extensions of G by A. Indeed we can: We pull back the Cartesian product to the diagonal of $G \times G$ and then push forward by the multiplication map $\mu : A \times A \to A$. That is, suppose we have two central extensions:

$$\mathcal{E}_1 : \qquad 1 \to A \overset{\iota_1}{\to} \tilde{G}_1 \overset{\pi_1}{\to} G \to 1 \tag{15.108}$$

$$\mathcal{E}_2 : \qquad 1 \to A \overset{\iota_2}{\to} \tilde{G}_2 \overset{\pi_2}{\to} G \to 1 \tag{15.109}$$

The Cartesian product $\mathcal{E}_1 \times \mathcal{E}_2$ is the extension of $G \times G$ by $A \times A$ using the group $\tilde{G}_1 \times \tilde{G}_2$ with the Cartesian product of the group homomorphisms. We want an extension of G by A, corresponding, under the 1-1 correspondence of the above theorem to the natural group structure on $H^2(G, A)$. To construct it, let

$$\Delta : G \to G \times G \tag{15.110}$$

be the diagonal homomorphism: $\Delta : g \mapsto (g, g)$. Then we claim that the product extension $\mathcal{E}_1 \cdot \mathcal{E}_2$ can be identified as

$$\mathcal{E}_1 \cdot \mathcal{E}_2 = \mu_* \Delta^* (\mathcal{E}_1 \times \mathcal{E}_2) \tag{15.111}$$

where $\Delta^*(\mathcal{E}_1 \times \mathcal{E}_2)$ is the pull-back extension under Δ (see equation (15.34)), an extension of G by $A \times A$, and μ_* is the pushforward extension. In concrete terms the pullback extension under Δ^* is:

$$1 \to A \times A \overset{(\iota_1, \iota_2)}{\to} \widehat{G_{12}} \overset{\pi_{12}}{\to} G \to 1 \tag{15.112}$$

where

$$\widehat{G_{12}} := \{(\tilde{g}_1, \tilde{g}_2) | \pi_1(\tilde{g}_1) = \pi_2(\tilde{g}_2)\} \subset \tilde{G}_1 \times \tilde{G}_2 \tag{15.113}$$

We can define $\pi_{12}(\tilde{g}_1, \tilde{g}_2) := \pi_1(\tilde{g}_1) = \pi_2(\tilde{g}_2)$. Now consider the "anti-diagonal"

$$A^{\rm anti} := \ker(\mu) = \{(a, a^{-1})\} \subset A \times A \tag{15.114}$$

and its image:

$$N := \{(\iota_1(a), \iota_2(a^{-1})) | a \in A\} \subset \widehat{G_{12}} \tag{15.115}$$

Because we are working with central extensions this will be a normal subgroup. Then we let

$$\tilde{G}_{12} := \widehat{G_{12}}/N \tag{15.116}$$

Since N is in the kernel of π_{12} and since it is central the homomorphism π_{12} descends to a surjective homomorphism which we will also call $\pi_{12} : \tilde{G}_{12} \to G$. Now we have an exact sequence

$$1 \to A \stackrel{\iota_{12}}{\to} \tilde{G}_{12} \stackrel{\pi_{12}}{\to} G \to 1 \tag{15.117}$$

where $\iota_{12}(a) := [(\iota_1(a), \iota_2(1))] = [(\iota_1(1), \iota_2(a))]$. Given sections s_1, s_2 of π_1, π_2 respectively we can define a section $s_{12}(g) := [(s_1(g), s_2(g))]$ and one can check that the resulting cocycle is indeed in the cohomology class of $f_{s_1} \cdot f_{s_2}$. The extension (15.117) represents the product of extensions (15.108) and (15.109). The point of this construction is that it is canonical: We did not make any choices of sections to define the product extension.

3. *Trivial vs. Trivializable.* Above we defined the trivial cocycle to be the one with $f(g_1, g_2) = 1_A$ for all g_1, g_2. We define a cocycle to be *trivializable* if it is cohomologous to the trivial cocycle. Note that a trivializable cocycle f could be trivialized in multiple ways. Suppose both b and $\tilde{b}$ trivialize f. Then you should show that $\tilde{b}$ and b "differ" by a group homomorphism $\phi : G \to A$ in the sense that

$$\tilde{b}(g) = \phi(g) b(g) \tag{15.118}$$

There are situations where a cohomological obstruction vanishes and the choice of trivialization has physical significance.

♣Should give some examples: H^3 is obstruction to orbifolding CFT and choice of trivialization is H^2 - hence discrete torsion. There are bundle examples. Find example where class in H^2 is zero but trivialization has physical meaning. ♣

4. *An analogy to gauge theory*: Changing a cocycle by a coboundary is strongly analogous to making a gauge transformation in a gauge theory. In Maxwell's theory we can make a change of gauge of the vector potential A_μ by

$$A'_\mu(x) = A_\mu(x) - {\rm i} g^{-1}(x) \partial_\mu g(x) \tag{15.119}$$

where $g(x) = e^{{\rm i}\epsilon(x)}$ is a function on spacetime valued in $U(1)$. In the case of electromagnetism we would say that A_μ is trivializable if there is a gauge transformation $g(x)$ that simplifies it to 0. (For valid gauge transformations $g(x)$ must be a single-valued function on spacetime.) If we are presented with $A_\mu(x)$ and we want to know if it is trivializable then we should check whether gauge invariant quantities vanish.

One such quantity is the fieldstrength tensor $F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu$, but this is not a complete gauge invariant. The isomorphism class of a field is completely specified by the holonomies $\exp i \oint_\gamma A$ around all the closed cycles γ in spacetime. Even when $A_\mu(x)$ is not trivializable, it is very often useful to use gauge transformations to try to simplify A_μ. In the next remark we do the same for cocycles.

5. *Simplifying Cocycles Using Coboundaries.* Using a coboundary one can usefully simplify cocycles. Since this topic will be unfamiliar to some readers we explain this in excruciating detail. Those who are familiar with cohomology can safely skip the rest of this remark. To begin, note that a coboundary modification takes a cochain f to $f^{(1)}$ satisfying:

$$f^{(1)}(1,1) = f(1,1)\frac{t(1)t(1)}{t(1\cdot 1)} = f(1,1)t(1) \tag{15.120}$$

so by choosing any function t such that $t(1) = f(1,1)^{-1}$ we get a new cochain satisfying $f^{(1)}(1,1) = 1$. Choose any such function. (The simplest thing to do is set $t(g) = 1$ for all other $g \neq 1$. We will make this choice, but it is really not necessary.) Now recall that if f is a cocycle then a modification of f by any coboundary produces a new cochain $f^{(1)}$ that is also a cocycle. So now, if f is a cocycle and we have set $f^{(1)}(1,1) = 1$ then, by (15.92) we have $f^{(1)}(g,1) = f^{(1)}(1,g) = 1$ for all g. Now, we can continue to make modifications by coboundaries to simplify further our cocycle $f^{(1)}$. In order not to undo what we have done we require that the new coboundaries we use, say, $\tilde{t}$ satisfy $\tilde{t}(1) = 1$. We may say that we are "partially choosing a gauge" by choosing representatives so that $f^{(1)}(1,1) = 1$ and then the further coboundaries $\tilde{t}$ must "preserve that gauge." Now suppose that $g \neq 1$. Then (using our particular choice of t above):

$$f^{(1)}(g,g^{-1}) = f(g,g^{-1})\frac{1}{t(1)} = f(g,g^{-1})f(1,1) \tag{15.121}$$

is not particularly special. (Remember that we are making the somewhat arbitrary choice that $t(g) = 1$ for $g \neq 1$.) So we have not simplified these quantities. However, we still have plenty of gauge freedom left and we can try to simplify the values as follows: Suppose, first, that $g \neq g^{-1}$, equivalently, suppose $g^2 \neq 1$ so g is not an involution. Then we can make another "gauge transformation" by a coboundary function $\tilde{t}$ to produce:

$$f^{(2)}(g,g^{-1}) = f^{(1)}(g,g^{-1})\frac{\tilde{t}(g)\tilde{t}(g^{-1})}{\tilde{t}(g\cdot g^{-1})} = \hat{f}(g,g^{-1})\tilde{t}(g)\tilde{t}(g^{-1}) \tag{15.122}$$

where in the second equality we used the "gauge-preserving" property that $\tilde{t}(1) = 1$. Now, in any way you like, divide the non-involution elements of G into two disjoint sets $S_1 \amalg S_2$ so that no two group elements in S_1 are related by $g \to g^{-1}$. Then, if $g \in S_2$ we have $g^{-1} \in S_1$ and vice versa. Then we can choose a function $\tilde{t}$ so that for every $g \in S_2$ we have

$$\tilde{t}(g) = (\tilde{t}(g^{-1}))^{-1}(f^{(1)}(g,g^{-1}))^{-1} \tag{15.123}$$

Consequently:

$$f^{(2)}(g, g^{-1}) = 1 \qquad \forall g \in S_2 \tag{15.124}$$

It doesn't really matter what we choose for $\tilde{t}$ on S_1. For definiteness we choose it to be $= 1$. But if we had made another choice the above procedure would still lead to equation (15.124). Now recall from (15.93) that any cocycle f satisfies $f(g, g^{-1}) = f(g^{-1}, g)$ for all g. Since $f^{(2)}$ is a cocycle (if we started with a cocycle f) then we conclude that for all the non-involutions:

$$f^{(2)}(g, g^{-1}) = f^{(2)}(g^{-1}, g) = 1 \qquad \forall g \in S_1 \amalg S_2 \tag{15.125}$$

Note that there is still a lot of "gauge freedom": We have not yet constrained $\tilde{t}(g)$ for $g \in S_1$, nor have we constrained $\tilde{t}(g)$ for the involutions, that is, the group elements g with $g^2 = 1$. What can we say about $f^{(2)}(g, g)$ for g an involution? we have

$$f^{(2)}(g, g) = f^{(1)}(g, g)\frac{\tilde{t}(g)^2}{\tilde{t}(g^2)} = f^{(1)}(g, g)(\tilde{t}(g))^2 \tag{15.126}$$

Now, it might, or might not be the case that $f^{(1)}(g, g)$ is a perfect square in the group. If it is not a perfect square then we are out of luck: We cannot make any further gauge transformations to set $f^{(2)}(g, g) = 1$. Now one can indeed check that the property of $f(g, g)$ being a perfect square, or not, for an involution g is a truly "gauge invariant" condition. Therefore we have proven: *If $f(g, g)$ is not a perfect square for some nontrivial involution g then we know that f is not "gauge equivalent" - that is, is not cohomologous to - the trivial cocycle. That is, $[f]$ is a nontrivial cohomology class.* Such cocycles will define nontrivial central extensions.

Example 1 . *Extensions of $\mathbb{Z}_2$ by $\mathbb{Z}_2$.* WLOG we can take $f(1, 1) = f(1, \sigma) = f(\sigma, 1) = 1$. Then we have two choices: $f(\sigma, \sigma) = 1$ or $f(\sigma, \sigma) = \sigma$. Each of these choices satisfies the cocycle identity and they are not related by a coboundary. Indeed σ is an involution and also σ is not a perfect square, so by our discussion above a cocycle with $f(\sigma, \sigma) = \sigma$ cannot be gauged to the trivial cocycle. In other words $H^2(\mathbb{Z}_2, \mathbb{Z}_2) = \mathbb{Z}_2$. For the choice $f = 1$ we obtain $\tilde{G} = \mathbb{Z}_2 \times \mathbb{Z}_2$. For the nontrivial choice $f(\sigma, \sigma) = \sigma$ we obtain $\tilde{G} \cong \mathbb{Z}_4$. Let us see this in detail. We'll let $\sigma_1 \in A \cong \mathbb{Z}_2$ and $\sigma_2 \in G \cong \mathbb{Z}_2$ be the nontrivial elements so we should write $f(\sigma_2, \sigma_2) = \sigma_1$. Note that $(\sigma_1, 1)$ has order 2, but then

$$(1, \sigma_2) \cdot (1, \sigma_2) = (f(\sigma_2, \sigma_2), 1) = (\sigma_1, 1) \tag{15.127}$$

shows that $(1, \sigma_2)$ has order 4. Moreover $(\sigma_1, \sigma_2) = (\sigma_1, 1)(1, \sigma_2) = (1, \sigma_2)(\sigma_1, 1)$. Thus,

$$\begin{aligned} \Psi &: (\sigma_1, 1) \to \omega^2 = -1 \\ \Psi &: (1, \sigma_2) \to \omega \end{aligned} \tag{15.128}$$

where ω is a primitive 4th root of 1 defines an isomorphism with the group of fourth roots of unity. In conclusion, the nontrivial central extension of $\mathbb{Z}_2$ by $\mathbb{Z}_2$ is:

$$1 \to \mathbb{Z}_2 \to \mathbb{Z}_4 \to \mathbb{Z}_2 \to 1 \tag{15.129}$$

Recall that $\mathbb{Z}_4$ is not isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. The square of this extension is the trivial extension.

Example 2. *Extensions of $\mathbb{Z}_p$ by $\mathbb{Z}_p$.* The generalization of the previous example to the extension of $\mathbb{Z}_p$ by $\mathbb{Z}_p$ for an odd prime p is extremely instructive. So, let us study in detail the extensions

$$1 \to \mathbb{Z}_p \to G \to \mathbb{Z}_p \to 1 \tag{15.130}$$

In this example we will write our cyclic groups multiplicatively. Now, using methods of topology one can show that [171]

$$H^2(\mathbb{Z}_p, \mathbb{Z}_p) \cong \mathbb{Z}_p. \tag{15.131}$$

The result (15.131) should puzzle you. After all we know that G must be a group of order p^2, and we know from the class equation and Sylow's theorems that there are exactly two groups of order p^2, up to isomorphism! How is that compatible with the p distinct extensions predicted by equation (15.131) !? The answer is that there can be nonisomorphic extensions (15.22) involving the same group $\tilde{G}$. Let us see how this works in the present example by examining in detail the possible extensions:

$$1 \to \mathbb{Z}_p \xrightarrow{\iota} \mathbb{Z}_{p^2} \xrightarrow{\pi} \mathbb{Z}_p \to 1 \tag{15.132}$$

We write the first, second and third groups in this sequence as

$$\begin{aligned} \mathbb{Z}_p &= \langle \sigma_1 | \sigma_1^p = 1 \rangle \\ \mathbb{Z}_{p^2} &= \langle \alpha | \alpha^{p^2} = 1 \rangle \\ \mathbb{Z}_p &= \langle \sigma_2 | \sigma_2^p = 1 \rangle \end{aligned} \tag{15.133}$$

respectively.

For the injection ι we have

$$\iota(\sigma_1) = \alpha^x \tag{15.134}$$

for some x. For this to be a well-defined homomorphism we must have

$$\iota(1) = \iota(\sigma_1^p) = \alpha^{px} = 1 \tag{15.135}$$

and therefore $px = 0 \mod p^2$ and therefore $x = 0 \mod p$. But since ι must be an injection it must be of the form

$$\iota_k(\sigma_1) := \alpha^{kp} \tag{15.136}$$

[171] You can also show it by examining the cocycle equation directly. We will write down the nontrivial cocycles presently.

where k is relatively prime to p. We can take

$$1 \leq k \leq p-1 \tag{15.137}$$

or (preferably) we can regard $k \in \mathbb{Z}_p^*$.

Similarly, for π we must have $\pi(\alpha) = \sigma_2^y$ for some y. Now, since π has to be surjective, σ_2^y must be a generator and hence π must be of the form

$$\pi_r(\alpha) = \sigma_2^r \qquad\qquad 1 \leq r \leq p-1 \tag{15.138}$$

where again we should really regard r as an element of $\mathbb{Z}_p^*$.

Note that the kernel of π_r is the set of elements α^ℓ with $\sigma_2^{\ell r} = 1$. This implies $\ell r = 0 \mod p$ and therefore $\ell = 0 \mod p$ so

$$\ker(\pi_r) = \{1, \alpha^p, \alpha^{2p}, \ldots, \alpha^{(p-1)p}\} \tag{15.139}$$

Since $k \in \mathbb{Z}_p^*$ we have

$$\ker(\pi_r) = \mathrm{im}(\iota_k) \tag{15.140}$$

so our sequence is exact for any choice of $r, k \in \mathbb{Z}_p^*$. We have now described all the extensions of $\mathbb{Z}_p$ by $\mathbb{Z}_p$. Let us find a representative cocycle $f_{k,r}$ for each of these extensions.

To find the cocycle we choose a section of π_r. It is instructive to try to make it a homomorphism. Therefore we must take $s(1) = 1$. What about $s(\sigma_2)$? It must be of the form $s(\sigma_2) = \alpha^x$ for some x, and since $\pi_r(s(\sigma_2)) = \sigma_2$ we must have

$$\sigma_2^{xr} = \sigma_2 \tag{15.141}$$

so that

$$xr = 1 \mathrm{mod}\ p \tag{15.142}$$

Recall that $r \in \mathbb{Z}_p^*$ and let r^* be the integer $1 \leq r^* \leq p-1$ such that

$$rr^* = 1 \mathrm{mod}\ p \tag{15.143}$$

Then we have that $x = r^* + \ell p$ for any ℓ. That is, $s(\sigma_2)$ could be any of

$$\alpha^{r^*}, \alpha^{r^*+p}, \alpha^{r^*+2p}, \ldots, \alpha^{r^*+(p-1)p} \tag{15.144}$$

Here we will make the simplest choice $s(\sigma_2) = \alpha^{r^*}$. The reader can check that the discussion is not essentially changed if we make one of the other choices. (After all, this will just change our cocycle by a coboundary!)

Now that we have chosen $s(\sigma_2) = \alpha^{r^*}$, if s were a homomorphism then we would be forced to take:

$$\begin{aligned} s(1) &= 1 \\ s(\sigma_2) &= \alpha^{r^*} \\ s(\sigma_2^2) &= \alpha^{2r^*} \\ \vdots \qquad &\quad \vdots \\ s(\sigma_2^{p-1}) &= \alpha^{(p-1)r^*} \end{aligned} \tag{15.145}$$

But now we are stuck! The property that s is a homomorphism requires two contradictory things. On the one hand, we must have $s(1) = 1$ for any homomorphism. On the other hand, from the above equations we also must have $s(\sigma_2^p) = \alpha^{pr^*}$. But because $1 \leq r^* \leq p-1$ we know that $\alpha^{pr^*} \neq 1$. So the conditions for s being a homomorphism are impossible to meet. Therefore, with this choice of section we find a nontrivial cocycle as follows:

$$s(\sigma_2^x)s(\sigma_2^y)s(\sigma_2^{x+y})^{-1} = \begin{cases} 1 & x+y \leq p-1 \\ \alpha^{r^*p} & p \leq x+y \end{cases} \tag{15.146}$$

Here we computed:

$$\alpha^{r^*x}\alpha^{r^*y}\alpha^{-r^*(x+y-p)} = \alpha^{r^*p} \tag{15.147}$$

where you might note that if $p \leq x+y \leq 2p-2$ then $0 \leq x+y-p \leq p-2$. Therefore, our cocycle is $f_{k,r}$ where

$$f_{k,r}(\sigma_2^x, \sigma_2^y) := \begin{cases} 1 & x+y \leq p-1 \\ \sigma_1^{k^*r^*} & p \leq x+y \end{cases} \tag{15.148}$$

since

$$\iota_k(\sigma_1^{k^*r^*}) = \alpha^{k^*r^*kp} = \alpha^{r^*p} \tag{15.149}$$

and here we have introduced an integer $1 \leq k^* \leq p-1$ so that

$$kk^* = 1 \bmod p \tag{15.150}$$

Although it is not obvious from the above formula for $f_{k,r}$, we know that $f_{k,r}$ will satisfy the cocycle equation because we constructed it from a section of a group extension.

Now, we know the cocycle is nontrivial because $\mathbb{Z}_p \times \mathbb{Z}_p$ is not isomorphic to $\mathbb{Z}_{p^2}$. But let us try to trivialize our cocycle by a coboundary. So we modify our section to

$$\tilde{s}(\sigma_2^x) = \iota(t(\sigma_2^x))s(\sigma_2^x) \tag{15.151}$$

We can always write our function t in the form

$$t(\sigma_2^x) = \sigma_1^{\tau(x)} \tag{15.152}$$

for some function $\tau(x)$ valued in $\mathbb{Z}/p\mathbb{Z}$. We are trying to find a function $\tau(x)$ so that the new cocycle $f_{\tilde{s}}$ is identically 1. We certainly need $\tilde{s}(1) = 1$ and hence $\tau(\bar{0}) = \bar{0}$. But now, because $f(\sigma_2^x, \sigma_2^y) = 1$ already holds for $x+y \leq p-1$ don't want to undo that so we learn that

$$\tau(x) + \tau(y) - \tau(x+y) = 0 \text{mod} p \tag{15.153}$$

for $x+y \leq p-1$. This means we must take

$$\tau(x) = x\tau(1) \qquad\qquad 1 \leq x \leq p-1 \tag{15.154}$$

So, our coboundary is completely fixed up to a choice of $\tau(1)$. But now let us compute for $x+y \geq p-1$:

$$\tilde{s}(\sigma_2^x)\tilde{s}(\sigma_2^y)\tilde{s}(\sigma_2^{x+y})^{-1} = \alpha^{r^*p}\iota(\sigma_1^{\tau(x)+\tau(y)-\tau(x+y)}) = \alpha^{r^*p} \tag{15.155}$$

So, we cannot gauge the cocycle to one, confirming what we already knew: The cocycle is nontrivial.

Now let us see when the different extensions defined by $k, r \in \mathbb{Z}_p^*$ are actually equivalent. To see this let us try to construct φ so that

$$\begin{array}{ccccccccc} & & & & \langle\alpha\rangle & & & & \\ & & & \overset{\iota_{k_1}}{\nearrow} & & \overset{\pi_{r_1}}{\searrow} & & & \\ 1 & \longrightarrow & \langle\sigma_1\rangle & & \downarrow{\scriptstyle\varphi} & & \langle\sigma_2\rangle & \longrightarrow & 1 \\ & & & \underset{\iota_{k_2}}{\searrow} & & \underset{\pi_{r_2}}{\nearrow} & & & \\ & & & & \langle\alpha\rangle & & & & \end{array} \tag{15.156}$$

Now φ, being a homomorphism, must be of the form

$$\varphi(\alpha) = \alpha^y \tag{15.157}$$

for some y. We know this must be an isomorphism so y must be relatively prime to p. Moreover commutativity of the diagram implies

$$\pi_{r_2}(\varphi(\alpha)) = \pi_{r_1}(\alpha) \qquad \Rightarrow \qquad r_2 y = r_1 \bmod p \tag{15.158}$$

$$\varphi(\iota_{k_1}(\sigma_1)) = \iota_{k_2}(\sigma_1) \qquad \Rightarrow \qquad k_1 p y = k_2 p \bmod p^2 \qquad \Rightarrow \qquad k_1 y = k_2 \bmod p \tag{15.159}$$

Putting these equations together, and remembering that y is multiplicatively invertible modulo p we find that there exists a morphism of extensions iff

$$k_1 r_1 = k_2 r_2 \mathrm{mod} p \tag{15.160}$$

Note that the cocycles $f_{k,r}$ constructed in (15.148) indeed only depend on $kr\mathrm{mod}p$. Equivalently, we can label their cohomology class by $(kr)^* = k^* r^* \mathrm{mod} p$.

The conclusion is that $kr \in \mathbb{Z}_p^*$ is the invariant quantity. Extensions with the same group $\tilde{G} = \mathbb{Z}_{p^2}$ in the middle, but with different $kr \in \mathbb{Z}_p^*$, define inequivalent extensions of $\mathbb{Z}_p$ by $\mathbb{Z}_p$.

Now let us examine the group structure on the group cohomology. Just multiplying the cocycles we get:

$$\left(f_{k_1,r_1} \cdot f_{k_2,r_2}\right)(\sigma_2^x, \sigma_2^y) = \begin{cases} 1 & x + y \leq p - 1 \\ \sigma_1^{(k_1 r_1)^* + (k_2 r_2)^*} & p \leq x + y \end{cases} \tag{15.161}$$

Thus if we map

$$[f_{k,r}] \mapsto (kr)^* \mathrm{mod} \mathbb{Z}_p \tag{15.162}$$

we have a homomorphism of $H^2(G, A)$ to the <u>additive</u> group $\mathbb{Z}/p\mathbb{Z}$, with the trivializable cocycle representing the direct product and mapping to $\bar{0} \in \mathbb{Z}/p\mathbb{Z}$.

In conclusion, we describe the *group* of isomorphism classes of central extensions of $\mathbb{Z}_p$ by $\mathbb{Z}_p$ as follows: The identity element is the trivial extension

$$1 \to \mathbb{Z}_p \to \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p \to 1 \tag{15.163}$$

and then there is an orbit of $(p-1)$ nontrivial extensions of the form

$$1 \to \mathbb{Z}_p \to \mathbb{Z}_{p^2} \to \mathbb{Z}_p \to 1 \tag{15.164}$$

acted on by $\mathrm{Aut}(\mathbb{Z}_p) = \mathbb{Z}_p^*$.

Example 3:*Prime Powers.* Once we start to look at prime powers things start to get more complicated. We will content ourselves with extensions of $\mathbb{Z}_4$ by $\mathbb{Z}_2$. Here it can be shown that

$$H^2(\mathbb{Z}_4, \mathbb{Z}_2) \cong \mathbb{Z}_2 \tag{15.165}$$

so there should be two inequivalent extensions. One is the direct product and the other is

♣Do this more systematically so show that there are precisely two extensions of $\mathbb{Z}_4$ by $\mathbb{Z}_2$. ♣

$$1 \to \mathbb{Z}_2 \to \mathbb{Z}_8 \to \mathbb{Z}_4 \to 1 \tag{15.166}$$

We will think of these as multiplicative groups of roots of unity, with generators $\sigma = -1$ for $\mathbb{Z}_2$, $\alpha = \exp[2\pi i/8]$ for $\mathbb{Z}_8$, and $\omega = \exp[2\pi i/4]$ for $\mathbb{Z}_4$.

The inclusion map $\iota : \sigma \to \alpha^4$, while the projection map takes $\pi : \alpha \to \alpha^2 = \omega$.

Let us try to find a section. Since we want a normalized cocycle we must choose $s(1) = 1$. Now, $\pi(s(\omega)) = \omega$ implies $s(\omega)^2 = \omega$, and this equation has two solutions: $s(\omega) = \alpha$ or $s(\omega) = \alpha^5$. Let us choose $s(\omega) = \alpha$. (The following analysis for α^5 is similar.) If we try to make s into a homomorphism then we are forced to choose

$$\begin{aligned} s(\omega) &= \alpha \\ s(\omega^2) &= \alpha^2 \\ s(\omega^3) &= \alpha^3 \end{aligned} \tag{15.167}$$

but now we have no choice - we *must* set $s(\omega^4) = s(1) = 1$. On the other hand, if s *were* to have been a homomorphism we would have wanted to set $s(\omega^4) = s(\omega)^4 = \alpha^4$, but, as we just said, we cannot do this. With the above choice of section we get the symmetric cocycle whose nontrivial entries are

$$f(\omega, \omega^3) = f(\omega^2, \omega^2) = f(\omega^2, \omega^3) = f(\omega^3, \omega^3) = \alpha^4 = \sigma. \tag{15.168}$$

♣Probably should just describe all extensions with $Q = \mathbb{Z}_n$ ♣

Example 4.. *Products Of Cyclic Groups.* Another natural generalization is to consider products of cyclic groups. For simplicity we will only consider the case

$$G = \mathbb{Z}_p \oplus \cdots \oplus \mathbb{Z}_p \tag{15.169}$$

where there are k summands, but p is prime. We will think of our group additively and moreover we will think of $\mathbb{Z}_p$ as a ring in this example. If we write elements as $\vec{x} = (x_1, \ldots, x_k)$ with $x_i \in \mathbb{Z}_p$ and our cocycle $f(\vec{x}, \vec{y})$ is also valued in $\mathbb{Z}_p$, so that we are considering central extensions:

$$0 \to \mathbb{Z}_p \to \widetilde{G} \to \mathbb{Z}_p^{\oplus k} \to 0 \tag{15.170}$$

then the cocycle condition becomes:

$$f(\vec{x},\vec{y}) + f(\vec{x}+\vec{y},\vec{z}) = f(\vec{x},\vec{y}+\vec{z}) + f(\vec{y},\vec{z}) \tag{15.171}$$

An obvious way to satisfy this condition is to use a bilinear form:

$$f(\vec{x},\vec{y}) = A_{ij}x_i y_j \tag{15.172}$$

where the matrix elements $A_{ij} \in \mathbb{Z}_n$. We can modify by a coboundary:

$$f(\vec{x},\vec{y}) \to f(\vec{x},\vec{y}) + q(\vec{x}+\vec{y}) - q(\vec{x}) - q(\vec{y}) \tag{15.173}$$

Notice a linear term cancels out. It we want to restrict attention to expressions which are quadratic then we can modify

$$A_{ij} \to A_{ij} - (q_{ij} + q_{ji}) \tag{15.174}$$

where q_{ij} is any matrix with values in $\mathbb{Z}_p$.

Now we must distinguish the case $p = 2$ from p an odd prime. If $p = 2$ we can use the coboundary to make the off-diagonal part asymmetric, and WLOG we can agree that for each $i < j$ either $A_{ij} = A_{ji} = 0$ or $A_{ij} = 0$ and $A_{ji} = 1$. Note that the diagonal matrix elements are gauge invariant since $q_{ii} + q_{ii} = 2q_{ii} = 0$. Therefore we can produce in this way $\frac{1}{2}k(k+1)$ independent cocycles.

If p is an odd prime we can require that the matrix is "anti-symmetric" in the sense that $A_{ij} + A_{ji} = 0$ for all i, j, because 2 is invertible. In this way we only produce $\frac{1}{2}k(k-1)$ independent cocycles.

On the other hand, using methods of topology (See section **** below for hints) one can prove that

$$H^2(\mathbb{Z}_p^{\oplus k}, \mathbb{Z}_p) \cong \mathbb{Z}_p^{\frac{1}{2}k(k+1)} \tag{15.175}$$

for any prime p. What are the k "missing" cocycles for p an odd prime? They are exactly the extensions we discussed in detail in Example 2 above!

Example 5.. As a special case of the above, consider extensions of $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ by $\mathbb{Z}_2$. This will be a group of order 8. As we will see, there are five groups of order 8 up to isomorphism:

$$\mathbb{Z}_8, \quad \mathbb{Z}_2 \times \mathbb{Z}_4, \quad \mathbb{Z}_2^3, \quad Q, \quad D_4 \tag{15.176}$$

where Q and D_4 are the quaternion and dihedral groups, respectively. Now $\mathbb{Z}_8$ cannot sit in an extension of $\mathbb{Z}_2 \times \mathbb{Z}_2$. (Why not? [172]) This leaves 4 isomorphisms classes of groups which do fit in extensions of $\mathbb{Z}_2 \times \mathbb{Z}_2$ by $\mathbb{Z}_2$ and it happens they are all central extensions. They are:

$$\begin{aligned} &1 \to \mathbb{Z}_2 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to 1 \\ &1 \to \mathbb{Z}_2 \to \mathbb{Z}_2 \times \mathbb{Z}_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to 1 \\ &1 \to \mathbb{Z}_2 \to Q \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to 1 \\ &1 \to \mathbb{Z}_2 \to D_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to 1 \end{aligned} \tag{15.177}$$

[172] *Answer*: Because $\mathbb{Z}_2 \times \mathbb{Z}_2$ would have to be a quotient of $\mathbb{Z}_8$. But we can easily list the subgroups of $\mathbb{Z}_8$ and no quotient is of this form.

where Q is the quaternion group and D_4 the dihedral group. We have already met Q and D_4 above. One can define a homomorphism $\pi : Q \to \mathbb{Z}_2 \oplus \mathbb{Z}_2$ by

$$\begin{aligned} \pi(\pm 1) &= 0 = (0,0) \\ \pi(\pm \mathrm{i}\sigma^1) &= v_1 = (1,0) \\ \pi(\pm \mathrm{i}\sigma^2) &= v_2 = (0,1) \\ \pi(\pm \mathrm{i}\sigma^3) &= v_1 + v_2 = (1,1) \end{aligned} \tag{15.178}$$

where we are thinking of $\mathbb{Z}_2 \cong \mathbb{Z}/2\mathbb{Z}$ additively to make contact with the previous example. We can choose a section:

$$\begin{aligned} s(v_1) &= \mathrm{i}\sigma^1 \\ s(v_2) &= \mathrm{i}\sigma^2 \\ s(v_1+v_2) &= \mathrm{i}\sigma^3 \end{aligned} \tag{15.179}$$

and, computing the cocycle we find that it is given by the bilinear form (see the previous exercise):

$$A_Q = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{15.180}$$

Similarly, we can define a homomorphism $\pi : D_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2$ by

$$\begin{aligned} \pi(\pm 1) &= 0 = (0,0) \\ \pi(\pm R(\pi/2)) &= v_1 = (1,0) \\ \pi(\pm P) &= v_2 = (0,1) \\ \pi(\pm PR(\pi/2)) &= v_1 + v_2 = (1,1) \end{aligned} \tag{15.181}$$

We can choose a section:

$$\begin{aligned} s(v_1) &= R(\pi/2) \\ s(v_2) &= P \\ s(v_1+v_2) &= PR(\pi/2) \end{aligned} \tag{15.182}$$

and, computing the cocycle we find that it is given by the bilinear form (see the previous exercise):

$$A_{D_4} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \tag{15.183}$$

Now, on the other hand, using methods of topology one can prove that

$$H^2(\mathbb{Z}_2 \times \mathbb{Z}_2, \mathbb{Z}_2) = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \tag{15.184}$$

We can understand this group in terms of the bilinear quadratic forms mentioned in the previous example. Under $A_{ij} \to A_{ij} - (q_{ij} + q_{ji})$. Note that A_{11} and A_{22} are invariant, but we can modify the off-diagonal part of A_{ij} by a symmetric matrix. Thus, there are 8 possible values for $A_{11}, A_{22}, A_{12} \in \mathbb{Z}_2$.

In a way analogous to our discussion of extensions of $\mathbb{Z}_p$ by $\mathbb{Z}_p$, while there are only four different isomorphism classes of groups, there can be different extensions. An extension with group cocycle $A_{ij}x_1^i x_2^j$ defines a group of elements $(z, \vec{x})$. If we only care about the isomorphism class of the group we are free to consider an isomorphism

$$(z, \vec{x}) \mapsto (z, S\vec{x}) \tag{15.185}$$

where $S \in GL(2, \mathbb{Z}_2)$. This maps $A \to SAS^{tr}$. In general that will produce an isomorphic group, but a different extension.

♣Explain in detail how this takes us from 8 extensions to four isomorphism classes of groups using the explicit transforms by elements of $GL(2, \mathbb{Z}_2) \cong S_3$. ♣

In all our examples up to now the group $\tilde{G}$ has been Abelian, but in this example we have produced two nonisomorphic nonabelian groups Q and D_4 of order 8.

Example 5.. Nonabelian groups can also have central extensions. Indeed, we already saw this for $G = SO(3)$. Here is an example with G a nonabelian finite group. We take G to be the symmetric group S_n. It turns out that it has one nontrivial central extension by $\mathbb{Z}_2$:

$$H^2(S_n; \mathbb{Z}_2) \cong \mathbb{Z}_2 \tag{15.186}$$

To define it we let $\sigma_i = (i, i+1)$, $1 \le i \le n-1$ be the transpositions generating S_n. Then $\hat{S}_n$ is generated by $\hat{\sigma}_i$ and a central element z satisfying the relations:

♣Should add an exercise showing that $f(g,g)$ for g an involution determines the entire cocycle in this case. There are three nontrivial involutions, again giving 8 possible nonisomorphic cocycles. ♣

♣Should also consider central extensions of $\mathbb{Z}_2$ by $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. ♣

$$\begin{aligned} z^2 &= 1 \\ \hat{\sigma}_i^2 &= z \\ \hat{\sigma}_i \hat{\sigma}_{i+1} \hat{\sigma}_i &= \hat{\sigma}_{i+1} \hat{\sigma}_i \hat{\sigma}_{i+1} \\ \hat{\sigma}_i \hat{\sigma}_j &= z \hat{\sigma}_j \hat{\sigma}_i \qquad\qquad j > i+1 \end{aligned} \tag{15.187}$$

When restricted to the alternating group A_n we get an extension of A_n that can be elegantly described using spin groups.

Remarks:

1. One generally associates cohomology with the subject of topology. There is indeed a beautiful topological interpretation of group cohomology in terms of "classifying spaces."

2. In the case where G is itself abelian we can use more powerful methods of homological algebra to classify central extensions.

3. The special case $H^2(G, U(1))$ (or sometimes $H^2(G, \mathbb{C}^*)$, they are the same) is known as the *Schur multiplier*. It plays an important role in the study of projective representations of G. We will return to this important point.

4. We mentioned that a general extension (15.1) can be viewed as a principal N bundle over Q. Let us stress that trivialization of $\pi : G \to Q$ as a principal bundle is completely different from trivialization of the extension (by choosing a splitting).

These are different mathematical structures! For example, for finite groups the bundle is of course trivial because any global section is also continuous. However, as we have just seen the extensions might be nontrivial. It is true, quite generally, that if a central extension is trivial as a group extension then $\tilde{G} = A \times G$ and hence $\pi : \tilde{G} \to G$ is trivializable as an A-bundle.

♣In general a central extension by $U(1)$ is equivalent to a line bundle over the group and you should explain that here. ♣

Exercise

Suppose that the central extension (15.22) is equivalent to the trivial extension with $\tilde{G} = A \times G$, the direct product. Show that the possible splittings are in one-one correspondence with the set of group homomorphisms $\phi : G \to A$.

Exercise

Construct cocycles corresponding to each of the central extensions in (15.177) and show how the automorphisms of $\mathbb{Z}_2 \times \mathbb{Z}_2$ account for the the fact that there are only four entries in (15.177) while (15.184) is order 8.

Exercise *D_4 vs. Q*

a.) Show that D_4 and Q both fit in exact sequences

$$1 \to \mathbb{Z}_4 \to D_4 \to \mathbb{Z}_2 \to 1 \tag{15.188}$$

$$1 \to \mathbb{Z}_4 \to Q \to \mathbb{Z}_2 \to 1 \tag{15.189}$$

b.) Are these central extensions?

c.) Are D_4 and Q isomorphic? [173]

Exercise

Choosing the natural section $s : \sigma_i \to \hat{\sigma}_i$ in (15.187) and find the corresponding cocycle f_s.

---

[173] *Answer*: No. D_4 has 5 nontrivial involutions: The reflections in the four symmetry axes of the square and the rotation by π, while Q has only one nontrivial involution, namely -1.

Exercise *Due Diligence*

Show that the associative law for the twisted product (15.102) is equivalent to the cocycle condition on the 2-cochain f.

Exercise *Involution Criterion For A Nontrivial Cocycle*

Let g be a nontrvial involution. Show that the condition that $f(g,g)$ is, or is not, a perfect square is independent of which cocycle we use within a cohomology class.

Exercise *Group Commutator Criterion For A Nontrivial Cocycle*

a.) Show that if a central extension is defined by a cocycle f then the group commutator is:

$$[(a_1,g_1),(a_2,g_2)] = \left(\frac{f(g_1g_2,g_1^{-1}g_2^{-1})}{f(g_2g_1,g_1^{-1}g_2^{-1})}\frac{f(g_1,g_2)}{f(g_2,g_1)}, g_1g_2g_1^{-1}g_2^{-1}\right) \tag{15.190}$$

b.) Suppose G is abelian. Show that $\tilde{G}$ is abelian iff $f(g_1,g_2)$ is symmetric.

c.) In general the condition that f is symmetric: $f(g_1,g_2) = f(g_2,g_1)$ would not be preserved by a coboundary transformation. Show that it does make sense in this setting.

d.) Suppose $\tilde{G}$ is a central extension of a not-necessarily-Abelian group G by an Abelian group A. Show that if (g_1,g_2) is a commuting pair of elements in G and if $f(g_1,g_2)/f(g_2,g_1)$ is not the identity then the extension is nontrivial. [174]

15.4 Extended Example: Charged Particle On A Circle Surrounding A Solenoid

In the following extended example we will illustrate how classical symmetries can be centrally extended in the context of a very interesting quantum system. Along the way we will take the opportunity to introduce many ideas about quantum field theory in a very simple context.

15.4.1 Hamiltonian Analysis

Consider a particle of mass m confined to a ring of radius r in the xy plane. The position of the particle is described by an angle ϕ, so we identify $\phi \sim \phi + 2\pi$, and the action is

$$S = \int \frac{1}{2}mr^2\dot{\phi}^2 = \int \frac{1}{2}I\dot{\phi}^2 \tag{15.191}$$

[174] *Answer*: This follows because, as we saw above, a split central extension is a direct product. But the group commutator of $(1,g_1)$ and $(1,g_2)$ must then be the identity. On the other hand, $f(g_1,g_2)/f(g_2,g_1)$ is gauge invariant, so if it is nontrivial then the group commutator cannot be the identity.

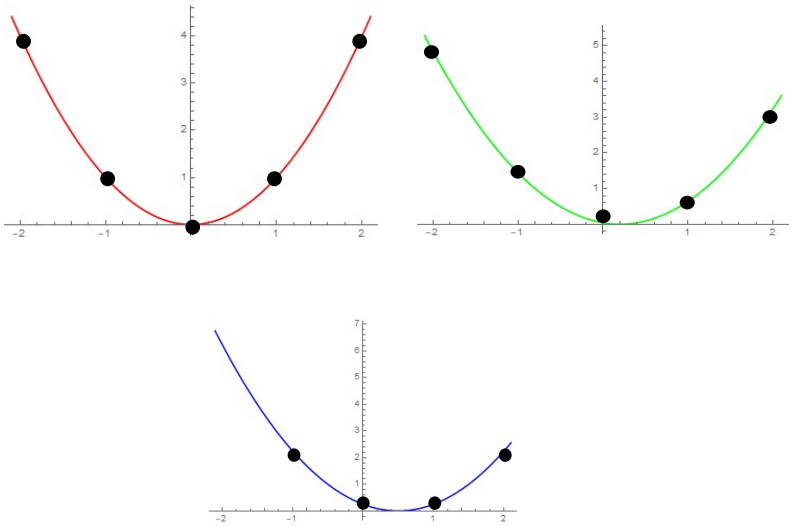


Figure 38: Spectrum of a particle on a circle as a function of $\mathcal{B} = eB/2\pi$. The upper left shows the low-lying spectrum for $\mathcal{B} = 0$. It is symmetric under $m \to -m$. The upper right shows the spectrum for $\mathcal{B} = 0.2$. There is no symmetry in the spectrum. The lower figure shows the spectrum for $\mathcal{B} = 1/2$. There is again a symmetry, but under $m \to 2\mathcal{B} - m = 1 - m$. In general there will be no symmetry unless $2\mathcal{B} \in \mathbb{Z}$. If $2\mathcal{B} \in \mathbb{Z}$ the spectrum is symmetric under $m \to 2\mathcal{B} - m$.

with $I = mr^2$ the moment of inertia.

Let us also suppose that our particle has electric charge e and that the ring is threaded by a solenoid with magnetic field B, so the particle moves in a zero B field, but there is a nonzero gauge potential [175]

$$A = \frac{B}{2\pi} d\phi \tag{15.192}$$

The action is therefore:

$$\begin{aligned} S &= \int \frac{1}{2} I \dot{\phi}^2 dt + \oint eA \\ &= \int \frac{1}{2} I \dot{\phi}^2 dt + \frac{eB}{2\pi} \dot{\phi} dt \end{aligned} \tag{15.193}$$

The second term is an example of a "topological term" or a "θ-term." Classically, the second term does not affect physical predictions, since it is a total derivative. However as we will soon see, quantum mechanically, it will have an important effect on physical predictions.

[175] For readers not familiar with differential form notation this means, in cylindrical coordinates that $A_z = 0$, $A_r = 0$ and $A_\phi = B/2\pi$.

We are going to analyze the symmetries of this system and compare their realization in the classical and quantum theories.

Classical Symmetries:

We begin by analyzing the classical symmetries. Because the θ-term does not affect the classical dynamics the classical system has $O(2)$ symmetry. We can rotate: $R(\alpha)$: $e^{i\phi} \to e^{i\alpha}e^{i\phi}$, or, if you prefer, translate $\phi \to \phi + \alpha$ (always bearing in mind that α and ϕ are only defined modulo addition of an integral multiple of 2π). If we think of the circle in the $x - y$ plane centered on the origin, with the solenoid along the z-axis then we could also take as usual:

$$R(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} . \tag{15.194}$$

Also we can make a "parity" or "charge conjugation" transformation $P : \phi \to -\phi$. The second term in the Lagrangian is not invariant but this "doesn't matter" because it is a total derivative. Put differently: $\phi \to -\phi$ is a symmetry of the equations of motion, and hence it is a classical symmetry.

Note that these group elements in $O(2)$ satisfy

$$\begin{aligned} R(\alpha)R(\beta) &= R(\alpha+\beta) \\ P^2 &= 1 \\ PR(\alpha)P &= R(-\alpha) \end{aligned} \tag{15.195}$$

and indeed, as we have seen, $O(2)$ is a semidirect product:

$$O(2) = SO(2) \rtimes \mathbb{Z}_2 \tag{15.196}$$

with $\omega : \langle P \rangle \cong \mathbb{Z}_2 \to \mathrm{Aut}(SO(2)) \cong \mathbb{Z}_2$ acting by taking the nontrivial element of $\mathbb{Z}_2$ to the outer automorphism that sends $R(\alpha) \to R(-\alpha)$.

Diagonalizing The Hamiltonian

Now let us consider the quantum mechanics with the "θ-term" added to the Lagrangian. Our goal is to see how that term affects the quantum theory.

We will first analyze the quantum mechanics in the Hamiltonian approach. See the remark below for some remarks on the path integral approach. The conjugate momentum is

$$L = I\dot{\phi} + \frac{eB}{2\pi} \tag{15.197}$$

We denote it by L because it can be thought of as angular momentum.

Note that the coupling to the flat gauge field has altered the usual relation of angular momentum and velocity. Now we obtain the Hamiltonian from the Legendre transform:

$$\int L\dot{\phi}dt - S = \int \frac{1}{2I}(L - \frac{eB}{2\pi})^2 dt \tag{15.198}$$

Upon quantization $L \to -\mathrm{i}\hbar\frac{\partial}{\partial\phi}$, so the Hamiltonian is

$$H_{\mathcal{B}} := \frac{\hbar^2}{2I}\left(-\mathrm{i}\frac{\partial}{\partial\phi} - \mathcal{B}\right)^2 \tag{15.199}$$

where $\mathcal{B} := \frac{eB}{2\pi\hbar}$.

The eigenfunctions of the Hamiltonian $H_{\mathcal{B}}$ are just

$$\Psi_m(\phi) = \frac{1}{\sqrt{2\pi}}e^{\mathrm{i}m\phi} \qquad m \in \mathbb{Z} \tag{15.200}$$

They give energy eigenstates with energy

$$E_m = \frac{\hbar^2}{2I}(m - \mathcal{B})^2 \tag{15.201}$$

There is just one energy eigenstate for each $m \in \mathbb{Z}$.

Before moving on with the analysis of the symmetries in this quantum mechanical problem let us take the opportunity to make a long list of:

Remarks:

1. The action (15.193) makes good sense for ϕ valued in the real line or for $\phi \sim \phi + 2\pi$, valued in the circle. Making this choice is important in the choice of what theory we are describing. Where - in the above analysis - did we make the choice that the target space is a circle? [176]

2. Taking $\phi \sim \phi + 2\pi$, even though the θ-term is a total derivative it has a nontrivial effect on the quantum physics as we can see since B has shifted the spectrum of the quantum Hamiltonian in a physically observable fashion: *This is how we see that topological terms matter.*

3. Note that when $2\mathcal{B}$ is even the energy eigenspaces are two-fold degenerate, except for the ground state at $m = \mathcal{B}$. On the other hand, when $2\mathcal{B}$ is odd all the energy eigenspaces are two-fold degenerate, including the ground state. If $2\mathcal{B}$ is not an integer all the energy eigenspaces are one-dimensional. See Figure 38.

4. The total spectrum is *periodic* in $\mathcal{B}$, and shifting $\mathcal{B} \to \mathcal{B}+1$ is equivalent to $m \to m+1$. To be more precise, we can define a unitary operator on the Hilbert space by its action on a basis:

   $$U\Psi_m = \Psi_{m+1} \tag{15.202}$$

   and

   $$UH_{\mathcal{B}}U^{-1} = H_{\mathcal{B}+1} \tag{15.203}$$

[176] *Answer*: If we took the case where ϕ is valued in $\mathbb{R}$ and not the circle then there would be no quantization on m and the spectrum of the Hamiltonian would be continuous. In this case the Chern-Simons term would not affect the physics in the quantum mechanical version as well.

5. The quantum mechanics problem (15.193) and the spectrum (15.201) arise in the discussion of the "Coulomb blockade" in physics of quantum dots. See Yoshimasa Murayama, *Mesoscopic Systems*, Section 10.10.

6. *Viewing the system as a field theory.* We have introduced this system as describing the quantum mechanics of a particle. However, it is important to note that it can also be viewed as a special case of a quantum field theory. In general, in a field theory [177] we have a spacetime M and the fields ϕ are functions on M valued in some *target space* $\mathcal{X}$. (So the term "target space" means nothing more or less than the codomain of the fields.) An important example is that of a nonrelativistic particle of mass m moving on a Riemannian manifold X with metric $ds^2 = g_{\mu\nu}(x)dx^\mu \otimes dx^\nu$. The action would be

$$S = \int dt \frac{m}{2} g_{\mu\nu}(x(t))\dot{x}^\mu \dot{x}^\nu dt \tag{15.204}$$

If, in addition, the particle has charge e and there is an electromagnetic potential $A_\mu(x)dx^\mu$ on X then the action is

$$S = \int dt \frac{m}{2} g_{\mu\nu}(x(t))\dot{x}^\mu \dot{x}^\nu dt + \int e A_\mu(x(t))\dot{x}^\mu dt \tag{15.205}$$

Here M is the manifold of time. It could be $M = \mathbb{R}$ if we describe the entire history of the particle, or $M = [t_{in}, t_{fin}]$ if we describe only the motion in a finite time interval. As we will soon see, it can also be interesting to let $M = S^1$. The "field" is a suitably differentiable map

$$x : M \to X \tag{15.206}$$

describing the position of the particle as a function of time. This is an example of a "$0+1$ dimensional field theory." A generalization would be a theory of maps from a d-dimensional spacetime with metric $h_{ab}d\sigma^a d\sigma^b$ and action

$$S = \int d^{d+1}\sigma \sqrt{|\det h|} h^{ab}(\sigma) \frac{1}{2} m g_{\mu\nu}(x(t))\partial_a x^\mu \partial_b x^\nu \tag{15.207}$$

and the "field" would be a suitably differentiable map:

$$x : M \to X \tag{15.208}$$

Equations (15.205) and (15.207) are examples of what is known as a "nonlinear sigma model." [178] In our case our fields are maps

$$e^{i\phi} : M \to S^1 \tag{15.212}$$

[177] As traditionally conceived. The topic of topological field theory generalizes the next few lines considerably.

[178] For the mathematically sophisticated reader we note that, for general nonlinear sigma models

$$dx : T_\sigma M \to T_{x(\sigma)} X \tag{15.209}$$

is a linear map between two inner-product spaces. We can use the inner products to define $(dx)^\dagger$ and then the kinetic term is

$$\int_M \mathrm{Tr}((dx)(dx)^\dagger)\mathrm{vol}\,(h) \tag{15.210}$$

We have been referring to $\phi \to -\phi$ as "parity" because that is the appropriate term in the context of the quantum mechanics of a particle constrained to a circle in the plane. The parity operation is just reflection around some line in the plane. However, if we take the point of view that we are discussing a $0+1$ dimensional "field theory" then it would be better to refer to the operation as "charge conjugation" because it complex conjugates the $U(1)$-valued field $e^{\mathrm{i}\phi}$.

In addition there are (in the field theory interpretation) "worldvolume symmetries" of time translation invariance and time reversal. These form the group $\mathbb{R} \rtimes \mathbb{Z}_2$. We will put those aside. (Note that time reversal is not a symmetry of the second term in the Lagrangian but is a symmetry of the space of solutions of the equations of motion.)

7. *Relations to higher dimensional field theories and string theory.* The θ-term we have added has a very interesting analog in $1+1$ dimensional field theory, where it is known as a coupling to the B-field. It can also be obtained from a Kaluza-Klein reduction of $1+1$ dimensional Maxwell theory:

$$\begin{aligned} S &= \frac{1}{e^2}\int dx^0 dx^1 F_{01}^2 + \int \frac{\theta}{2\pi} F_{01} dx^0 \wedge dx^1 \\ &= \frac{1}{e^2}\int F * F + \int \frac{\theta}{2\pi} F \end{aligned} \tag{15.213}$$

In $1+1$ dimensional theory we can choose $A_0 = 0$ gauge and gauge away the x^1 dependence so that on $S^1 \times \mathbb{R}$ the only gauge invariant quantity is

$$e^{\mathrm{i}\phi(t)} = e^{\mathrm{i}\oint_{S^1} A} = e^{\mathrm{i}\oint_{S^1} A_1 dx^1} \tag{15.214}$$

With this in mind we can say

$$\theta = 2\pi\mathcal{B} \tag{15.215}$$

Remark: More generally, in $1+1$ dimensional Yang-Mills theory on $S^1 \times \mathbb{R}$ we can always go to $A_0 = 0$ gauge and then the only gauge invariant observable is the conjugacy class of the holonomy around the circle.

The theta term also has a close analog in $3+1$-dimensional gauge theory. In the case of $3+1$ dimensional Maxwell theory we can write

$$\begin{aligned} S &= \int d^4x \frac{1}{4e^2} F_{\mu\nu}F^{\mu\nu} + \int \frac{\theta}{8\pi}\epsilon^{\mu\nu\lambda\rho}F_{\mu\nu}F_{\lambda\rho}d^4x \\ &= \int \frac{1}{2e^2} F * F + \int \frac{\theta}{(4\pi)} F \wedge F \end{aligned} \tag{15.216}$$

More to the point dx is a section of $(TM)^\vee \otimes f^*(TX)$ and we can use the metric on this bundle to write $\int_M \parallel dx \parallel^2$. In the case of the charged particle moving on the Riemannian manifold X, there is also the data of a principal $U(1)$ bundle with connection $d + A$ and the topological term is based on the holonomy of the pulled back connection:

$$S = \int \frac{m}{2} \parallel dx \parallel^2 dt + \oint ex^*(A) \tag{15.211}$$

There are similar topological terms for the $d > 0$ sigma models.

In fact, in the effective theory of electromagnetism in the presence of an insulator a very similar action arises with a θ term. If a parity- and/or time-reversal symmetry is present then θ is zero or π, corresponding to our case $2\mathcal{B} \in \mathbb{Z}$. The difference between a normal and a topological insulator is then, literally, the difference between $2\mathcal{B}$ being even (normal) and odd (topological), respectively. Finally, in the 3+1-dimensional Yang-Mills theories that describe the standard model of electro-weak and strong interactions one can add an analogous θ-term. Topological terms matter, and in this case the topological term for the strong gauge field leads to the prediction of an intrinsic electric dipole moment of the neutron. However, to excellent accuracy it is known that if the neutron dipole moment it is very small and

$$|\theta| < 10^{-9} \tag{15.217}$$

One of the great unsolved mysteries about nature is why the (effective) theta angle for the strong interactions in the standard model is so small. [179]

Now let us get back to the symmetries of the particle on the ring. We have seen that the classical "internal" symmetry group - the "internal" symmetry group of the equations of motion - is $O(2)$. Now let us analyze how the symmetries are implemented in the quantum theory:

In quantum mechanics the $SO(2)$ shift symmetry $\phi \to \phi + \alpha$ is realized by a translation operator $\rho(R(\alpha)) = \mathcal{R}(\alpha)$ and acting on Ψ_m we have

$$(\mathcal{R}(\alpha) \cdot \Psi_m) = e^{\mathrm{i}m\alpha}\Psi_m \tag{15.218}$$

Can we also represent $\rho(P) = \mathcal{P}$ on the Hilbert space? Classically, parity symmetry P just takes $\phi \to -\phi$. If we make this substitution in the Hamiltonian $H_{\mathcal{B}}$ we see that the naive parity operation takes

$$\mathcal{P} H_{\mathcal{B}} \mathcal{P}^{-1} = H_{-\mathcal{B}} \tag{15.219}$$

For general values of $\mathcal{B}$ the operator $H_{\mathcal{B}}$ is not unitarily equivalent to $H_{-\mathcal{B}}$. However, thanks to (15.203) it is clear that when $2\mathcal{B} \in \mathbb{Z}$ they are unitarily equivalent and the naive operation of taking $\phi \to -\phi$, which takes $m \to -m$ on eigenvectors of $H_{\mathcal{B}}$, should be accompanied by $U^{2\mathcal{B}}$. Therefore $\mathcal{P}$ should map the eigenspace associated with m to that associated with with $2\mathcal{B} - m$. As a sanity check note that indeed $E_m = E_{2\mathcal{B}-m}$. Therefore we should define a parity operation:

$$\mathcal{P} \cdot \Psi_m = \xi_m \Psi_{2\mathcal{B}-m} \tag{15.220}$$

where ξ_m is a phase which we can take to be 1. Note that the operator $\mathcal{P}$ so defined commutes with the Hamiltonian: Indeed, it takes eigenvectors to eigenvectors with the same eigenvalue.

♣You should allow the possibility of a phase in the definition of $\mathcal{P}$ and show in detail it doesn't matter. ♣

If $2\mathcal{B}$ is not an integer the parity symmetry is broken and the quantum symmetry group is just $SO(2)$.

[179] For much more about this see M. Dine's TASI lectures https://arxiv.org/pdf/hep-ph/0011376.pdf.

Now consider the case when $2\mathcal{B} \in \mathbb{Z}$ and let us study the relations obeyed by the operators $\mathcal{R}(\alpha)$ and $\mathcal{P}$ and compare them with the classical relations (15.195). We still have $\mathcal{R}(\alpha)\mathcal{R}(\beta) = \mathcal{R}(\alpha+\beta)$ and $\mathcal{P}^2 = 1$ but now the third line of (15.195) is modified to:

$$\mathcal{P}\mathcal{R}(\alpha)\mathcal{P} = e^{\mathrm{i}2\mathcal{B}\alpha}\mathcal{R}(-\alpha) \tag{15.221}$$

Exercise *Due Diligence*

a.) Check that (15.221) is well-defined, even though α is only defined up to a shift by an integral multiple of 2π.

b.) Check the operator relation (15.221)!

We now consider the group of operators generated by the operators $\mathcal{P}$, $\mathcal{R}(\alpha)$, and $z1_{\mathcal{H}}$ where $z \in U(1)$. (Do not forget that we identify $\alpha \sim \alpha + 2\pi$. This will be quite important in what follows.) Denote this group of operators by $\mathcal{G}_{\mathcal{B}}$. Naively we might have expected this group of operators on Hilbert space to be isomorphic to $U(1) \times O(2)$ where $O(2)$ is our classical symmetry group and $U(1)$ is just the group of phases acting on wavefunctions by scalar multiplication. However, equation (15.221) is not satisfied by a direct product. So, how is $\mathcal{G}_{\mathcal{B}}$ related to $U(1)$ and $O(2)$? General principles tell us it will be an extension

$$1 \longrightarrow U(1) \longrightarrow \mathcal{G}_B \longrightarrow O(2) \longrightarrow 1 \tag{15.222}$$

But what extension?

Now, when $\mathcal{B}$ is an integer we can indeed define an isomorphism of $\mathcal{G}_{\mathcal{B}}$ with $U(1) \times O(2)$ by setting

$$\tilde{\mathcal{R}}(\alpha) := e^{-\mathrm{i}\mathcal{B}\alpha}\mathcal{R}(\alpha) \tag{15.223}$$

We now recover the standard relations of $O(2)$, so the classical $O(2)$ symmetry is not modified quantum mechanically. However, when $\mathcal{B}$ is a half-integer, $\tilde{\mathcal{R}}$ is not well-defined since we must identify $\alpha \sim \alpha + 2\pi$. In this case the group $\mathcal{G}_{\mathcal{B}}$ is really different.

To understand what happens when $\mathcal{B}$ is half-integral we introduce a new group called Spin(2). As an abstract group it is isomorphic to $SO(2)$, and $U(1)$, and $\mathbb{R}/\mathbb{Z}$. The groups are all isomorphic. What makes Spin(2) nontrivial is its relation to $SO(2)$. The group elements in Spin(2) can be parametrized by $\hat{\alpha}$ with $\hat{\alpha} \sim \hat{\alpha} + 2\pi$. Let us call the elements of the spin group $\hat{R}(\hat{\alpha})$. You can think of it in terms of Pauli matrices as

$$\hat{R}(\hat{\alpha}) = \exp[\hat{\alpha}\sigma^1\sigma^2] = \cos(\hat{\alpha}) + \mathrm{i}\sin(\hat{\alpha})\sigma^3 \tag{15.224}$$

But it is called the *spin group* because it comes with a nontrivial double cover:

$$\pi : \mathrm{Spin}(2) \to SO(2) \tag{15.225}$$

the double covering is given by restricting our standard projection $\pi : SU(2) \to SO(3)$ to the subgroup of $SU(2)$ in (15.224). In this way we get a double cover of the rotation group around the z axis:

$$\pi : \hat{R}(\hat{\alpha}) \mapsto R(2\hat{\alpha}) \tag{15.226}$$

See equation (15.48) above.

Now, taking $\mathbb{Z}_2$ to act on Spin(2) by the nontrivial outer automorphism. So, denoting the nontrivial element of $\mathbb{Z}_2$ by $\hat{P}$ we use the homomorphism $\alpha : \mathbb{Z}_2 \to \mathrm{Aut}(\mathrm{Spin}(2))$ defined by

$$\alpha(\hat{P}) : \hat{R}(\hat{\alpha}) \to (R(\hat{\alpha}))^{-1} = \hat{R}(-\hat{\alpha}) \tag{15.227}$$

Then, one definition of the group $\mathrm{Pin}^+(2)$ is that it is the semidirect product:

$$\mathrm{Pin}^+(2) \cong \mathrm{Spin}(2) \rtimes_\alpha \mathbb{Z}_2 \tag{15.228}$$

(We will give a slightly different definition below.) There is a generalization of Spin and Pin groups to higher dimensions. They double cover $SO(d)$ and $O(d)$, respectively. See the remark below for a brief description and Chapters *** and **** for full details.

Now, when $2\mathcal{B}$ is an odd integer the group $\mathcal{G}_B$ is generated by

$$\begin{aligned} z &\mathbf{1}_{\mathcal{H}} \\ \rho(\hat{R}(\hat{\alpha})) &:= e^{-\mathrm{i}(2\mathcal{B})\hat{\alpha}}\mathcal{R}(2\hat{\alpha}) \qquad\qquad 0 \le \hat{\alpha} < 2\pi \\ \rho(\hat{P}) &:= \mathcal{P} \end{aligned} \tag{15.229}$$

where we take $\hat{P}$ to be the nontrivial element in $\mathbb{Z}_2$ in the semidirect product that defines $\mathrm{Pin}^+(2)$, so that $\hat{R}(\hat{\alpha})$ and $\hat{P}$ generate $\mathrm{Pin}^+(2)$. One checks that ρ is a homomorphism and the image under ρ is an isomorphic copy of $\mathrm{Pin}^+(2)$ inside $\mathcal{G}_B$, and we have:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathbb{Z}_2 & \longrightarrow & \mathrm{Pin}^+(2) & \longrightarrow & O(2) & \longrightarrow & 1 \\ & & \downarrow & & \downarrow{\scriptstyle \rho} & & \downarrow{\scriptstyle Id} & & \\ 1 & \longrightarrow & U(1) & \longrightarrow & \mathcal{G}_B & \longrightarrow & O(2) & \longrightarrow & 1 \end{array} \tag{15.230}$$

where $\mathbb{Z}_2 \cong \{\pm 1\} \subset U(1)$. When $2\mathcal{B}$ is odd ρ has no kernel. (When $2\mathcal{B}$ is even there is a kernel.)

In conclusion:

1. The classical theory has an $O(2)$ symmetry.

2. In the quantum theory when $2\mathcal{B}$ is not an integer the symmetry is broken to $SO(2)$.

3. In the quantum theory, when $2\mathcal{B}$ is an even integer the theory still has $O(2)$ symmetry. The sequence
$$1 \longrightarrow U(1) \longrightarrow \mathcal{G}_B \longrightarrow O(2) \longrightarrow 1 \tag{15.231}$$
splits and $\mathcal{G}_B \cong U(1) \times O(2)$.

4. In the quantum theory, when $2\mathcal{B}$ is an odd integer,
$$1 \longrightarrow U(1) \longrightarrow \mathcal{G}_B \longrightarrow O(2) \longrightarrow 1 \tag{15.232}$$
does not split. When we try to realize the classical $O(2)$ symmetry on the Hilbert space we are forced to implement the pin double cover $\mathrm{Pin}^+(2)$, a central extension of $O(2)$ by $\mathbb{Z}_2$. It is related to $\mathcal{G}_{\mathcal{B}}$ as in (15.230).

We conclude with some remarks:

1. We stress that the particle we put on the ring did NOT have any intrinsic spin!! Having said that, if we define an angular momentum $\mathcal{L}$ so that $H = \frac{\mathcal{L}^2}{2I}$ then indeed when $\mathcal{B}$ is half-integral the angular momentum has half-integral eigenvalues, as one expects for a spin representation. So, what we are finding is that the half flux quantum is inducing a half-integral spin of the system so that the classical $O(2)$ symmetry of the classical system is implemented as a $\mathrm{Pin}^+(2)$ symmetry in the quantum theory. This is an intriguing phenomenon appearing in quantum symmetries with nontrivial gauge fields and topological terms: The statistics and spins of particles can be shifted from their classical values, often in ways that involve curious fractions.

2. *Spin And Pin Groups.* Enquiring minds will wonder about the definition of $\mathrm{Pin}^{\pm}(d)$. These groups are defined using Clifford algebras. For much more detail and motivation see the two chapters on Clifford algebras and Spin groups. In brief, consider the Clifford algebra generated by $\{\gamma_i, \gamma_j\} = 2Q_{ij}$ where Q_{ij} is an invertible $d \times d$ symmetric matrix. For a vector v^i define $\gamma(v) := v^i\gamma_i$. Assume that $Q_{ij} = \delta_{ij}$. Then $\mathrm{Pin}^+(d)$ is the group of expressions of the form
$$\pm\gamma(v_1)\cdots\gamma(v_r) \tag{15.233}$$
for some r. The group $\mathrm{Pin}^-(d)$ is similarly defined with $Q_{ij} = -\delta_{ij}$. The group $\mathrm{Spin}(d)$ is the subgroup of such expressions where r is even. The projection $\pi : \mathrm{Pin}^{\pm}(d) \to O(d)$ is defined by the equation:
$$\gamma(\pi(g)\cdot w) = (-1)^r g\gamma(w)g^{-1} \tag{15.234}$$
The key idea here is that
$$-\gamma(v)\gamma(w)\gamma(v)^{-1} = \gamma(R_v(w)) \tag{15.235}$$
where $R_v(w)$ is the reflection of w through the plane orthogonal to v, as the reader can easily check in an exercise below. Then use the fact that all elements of $O(d)$ are products of reflections. The restriction to $\mathrm{Spin}(d)$ defines $\pi : \mathrm{Spin}(d) \to SO(d)$. This is a generalization of our standard double-covering $\pi : SU(2) \to SO(3)$. Although $\mathrm{Spin}(3) \cong SU(2)$ for $d > 3$ $\mathrm{Spin}(d)$ is not isomorphic to a unitary or orthogonal group. The difference between $\mathrm{Pin}^+(d)$ and $\mathrm{Pin}^-(d)$ is whether the lift of a reflection will square to $+1$ or -1 respectively. As a group $\mathrm{Pin}^-(d)$ is isomorphic to $(\mathrm{Spin}(2) \rtimes \mathbb{Z}_4)/\mathbb{Z}_2$.

3. It is instructive to study the representation of $\mathrm{Pin}^+(2)$ on the two-dimensional space of ground states, $\mathcal{H}_{\mathrm{grnd}}$, when $\mathcal{B} = 1/2$. In this case we can choose the ordered basis $\{\Psi_0, \Psi_1\}$ for $\mathcal{H}_{\mathrm{grnd}}$, and, relative to this basis we have a matrix representation:
$$\begin{aligned}\rho(\widehat{R}(\hat{\alpha}))|_{\mathcal{H}_{\mathrm{grnd}}} &= \begin{pmatrix} e^{-\mathrm{i}\hat{\alpha}} & 0 \\ 0 & e^{\mathrm{i}\hat{\alpha}} \end{pmatrix} \\ \rho(\widehat{P})|_{\mathcal{H}_{\mathrm{grnd}}} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\end{aligned} \tag{15.236}$$

We stress the appearance of $\hat{\alpha}$ in the representation matrix. This transformation corresponds to a translation of ϕ by $\alpha = 2\hat{\alpha}$. Had we tried to express the representation in terms of α we would encounter the phase $e^{\pm i\alpha/2}$ which is not well-defined because α is only defined modulo $\alpha \sim \alpha + 2\pi$.

4. As pointed out in a recent paper [180] this extension of the symmetry group at half-integral θ is an excellent baby model for how one can learn about nontrivial dynamics of quantum systems (in particular, QCD) by thinking carefully about group extensions. For example, if we were to add a potential $U(\phi)$ to the problem we just discussed we could no longer solve exactly for the eigenstates. Also, a generic potential would be of the form

$$U^{generic}(\phi) = \sum_{n\in\mathbb{Z}} c_n \cos(n\phi) + \sum_{n\in\mathbb{Z}} s_n \sin(n\phi) \tag{15.237}$$

Potentials with generic coefficients will explicitly break all of the $O(2)$ symmetry. Suppose however, that we can restrict attention to a special class of potentials with only cosine Fourier coefficients that are 0 mod 2:

$$U^{special}(\phi) = \sum_{n} u_n \cos(2n\phi) \tag{15.238}$$

Then, even though we cannot solve the spectrum of the Hamiltonian exactly we can make an interesting statement about it. For such potentials the classical $O(2)$ symmetry is explicitly broken to $\mathbb{Z}_2 \times \mathbb{Z}_2$ generated by $P : \phi \to -\phi$ and $r : \phi \to \phi + \pi$. We have shown that when $2\mathcal{B}$ is odd and the potential is zero the $O(2)$ symmetry is centrally extended and realized as the double-cover $\mathrm{Pin}^+(2)$ on the Hilbert space. The double cover of the subgroup group $\langle P, r\rangle \subset O(2)$ acting on Hilbert space is described by the pullback diagram:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathbb{Z}_2 & \xrightarrow{\iota} & D_4 & \xrightarrow{\pi_1} & \mathbb{Z}_2 \times \mathbb{Z}_2 & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle \mathrm{Id}} & & \downarrow{\scriptstyle \iota} & & \downarrow{\scriptstyle \iota} & & \\ 1 & \longrightarrow & \mathbb{Z}_2 & \longrightarrow & \mathrm{Pin}^+(2) & \xrightarrow{\pi_2} & O(2) & \longrightarrow & 1 \end{array} \tag{15.239}$$

To check this note that P lifts to the operators $\pm\mathcal{P}$ and $r = R(\pi)$ lifts to the operators

$$\pm\hat{r} = e^{i\mathcal{B}\pi}\mathcal{R}(\pi) \tag{15.240}$$

One checks that $\langle \mathcal{P}, \hat{r}\rangle$ generates a group isomorphic to D_4. Indeed, these operators satisfy the defining relations of D_4: $\mathcal{P}^2 = 1$, $\hat{r}^4 = 1$, and $\mathcal{P}\hat{r}\mathcal{P} = \hat{r}^{-1}$. Note that, in addition $\hat{r}^2 = -1$ on the entire Hilbert space. The representation on the Qbit groundstate in (15.236) (with $\hat{\alpha} = \pm\pi/2$) is a two-dimensional irrep of D_4. In fact all the doubly degenerate energy eigenspaces are two-dimensional irreps of D_4.

[180] D. Gaiotto, A. Kapustin, Z. Komargodski and N. Seiberg, "Theta, Time Reversal, and Temperature," https://arxiv.org/pdf/1703.00501.pdf

It is reasonable to assume that when we turn on a weak potential of the form (15.238) the classically preserved $\mathbb{Z}_2 \times \mathbb{Z}_2$ subgroup again lifts to a D_4 action on the Hilbert space, even though we can no longer construct the operators in the D_4 group explicitly. The cocycle is discrete: So if it is a continuous function of the parameters u_n at $u_n = 0$ (this is an assumption) then the classical $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry must be realized by D_4 on the Hilbert space.

Now, we saw that, in the absence of the potential, there is a Qbit giving a two-dimensional representation of $\mathrm{Pin}^+(2)$. This representation restricts to an irreducible two-dimensional representation of D_4. (See equation (15.236) with $\hat{\alpha} = \pm\pi/2$.) Now, D_4 has four one-dimensional irreducible representations $\mathbf{1}_{\pm,\pm}$ and, we will show later, exactly one two-dimensional irreducible representation. In particular, the set of representations of D_4 is discrete. Again, it is reasonable to suppose that the representation is a continuous function of u_n. Again, this is an assumption. But granting this, turning on a weak potential cannot change the decomposition of the energy eigenspaces into irreducible representations. This leads to a striking prediction: The two-fold groundstate degeneracy is not broken by potentials of the form (15.238) when $2\mathcal{B}$ is odd! This is remarkable when one compares to the standard discussion of the double-well potential of one-dimensional quantum mechanics. In that standard case one has a two-fold classical degeneracy broken by tunneling (instanton) effects so that there is a unique ground state. For potentials of the form (15.238) there are (generically) four stationary points of the potential, at $\phi = 0, \pm\pi/2, \pi$. Generically, two will be maxima and two will be minima. So, classically, and perturbatively in quantum mechanics, for a generic potential of the form (15.238) there will be a two-fold degenerate groundstate. However, unlike the textbook discussion of the double-well potential, the degeneracy will not be lifted by nonperturbative tunneling effects.

Exercise

Show that the ground state energy is

$$E_{\text{ground}} = Min_{m\in\mathbb{Z}} \frac{\hbar^2}{2I}(m - \mathcal{B})^2 \tag{15.241}$$

and, using the floor function, give a formula for E_{ground} directly in terms of $\mathcal{B}$ (without requiring minimization).

Exercise *Pin Action*

a.) Show that (15.234) defines a homomorphism to $O(d)$. [181]

[181] *Answer*: The key equation to check is (15.235). To check this consider the two cases that w is parallel to v and that w is perpendicular to v. Then note that every element in $O(d)$ is a product of reflections.

b.) Show that the general definition of $\mathrm{Pin}^+(d)$ specializes to the definition of $\mathrm{Pin}^+(2)$ as a semidirect product. [182]

Exercise *One-Dimensional Representations Of D_4*

Show that there are four distinct one-dimensional representations of D_4. [183]

Exercise *A Cocycle Puzzle*

Note that had we defined $\mathcal{P}$ with an extra factor of i we would have concluded that it is order 4, not order 2. Now, we know that the sequence $1 \to \mathbb{Z}_2 \to \mathbb{Z}_4 \to \mathbb{Z}_2 \to 1$ is not split and has a nontrivial cocycle. When then, can we define a parity operation of order two? [184]

15.4.2 Remarks About The Quantum Statistical Mechanics Of The Particle On The Ring

In quantum statistical mechanics a central object of study is the partition function:

$$Z := \mathrm{Tr}_{\mathcal{H}} e^{-\beta H} \tag{15.243}$$

Here $\beta = 1/(kT)$ where k is Boltzmann's constant and T is the absolute temperature.

For simplicity we will henceforth set $\hbar = k = 1$ (as can always be done by a suitable choice of units).

Since we have diagonalized the Hamiltonian exactly we can immediately say that

$$Z = \sum_{m \in \mathbb{Z}} e^{-\frac{\beta}{2I}(m-\mathcal{B})^2} \tag{15.244}$$

This is in fact a very interesting function of $\frac{\beta}{2I}$ and $\mathcal{B}$. Some immediate facts we can note are

---

[182] *Answer*: First reproduce Spin(2) using the general definition. We can represent the $d = 2$ Clifford algebra by σ^1, σ^2 and then the product of two vectors of the form $\gamma(v)$ with $v^2 = 1$ is a matrix of the form $(x^1\sigma^1 + x^2\sigma^2)(y^1\sigma^1 + y^2\sigma^2)$ where $v_1 = (x^1, x^2)$ and $v_2 = (y^1, y^2)$ are unit vectors in $\mathbb{R}^2$. Multiplying this out we get

$$\cos\theta + \sin\theta\sigma^1\sigma^2 \tag{15.242}$$

where θ is the angle between v_1 and v_2. Now let $\hat{P}$ be represented by σ^1 (or $\gamma(w)$ for any unit vector w).

[183] *Answer*: D_4 has generators x, y with $x^2 = 1$ and $y^4 = 1$ and $xyx = y^{-1}$. In a one-dimensional representation x, y will be represented by complex numbers. So we solve the above equations with $x, y \in \mathbb{C}$. Clearly $x \in \{\pm 1\}$ and then $y^2 = 1$ so $y \in \{\pm 1\}$ and there is no correlation between the choice of sign for x and the choice of sign for y.

[184] *Hint*:It is important to think about which group the cocycle and coboundaries take values in.

1. The expression is manifestly periodic under integer shifts of $\mathcal{B}$, illustrating the general claim above that the theory is invariant under integral shifts of the "theta angle" $\mathcal{B}$.

2. Moreover, at low temperature, $\beta \to \infty$ there is a single dominant term from the sum, unless $\mathcal{B}$ is a half-integer, in which there are two equally dominant terms - this reflects the double degeneracy of the ground state when $2\mathcal{B}$ is odd: The ground state is a Qbit. A standard technique in field theory is to study the IR behavior of a partition function to learn about the ground states of the system.

We are going to see that this system in fact has a very interesting high/low temperature duality and use this to understand better the θ-dependence of the previous example in terms of path integrals.

To relate Z to a path integral we observe that we can write:

$$Z = \int_0^{2\pi} d\phi \langle \phi | e^{-\beta H} | \phi \rangle \tag{15.245}$$

Now, we can interpret $\langle \phi | e^{-\beta H} | \phi \rangle$ as a specialization of the matrix elements of the Euclidean time propagator

$$\langle \phi_2 | e^{-t_E H_{\mathcal{B}}} | \phi_1 \rangle = \frac{1}{2\pi} \sum_{m \in \mathbb{Z}} e^{-\frac{t_E}{2I}(m-\mathcal{B})^2 + \mathrm{i}m(\phi_1 - \phi_2)} \tag{15.246}$$

Indeed, the usual propagator in quantum mechanics is

$$U(t) = e^{-\mathrm{i}tH/\hbar} \tag{15.247}$$

Under good conditions this family of operators for $t \in \mathbb{R}$ has an "analytic continuation" to part of the complex plane. What this means is that there is a well-defined family of operators

$$e^{-\mathrm{i}zH/\hbar} \tag{15.248}$$

where z takes values in a region $\mathcal{R} \subset \mathbb{C}$. The region $\mathcal{R}$ should, at least contain the real axis of time on its boundary (or closure). To see that there might be restrictions on $\mathcal{R}$ suppose that $\mathcal{R} = \mathbb{C}$. Then we can consider the restriction to the imaginary axis, setting

$$z = -\mathrm{i}t_E \qquad t_E \in \mathbb{R} \tag{15.249}$$

Here t_E is called *Euclidean time* because if we were to make a substitution $t \to -\mathrm{i}t_E$ in the Lorentz metric then we would get a metric of definite signature. (If we take signature $(-,+^d)$ we would get the Euclidean metric.) If the Hamiltonian is bounded below $\exp[-t_E H]$ should make sense for t_E positive, but if the spectrum of H grows rapidly the operator will be unbounded and certainly not traceclass for t_E negative. So, for Hamiltonians, such as the one we are considering the region $\mathcal{R}$ can be taken to be the negative half-plane. Defining such an analytic family of operators and restricting to the negative imaginary axis (or some other part of the imaginary axis) is called *Wick rotation.*

Now, it is well-known that the propagator: $\langle \phi_2 | e^{-\frac{\mathrm{i}tH}{\hbar}} | \phi_1 \rangle$ can be represented by a Feynman path integral. After Wick rotation we still have a path integral representation.

Feynman's argument proceeds just as well with $e^{-t_E H/\hbar}$. In fact, formally, it is better since the integral has better (formal) convergence properties for Euclidean actions whose real part is bounded below.

Now, by setting $\phi_1 = \phi_2 = \phi$ and integrating over ϕ we are making Euclidean time periodic, with period β and computing the path integral on a compact spacetime, namely, the circle. The path integral for $\phi_1 = \phi_2$ is done with boundary conditions on the fields so that $\phi(0) = \phi(\beta)$. This is precisely the kind of boundary condition that says that $\phi(t)$ is defined on a circle. More details on path integrals are available in many textbooks. See, for examples:

1. Feynman and Hibbs, *Quantum Mechanics and Integrals*
2. Feynman, *Statistical Mechanics*
3. C. Itzykson and J.B. Zuber, *Quantum Field Theory*,
4. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*,

In this subsubsection we will henceforth drop the subscript E on t_E and just use t for the real Euclidean time coordinate.

In the Wick rotation to Euclidean space the "θ-angle" $\oint eA$ remains real so the matrix elements of the Euclidean time propagator have the path integral representation:

$$\langle \phi_2 | e^{-\beta H} | \phi_1 \rangle = Z(\phi_2, \phi_1 | \beta) := \int [d\phi(t)]_{\phi(0)=\phi_1}^{\phi(\beta)=\phi_2} e^{-\frac{1}{\hbar}\int_0^\beta \frac{1}{2} I \dot{\phi}^2 dt - \mathrm{i}\int \mathcal{B}\dot{\phi} dt} \tag{15.250}$$

(One must be careful with the sign of the imaginary term, and it matters.)

Viewed as a field theory, this is a free field theory and the path integral can be done exactly by semiclassical techniques:

The equation of motion is simply $\ddot{\phi} = 0$. Again, the θ-term has not changed it.

Thus, the classical solutions to the equations of motion with boundary conditions $\phi(0) = \phi_1, \phi(\beta) = \phi_2$ are:

$$\phi_{cl}(t) = \phi_1 + \left(\frac{\phi_2 - \phi_1 + 2\pi w}{\beta} \right) t \qquad\qquad w \in \mathbb{Z} \tag{15.251}$$

or more to the point:

$$e^{\mathrm{i}\phi_{cl}(t)} = e^{\mathrm{i}\left((1-\frac{t}{\beta})\phi_1 + \frac{t}{\beta}\phi_2 \right) + \frac{2\pi i t w}{\beta}} \tag{15.252}$$

These are solutions of the Euclidean equations of motion, and are known as "instantons" for historical reasons. Notice that because of the compact nature of the spacetime on which we define our $0+1$ dimensional field theory there are infinitely many solutions labeled by $w \in \mathbb{Z}$. There are two circles in the game: The spacetime of this $0+1$-dimensional field theory is the Euclidean time circle. Then the target space of the field theory is also a circle. The quantum field $e^{\mathrm{i}\phi(t)}$ is a map $M \to \mathcal{X}$. M, which is spacetime is S^1 and $\mathcal{X}$, which is the target space is also $\mathcal{X}$. As we saw: $\pi_1(S^1) \cong \mathbb{Z}$. There can be topologically inequivalent field configurations. That is the space of maps $Map(M \to \mathcal{X})$ has different connected components. The different topological sectors are uniquely labelled by the winding number of the map (15.252). In the path integral we sum over all field configurations so we should sum over all these instanton configurations.

We now write

$$\phi = \phi_{cl} + \phi_q \tag{15.253}$$

where ϕ_{cl} is an instanton solution, as in (15.251), and ϕ_q is the quantum fluctuation with $\phi_q(0) = \phi_q(\beta) = 0$, and, moreover, $\phi_q(t)$ is in the topologically trivial component of $Map(S^1, \mathcal{X})$. The action is a sum $S(\phi_{cl}) + S(\phi_q)$, precisely because ϕ_{cl} solves the equation of motion. Indeed the integral factorizes and we just get:

$$Z(\phi_2, \phi_1|\beta) = Z_q \sum_{w\in\mathbb{Z}} e^{-\frac{2\pi^2 I}{\beta}(w+\frac{\phi_2-\phi_1}{2\pi})^2 + 2\pi i\mathcal{B}(w+\frac{\phi_2-\phi_1}{2\pi})} \tag{15.254}$$

The summation runs over classical solutions, weighted by the value of the classical action on that solution.

Z_q is the path integral over ϕ_q:

$$Z_q = \int [d\phi_q(t)]_{\phi_q(0)=0}^{\phi(\beta)=0} e^{-\int_0^\beta \frac{1}{2} I\dot{\phi}^2 dt} \tag{15.255}$$

We are integrating over the space of "all" maps $\phi_q : [0,1] \to U(1)$ with $\phi_q(0) = \phi_q(1) = 0$ that are homotopically trivial. We can do it by noticing that this is a Gaussian integral.

Now in finite dimensions we have the integral

$$\int \prod_{i=1}^n \frac{dx^i}{\sqrt{2\pi}} e^{-\frac{1}{2}x^i A_{ij} x^j + b_i x^i} = \frac{1}{\sqrt{\det A}} e^{\frac{1}{2}b_i (A^{-1})^{ij} b_j} \tag{15.256}$$

where $Re(A) > 0$ is a symmetric matrix. When A can be diagonalized by a real orthogonal transformation we can replace

$$\det A = \prod_{i=1}^n \lambda_i \tag{15.257}$$

where the product runs over the eigenvalues of A. Thus, we need to generalize this expression to the determinant of an infinite-dimensional "matrix"

$$\int [d\phi_q] \exp[-\int_0^1 \phi_q(-\frac{I}{2\beta}\frac{d^2}{d\tau^2})\phi_q] = (2\pi)\mathrm{Det}'^{-1/2}(\mathcal{O}) \tag{15.258}$$

Here the prime on the determinant means that we have omitted the zero-mode and the analog of A is the operator $\mathcal{O} = -\frac{I}{2\hbar\beta}\frac{d^2}{d\tau^2}$.

One way to make sense of $\mathrm{Det}\mathcal{O}$ for an operator $\mathcal{O}$ on Hilbert space is known as "ζ-function regularization." (It will only work for a suitable class of operators.) Note that

$$\frac{d}{ds}|_{s=0}\lambda^{-s} = -\log\lambda \tag{15.259}$$

So if we define

$$\zeta_{\mathcal{O}}(s) := \sum_\lambda \lambda^{-s} \tag{15.260}$$

where we take the sum over the spectrum of $\mathcal{O}$ (and we assume $\mathcal{O}$ is diagonalizable with discrete spectrum) then, formally:

$$\prod_\lambda \lambda = \exp[-\zeta'_{\mathcal{O}}(0)] \tag{15.261}$$

For good operators $\mathcal{O}$ the spectrum goes to infinity sufficiently fast that $\zeta_{\mathcal{O}}(s)$ exists as an analytic function of s in a half plane $Re(s) > N$ for some N. Moreover, $\zeta_{\mathcal{O}}(s)$ also admits an analytic continuation in s to an open region around $s = 0$. In this case, we can define the determinant by the RHS of (15.261).

For $\mathcal{O} = -\frac{I}{2\hbar\beta}\frac{d^2}{d\tau^2}$ we have $\zeta_{\mathcal{O}}(s) = 2\big(\frac{I\pi^2}{2\hbar\beta}\big)^s\zeta(2s)$ where $\zeta(s)$ is the standard Riemann ζ-function, and since

$$\zeta(s) = -\frac{1}{2} + s\log\big(\frac{1}{\sqrt{2\pi}}\big) + \mathcal{O}(s^2) \tag{15.262}$$

we have

$$\mathrm{Det}'(\mathcal{O}) := \exp[-\zeta'_{\mathcal{O}}(0)] = \frac{\beta}{I} \tag{15.263}$$

(There are some factors of 2 and π that need to be fixed in this equation.)

We can understand this result nicely as follows. Let us study the $\beta \to 0$ behavior of the path integral. Then for $|\phi_1 - \phi_2| < \pi$,

$$Z \to Z_q e^{-\frac{I}{2\beta}(\phi_2-\phi_1)^2 + i\mathcal{B}(\phi_2-\phi_1)}\Big(1 + \mathcal{O}(e^{-\kappa/\beta})\Big) \tag{15.264}$$

where $\kappa > 0$. In plain English: the instantons are only important at *large* β. This is intuitively very satisfying: At very small times β it must cost a lot of action for $\phi(t)$ to make a nonzero number of circuites around the circle because the velocity must then be large, and large velocity means large action. So for physical quantities based on such small fluctuations the topologically nontrivial field configurations must contribute subleading effects. Let us therefore compare $Z(\phi_2, \phi_1|\beta)$ as $\beta \to 0$ with the standard quantum mechanical propagator. For small ϕ we can remove the phase from the B-field via $\psi(\phi) \to e^{-i\mathcal{B}\phi}\psi(\phi)$, so Z_q should not depend on $\mathcal{B}$. (Note this transformation is not globally defined in ϕ for generic $\mathcal{B}$ so we cannot use it to remove $\mathcal{B}$ from the problem when we treat the full quantity Z exactly.) After this transformation we expect to recover the standard propagator of a particle of mass $M = I$ on the line. Rotated to Euclidean space this would be:

$$\sqrt{\frac{M}{2\pi\hbar\beta}}e^{-\frac{M(\phi_2-\phi_1)^2}{2\hbar\beta}} \tag{15.265}$$

so

$$Z_q = \sqrt{\frac{I}{2\pi\hbar\beta}} \tag{15.266}$$

The net result is that

$$Z(\phi_2, \phi_1|\beta) = \sqrt{\frac{I}{2\pi\beta}}\sum_{w\in\mathbb{Z}} e^{-\frac{2\pi^2 I}{\beta}(w+\frac{\phi_2-\phi_1}{2\pi})^2 - 2\pi i\mathcal{B}(w+\frac{\phi_2-\phi_1}{2\pi})} \tag{15.267}$$

Now compare (15.246) (with $t_E = \beta$) with (15.267). These expressions look very different! One involves a sum of exponentials with β in the numerator and the other with β in the demoninator. One is well-suited to discussing the asymptotic behavior for $\beta \to \infty$ (low temperature) and the other for $\beta \to 0$ (high temperature), respectively. Nevertheless, we have computed the same physical quantity, just using two different methods. So they must

be the same. But the mathematical identity that says they are the same appears somewhat miraculous. We now explain how to verify the two expressions are indeed identical using a direct mathematical argument.

The essential fact is the Poisson summation formula discussed in section

In our case, the expression computed directly from the diagonalization of the Hamiltonian is

$$\begin{aligned} Z(\phi_2,\phi_1|\beta) &= \frac{1}{2\pi}\sum_{m\in\mathbb{Z}} e^{-\frac{\beta}{2I}(m-\mathcal{B})^2+\mathrm{i}m(\phi_1-\phi_2)} \\ &= \frac{1}{2\pi} e^{\mathrm{i}\mathcal{B}(\phi_1-\phi_2)}\vartheta[\begin{smallmatrix}\theta\\ \phi\end{smallmatrix}](0|\tau) \end{aligned} \tag{15.268}$$

where we have written it in terms of the Riemann theta function (11.237) with:

$$\begin{aligned} \tau &= \mathrm{i}\frac{\beta}{2\pi I} \\ \theta &= -\mathcal{B} \\ \phi &= \frac{\phi_1-\phi_2}{2\pi} \end{aligned} \tag{15.269}$$

On the other hand, the expression that emerges naturally from the semiclassical evaluation of the Euclidean path integral is

$$\begin{aligned} Z(\phi_2,\phi_1|\beta) &= \sqrt{\frac{I}{2\pi\beta}}\sum_{w\in\mathbb{Z}} e^{-\frac{2\pi^2 I}{\beta}(w+\frac{\phi_2-\phi_1}{2\pi})^2-2\pi\mathrm{i}\mathcal{B}(w+\frac{\phi_2-\phi_1}{2\pi})} \\ &= \sqrt{\frac{I}{2\pi\beta}}\vartheta[\begin{smallmatrix}\theta'\\ \phi'\end{smallmatrix}](0|\tau') \end{aligned} \tag{15.270}$$

where we have written it in terms of the Riemann theta function (**??**):

$$\begin{aligned} \tau' &= \mathrm{i}\frac{2\pi I}{\beta} \\ \theta' &= -\frac{\phi_1-\phi_2}{2\pi} \\ \phi' &= -\mathcal{B} \end{aligned} \tag{15.271}$$

Note that the modular transformation law of the Riemann theta function, equation (**??**) is equivalent to the relation between the expressions naturally arising from the Hamiltonian and Lagrangian approaches to evaluation of the matrix elements of the Euclidean time propagator!

Note in particular that for the partition function proper we have, as $\beta \to 0$:

$$\begin{aligned} Z(S^1) &= \left(\frac{2\pi I}{\beta\hbar}\right)^{1/2}\sum_{n\in\mathbb{Z}} e^{-2\pi^2 n^2\frac{I}{\beta\hbar}+2\pi i n\mathcal{B}} \\ &\sim_{\beta\to 0} \left(\frac{2\pi I}{\beta\hbar}\right)^{1/2}\left(1+2e^{-2\pi^2\frac{I}{\beta\hbar}}\cos(2\pi\mathcal{B})+\cdots\right) \end{aligned} \tag{15.272}$$

The overall factor of $\beta^{-1/2}$ gives the expected divergence. The first correction term to the factor in parentheses is an instanton effect.

Note that in the Hamiltonian version the only thing that is manifest about the high-temperature, $\beta \to 0$, limit is that Z diverges. Note that for $\beta \to 0$ all the terms in the sum contribute about equally and the sum diverges. The modular transformation reveals an interesting duality: Once we factor out this multiplicative divergence we discover another theta function.

Exercise *A Parity Puzzle*
Suppose $2\mathcal{B}$ is an odd integer.
a.) Show that as $\beta \to \infty$ we have

$$\langle \phi_2 | e^{-\frac{\beta H}{\hbar}} | \phi_1 \rangle \sim 2 e^{\frac{\mathrm{i}}{2}(\phi_1 - \phi_2)} \cos\left(\frac{\phi_1 - \phi_2}{2} \right) e^{-\beta E_{\text{ground}}} + \cdots \tag{15.273}$$

b.) Note that this expression is not invariant under $\phi \to -\phi$. But in $\mathrm{Pin}^+(2)$ there is an element P which corresponds to $\phi \to -\phi$. How is this compatible with our argument that $\mathrm{Pin}^+(2)$ is a valid symmetry of the quantum theory? [185]

15.4.3 Gauging The Global $SO(2)$ Symmetry, Chern-Simons Terms, And Anomalies

When a theory has a symmetry one can implement a procedure called "gauging the symmetry." This is a two-step process:

1. Make the symmetry local and couple to a gauge field.
2. Integrate over "all possible" gauge fields consistent with the symmetry.

It is not necessary to proceed to step (2) after completing step (1). In this case, we say that we are coupling to nondynamical external gauge fields. It makes perfectly good sense to introduce nondynamical, external gauge fields for a symmetry. We do this all the time in quantum mechanics courses where we couple our quantum system to an electromagnetic field, but do not try to quantize the electromagnetic field.

For the more mathematically sophisticated reader the two-step process can be summarized, somewhat more concisely and precisely, as saying that we:

1. Identify the symmetry group with the structure group of a principal bundle and we change the bordism category in the domain of the field theory functor to include G-bundles with connection (where G is the symmetry we are gauging).

[185] *Answer*: Show that

$$\mathcal{P} \cdot |\phi\rangle = e^{-\mathrm{i}\phi} | -\phi \rangle \tag{15.274}$$

You can prove this by expanding $|\phi\rangle = \sum_{m \in \mathbb{Z}} \langle \Psi_m | \phi \rangle \Psi_m$. Now, using this expression check that the propagator indeed transforms correctly.

2. Sum over isomorphism classes of principal bundles and integrate over the isomorphism classes of connections on those bundles.

In the present simple example of the charged particle on a ring surrounding a solenoid we can "gauge" the global $SO(2)$ symmetry $\phi \to \phi + \alpha$ that is present for all values of $\mathcal{B}$. It is then interesting to see how coupling to the external gauge field tells us about the subtleties of combining $SO(2)$ symmetry with charge conjugation symmetry that we studied above. (The following discussion was inspired by Appendix D of. [186])

So in our simple example we implement Step 1 above as follows: We seek to make the shift symmetry local, that is, we attempt to make

$$\phi(t) \to \phi(t) + \alpha(t) \tag{15.275}$$

into a symmetry where $\alpha(t)$ is not a constant but an "arbitrary" function of time. When $\alpha(t)$ is time dependent the action $\sim \int \dot{\phi}^2$ is not invariant under such transformations. To compensate for this we introduce an extra function of time into the problem, call it $A^{(e)}(t)$. Here the superscript e - for "external" - reminds us that this is an "external" or "background" field: We will <u>not</u> do a path integral over these functions (unless we proceed to Step 2 above). By contrast, we <u>will</u> do a path integral over the "dynamical" field $\phi(t)$ or $\Phi(t) = e^{\mathrm{i}\phi(t)}$.

The gauged action is

$$\begin{aligned} S &= \int \frac{1}{2} I(\dot{\phi} + A^{(e)})^2 dt + \oint \mathcal{B}(\dot{\phi} + A^{(e)}) dt \\ &= \int \frac{1}{2} I \left(\Phi(t)^{-1}(-\mathrm{i}\frac{d}{dt} + A^{(e)})\Phi(t) \right)^2 dt + \oint \mathcal{B}\Phi(t)^{-1}(-\mathrm{i}\frac{d}{dt} + A^{(e)})\Phi(t) dt \end{aligned} \tag{15.276}$$

This action is a functional of both the nondynamical field $A^{(e)}(t)$ and the dynamical field $\phi(t)$. Note that the action is invariant under the gauge transformation:

$$\begin{aligned} \phi(t) &\to \phi(t) + \alpha(t) \\ A^{(e)}(t) &\to A^{(e)}(t) - \partial_t \alpha(t) \end{aligned} \tag{15.277}$$

where, for the moment, we ignore boundary terms.

It turns out that it is better to regard the function $A^{(e)}(t)$ as a component of a 1-form:

$$A^{(e)} := A^{(e)}(t) dt \tag{15.278}$$

and, better still, $A^{(e)}$ is the local one-form associated to a connection on a (locally trivialized) principal $SO(2)$ bundle over the time manifold M. We stress that the gauge field $A^{(e)}$ is <u>NOT</u> the gauge field of electromagnetism. (That field has already produced our theta term.) Rather, it is a new field in our system: It is a gauge field for the shift symmetry of the field $\phi(t)$. [187]

[186] D. Gaiotto, A. Kapustin, Z. Komargodski and N. Seiberg, "Theta, Time Reversal, and Temperature," https://arxiv.org/pdf/1703.00501.pdf

[187] The physical interpretation of the gauging process in terms of our original charged particle on a ring is not completely clear to the author. But the mathematical structure makes sense and is a good toy model for other field theoretic systems where the physical interpretation of the gauging is clear.

Note that we could also write the gauge transformation in the form:

$$\begin{aligned} e^{\mathrm{i}\phi(t)} &\to e^{\mathrm{i}\phi(t)} e^{\mathrm{i}\alpha(t)} \\ d - \mathrm{i}A^{(e)} &\to e^{\mathrm{i}\alpha(t)} (d + \mathrm{i}A^{(e)}) e^{-\mathrm{i}\alpha(t)} \end{aligned} \tag{15.279}$$

This is better because is captures better the geometric content, Consequently, it makes more sense when working on topologically nontrivial spacetimes, such as the Euclidean time circle. One can make choices of gauge group so that it becomes important that $e^{\mathrm{i}\alpha(t)}$ can be single-valued even when $\alpha(t)$ is not.

What about charge conjugation symmetry?

First consider the classical theory. In the absence of the external gauge field we noted that there is an $O(2)$ symmetry of the equations of motion, even though under $\phi(t) \to -\phi(t)$ the theta term in the action flips sign. In the presence of the external gauge field the equations of motion are modified, however, as we will see below, we can gauge $A^{(e)}(t)$ to be a constant, and in this case they are not modified. So we still have an $O(2)$ symmetry.

Now consider the quantum theory. One can show that, appropriately defined, the quantum Hamiltonian is still $H_{\mathcal{B}}$. Under charge conjugation we must flip $\mathcal{B}$ and then we change the Hamiltonian (unless $\mathcal{B} = 0$). But, as noted above, if $2\mathcal{B} \in \mathbb{Z}$ in that case $H_{\mathcal{B}}$ is unitarily equivalent to $H_{-\mathcal{B}}$ and we can implement a unitary operator $\mathcal{P}$ corresponding to the charge conjugation operation. In the path integral the value of the action matters. The action (15.276) is invariant under the charge conjugation transformation if we take

$$\begin{aligned} \phi(t) &\to -\phi(t) \\ A^{(e)} &\to -A^{(e)} \\ \mathcal{B} &\to -\mathcal{B} \end{aligned} \tag{15.280}$$

and consequently if we change $\mathcal{B} \to -\mathcal{B}$ we must also take $A^{(e)} \to -A^{(e)}$ as noted above. We will return to the quantum implementation of charge conjugation symmetry.

Now let us re-examine the periodicity of the physics as a function of $\mathcal{B}$. In the absence of the external gauge field $A^{(e)}$ we found that physical quantities are periodic functions of $\mathcal{B}$ with period one. However, in the presence of a nonzero $A^{(e)}$, the term $\int \mathcal{B} A^{(e)}(t) dt$ spoils the periodicity in $\mathcal{B}$, because the value of the action matters in the quantum theory.

We can restore a kind of periodicity in $\mathcal{B}$ by adding a Chern-Simons term to the action. We will comment in detail on the Chern-Simons term below. In Euclidean space the new action is:

$$e^{-S} = e^{-\int \frac{1}{2} I(\dot{\phi} + A_t^{(e)})^2 dt - \mathrm{i} \oint \mathcal{B}(\dot{\phi} + A_t^{(e)}) dt} e^{\mathrm{i}k \int A_t^{(e)} dt} \tag{15.281}$$

and the last factor is the Chern-Simons term. By introducing the Chern-Simons term we have introduced yet another parameter, the level k, into our theory. Classically the action with $(\mathcal{B}, k)$ is equivalent to the action with $(\mathcal{B} + r, k + r)$ where $r \in \mathbb{R}$ is any real number.

Now that we have restored some kind of periodicity we can ask about quantum implementation of charge conjugation symmetry. We must take $\mathcal{B} \to -\mathcal{B}$, but quantum mechanically the theory with $\mathcal{B}$ is only equivalent to that with $-\mathcal{B}$ when $2\mathcal{B} \in \mathbb{Z}$. So, in

the quantum theory we can only hope to have charge conjugation invariance if there is an integer N so that

$$(\mathcal{B} + N, k + N) = (-\mathcal{B}, -k) \tag{15.282}$$

In other words $k = \mathcal{B} = N/2 \in \mathbb{Z}/2$.

The introduction of the Chern-Simons term raises a new issue: When one sees gauge potentials in an action that do not enter through fieldstrengths or covariant derivatives it is important to ask about gauge invariance. In order to discuss the gauge invariance of the Chern-Simons term properly we need first to discuss more carefully the space of gauge fields and the group of gauge transformations.

The Space Of Gauge Fields And The Group Of Gauge Transformations

In our simple setting the space $\mathcal{A}$ of gauge fields can be identified with the space of single-valued, continuous real-valued functions $A^{(e)}(t)$ on M.

We now must choose a *gauge group* G. This will be a Lie group - typically finite-dimensional, although not necessarily connected. In our case, there are two natural choices: We could take $G = \mathbb{R}$ or we could take $G = U(1)$. Then the *group of gauge transformations* is a group of maps

$$\mathcal{G} = Map[M \to G] \tag{15.283}$$

If M and G have positive dimension the group of gauge transformations will be an infinite-dimensional Lie group.

Of particular interest in gauge theory is the quotient space $\mathcal{A}/\mathcal{G}$, the space of gauge orbits, or, equivalently, the space of gauge-inequivalent field configurations.

Let us examine a few examples of $\mathcal{A}/\mathcal{G}$:

1. Let us first consider what happens when $G = \mathbb{R}$ and M is an interval or the real line. Then the space $\mathcal{A}$ of gauge fields can be identified with the space of real-valued continuous functions on M. The group $\mathcal{G}$ is the space of real-valued C^1 functions on M, $t \mapsto \alpha(t) \in \mathbb{R}$. The group $\mathcal{G}$ acts on $\mathcal{A}$ via

$$A^{(e)}(t) \to A^{(e)}(t) - \partial_t \alpha(t) \tag{15.284}$$

If $M = [t_1, t_2]$ with free boundary conditions on $\mathcal{A}$ and $\mathcal{G}$ then we can always solve

$$\partial_t \alpha(t) = A^{(e)}(t) \tag{15.285}$$

for some $\alpha(t)$ and hence we can always gauge $A^{(e)}(t)$ to zero. So $\mathcal{A}/\mathcal{G}$ is just a point. [188]

For the discussion below it is useful to note here that the expression

$$\exp[\mathrm{i} \int_{t_1}^{t_2} A_t^{(e)}(t') dt'] \tag{15.286}$$

[188] Actually, this is too naive: $\mathcal{A}/\mathcal{G}$ is more properly thought of as a stack. See the section below on groupoids.

is in general not gauge invariant. Rather:

$$\exp[\mathrm{i}\int_{t_1}^{t_2} A_t^{(e)}(t')dt'] \to e^{\mathrm{i}\alpha(t_1)}\exp[\mathrm{i}\int_{t_1}^{t_2} A_t^{(e)}(t')dt']e^{-\mathrm{i}\alpha(t_2)} \tag{15.287}$$

2. Now, continuing to take $G = \mathbb{R}$ let us consider what happens if $M = \mathbb{R}$ and we impose boundary conditions that $\alpha(t) \to 0$ at $t \to \pm\infty$. In this case

$$\int_{-\infty}^{+\infty} A^{(e)}(t)dt \tag{15.288}$$

is gauge invariant. There are no "local" invariants. From the previous discussion we see that we can gauge $A^{(e)}(t)$ to zero in any compact region. In this case

$$\mathcal{A}/\mathcal{G} \cong \mathbb{R} \tag{15.289}$$

and the integral (15.288) fully determines the gauge equivalence class.

3. Now let us consider the case $G = U(1)$, and let us also take M to be the Euclidean time circle so

$$\mathcal{G} = \mathrm{Map}(S^1_{s.t.} \to U(1)) \tag{15.290}$$

Now, just viewing the gauge transformation as a set of continuous maps $S^1 \to S^1$ there is a winding number. If this winding number is nonzero there is an obstruction to finding a single-valued function $\alpha(t)$ so that $g(t) = e^{\mathrm{i}\alpha(t)}$.

There is a normal subgroup $\mathcal{G}_0$ of *small gauge transformations* for which $g(t)$ admits a well-defined logarithm. That is, the gauge transformations $g(t) \in \mathcal{G}_0$ are of the form $g(t) = e^{\mathrm{i}\alpha(t)}$ where $\alpha(t)$ is single-valued. Then

$$1 \to \mathcal{G}_0 \to \mathcal{G} \quad \stackrel{\pi}{\to} \quad \mathbb{Z} \to 1 \tag{15.291}$$

where the map π can be viewed as the winding number. Gauge transformations in $\mathcal{G}_0$ are known as *small gauge transformations*. Those which have nonzero winding numbers are known as *large gauge transformations*. It is worth noting that the above sequence splits: For $w \in \mathbb{Z}$ we can take $s(w) = g_w$ to be the gauge transformation:

$$g_w(t) = \exp[2\pi\mathrm{i}wt/\beta] \tag{15.292}$$

and $g_w(t)g_{w'}(t) = g_{w+w'}(t)$. Note that for these transformations if we tried to define $\alpha(t)$ it would be $\alpha(t) = 2\pi wt/\beta$ and would not be single-valued when M is the circle.

Now, referring to (15.295) it is clear that if $\alpha(t)$ is single valued then

$$\exp[\mathrm{i}\oint_{S^1} A_t^{(e)}(t')dt'] \tag{15.293}$$

is gauge invariant. However when we have a large gauge transformation $g_w(t)$ we can cut the circle say at $t = 0$ and $t = \beta$ and then, with $\alpha(t) = 2\pi wt/\beta$ with $w \in \mathbb{Z}$ the

holonomy is still gauge invariant. Note that the large gauge transformations $g_w(t)$ take

$$A_t^{(e)}(t') \to A_t^{(e)}(t') + w/\beta \tag{15.294}$$

but preserves the holonomy (15.293). Put differently, (15.295) is generalized to

$$\exp[\mathrm{i}\int_{t_1}^{t_2} A_t^{(e)}(t')dt'] \to g(t_2)^{-1}\exp[\mathrm{i}\int_{t_1}^{t_2} A_t^{(e)}(t')dt']g(t_1) \tag{15.295}$$

so that on the circle the *holonomy* (15.293) is gauge invariant. Equation (15.295) generalizes nicely to the case of nonabelian groups on arbitrary spacetimes.

We can ask if there are other independent gauge invariant functions of $A^{(e)}(t)$ besides the holonomy. Since $A^{(e)}(t)$ is periodic we can decompose $A^{(e)}(t)$ in a Fourier expansion. Write $\tilde{A}_t^{(e)}(t)$ for the sum of the nonzero frequency modes. Then we can solve the differential equation

$$\partial_t \alpha(t) = \tilde{A}_t^{(e)}(t) \tag{15.296}$$

with a single-valued $\alpha(t)$ to choose a gauge so that

$$A^{(e)}(t) = \mu/\beta \tag{15.297}$$

is constant. Put differently, $\mathcal{A}/\mathcal{G}_0$ can be identified with the space of real numbers, given by the constant μ. We will denote $\mathcal{A}^{\rm red} = \mathcal{A}/\mathcal{G}_0$. Then the "large" gauge transformations $g_w(t)$ shift $\mu \to \mu + 2\pi w$, with $w \in \mathbb{Z}$. Therefore we have

$$\mathcal{A}/\mathcal{G} \cong \mathcal{A}^{\rm red}/\mathbb{Z} \cong \{[\mu] = [\mu + 2\pi w] \qquad w \in \mathbb{Z}\} \cong U(1) \tag{15.298}$$

and the holonomy $e^{\mathrm{i}\mu}$ around the circle is a complete gauge invariant.

4. Finally, let us take the gauge group to be $G = O(2) = SO(2) \rtimes \mathbb{Z}_2$. Let $\mathcal{G}_{SO(2)}$ be the group of gauge transformations when the gauge group is $SO(2)$ and $\mathcal{G}_{O(2)}$ be the group of gauge transformations when the gauge group is $O(2)$. Then $\mathcal{G}_{O(2)} = \mathcal{G}_{SO(2)} \rtimes \mathbb{Z}_2$ and we have the exact sequence

$$1 \to \mathcal{G}_0 \to \mathcal{G}_{O(2)} \to \mathbb{Z} \rtimes \mathbb{Z}_2 \cong D_\infty \to 1 \tag{15.299}$$

Again the sequence splits and the infinite dihedral group $D_\infty \cong \mathbb{Z} \rtimes \mathbb{Z}_2$ preserves the space $\mathcal{A}^{\rm red}$ of constant gauge fields with generators acting as

$$\begin{aligned} \sigma &: \mu \to -\mu \\ s &: \mu \to \mu + 2\pi \end{aligned} \tag{15.300}$$

Comments On The Chern-Simons Term

Now we are ready to discuss the gauge invariance of the Chern-Simons term. The Chern-Simons term on $M = S^1$ is invariant under $\mathcal{G}_0$ for any value of k. However, under the large gauge transformations $g_w(t)$ with $w \neq 0$:

$$\exp[\mathrm{i}k \oint_{S^1} A_t^{(e)}(t')dt'] \to e^{2\pi \mathrm{i} w k} \exp[\mathrm{i}k \oint_{S^1} A_t^{(e)}(t')dt'] \tag{15.301}$$

and therefore, if we are going to allow our theory to make sense on a circle with the gauge group $SO(2) \cong U(1)$ then k should be quantized to be an integer. Note there would be no such quantization of k if the gauge group is taken to be $\mathbb{R}$.

The above observation is related to two extremely important conceptual points that are essential to all discussions of the use of Chern-Simons terms in quantum physics:

It is not necessary for the action to be invariant. All that is necessary for a well-defined path integral is that the exponentiated action must be invariant.
Gauge invariance of the "Chern-Simons term" under large gauge transformations implies that the level k, one of the couplings of the theory, is quantized: $k \in \mathbb{Z}$.

In our case the action in equation (15.281) is not gauge invariant! Under large gauge transformations with winding number w we have $S \to S + 2\pi \mathrm{i} k w$ with $w \in \mathbb{Z}$. However, having a well-defined measure in the path integral only requires e^{-S} to be well-defined, and this will be the case if, and only if, $k \in \mathbb{Z}$.

Note that

$$\exp[\mathrm{i}k \oint_{S^1} A_t^{(e)}(t')dt'] = e^{\mathrm{i}k\mu} \tag{15.302}$$

and, since $k \in \mathbb{Z}$ is quantized this is properly periodic under $\mu \to \mu + 2\pi$ and the exponentiated Chern-Simons term descends to a well-defined function on $\mathcal{A}/\mathcal{G}$. For the group $O(2)$ we would have to consider $\cos(\mu)$. (Of course, $\cos(n\mu)$ is a Tchebyshev polynomial of the basic invariant $\cos(\mu)$.)

Anomalies

We can now discuss, very generally, the notion of anomalies. In quantum systems we typically have both "dyanmical variables" such as dynamical fields, degrees of freedom, etc. as well as "external" or "background" or "control" variables. We will denote generic "background fields" by ϕ^{bck} and generic "dynamical fields" by ϕ^{dyn}. Any parameter of the theory should be considered a "field." The space of all fields is then fibered:

$$\begin{array}{ccc} \mathcal{F}^{dyn} & \longrightarrow & \mathcal{F} \\ & & \downarrow \\ & & \mathcal{F}^{bck} \end{array} \tag{15.303}$$

In the very simple situation we are discussing here the fibration is just a Cartesian product.

In the computation of physical quantities we will typically integrate over $\mathcal{F}^{dyn}$ thus producing a function (or, more generally, a section of a bundle) on $\mathcal{F}^{bck}$. We then study the physical quantity as a function on $\mathcal{F}^{bck}$.

In our example $\mathcal{F}^{dyn}$ can be taken to be the set of functions $\Phi(t) : M \to U(1)$ and $\mathcal{F}^{bck}$ can be taken to be the set of functions $A^{(e)}(t)$, or better, the connections on a principal G-bundle over M where in our present examples $G = \mathbb{R}, SO(2)$ or $O(2)$. [189]

Now suppose that there is a group $\mathcal{G}$ acting on $\mathcal{F}$ so that physical quantities are formally invariant. For example, if we have an invariant action $S[\phi^{dyn}; \phi^{bck}]$ and a formally invariant measure, then the path integral will be formally invariant. Then, physical quantities such as the partition function:

$$Z[\phi^{bck}] = \int_{\mathcal{F}^{dyn}} e^{-S[\phi^{dyn};\phi^{bck}]} \mathrm{vol}\,(\phi^{dyn}) \tag{15.304}$$

will, formally define a $\mathcal{G}$-invariant function on $\mathcal{F}^{bck}$. However, it can happen that when one defines the path integral carefully the partition function fails to be $\mathcal{G}$-invariant. In that case we say that there is a *potential anomaly*. Sometimes potential anomalies can be removed by physically unimportant redefinitions. When this cannot be done we say there is an *anomaly*.

If we tried to consider the Chern-Simons term for $k \notin \mathbb{Z}$ we would say it has an *anomaly*. For any value of k it descends to a function on $\mathcal{A}/\mathcal{G}_0$. However, only when $k \in \mathbb{Z}$ does it desend to a well-defined function on $\mathcal{A}/\mathcal{G}$.

It can happen that there can be different subgroups $\mathcal{H}_1 \subset \mathcal{G}$ and $\mathcal{H}_2 \subset \mathcal{G}$ such that there are different definitions of the path integral so that it is invariant either under $\mathcal{H}_1$ or under $\mathcal{H}_2$ but there is no definition so that it is invariant simultaneously under both $\mathcal{H}_1$ and $\mathcal{H}_2$. In this case we say there is a *mixed anomaly*.

The Partition Function As Function On $\mathcal{A}$ And Its Behavior Under The Action Of $\mathcal{G}$

Let us now illustrate some of the above ideas about anomalies by examining the partition function in our example of the gauged particle on a ring.

There will not be any interesting anomalies under $\mathcal{G}_0$. As we have explained we can always use $\mathcal{G}_0$ to gauge $A^{(e)}$ to be a constant 1-form, and we will henceforth take our gauge field to be constant. Then the equation of motion is the same as before, and performing the path integral just as in the previous section we find

$$\begin{aligned} Z(\mu) &= e^{\mathrm{i}k\mu} Z_q \sum_{w\in\mathbb{Z}} e^{-\frac{2\pi^2 I}{\beta}(w+\frac{\mu}{2\pi})^2 - 2\pi\mathrm{i}\mathcal{B}(w+\frac{\mu}{2\pi})} \\ &= e^{\mathrm{i}k\mu} Z_q \vartheta[{}^{\mu/2\pi}_{-\mathcal{B}}](0|\tau) \end{aligned} \tag{15.305}$$

[189] One can also promote $\mathcal{B}$ to be a field for fun and profit. This has been discussed in many places. See e-Print: 1905.09315 for a recent discussion.

with $\tau = \mathrm{i}\frac{2\pi I}{\beta}$. All we need to do here is replace the value of the classical action for solutions with $\dot{\phi} = 2\pi w/\beta$ by making the substitution $w \to w + \mu/2\pi$.

As in the case without the external gauge field there is a Hamiltonian interpretation. Performing the Poisson summation (or using the modular transformation law of the theta function) we get:

$$Z(\mu) = e^{\mathrm{i}(k-\mathcal{B})\mu}\frac{1}{2\pi}\sum_{m\in\mathbb{Z}} e^{-\frac{\beta}{2I}(m-\mathcal{B})^2 - \mathrm{i}(m-\mathcal{B})\mu} \tag{15.306}$$

It can be shown that the Euclidean path integral with action (15.193) is in fact equal to

$$Z(\mu) = e^{\mathrm{i}k\mu}\mathrm{Tr}e^{-\beta H_{\mathcal{B}}}e^{\mathrm{i}\mu Q} \tag{15.307}$$

where Q is the operator measuring the charge of the $SO(2)$ symmetry we gauged. In our case $Q\Psi_m = m\Psi_m$. [190] As noted above we still have

$$H_{\mathcal{B}} = \frac{1}{2I}\left(-\mathrm{i}\frac{\partial}{\partial\phi} - \mathcal{B}\right)^2 \tag{15.308}$$

acting on $L^2(\mathcal{X}) = L^2(S^1)$. One easily checks the equality of (15.306) and (15.307).

From either point of view $Z(\mu)$ is a periodic function of μ and there is no anomaly under the group $\mathbb{Z}$ of large gauge transformations, so long as $k \in \mathbb{Z}$.

The Gauge Group $O(2)$ and Mixed Anomalies

What happens if we try to extend the gauge group to gauge the full $O(2)$? Then, as we have seen, the quotient group $\mathcal{G}/\mathcal{G}_0$ is the infinite dihedral group generated by σ and s defined in equation (15.300) above.

If $2\mathcal{B}$ is even then we can take $k = \mathcal{B} \in \mathbb{Z}$. The partition function is invariant under $\mu \to -\mu$ and has the expected periodicity $\mu \sim \mu + 2\pi$. In other words $Z(\mu)$ is invariant under the group of large gauge transformations isomorphic to D_∞ and generated by σ and s, so it descends to a function on $\mathcal{A}/\mathcal{G}$ and there is no anomaly.

Things are much more subtle when $2\mathcal{B}$ is odd. As we saw, we can only expect charge conjugation symmetry when

$$k = \mathcal{B} \in \mathbb{Z} + \frac{1}{2} \tag{15.309}$$

But this clashes with the constraint $k \in \mathbb{Z}$. So we see an example of a mixed anomaly.

It is interesting to see how the mixed anomaly is manifested in the partition function. The main point can be seen most easily by considering the leading term in the $\beta \to \infty$ expansion which is (taking $\mathcal{B} = 1/2$ for simplicity):

$$Z \to \frac{e^{-\beta E_{ground}}}{2\pi}e^{\mathrm{i}(k-\frac{1}{2})\mu}\left(e^{\mathrm{i}\mu/2} + e^{-\mathrm{i}\mu/2}\right) + \cdots \tag{15.310}$$

[190] To give a first-principles proof of why this should be so we gauge away $A^{(e)}$ in the path integral. The result is an identification of the fields at $t = 0$ with the fields at $t = \beta$ accompanied by a gauge transformation by the holonomy $e^{\mathrm{i}\mu}$ which takes $e^{\mathrm{i}\phi} \to e^{\mathrm{i}(\phi+\mu)}$. So $\Psi_m \to e^{\mathrm{i}m\mu}\Psi_m = e^{\mathrm{i}\mu Q}\Psi_m$.

If $k = 0$ then

$$Z \to \frac{e^{-\beta E_{ground}}}{2\pi} \left(1 + e^{-\mathrm{i}\mu}\right) + \cdots \tag{15.311}$$

the expression is properly periodic in μ, but not invariant under the analog of charge conjugation: $\mu \to -\mu$. This is not surprising since $k \neq \mathcal{B}$.

As we will discuss below, by changing the physical system (yet again!) there is a way to make sense of the half-integral level Chern-Simons term. If we just go ahead and mindlessly substitute $k = \mathcal{B} = 1/2$ in the above formula for $Z(\mu)$ we get:

$$Z \to \frac{e^{-\beta E_{ground}}}{2\pi} \left(e^{\mathrm{i}\mu/2} + e^{-\mathrm{i}\mu/2}\right) + \cdots \tag{15.312}$$

The action is now invariant under the generator $\sigma : \mu \to -\mu$ of D_∞ but it is no longer invariant under the generator $s : \mu \to \mu + 2\pi$.

Provided we view the different choices of Chern-Simons terms as different definitions of the theory, we can define the theory to be invariant under the group generated by s, but with that definition σ is anomalous, or we can define the theory to be invariant udner the group generated by σ, but then with that definition s is anomalous. So in this sense there is a mixed anomaly of $\mathbb{Z}$ and $\mathbb{Z}_2$ in the D_∞ subgroup of global gauge transformations.

Making sense of Chern-Simons terms with half-integer level

There is a way to make sense of the half-integer quantized Chern-Simons term by viewing the $0+1$ dimensional theory as the boundary of a well-defined $1+1$ dimensional theory. By Stokes' theorem we have:

$$\exp[ik \oint_{S^1} A_t^{(e)} dt] = \exp[\mathrm{i}k \int_\Sigma F^{(e)}] \tag{15.313}$$

where $F^{(e)} = dA^{(e)}$. The RHS makes sense even if k is not an integer, but now the expression depends on details of the gauge field in the "bulk" of the $1+1$ dimensional spacetime Σ.

A very analogous phenomenon is observed in real condensed matter systems where the boundary theory of a 3+1 dimensional topological insulator is described by a Chern-Simons theory with half-integral level. (That is, half the level allowed by naive gauge invariance.)

Such half-integral level Chern-Simons terms come up in many interesting physical systems. For example, half-integral (spin) Chern-Simons theory is needed to describe the topological features of the fractional quantum Hall effect. In supersymmetric field theories and string theories many of the supergravity effective actions and brane effective actions involve half-integrally quantized Chern-Simons terms.

Exercise *Puzzle*

Warning: This exercise requires some knowledge of topology.

Resolve the following paradox:

We first argued that, if $k \notin \mathbb{Z}$ then the LHS of (15.313) is not invariant under large gauge transformations. Then we proceeded to define the LHS by the expression on the RHS which is manifestly gauge invariant.

How can these two statements be compatible? [191]

15.5 Heisenberg Extensions

Consider again a central extension

$$1 \to A \to \widetilde{G} \to G \to 1 \tag{15.314}$$

In many of the examples above we had G Abelian and $\widetilde{G}$ was also Abelian. However, as our examples with Q and D_4 have shown, in general $\widetilde{G}$ need not be Abelian. (See equation (15.177).) In this section we focus on an important class of examples where G is Abelian and $\widetilde{G}$ is non-Abelian. They are known as *Heisenberg groups* and *Heisenberg extensions*. In fact in the literature closely related but slightly different things are meant by "Heisenberg extensions" and "Heisenberg groups." These kinds of extensions show up all the time in physics, in many different ways. They are very basic in quantum field theory and other areas of physics, so we are going to dwell upon them a bit.

15.5.1 Heisenberg Groups: The Basic Motivating Example

Those who have taken quantum mechanics will be familiar with the relation between position and momentum operators for the quantum mechanics of a particle on the real line:

$$[\hat{q}, \hat{p}] = \mathrm{i}\hbar \tag{15.315}$$

One realization of these operator relations is in terms of normalizable wavefunctions $\psi(q)$ where we write:

$$\begin{aligned} (\hat{q} \cdot \psi)(q) &= q\psi(q) \\ (\hat{p} \cdot \psi)(q) &= -\mathrm{i}\hbar \frac{d}{dq}\psi(q) \end{aligned} \tag{15.316}$$

Now, let us consider the operators

$$\begin{aligned} U(\alpha) &:= \exp[\mathrm{i}\alpha\hat{p}] \\ V(\beta) &:= \exp[\mathrm{i}\beta\hat{q}] \end{aligned} \tag{15.317}$$

These are unitary when α, β are real. When α is real $U(\alpha)$ implements translation in position space by $\hbar\alpha$. When β is real $V(\beta)$ implements translation in momentum space by $-\hbar\beta$.

[191] *Answer*: The gauge transformation $e^{\mathrm{i}\alpha(t)}$ must extend to a continuous map $\Sigma \to U(1)$. If Σ is a smooth manifold whose only boundary is S^1, as we have tacitly assumed in writing equation (15.313), then such maps always restrict to small gauge transformations on the bounding S^1.

The group of operators $\{U(\alpha)|\alpha \in \mathbb{R}\}$ is isomorphic to $\mathbb{R}$ because $U(\alpha_1)U(\alpha_2) = U(\alpha_1+\alpha_2)$. A similar statement holds for the group of operators $V(\beta)$. But when we take products of both $U(\alpha)$ and $V(\beta)$ operators we do not get the group $\mathbb{R}\oplus\mathbb{R}$ of translations in position and momentum, separately. Rather, one can show in a number of ways that:

$$U(\alpha)V(\beta) = e^{\mathrm{i}\hbar\alpha\beta}V(\beta)U(\alpha) \tag{15.318}$$

This is an extremely important equation. We can understand it in many different ways. We will explain three ways to derive it. First, it immediately follows from the BCH formula since $[\hat{q},\hat{p}]$ is central.

A second way to derive (15.318) is to evaluate both operators on a wavefunction in the position representation. So, on the one hand:

$$\begin{aligned}((U(\alpha)V(\beta))\cdot\psi)(q) &= (V(\beta)\cdot\psi)(q+\hbar\alpha)\\ &= e^{\mathrm{i}\beta(q+\hbar\alpha)}\psi(q+\hbar\alpha)\end{aligned} \tag{15.319}$$

On the other hand

$$\begin{aligned}((V(\beta)U(\alpha))\cdot\psi)(q) &= e^{\mathrm{i}\beta q}(U(\alpha)\cdot\psi)(q+\hbar\alpha)\\ &= e^{\mathrm{i}\beta q}\psi(q+\hbar\alpha)\end{aligned} \tag{15.320}$$

Comparing (15.319) with (15.320) we arrive at (15.320). The reader should compare this with our discussion of quantum mechanics with a finite number of degrees of freedom, especially the derivation of (11.251).

Here is a third derivation of (15.318): Using (11.441) it follows that

$$U(\alpha)\hat{q}U(\alpha)^{-1} = e^{\mathrm{i}\alpha\mathrm{Ad}(\hat{p})}\hat{q} = \hat{q}+\hbar\alpha \tag{15.321}$$

$$V(\beta)\hat{p}V(\beta)^{-1} = e^{\mathrm{i}\beta\mathrm{Ad}(\hat{q})}\hat{p} = \hat{q}-\hbar\beta \tag{15.322}$$

Now using (11.445) we obtain (15.318).

Returning to the group generated by the operators $U(\alpha)$ and $V(\alpha)$ for $\alpha\in\mathbb{R}$, which we'll denote $\mathrm{Heis}(\mathbb{R}\times\mathbb{R})$, it fits in a central extension:

$$1 \to U(1) \to \mathrm{Heis}(\mathbb{R}\times\mathbb{R}) \to \mathbb{R}\times\mathbb{R} \to 1 \tag{15.323}$$

With one (nice) choice of cocycle we can write the group law as:

$$(z_1,(\alpha_1,\beta_1))\cdot(z_2,(\alpha_2,\beta_2)) = (z_1z_2e^{\frac{i}{2}\hbar(\alpha_1\beta_2-\alpha_2\beta_1)},(\alpha_1+\alpha_2,\beta_1+\beta_2)) \tag{15.324}$$

Notice that the cocycle is expressed in terms of the anti-symmetric form

$$\omega(v_1,v_2) := \alpha_1\beta_2-\alpha_2\beta_1 = \begin{pmatrix}\alpha_1 & \beta_1\end{pmatrix}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}\alpha_2\\ \beta_2\end{pmatrix} = v_1^{tr}Jv_2 \tag{15.325}$$

where

$$v = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} \tag{15.326}$$

and the matrix

$$J = \begin{pmatrix} 0 & 1_n \\ -1_n & 0 \end{pmatrix} \tag{15.327}$$

was used at the very beginning of the course (see equation (2.21)) to define the symplectic group $Sp(2n, \kappa)$ as the group of matrices such that $A^{tr} J A = J$.

ω is called a *symplectic form* Note that

$$\omega(A\vec{v}_1, A\vec{v}_2) = \omega(\vec{v}_1, \vec{v}_2) \tag{15.328}$$

for $A \in Sp(2, \mathbb{R})$. We say that ω is invariant under symplectic transformations.

Exercise

Referring to equations (15.318) and (15.316) et. seq.

a.) Show that the choice of section

$$s(\alpha, \beta) = U(\alpha)V(\beta) \tag{15.329}$$

leads to the cocycle

$$f((\alpha_1, \beta_1), (\alpha_2, \beta_2) = e^{i\hbar(\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_1\beta_2)} \tag{15.330}$$

b.) Show that the choice of section

$$s(\alpha, \beta) = \exp[\mathrm{i}(\alpha\hat{p} + \beta\hat{q})] \tag{15.331}$$

leads to the cocycle

$$f((\alpha_1, \beta_1), (\alpha_2, \beta_2) = e^{\frac{i}{2}\hbar(\alpha_1\beta_2 - \alpha_2\beta_1)} \tag{15.332}$$

c.) Find an explicit coboundary that relates the cocycle (15.330) to (15.332).

♣NEED TO PROVIDE ANSWER HERE. ♣

Exercise *Generalization To* Heis($\mathbb{R}^n \oplus \mathbb{R}^n$)

Consider the group generated by exponentiating linear combinations of $\hat{q}^i$ and $\hat{p}_i$, $i = 1, \ldots, n$ with the nonzero commutators being:

$$[\hat{q}^i, \hat{p}_j] = \mathrm{i}\hbar\delta^i{}_j \tag{15.333}$$

Then for

$$\vec{v} = \begin{pmatrix} \alpha_i \\ \beta^j \end{pmatrix} \tag{15.334}$$

we can choose section:

$$s(\vec{v}) = \exp\left[\mathrm{i}(\alpha^i\hat{p}_i + \beta_i\hat{q}^i)\right] \tag{15.335}$$

Show that the resulting group is

$$(z_1, \vec{v}_1) \cdot (z_2, \vec{v}_2) = (z_1 z_2 e^{\mathrm{i}\frac{\hbar}{2}\omega(\vec{v}_1, \vec{v}_2)}, \vec{v}_1 + \vec{v}_2) \tag{15.336}$$

with ω defined by equation (11.470).

15.5.2 Example: The Magnetic Translation Group For Two-Dimensional Electrons

In the presence of an electromagnetic field the group of translations acting on charged particles definitely becomes centrally extended. This shows up naturally when discussing a charged nonrelativistic particle confined to two spatial dimensions and moving in a constant magnetic field B. In one convenient gauge the Hamiltonian is

$$H = \frac{1}{2m}\left((p_1 + \frac{eBx_2}{2})^2 + (p_2 - \frac{eBx_1}{2})^2\right) = \frac{1}{2m}(\tilde{p}_1^2 + \tilde{p}_2^2) \tag{15.337}$$

where the gauge invariant momenta are $\tilde{p}_i := p_i - eA_i$ are

$$\begin{aligned} \tilde{p}_1 &= p_1 + \frac{eB}{2}x_2 \\ \tilde{p}_2 &= p_2 - \frac{eB}{2}x_1 \end{aligned} \tag{15.338}$$

Ordinary translations are generated by p_1, p_2 and do not commute with the Hamiltonian: We have lost translation invariance. Nevertheless we can define the *magnetic translation operators*:

$$\pi_1 := p_1 - \frac{eBx_2}{2} \qquad \pi_2 := p_2 + \frac{eBx_1}{2} \tag{15.339}$$

Compare this carefully with the definitions of $\tilde{p}_i$. Note the relative signs! These operators satisfy $[\pi_i, \tilde{p}_j] = 0$. In particular they are translation-like operators that commute with the Hamiltonian: $[\pi_i, H] = 0$. Hence the name. While they are called "translation operators" note that they do not commute:

$$[\pi_1, \pi_2] = -i\hbar eB \tag{15.340}$$

The "magnetic translation group" is generated by the operators

$$\begin{aligned} U(a_1) &= \exp[ia_1\pi_1/\hbar] \\ V(a_2) &= \exp[ia_2\pi_2/\hbar] \end{aligned} \tag{15.341}$$

The operators $U(a_1), V(a_2)$ satisfy the relations:

$$U(a_1)V(a_2) = \exp[\mathrm{i}eBa_1a_2/\hbar]V(a_2)U(a_1) \tag{15.342}$$

If we are interested in quantized values of a_1, a_2 (as, for example, if the charged particle is moving in a lattice, or is confined to a torus) then we obtain the basic relations (15.522). Note that

$$\exp[\mathrm{i}eBa_1a_2/\hbar] = \exp[2\pi\mathrm{i}\Phi/\Phi_0] \tag{15.343}$$

where $\Phi = Ba_1a_2$ is the flux through an area element a_1a_2 and $\Phi_0 = h/e$ is known as the *magnetic flux quantum.* [192]

[192] One should be careful about a factor of two here since in superconductivity the condensing field has charge $2e$ and hence the official definition of the term "flux quantum" used, for example, by NIST is $\Phi_0 = h/2e$, half the value we use.

Remark: The group generated by the operators $U = U(a_1)$ and $V = V(a_2)$ is a Heisenberg group, but it is also interesting to consider the *algebra* of operators generated by U, V. This algebra admits a $C*$-algebra structure and is sometimes referred to as the *algebra of functions on the noncommutative torus* or the *irrational rotation algebra.* Abstractly it is the C^* algebra generated by unitary operators U, V satisfying $UV = e^{2\pi i\theta}VU$ for some θ. The properties of the algebra are very different for θ rational and irrational. The algebra $\mathcal{A}_\theta$ figures prominantly in applications of noncommutative geometry to the QHE and in applications of noncommutative geometry to toroidal compactifications of string theory.

15.5.3 The Commutator Function And The Definition Of A General Heisenberg Group

Let us now step back and think more generally about central extensions of G by A where G is *also abelian.* From the exercise (15.190) we know that for G abelian the commutator is

$$[(a_1, g_1), (a_2, g_2)] = \left(\frac{f(g_1, g_2)}{f(g_2, g_1)}, 1\right) \tag{15.344}$$

(We are writing $1/f(g_2, g_1)$ for $f(g_2, g_1)^{-1}$ and since A is abelian the order doesn't matter, so we write a fraction as above.)

The function $\kappa : G \times G \to A$ defined by

$$\kappa(g_1, g_2) = \frac{f(g_1, g_2)}{f(g_2, g_1)} \tag{15.345}$$

is known as the *commutator function.*

Note that:

1. The commutator function is *gauge invariant*, in the sense that it does not change under the change of 2-cocycle f by a coboundary. (Check that! This uses the property that G is abelian). It is therefore a more intrinsic quantity associated with the central extension.

2. $\kappa(g, 1) = \kappa(1, g) = 1$. (This follows from exercise (15.92) above.)

3. The extension $\tilde{G}$ is abelian iff $\kappa(g_1, g_2) = 1$, that is, iff there exists a symmetric cocycle f. [193]

4. κ is *skew*:
$$\kappa(g_1, g_2) = \kappa(g_2, g_1)^{-1} \tag{15.346}$$

5. κ is *alternating*:
$$\kappa(g, g) = 1 \tag{15.347}$$

6. κ is *bimultiplicative*:
$$\kappa(g_1 g_2, g_3) = \kappa(g_1, g_3)\kappa(g_2, g_3) \tag{15.348}$$
$$\kappa(g_1, g_2 g_3) = \kappa(g_1, g_2)\kappa(g_1, g_3) \tag{15.349}$$

[193] Note that in our example of Q, D_4 as extensions the cocycle we computed was not symmetric.

All of these properties except, perhaps, the last, are obvious. To prove the bimultiplicative properties (it suffices to prove just one) we rewrite (15.348) as

$$f(g_1g_2, g_3)f(g_3, g_2)f(g_3, g_1) = f(g_2, g_3)f(g_1, g_3)f(g_3, g_1g_2) \tag{15.350}$$

Now multiply the equation by $f(g_1, g_2)$ and use the fact that A is abelian to write

$$(f(g_1, g_2)f(g_1g_2, g_3))f(g_3, g_2)f(g_3, g_1) = f(g_2, g_3)f(g_1, g_3)(f(g_1, g_2)f(g_3, g_1g_2)) \tag{15.351}$$

We apply the cocycle identity on both the LHS and the RHS (and also use the fact that G is abelian) to get

$$f(g_2, g_3)f(g_1, g_2g_3)f(g_3, g_2)f(g_3, g_1) = f(g_2, g_3)f(g_1, g_3)f(g_3, g_1)f(g_3g_1, g_2) \tag{15.352}$$

Now canceling some factors and using that A is abelian we have

$$f(g_1, g_2g_3)f(g_3, g_2) = f(g_1, g_3)f(g_3g_1, g_2) \tag{15.353}$$

Now use the fact that G is abelian to write this as

$$f(g_1, g_3g_2)f(g_3, g_2) = f(g_1, g_3)f(g_1g_3, g_2) \tag{15.354}$$

which is the cocycle identity. This proves the bimultiplicative property (15.348). ♠

We now define the *Heisenberg extensions*. The function κ is said to be *nondegenerate* if for all $g_1 \neq 1$ there is a g_2 with $\kappa(g_1, g_2) \neq 1$. When this is the case the center of $\tilde{G}$ is precisely A:

$$Z(\widetilde{G}) \cong A \ . \tag{15.355}$$

This follows immediately from equation (15.344). If κ is degenerate the center will be larger. In the extreme case that $\kappa(g_1, g_2) = 1$ for all g_1, g_2 we get the direct product $\widetilde{G} = A \times G$ and

$$Z(\widetilde{G}) = \widetilde{G} \ . \tag{15.356}$$

In general, we will have an intermediate situation and A will be a proper subgroup of $Z(\tilde{G})$.

One definition which is used in the literature is

Definition: A *Heisenberg extension* is a central extension of an *abelian* group G by an *abelian* group A where the commutator function κ is nondegenerate.

Exercise *Alternating implies skew*

Show that a map $\kappa : G \times G \to A$ which satisfies the bimultiplicative identity (15.348) and the alternating identity (15.347) is also skew, that is, satisfies (15.346).

Exercise *Commutator function for* Heis($\mathbb{R} \oplus \mathbb{R}$)

a.) Show that both of the cocycles (15.330) and (15.332) defining groups isomorphic to Heis($\mathbb{R} \oplus \mathbb{R}$) have the same commutator function

$$\kappa((\alpha_1, \beta_1), (\alpha_2, \beta_2) = e^{\mathrm{i}\hbar(\alpha_1\beta_2 - \alpha_2\beta_1)} = e^{\mathrm{i}\hbar\omega(v_1, v_2)} \tag{15.357}$$

b.) Similarly show that for Heis($\mathbb{R}^n \oplus \mathbb{R}^n$) we have

$$\kappa(v_1, v_2) = \exp\left(\mathrm{i}\hbar\omega(v_1, v_2)\right) \tag{15.358}$$

where

$$\omega(v_1, v_2) = v_1^{tr} J v_2 \tag{15.359}$$

15.5.4 Classification Of $U(1)$ Central Extensions Using The Commutator Function

For a large class of abelian groups G, there is a nice theorem regarding arbitrary central extensions by $U(1)$. We consider

1. Finitely generated Abelian groups. As we will prove in sections 16.2 and 16.3 below, these can be (noncanonically) written as products of cyclic groups $\mathbb{Z}_n$, for various n, and a lattice $\mathbb{Z}^d$ for some d (possibly $d = 0$).
2. Vector spaces. [194]
3. Tori. These are isomorphic to $V/\mathbb{Z}^d$ where V is a d-dimensional real vector space.
4. Direct products of the above three.

Remark: This class of groups can be characterized as the set of Abelian groups A which are topological groups so there is an exact sequence:

$$0 \to \pi_1(A) \to \mathrm{Lie}(A) \to A \to \pi_0(A) \to 0 \tag{15.360}$$

where Lie(A) is a vector space that projects to A by an exponential map.

For this class of groups we have the following theorem:

Theorem Let G be a topological Abelian group of the above class. The isomorphism classes of central extensions of G by $U(1)$ are in one-one correspondence with continuous bimultiplicative maps

$$\kappa : G \times G \to U(1) \tag{15.361}$$

which are alternating (and hence skew).

[194] Topological, separable.

I do not know who originally proved this theorem, but one proof can be found in [195]

Remarks

1. In other words, given the commutator function κ one can always find a corresponding cocycle f. This theorem is useful because κ is invariant under change of f by a coboundary, and moreover the bimultiplicative property is simpler to check than the cocycle identity. (In fact, one can show that it is always possible to find a cocycle f which is bimultiplicative. This property automatically ensures the cocycle relation.)

2. It is important to realize that κ only characterizes $\tilde{G}$ up to *noncanonical* isomorphism: to give a definite group one must choose a definite cocycle. ♣Explain this comment more. ♣

3. In this theorem we can replace $U(1)$ by any subgroup of $U(1)$, such as $\mathbb{Z}_n$ realized as the group of nth roots of unity.

15.5.5 Pontryagin Duality And The Stone-von Neumann-Mackey Theorem

The theorem of section 15.5.4 is nicely illustrated by the Heisenberg of extension of the Abelian group $G = S \times \widehat{S}$ by $A = U(1)$. Here S is any locally compact Abelian group and $\widehat{S}$ is the Pontryagin dual.

Now, there is a very natural alternating, skew, bilinear map on the product $S \times \widehat{S}$ defined by

$$\kappa\left((s_1, \chi_1), (s_2, \chi_2)\right) := \frac{\chi_2(s_1)}{\chi_1(s_2)} \tag{15.362}$$

and according to the above theorem this defines a general Heisenberg extension

$$1 \to U(1) \to \mathrm{Heis}(S \times \widehat{S}) \to S \times \widehat{S} \to 1 \tag{15.363}$$

at least up to isomorphism.

Remarks

1. Note that one natural cocycle giving the commutator function (15.362) is

$$f\left((s_1, \chi_1), (s_2, \chi_2)\right) := \frac{1}{\chi_1(s_2)} \tag{15.364}$$

2. There is a very natural representation of (15.363). We need a Haar measure on S so that we can define $V = L^2(S)$, a Hilbert space of complex-valued functions on S.

[195] D. Freed, G. Moore, G. Segal, "The uncertainty of fluxes," Commun.Math.Phys. 271 (2007) 247-274, arXiv:hep-th/0605198, Proposition A.1.

For our key examples we have

$$\begin{aligned}\langle\psi_1,\psi_2\rangle &= \int_{\mathbb{R}} \psi_1^*(x)\psi_2(x)dx \qquad S=\mathbb{R}\\ &= \sum_{n\in\mathbb{Z}} \psi_1^*(n)\psi_2(n) \qquad S=\mathbb{Z}\\ &= \frac{1}{2\pi}\int_0^{2\pi} \psi_1^*(e^{\mathrm{i}\theta})\psi_2(e^{\mathrm{i}\theta})d\theta \qquad S=U(1)\\ &= \frac{1}{n}\sum_{k=0}^{n-1} \psi_1^*(\bar{k})\psi_2(\bar{k}) \qquad S=\mathbb{Z}/n\mathbb{Z}\end{aligned} \tag{15.365}$$

We represent $s_0 \in S$ as a translation operator:

$$(T_{s_0}\cdot\Psi)(s) := \Psi(s+s_0) \tag{15.366}$$

and we represent $\chi_0 \in \widehat{S}$ as a multiplication operator

$$(M_{\chi_0}\cdot\Psi)(s) := \chi_0(s)\Psi(s) \tag{15.367}$$

Then one checks that this does <u>not</u> define a representation of the direct product $S\times\widehat{S}$ but rather we have the operator equation:

$$T_{s_0}M_{\chi_0} = \chi_0(s_0)M_{\chi_0}T_{s_0} \tag{15.368}$$

If $\mathcal{O}$ is the group of operators generated by T_s, M_χ and $z\in U(1)$ acting on V then the map

$$(z;(s,\chi)) \to zT_sM_\chi \tag{15.369}$$

is a homomorphism, and in fact an isomorphism of $\mathrm{Heis}(S\times\widehat{S})$ with $\mathcal{O}$ where we use the cocycle (15.364).

3. This construction is extremely important in quantum mechanics and also in the description of free quantum field theories. In these cases we take S to be a vector space. For example, for the quantum mechanics of a particle in $\mathbb{R}^n$ we have $V=\mathbb{R}^n$ as an Abelian group. For the quantum field theory of a free real scalar field in $d+1$ dimensions we would take V to be a suitable space of real-valued functions on a d-dimensional spatial slice. [196] Then we introduce the dual vector space $V^\vee\cong\widehat{S}$, and the canonical pairing $V\times V^\vee\to\mathbb{R}$ gives the character:

$$\chi_k(v) = e^{\mathrm{i}k\cdot v} \tag{15.370}$$

Then, T_s in our general discussion is the operator $U(s)$ of the basic motivating example, and M_{χ_k} is the operator $V(k)$ of the basic motivating example and the general identity (15.368) becomes our starting point:

$$U(s)V(k) = e^{\mathrm{i}k\cdot s}V(k)U(s) \tag{15.371}$$

[196] Technical point: This will not be locally compact. So, if one wishes to be rigorous, further considerations are required.

4. *Stone-von Neumann-Mackey Theorem.* Up to isomorphism (equivalence) there is a unique irreducible unitary representation of $\mathrm{Heis}(S \times \widehat{S})$ such that $A = U(1)$ acts by scalar multiplication. That is, if $\xi \in U(1)$ then we require $\rho(\xi) = \xi \mathbf{1}_V$. In addition we need some further technical hypotheses. [197] This is called the Stone-von Neumann theorem or sometimes the Stone-von Neumann-Mackey theorem. For a relatively short proof see. [198] The main idea is to consider the maximal Abelian subgroups of $\mathrm{Heis}(S \times \widehat{S})$. One such subgroup is isomorphic to $U(1) \times S$, another is $U(1) \times \widehat{S}$. Let us consider $U(1) \times \widehat{S}$. Over the subgroup $\{1\} \times \widehat{S}$ we can split the sequence and consider the elements:

$$M_\chi := \rho(1, (0, \chi)) \tag{15.372}$$

where $1 \in U(1)$, $0 \in S$, and $\chi \in \widehat{S}$. These operators commute for different choices of χ and are simultaneously diagonalizable so we have a complete basis ψ_α of simultaneous eigenvectors for the representation with eigenvalues: [199]

$$M_\chi \psi_\alpha = \lambda_\alpha(\chi) \psi_\alpha \tag{15.373}$$

Clearly $\chi \mapsto \lambda_\alpha(\chi)$ must be a character on $\widehat{S}$, so the eigenvalues can be identified with characters on $\widehat{S}$ and therefore, by Pontryagin duality, can be identified with elements $s_\alpha \in S$. So there is some set $\{s_\alpha\}$ whose elements are drawn from S with corresponding eigenbasis ψ_α with

$$M_\chi \psi_\alpha = \chi(s_\alpha) \psi_\alpha \tag{15.374}$$

Now, for any $s \in S$ define the operator

$$T_s := \rho(1, (s, 1)) \tag{15.375}$$

Choose some particular α_0 and consider the vectors $T_s \psi_{\alpha_0}$ for $s \in S$. Using the Heisenberg relations and the fact that the central $U(1)$ group just acts by scalars we know that:

$$M_\chi \left(T_s \psi_{\alpha_0} \right) = \chi(s - s_{\alpha_0}) \left(T_s \psi_{\alpha_0} \right) \tag{15.376}$$

Therefore the span of $\{T_s \psi_{\alpha_0}\}_{s \in S}$, which is the span of $\{T_{s+s_{\alpha_0}} \psi_{\alpha_0}\}_{s \in S}$ is a copy of the representation constructed above. So if (V, ρ) is irreducible, it must be equivalent to our representation constructed above. This demonstrates uniqueness of the irreducible representation.

5. *Simple Proof Of Irreducibility for $S = \mathbb{R}^n$.* For $v = (\alpha, \beta) \in \mathbb{R}^{2n}$ we introduced the section

$$s(v) := \exp[\mathrm{i}(\alpha \hat{q} + \beta \hat{p})] \tag{15.377}$$

Now, consider the standard representation of $\mathrm{Heis}(\mathbb{R}^n \times \widehat{\mathbb{R}^n})$ on $\mathcal{H} = L^2(\mathbb{R}^n)$. For

♣This remark assumes at least a little bit of knowledge from the linear algebra chapter 2 and the idea of an irreducible representation, from chapter 4. ♣

♣The α, β here are reversed from the convention in previous section. ♣

[197] We need the representation to be continuous in the norm topology and we need S to have a translation-invariant measure so that $L^2(A)$ makes sense. Then we replace V above by the Hilbert space $L^2(S)$.

[198] A. Prasad, "An easy proof of the Stone-von Neumann-Mackey Theorem," arXiv:0912.0574.

[199] It is exactly at this point that we are using unitarity, and a careful discussion requires more functional analysis.

any two vectors $\psi_1, \psi_2 \in \mathcal{H}$ define the *Wigner function*:

$$W(\psi_1, \psi_2)(v) := \langle s(v)\psi_1, \psi_2 \rangle \tag{15.378}$$

This is a function on the phase space $v \in \mathbb{R}^{2n}$. Now we compute:

$$(s(v)\psi_1)(q) = e^{\mathrm{i}\frac{\alpha\beta}{2}} e^{\mathrm{i}\alpha q}\psi_1(q+\beta) \tag{15.379}$$

Shifting the integration variable by $\beta/2$ in the inner product we have

$$W(\psi_1, \psi_2)(v) = \int_{\mathbb{R}^n} e^{-\mathrm{i}\alpha q}\psi_1^*(q+\beta/2)\psi_2(q-\beta/2)dq \tag{15.380}$$

Now, an elementary computation shows that, on $L^2(\mathbb{R}^{2n})$ with the standard measure we have

$$\| W(\psi_1, \psi_2) \|^2 = \| \psi_1 \|^2 \| \psi_2 \|^2 \tag{15.381}$$

Here are some details

♣Need to clean up some 2π's here... ♣

$$\begin{aligned} \| W(\psi_1, \psi_2) \|^2 &= \int_{\mathbb{R}^{2n}} |W(\psi_1, \psi_2)|^2 d^{2n}v \\ &= \int d\alpha d\beta dq_1 dq_2 e^{\mathrm{i}\alpha(q_1-q_2)}\psi_1(q_1+\beta/2)\psi_2^*(q_1-\beta/2)\psi_1^*(q_2+\beta/2)\psi_2(q_2-\beta/2) \\ &= \int dq d\beta \psi_1^*(q+\beta/2)\psi_1(q+\beta/2)\psi_2^*(q-\beta/2)\psi_2(q-\beta/2) \\ &= \| \psi_1 \|^2 \| \psi_2 \|^2 \end{aligned} \tag{15.382}$$

Now with the key result (15.381) we can show that $\mathcal{H}$ is irreducible. [200] Suppose that $\mathcal{H}_0 \subset \mathcal{H}$ is preserved preserved by the Heisenberg group. Suppose there is a vector $\psi_\perp \notin \mathcal{H}_0$. WLOG we can take $\psi_\perp$ to be perpendicular to $\mathcal{H}_0$ (hence the notation). But then,

$$W(\psi, \psi_\perp)(v) = \langle s(v)\psi, \psi_\perp \rangle = 0 \tag{15.383}$$

for all $v \in \mathbb{R}^{2n}$ and all $\psi \in \mathcal{H}_0$, because $s(v)\psi \in \mathcal{H}_0$. But then by (15.381) we know that $\| \psi \|^2 \| \psi_\perp \|^2 = 0$. Therefore either $\psi = 0$ or $\psi_\perp = 0$. If $\mathcal{H}_0$ is not the zero vector space then we can always choose $\psi \neq 0$ and hence $\psi_\perp = 0$. But if $\mathcal{H}_0$ were proper then there would be a nonzero choice for $\psi_\perp$. Therefore, there is no nonzero and proper subspace $\mathcal{H}_0 \subset \mathcal{H}$ preserved by the group action of Heis. Therefore $\mathcal{H}$ is irreducible.

6. *Stone von-Neumann And Fourier* Now let us combine the Stone-von-Neumann theorem with Pontryagin duality. Because $\widehat{\widehat{S}} \cong S$ we can write

$$\mathrm{Heis}(S \times \widehat{S}) \cong \mathrm{Heis}(\widehat{S} \times \widehat{\widehat{S}}) \tag{15.384}$$

[200] See chapter 4 for a thorough discussion of reducible vs. irreducible representations. Briefly - if a representation $\mathcal{H}$ has a nonzero and proper subspace $\mathcal{H}_0$ preserved by the group action then it is said to be *reducible*. A representation which is not reducible is said to be *irreducible*.

so, we could equally well give a unitary representation of the group by taking $\widehat{V} := Fun(\widehat{S} \to \mathbb{C})$ with inner product

$$\langle \hat{\psi}_1, \hat{\psi}_2 \rangle := \int_{\widehat{S}} d\chi \hat{\psi}_1^*(\chi)\hat{\psi}_2(\chi) \tag{15.385}$$

Now we represent translation and multiplication operators by

$$(\hat{T}_{\chi_0}\hat{\psi})(\chi) := \hat{\psi}(\chi_0\chi) \tag{15.386}$$

$$(\hat{M}_{s_0}\hat{\psi})(\chi) := \chi(s_0)\hat{\psi}(\chi) \tag{15.387}$$

and check the commutator function. So, by SvN there must be a unitary isomorphism

$$\mathcal{S} : V \to \widehat{V} \tag{15.388}$$

mapping

$$\begin{aligned} \mathcal{S}T_{s_0}\mathcal{S}^{-1} &= \hat{M}_{s_0} \\ \mathcal{S}M_{\chi_0}\mathcal{S}^{-1} &= \hat{T}_{\chi_0} \end{aligned} \tag{15.389}$$

Indeed there is: It is the Fourier transform:

$$\psi \mapsto \hat{\psi}(\chi) := \int_S \chi(s)^*\psi(s)ds \tag{15.390}$$

Moreover, $\mathcal{S}$ is an isometry by the Plancherel/Parseval theorem noted above.

7. Of course, there are many different Lagrangian subgroups of G. For example, if G is a symplectic vector space there will be many different Lagrangian subspaces, all related by the linear action of the symplectic group. Each choice of Lagrangian subspace gives a representation of the Heisenberg group, but they all must be isomorphic, by the Stone-von Neumann-Mackey theorem. This leads to interesting isomorphisms between seemingly different representations and interesting (projective) actions of symplectic groups. These kinds of considerations are central to some simple examples of "duality symmetries" in quantum field theory. We will next investigate how the groups of symplectic automorphisms lift to automorphism groups of Heisenberg groups.

Exercise *Bimultiplicative cocycle*
a.) Check that (15.364) satisfies the cocycle relation.
b.) Show that, in fact, (15.364) is bimultiplicative.

Exercise *Irreducibility Of The Stone-von Neumann Representation Of* $\mathrm{Heis}(S \times \widehat{S})$

Generalize the proof of irreducibility for $S = \mathbb{R}^n$ to other locally compact Abelian groups. [201]

15.5.6 Some More Examples Of Heisenberg Extensions

For this section the reader might wish to consult section [**** 2.1 *** ?] of chapter two for the definition of a ring. The reader won't lose much by taking $R = \mathbb{Z}$ or $R = \mathbb{Z}/N\mathbb{Z}$ with abelian group structure $+$ and extra multiplication structure $\bar{n}_1 \bar{n}_2 = \overline{n_1 n_2}$.

Example 1: Suppose R is a commutative ring with identity. Then we can consider the group of 3×3 matrices over R of the form

$$M(a,b,c) := \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \tag{15.391}$$

The multiplication law is easily worked out to be

$$M(a,b,c)M(a',b',c') = M(a+a', b+b', c+c'+ab') \tag{15.392}$$

Therefore, if we define $\mathrm{Heis}(R \times R)$ to be the group of matrices $M(a,b,c)$ and take

$$\pi : M(a,b,c) \to (a,b) \tag{15.393}$$

and

$$\iota : c \mapsto M(0,0,c) \tag{15.394}$$

we have an extension

$$0 \to R \to \mathrm{Heis}(R \times R) \to R \oplus R \to 0 \tag{15.395}$$

with cocycle $f((a,b),(a',b')) = ab'$. Note that we are writing our Abelian group R additively so the cocycle identity becomes

$$f(v_1,v_2) + f(v_1+v_2,v_3) = f(v_1,v_2+v_3) + f(v_2,v_3) \tag{15.396}$$

where $v = (a,b) \in \mathbb{R} \oplus \mathbb{R}$. In this additive notation the commutator function is

$$\kappa((a,b),(a',b')) = ab' - a'b \tag{15.397}$$

In the literature one will sometimes find the above class of groups defined as the "Heisenberg groups." It is a special case of what we have defined as general Heisenberg groups.

---

[201] *Answer* For the general case define the Wigner function as the function $W(\psi_1,\psi_2) : S \times \widehat{S} \to \mathbb{C}$ by $W(\psi_1,\psi_2)(s,\chi) := \langle T_s M_\chi \psi_1, \psi_2 \rangle$. Show that (15.381) continues to hold. You will need to use the orthogonality relation for characters.

As a special case of the above construction let us take $R = \mathbb{Z}/n\mathbb{Z}$. We will now show that we recover the group $\mathrm{Heis}(\mathbb{Z}_n \times \mathbb{Z}_n)$ discussed in section 11.11 in the context of a particle on a discrete approximation to a circle.

First consider $\mathbb{Z}_n$ written additively. So if $a \in \mathbb{Z}$, then $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$ is just $\bar{a} = a + n\mathbb{Z}$ is the coset. Then we define

$$U = \begin{pmatrix} \bar{1} & \bar{1} & 0 \\ 0 & \bar{1} & 0 \\ 0 & 0 & \bar{1} \end{pmatrix} \qquad V = \begin{pmatrix} \bar{1} & 0 & 0 \\ 0 & \bar{1} & \bar{1} \\ 0 & 0 & \bar{1} \end{pmatrix} \qquad q = \begin{pmatrix} \bar{1} & 0 & \bar{1} \\ 0 & \bar{1} & 0 \\ 0 & 0 & \bar{1} \end{pmatrix} \tag{15.398}$$

We easily check that for $a \in \mathbb{Z}$,

$$U^a = \begin{pmatrix} \bar{1} & \bar{a} & 0 \\ 0 & \bar{1} & 0 \\ 0 & 0 & \bar{1} \end{pmatrix} \qquad V^a = \begin{pmatrix} \bar{1} & 0 & 0 \\ 0 & \bar{1} & \bar{a} \\ 0 & 0 & \bar{1} \end{pmatrix} \qquad q^a = \begin{pmatrix} \bar{1} & 0 & \bar{a} \\ 0 & \bar{1} & 0 \\ 0 & 0 & \bar{1} \end{pmatrix} \tag{15.399}$$

so

$$U^n = V^n = q^n = 1 \tag{15.400}$$

Moreover:

$$UV = qVU \qquad qU = Uq \qquad qV = Vq \tag{15.401}$$

Thus we obtain the presentation:

$$\mathrm{Heis}(\mathbb{Z}_n \times \mathbb{Z}_n) = \langle U, V, q | U^n = V^n = q^n = 1, \quad UV = qVU, \quad Uq = qU, \quad Vq = qV \rangle \tag{15.402}$$

we saw before.

A simple and useful generalization of the previous construction is to take any bilinear map $c : R \times R \to \mathcal{Z}$ where $\mathcal{Z}$ is Abelian. Thus $c(a_1 + a_2, b) = c(a_1, b) + c(a_2, b)$ and $c(a, b_1 + b_2) = c(a, b_1) + c(a, b_2)$. Then we can define a central extension

$$0 \to \mathcal{Z} \to \tilde{G} \to R \oplus R \to 0 \tag{15.403}$$

by the law

$$(z_1, (a, b)) \cdot (z_2, (a', b')) = (z_1 + z_2 + c(a, b'), (a + a', b + b')) \tag{15.404}$$

The corresponding group cocycle is $f((a, b), (a', b')) = c(a, b')$. The cocycle relation is satisfied simply by virtue of c being bilinear. It will be a Heisenberg extension if $\kappa : (R \times R) \times (R \times R) \to \mathcal{Z}$ given by $\kappa((a, b), (a', b')) = c(a, b') - c(a', b)$ is nondegenerate. In particular, if we take $\mathcal{Z} = R$ and $c(a, b') = ab'$ using the ring multiplication then we recover (15.391).

Example 2: *Clifford algebra representations and Extra-special groups.* Suppose we have a set of matrices γ_i, $1 \leq i \leq n$ such that

$$\{\gamma_i, \gamma_j\} = 2\delta_{ij} \tag{15.405}$$

Such a set of matrices can indeed be constructed, by taking suitable tensor products of Pauli matrices. They are called "gamma matrices." They form a matrix representation of what is called a *Clifford algebra* and we will study them in more detail and more abstractly in chapter ****. For the moment the reader should be content with the explicit representation: [202]

$$\begin{aligned}\gamma_1 &= \sigma^1\\ \gamma_2 &= \sigma^2\end{aligned} \tag{15.406}$$

for $n = 2$,

$$\begin{aligned}\gamma_1 &= \sigma^1\\ \gamma_2 &= \sigma^2\\ \gamma_3 &= \sigma^3\end{aligned} \tag{15.407}$$

for $n = 3$,

$$\begin{aligned}\gamma_1 &= \sigma^1 \otimes \sigma^1\\ \gamma_2 &= \sigma^1 \otimes \sigma^2\\ \gamma_3 &= \sigma^1 \otimes \sigma^3\\ \gamma_4 &= \sigma^2 \otimes 1\end{aligned} \tag{15.408}$$

for $n = 4$,

$$\begin{aligned}\gamma_1 &= \sigma^1 \otimes \sigma^1\\ \gamma_2 &= \sigma^1 \otimes \sigma^2\\ \gamma_3 &= \sigma^1 \otimes \sigma^3\\ \gamma_4 &= \sigma^2 \otimes 1\\ \gamma_5 &= \sigma^3 \otimes 1\end{aligned} \tag{15.409}$$

for $n = 5$,

$$\begin{aligned}\gamma_1 &= \sigma^1 \otimes \sigma^1 \otimes \sigma^1\\ \gamma_2 &= \sigma^1 \otimes \sigma^1 \otimes \sigma^2\\ \gamma_3 &= \sigma^1 \otimes \sigma^1 \otimes \sigma^3\\ \gamma_4 &= \sigma^1 \otimes \sigma^2 \otimes 1\\ \gamma_5 &= \sigma^1 \otimes \sigma^3 \otimes 1\\ \gamma_6 &= \sigma^2 \otimes 1 \otimes 1\end{aligned} \tag{15.410}$$

for $n = 6$, and so on. So for the Clifford algebra with n generators we have constructed a representation by $2^{[n/2]} \times 2^{[n/2]}$ matrices.

Of course, the above choice of matrices is far from a unique choice of matrices satisfying the Clifford relations (15.405). If the $\gamma_i \in \mathrm{Mat}_d(\mathbb{C})$ then for any $S \in GL(d, \mathbb{C})$ we can

[202] See Chapter 2, section 5.3 for a detailed discussion of tensor product $\otimes$.

change $\gamma_i \to S\gamma_i S^{-1}$. These give *equivalent* representations of the Clifford algebra. We could also modify $\gamma_i \to \epsilon_i \gamma_i$ where $\epsilon_i \in \{\pm 1\}$ and still get a representation, although it might not be an equivalent one.

For example, note that for $n = 3$

$$\gamma_1\gamma_2\gamma_3 = \mathrm{i}\mathbf{1}_{2\times 2} \tag{15.411}$$

and for $n = 5$

$$\gamma_1 \cdots \gamma_5 = -\mathbf{1}_{4\times 4} \tag{15.412}$$

We cannot change the sign on the RHS by conjugating with S. So in this case we conclude that there are at least two inequivalent representations of the Clifford algebra.

The general story, proved in detail in Chapters 11-12 is that

1. If n is even there is a unique irreducible representation of dimension $d = 2^{n/2}$.
2. If n is odd there are precisely two irreducible representation of dimension $d = 2^{(n-1)/2}$ and they are distinguished by the sign of the "Clifford volume element" $\omega = \gamma_1 \cdots \gamma_n$.

In Chapter 11 these statements, and more are generalized to real Clifford algebras for a quadratic form of any signature,

For $w \in \mathbb{Z}_2^n$ (where we will think of $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ as a ring in this example) we define

$$\gamma(w) := \gamma_1^{w_1} \cdots \gamma_n^{w_n} \tag{15.413}$$

Then

$$\gamma(w)\gamma(w') = \epsilon(w, w')\gamma(w + w') = \kappa(w, w')\gamma(w')\gamma(w) \tag{15.414}$$

where

$$\epsilon(w, w') = (-1)^{\sum_{i>j} w_i w'_j} \tag{15.415}$$

defines a cocycle with commutator function

$$\kappa(w, w') = (-1)^{\sum_{i\neq j} w_i w'_j} \tag{15.416}$$

When is the commutator function κ nondegenerate? We need to consider two cases:

Case 1a $\sum_i w_i = 1 \mathrm{mod}\ 2$ and there is an i_0 so that $w_{i_0} = 0$. Then $\gamma(w)$ anticommutes with γ_{i_0}.

Case 1b $\sum_i w_i = 1 \mathrm{mod}\ 2$ and $w_i = 1$ for all i. In this case n must be odd. Then in fact $\gamma_1 \cdots \gamma_n$ commutes with all the γ_i and the commutator function is degenerate.

Case 2. $\sum_i w_i = 0 \mod 2$ and some $w_{i_0} \neq 0$. Then γ_{i_0} anticommutes with $\gamma(w)$.

We conclude that for the even degree Clifford algebras κ is nondegenerate and the group generated by taking products of the matrices $\pm\gamma(w)$ defined by an irreducible representation in fact defines a Heisenberg extension:

$$1 \to \mathbb{Z}_2 \to \mathcal{E}_{2m} \to \mathbb{Z}_2^{2m} \to 1 \tag{15.417}$$

In finite group theory this group is an example of what is known as an "extra-special group" and is denoted $\mathcal{E}_{2m} = 2_+^{1+2m}$. [203]

In the case when n is odd then in fact the Clifford volume form $\gamma_1 \cdots \gamma_n$ commutes with all the γ_i and the cocycle is degenerate.

Example 3: *Heisenberg Construction Of Nontrivial $U(1)$ Bundle Over Symplectic Tori.* Consider a torus $T := \mathbb{R}^n/\Gamma$ where Γ is an integral lattice. Suppose we have an integral-valued symplectic form on Γ. This is a blinear, anti-symmetric, nondegenerate map:

$$\Omega : \Gamma \times \Gamma \to \mathbb{Z} \tag{15.418}$$

It is shown in the Linear Algebra User's Manual that we can choose an ordered basis $\{\gamma_1, \ldots, \gamma_n\}$ for Γ so that the matrix $\Omega(\gamma_i, \gamma_j)$ is of the form

$$\begin{pmatrix} 0 & d_1 \\ -d_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & d_2 \\ -d_2 & 0 \end{pmatrix} \cdots \tag{15.419}$$

where the d_i are integers. We will assume that Ω is nondegenerate so therefore $n = 2m$ is even and all the integer d_i are nonzero.

If Γ is full rank we can extend Ω to a bilinear antisymmetric form on [204]

$$V = \Gamma \otimes_{\mathbb{Z}} \mathbb{R} \cong \mathbb{R}^{2m}$$

to get an antisymmetric bilinear form:

$$\Omega : V \times V \to \mathbb{R} \tag{15.420}$$

If all the d_i are nonzero then this is a symplectic form. Using an invertible matrix S we can bring $S^{tr}\Omega S$ to the standard form J. We can define a commutator function on $\mathbb{R}^n$:

$$\kappa(v_1, v_2) = e^{2\pi \mathrm{i}\Omega(v_1, v_2)} \tag{15.421}$$

and this defines the isomorphism class of a Heisenberg extension of $V \cong \mathbb{R}^{2m}$. To be concrete our extension is

$$1 \to U(1) \to \mathrm{Heis}(V, \Omega) \to V \to 0 \tag{15.422}$$

[203] For any prime p an *extra-special group* is a group G that fits in a central extension $1 \to \mathbb{Z}_p \to G \to \mathbb{Z}_p^n \to 1$ where the center is minimal, that is, is isomorphic to $\mathbb{Z}_p$. For example, for $p = 2$ we have seen that both D_4 and Q are extra-special groups. Up to isomorphism there are two such groups, sometimes denoted $p_\pm^{1+n}$.

[204] The symbol $\otimes_{\mathbb{Z}}$ is explained in the Linear Algebra User's Manual. It means we are taking a tensor product of $\mathbb{Z}$-modules, i.e. Abelian groups.

where we make the explicit choice of cocycle $f(v_1, v_2) = e^{\mathrm{i}\pi\Omega(v_1,v_2)}$. So, Heis$(V, \Omega)$ is the group of pairs (z, v) with $z \in U(1)$ and $v \in V$ with multiplication:

$$(z_1, v_1) \cdot (z_2, v_2) := (z_1 z_2 e^{\mathrm{i}\pi\Omega(v_1,v_2)}, v_1 + v_2) \tag{15.423}$$

We stress that this gives a Heisenberg group and in particular the sequence does not split.

Now consider the pullback of the sequence under the inclusion $\iota : \Gamma \to V$. We claim that the pulled-back sequence splits: Let us try to choose a section

$$s(\gamma) = (\epsilon_\gamma, \gamma) \tag{15.424}$$

then to split the sequence we will need

$$\begin{aligned}(\epsilon_{\gamma_1+\gamma_2}, \gamma_1 + \gamma_2) &= s(\gamma_1 + \gamma_2) \\ &= s(\gamma_1)s(\gamma_2) \\ &= (\epsilon_{\gamma_1}\epsilon_{\gamma_2} e^{\mathrm{i}\pi\Omega(\gamma_1,\gamma_2)}, \gamma_1 + \gamma_2)\end{aligned} \tag{15.425}$$

In other words, to split the sequence over Γ we need to find a function $\epsilon : \Gamma \to U(1)$ so that

$$\epsilon_{\gamma_1}\epsilon_{\gamma_2} = e^{-\mathrm{i}\pi\Omega(\gamma_1,\gamma_2)}\epsilon_{\gamma_1+\gamma_2} \tag{15.426}$$

It is indeed possible to find such functions. See the exercise.

Since the sequence splits over Γ we consider the Abelian subgroup $s(\Gamma) \subset \tilde{V}$. Then define the quotient space:

$$P(T, \Omega, \epsilon) := \mathrm{Heis}(V, \Omega)/s(\Gamma) \tag{15.427}$$

Explicitly $P(T, \Omega, \epsilon)$ is the quotient $(U(1) \times V)/\Gamma$ with the equivalence relation

$$(z, v) \sim (z, v) \cdot (\epsilon_\gamma, \gamma) = (z\epsilon_\gamma e^{\mathrm{i}\pi\Omega(v,\gamma)}, v + \gamma) \tag{15.428}$$

for all $\gamma \in \Gamma$.

Note that there is a continuous map

$$\pi : P(T, \Omega, \epsilon) \to T \tag{15.429}$$

whose fiber is $U(1)$. This space, together with its projection map is an example of a principal $U(1)$ bundle over the torus T: Each fiber is a principal homogeneous space for the group $U(1)$, under the natural action of $U(1)$ on $P(T, \Omega, \epsilon)$. (Since the $U(1)$ is central we can consider it either as a left- or right- action.) Our construction of the bundle depended on a choice of splitting ϵ_γ, but a change of splitting defines isomorphic bundles.

Remarks

1. The above construction comes up, either explicitly or implicitly in discussions of the quantum Hall efect, Chern-Simons theory, and the quantization of p-form gauge theories.

2. Note that while $s(\Gamma)$ is a subgroup of $\mathrm{Heis}(V,\Omega)$ it is <u>not</u> a normal subgroup, so that, while, $P(T,\Omega,\epsilon)$ is a bundle, it is not a group.

3. If we view $\mathrm{Heis}(V,\Omega)$ as a principal $U(1)$ bundle over V then we can construct a very natural connection on this bundle. Parallel transport of the point $(z,v_0)\in P(T,\Omega,\epsilon)$ over a straightline path $\wp_{v_0,w}:=\{v_0+tw|0\le t\le 1\}$ in V is defined by left-multiplication by $(1,w)$:

$$U(\wp_{v_0,w}):(z,v_0)\to(1,w)\cdot(z,v_0)=(ze^{\mathrm{i}\pi\Omega(w,v_0)},w+v_0) \tag{15.430}$$

Now, if one considers parallel transport around a small square loop starting at v_0 by composing paths

$$\wp_{v,w_1,w_2}:=\wp_{v_0,w_1}\star\wp_{v_0+w_1,w_2}*\wp_{v_0+w_1+w_2,-w_1}*\wp_{v_0+w_2,-w_2} \tag{15.431}$$

one obtains the holonomy

$$U(\wp_{v_0,w_1,w_2}):(z,v_0)\to(ze^{2\pi\mathrm{i}\Omega(w_1,w_2)},v_0) \tag{15.432}$$

showing that the curvature of this connection is Ω regarded as a 2-form on V. The connection and 2-form descend to the principal $U(1)$ bundle $P(T,\Omega,\epsilon)$. The cohomology class of $[\Omega]$, which is the first Chern class of $P(T,\Omega,\epsilon)$ is characterized by the integers $\vec{d}=(d_1,d_2,\dots)$. A nonzero value of $\vec{d}$ obstructs topological triviality.

Exercise

We illustrated how Q and D_4 are the only two non-Abelian groups that sit in an extension of $\mathbb{Z}_2\times\mathbb{Z}_2$ by $\mathbb{Z}_2$. Which one is the Heisenberg extension?

Exercise *Finite Heisenberg group in multiplicative notation*

It is interesting to look at the Heisenberg extension

$$1\to\mathbb{Z}_n\to\mathrm{Heis}(\mathbb{Z}_n\times\mathbb{Z}_n)\to\mathbb{Z}_n\times\mathbb{Z}_n\to 1 \tag{15.433}$$

where we think of $\mathbb{Z}_n$ as the *multiplicative* group of nth roots of unity. Let $\omega=\exp[2\pi\mathrm{i}/n]$. We distinguish the three $\mathbb{Z}_n$ factors by writing generators as $\omega_1,\omega_2,\omega_3$.

a.) Show that one natural choice of cocycle is:

$$f\left((\omega_1^s,\omega_2^t),(\omega_1^{s'},\omega_2^{t'})\right):=\omega_3^{st'} \tag{15.434}$$

b.) Compute the commutator function

$$\kappa\left((\omega_1^s,\omega_2^t),(\omega_1^{s'},\omega_2^{t'})\right):=\omega_3^{st'-ts'} \tag{15.435}$$

c.) Connect to our general theory of extensions by defining $U := (1,(\omega_1,1))$, $V := (1,(1,\omega_2))$ and computing

$$\begin{aligned} UV &= (f((\omega_1,1),(1,\omega_2)),(\omega_1,\omega_2)) \\ &= (\omega_3,(\omega_1,\omega_2)) \\ VU &= (f((1,\omega_2),(\omega_1,1)),(\omega_1,\omega_2)) \\ &= (1,(\omega_1,\omega_2)) \end{aligned} \tag{15.436}$$

or in other words, since the center is generated by $q = (\omega_3,(1,1))$ we can write:

$$UV = qVU \tag{15.437}$$

Exercise *Degenerate Heisenberg extensions*

Suppose $n = km$ is composite and suppose we use the function $c_k(a,b') = kab'$ in defining an extension of $\mathbb{Z}_n \times \mathbb{Z}_n$.

a.) Show that the commutator function is now degenerate.

b.) Show that the center of the central extension is larger than $\mathbb{Z}_n$. Compute it. [205]

While these are not - strictly speaking - Heisenberg extensions people will often refer to them as Heisenberg extensions. We might call them "degenerate Heisenberg extensions."

Exercise *Constructing The Splitting* (15.426)

Give an explicit construction of a function $\epsilon : \Gamma \to U(1)$ satisfying (15.426). [206]

Exercise *Different Splittings Give Isomorphic Bundles*

a.) Describe the relation between two splittings of the pullback of the sequence (15.422) to Γ.

b.) Let ϵ_1, ϵ_2 denote two splittings of the pullback of the sequence (15.422) to Γ. Show that the bundles $P(T,\Omega,\epsilon_1)$ with $P(T,\Omega,\epsilon_2)$.

[205] *Answer*: The center is generated by q, U^m, V^m and is $\mathbb{Z}_n \times \mathbb{Z}_k \times \mathbb{Z}_k$.

[206] *Answer*: Choose an ordered basis $\gamma_1,\ldots,\gamma_n$ for Γ and define

$$\epsilon_\gamma := e^{-i\pi \sum_{i<j} n_i n_j \Omega_{ij}} \tag{15.438}$$

where $\gamma = \sum_i n_i\gamma_i$ and $\Omega_{ij} = \Omega(\gamma_i,\gamma_j)$. One can check this satisfies the desired identity. Note that it is crucial that Ω_{ij} and n_i are integral.

Exercise *Two Dimensions*
a.) Suppose $T = \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ with $\mathrm{Im}\tau > 0$. So $\Gamma = \mathbb{Z} + \tau\mathbb{Z}$. Choose

$$\Omega(1,\tau) = -\Omega(\tau,1) = k \in \mathbb{Z}. \tag{15.439}$$

b.) Show that

$$\Omega(z_1,z_2) = k\frac{\mathrm{Im}(\bar{z}_1 z_2)}{\mathrm{Im}\tau} \tag{15.440}$$

15.5.7 Lagrangian Subgroups And Induced Representations

Let us compare the general Heisenberg extension

$$1 \to \mathcal{Z} \to \tilde{G} \to G \to 0 \tag{15.441}$$

with (15.363) and (15.403) with $\mathcal{Z}$ any subgroup of $U(1)$. The difference from the general case is that in these examples G is explicitly presented as a product of subgroups $G = L \times L'$ where L and L' are *maximal Lagrangian subgroups*. A subgroup $L \subset G$ is said to be a *Lagrangian subgroup* if $\kappa(g_1,g_2) = 1$ for all pairs $(g_1,g_2) \in L$ and similarly for L'.

When discussing Heisenberg groups of the form $\mathrm{Heis}(S \times \widehat{S})$ the group $S \times \widehat{S}$ has two canonical Lagrangian subgroups, namely S and $\widehat{S}$. With the choice of κ we made above these are maximal Lagrangian subgroups.

The case of $G = S \times \widehat{S}$ should be contrasted with other examples where G is $\mathbb{R}^{2n}$ or $\mathbb{Z}_2^{2n}$. These groups certainly can be presented as products of maximal Lagrangian subgroups, but there is no canonical decomposition. Consider, for example the Heisenberg extension $\mathrm{Heis}(\mathbb{Z}_2^{2n})$ constructed using the gamma-matrices. Note that $\mathbb{Z}_2 = \mathbb{F}_2$ is a field, and we can consider $\mathbb{F}_2^{2n}$ to be a vector space over $\kappa = \mathbb{F}_2$. We could take L to be any half-dimensional Lagrangian subspace.

In the general situation, with no canonical choice of L one would often like to construct an explicit unitary representation of the Heisenberg group. One way to do this is the following:

Choose a maximal Lagrangian subgroup $L \subset G$. The inverse image $\tilde{L} \subset \tilde{G}$ is a maximal commutative subgroup of $\tilde{G}$.

We now choose a character of $\tilde{L}$ such that $\rho(z,x) = z\rho(x)$. Note that such a character must satisfy

$$\rho(x)\rho(x') = f(x,x')\rho(x+x') \qquad \forall x,x' \in L \tag{15.442}$$

Note that f need not be trivial on L, but it does define an Abelian extension $\tilde{L}$ which must therefore be isomorphic to a product $L \times U(1)$, albeit noncanonically. Different choices of ρ are different choices of splitting. Indeed note that (15.442) says that, when restricted to $L \times L$ the cocycle is trivialized by ρ.

♣This is bad notation, because $\rho(z,x)$ is later used for the full SvN representation. The letter here should be changed and it propagates through the discussion. ♣

The carrier space of our representation will be the space $\mathcal{F}$ of functions $\psi : \tilde{G} \to \mathbb{C}$ such that

$$\psi\left((z,x)(z',x')\right) = \rho(z',x')^{-1}\psi\left(z,x\right) \qquad \forall (z',x') \in \tilde{L} \tag{15.443}$$

Setting $(z',x') = (z^{-1},0)$ we note that equation (15.443) implies $\psi(z,x) = z^{-1}\psi(1,x)$, so defining $\Psi(x) := \psi(1,x)$ we can simplify the description of $\mathcal{F}$ by identifying it with the space of functions $\Psi : G \to \mathbb{C}$ such that:

$$\Psi\left(x + x'\right) = \frac{f(x,x')}{\rho(x')}\Psi(x) \qquad \forall x' \in L. \tag{15.444}$$

If our group is continuous or noncompact we should state an L^2 condition. We take ρ to be a unitary character so that $|\Psi(x)|^2$ descends to a function on G/L and we demand:

$$\int_{G/L} |\Psi(x)|^2 dx < +\infty. \tag{15.445}$$

The group action is simply left-action of $\tilde{G}$ on the functions $\psi(z,x)$. When written in terms of $\Psi(x)$ the representation of $(x,z) \in \tilde{G}$ is:

$$\begin{aligned}(\rho(z,x)\cdot\Psi)(y) &= (\rho(z,x)\cdot\psi)(1,y)\\ &= \psi\left((z,x)^{-1}\cdot(1,y)\right)\\ &= \psi\left((z^{-1}f(x,-x)^{-1}f(-x,y), y-x)\right)\\ &= zf(x,y-x)\Psi(y-x)\end{aligned} \tag{15.446}$$

where in the last line we assumed that f is a normalized cocycle so that $f(0,y) = 1$.

Let us see how we recover the standard Stone-von Neumann representation of Heis$(S \times \widehat{S})$ from this viewpoint. Let us choose $L = \widehat{S}$. Then the equivariance condition (15.444) becomes

$$\Psi(s,\chi\chi') = \frac{1}{\rho(\chi')}\Psi(s,\chi) \tag{15.447}$$

Now set $\chi' = 1/\chi$ and conclude that

$$\Psi(s,\chi) = \frac{1}{\rho(\chi)}\Psi(s,1) \tag{15.448}$$

So the dependence on $\chi \in \widehat{S}$ is completely fixed by equivariance. Defining $\tilde{\psi}(s) := \Psi(s,1)$ we obtain a vector $\tilde{\psi} \in L^2(S)$, and thus if we take $L = \widehat{S}$ then our space of equivariant functions is naturally identified with $L^2(S)$.

We can now work out the representations of

$$\begin{aligned}T_s &= \rho(1,(s,1))\\ M_\chi &= \rho(1,(0,\chi))\end{aligned} \tag{15.449}$$

on $L^2(S)$. Working through the above definitions it should not be surprising that one recovers:

$$\begin{aligned}(T_s\tilde{\psi})(s_0) &= \tilde{\psi}(s_0 - s)\\ (M_\chi\tilde{\psi})(s_0) &= \frac{\rho(\chi)}{\chi(s_0)}\tilde{\psi}(s_0)\end{aligned} \tag{15.450}$$

Example. Consider $\mathbb{F}_2^{2m}$ with $\kappa(w,w') = (-1)^{\sum_{i\neq j} w_i w'_j}$. We can give a Lagrangian decomposition

$$\mathbb{F}_2^{2m} \cong L \oplus N \tag{15.451}$$

in many different ways. For special values of m there are special Lagrangian subspaces provided by classical error correcting codes. A Heisenberg representation can be given by taking V to be the space of functions $\psi : L \to \mathbb{C}$. This has complex dimension 2^m. N can be identified with the group of characters on L since we can set

$$\chi_n(\ell) = \kappa(n,\ell) \tag{15.452}$$

The usual translation and multiplication operators T_ℓ and M_n generate an algebra isomorphic to $Mat_d(\mathbb{C})$. V is also a representation of the Clifford algebra (and hence the extra-special group). So the Clifford representation matrices γ_i can be expressed in terms of these, and vice versa.

Remarks

1. The representation is, geometrically, just the space of L^2-sections of the associated line bundle $\tilde{G} \times_{\tilde{L}} \mathbb{C}$ defined by ρ. The representation is independent of the choice of ρ, and any two choices are related by an automorphism of L given by the restriction of an inner automorphism of $\tilde{G}$.

2. This is an example of an *induced representation*, $Ind_{\tilde{L}}^{\tilde{G}}(\mathbb{C})$ which we will study more systematically in chapter 4.

Exercise

Construct explicit Lagrangian subspaces of $\mathbb{F}_2^{2m}$ for small values of m and write out the matrices of the Heisenberg representation. [207]

15.5.8 Automorphisms Of Heisenberg Extensions

In several of our examples above, such as $\mathrm{Heis}(V,\Omega)$ and the extraspecial group we have noted that to define a representation one must make a choice of a Lagrangian subgroup, where κ restricts to 1. On the other hand, we also noted that in some situations there is no natural choice of such a Lagrangian group. There are many such Lagrangian subgroups and they are related by "symplectic automorphisms" of G.

We say that *an automorphism $\alpha \in \mathrm{Aut}(G)$ is symplectic* if it preserves the commutator function:

$$\begin{aligned} \alpha^*\kappa(g_1,g_2) &:= \kappa(\alpha(g_1),\alpha(g_2)) \\ &= \kappa(g_1,g_2) \end{aligned} \tag{15.453}$$

[207] *Answer.* Start with $m = 1$. Any vector is isotropic so let L be spanned by $\ell = (1,0)$ and N spanned by $n = (0,1)$. Choose a basis for V by taking the delta function supported on basis vector e_i. Then $T_\ell = \sigma^1$ and $M_n = \sigma_3$ relative to this basis.********** CONTINUE **********.

In the first line we defined the general notion of pullback $\kappa \to \alpha^*\kappa$ and the second line is the invariance condition. In physics, such symplectic transformations are relevant to canonical transformations. We can ask whether such automorphisms of "phase space" actually lift to automorphisms of the full Heisenberg group, and then whether and how this lifted group acts on the representations of the Heisenberg group. This would be the "quantum mechanical implementation of symplectic transformations." In this section we will investigate those questions from the group-theoretical viewpoint.

Quite generally, (there is no need for G to be Abelian in this paragraph), if $\pi : \widetilde{G} \to G$ is a homomorphism and $\alpha \in \mathrm{Aut}(G)$ is an automorphism of G we say that α *lifts to an automorphism of* $\widetilde{G}$ if there is an automorphism $\widetilde{\alpha} \in \mathrm{Aut}(\widetilde{G})$ that completes the diagram:

$$\begin{array}{ccc} \widetilde{G} & \xrightarrow{\pi} & G \\ \vdots & & \\ \downarrow \widetilde{\alpha} & & \downarrow \alpha \\ \widetilde{G} & \xrightarrow{\pi} & G \end{array} \tag{15.454}$$

Or, in equations, for every $\alpha \in \mathrm{Aut}(G)$ we seek a corresponding $\widetilde{\alpha} \in \mathrm{Aut}(\widetilde{G})$ such that

$$\pi(\widetilde{\alpha}(\widetilde{g})) = \alpha(\pi(\widetilde{g})) \tag{15.455}$$

for all $\widetilde{g} \in \widetilde{G}$.

Some General Theory [208]
Consider a group extension

$$1 \to A \to \widetilde{G} \to G \to 1 \tag{15.456}$$

where A and G are Abelian and we write the group operations on both A and G additively so $\widetilde{G}$ is the group of pairs (a, g) with group multiplication

$$(a_1, g_1)(a_2, g_2) = (a_1 + a_2 + f(g_1, g_2), g_1 + g_2) \tag{15.457}$$

and $f(g_1, g_2)$ is a cocycle satisfying the additive version of the cocycle identity:

$$f(g_1 + g_2, g_3) + f(g_1, g_2) = f(g_1, g_2 + g_3) + f(g_2, g_3) \tag{15.458}$$

Now suppose $\alpha \in \mathrm{Aut}(G)$. If α preserves the commutator function then we can hope to lift it to an automorphism T_α of $\widetilde{G}$. As explained above, "lifting" means that

$$\pi(T_\alpha(a, g)) = \alpha(\pi(a, g)) = \alpha(g) \ . \tag{15.459}$$

Therefore, T_α must be of the form:

$$T_\alpha(a, g) = (\xi_\alpha(a, g), \alpha(g)) \tag{15.460}$$

[208] What follows is an elaboration of the ideas from Appendix A of arXiv:1707.08888. I also got some useful help from Graeme Segal.

where ξ_α is some function $\xi_\alpha : A \times G \to A$. We can write constraints on this function from the requirement that T_α must be an automorphism with the group law defined by the cocycle f. In particular T_α must be a group homomorphism:

$$T_\alpha((a_1, g_1), (a_2, g_2)) = T_\alpha(a_1, g_1) \cdot T_\alpha(a_2, g_2) \tag{15.461}$$

which is true iff

$$\xi_\alpha(a_1 + a_2 + f(g_1, g_2), g_1 + g_2) = \xi_\alpha(a_1, g_1) + \xi_\alpha(a_2, g_2) + f(\alpha(g_1), \alpha(g_2)) \tag{15.462}$$

Now, specialize this equation by putting $g_1 = 0$ and assuming (WLOG) that we have a normalized cocycle, so that $f(g, 0) = f(0, g) = 0$. Then equation (15.462) simplifies to

$$\xi_\alpha(a_1 + a_2, g) = \xi_\alpha(a_1, 0) + \xi_\alpha(a_2, g) \tag{15.463}$$

Putting $a_2 = 0$ in (15.463) we now learn that

$$\xi_\alpha(a, g) = \xi_\alpha(a, 0) + \xi_\alpha(0, g) \tag{15.464}$$

On the other hand, putting $g = 0$ in (15.463) we now learn that $a \mapsto \xi_\alpha(a, 0)$ is just an automorphism of A. Composing lifts of α with such automorphisms is an inherent ambiguity in lifting α. Thus, it is useful to make the simplifying assumption that $\xi_\alpha(a, 0) = a$, since we can always arrange this by composition with an automorphism of A. Therefore we can write equation (15.464) in the general form

$$T_\alpha(a, g) = (a + \tau_\alpha(g), \alpha(g)) \tag{15.465}$$

If we restrict to automorphisms of the type (15.465) then one easily checks that T_α is indeed a group homomorphism iff

$$(\alpha^* f - f)(g_1, g_2) = \tau_\alpha(g_1 + g_2) - \tau_\alpha(g_1) - \tau_\alpha(g_2) \tag{15.466}$$

where

$$\alpha^* f(g_1, g_2) := f(\alpha(g_1), \alpha(g_2)) \tag{15.467}$$

is known as the "pulled-back cocycle." The conceptual meaning of this equation is the following: $\alpha \in \mathrm{Aut}(G)$ is a "symmetry of G." We are asking how badly the cocycle f breaks that symmetry. We say that "f is invariant under pullback" if $\alpha^* f = f$ and in that case we can take $\tau_\alpha = 0$ and we can easily lift the group $\mathrm{Aut}(G)$ to a group of automorphisms of $\widetilde{G}$. The more general condition (15.466) says that the amount by which f is not symmetric under α, that is $\alpha^* f - f$, must be a trivializable cocycle. Put this way, it is clear that the condition is unchanged under shifting f by a coboundary. Indeed, if we change f by a coboundary so

$$\tilde{f}(g_1, g_2) = f(g_1, g_2) + b(g_1 + g_2) - b(g_1) - b(g_2) \tag{15.468}$$

then we can solve (15.466) by taking

$$\tilde{\tau}_\alpha(g) = \tau_\alpha(g) + (\alpha^* b - b)(g) \tag{15.469}$$

so the existence of a solution to (15.466) is gauge invariant. Put more simply, *the cohomology class* $[f] \in H^2(G,A)$ *must be invariant under the action of* α^* *on* $H^2(G,A)$.

♣The above paragraph should be said more succinctly. ♣

Thus, in general we cannot lift all automorphisms of G, only those for which (15.466) holds. The set of such automorphisms forms a subgroup of $\mathrm{Aut}(G)$ that we will denote as $\mathrm{Aut}_0(G)$. Note that, in our additive notation we have

$$\kappa(g_1,g_2) = f(g_1,g_2) - f(g_2,g_1) \tag{15.470}$$

which, as mentioned above, is a generalization of a symplectic form. Any automorphism satisfying (15.466) automatically satisfies $\alpha^*(\kappa) = \kappa$ and is therefore a "symplectic automorphism." However, $\mathrm{Aut}_0(G)$ is in general a subgroup of the group of symplectic automorphisms of G. We will give an example with $G = SL(2,\mathbb{Z}_n)$ below.

We now restrict attention to $\alpha \in \mathrm{Aut}_0(G)$. If, in addition

$$\tau_{\alpha_1\circ\alpha_2}(g) = \tau_{\alpha_1}(\alpha_2(g)) + \tau_{\alpha_2}(g) \tag{15.471}$$

that is,

$$\tau_{\alpha_1\circ\alpha_2} = \alpha_2^*\tau_{\alpha_1} + \tau_{\alpha_2} \tag{15.472}$$

then in fact $T_{\alpha_1}\circ T_{\alpha_2} = T_{\alpha_1\circ\alpha_2}$ generate a subgroup of $\mathrm{Aut}(\widetilde{G})$ isomorphic to $\mathrm{Aut}(G)$.

♣ Relate this to equation (15.542)) the condition for a twisted homomorphism $\tau : \mathrm{Aut}(G) \to A$. ♣

In general, even if we can find a solution to (15.466) the criterion (15.472) will not hold. Nevertheless, the automorphisms T_α will generate a subgroup of $\mathrm{Aut}(\widetilde{G})$. To see what subgroup it is we introduce, for $\ell \in \mathrm{Hom}(G,A)$, the automorphism

$$P_\ell(a,g) = (a + \ell(g), g) \tag{15.473}$$

Then we note that

$$T_{\alpha_1}\circ T_{\alpha_2}(a,g) = T_{\alpha_1\circ\alpha_2}\circ P_{\ell_{\alpha_1,\alpha_2}} \tag{15.474}$$

where

$$\ell_{\alpha_1,\alpha_2}(g) := \tau_{\alpha_1}(\alpha_2(g)) + \tau_{\alpha_2}(g) - \tau_{\alpha_1\circ\alpha_2}(g) \tag{15.475}$$

A little computation shows that $\ell_{\alpha_1,\alpha_2} \in \mathrm{Hom}(G,A)$ is indeed a homomorphism from G to A. A little more computation reveals

$$\begin{aligned} P_{\ell_1}\circ P_{\ell_2} &= P_{\ell_1+\ell_2} \\ T_\alpha \circ P_\ell &= P_{\alpha^*(\ell)}\circ T_\alpha \end{aligned} \tag{15.476}$$

Therefore, we can write any word in P's and T's in the form $P_{\ell'}\circ T_{\alpha'}$ for some (ℓ',α'). Altogether, equations (15.474), (15.475), and (15.476) mean that T_α generate a subgroup of $\mathrm{Aut}(\widetilde{G})$. Including all transformations P_ℓ defines a subgroup $\widetilde{\mathrm{Aut}(G)}$ of $\mathrm{Aut}(\widetilde{G})$ which fits in an exact sequence:

$$1 \to \mathrm{Hom}(G,A) \to \widetilde{\mathrm{Aut}(G)} \to \mathrm{Aut}_0(G) \to 1 \tag{15.477}$$

Finally, restoring $\mathrm{Hom}(G,A) \to \mathrm{Hom}(G,A) \rtimes \mathrm{Aut}(A)$ gives the group of (possible) lifts of automorphisms of G to automorphisms of $\tilde{G}$.

Example: Heis($\mathbb{R} \oplus \mathbb{R}$)

Let us consider the basic example from quantum mechanics based on a phase space $\mathbb{R} \oplus \mathbb{R}$ with symplectic form defined by J. In this case there is a very nice way of thinking of the matrix group $Sp(2, \mathbb{R})$. We note that if

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{15.478}$$

then

$$\begin{aligned} A^{tr} J A &= \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= (ad - bc) J \end{aligned} \tag{15.479}$$

Therefore, $A \in Sp(2, \mathbb{R})$ iff $ad - bc = 1$. But this is precisely the condition that defines $SL(2, \mathbb{R})$. Therefore

$$SL(2, \mathbb{R}) = Sp(2, \mathbb{R}) \tag{15.480}$$

are identical as matrix groups. The same argument applies if we replace $\mathbb{R}$ by any ring R. This kind of isomorphism is definitely not true if we consider higher rank groups $SL(n, \mathbb{R})$ and $Sp(2n, \mathbb{R})$.

Now, as we have seen, the section $s(\alpha, \beta) = \exp[\mathrm{i}(\alpha \hat{p} + \beta \hat{q})]$ leads to the choice of section, written additively

$$f\left((\alpha_1, \beta_1), (\alpha_2, \beta_2)\right) = \frac{\hbar}{2}(\alpha_1 \beta_2 - \alpha_2 \beta_1) \tag{15.481}$$

This is symplectic invariant, so we can take $\tau_\alpha(g) = 0$ in the equation (15.466) and we conclude that the symplectic group acts as a group of automorphisms on the Heisenberg group Heis($\mathbb{R}^{2n}$).

It is interesting however, that the symplectic group only acts projectively on the Stone-von-Neumann representation of the Heisenberg group. We now explain this point.

As we will discuss in detail in the chapter on Lie groups, $SL(2, \mathbb{R})$ and $Sp(2, \mathbb{R})$ are examples of Lie groups. It is useful to look at group elements infinitesimally close to the identity matrix. These can be written as

$$A = 1 + \epsilon m + \mathcal{O}(\epsilon^2) \tag{15.482}$$

We learn from the defining equation that

$$\mathrm{Tr}(m) = 0 \tag{15.483}$$

is required to satisfy the defining conditions to order ϵ. (Exercise: Prove this using both the definition of $Sp(2, \mathbb{R})$ and of $SL(2, \mathbb{R})$.)

The infinitesimal group elements are thus characterized by the vector space $\mathfrak{sp}(2, \mathbb{R}) = \mathfrak{sl}(2, \mathbb{R})$, which is the vector space of 2×2 real traceless matrices. These form a Lie algebra with the standard matrix commutator. Generic (but not all) group elements are obtained

by exponentiating such matrices. In particular, in a neighborhood of the identity all group elements are obtained by exponentiating elements of the Lie algebra.

Recall the standard basis of $\mathfrak{sl}(2,\mathbb{R})$:

$$e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad h = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad f = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \tag{15.484}$$

(Check the signs carefully!) Compute:

$$[h,e] = -2e \qquad [e,f] = h \qquad [h,f] = +2f \tag{15.485}$$

From this one can in principle multiply exponentiated matrices using the BCH formula.

We now consider the quantum implementation of these operators on $L^2(\mathbb{R})$ with $\rho(e) = \hat{e}$ etc. with:

$$\hat{e} := \frac{\mathrm{i}}{2\hbar}\hat{p}^2 \qquad \hat{h} := \frac{\mathrm{i}}{2\hbar}(\hat{q}\hat{p} + \hat{p}\hat{q}) \qquad \hat{f} := \frac{\mathrm{i}}{2\hbar}\hat{q}^2 \tag{15.486}$$

Now, using the useful identities: [209]

$$\begin{aligned} [AB,CD] &= A[B,C]D + [A,C]BD + CA[B,D] + C[A,D]B \\ &= AC[B,D] + A[B,C]D + C[A,D]B + [A,C]DB \end{aligned} \tag{15.488}$$

we can check that this is indeed a representation:

$$[\hat{h},\hat{e}] = -2\hat{e} \qquad [\hat{e},\hat{f}] = \hat{h} \qquad [\hat{h},\hat{f}] = +2\hat{f} \tag{15.489}$$

However, we have to be careful about exponentiating these operators.

Let us consider the one-parameter subgroup of $SL(2,\mathbb{R})$:

$$\exp[\theta(e+f)] = \cos\theta + \sin\theta \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{15.490}$$

This is in fact a maximal compact subgroup of $SL(2,\mathbb{R})$. (See chapter **** on 2x2 matrix groups.) Note carefully that it has period $\theta \sim \theta + 2\pi$.

The quantum implementation of $e+f$ is just the standard harmonic oscillator Hamiltonian!

$$\hat{e} + \hat{f} = \frac{\mathrm{i}}{2}(\hat{p}^2 + \hat{q}^2) = \mathrm{i}(\bar{a}a + \frac{1}{2}) \tag{15.491}$$

where

$$\begin{aligned} a &= \frac{1}{\sqrt{2}}(q + \mathrm{i}p) \\ \bar{a} &= \frac{1}{\sqrt{2}}(q - \mathrm{i}p) \end{aligned} \tag{15.492}$$

[209] These identities are very useful, but a bit hard to remember. If you are on a desert island you can easily reconstruct them from the special cases:

$$\begin{aligned} [AB,C] &= A[B,C] + [A,C]B \\ [A,CD] &= C[A,D] + [A,C]D \end{aligned} \tag{15.487}$$

both of which are easy to remember.

Now, in the Stone-von-Neumann representation $\frac{\mathrm{i}}{2}(\hat{p}^2+\hat{q}^2)$ has the spectrum $\mathrm{i}(n+\frac{1}{2})$, $n=0,1,2,\ldots$. Therefore, the one-parameter subgroup $\exp[\theta(\hat{e}+\hat{f})]$ has period $\theta\sim\theta+4\pi$. We see that the group generated by $\hat{e},\hat{f},\hat{h}$ is at least a double cover of $Sp(2,\mathbb{R})$. In fact, it turns out to be exactly a double cover, and it is known as the *metaplectic group*.

One very interesting aspect of the metaplectic group is that this is a Lie group with no finite-dimensional faithful representation. We now explain that fact, [210] and a few other important things in the following remarks:

Remarks

1. *Application to the metaplectic group.* The Lie algebra of the metaplectic group is $\mathfrak{sl}(2,\mathbb{R})$. Any finite dimensional representation of the metaplectic group would give a finite-dimensional representation of the Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ and we discussed in detail what these are above. Note that $e+f$ corresponds to $\mathrm{i}J^1$. So: In any finite-dimensional complex representation of $\mathfrak{sl}(2,\mathbb{R})$, the operator $\rho(e)+\rho(f)$ is diagonalizable and all the eigenvalues are of the form $\mathrm{i}\ell$, where ℓ is an integer. So $\exp[\theta\rho(e+f)]$ has period $\theta\sim\theta+2\pi$. Therefore no finite dimensional representation of Mpl can be faithful. In particular, Mpl is an example of a Lie group which is not a matrix group: It cannot be embedded as a subgroup of $GL(N,\mathbb{C})$ for any N.

2. It is very interesting to consider the action of the one-parameter family $\exp[\theta(\hat{e}+\hat{f})]$ in the standard "position space" representation $L^2(\mathbb{R})$ with $\hat{p}=-\mathrm{i}\frac{d}{dq}$. Let us compute the integral kernel:
$$\langle x|\exp[\theta(\hat{e}+\hat{f})]|y\rangle \tag{15.493}$$
since
$$\left(e^{\theta(\hat{e}+\hat{f})}\psi\right)(x)=\int_{-\infty}^{+\infty}\langle x|\exp[\theta(\hat{e}+\hat{f})]|y\rangle\psi(y)dy \tag{15.494}$$
Clearly, to evaluate (15.493) we should insert a complete set of eigenstates of the harmonic oscillator Hamiltonian. So we introduce the *Hermite functions*: [211]
$$\psi_n(x):=(2^n n!\sqrt{\pi})^{-1/2}e^{-\frac{x^2}{2}}H_n(x)\qquad n\in\mathbb{Z}_+ \tag{15.495}$$
where $H_n(x)$ is the Hermite polynomial
$$H_n(x)=e^{x^2}\left(-\frac{d}{dx}\right)^n e^{-x^2} \tag{15.496}$$
The Hermite functions $\psi_n(x)$ satisfy
$$\left(-\frac{d^2}{dx^2}+x^2\right)\psi_n=(2n+1)\psi_n \tag{15.497}$$

[210] Our demonstration of this surprising fact follows the discussion in G. Segal in *Lectures On Lie Groups and Lie Algebras.*

[211] What follows here are completely standard facts. We used Wikipedia.

and we have normalized them so that so that

$$\int_{-\infty}^{+\infty} \psi_n(x)\psi_m(x)dx = \delta_{n,m} \tag{15.498}$$

Now we have

$$\sum_{n=0}^{\infty} u^n \psi_n(x)\psi_m(x) = \frac{1}{\sqrt{\pi(1-u^2)}} \exp\left[-\frac{1-u}{1+u}\left(\frac{x+y}{2}\right)^2 - \frac{1+u}{1-u}\left(\frac{x-y}{2}\right)^2\right] \tag{15.499}$$

To prove this write

$$H_n(x) = e^{x^2}\left(-\frac{d}{dx}\right)^n e^{-x^2} = e^{x^2}\left(-\frac{d}{dx}\right)^n \left(\frac{1}{\sqrt{4\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{4}s^2 + \mathrm{i}sx} ds\right) \tag{15.500}$$

Apply this to both $\psi_n(x)$ and $\psi_n(y)$ apply the derivatives, exchange integration and sum and get

$$\begin{aligned}\sum_{n=0}^{\infty} u^n \psi_n(x)\psi_m(x) &= \frac{e^{\frac{1}{2}(x^2+y^2)}}{4\pi^{3/2}} \int dsdt e^{-\frac{1}{4}(s^2+t^2) - \frac{1}{2}stu + \mathrm{i}sx + \mathrm{i}ty} \\ &= \frac{1}{\sqrt{\pi(1-u^2)}} \exp\left[-\frac{1-u}{1+u}\left(\frac{x+y}{2}\right)^2 - \frac{1+u}{1-u}\left(\frac{x-y}{2}\right)^2\right]\end{aligned} \tag{15.501}$$

Now we have

$$\langle x|\exp[\theta(\hat{e}+\hat{f})]|y\rangle = e^{\mathrm{i}\theta/2}\sum_{n=0}^{\infty} e^{\mathrm{i}n\theta}\psi_n(x)\psi_n(y) \tag{15.502}$$

so we apply the above identity with $u = e^{\mathrm{i}\theta}$. We should be careful about convergence: The Gaussian integral in (15.501) has quadratic form

$$A = \frac{1}{4}\begin{pmatrix} 1 & u \\ u & 1 \end{pmatrix} \tag{15.503}$$

which has eigenvalues $\frac{1}{4}(1 \pm u)$. The zero-modes at $u = \pm 1$ indicate a divergent Gaussian. In fact we have

$$\begin{aligned}\sum_{n=0}^{\infty} \psi_n(x)\psi_n(y) &= \delta(x-y) \\ \sum_{n=0}^{\infty} (-1)^n \psi_n(x)\psi_n(y) &= \delta(x+y)\end{aligned} \tag{15.504}$$

The second line follows easily from the first since the parity of $\psi_n(x)$ as a function of x is the parity of n.

The quadratic form A has a positive definite real part for $|u| \leq 1$ except for $u = \pm 1$. The values for $|u| > 1$ have to be defined by analytic continuation and there is a branch point at $u = \pm 1$. Note that at $\theta = \pi/2$ we have

$$\langle x|\exp[\frac{\pi}{2}(\hat{e}+\hat{f})]|y\rangle = \frac{e^{\mathrm{i}\pi/4}}{\sqrt{2\pi}} e^{\mathrm{i}xy} \tag{15.505}$$

and we recognize the kernel for the Fourier transform:

$$e^{\frac{\pi}{2}(\hat{e}+\hat{f})}\psi = e^{\mathrm{i}\pi/4}\mathcal{F}(\psi) \tag{15.506}$$

where $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is the Fourier transform. This is the quantum implementation of the canonical transformation exchanging position and momenta. Note that $\theta = \pi$ is the square of the Fourier transform (up to a scalar multiplication by i) but this is not a scalar operator on the space of functions. Indeed at $\theta = \pi$ we have

$$\left(e^{\pi(\hat{e}+\hat{f})}\psi\right)(x) = \mathrm{i}\psi(-x) \tag{15.507}$$

The Fourier transform is of order four not order two. Also note that at $\theta = 2\pi$ the operator is just multiplication by -1.

3. We note that there are beautiful general formulae for expectation value operators defined by exponentiating general quadratic forms in the $\hat{p}_i$ and $\hat{q}^i$, or, equivalently in a's and $a^\dagger$'s. This is useful when working with coherent states and squeezed states. But it is best presented in the Bargmann or geometric quantization formalism. ♣say more? ♣

Exercise

a.) Check that, for 2×2 matrices the condition $\mathrm{Tr}(m) = 0$ is identical to the condition $(mJ)^{tr} = mJ$.

b.) Show that for any $n \times n$ matrix, the infinitesimal version of the condition $\det A = 1$ is that $A = 1 + \epsilon m + \mathcal{O}(\epsilon^2)$ with $\mathrm{Tr}(m) = 0$.

c.) Show that for any $n \times n$ matrix, the infinitesimal version of the condition $A^{tr}JA = J$ is that $A = 1 + \epsilon m + \mathcal{O}(\epsilon^2)$ with $(mJ)^{tr} = mJ$.

d.) Show that the conditions $\mathrm{Tr}(m) = 0$ and $(mJ)^{tr} = mJ$ are inequivalent different for $n > 2$.

♣Following exercise belongs in the Linear Algebra chapter. ♣

Exercise *Linear independence of eigenvectors with distinct eigenvalues*

Suppose a set of nonzero vectors $v_1, \ldots, v_n$ are eigenvectors of some operator A with distinct eigenvalues. Show that they are linearly independent. [212]

---

[212] *Answer*: Suppose $\sum_s c_s v_s = 0$ for some coefficients c_s. Applying powers of A we determine that $\sum c_s \lambda_s^k v_s = 0$. If all the c_s are nonzero then we learn that the matrix the matrix VC must have determinant zero where $V_{ij} = \lambda_i^j$ and C is the diagonal matrix with entries $c_1, \ldots, c_n$. If some of the c_s are nonzero then we have a minor of the matrix V times the diagonal matrix of the nonzero c_s. In any case, none of the minors of V_{ij} have zero determinant, provided the λ_i are distinct. Therefore, VC (or the appropriate minor) has nonzero determinant and hence no kernel. So $\{v_s\}$ is a linearly independent set.

Exercise $\mathfrak{su}(2)$ *vs.* $\mathfrak{sl}(2,\mathbb{R})$

A basis for the real Lie algebra of 2×2 traceless anti-Hermitian matrices is

$$T^i = -\frac{\mathrm{i}}{2}\sigma^i \qquad\qquad i=1,2,3 \tag{15.508}$$

with Lie algebra

$$[T^i, T^j] = \epsilon^{ijk}T^k \tag{15.509}$$

Can one make real linear combinations of T^i to produce the generators e, h, f of $\mathfrak{sl}(2,\mathbb{R})$ above?

Exercise *Representation matrices*

a.) Show that if v_0 is the vector used above with $\rho(e)v_0 = 0$ and $\rho(h)v_0 = -Nv_0$ then [213]

$$\rho(e)\rho(f)^\ell v_0 = -\ell(N+1-\ell)\rho(f)^{\ell-1}v_0 \tag{15.510}$$

b.) Suppose that we choose a highest weight vector v_h so that $\rho(f)w = 0$ and $\rho(h)w = +Nw$. Write the representation matrices in the ordered basis

$$w, \rho(e)w, \ldots, \rho(e)^N w \tag{15.511}$$

c.) Put a unitary structure on the vector space V so that $\rho(T^j)$ are anti-Hermitian matrices and relate the above bases to the standard basis $|j,m\rangle$, $m = -N/2, -N/2+1, \ldots, N/2-1, N/2$ appearing in quantum-mechanics textbooks. That is, find a rescaling of the vectors $\rho(f)^k v_0$ so that in the new basis, after defining $\rho(T^i)$ using (11.557) one obtains anti-Hermitian matrices for $\rho(T^i)$.

Example: $SL(2,\mathbb{Z})$ action on $Heis(\mathbb{Z}_n \times \mathbb{Z}_n)$.

Consider again the example of the quantum mechanics of a particle on a discrete approximation to a ring.

1. Because the position is periodic, the momentum is quantized.
2. Because the position is quantizied, the momentum is periodic.

So, the momentum is both periodic and discrete, just like the position. Recall the position operator Q and the momentum operator P both had a spectrum which is given by n^{th} roots of unity.

[213] *Hint*: Use $[\rho(e), \rho(f)^\ell] = \sum_{i=0}^{\ell-1}\rho(e)^i[\rho(e),\rho(f)]\rho(e)^{\ell-1-i}$.

So there is a symmetry between momentum and position. This is part of a kind of symplectic symmetry in this discrete system related to $SL(2,\mathbb{Z}_n)$. All such matrices arise from reduction modulo n of matrices in $SL(2,\mathbb{Z})$. Recall from Section [**** 8.3 ****] that $SL(2,\mathbb{Z})$ is generated by S and T with relations

$$(ST)^3 = S^2 = -1 \tag{15.512}$$

Therefore, S and T (reduced modulo n) will generate $SL(2,\mathbb{Z}_n)$, although there will be further relations, such as $T^n = 1$. The $SL(2,\mathbb{Z}_n)$ symmetry plays and important role in string theory and Chern-Simons theory and illustrates nicely some ideas of duality.

We now take the cocycle for Heis$(\mathbb{Z}_n \times \mathbb{Z}_n)$ to be

$$f((a_1,b_1),(a_2,b_2)) = a_1 b_2 \in \mathbb{Z}_n \tag{15.513}$$

so the commutator function is

$$\kappa((a_1,b_1),(a_2,b_2)) = a_1 b_2 - a_2 b_1 = v_1^{tr} J v_2 \tag{15.514}$$

which we can recognize as a symplectic form on $\mathbb{Z}_n \oplus \mathbb{Z}_n$.

Now an important subtlety arises. In general we cannot find an equivalent cocycle so that

$$\tilde{f}((a_1,b_1),(a_2,b_2)) = \frac{1}{2}(a_1 b_2 - a_2 b_1) \tag{15.515}$$

because, in general, we are not allowed to divide by 2 in $\mathbb{Z}_n$. After all, $x = \frac{1}{2}(a_1 b_2 - a_2 b_1)$ should be the solution to $2x = (a_1 b_2 - a_2 b_1)$, but, if n is even, then if x is a solution $x + n/2$ is a different solution, so the expression is ambiguous. If n is odd then 2 is invertible and we can divide by 2. Therefore, it is not obvious if $Sp(2,\mathbb{Z}_n)$ will lift to automorphisms of the finite Heisenberg group.

Consider the transformation:

$$S : (a,b) \to (b,-a) \tag{15.516}$$

This is a symplectic transformation: $S^*\kappa = \kappa$, and it satisfies $S^2 = -1$. Now we compute

$$(S^* f - f)((a_1,b_1),(a_2,b_2)) = -(a_1 b_2 + b_1 a_2) \tag{15.517}$$

Although the cocycle is not invariant under S nevertheless the difference $S^* f - f$ can indeed be trivialized by

$$\tau_S(a,b) = -ab \tag{15.518}$$

Now consider the transformation

$$T : (a,b) \to (a+b,b) \tag{15.519}$$

We compute

$$(T^* f - f)((a_1,b_1),(a_2,b_2)) = b_1 b_2 \tag{15.520}$$

This can be trivialized by

$$\tau_T(a,b) = \frac{1}{2} b^2 \tag{15.521}$$

PROVIDED we are able to divide by 2!! This is possible if n is odd, but not when n is even. Indeed, when $n = 2$ the cocycle $(T^*f - f)$ is not even a trivializable cocycle! (Why not? Apply the triviality test described in [**** Remark 5, section 11.3 ****] above.)

One way to determine the lifted group, and how the group lifts when n is even is the following. Suppose we have any two operators U, V that satisfy

$$UV = e^{2\pi i\theta} VU \tag{15.522}$$

for some θ (which, at this point, need not even be rational). If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \tag{15.523}$$

then

$$\widetilde{U} := U^a V^c \qquad\qquad \widetilde{V} = U^b V^d \tag{15.524}$$

♣Need to add phases here to get a proper group action. These don't compose properly yet. ♣

also satisfy

$$\widetilde{U}\widetilde{V} = e^{2\pi i\theta}\widetilde{V}\widetilde{U} \tag{15.525}$$

Now suppose, in addition that $\theta = 1/n$ and $U^n = V^n = 1$. Then we can compute

$$\begin{aligned} \widetilde{U}^n &= e^{-i\pi\theta n(n-1)ac} = e^{-i\pi(n-1)ac} \\ \widetilde{V}^n &= e^{-i\pi\theta n(n-1)bd} = e^{-i\pi(n-1)bc} \end{aligned} \tag{15.526}$$

Now, when n is odd, the conditions (15.526) place no restriction on A. In that case, the group $SL(2,\mathbb{Z})$ acts on $\mathrm{Heis}(\mathbb{Z}_n \oplus \mathbb{Z}_n)$ as a group of automorphisms, but the normal subgroup

$$\Gamma(n) := \{A \in SL(2,\mathbb{Z}) | A = \mathbf{1} \mathrm{mod} n\} \tag{15.527}$$

acts trivially so that

$$SL(2,\mathbb{Z})/\Gamma(n) \cong SL(2,\mathbb{Z}_n) \tag{15.528}$$

indeed acts as a group of automorphisms.

However, when n is even, we must consider the subgroup of $SL(2,\mathbb{Z})$ with the extra conditions $ac = bd = 0\mathrm{mod}2$. Only this subgroup acts as a group of automorphisms. Again a finite quotient group acts effectively.

Lifting Symplectomorphisms In Geometric Quantization

****

****

TO BE WRITTEN OUT IN DETAIL: A good example of the general problem of lifting automorphisms is to consider a line bundle with connection (L, ∇) over a symplectic manifold such that the curvature of the connection is the symplectic form Ω. Then, in geometric quantization we can attempt to lift the symplectomorphisms which preserve the line bundle with connection. We will see all the above phenomena, and only the "Hamiltonian automorphisms" will lift. For some discussion see section 6 of:

https://arxiv.org/pdf/hep-th/0605200.pdf

IN PARTICULAR EXPLAIN THE RELATION OF $\mathrm{Aut}_0(G)$ to the kernel of a homomorphism $\mathrm{Aut}(G) \to \mathrm{Hom}(G, H^1(G, A))$.

****

****

15.5.9 Coherent State Representations Of Heisenberg Groups: The Bargmann Representation

EXPLAIN BARGMANN REPRESENTATION.

NICE FORMULAE FOR COHERENT STATES AND VEV'S OF EXPONENTIATED QUADRATIC EXPRESSIONS IN OSCILLATORS.

15.5.10 Some Remarks On Chern-Simons Theory

In Chern-Simons theory (and similar topological field theories) it is quite typical for the Wilson line operators to generate finite Heisenberg groups. For example for $U(1)$ Chern-Simons of level k on a torus we have an action

$$S = \frac{k}{4\pi} \int_{\mathbb{R}} dt \int_{T^2} A_1 \partial_t A_2 + \cdots \tag{15.529}$$

so upon quantization $A_2 \sim \frac{4\pi}{k} \frac{\delta}{\delta A_1}$. The consequence is that Wilson lines along the a- and b-cycles generate a finite Heisenberg group with $q = e^{2\pi \mathrm{i}/k}$. The Hilbert space of states is a finite-dimensional irreducible representation of this group.

****

EXPLAIN MORE. THETA FUNCTIONS AND METAPLECTIC REPRESENTATION.

****

15.6 Non-Central Extensions Of A General Group G By An Abelian Group A: Twisted Cohomology

Let us now generalize central extensions to extensions of the form:

$$1 \to A \quad \stackrel{\iota}{\to} \quad \tilde{G} \quad \stackrel{\pi}{\to} \quad G \to 1 \tag{15.530}$$

Here G can be any group, not necessarily Abelian. We continue to assume that $N = A$ is Abelian, but now we no longer assume $\iota(A)$ is central in $\tilde{G}$. So we allow for the possibility of non-central extensions by an Abelian group.

Much of our original story goes through, but now the map

$$\omega : G \to \mathrm{Aut}(A) \tag{15.531}$$

of our general discussion (defined in equations (15.12) and (15.14)) is canonically defined and is actually a group homomorphism. As we stressed below (15.14), in general it is not a group homomorphism. There are two ways to understand that:

1. $\tilde{G}$ acts on A by conjugation on the isomorphic image of A in $\tilde{G}$ which, because the sequence is exact, is still a normal subgroup. In equations, we can define

$$\iota(\tilde{\omega}_{\tilde{g}}(a)) := \tilde{g}\iota(a)\tilde{g}^{-1} \tag{15.532}$$

But now $\tilde{\omega}_{\tilde{g}}$ only depends on the equivalence class $[\tilde{g}] \in \tilde{G}/\iota(A)$ beccause

$$(\tilde{g}\iota(a_0))\,\iota(a)\,(\tilde{g}\iota(a_0))^{-1} = \tilde{g}\iota(a)\tilde{g}^{-1} \tag{15.533}$$

so $\tilde{\omega}_{\tilde{g}\iota(a_0)} = \tilde{\omega}_{\tilde{g}}$ and since $\tilde{G}/\iota(A) \cong G$ we can use this to define ω_g. However, from this definition it is clear that $g \mapsto \omega_g$ is a group homomorphism.

2. Or you can just choose a section and define ω_g exactly as in (15.14). To stress the dependence on s we write

$$\iota(\omega_{g,s}(a)) = s(g)\iota(a)s(g)^{-1} \tag{15.534}$$

However, now if we change section so that [214] $\hat{s}(g) = \iota(t(g))s(g)$ is another section then we compute

$$\begin{aligned}
\iota\left(\omega_{g,\hat{s}}(a)\right) &:= \{\iota(t(g))s(g)\} \cdot \iota(a) \cdot \{\iota(t(g))s(g)\}^{-1} \\
&= \iota(t(g)) \cdot \iota(\omega_{g,s}(a)) \cdot \iota(t(g))^{-1} \\
&= \iota\left\{t(g) \cdot \omega_{g,s}(a) \cdot (t(g))^{-1}\right\} \\
&= \iota(\omega_{g,s}(a))
\end{aligned} \tag{15.535}$$

and since ι is injective $\omega_{g,s}$ is independent of section and we can just denote it as ω_g.

Note carefully that only in the very last line did we use the assumption that A is Abelian. We will come back to this when we discuss general extensions in section 15.7.

Moreover, given a choice of section we can define $f_s(g_1, g_2)$ just as we did in equation (15.90). This definition works for all group extensions:

$$s(g_1)s(g_2) = \iota(f_s(g_1, g_2))s(g_1g_2) \tag{15.536}$$

We can now compute, just as in (15.15):

$$\begin{aligned}
\iota\left(\omega_{g_1} \circ \omega_{g_2}(a)\right) &= s(g_1)\iota(\omega_{g_2}(a))s(g_1)^{-1} \\
&= s(g_1)s(g_2)\iota(a)(s(g_1)s(g_2))^{-1} \\
&= \iota(f_s(g_1, g_2)) \cdot \iota(\omega_{g_1g_2}(a)) \cdot \iota(f_s(g_1, g_2))^{-1} \\
&= \iota\left\{f_s(g_1, g_2))\omega_{g_1g_2}(a)f_s(g_1, g_2))^{-1}\right\} \\
&= \iota\left(\omega_{g_1g_2}(a)\right)
\end{aligned} \tag{15.537}$$

and again notice that only in the very last line did we use the hypothesis that A is Abelian. Again, since ι is injective, we conclude that $\omega_{g_1} \circ \omega_{g_2} = \omega_{g_1g_2}$ so that the map ω is a group homomorphism.

[214] Note that here the order of the two factors on the RHS matters, since $\iota(A)$ is not necessarily central in $\widetilde{G}$

Now, computing $s(g_1)s(g_2)s(g_3)$ in two ways, just as before, we derive the *twisted cocycle relation*:

$$\omega_{g_1}(f_s(g_2,g_3))f_s(g_1,g_2g_3) = f_s(g_1,g_2)f_s(g_1g_2,g_3) \tag{15.538}$$

Conversely, given a homomorphism $\omega : G \to \mathrm{Aut}(A)$ and a twisted cocycle for ω we can define a group law on the set $A \times G$:

$$(a_1,g_1)\cdot(a_2,g_2) = (a_1\omega_{g_1}(a_2)f(g_1,g_2), g_1g_2) \tag{15.539}$$

The reader should check that this really does define a valid group law on the set $A \times G$.

Remark: Note that (15.539) simultaneously generalizes the twisted product of a semidirect product (14.2) and the twisted product of a central extension (15.102).

Now suppose that we change section from s to $\hat{s}(g) := \iota(t(g))s(g)$ using some arbitrary function $t : G \to A$. Then one can compute that the new cocycle is related to the old one by

$$f_{\hat{s}}(g_1,g_2) = t(g_1)\omega_{g_1}(t(g_2))f_s(g_1,g_2)t(g_1g_2)^{-1} \tag{15.540}$$

Note that since A is Abelian the order of the factors on the RHS do not matter, but in the analogous formula for general extensions, equation (15.620) below, the order definitely does matter.

We say two different twisted cocycles are related by a twisted coboundary if they are related as in (15.540) for some function $t : G \to A$. One can check that if f is a twisted cocycle and we define f' as in (15.540) then f' is also a twisted cocycle. We again have an equivalence relation and we define the *twisted cohomology group* $H^{2+\omega}(G,A)$ to be the abelian group of equivalence classes. It is again an Abelian group, as in the untwisted case, as one shows by a similar argument.

The analog of the main theorem of section 15.3 above is:

Theorem: Let $\omega : G \to \mathrm{Aut}(A)$ be a fixed group homomorphism. Denote the set of isomorphism classes of extensions of the form

$$1 \to A \to \tilde{G} \to G \to 1 \tag{15.541}$$

which induce ω by $\mathrm{Ext}^\omega(G,A)$. Then the set $\mathrm{Ext}^\omega(G,A)$ is in 1-1 correspondence with the twisted cohomology group $H^{2+\omega}(G,A)$.

The proof is very similar to the untwisted case and we will skip it. Now the trivial element of the Abelian group $H^{2+\omega}(G,A)$ corresponds to the semidirect product determined by ω.

Now we can observe an interesting phenomenon which happens often in cohomology theory: Suppose that a twisted cocyle f is trivializable so that $[f] = 0$. Then our group extension is equivalent to a semidirect product. Nevertheless, the sequence (15.530) can be split in many different ways: There are many distinct trivializations and the different

trivializations have meaning. Equivalently, there are many different coboundary transformations that preserve the trivial cocycle. A glance at (15.540) reveals that this will happen when

$$t(g_1 g_2) = t(g_1)\omega_{g_1}(t(g_2)) \tag{15.542}$$

This is known as a *twisted homomorphism*. Of course, in the case that $\omega : G \to \mathrm{Aut}(A)$ takes every $g \in G$ to the identity automorphism of $\mathrm{Aut}(A)$ (that is, the identity element of $\mathrm{Aut}(A)$, the condition specializes to the definition of a homomorphism.

For the later discussion of group cohomology is useful:

A 1-cochain $t \in C^1(G, A)$ is simply a map $t : G \to A$.

A twisted homomorphism is also known as a *twisted one-cocycle*. That is, a 1-cocycle $t \in Z^{1+\omega}(G, A)$ with twisting ω is a 1-cochain that satisfies (15.542).

To define group cohomology $H^{1+\omega}(G, A)$ we need an appropriate notion of equivalence of one-cocycles. This is motivated by noting that if $s : G \to \tilde{G}$ is a section that is also a homomorphism (that is, a splitting) then for any $a \in A$ we can produce a new splitting

$$\tilde{s}(g) = \iota(a)s(g)\iota(a)^{-1} \tag{15.543}$$

This corresponds to the change of section $\tilde{s}(g) = \iota(t(g))s(g)$ where the function $t(g)$ is:

$$t(g) = t_a(g) := a\omega_g(a)^{-1} \quad . \tag{15.544}$$

To check this you write

$$\begin{aligned} \tilde{s}(g) &= \iota(a)s(g)\iota(a)^{-1} \\ &= \iota(a) \cdot \left(s(g)\iota(a)^{-1}s(g)^{-1}\right) \cdot s(g) \\ &= \iota(a) \cdot \left(\iota(\omega_g(a^{-1}))\right) \cdot s(g) \\ &= \left(\iota(a\omega_g(a^{-1}))\right) \cdot s(g) \end{aligned} \tag{15.545}$$

One easily checks that if t is a one-cocycle, then $t \cdot t_a$ is also a one-cocycle. So, in defining the cohomology group $H^{1+\omega}(G, A)$ we use the equivalence relation $t \sim t'$ if there exists an a so that $t = t' t_a$.

Theorem: When the sequence (15.530) splits, that is, when the cohomology class of the twisted cocycle is trivial $[f] = 0$, then the inequivalent splittings are in one-one correspondence with the inequivalent trivializations of a trivializable cocycle, and these are in one-one correspondence with the cohomology group $H^{1+\omega}(G, A)$.

Example 1: Consider the sequence associated with the Euclidean group

$$0 \to \mathbb{R}^d \xrightarrow{\iota} \mathrm{Euc}(d) \xrightarrow{\pi} O(d) \to 1 \tag{15.546}$$

Recall that if $v \in \mathbb{R}^d$ then $\iota(v) = T_v$ is the translation operator on affine space $\mathbb{A}^d$. We have $T_v(p) = p + v$. As we saw in (14.74) and (14.75) and the discussion preceding that exercise, for any $p \in \mathbb{A}^d$ we have a section $R \mapsto s_p(R) \in \mathrm{Euc}(d)$ where $s_p(R)$ is the transformation that takes

$$s_p(R) : p + v \mapsto p + Rv \tag{15.547}$$

In other words, we define rotation-reflections by choosing p as the origin. Then from

$$T_{\omega_R(v_0)} = s_p(R) T_{v_0} s_p(R)^{-1} \tag{15.548}$$

we compute that

$$\omega_R(v_0) = R v_0 \tag{15.549}$$

thus $\omega_R \in \mathrm{Aut}(\mathbb{R}^d)$, and indeed $R \mapsto \omega_R$ is a group homomorphism. If we have two difference sections $s_{p'}$ and s_p then

$$s_{p'}(R) = T_{t(R)} s_p(R) \tag{15.550}$$

where

$$t(R) = (1-R)(p'-p) = (1-R)w \tag{15.551}$$

where we have put $p' = p + w$, $w \in \mathbb{R}^d$.

Note that One easily checks that for fixed $w \in \mathbb{R}^d$

$$R \mapsto t(R) := (1-R)w \tag{15.552}$$

is indeed a twisted homomorphism $O(d) \to \mathbb{R}^d$. (It is <u>not</u> a homomorphism.) However, one also checks that it is of the form $t_w(R) = w - \omega_R(w)$, so it is a trivial one-cocycle: All the splittings are equivalent in the sense defined above.

Example 2: Now restrict the sequence (15.546) to

$$0 \to \mathbb{Z}^d \to G \to \{1, \sigma\} \to 1 \tag{15.553}$$

where $\{1, \sigma\} \subset O(d)$ is a subgroup isomorphic to $\mathbb{Z}_2$ with $\sigma = -\mathbf{1}_{d\times d}$. Then $\omega : \mathbb{Z}_2 \to \mathrm{Aut}(\mathbb{Z}^d)$ is inherited from ω_R in (15.546) and $\omega_\sigma(\vec{n}) = -\vec{n}$. Now the sequence splits and the most general possible splitting is $s_{\vec{n}_0}$ where $s_{\vec{n}_0}(1) = \{0|\mathbf{1}\}$ and

$$s_{\vec{n}_0}(\sigma) = \{\vec{n}_0|\sigma\} \tag{15.554}$$

for some $\vec{n}_0 \in \mathbb{Z}^d$. Indeed one checks $s_{\vec{n}_0}(\sigma)^2 = 1$. Now for $\vec{a} \in \mathbb{R}^d$ we have

$$t_{\vec{a}}(\sigma) = \vec{a} - \omega_\sigma(\vec{a}) = 2\vec{a} \in 2\mathbb{Z}^d \tag{15.555}$$

So not all splittings are equivalent! The equivalent ones have $\vec{n}_0 - \vec{n}_0' \in 2\mathbb{Z}^d$. Therefore

$$H^{1+\omega}(\mathbb{Z}_2, \mathbb{Z}^d) \cong \mathbb{Z}^d/2\mathbb{Z}^d \cong (\mathbb{Z}_2)^d \tag{15.556}$$

Remarks

1. Different trivializations of something trivializable can have physical meaning. In the discussion on crystallographic groups below the different trivializations are related to a choice of origin for rotation-reflection symmetries of the crystal.

2. An analogy to bundle theory might help some readers: Let G be a compact Lie group. Then the isomorphism classes of principal G-bundles over S^3 are in 1-1 correspondence with $\pi_2(G)$ and a theorem states that $\pi_2(G) = 0$ for all compact Lie groups. Therefore, every principal G-bundle over S^3 is trivializable. Distinct trivializations differ by maps $t : S^3 \to G$ and the set of inequivalent trivializations is classified by $\pi_3(G)$, which is, in general nontrivial. This can have physical meaning. For example, in Yang-Mills theory in $3+1$ dimensions on $S^3 \times \mathbb{R}$ the principal G-bundle on space S^3 is trivializable. But if there is an instanton between two time slices then the trivialization jumps by an element of $\pi_3(G)$.

Exercise *Due diligence*

Derive equation (15.538) and show that if we change f by a coboundary using (15.540) then indeed we produce another twisted cocycle.

Exercise

Suppose that a twisted cocycle $f(g_1, g_2)$ can be trivialized by two different functions $t_1, t_2 : G \to A$. Show that $t_{12}(g) := t_1(g)/t_2(g)$ is a trivialization that preserves the trivial cocycle. That is, show that t_{12} is a twisted 1-cocycle.

15.6.1 Crystallographic Groups

A *crystal* is a subset of affine space $C \subset \mathbb{A}^d$ that is invariant under translations by a lattice $L \subset \mathbb{R}^d$ (actually, that's an embedded lattice). As an example, see Figure 39. Then restricting the exact sequence of the Euclidean group (equation (15.546) above) to the subgroup $G(C) \subset \mathrm{Euc}(d)$ of those transformations that preserve C we have an exact sequence

$$1 \to L(C) \to G(C) \to P(C) \to 1 \tag{15.557}$$

where $P(C) \cong G(C)/L(C)$ is a subgroup of $O(d)$ known as the *point group of the crystal*.

Remark: In solid state physics when the sequence (15.557) does not split the crystallographic group $G(C)$ is said to be *nonsymmorphic*.

Example 1: Take $C = \mathbb{Z} \amalg (\mathbb{Z} + \delta) \subset \mathbb{R}$ where $0 < \delta < 1$. Then of course $L(C) = \mathbb{Z}$ acts by translations, preserving the crystal. But note that it is also true that

$$\begin{aligned} \{\delta|\sigma\} &: n \mapsto \delta - n = -n + \delta \\ &: n + \delta \mapsto \delta - (n + \delta) = -n \end{aligned} \tag{15.558}$$

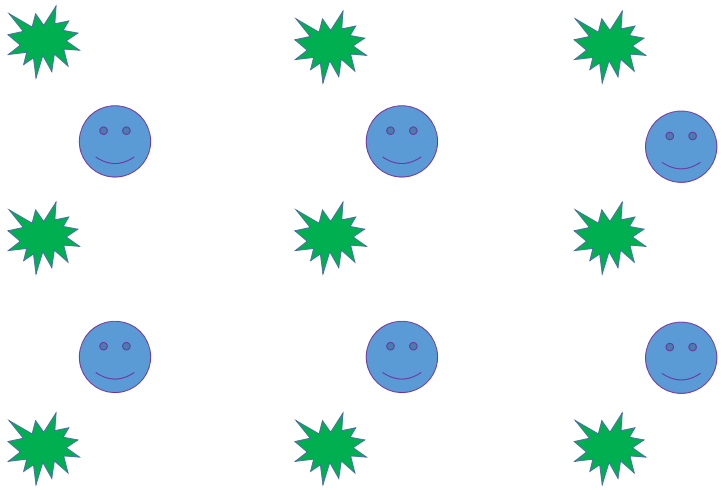


Figure 39: A portion of a crystal in the two-dimensional plane.

where $\sigma \in O(1)$ is the reflection around 0, $\sigma : x \to -x$ in $\mathbb{R}$. The transformation $\{\delta|\sigma\}$ maps $\mathbb{Z}$ to $\mathbb{Z}+\delta$ and $\mathbb{Z}+\delta$ to $\mathbb{Z}$ so that the whole crystal is preserved. Since $O(1) = \mathbb{Z}_2$, this is all we can do. We thus find that $G(C)$ fits in a sequence

$$0 \to L(C) \cong \mathbb{Z} \to G(C) \to O(1) \cong \mathbb{Z}_2 \to 1 \tag{15.559}$$

But we can split this sequence by choosing a section $s(\sigma) = \{\delta|-1\}$. Note that

$$\{\delta|\sigma\} \cdot \{\delta|\sigma\} = \{0|1\} \tag{15.560}$$

so $s : O(1) \to G(C)$ is a homomorphism. Another way of thinking about this is that $s(\sigma)$ is just reflection, not around the origin, but around the point $\frac{1}{2}\delta$. So, by a shift of origin for defining our rotation-inversion group $O(1)$ we just have reflections and integer translations. In any case we can recognize $G(C)$ as the infinite dihedral group.

Example 2: More generally, consider a lattice $L \subset \mathbb{R}^d$ and a generic vector $\vec{\delta} \in \mathbb{R}^d$. Consider the crystal

$$C = L \amalg (L + \vec{\delta}) \tag{15.561}$$

If L and $\vec{\delta}$ are generic then the point group is just $\mathbb{Z}_2$ generated by $-1 \in O(d)$. Denoting the action of -1 on $\mathbb{R}^d$ by σ we can lift this to the involution

$$\{\vec{\delta}|\sigma\} \in G(C) \tag{15.562}$$

which exchanges L with $(L+\vec{\delta})$. This group is symmorphic (because $\{\vec{\delta}|\sigma\}$ is an involution). In fact, this operation is just inversion about the new origin $\frac{1}{2}\vec{\delta}$:

$$\{\vec{\delta}|\sigma\} : \frac{1}{2}\vec{\delta} + \vec{y} \to \frac{1}{2}\vec{\delta} - \vec{y} \tag{15.563}$$

♣NEED TO HAVE A FIGURE HERE. THIS WOULD HELP. ♣

Example 3: For another very similar example consider

$$C = L \amalg (L + \vec{\delta}) \subset \mathbb{R}^2 \tag{15.564}$$

where

$$L = a_1\mathbb{Z} \oplus a_2\mathbb{Z} \subset \mathbb{R}^2 \tag{15.565}$$

As we have just discussed, for generic a_1, a_2 and $\vec{\delta}$ the symmetry group will be isomorphic to the semidirect product $\mathbb{Z}^2 \rtimes \mathbb{Z}_2$.

However, now let $0 < \delta < \frac{1}{2}$ and specialize $\vec{\delta}$ to $\vec{\delta} = (\delta a_1, \frac{1}{2}a_2)$. Then the crystal has more symmetry and in particular the point group is enhanced from $\mathbb{Z}_2$ to $\mathbb{Z}_2 \times \mathbb{Z}_2$:

$$1 \to \mathbb{Z}^2 \to G(C) \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to 1 \tag{15.566}$$

To see this let σ_1, σ_2 be generators of $\mathbb{Z}_2 \times \mathbb{Z}_2$ acting by reflection around the x_2 and x_1 axes, respectively. Then the operations:

$$\hat{\sigma}_1 : (x_1, x_2) \mapsto \vec{\delta} + (-x_1, x_2) \tag{15.567}$$

$$\hat{\sigma}_2 : (x_1, x_2) \mapsto (x_1, -x_2) \tag{15.568}$$

are symmetries of the crystal $G(C)$. In Seitz notation (or rather, its improvement - see equations (14.31) and (14.32) above) we have:

$$\hat{\sigma}_1 = \{\vec{\delta}| \begin{pmatrix} -1 & \\ & 1 \end{pmatrix}\} \tag{15.569}$$

$$\hat{\sigma}_2 = \{0| \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}\} \tag{15.570}$$

Now, we can define a section $s(\sigma_1) = \hat{\sigma}_1$ and $s(\sigma_2) = \hat{\sigma}_2$. Note that the square of the lift

$$\hat{\sigma}_1^2 = \{(0, a_2)|1\} \tag{15.571}$$

is a nontrivial translation. Thus $\sigma_i \to \hat{\sigma}_i$ is *not* a splitting. Moreover, $\hat{\sigma}_1$ does not have finite order. Therefore, it cannot be in a discrete group of rotations about any point!

Just because we chose a section that wasn't a splitting doesn't mean that a splitting doesn't actually exist. Here is how we can prove that in fact no splitting exists: The most general section is of the form

$$s(\sigma_1) = \{\vec{\delta} + \vec{v}|\sigma_1\} \tag{15.572}$$

where $\vec{v} = (n_1 a_1, n_2 a_2) \in L$ where $n_1, n_2 \in \mathbb{Z}$. Now consider the square:

$$s(\sigma_1)^2 = \{(0, a_2(1+2n_2)|1\}. \tag{15.573}$$

Since $n_2 \in \mathbb{Z}$ there is no lifting that makes this an involution. Therefore, there is no section. Therefore the sequence (15.566) does not split.

Example 4: It is interesting to see what happens to the previous example when $a_1 = a_2 = a$ and we take $\delta = a(\frac{1}{2}, \frac{1}{2})$. Then, clearly

$$C = L \amalg (L + \vec{\delta}) \subset \mathbb{R}^2 \tag{15.574}$$

has a point group symmetry D_4. So this becomes a symmorphic crystal. In fact, this is just a square lattice in disguise! We can take basis vectors δ and $R(\pi/2)\delta$.

****

NEED TO RELATE THE ABOVE FACTS MORE DIRECTLY TO THE PREVIOUS DISCUSSION OF GROUP COHOMOLOGY. SHOULD DO MORE ON CASE WHERE THE SEQUENCE SPLITS BUT THERE ARE INEQUIVALENT SPLITTINGS: PROBABLY A good example is Zincblend structure with tetrahedral symmetry. For example GaAs has this structure. There are two tetrahedra around the Ga and As but they are rotated.

****

Exercise

Why does the argument of example 3 fail in the special case of example 4? [215]

Exercise *Honeycomb*

Consider a honeycomb crystal in the plane. Discuss the crystal group, the point group, and decide if it is symmorphic or not.

♣Need to provide answer in a footnote ♣

15.6.2 Time Reversal

A good example of a physical situation in which it is useful to know about how twisted cocycles define non-central extensions is when there are anti-unitary symmetries in a quantum mechanical system. A typical example where this happens is when there is a time-orientation-reversing symmetry. In this case there is a homomorphism

$$\tau : G \to \{\pm 1\} \cong \mathbb{Z}_2 \tag{15.575}$$

[215] *Answer*: The wrong step is in equation (15.572). When δ takes the special form $\frac{1}{2}(a, a)$ this is not the most general lifting. One has now translation symmetry by multiples of $\vec{\delta}$, so there is an obvious lifting of σ_1.

telling us whether the symmetry $g \in G$ preserves or reverses the orientation of time.

In quantum mechanics it is often (but not always! - see below) the case that time-reversal is implemented as an anti-unitary operator (see Chapter 2 below for a precise definition of this term) and therefore when looking at the way the symmetry is implemented quantum mechanically we should consider the nontrivial automorphism of $U(1)$ defined by complex conjugation.

Recall that

$$\mathrm{Aut}(U(1)) \cong \mathrm{Out}(U(1)) \cong \mathbb{Z}_2 \tag{15.576}$$

and the nontrivial element of $\mathrm{Aut}(U(1)$ is the automorphism $z \to z^* = z^{-1}$.

So, when working with a symmetry group G that includes time-orientation-reversing symmetries we will need to consider the group homomorphism

$$\omega : G \to \mathrm{Aut}(U(1)) \tag{15.577}$$

where:

$$\omega(g)(z) = \begin{cases} z & \tau(g) = +1 \\ z^{-1} & \tau(g) = -1 \end{cases} \tag{15.578}$$

Example 1. The simplest example is where we have a symmetry group $G = \mathbb{Z}_2$ interpreted as time reversal. It will be convenient to denote $M_2 = \{1, \bar{T}\}$, with $\bar{T}^2 = 1$. Of course, $M_2 \cong \mathbb{Z}_2$. In quantum mechanics the representation of $\bar{T}$ will be an operator $\rho(T) := \tilde{T}$ on the Hilbert space and we will get a possibly twisted central extension of M_2. Let $\omega : M_2 \to \mathrm{Aut}(U(1))$. There are two possibilities: $\omega(\bar{T}) = 1$ (so the operation is unitary) and $\omega(\bar{T})$ is the complex conjugation automorphism (so the operation is anti-unitary). Assuming the anti-unitary case is the relevant one, so that ω is the nontrivial homomorphis $M_2 \to \mathrm{Aut}(U(1))$ (both are isomorphic to $\mathbb{Z}_2$ so ω is just the identity homomorphism of $\mathbb{Z}_2$) we need the group cohomology:

$$H^{2+\omega}(\mathbb{Z}_2, U(1)) = \mathbb{Z}_2 \tag{15.579}$$

To prove this we look at the twisted cocycle identity. Exactly the same arguments as in Remark 5 of section 15.3 show that we can choose a gauge with $f(g,1) = f(1,g) = 1$ for all g. This leaves only $f(\bar{T}, \bar{T})$ to be determined. Now, the twisted cocycle relation for the case $g_1 = g_2 = g_3 = \bar{T}$ says that

$$f(\bar{T}, \bar{T})^* = \omega_{\bar{T}}(f(\bar{T}, \bar{T})) = f(\bar{T}, \bar{T}) \tag{15.580}$$

and since $f(\bar{T}, \bar{T}) \in U(1)$ this means $f(\bar{T}, \bar{T}) \in \{\pm 1\}$. We need to check that we can't gauge $f(\bar{T}, \bar{T})$ to one using a twisted coboundary relation. That relation says that we can gauge f to

$$\tilde{f}(\bar{T}, \bar{T}) = t(\bar{T})\omega_{\bar{T}}(t(\bar{T}))f(\bar{T}, \bar{T})/t(1) \tag{15.581}$$

Now $t(1) = 1$ since we want to preserve the gauge $f(g,1) = f(1,g) = 1$ and $t(\bar{T})\omega_{\bar{T}}(t(\bar{T})) = |t(\bar{T})|^2 = 1$ so $f(\bar{T}, \bar{T})$ is gauge invariant.

(Note that $\bar{T}$ is an involution so our old criterion from Remark 5 of section 15.3 would ask us to find a square root of $f(\bar{T},\bar{T}) = -1$. Indeed such a square root exists, it is $\pm$i, but our old criterion no longer applies because we are in the twisted case.)

So, choosing ω to be the nontrivial homomorphism $M_2 \to \mathrm{Aut}(U(1))$ there are two extensions:

$$1 \longrightarrow U(1) \longrightarrow M_2^{\pm} \xrightarrow{\tilde{\pi}} M_2 \longrightarrow 1 \tag{15.582}$$

Let us write these out more explicitly:

Choose a lift $\tilde{T}$ of $\bar{T}$. Then $\pi(\tilde{T}^2) = 1$, so $\tilde{T}^2 = z \in U(1)$. But, then

$$\tilde{T}z = \tilde{T}\tilde{T}^2 = \tilde{T}^2\tilde{T} = z\tilde{T} \tag{15.583}$$

On the other hand, since we take $\omega(\bar{T})$ to be the nontrivial automorphism of $U(1)$ then

$$\tilde{T}z = z^{-1}\tilde{T} \tag{15.584}$$

Therefore $z^2 = 1$, so $z = \pm 1$, and therefore $\tilde{T}^2 = \pm 1$. Thus the two groups are

$$M_2^{\pm} = \{z\tilde{T} | z\tilde{T} = \tilde{T}z^{-1} \qquad \& \qquad \tilde{T}^2 = \pm 1\} \tag{15.585}$$

These possibilities are really distinct: If $\tilde{T}'$ is another lift of $\bar{T}$ then $\tilde{T}' = \mu\tilde{T}$ for some $\mu \in U(1)$ and so

$$(\tilde{T}')^2 = (\mu\tilde{T})^2 = \mu\bar{\mu}\tilde{T}^2 = \tilde{T}^2 \tag{15.586}$$

So the sign of the square of the lift of the time-reversing symmetry is an invariant.

The extension corresponding to the identity element of $H^{2+\omega}(\mathbb{Z}_2, U(1))$ is the semidirect product. This is just $O(2)$, using $SO(2) \cong U(1)$:

$$O(2) = SO(2) \rtimes \mathbb{Z}_2 \tag{15.587}$$

But the nontrivial extension is a new group for us. It double-covers $O(2)$ and is known as $\mathrm{Pin}^-(2)$. Indeed we can define homomorphisms

$$\pi^{\pm} : M_2^{\pm} \to O(2) \tag{15.588}$$

where $\pi^{\pm}(\tilde{T}) = P \in O(2)$ and $\pi^{\pm}(z = e^{\mathrm{i}\alpha}) = R(2\alpha)$. Note that $-1 = e^{\mathrm{i}\pi} \mapsto R(e^{2\mathrm{i}\pi}) = +\mathbf{1}$. In $\mathrm{Pin}^+(2)$ the double cover of a reflection, $\tilde{T} = (\pi^+)^{-1}(P)$, squares to one. In $\mathrm{Pin}^-(2)$ the double cover of a reflection, $\tilde{T} = (\pi^-)^{-1}(P)$ squares to -1.

Remark: In QM textbooks it is shown that if we write Schrödinger equation for an electron in a potential with spin-orbit coupling then there is a time-reversal symmetry:

$$(\tilde{T} \cdot \Psi)(\vec{x}, t) = \mathrm{i}\sigma^2(\Psi(\vec{x}, -t)^* \tag{15.589}$$

where here Ψ is a 2-component spinor function of $(\vec{x}, t)$. [216] Note that this implies:

$$\begin{aligned}(\tilde{T}^2 \cdot \Psi)(\vec{x}, t) &= \mathrm{i}\sigma^2 \cdot \Big((\tilde{T} \cdot \Psi)(\vec{x}, -t)\Big)^* \\ &= \mathrm{i}\sigma^2 \cdot \big(\mathrm{i}\sigma^2 \cdot (\Psi(\vec{x}, t))^*\big)^* \\ &= \mathrm{i}\sigma^2\mathrm{i}\sigma^2\Psi(\vec{x}, t) \\ &= -\Psi(\vec{x}, t)\end{aligned} \tag{15.590}$$

[216] This is most elegantly derived from the time-reversal transformation on the Dirac equation.

So, in this example, $\tilde{T}^2 = -1$. More generally, in analogous settings for spin j particles $\tilde{T}^2 = (-1)^{2j}$. See section 15.6.3 below for an explanation. The fact that $\tilde{T}^2 = (-1)^{2j}$ in the spin j representation has a very important consequence known as *Kramer's theorem*: In these situations the energy eigenspaces must have even degeneracy. For if Ψ is an energy eigenstate $H\Psi = E\Psi$ and we have a time-reversal invariant system then $\tilde{T} \cdot \Psi$ is also an energy eigenstate. We can prove that it is linearly independent of Ψ as follows: Suppose to the contrary that

$$\tilde{T} \cdot \Psi = z\Psi \tag{15.591}$$

for some complex number z. Then act with $\tilde{T}$ again and use the fact that it is anti-unitary and squares to -1:

$$-\Psi = z^* \tilde{T} \cdot \Psi \tag{15.592}$$

but this implies that $z = -1/z^*$ which implies $|z|^2 = -1$, which is impossible. Therefore, (15.591) is impossible. Therefore Ψ and $\tilde{T} \cdot \Psi$ are independent energy eigenstates. A slight generalization of the argument shows that the dimension of the energy eigenspace must be even. A more conceptual way of understanding this is that the energy eigenspace must be a quaternionic vector space because we have an anti-linear operator on it that squares to -1. See the discussion of real, complex, and quaternionic vector spaces in Chapter 2 below.

Example 2: In general a system can have time-orientation reversing symmetries but the simple transformation $t \to -t$ is not a symmetry. Rather, it must be accompanied by other transformations so that the symmetry group is <u>not</u> of the simple form $G = G_0 \times \mathbb{Z}_2$ where G_0 is a group of time-orientation-preserving symmetries. (Such a structure is often assumed in the literature.) As a simple example consider a crystal

$$C = \big(\mathbb{Z}^2 + (\delta_1, \delta_2)\big) \amalg \big(\mathbb{Z}^2 + (-\delta_2, \delta_1)\big) \amalg \big(\mathbb{Z}^2 + (-\delta_1, -\delta_2)\big) \amalg \big(\mathbb{Z}^2 + (\delta_2, -\delta_1)\big) \tag{15.593}$$

where $\vec{\delta}$ is generic so, as we saw above we have a symmorphic crystal with $P(C) \cong D_4$. The action of D_4 is just given by rotation around the origin $\{0|R(\frac{\pi}{2})\}$ which we will denote by R and reflection, say, in the y-axis, which we will denote by P. So $R^4 = 1$, $P^2 = 1$, and $PRP = R^{-1}$. We have

$$G(C) = \mathbb{Z}^2 \rtimes D_4 \tag{15.594}$$

But now suppose there is a dipole moment, or spin S. We model this with a set of two elements $\mathcal{S} = \{S, -S\}$ for dipole moment up and down and now our crystal with spin is a subset of $\mathbb{R}^2 \times \mathcal{S}$. This subset is of the form

$$\widehat{C} = \widehat{C}_+ \amalg \widehat{C}_- \tag{15.595}$$

with

$$\widehat{C}_+ = \big(\mathbb{Z}^2 + (\delta_1, \delta_2)\big) \times \{S\} \amalg \big(\mathbb{Z}^2 + (-\delta_1, -\delta_2)\big) \times \{S\} \tag{15.596}$$

but a spin $-S$ on points of the complementary sub-crystal

$$\widehat{C}_- = \big(\mathbb{Z}^2 + (-\delta_2, \delta_1)\big) \times \{-S\} \amalg \big(\mathbb{Z}^2 + (\delta_2, -\delta_1)\big) \times \{-S\} \tag{15.597}$$

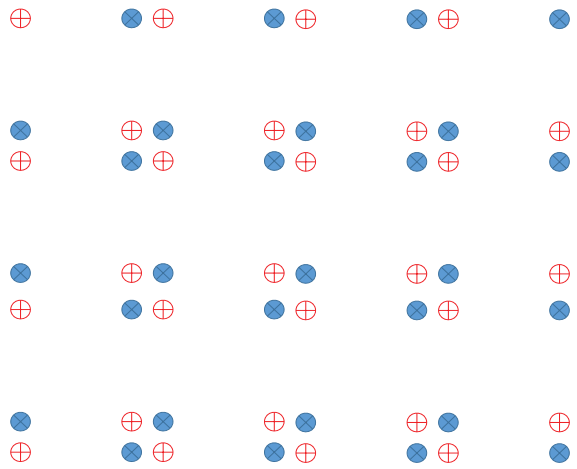


Figure 40: In this figure the blue crosses represent an atom with a local magnetic moment pointing up while the red crosses represent an atom with a local magnetic moment pointing down. The magnetic point group is isomorphic to D_4 but the homomorphism τ to $\mathbb{Z}_2$ has a kernel $\mathbb{Z}_2 \times \mathbb{Z}_2$ (generated by π rotation around a lattice point together with a reflection in a diagonal). Since D_4 is nonabelian the sequence $1 \to \widehat{P}_0 \to \widehat{P} \xrightarrow{\tau} \mathbb{Z}_2 \to 1$ plainly does not split.

Now let $\mathbb{Z}_2 = \{1, \sigma\}$ act on $\mathbb{R}^2 \times \mathcal{S}$ by acting trivially on the first factor and $\sigma : S \to -S$ on the second factor. Now reversal of time orientation exchanges S with $-S$. So the symmetries of the crystal with dipole is a subgroup $\widehat{G(C)} \subset \mathrm{Euc}(2) \times \mathbb{Z}_2$ known as the *magnetic crystallographic group*. The subgroup of translations by the lattice is still a normal subgroup and the quotient by the lattice of translations is the *magnetic point group*. In the present example:

$$0 \to \mathbb{Z}^2 \to \widehat{G(C)} \to \widehat{P(C)} \to 1 \tag{15.598}$$

The elements in $\widehat{P(C)}$ are

$$\{(1,1), (R,\sigma), (R^2,1), (R^3,\sigma), (P,\sigma), (PR,1), (PR^2,\sigma), (PR^3,1)\} \tag{15.599}$$

This magnetic point group is isomorphic to D_4 but the time reversal homomorphism takes $\tau(R,\sigma) = -1$ and $\tau(P,\sigma) = -1$ so that we have

$$1 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to \widehat{P(C)} \xrightarrow{\tau} \mathbb{Z}_2 \to 1 \tag{15.600}$$

The induced automorphism on $\mathbb{Z}_2 \times \mathbb{Z}_2$ is trivial so clearly this sequence does not split, since $\widehat{P(C)} \cong D_4$ is nonabelian.

Remarks:

1. With the possible exception of exotic situations in which quantum gravity is important, physics takes place in space and time. Except in unusual situations associated with nontrivial gravitational fields we can assume our spacetime is time-orientable. Then, any physical symmetry group G must be equipped with a homomorphism

$$\tau : G \to \mathbb{Z}_2 \tag{15.601}$$

telling us whether the symmetry operations preserve or reverse the orientation of time. That is $\tau(g) = +1$ are symmetries which preserve the orientation of time while $\tau(g) = -1$ are symmetries which reverse it.

Now, suppose that G is a symmetry of a quantum system. Then Wigner's theorem gives G another grading $\phi : G \to \mathbb{Z}_2$, telling us whether the operator $\rho(g)$ implementing the symmetry transformation g on the Hilbert space is unitary or anti-unitary. Thus, on very general grounds, a symmetry of a quantum system should be *bigraded* by a pair of homomorphisms (ϕ, τ), or what is the same, a homomorphism to $\mathbb{Z}_2 \times \mathbb{Z}_2$.

It is natural to ask whether ϕ and τ are related. A natural way to try to relate them is to study the dynamical evolution.

In quantum mechanics, time evolution is described by unitary evolution of states. That is, there should be a family of unitary operators $U(t_1, t_2)$, strongly continuous in both variables and satisfying composition laws $U(t_1, t_3) = U(t_1, t_2)U(t_2, t_3)$ so that the density matrix ϱ evolves according to:

$$\varrho(t_1) = U(t_1, t_2)\varrho(t_2)U(t_2, t_1) \tag{15.602}$$

Let us - for simplicity - make the assumption that our physical system has time-translation invariance so that $U(t_1, t_2) = U(t_1 - t_2)$ is a strongly continuous group of unitary transformations. [217]

By Stone's theorem, $U(t)$ has a self-adjoint generator H, the Hamiltonian, so that we may write

$$U(t) = \exp\left(-\frac{it}{\hbar}H\right) \tag{15.603}$$

♣There is an obvious generalization of this statement for $U(t_1, t_2)$. Is it proved rigorously somewhere? ♣

Now, suppose we have a group [218] of operators on the Hilbert space: $\rho : G \to \mathrm{Aut}_{\mathbb{R}}(\mathcal{H})$. We say this group action is a *symmetry of the dynamics* if for all $g \in G$:

$$\rho(g)U(t)\rho(g)^{-1} = U(\tau(g)t) \tag{15.604}$$

where $\tau : G \to \mathbb{Z}_2$ is the indicator of time-orientation-reversal.

[217] In the more general case we would need an analog of Stone's theorem to assert that there is a family of self-adjoint operators with $U(t_1, t_2) = \mathrm{Pexp}[-\frac{i}{\hbar}\int_{t_1}^{t_2} H(t')dt']$. Then, the argument we give below would lead to $\rho(g)H(t)\rho(g)^{-1} = \phi(g)\tau(g)H(t)$ for all t.

[218] As explained in the 2012 article of Freed and Moore, this group might be an extension of the original group of quantum symmetries $\bar{\rho} : \bar{G} \to \mathrm{Aut}_{\mathrm{qtm}}(\mathbb{P}\mathcal{H})$.

Now, substituting (15.603) and paying proper attention to ϕ we learn that the condition for a symmetry of the dynamics (15.604) is equivalent to

$$\phi(g)\rho(g)H\rho(g)^{-1} = \tau(g)H \tag{15.605}$$

in other words,

$$\rho(g)H\rho(g)^{-1} = \phi(g)\tau(g)H \tag{15.606}$$

Thus, the answer to our question is that ϕ and τ are *unrelated* in general. We should therefore define a third homomorphism $\chi : G \to \mathbb{Z}_2$

$$\chi(g) := \phi(g)\tau(g) \in \{\pm 1\} \tag{15.607}$$

Note that

$$\phi \cdot \tau \cdot \chi = 1 \tag{15.608}$$

2. It is very unusual for physical systems to have nontrivial homomorphisms χ. That is, it is very unusual to have physical systems with time-orientation-reversing symmetries which are $\mathbb{C}$-linear or time-orientation-preserving symmetries which act $\mathbb{C}$-anti-linearly. But it is not impossible. To see why it is unusual note that:

$$\rho(g)H\rho(g)^{-1} = \chi(g)H \tag{15.609}$$

implies that if any group element has $\chi(g) = -1$ then the spectrum of H must be symmetric around zero. In particular, if the spectrum is bounded below but not above this condition must fail. In many problems, e.g. in the standard Schrödinger problem with potentials which are bounded below, or in relativistic QFT with H bounded below we must have $\chi(g) = 1$ for all g and hence $\phi(g) = \tau(g)$, which is what one reads in virtually every physics textbook: "A symmetry is anti-unitary iff it reverses the orientation of time." Not true, in general.

3. However, there *are* physical examples where $\chi(g)$ can be non-trivial, that is, there can be symmetries which are both anti-unitary and time-orientation preserving. An example are the so-called "particle-hole" symmetries in free fermion systems.

15.6.3 $T^2 = (-1)^{2j}$ and the Clebsch-Gordon Decomposition

Above we checked that $T^2 = -1$ on spin 1/2 particles whose wavefunction obeys the usual Schrödinger equation with spin-orbit coupling. The generalization to spin j particles is $T^2 = (-1)^{2j}$.

One simple way to see this is to note that the spin j representation is obtained by decomposing the tensor product of $(2j)$ copies of the spin 1/2 representation. In general we have the very important Clebsch-Gordon decomposition:

$$V(j_1) \otimes V(j_2) \cong V(|j_1 - j_2|) \oplus V(|j_1 - j_2| + 1) \oplus \cdots\cdots \oplus V(j_1 + j_2) \tag{15.610}$$

Note that every representation on the RHS has the same parity of $(-1)^{2j}$. Also note the triangular structure of the Clebsch-Gordon decomposition of $V(\frac{1}{2})^{\otimes n}$ allowing for an

inductive proof. Finally $\tilde{T}^2$ on $V(\frac{1}{2})^{\otimes n}$ is just $(-1)^n$, so it is $(-1)^n$ on the highest summand $V(n/2)$.

Let us give a proof of (15.610). We need some general facts about representation theory. See Chapter 4 for full explanations:

15.7 General Extensions

Let us briefly return to the general extension (15.1). Thus, we are now not assuming that N or Q is abelian. We might ask what happens if we try to continue following the reasoning of section (15.1) in this general case, but now keeping in mind the nice classification of central extensions using group cohomology.

What we showed is that for *any* group extension a choice of a section $s : Q \to G$ automatically gives us two maps:

1. $\omega_s : Q \to \mathrm{Aut}(N)$
2. $f_s : Q \times Q \to N$

These two maps are defined by

$$\iota(\omega_{s,q}(n)) := s(q)\iota(n)s(q)^{-1} \tag{15.611}$$

and

$$s(q_1)s(q_2) := \iota(f_s(q_1, q_2))s(q_1 \cdot q_2) \tag{15.612}$$

respectively.

Now (15.611) defines an element of $\mathrm{Aut}(N)$ for fixed s and q, but the map $q \mapsto \omega_{s,q}$ need not be a homomorphism, as we have repeatedly stressed. Rather, using (15.611) and (15.612) we can derive a twisted version of the homomorphism rule:

$$\omega_{s,q_1} \circ \omega_{s,q_2} = I(f_s(q_1, q_2)) \circ \omega_{s,q_1 q_2} \tag{15.613}$$

Recall that for $a \in N$, $I(a) \subset \mathrm{Aut}(N)$ denotes the inner automorphism given by conjugation by a. The proof of (15.613) follows exactly the same steps as (15.537), except for the very last line.

Moreover, using (15.612) to relate $s(q_1)s(q_2)s(q_3)$ to $s(q_1q_2q_3)$ in two ways gives a twisted cocycle relation:

$$\omega_{s,q_1}(f_s(q_2, q_3))f_s(q_1, q_2q_3) = f_s(q_1, q_2)f_s(q_1q_2, q_3) \tag{15.614}$$

Note this is the same as (15.538), but unlike that equation now order of the terms is very important since we no longer assume that N is abelian.

To summarize: Given a general extension (15.1) there exist maps (ω_s, f_s), associated with any section s and defined by (15.611) and (15.612). The maps (ω_s, f_s) automatically satisfy the identities (15.613) and (15.614).

We now consider, more generally, functions satisfying identities (15.613) and (15.614). That is, we assume we are given two maps (not necessarily derived from some section):

1. A map $f : Q \times Q \to N$

2. A map $\omega : Q \to \mathrm{Aut}(N)$

And we suppose the data (ω, f) satisfy the two conditions

$$\omega_{q_1} \circ \omega_{q_2} = I(f(q_1, q_2)) \circ \omega_{q_1 q_2} \tag{15.615}$$

$$\omega_{q_1}(f(q_2, q_3)) f(q_1, q_2 q_3) = f(q_1, q_2) f(q_1 q_2, q_3) \tag{15.616}$$

then we can construct an extension (15.1) with the multiplication law:

$$(n_1, q_1) \cdot_{f,\omega} (n_2, q_2) := (n_1 \omega_{q_1}(n_2) f(q_1, q_2), q_1 q_2) \tag{15.617}$$

This is very similar to (15.539) but we stress that since N might be nonabelian, the order of the factors in the first entry on the RHS matters!

With a few lines of algebra, using the identities (15.615) and (15.616) one can check the associativity law and the other group axioms. We have already seen this simultaneous generalization of the semidirect product (14.2) and the twisted product of a central extension (15.102) in our discussion of the case where $N = A$ is abelian. (See equation (15.539) above.) The new thing we have now learned is that this is the most general way of putting a group structure on a product $N \times Q$ so that the result fits in an extension of Q by N.

Now, suppose again that we are given a group extension. As we showed, a choice of section s gives us a pair of functions (ω_s, f_s) satisfying (15.615) and (15.616). Any other section $\tilde{s}$ is related to s by a function $t : Q \to N$. Indeed that function t is defined by:

$$\tilde{s}(q) = \iota(t(q)) s(q) \tag{15.618}$$

and one easily computes that we now have

$$\omega_{\tilde{s},q} = I(t(q)) \circ \omega_{s,q} \tag{15.619}$$

$$f_{\tilde{s}}(q_1, q_2) = t(q_1) \omega_{s,q_1}(t(q_2)) f_s(q_1, q_2) t(q_1 q_2)^{-1} \tag{15.620}$$

The proof of (15.619) follows exactly the same steps as (15.535). To prove (15.620) we patiently combine the definition (15.618) with the definition (15.612).

♣We also skipped this proof for $N = A$ abelian. Probably should show the steps. ♣

These formulae for how (ω_s, f_s) change as we change the section now motivate the following:

Suppose we are given a pair (ω, f) satisfying (15.615) and (15.616) and an arbitrary function $t : Q \to N$. We can now define a new pair (ω', f') by the equations:

$$\omega'_q = I(t(q)) \circ \omega_q \tag{15.621}$$

$$f'(q_1, q_2) = t(q_1) \omega_{q_1}(t(q_2)) f(q_1, q_2) t(q_1 q_2)^{-1} \tag{15.622}$$

Now, with some algebra (DO IT!) one can check that indeed (ω', f') really do satisfy (15.615) and (15.616) as well. Equations (15.621) and (15.622) generalize the coboundary relation (15.97) of central extension theory. Note that the equations relating ω and f back to ω' and f' are of the same form with $t(q) \to t(q)^{-1}$.

The relations (15.621) and (15.622) define an equivalence relation on the set of pairs (ω, f) satisfying (15.615) and (15.616). Moreover, if (ω, f) and (ω', f') are related by (15.621) and (15.622) then we can define a group structure on the set $N \times Q$ in two ways using the equation (15.617) for each pair. Nevertheless, there is a morphism between these two extensions in the sense of (15.4) above where we define

$$\varphi(n, q) := (nt(q)^{-1}, q) \tag{15.623}$$

So, to check this you need to check

$$\varphi((n_1, q_1) \cdot_{f,\omega} (n_2, q_2)) = \varphi(n_1, q_1) \cdot_{f',\omega'} \varphi(n_2, q_2) \tag{15.624}$$

Then note that $\varphi^{-1}(n, q) = (nt(q), q)$ is an inverse morphism of extensions, and hence we have an isomorphism of extensions.

Now we would like to state all this a little more conceptually. The first point to note is that a map $q \mapsto \omega_q \in \mathrm{Aut}(N)$ that satisfies (15.615) in fact canonically defines a homomorphism $\bar{\omega} : Q \to \mathrm{Out}(N)$ of Q into the group of outer automorphisms of N. This homomorphism is defined more conceptually as the unique map that makes the diagram:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & N & \xrightarrow{\iota} & G & \xrightarrow{\pi} & Q & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle I} & & \downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \bar{\omega}} & & \\ 1 & \longrightarrow & \mathrm{Inn}(N) & \longrightarrow & \mathrm{Aut}(N) & \longrightarrow & \mathrm{Out}(N) & \longrightarrow & 1 \end{array} \tag{15.625}$$

Here $I : N \to \mathrm{Inn}(N)$ is the map that takes n to the inner automorphism $I(n) : n' \mapsto nn'n^{-1}$ and ψ is the map from $G \to \mathrm{Aut}(N)$ defined by

$$\iota(\psi(g)(n)) = g\iota(n)g^{-1} \tag{15.626}$$

Now, we can ask the converse question: *Given an arbitrary homomorphism $\bar{\omega} : Q \to \mathrm{Out}(N)$ is there an extension of Q by N that induces it as in (15.625)?*

The most obvious thing to try when trying to answer this question is to use $\bar{\omega} : Q \to \mathrm{Out}(N)$ and the pullback construction (15.34) of the canonical exact sequence given by the lower line of (15.625). But this will only give an extension of Q by $\mathrm{Inn}(N)$. Note that $\mathrm{Inn}(N) \cong N/Z(N)$, and so the center of N might cause some trouble. That is in fact what happens: The answer to the above question is, in general, "NO," and the obstruction has to do with the third cohomology group $H^{3+\bar{\omega}}(Q, Z(N))$ where $Z(N)$ is the center of N. See section 15.8.5 below.

But for now, let us suppose we have a choice of $\bar{\omega}$ such that extensions inducing it do exist. What can we say about the set $\mathrm{Ext}^{\bar{\omega}}(Q, N)$ of equivalence classes of such extensions?

To answer this we choose a lifting of the homomorphism, that is, a map $q \mapsto \omega_q \in \mathrm{Aut}(N)$. Now, if we have two extensions both inducing $\bar{\omega}$ and we choose two liftings $\omega_q^{(1)}$ and $\omega_q^{(2)}$ then they will be related by

$$\omega_q^{(1)} = I(t(q)) \circ \omega_q^{(2)} \tag{15.627}$$

for some function $t : Q \to N$. Note, please, that while this equation is formally very similar to (15.619) it is conceptually different. Nothing has been said about the relation of the two extensions, other than that they induce the same $\bar{\omega}$.

Now we try to relate the corresponding functions $f^{(1)}(q_1, q_2)$ and $f^{(2)}(q_1, q_2)$. To do that we compute

$$\begin{aligned}
\omega^{(1)}_{q_1} \circ \omega^{(1)}_{q_2}(n) &= t(q_1)\omega^{(2)}_{q_1}\left(t(q_2)\omega^{(2)}_{q_2}(n)t(q_2)^{-1}\right)t(q_1)^{-1} \\
&= t(q_1)\omega^{(2)}_{q_1}(t(q_2))\left(\omega^{(2)}_{q_1} \circ \omega^{(2)}_{q_2}(n)\right)\omega^{(2)}_{q_2}(t(q_2)^{-1})t(q_1)^{-1} \\
&= t(q_1)\omega^{(2)}_{q_1}(t(q_2))\left(f^{(2)}(q_1, q_2)\omega^{(2)}_{q_1q_2}(n)f^{(2)}(q_1, q_2)^{-1}\right)\omega^{(2)}_{q_2}(t(q_2)^{-1})t(q_1)^{-1} \\
&= \left\{t(q_1)\omega^{(2)}_{q_1}(t(q_2))f^{(2)}(q_1, q_2)t(q_1q_2)^{-1}\right\} \cdot \omega^{(1)}_{q_1q_2}(n)\left\{t(q_1)\omega^{(2)}_{q_1}(t(q_2))f^{(2)}(q_1, q_2)t(q_1q_2)^{-1}\right\}^{-1} \\
&= \hat{f}^{(2)}(q_1, q_2) \cdot \omega^{(1)}_{q_1q_2}(n)\hat{f}^{(2)}(q_1, q_2)^{-1}
\end{aligned} \tag{15.628}$$

where we define

$$\hat{f}^{(2)}(q_1, q_2) := t(q_1)\omega^{(2)}_{q_1}(t(q_2))f^{(2)}(q_1, q_2)t(q_1q_2)^{-1} \tag{15.629}$$

On the other hand, we know that

$$\omega^{(1)}_{q_1} \circ \omega^{(1)}_{q_2}(n) = f^{(1)}(q_1, q_2)\omega^{(1)}_{q_1q_2}(n)f^{(1)}(q_1, q_2)^{-1} \tag{15.630}$$

Can we conclude that $\hat{f}^{(2)}(q_1, q_2) = f^{(1)}(q_1, q_2)$? Certainly not! Provided $\omega^{(1)}_{q_1q_2}(n)$ is sufficiently generic all we can conclude is that

$$\hat{f}^{(2)}(q_1, q_2) = f^{(1)}(q_1, q_2)\zeta(q_1, q_2) \tag{15.631}$$

for some function $\zeta : Q \times Q \to Z(N)$. These two functions are <u>not</u> necessarily related by a coboundary and the extensions are <u>not</u> necessarily equivalent!

What is true is that if $\hat{f}^{(2)}$ and $f^{(1)}$ satisfy the twisted cocycle relation then $\zeta(q_1, q_2)$ in (15.631) also satisfies the twisted cocycle relation. (This requires a lot of patient algebra....) It follows that

$$\zeta \in Z^{2+\bar{\omega}}(Q, Z(N)) \tag{15.632}$$

Moreover, going the other way, given one extension and corresponding $(\omega^{(1)}, f^{(1)})$, and a $\zeta \in Z^{2+\bar{\omega}}(Q, Z(N))$ we can change f as in (15.631). If $[z] \in H^{2+\bar{\omega}}(Q, Z(N))$ is nontrivial we will in general get a new, nonequivalent extension.

All this is summarized by the theorem:

Theorem: Let $\mathrm{Ext}^{\bar{\omega}}(Q, N)$ be the set of inequivalent extensions of Q by N inducing $\bar{\omega}$. Then either this set it is empty or it is a torsor [219] for $H^{2+\bar{\omega}}(Q, Z(N))$.

****

[219] A *torsor* X for a group G is a set X with a G-action on it so that given any pair $x, x' \in X$ there is a unique $g \in G$ that maps x to x'. In this chapter we have discussed an important example of a torsor quite extensively: Affine space $\mathbb{A}^d$ is a torsor for $\mathbb{R}^d$ with the natural action of $\mathbb{R}^d$ on $\mathbb{A}^d$ by translation.

NEED SOME EXAMPLES HERE. AND NEED SOME MORE INTERESTING EXERCISES.

****

Exercise *Checking the group laws*

Show that (15.617) really defines a group structure.

a.) Check the associativity relation.

b.) What is the identity element? [220]

c.) Check that every element has an inverse.

Exercise

a.) Check that (15.623) really does define a homomorphism of the group laws (15.617) defined by (ω, f) and (ω', f') if (ω', f') is related to (ω, f) by (15.621) and (15.622).

b.) Check that the diagram (15.4) really does commute if we use (15.623).

15.8 Group cohomology in other degrees

Motivations:

a.) The word "cohomology" suggests some underlying chain complexes, so we will show that there is such a formulation.

b.) There has been some discussion of higher degree group cohomology in physics in

1. The theory of anomalies (Faddeev-Shatashvili; Segal; Carey et. al.; Mathai et. al.; ...)

2. Classification of rational conformal field theories (Moore-Seiberg; Dijkgraaf-Vafa-Verlinde-Verlinde; Dijkgraaf-Witten; Kapustin-Saulina)

3. Chern-Simons theory and topological field theory (Dijkgraaf-Witten,...)

4. Condensed matter/topological phases of matter (Kitaev; Wen et. al.; Kapustin et. al.; Freed-Hopkins;....)

5. Three-dimensional supersymmetric gauge theory.

Here we will be brief and just give the basic definitions:

[220] *Answer*: $(f(1,1)^{-1}, 1_Q)$.

15.8.1 Definition

Suppose we are given any group G and an Abelian group A (written additively in this sub-section) and a homomorphism

$$\omega : G \to \mathrm{Aut}(A) \tag{15.633}$$

Definition: An n-cochain is a function $\phi : G^{\times n} \to A$. The space of n-cochains is denoted $C^n(G, A)$. It is also useful to speak of 0-cochains. We interpret a 0-cochain ϕ_0 to be some element $\phi_0 = a \in A$.

Note that $C^n(G, A)$, for $n \geq 0$, is an abelian group using the abelian group structure of A on the values of ϕ, that is: $(\phi_1 + \phi_2)(\vec{g}) := \phi_1(\vec{g}) + \phi_2(\vec{g})$.

Define a group homomorphism: $d : C^n(G, A) \to C^{n+1}(G, A)$

$$\begin{aligned}(d\phi)(g_1, \ldots, g_{n+1}) &:= \omega_{g_1}\left(\phi(g_2, \ldots, g_{n+1})\right) \\ -\phi(g_1g_2, g_3, \ldots, g_{n+1}) &+ \phi(g_1, g_2g_3, \ldots, g_{n+1}) \pm \cdots + (-1)^n \phi(g_1, \ldots, g_{n-1}, g_ng_{n+1}) \\ &+ (-1)^{n+1}\phi(g_1, \ldots, g_n)\end{aligned} \tag{15.634}$$

Then we have, for $n = 0$:

$$(d\phi_0)(g) = \omega_g(a) - a \tag{15.635}$$

For $n = 1$, $n = 2$ and $n = 3$ the formula written out looks like:

$$(d\phi_1)(g_1, g_2) = \omega_{g_1}\left(\phi_1(g_2)\right) - \phi_1(g_1g_2) + \phi_1(g_1) \tag{15.636}$$

$$(d\phi_2)(g_1, g_2, g_3) = \omega_{g_1}\left(\phi_2(g_2, g_3)\right) - \phi_2(g_1g_2, g_3) + \phi_2(g_1, g_2g_3) - \phi_2(g_1, g_2) \tag{15.637}$$

$$(d\phi_3)(g_1, g_2, g_3, g_4) = \omega_{g_1}\left(\phi_3(g_2, g_3, g_4)\right) - \phi_3(g_1g_2, g_3, g_4) + \phi_3(g_1, g_2g_3, g_4) - \phi_3(g_1, g_2, g_3g_4) + \phi_3(g_1, g_2, g_3) \tag{15.638}$$

Next, one can check that for any ϕ, we have the absolutely essential equation:

$$\boxed{d(d\phi) = 0} \tag{15.639}$$

We will give a simple proof of (15.639) below but let us just look at how it works for the lowest degrees: If $\phi_0 = a \in A$ is a 0-cochain then

$$\begin{aligned}(d^2\phi_0)(g_1, g_2) &= \omega_{g_1}(d\phi_0(g_2)) - d\phi_0(g_1 \cdot g_2) + d\phi_0(g_1) \\ &= \omega_{g_1}\left(\omega_{g_2}(a) - a\right) - \left(\omega_{g_1g_2}(a) - a\right) + \left(\omega_{g_1}(a) - a\right) \\ &= \omega_{g_1}(\omega_{g_2}(a)) - \omega_{g_1g_2}(a) \\ &= 0\end{aligned} \tag{15.640}$$

if ϕ_1 is any 1-cochain then we compute:

$$\begin{aligned}(d^2\phi_1)(g_1,g_2,g_3) &= \omega_{g_1}(d\phi_1(g_2,g_3)) - (d\phi_1)(g_1g_2,g_3) + (d\phi_1)(g_1,g_2g_3) - (d\phi_1)(g_1,g_2)\\ &= \omega_{g_1}\left(\omega_{g_2}(\phi_1(g_3)) - \phi_1(g_2g_3) + \phi_1(g_2)\right)\\ &- \left(\omega_{g_1g_2}(\phi_1(g_3)) - \phi_1(g_1g_2g_3) + \phi_1(g_1g_2)\right)\\ &+ \left(\omega_{g_1}(\phi_1(g_2g_3) - \phi_1(g_1g_2g_3) + \phi_1(g_1)\right)\\ &- \left(\omega_{g_1}(\phi_1(g_2)) - \phi_1(g_1g_2) + \phi_1(g_1)\right)\\ &= 0\end{aligned} \tag{15.641}$$

where you can check that all terms cancel in pairs, once you use $\omega_{g_1}\circ\omega_{g_2} = \omega_{g_1g_2}$.

The set of (ω-twisted) n-cocycles is defined to be the subgroup $Z^{n+\omega}(G,A)\subset C^n(G,A)$ of cochains that satisfy $d\phi_n = 0$.

Thanks to (15.639) we can define a subgroup $B^{n+\omega}(G,A)\subset Z^{n+\omega}(G,A)$, called the subgroup of coboundaries:

$$B^{n+\omega}(G,A) := \{\phi_n | \exists \phi_{n-1} \qquad s.t. \qquad d\phi_{n-1} = \phi_n\} \tag{15.642}$$

then, since $d^2 = 0$ we have $B^{n+\omega}(G,A)\subset Z^{n+\omega}(G,A)$.

Then the group cohomology is defined to be the quotient

$$H^{n+\omega}(G,A) = Z^{n+\omega}(G,A)/B^{n+\omega}(G,A) \tag{15.643}$$

Example: Let us take $G = \mathbb{Z}_2 = \{1,\sigma\}$ and $A = \mathbb{Z}$. Recall that

$$\mathrm{Aut}(A) = \mathrm{Aut}(\mathbb{Z}) = \{\mathrm{Id}_{\mathbb{Z}}, \mathcal{P}\} \cong \mathbb{Z}_2 \tag{15.644}$$

where $\mathcal{P}$ is the automorphism that takes $\mathcal{P}: n\to -n$. Now $\mathrm{Hom}(G,\mathrm{Aut}(\mathbb{Z}))\cong\mathbb{Z}_2$. Of course, $\omega_1 = \mathrm{Id}_{\mathbb{Z}}$ always and now we have two possibilities for ω_σ. Either $\omega_\sigma = \mathrm{Id}_{\mathbb{Z}}$ in which case we denote $\omega = T$ ("T" for trivial) or $\omega_\sigma = \mathcal{P}$ which we will denote $\mathcal{I}$. Let us compute $H^{1+\omega}(\mathbb{Z}_2,\mathbb{Z})$ for these two possibilities. First look at the subgroup of coboundaries. If $\phi_0 = n_0\in\mathbb{Z}$ is some integer then

$$\begin{aligned}&(d\phi_0)(1) = 0\\ &(d\phi_0)(\sigma) = \omega_\sigma(n_0) - n_0 = \begin{cases}0 & \omega = T\\ -2n_0 & \omega = \mathcal{I}\end{cases}\end{aligned} \tag{15.645}$$

Now consider the differential of a one-cochain:

$$\begin{aligned}(d\phi_1)(1,1) &= \omega_1(\phi_1(1)) - \phi_1(1) + \phi_1(1) = \phi_1(1)\\ (d\phi_1)(1,\sigma) &= \omega_1(\phi_1(\sigma)) - \phi_1(\sigma) + \phi_1(1) = \phi_1(1)\\ (d\phi_1)(\sigma,1) &= \omega_\sigma(\phi_1(1)) - \phi_1(\sigma) + \phi_1(\sigma) = \omega_\sigma(\phi_1(1))\\ (d\phi_1)(\sigma,\sigma) &= \omega_\sigma(\phi_1(\sigma)) - \phi_1(1) + \phi_1(\sigma)\end{aligned} \tag{15.646}$$

Now the cocycle condition implies $\phi_1(1) = 0$, making the first three lines of (15.646) vanish. Using this the fourth line becomes:

$$(d\phi_1)(\sigma,\sigma) = \omega_\sigma(\phi_1(\sigma)) + \phi_1(\sigma) = \begin{cases} 2\phi_1(\sigma) & \omega = T \\ 0 & \omega = \mathcal{I} \end{cases} \tag{15.647}$$

Now, when is ϕ_1 a cocycle? When $\omega = T$ is trivial then we must take $\phi_1(\sigma) = 0$ and hence $\phi_1 = 0$ moreover, there are no coboundaries. We find $H^{1+T}(\mathbb{Z}_2, \mathbb{Z}) = 0$ in this case, reproducing the simple fact that there are no nontrivial group homomorphisms from $\mathbb{Z}_2$ to $\mathbb{Z}$.

On the other hand, when $\omega = \mathcal{I}$ we can take $\phi_1(\sigma) = a$ to be any integer $a \in \mathbb{Z}$. The group of twisted cocycles is isomorphic to $\mathbb{Z}$. However, now there are nontrivial coboundaries, as we see from (15.645). We can shift a by any even integer $a \to a - 2n_0$. So

$$H^{1+\mathcal{I}}(\mathbb{Z}_2, \mathbb{Z}) \cong \mathbb{Z}_2 \tag{15.648}$$

In addition to the interpretation in terms of splittings, this has a nice interpretation in topology in terms of the unorientability of even-dimensional real projective spaces.

Remarks:

1. Previously we were denoting the cohomology groups by $H^{n+\omega}(G, A)$. In the equations above the ω is still present, (see the first term in the definition of $d\phi$) but we leave the ω implicit in the notation. Nevertheless, we are talking about the same groups as before, but now generalizing to arbitrary degree n.

2. Remembering that we are now writing our abelian group A additively, we see that the equation $(d\phi_2) = 0$ is just the twisted 2-cocycle conditions, and $\phi_2' = \phi_2 + d\phi_1$ are two different twisted cocycles related by a coboundary. See equations (15.538) and (15.540) above. Roughly speaking, you should "take the logarithm" of these equations.

3. *Homological Algebra*: What we are discussing here is a special case of a topic known as homological algebra. Quite generally, a *chain complex* is a sequence of Abelian groups $\{C_n\}_{n\in\mathbb{Z}}$ equipped with group homomorphisms

$$\partial_n : C_n \to C_{n-1} \tag{15.649}$$

such that $\partial_n \circ \partial_{n+1} = 0$ for all $n \in \mathbb{Z}$. A *cochain complex* is similarly a sequence of Abelian groups $\{C^n\}_{n\in\mathbb{Z}}$ with group homomorphisms $d_n : C^n \to C^{n+1}$ so that $d_{n+1} \circ d_n = 0$ for all $n \in \mathbb{Z}$. Note that these are NOT exact sequences. Indeed the failure to be an exact sequence is measured by the *homology groups* of the chain complex

$$H_n(C_*, \partial_*) := \ker(\partial_n)/\mathrm{im}(\partial_{n+1}) \tag{15.650}$$

and the *cohomology groups* of the cochain complex:

$$H^n(C^*, d_*) := \ker(d_n)/\mathrm{im}(d_{n-1}) \tag{15.651}$$

4. *Homogeneous cocycles*: A nice way to prove that $d^2 = 0$ is the following. We define *homogeneous* n-cochains to be maps $\varphi : G^{n+1} \to A$ which satisfy

$$\varphi(hg_0, hg_1, \ldots, hg_n) = \omega_h\left(\varphi(g_0, g_1, \ldots, g_n)\right) \tag{15.652}$$

Let $\mathcal{C}^n(G, A)$ denote the abelian group of such homogeneous group cochains. (Warning! Elements of $\mathcal{C}^n(G, A)$ have $(n+1)$ arguments!) Define

$$\delta : \mathcal{C}^n(G, A) \to \mathcal{C}^{n+1}(G, A) \tag{15.653}$$

by

$$\delta\varphi(g_0, \ldots, g_{n+1}) := \sum_{i=0}^{n+1} (-1)^i \varphi(g_0, \ldots, \widehat{g_i}, \ldots, g_{n+1}) \tag{15.654}$$

where $\widehat{g_i}$ means the argument is omitted. Clearly, if φ is homogeneous then $\delta\varphi$ is also homogeneous. It is then very straightforward to prove that $\delta^2 = 0$. Indeed, if $\varphi \in \mathcal{C}^{n-1}(G, A)$ we compute:

$$\begin{aligned}
\delta^2\varphi(g_0, \ldots, g_{n+1}) &= \sum_{i=0}^{n+1} (-1)^i \Bigg\{ \sum_{j=0}^{i-1} (-1)^j \varphi(g_0, \ldots, \widehat{g_j}, \ldots, \widehat{g_i}, \ldots, g_{n+1}) \\
&\qquad - \sum_{j=i+1}^{n+1} (-1)^j \varphi(g_0, \ldots, \widehat{g_i}, \ldots, \widehat{g_j}, \ldots, g_{n+1}) \Bigg\} \\
&= \sum_{0 \le j < i \le n+1} (-1)^{i+j} \varphi(g_0, \ldots, \widehat{g_j}, \ldots, \widehat{g_i}, \ldots, g_{n+1}) \\
&\qquad - \sum_{0 \le i < j \le n+1} (-1)^{i+j} \varphi(g_0, \ldots, \widehat{g_i}, \ldots, \widehat{g_j}, \ldots, g_{n+1}) \\
&= 0
\end{aligned} \tag{15.655}$$

Now, we can define an isomorphism $\psi : \mathcal{C}^n(G, A) \to C^n(G, A)$ by defining

$$\phi_n(g_1, \ldots, g_n) := \varphi_n(1, g_1, g_1 g_2, \ldots, g_1 \cdots g_n) \tag{15.656}$$

That is, when ϕ_n and φ_n are related this way we say $\phi_n = \psi(\varphi_n)$. Now one can check that the simple formula (15.654) becomes the more complicated formula (15.634). Put more formally: there is a unique d so that $d\psi = \psi\delta$, or even more formally, there is a unique group homomorphism d such that we have a commutative diagram:

$$\begin{array}{ccc}
\mathcal{C}^n(G, A) & \xrightarrow{\delta} & \mathcal{C}^{n+1}(G, A) \\
\downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \psi} \\
C^n(G, A) & \xrightarrow{d} & C^{n+1}(G, A)
\end{array} \tag{15.657}$$

For example, if

$$\phi_1(g) = \psi(\varphi_1)(g) = \varphi_1(1, g) \tag{15.658}$$

then we can check that

$$\begin{aligned}(d\phi_1)(g_1,g_2) &= d(\psi(\varphi_1))(g_1,g_2)\\ &= \psi(\delta\varphi_1)(g_1,g_2)\\ &= \delta\varphi_1(1,g_1,g_1g_2)\\ &= \varphi_1(g_1,g_1g_2)-\varphi_1(1,g_1g_2)+\varphi_1(1,g_1)\\ &= \omega_{g_1}(\varphi_1(1,g_2))-\varphi_1(1,g_1g_2)+\varphi_1(1,g_1)\\ &= \omega_{g_1}(\phi_1(g_2))-\phi_1(g_1g_2)+\phi_1(g_1)\end{aligned} \tag{15.659}$$

in accord with the previous definition!

5. Where do all these crazy formulae come from? The answer is in topology. We will indicate it briefly in our discussion of categories and groupoids below.

6. The reader will probably find these formulae a bit opaque. It is therefore good to stop and think about what the cohomology is measuring, at least in low degrees.

Exercise
Derive the formula for the differential on an inhomogeneous cochain $d\phi_2$ starting with the definition on the analogous homogeneous cochain φ_3

Exercise
If (C_n, ∂_n) is a chain complex show that one can define a cochain complex with groups:

$$C^n := \mathrm{Hom}(C_n, \mathbb{Z}) \tag{15.660}$$

15.8.2 Interpreting the meaning of $H^{0+\omega}$

A zero-cocycle is an element $a \in A$ so that for all g

$$\omega_g(a) = a \tag{15.661}$$

There are no coboundaries to worry about, so $H^0(G,A)$ is just the set of fixed points of the G action on A.

15.8.3 Interpreting the meaning of $H^{1+\omega}$

We have interpreted $H^{1+\omega}(G,A)$ above as the set of nontrivial splittings of the semidirect product defined by ω:

$$0 \to A \to A \rtimes G \to G \to 1 \tag{15.662}$$

15.8.4 Interpreting the meaning of $H^{2+\omega}$

Again, we have interpreted $H^{2+\omega}(G,A)$ as $\mathrm{Ext}^{\omega}(G,A)$, the set of equivalence classes of extensions

$$0 \to A \to \tilde{G} \to G \to 1 \tag{15.663}$$

inducing a fixed $\omega : G \to \mathrm{Aut}(A)$. The trivial element of the cohomology group corresponds to the semi-direct product and the set of inequivalent trivializations is the group $H^{1+\omega}(G,A)$ of splittings of the semi-direct product.

More generally, $\mathrm{Ext}^{\bar{\omega}}(Q,N)$ is a torsor for $H^{2+\bar{\omega}}(Q,Z(N))$.

15.8.5 Interpreting the meaning of H^3

To see one interpretation of H^3 in terms of extension theory let us return to the analysis of general extensions in §15.7.

Recall that, as we have discussed using (15.625), a general extension (15.1) has a canonically associated homomorphism

$$\bar{\omega} : Q \to \mathrm{Out}(N) \tag{15.664}$$

where $\mathrm{Out}(N)$ is the group of outer automorphisms of N.

The natural question arises: *Given a homomorphism $\bar{\omega}$ as in (15.664) is there a corresponding extension of Q by N inducing $\bar{\omega}$ as in equation (15.625) ?*

To answer this question we could proceed by *choosing* for each $q \in Q$ an automorphism $\xi_q \in \mathrm{Aut}(N)$ such that $[\xi_q] = \bar{\omega}_q$ in $\mathrm{Out}(N)$. To do this, choose a section s of $\pi : \mathrm{Aut}(N) \to \mathrm{Out}(N)$ and let $\xi_q := s(\bar{\omega}_q)$. If we cannot split the sequence

$$1 \to \mathrm{Inn}(N) \to \mathrm{Aut}(N) \to \mathrm{Out}(N) \to 1 \tag{15.665}$$

then $q \mapsto \xi_q$ will not be a group homomorphism. But we do know that for all $q_1, q_2 \in Q$

$$\xi_{q_1} \circ \xi_{q_2} \circ \xi_{q_1 q_2}^{-1} \in \mathrm{Inn}(N) \tag{15.666}$$

Therefore, for every q_1, q_2 we may *choose* an element $f(q_1,q_2) \in N$ so that

$$\xi_{q_1} \circ \xi_{q_2} \circ \xi_{q_1 q_2}^{-1} = I(f(q_1,q_2)) \tag{15.667}$$

i.e.

$$\xi_{q_1} \circ \xi_{q_2} = I(f(q_1,q_2)) \circ \xi_{q_1 q_2} \tag{15.668}$$

Of course, the choice of $f(q_1,q_2)$ is ambiguous by an element of $Z(N)$!

Equation (15.668) is of course just (15.615) written in slightly different notation. Therefore, as we saw in §15.7, if $f(q_1,q_2)$ were to satisfy the the "twisted cocycle condition" (15.616) then we could use (15.617) to define an extension inducing $\bar{\omega}$.

Therefore, let us check if some choice of $f(q_1,q_2)$ actually does satisfy the twisted cocycle condition (15.616). Looking at the RHS of (15.616) we compute:

$$\begin{aligned} I(f(q_1,q_2)f(q_1q_2,q_3)) &= I(f(q_1,q_2))I(f(q_1q_2,q_3)) \\ &= \left(\xi_{q_1} \circ \xi_{q_2} \circ \xi_{q_1q_2}^{-1}\right) \circ \left(\xi_{q_1q_2} \circ \xi_{q_3} \circ \xi_{q_1q_2q_3}^{-1}\right) \\ &= \xi_{q_1} \circ \xi_{q_2} \circ \xi_{q_3} \circ \xi_{q_1q_2q_3}^{-1} \end{aligned} \tag{15.669}$$

On the other hand, looking at the LHS of (15.616) we compute:

$$\begin{aligned} I(\xi_{q_1}(f(q_2,q_3))f(q_1,q_2q_3)) &= I(\xi_{q_1}(f(q_2,q_3)))I(f(q_1,q_2q_3)) \\ &= \xi_{q_1}\circ I((f(q_2,q_3)))\circ\xi_{q_1}^{-1}\circ I(f(q_1,q_2q_3) \\ &= \xi_{q_1}\circ\left(\xi_{q_2}\circ\xi_{q_3}\circ\xi_{q_2q_3}^{-1}\right)\circ\xi_{q_1}^{-1}\circ\left(\xi_{q_1}\circ\xi_{q_2q_3}\circ\xi_{q_1q_2q_3}^{-1}\right) \\ &= \xi_{q_1}\circ\xi_{q_2}\circ\xi_{q_3}\circ\xi_{q_1q_2q_3}^{-1} \end{aligned} \tag{15.670}$$

Therefore, comparing (15.669) and (15.670) we conclude that

$$I(\xi_{q_1}(f(q_2,q_3))f(q_1,q_2q_3)) = I(f(q_1,q_2)f(q_1q_2,q_3)) \tag{15.671}$$

We cannot conclude that f satisfies the twisted cocycle equation from this identity because inner transformations are trivial for elements in the center $Z(N)$. Rather, what we can conclude is that for every q_1,q_2,q_3 there is an element $z(q_1,q_2,q_3)\in Z(N)$ such that

$$f(q_1,q_2)f(q_1q_2,q_3) = z(q_1,q_2,q_3)\xi_{q_1}(f(q_2,q_3))f(q_1,q_2q_3) \tag{15.672}$$

Now, one can check (with a lot of algebra) that

1. z is a cocycle in $Z^{3+\bar{\omega}}(Q,Z(N))$. (We are using $\mathrm{Aut}(Z(N))\cong\mathrm{Out}(Z(N))$.)
2. Changes in choices of ξ_q and $f(q_1,q_2)$ lead to changes in z by a coboundary.

and therefore we conclude:

Theorem 15.8.5.1 : Given $\bar{\omega}:Q\to\mathrm{Out}(N)$ there exists an extension of Q by N iff the cohomology class $[z]\in H^3(Q,Z(N))$ vanishes.

Moreover, as we have seen, if $[z]=0$ then the trivializations of z are in 1-1 correspondence with elements $H^2(Q,Z(N))$ and are hence in 1-1 correspondence with isomorphism classes of extensions of Q by N. This is the analogue, one step up in degree, of our interpretation of $H^1(G,A)$.

Examples: As an example [221] where a degree three cohomology class obstructs the existence of an extension inducing a homomorphism $\bar{\omega}:Q\to\mathrm{Out}(N)$ we can take N to be the generalized quaternion group of order 16. It is generated by x and y satisfying:

$$x^4=y^2 \qquad x^8=1 \qquad yxy^{-1}=x^{-1} \tag{15.673}$$

Using these relations every word in $x^{\pm1}$ and $y^{\pm1}$ can be reduced to either x^m, or yx^m, with $m=0,\ldots,7$, and these words are all different. One can show the outer automorphism group $\mathrm{Out}(N)\cong\mathbb{Z}_2\times\mathbb{Z}_2$ with generators α,β acting by

$$\alpha(x)=x^3,\alpha(y)=y \qquad\qquad \beta(x)=x,\beta(y)=yx \tag{15.674}$$

Then there is is no group extension with group G fitting in

$$1\to N\to G\to\mathbb{Z}_2\to 1 \tag{15.675}$$

[221] I learned these nice examples from Clay Cordova. They will appear in a forthcoming paper with Po-Shen Hsin and Francesco Benini.

inducing the homomorphism $\bar{\omega} : \mathbb{Z}_2 \to \mathrm{Out}(N)$ defined by $\bar{\omega}(\sigma) = \alpha \circ \beta$ where σ is the nontrivial element of $\mathbb{Z}_2$. One way to prove this is to look up the list of groups of order 32 and search for those with maximal normal subgroup given by N. [222] There are five such. Then one computes $\bar{\omega}$ for each such extension and finds that it is never of the above type. A similar example can be constructed by taking N to be a dihedral group of order 16.

♣Really should explain four-term sequences and crossed modules here.... ♣

Remark There is an interpretation of $H^3(Q, Z(N))$ as a classification of four-term exact sequences, and there are generalizations of this to higher degree. See

1. K. Brown, *Group Cohomology*.
2. C. A. Weibel, *An introduction to homological algebra*, chapter 6

15.9 Some references

Some online sources with links to further material are

1. http://en.wikipedia.org/wiki/Group-extension
2. http://ncatlab.org/nlab/show/group+extension

3 http://terrytao.wordpress.com/2010/01/23/some-notes-on-group-extensions/

4. Section 15.8.5, known as the Artin-Schreier theory, is based on a nice little note by P.J. Morandi,

http://sierra.nmsu.edu/morandi/notes/GroupExtensions.pdf

5. Jungmann, Notes on Group Theory

6. S. MacLane, "Topology And Logic As A Source Of Algebra," Bull. Amer. Math. Soc. 82 (1976), 1-4.

Textbooks:

1. K. Brown, Group Cohomology
2. Karpilovsky, The Schur Multiplier
3. C. A. Weibel, *An introduction to homological algebra*, chapter 6

16. Overview of general classification theorems for finite groups

In general if a mathematical object proves to be useful then there is always an associated important problem, namely the *classification* of these objects.

For example, with groups we can divide them into classes: finite and infinite, abelian and nonabelian producing a four-fold classication:

Finite abelian	Finite nonabelian
Infinite abelian	Infinite nonabelian

[222] See, for example, B. Shuster, "Morava K-theory of groups of order 32," Algebr. Geom. Topol. **11** (2011) 503-521.

But this is too rough, it does not give us a good feeling for what the examples really are.

Once we have a "good" criterion we often can make a nontrivial statement about the general structure of objects in a given class. Ideally, we should be able to construct all the examples algorithmically, and be able to distinguish the ones which are not isomorphic. Of course, finding such a "good" criterion is an art. For example, classification of infinite nonabelian groups is completely out of the question. But in Chapter *** we will see that an important class of infinite nonabelian groups, the simply connected compact simple Lie groups, have a very beautiful classification: There are four infinite sequences of classical matrix groups: $SU(n), Spin(n), USp(2n)$ and then five exceptional cases with names G_2, F_4, E_6, E_7, E_8. [223]

One might well ask: Can we classify finite groups? In this section we survey a little of what is known about this problem.

16.1 Brute force

If we just start listing groups of low order we soon start to appreciate what a jungle is out there.

But let us try, if only as an exercise in applying what we have learned so far. First, let us note that for groups of order p where p is prime we automatically have the unique possibility of the cyclic group $\mathbb{Z}/p\mathbb{Z}$. Similarly, for groups of order p^2 there are precisely two possibilities: $\mathbb{Z}/p^2\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. This gets us through many of the low order cases.

Given this remark the first nontrivial order to work with is $|G| = 6$. By Cauchy's theorem there are elements of order 2 and 3. Call them b, with $b^2 = 1$ and a with $a^3 = 1$. Then $(bab)^3 = 1$, so either

1. $bab = a$ which implies $ab = ba$ which implies $G = \mathbb{Z}_2 \times \mathbb{Z}_3 = \mathbb{Z}_6$
2. $bab = a^{-1}$ which implies $G = D_3$.

This is the first place we meet a nonabelian group. It is the dihedral group, the first of the series we saw before

$$D_n = \langle a, b | a^n = b^2 = 1, bab = a^{-1} \rangle \tag{16.1}$$

and has order $2n$. There is a special isomorphism $D_3 \cong S_3$ with the symmetric group on three letters.

The next nontrivial case is $|G| = 8$. Here we can invoke Sylow's theorem: If $p^k || |G|$ then G has a subgroup of order p^k. Let us apply this to 4 dividing $|G|$. Such a subgroup has index two and hence must be a normal subgroup, and hence fits in a sequence

$$1 \to N \to G \to \mathbb{Z}_2 \to 1 \tag{16.2}$$

Now, N is of order 4 so we know that $N \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ or $N \cong \mathbb{Z}_4$. If we have

$$1 \to \mathbb{Z}_4 \to G \to \mathbb{Z}_2 \to 1 \tag{16.3}$$

[223] $Spin(n)$ double covers the classical matrix group $SO(n)$.

then we have $\alpha : \mathbb{Z}_2 \to \mathrm{Aut}(\mathbb{Z}_4) \cong \mathbb{Z}_2$ and there are exactly two such homomorphisms. Moreover, for a fixed α there are two possibilities for the square $\tilde{\sigma}^2 \in \mathbb{Z}_4$ where $\tilde{\sigma}$ is a lift of the generator of $\mathbb{Z}_2$. Altogether this gives four possibilities:

♣Need to explain more here. ♣

$$1 \to \mathbb{Z}_4 \to \mathbb{Z}_2 \times \mathbb{Z}_4 \to \mathbb{Z}_2 \to 1 \tag{16.4}$$

$$1 \to \mathbb{Z}_4 \to \mathbb{Z}_8 \to \mathbb{Z}_2 \to 1 \tag{16.5}$$

$$1 \to \mathbb{Z}_4 \to D_4 \to \mathbb{Z}_2 \to 1 \tag{16.6}$$

$$1 \to \mathbb{Z}_4 \to \widetilde{D}_2 \to \mathbb{Z}_2 \to 1 \tag{16.7}$$

Here we meet the first of the series of *dicyclic* or *binary dihedral* groups defined by

$$\widetilde{D_n} := \langle a, b | a^{2n} = 1, a^n = b^2, b^{-1}ab = a^{-1} \rangle \tag{16.8}$$

It has order $4n$. There is a special isomorphism of $\widetilde{D}_2$ with the quaternion group.

The other possibility for N is $\mathbb{Z}_2 \times \mathbb{Z}_2$ and here one new group is found, namely $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

Thus there are 5 inequivalent groups of order 8.

The next few cases are trivial until we get to $|G| = 12$. By Cauchy's theorem there are subgroups isomorphic to $\mathbb{Z}_2$, so we can view G as an extension of D_3 or $\mathbb{Z}_6$ by $\mathbb{Z}_2$. There is also a subgroup isomorphic to $\mathbb{Z}_3$ so we can view it as an extension of an order 4 group by an order 3 group. We skip the analysis and just present the 5 distinct order 12 groups. In this way we find the groups forming the pattern at lower order:

♣Check this reasoning is correct. You need to know the subgroups are normal to say there is an extension. ♣

$$\mathbb{Z}_{12}, \quad , \mathbb{Z}_2 \times \mathbb{Z}_6, \quad , D_6, \quad , \tilde{D}_3 \tag{16.9}$$

And we find one "new" group: $A_4 \subset S_4$.

We can easily continue the table until we get to order $|G| = 16$. At order 16 there are 14 inequivalent groups! So we will stop here. [224]

[224] See, however, M. Wild, "Groups of order 16 made easy," American Mathematical Monthly, Jan 2005

Order	Presentation	name
1	$\langle a \mid a = 1 \rangle$	Trivial group
2	$\langle a \mid a^2 = 1 \rangle$	Cyclic $\mathbb{Z}/2\mathbb{Z}$
3	$\langle a \mid a^3 = 1 \rangle$	Cyclic $\mathbb{Z}/3\mathbb{Z}$
4	$\langle a \mid a^4 = 1 \rangle$	Cyclic $\mathbb{Z}/4\mathbb{Z}$
4	$\langle a, b \mid a^2 = b^2 = (ab)^2 = 1 \rangle$	Dihedral $D_2 \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, Klein
5	$\langle a \mid a^5 = 1 \rangle$	Cyclic $\mathbb{Z}/5\mathbb{Z}$
6	$\langle a, b \mid a^3 = 1, b^2 = 1, bab = a \rangle$	Cyclic $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$
6	$\langle a, b \mid a^3 = 1, b^2 = 1, bab = a^{-1} \rangle$	Dihedral $D_3 \cong S_3$
7	$\langle a \mid a^7 = 1 \rangle$	Cyclic $\mathbb{Z}/7\mathbb{Z}$
8	$\langle a \mid a^8 = 1 \rangle$	Cyclic $\mathbb{Z}/8\mathbb{Z}$
8	$\langle a, b \mid a^2 = 1, b^4 = 1, aba = b \rangle$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$
8	$\langle a, b, c \mid a^2 = b^2 = c^2 = 1, [a, b] = [a, c] = [b, c] = 1 \rangle$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$
8	$\langle a, b \mid a^4 = 1, b^2 = 1, bab = a^{-1} \rangle$	Dihedral D_4
8	$\langle a, b \mid a^4 = 1, a^2 = b^2, b^{-1}ab = a^{-1} \rangle$	Dicyclic $\widetilde{D_2} \cong Q$, quaternion
9	$\langle a \mid a^9 = 1 \rangle$	Cyclic $\mathbb{Z}/9\mathbb{Z}$
9	$\langle a, b \mid a^3 = b^3 = 1, [a, b] = 1 \rangle$	$\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$
10	$\langle a \mid a^{10} = 1 \rangle$	Cyclic $\mathbb{Z}/10\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$
10	$\langle a \mid a^5 = b^2 = 1, bab = a^{-1} \rangle$	Dihedral D_5
11	$\langle a \mid a^{11} = 1 \rangle$	Cyclic $\mathbb{Z}/11\mathbb{Z}$
12	$\langle a \mid a^{12} = 1 \rangle$	Cyclic $\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$
12	$\langle a, b \mid a^2 = 1, b^6 = 1, [a, b] = 1 \rangle$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$
12	$\langle a, b \mid a^6 = 1, b^2 = 1, bab = a^{-1} \rangle$	Dihedral D_6
12	$\langle a, b \mid a^6 = 1, a^3 = b^2, b^{-1}ab = a^{-1} \rangle$	Dicyclic $\widetilde{D_3}$
12	$\langle a, b \mid a^3 = 1, b^2 = 1, (ab)^3 = 1 \rangle$	Alternating A_4
13	$\langle a \mid a^{13} = 1 \rangle$	Cyclic $\mathbb{Z}/13\mathbb{Z}$
14	$\langle a \mid a^{14} = 1 \rangle$	Cyclic $\mathbb{Z}/14\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/7\mathbb{Z}$
14	$\langle a, b \mid a^7 = 1, b^2 = 1, bab = a^{-1} \rangle$	Dihedral D_7
15	$\langle a \mid a^{15} = 1 \rangle$	Cyclic $\mathbb{Z}/15\mathbb{Z} \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$

Remarks:

1. Explicit tabulation of the isomorphism classes of groups was initiated by by Otto Holder who completed a table for $|G| \leq 200$ about 100 years ago. Since then there has been much effort in extending those results. For surveys see

   1. J.A. Gallan, "The search for finite simple groups," Mathematics Magazine, vol. 49 (1976) p. 149. (This paper is a bit dated.)

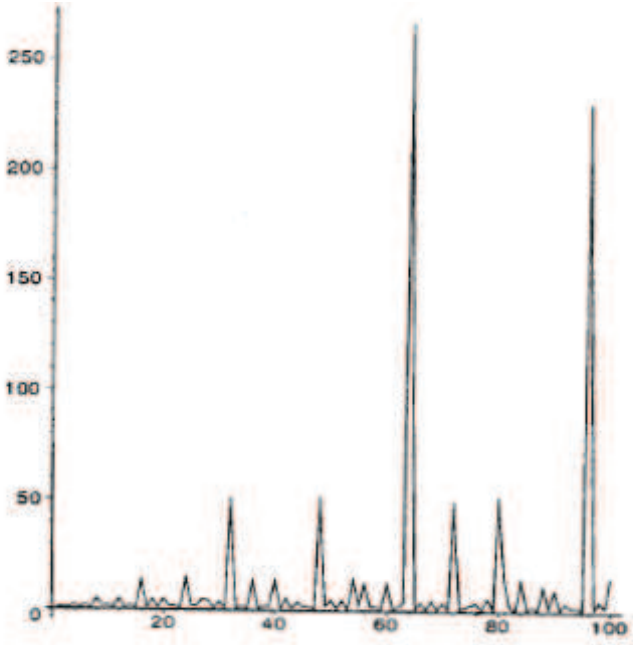


Figure 41: A plot of the number of nonisomorphic groups of order n. This plot was taken from the book by D. Joyner, *Adventures in Group Theory*.

2. H.U. Besche, B. Eick, E.A. O'Brian, "A millenium project: Constructing Groups of Small Order,"

2. There are also nice tables of groups of low order, in Joyner, *Adventures in Group Theory*, pp. 168-172, and Karpilovsky, *The Schur Multiplier* which go beyond the above table.

3. There are also online resources:

   1. http://www.gap-system.org/ for GAP
   2. http://hobbes.la.asu.edu/groups/groups.html for groups of low order.
   3. http://www.bluetulip.org/programs/finitegroups.html
   4. http://en.wikipedia.org/wiki/List-of-small-groups

4. The number of isomorphism types of groups jumps wildly. Apparently, there are $49,487,365,422$ isomorphism types of groups of order $2^{10} = 1024$. (Besche et. al. loc. cit.) The remarkable plot of Figure 41 from Joyner's book shows a plot of the number of isomorphism classes vs. order up to order 100. Figure 42 shows a log plot of the number of groups up to order 2000.

5. There is, however, a formula giving the asymptotics of the number $f(n,p)$ of isomorphism classes of groups of order p^n for $n \to \infty$ for a fixed prime p. (Of course, there are $p(n)$ Abelian groups, where $p(n)$ the the number of partitions of n. Here we are

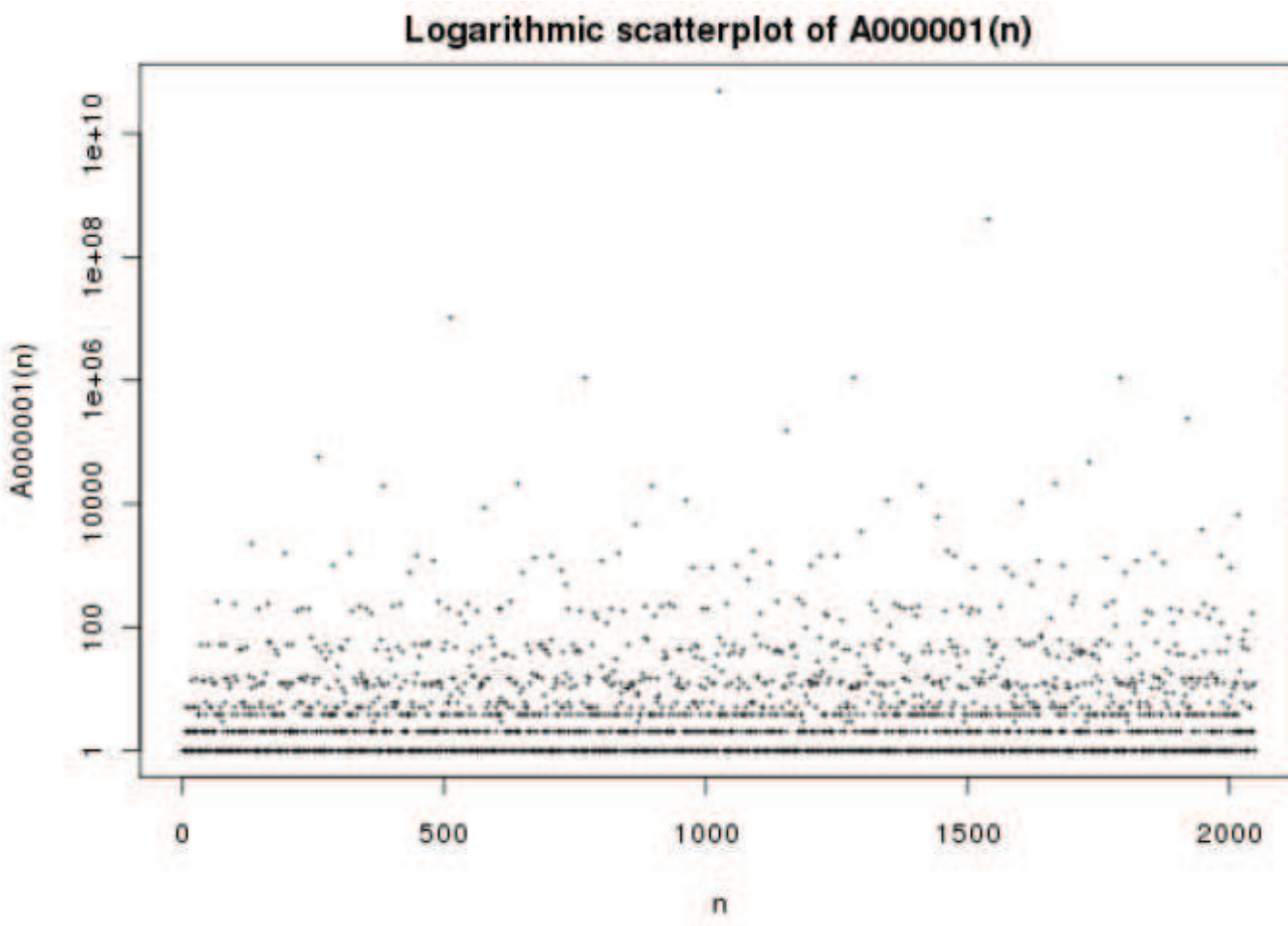


Figure 42: A logarithmic plot of the number of nonisomorphic groups of order n out to $n \leq 2000$. This plot was taken from online encyclopedia of integer sequences, OEIS.

talking about the number of all groups.) This is due to G. Higman [225] and C. Sims [226] and the result states that:

$$f(n,p) \sim p^{\frac{2}{27}n^3} \tag{16.10}$$

Note that the asymptotics we derived for $p(n)$ before had a growth like $e^{const.n^{1/2}}$ so, unsurprisingly, most of the groups are nonabelian.

Exercise *Relating the binary dihedral and dihedral groups*

[225] G. Higman, "Enumerating p-Groups," Proc. London Math. Soc. 3) 10 (1960)

[226] C. Sims, "Enumerating p-Groups," Proc. London Math. Soc. (3) IS (1965) 151-66

Show that $\widetilde{D}_n$ is a double-cover of D_n which fits into the exact sequence:

$$\begin{array}{ccccccccc} & & \mathbb{Z}_2 & = & \mathbb{Z}_2 & & & & \\ & & \downarrow & & \downarrow & & & & \\ 1 & \longrightarrow & \mathbb{Z}_{2n} & \longrightarrow & \widetilde{D}_n & \longrightarrow & \mathbb{Z}_2 & \longrightarrow & 1 \\ & & \| & & \downarrow & & \| & & \\ 1 & \longrightarrow & \mathbb{Z}_n & \longrightarrow & D_n & \longrightarrow & \mathbb{Z}_2 & \longrightarrow & 1 \end{array} \tag{16.11}$$

16.2 Finite Abelian Groups

The upper left box of our rough classification can be dealt with thoroughly, and the result is extremely beautiful.

In this subsection we will write our abelian groups *additively*.

Recall that we have shown that if p and q are positive integers then

$$0 \to \mathbb{Z}/gcd(p,q)\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/q\mathbb{Z} \to \mathbb{Z}/lcm(p,q)\mathbb{Z} \to 0 \tag{16.12}$$

and in particular, if p, q are relatively prime then

$$\mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/q\mathbb{Z} \cong \mathbb{Z}/pq\mathbb{Z}. \tag{16.13}$$

It thus follows that if n has prime decomposition

$$n = \prod_i p_i^{e_i} \tag{16.14}$$

then

$$\mathbb{Z}/n\mathbb{Z} \cong \oplus_i \mathbb{Z}/p_i^{e_i}\mathbb{Z} \tag{16.15}$$

This decomposition has a beautiful generalization to an arbitrary finite abelian group:

Kronecker Structure Theorem. Any finite abelian group is a direct product of cyclic groups of order a prime power. That is, we firstly have the decomposition:

$$\begin{aligned} G &= G_2 \oplus G_3 \oplus G_5 \oplus G_7 \oplus \cdots \\ &= \oplus_{p \text{ prime}} G_p \end{aligned} \tag{16.16}$$

where G_p has order p^n for some $n \geq 0$ (n can depend on p, and for all but finitely many p, $G_p = \{0\}$.) And, secondly, each nonzero factor G_p can be written:

$$G_p = \oplus_i \mathbb{Z}/(p^{n_i}\mathbb{Z}) \tag{16.17}$$

for some finite collection of positive integers n_i (depending on p).

Proof: The proof proceeds in two parts. The first, easy, part shows that we can split G into a direct sum of "p-groups" (defined below). The second, harder, part shows that an arbitrary abelian p-group is a direct sum of cyclic groups.

For part 1 of the proof let us consider an arbitrary finite abelian group G. We will write the group multiplication additively. Suppose n is an integer so that $ng = 0$ for all $g \in G$. To fix ideas let us take $n = |G|$. Suppose $n = m_1 m_2$ where m_1, m_2 are relatively prime integers. Then there are integers s_1, s_2 so that

$$s_1 m_1 + s_2 m_2 = 1 \tag{16.18}$$

Therefore any element g can be written as

$$g = s_1(m_1 g) + s_2(m_2 g) \tag{16.19}$$

Now $m_1 G$ and $m_2 G$ are subgroups and we claim that $m_1 G \cap m_2 G = \{0\}$. If $a \in m_1 G \cap m_2 G$ then $m_1 a = 0$ and $m_2 a = 0$ and hence (16.19) implies $a = 0$. Thus,

$$G = m_1 G \oplus m_2 G \tag{16.20}$$

Moreover, we claim that $m_1 G = \{g \in G | m_2 g = 0\}$. It is clear that every element in $m_1 G$ is killed by m_2. Suppose on the other hand that $m_2 g = 0$. Again applying (16.19) we see that $g = s_1 m_1 g = m_1(s_1 g) \in m_1 G$.

Thus, we can decompose

$$G = \oplus p \text{ prime} G_p \tag{16.21}$$

where G_p is the subgroup of G of elements whose order is a power of p.

If p is a prime number then a *p-group* is a group all of whose elements have order a power of p. Now for part 2 of the proof we show that any abelian p-group is a direct sum of the form (16.17). The proof of this statement proceeds by induction and is based on a systematic application of Cauchy's theorem: If p divides $|G|$ then there is an element of G of order precisely p. (Recall we proved this theorem in Section 9.

Now, note that any p-group G has an order which is a power p^n for some n. If not, then $|G| = p^n m$ where m is relatively prime to p. But then - by Cauchy's theorem - there would have to be an element of G whose order is a prime divisor of m.

Next we claim that if an abelian p-group has a *unique* subgroup H of order p then G itself is cyclic.

To prove this we again proceed by induction on $|G|$. Consider the subgroup defined by:

$$H = \{g | pg = 0\} \tag{16.22}$$

From Cauchy's theorem we see that H cannot be the trivial group, and hence this must be the unique subgroup of order p. On the other hand, H is manifestly the kernel of the homomorphism $\phi : G \to G$ given by $\phi(g) = pg$. Again by Cauchy, $\phi(G)$ has a subgroup of order p, but this must also be a subgroup of G, which contains $\phi(G)$, and hence $\phi(G)$ has a unique subgroup of order p. By the induction hypothesis, $\phi(G)$ is cyclic. But now $\phi(G) \cong G/H$, so let $g_0 + H$ be a generator of the cyclic group G/H. Next we claim

that $H \subset \langle g_0 \rangle$. Since G is a p-group the subgroup $\langle g_0 \rangle$ is a p-group and hence contains a subgroup of order p (by Cauchy) but (by hypothesis) there is a unique such subgroup in G and any subgroup of $\langle g_0 \rangle$ is a subgroup of G, so $H \subset \langle g_0 \rangle$. But now take any element $g \in G$. On the one hand it must project to an element $[g] \in G/H$. Thus must be of the form $[g] = kg_0 + H$, since $g_0 + H$ generates G/H. That means $g = kg_0 + h$, $h \in H$, but since $H \subset \langle g_0 \rangle$ we must have $h = \ell g_0$ for some integer ℓ. Therefore $G = \langle g_0 \rangle$ is cyclic.

The final step proceeds by showing that if G is a finite abelian p-group and M is a cyclic subgroup of maximal order then $G = M \oplus N$ for some subgroup N. Once we have established this the desired result follows by induction.

So, now suppose that that G has a cyclic subgroup of maximal order M. If G is cyclic then $N = \{0\}$. If G is not cyclic then we just proved that there must be at least two distinct subgroups of order p. One of them is in M. Choose another one, say K. Note that K must not be in M, because M is cyclic and has a unique subgroup of order p. Therefore $K \cap M = \{0\}$. Therefore $(M + K)/K \cong M$. Therefore $(M + K)/K$ is a cyclic subgroup of G/K. Any element $g + K$ has an order which divides $|g|$, and $|g| \leq |M|$ since M is a maximal cyclic subgroup. Therefore the cyclic subgroup $(M + K)/K$ is a maximal order cyclic subgroup of G/K. Now the inductive hypothesis implies $G/K = (M+K)/K \oplus H/K$ for some subgroup $K \subset H \subset G$. But this means $(M+K) \cap H = K$ and hence $M \cap H = \{0\}$ and hence $G = M \oplus H$. ♠

For other proofs see

1. S. Lang, *Algebra*, ch. 1, sec. 10.

2. I.N. Herstein, Ch. 2, sec. 14.

3. J. Stillwell, *Classical Topology and Combinatorial Group Theory.*

4. Our proof is based on G. Navarro, "On the fundamental theorem of finite abelian groups," Amer. Math. Monthly, Feb. 2003, vol. 110, p. 153.

One class of examples where we have a finite Abelian group, but it's Kronecker decomposition is far from obvious is the following: Consider the Abelian group $\mathbb{Z}^d$. Choose a set of d vectors $v_i \in \mathbb{Z}^d$, linearly independent as vectors in $\mathbb{R}^d$.

$$L := \{\sum_{i=1}^{d} n_i v_i | n_i \in \mathbb{Z}\} \tag{16.23}$$

is a subgroup. Then

$$A = \mathbb{Z}^d / L \tag{16.24}$$

is a finite Abelian group. For example if $v_i = ke_i$ where e_i is the standard unit vector in the ith direction then obviously $A \cong (\mathbb{Z}/k\mathbb{Z})^d$. But for a general set of vectors the decomposition is not obvious.

So, here is an algorithm for giving the Kronecker decomposition of a finite Abelian group:

1. Compute the orders of the various elements.

2. You need only consider the elements whose order is a prime power. (By the Bezout manipulation all the others will be sums of these.)

3. Focusing on one prime at a time. Take the element g_1 whose order is maximal. Then $G_p = \langle g_1 \rangle \oplus N$.

♣Have to say how to get N. ♣

4. Repeat for N.

Exercise

Show that an alternative of the structure theorem is the statement than any finite abelian group is isomorphic to

$$\mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k} \tag{16.25}$$

where

$$n_1 | n_2 \quad \& \quad n_2 | n_3 \quad \& \quad \cdots \quad \& \quad n_{k-1} | n_k \tag{16.26}$$

Write the n_i in terms of the prime powers in (16.17).

Exercise *p-groups*

a.) Show that $\mathbb{Z}_4$ is not isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

b.) Show more generally that if p is prime $\mathbb{Z}_{p^n}$ and $\mathbb{Z}_{p^{n-m}} \oplus \mathbb{Z}_{p^m}$ are not isomorphic if $0 < m < n$.

c.) How many nonisomorphic abelian groups have order p^n?

Exercise

Suppose $e_1, e_2 \in \mathbb{Z}^2$ are two linearly independent vectors (over $\mathbb{Q}$). Let $\Lambda = \langle e_1, e_2 \rangle \subset \mathbb{Z}^2$ be the sublattice generated by these vectors. Then $\mathbb{Z}^2/\Lambda$ is a finite abelian group. Compute its Kronecker decomposition in terms of the coordinates of e_1, e_2.

16.3 Finitely Generated Abelian Groups

It is hopeless to classify all infinite abelian groups, but a "good" criterion that leads to an interesting classification is that of *finitely generated* abelian groups.

Any abelian group has a canonically defined subgroup known as the *torsion subgroup*, and denoted Tors(G). This is the subgoup of elements of *finite order*:

$$\text{Tors}(G) := \{g \in G | \exists n \in \mathbb{Z} \qquad ng = 0\} \tag{16.27}$$

where we are writing the group G additively, so $ng = g + \cdots + g$.

One can show that any *finitely generated abelian group* fits in an exact sequence

$$0 \to \mathrm{Tors}(G) \to A \to \mathbb{Z}^r \to 0 \tag{16.28}$$

where $\mathrm{Tors}(G)$ is a *finite abelian group.*

For a proof, see, e.g., S. Lang, *Algebra* .

Moreover (16.28) is a split extension, that is, it is isomorphic to

$$\mathbb{Z}^r \oplus \mathrm{Tors}(G) \tag{16.29}$$

The integer r, called the *rank of the group*, and the finite abelian group $\mathrm{Tors}(G)$ are invariants of the finitely generated abelian group. Since we have a general picture of the finite abelian groups we have now got a general picture of the finitely generated abelian groups.

Remark:

Remarks

1. The groups $\mathbb{C}, \mathbb{R}, \mathbb{Q}$ under addition are abelian but not finitely generated. This is obvious for $\mathbb{C}$ and $\mathbb{R}$ since these are uncountable sets. To see that $\mathbb{Q}$ is not finitely generated consider any finite set of fractions $\{\frac{p_1}{q_1}, \ldots, \frac{p_s}{q_s}\}$. This set will will only generate rational numbers which, when written in lowest terms, have denominator at most $q_1 q_2 \cdots q_s$.

2. Note that a torsion abelian group need not be finite in general. For example $\mathbb{Q}/\mathbb{Z}$ is entirely torsion, but is not finite.

3. A rich source of finitely generated abelian groups are the integral cohomology groups $H^n(X; \mathbb{Z})$ of smooth compact manifolds.

4. We must stress that the presentation (16.29) of a finitely generated abelian group is not canonical! There are many distinct splittings of (16.28). They are in 1-1 correspondence with the group homomorphisms $\mathrm{Hom}(\mathbb{Z}^r, \mathrm{Tors}(G))$. For a simple example consider $\mathbb{Z}^d/\Lambda$ where Λ is a general sublattice of rank less than d.

5. In a nonabelian group the product of two finite-order elements can very well have infinite order. Examples include free products of cyclic groups and simple rotations by $2\pi/n$ around different axes in $SO(3)$. So, there is no straightforward generalization of $\mathrm{Tors}(G)$ to the case of nonabelian groups.

Exercise

Consider the finitely generated Abelian group [227]

$$L = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4 | \sum_i x_i = 0 \mathrm{mod} 2\} \tag{16.30}$$

and consider the subgroup S generated by

$$\begin{aligned} v_1 &= (1,1,1,1) \\ v_2 &= (1,1,-1,-1) \end{aligned} \tag{16.31}$$

a.) What is the torsion group of L/S ?

b.) Find a splitting of the sequence (16.28) and compare with the one found by other students in the course. Are they the same?

Exercise

Given a set of finite generators of an Abelian group A try to find an algorithm for a splitting of the sequence (16.28).

16.4 The Classification Of Finite Simple Groups

Kronecker's structure theorem is a very satisfying, beautiful and elegant answer to a classification question. The generalization to nonabelian groups is very hard. It turns out that a "good" criterion is that a finite group be a *simple* group. This idea arose from the Galois demonstration of (non)solvability of polynomial equations by radicals.

A key concept in abstract group theory is provided by the notion of a *composition series*. This is a sequence of subgroups

$$1 = G_{s+1} \triangleleft G_s \triangleleft \cdots \triangleleft G_2 \triangleleft G_1 = G \tag{16.32}$$

which have the property that G_{i+1} is a maximal normal subgroup of G_i. (Note: G_{i+1} need not be normal in G. Moreover, there might be more than one maximal normal subgroup in G_i.) As a simple example we shall see that we have

♣should give an example of this.... ♣

$$1 = G_4 \triangleleft G_3 = \mathbb{Z}_2 \times \mathbb{Z}_2 \triangleleft G_2 = A_4 \triangleleft G_1 = S_4 \tag{16.33}$$

but

$$G_3 = 1 \triangleleft G_2 = A_n \triangleleft G_1 = S_n \qquad\qquad n \geq 5 \tag{16.34}$$

Not every group admits a composition series. For example $G = \mathbb{Z}$ does not admit a composition series. (Explain why!) However, it can be shown that every finite group admits a composition series.

♣Give a reference. ♣

[227] It is the root lattice of $\mathfrak{so}(8)$.

It follows that in a composition series the quotient groups G_i/G_{i+1} are *simple groups*: By definition, a simple group is one whose only normal subgroups are 1 and itself. From what we have learned above, that means that a simple group has no nontrivial homomorphic images. It also implies that the center is trivial or the whole group.

Let us prove that the G_i/G_{i+1} are simple: In general, if $N \triangleleft G$ is a normal subgroup then there is a 1-1 correspondence:

Subgroups H between N and G: $N \subset H \subset G \Leftrightarrow$ Subgroups of G/N

Moreover, under this correspondence:

Normal subgroups of $G/N \Leftrightarrow$ Normal subgroups $N \subset H \triangleleft G$. If $H/G_{i+1} \subset G_i/G_{i+1}$ were normal and $\neq 1$ then $G_{i+1} \subset H \subset G_i$ would be normal and and properly contain G_{i+1}, contradicting maximality of G_{i+1}. ♠

♣Make this an exercise in an earlier section. ♣

A composition series is a nonabelian generalization of the Kronecker decomposition. It is not unique (see exercise below) but the the following theorem, known as the Jordan-Hölder theorem states that there are some invariant aspects of the decomposition:

Theorem: Suppose there are two different composition series for G:

$$1 = G_{s+1} \triangleleft G_s \triangleleft \cdots \triangleleft G_2 \triangleleft G_1 = G \tag{16.35}$$

$$1 = G'_{s'+1} \triangleleft G'_s \triangleleft \cdots \triangleleft G'_2 \triangleleft G'_1 = G \tag{16.36}$$

Then $s = s'$ and there is a permutation $i \to i'$ so that $G_i/G_{i+1} \cong G'_{i'}/G'_{i'+1}$. That is: The length and the unordered set of quotients are both invariants of the group and do not depend on the particular composition series.

For a proof see Jacobsen, Section 4.6.

The classification of all finite groups is reduced to solving the extension problem in general, and then classifying finite simple groups. The idea is that if we know $G_i/G_{i+1} = S_i$ is a finite simple group then we construct G_i from G_{i+1} and the extension:

$$1 \to G_{i+1} \to G_i \to S_i \to 1 \tag{16.37}$$

We have discussed the extension problem thoroughly above. One of the great achievements of 20th century mathematics is the complete classification of finite simple groups, so let us look at the finite simple groups:

First consider the abelian ones. These cannot have nontrivial subgroups and hence must be of the form $\mathbb{Z}/p\mathbb{Z}$ where p is prime.

So, now we search for the nonabelian finite simple groups. A natural source of nonabelian groups are the symmetric groups S_n. Of course, these are not simple because $A_n \subset S_n$ are normal subgroups. Could the A_n be simple? The first nonabelian example is A_4 and it is not a simple group! Indeed, consider the cycle structures $(2)^2$. There are three nontrivial elements: (12)(34), (13)(24), and (14)(23), they are all involutions, and

$$((12)(34)) \cdot ((13)(24)) = ((13)(24)) \cdot ((12)(34)) = (14)(23) \tag{16.38}$$

and therefore together with the identity they form a subgroup $K \subset A_4$ isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. Since cycle-structure is preserved under conjugation, this is obviously a normal subgroup of A_4!. After this unpromising beginning you might be surprised to learn:

Theorem A_n is a simple group for $n \geq 5$.

Sketch of the proof:

We first observe that A_n is generated by cycles of length three: (abc). The reason is that $(abc) = (ab)(bc)$, so any word in an even number of distinct transpositions can be rearranged into a word made from a product of cycles of length three. Therefore, the strategy is to show that any normal subgroup $K \subset A_n$ which is larger than 1 must contain at least one three-cycle (abc). WLOG let us say it is (123). Now we claim that the entire conjugacy class of three-cycles must be in K. We consider a permutation ϕ which takes

$$\phi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \cdots \\ i & j & k & l & m & \cdots \end{pmatrix} \tag{16.39}$$

Then $\phi(123)\phi^{-1} = (ijk)$. If $\phi \in A_n$ we are done, since K is normal in A_n so then $(ijk) \in K$. If ϕ is an odd permutation then $\tilde{\phi} = \phi(45)$ is even and $\tilde{\phi}(123)\tilde{\phi}^{-1} = (ijk)$.

Thus, we need only show that some 3-cycle is in K. For $n = 5$ this can be done rather explicitly. See the exercise below. Once we have established that A_5 is simple we can proceed by induction as follows.

We first establish a lemma: If $n \geq 5$ then for any $\sigma \in A_n$, $\sigma \neq 1$ there is a conjugate element (in A_n) σ' with $\sigma' \neq \sigma$ such that there is an $i \in \{1, \ldots, n\}$ so that $\sigma(i) = \sigma'(i)$.

To prove the lemma choose any $\sigma \neq 1$ and for σ choose a cycle of maximal length, say r so that $\sigma = (12 \ldots r)\pi$ with π fixing $\{1, \ldots, r\}$. If $r \geq 3$ then consider the conjugate:

$$\sigma' = (345)\sigma(345)^{-1} = (345)(123\cdots)\pi(354) \tag{16.40}$$

We see that $\sigma(1) = \sigma'(1) = 2$, while $\sigma(2) = 3$ and $\sigma'(2) = 4$. We leave the case $r = 2$ to the reader.

Now we proceed by induction: Suppose A_j is simple for $5 \leq j \leq n$. Consider A_{n+1} and let $N \triangleleft A_{n+1}$. Then choose $\sigma \in N$ and using the lemma consider $\sigma' \in A_{n+1}$ with $\sigma' \neq \sigma$ and $\sigma'(i) = \sigma(i)$ for some i. Let $H_i \subset A_{n+1}$ be the subgroup of permutations fixing i. It is isomorphic to A_n. Now, $\sigma' \in N$ since it is a conjugate of $\sigma \in N$ and N is assumed to be normal. Therefore $\sigma^{-1}\sigma' \in N$, and $\sigma^{-1}\sigma' \neq 1$. Therefore $N \cap H_i \neq 1$. But $N \cap H_i$ must be normal in H_i. Since $H_i \cong A_n$ it follows that $N \cap H_i = H_i$. But H_i contains 3-cycles. Therefore N contains 3-cycles and hence $N \cong A_{n+1}$. ♠

Remark: For several other proofs of the same theorem and other interesting related facts see

http://www.math.uconn.edu/kconrad/blurbs/grouptheory/Ansimple.pdf.

Digressive Remark: A group is called *solvable* if the G_i/G_{i+1} are abelian (and hence $\mathbb{Z}/p\mathbb{Z}$ for some prime p). The term has its origin in Galois theory, which in turn was the original genesis of group theory. Briefly, in Galois theory one considers a polynomial $P(x)$ with coefficients drawn from a field F. (e.g. consider $F = \mathbb{Q}$ or $\mathbb{R}$). Then the roots of the polynomial θ_i can be adjoined to F to produce a bigger field $K = F[\theta_i]$. The *Galois group of the polynomial* $Gal(P)$ is the group of automorphisms of K fixing F. Galois theory

sets up a beautiful 1-1 correspondence between subgroups $H \subset Gal(P)$ and subfields $F \subset K_H \subset K$. The intuitive notion of solving a polynomial by radicals corresponds to finding a series of subfields $F \subset F_1 \subset F_2 \subset \cdots \subset K$ so that F_{i+1} is obtained from F_i by adjoining the solutions of an equation $y^d = z$. Under the Galois correspondence this translates into a composition series where $Gal(P)$ is a solvable group - hence the name. If we take $F = \mathbb{C}[a_0, \ldots, a_{n-1}]$ for an nth order polynomial

$$P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \tag{16.41}$$

then the roots θ_i are such that a_j are the jth elementary symmetric polynomials in the θ_i (See Chapter 2 below). The Galois group is then S_n. For $n \geq 5$ the only nontrivial normal subgroup of S_n is A_n, and this group is simple, hence certainly not solvable. That is why there is no solution of an nth order polynomial equation in radicals for $n \geq 5$.

Returning to our main theme, we ask: What other finite simple groups are there? The full list is known. The list is absolutely fascinating: [228]

1. $\mathbb{Z}/p\mathbb{Z}$ for p prime.
2. The subgroup $A_n \subset S_n$ for $n \geq 5$.
3. "Simple Lie groups over finite fields."
4. 26 "sporadic oddballs"

We won't explain example 3 in great detail, but it consists of a few more infinite sequences of groups, like 1,2 above. To get a flavor of what is involved note the following: The additive group $\mathbb{Z}/p\mathbb{Z}$ where p is prime has more structure: One can multiply elements, and if an element is nonzero then it has a multiplicative inverse, in other words, it is a *finite field*. One can therefore consider the group of invertible matrices over this field $GL(n,p)$, and its subgroup $SL(n,p)$ of matrices of unit determinant. Since $\mathbb{Z}/p\mathbb{Z}$ has a finite number of elements it is a finite group. This group is not simple, because it has a nontrivial center, in general. For example, if n is even then the group $\{\pm 1\}$ is a normal subgroup isomorphic to $\mathbb{Z}_2$. If we divide by the center the we get a group $PSL(n,p)$ which, it turns out, is indeed a simple group. This construction can be generalized in a few directions. First, there is a natural generalization of $\mathbb{Z}/p\mathbb{Z}$ to finite fields $\mathbb{F}_q$ of order a prime power $q = p^k$. Then we can similarly define $PSL(n,q)$ and it turns out these are simple groups except for some low values of n, q. Just as the Lie groups $SL(n,\mathbb{C})$ have counterparts $O(n), Sp(n)$ etc. one can generalize this construction to groups of type B, C, D, E. This construction can be used to obtain the third class of finite simple groups.

♣Double check. Does this figure leave out a subgroup relation? ♣

[228] See the *Atlas of Finite Simple Groups*, by Conway and Norton

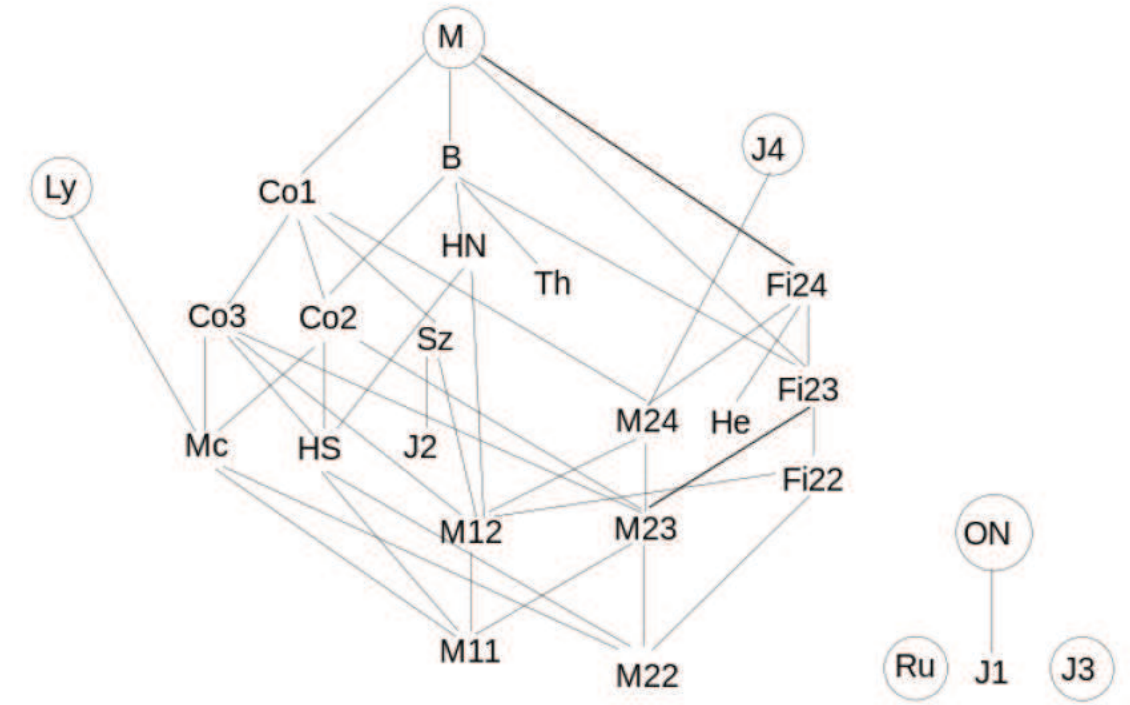


Figure 43: A table of the sporadic groups including subgroup relations. Source: Wikipedia.

It turns out that there are exactly 26 oddballs, known as the "sporadic groups." Some relationships between them are illustrated in Figure 43. The sporadic groups first showed up in the 19th century via the Mathieu groups

$$M_{11}, M_{12}, M_{22}, M_{23}, M_{24}. \tag{16.42}$$

M_n is a subgroup of the symmetric group S_n. M_{11}, which has order $|M_{11}| = 7920$ was discovered in 1861. We met M_{12} when discussing card-shuffling. The last group M_{24}, with order $\sim 10^9$ was discovered in 1873. All these groups may be understood as automorphisms of certain combinatorial objects called "Steiner systems."

It was a great surprise when Janko constructed a new sporadic group J_1 of order $175,560$ in 1965, roughly 100 years after the discovery of the Mathieu groups. The list of sporadic groups is now thought to be complete. The largest sporadic group is called the Monster group and its order is:

$$\begin{aligned} |Monster| &= 2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71 \\ &= 808017424794512875886459904961710757005754368000000000 \\ &\cong 8.08 \times 10^{53} \end{aligned} \tag{16.43}$$

but it has only 194 conjugacy classes! (Thus, by the class equation, it is "very" nonabelian. The center is trivial and $Z(g)$ tends to be a small order group.)

The history and status of the classification of finite simple groups is somewhat curious:[229]

1. The problem was first proposed by Hölder in 1892. Intense work on the classification begins during the 20th century.

2. Feit and Thompson show (1963) that any finite group of odd order is solvable. In particular, it cannot be a simple group.

3. Janko discovers (1965) the first new sporadic group in almost a century.

4. Progress is then rapid and in 1972 Daniel Gorenstein (of Rutgers University) announces a detailed outline of a program to classify finite simple groups.

5. The largest sporadic group, the Monster, was first shown to exist in 1980 by Fischer and Griess. It was explicitly constructed (as opposed to just being shown to exist) by Griess in 1982.

6. The proof is completed in 2004. It uses papers from hundreds of mathematicians between 1955 and 2004, and largely follows Gorenstein's program. The proof entails tens of thousands of pages. Errors and gaps have been found, but so far they are just "local."

Compared to the simple and elegant proof of the classification of simple Lie algebras (to be covered in Chapter **** below) the proof is obviously terribly unwieldy.

It is conceivable that physics might actually shed some light on this problem. The simple groups are probably best understood as automorphism groups of some mathematical, perhaps even geometrical object. For example, the first nonabelian simple group, A_5 is the group of symmetries of the icosahedron, as we will discuss in detail below. A construction of the monster along these lines was indeed provided by Frenkel, Lepowsky, Meurman, (at Rutgers) using vertex operator algebras, which are important in the description of perturbative string theory. More recently the mystery has deepened with interesting experimental discoveries linking the largest Mathieu group M_{24} to nonlinear sigma models with K3 target spaces. For more discussion about the possible role of physics in this subject see:

1. Articles by Griess and Frenkel et. al. in *Vertex Operators in Mathematics and Physics*, J. Lepowsky, S. Mandelstam, and I.M. Singer, eds.

2. J. Harvey, "Twisting the Heterotic String," in *Unified String Theories*, Green and Gross eds.

[229] Our source here is the Wikipedia article on the classification of finite simple groups. See also: Solomon, Ronald, "A brief history of the classification of the Finite simple groups," American Mathematical Society. Bulletin. New Series, 38 (3): 315-352 (2001).

3. L.J. Dixon, P.H. Ginsparg, and J.A. Harvey, "Beauty And The Beast: Superconformal Symmetry In A Monster Module," Commun.Math.Phys. 119 (1988) 221-241

4. M.C.N. Cheng, J.F.R. Duncan, and J.A. Harvey, "Umbral Moonshine," e-Print: arXiv:1204.2779 [math.RT]

Exercise *Completing the proof that A_5 is simple*

Show that any nontrivial normal subgroup of A_5 must contain a 3-cycle as follows:

a.) If $N \triangleleft A_5$ is a normal subgroup containing no 3-cycles then the elements must have cycle type $(ab)(cd)$ or $(abcde)$.

b.) Compute the group commutators (a, b, c, d, e are all distinct):

$$[(abe), (ab)(cd)] = (aeb) \tag{16.44}$$

$$[(abc), (abcde)] = (abd) \tag{16.45}$$

c.) Use these facts to conclude that N must contain a 3-cycle.

Legend has it that Galois discovered this theorem on the night before his fatal duel.

Exercise *Conjugacy classes in A_n*

Note that conjugacy classes in A_n are different from conjugacy classes in S_n. For example, (123) and (132) are not conjugate in A_3.

Describe the conjugacy classes in A_n.

Exercise *Jordan-Hölder decomposition*

Work out JH decompositions for the order 8 quaternion group $\widetilde{D}_2$ and observe that there are several maximal normal subgroups.

Exercise *The simplest of the Chevalley groups*

a.) Verify that $SL(2, \mathbb{Z}/p\mathbb{Z})$ is a group.

b.) Show that the order of $SL(2, \mathbb{Z}/p\mathbb{Z})$ is $p(p^2 - 1)$. [230]

[230] Break up the cases into $d = 0$ and $d \neq 0$. When $d = 0$ you can solve $ad - bc = 1$ for a. When $d = 0$ you can have arbitrary a but you must have $bc = -1$.

c.) Note that the scalar multiples of the 2×2 identity matrix form a normal subgroup of $SL(2, \mathbb{Z}/p\mathbb{Z})$. Show that the number of such matrices is the number of solutions of $x^2 = 1 \mathrm{mod} p$. Dividing by this normal subgroup produces the group $PSL(2, \mathbb{Z}/p\mathbb{Z})$. Jordan proved that these are simple groups for $p \neq 2, 3$.

It turns out that $PSL(2, \mathbb{Z}_5) \cong A_5$. (Check that the orders match.) Therefore the next simple group in the series is $PSL(2, \mathbb{Z}_7)$. It has many magical properties.

d.) Show that $PSL(2, \mathbb{Z}_7)$ has order 168.

---

17. Categories: Groups and Groupoids

A rather abstract notion, which nevertheless has found recent application in string theory and conformal field theory is the language of categories. Many physicists object to the high level of abstraction entailed in the category language. Some mathematicians even refer to the subject as "abstract nonsense." (Others take it very seriously.) However, it seems to be of increasing utility in the further formal development of string theory and supersymmetric gauge theory. It is also essential for reading any of the literature on topological field theory.

We briefly illustrate some of that language here. Our main point here is to introduce a different viewpoint on what groups are that leads to a significant generalization: groupoids. Moreover, this point of view also provides some very interesting insight into the meaning of group cohomology. Related constructions have been popular in condensed matter physics and topological field theory.

Definition A *category* $\mathcal{C}$ consists of

a.) A set $Ob(\mathcal{C})$ of "objects"

b.) A collection $Mor(\mathcal{C})$ of sets $\hom(X, Y)$, defined for any two objects $X, Y \in Ob(\mathcal{C})$. The elements of $\hom(X, Y)$ are called the "morphisms from X to Y." They are often denoted as arrows:

$$X \stackrel{\phi}{\to} Y \tag{17.1}$$

c.) A composition law:

$$\hom(X, Y) \times \hom(Y, Z) \to \hom(X, Z) \tag{17.2}$$

$$(\psi_1, \psi_2) \mapsto \psi_2 \circ \psi_1 \tag{17.3}$$

Such that

1. A morphism ϕ uniquely determines its source X and target Y. That is, $\hom(X, Y)$ are disjoint for distinct ordered pairs (X, Y).

2. $\forall X \in Ob(\mathcal{C})$ there is a distinguished morphism, denoted $1_X \in \hom(X, X)$ or $\mathrm{Id}_X \in \hom(X, X)$, which satisfies:

$$1_X \circ \phi = \phi \qquad \psi \circ 1_X = \psi \tag{17.4}$$

for all morphisms $\phi \in \hom(Y, X)$ and $\psi \in \hom(X, Y)$ for all $Y \in Ob(\mathcal{C})$. [231]

[231] As an exercise, show that these conditions uniquely determine the morphism 1_X.

3. Composition of morphisms is associative:

$$(\psi_1 \circ \psi_2) \circ \psi_3 = \psi_1 \circ (\psi_2 \circ \psi_3) \tag{17.5}$$

An alternative definition one sometimes finds is that a category is defined by two sets $\mathcal{C}_0$ (the objects) and $\mathcal{C}_1$ (the morphisms) with two maps $p_0 : \mathcal{C}_1 \to \mathcal{C}_0$ and $p_1 : \mathcal{C}_1 \to \mathcal{C}_0$. The map $p_0(f) = x_1 \in \mathcal{C}_0$ is the *range* map and $p_1(f) = x_0 \in \mathcal{C}_0$ is the *domain* map. In this alternative definition a category is then defined by a composition law on the set of *composable morphisms*

$$\mathcal{C}_2 = \{(f, g) \in \mathcal{C}_1 \times \mathcal{C}_1 | p_0(f) = p_1(g)\} \tag{17.6}$$

which is sometimes denoted $\mathcal{C}_{1 p_1} \times_{p_0} \mathcal{C}_1$ and called the *fiber product*. The composition law takes $\mathcal{C}_2 \to \mathcal{C}_1$ and may be pictured as the composition of arrows. If $f : x_0 \to x_1$ and $g : x_1 \to x_2$ then the composed arrow will be denoted $g \circ f : x_0 \to x_2$. The composition law satisfies the axioms

1. For all $x \in X_0$ there is an identity morphism in X_1, denoted 1_x, or Id_x, such that $1_x f = f$ and $g 1_x = g$ for all suitably composable morphisms f, g.

2. The composition law is associative. If f, g, h are 3-composable morphisms then $(hg)f = h(gf)$.

Remarks:

1. When defining composition of arrows one needs to make an important notational decision. If $f : x_0 \to x_1$ and $g : x_1 \to x_2$ then the composed arrow is an arrow $x_0 \to x_2$. We will write $g \circ f$ when we want to think of f, g as functions and fg when we think of them as arrows.

♣Is this dual notation really a good idea?? ♣

2. It is possible to endow the data X_0, X_1 and p_0, p_1 with additional structures, such as topologies, and demand that p_0, p_1 have continuity or other properties.

3. A morphism $\phi \in \hom(X, Y)$ is said to be *invertible* if there is a morphism $\psi \in \hom(Y, X)$ such that $\psi \circ \phi = 1_X$ and $\phi \circ \psi = 1_Y$. If X and Y are objects with an invertible morphism between them then they are called *isomorphic objects*. One key reason to use the language of categories is that objects can have nontrivial automorphisms. That is, $\hom(X, X)$ can have invertible elements other than just 1_X in it. When this is true then it is tricky to speak of "equality" of objects, and the language of categories becomes very helpful. As a concrete example you might ponder the following question: "Are all real vector spaces of dimension n *the same*?"

Here are some simple examples of categories:

1. **SET**: The category of sets and maps of sets. [232]

[232] We take an appropriate collection of sets and maps to avoid the annoying paradoxes of set theory.

2. **TOP**: The category of topological spaces and continuous maps.

3. **TOPH**: The category of topological spaces and homotopy classes of continuous maps.

4. **MANIFOLD**: The category of manifolds and suitable maps. We could take topological manifolds and continuous maps of manifolds. Or we could take smooth manifolds and smooth maps as morphisms. The two choices lead to two (very different!) categories.

5. **BORD**(n): The bordism category of n-dimensional manifolds. Roughly speaking, the objects are n-dimensional manifolds without boundary and the morphisms are bordisms. A bordism Y from an n-manifold M_1 to and n-manifold M_2 is an $(n+1)$-dimensional manifold with a decomposition of its boundary $\partial Y = (\partial Y)_{in} \amalg (\partial Y)_{out}$ together with diffeomorphisms $\theta_1 : (\partial Y)_{in} \to M_1$ and $\theta_2 : (\partial Y)_{out} \to M_2$.

6. **GROUP**: the category of groups and homomorphisms of groups. Note that here if we took our morphisms to be isomorphisms instead of homomorphisms then we would get a very different category. All the pairs of objects in the category with nontrivial morphism spaces between them would be pairs of isomorphic groups.

7. **AB**: The (sub) category of abelian groups.

8. Fix a group G and let **G-SET** be the category of G-sets, that is, sets X with a G-action. For simplicity let us just write the G-action $\Phi(g,x)$ as $g \cdot x$ for x a point in a G-set X. We take the morphisms $f : X_1 \to X_2$ to satisfy satisfy $f(g \cdot x_1) = g \cdot f(x_1)$.

9. **VECT**$_\kappa$: The category of finite-dimensional vector spaces over a field κ with morphisms the linear transformations.

One use of categories is that they provide a language for describing precisely notions of "similar structures" in different mathematical contexts. When discussed in this way it is important to introduce the notion of "functors" and "natural transformations" to speak of interesting relationships between categories.

In order to state a relation between categories one needs a "map of categories." This is what is known as a functor:

Definition A *functor* between two categories $\mathcal{C}_1$ and $\mathcal{C}_2$ consists of a pair of maps $F_{\rm obj} : Obj(\mathcal{C}_1) \to Obj(\mathcal{C}_2)$ and $F_{\rm mor} : Mor(\mathcal{C}_1) \to Mor(\mathcal{C}_2)$ so that if

$$x \xrightarrow{f} y \ \in \hom(x,y) \tag{17.7}$$

then

$$F_{\rm obj}(x) \xrightarrow{F_{\rm mor}(f)} F_{\rm obj}(y) \ \in \hom(F_{\rm obj}(x), F_{\rm obj}(y)) \tag{17.8}$$

and moreover we require that $F_{\rm mor}$ should be compatible with composition of morphisms: There are two ways this can happen. If f_1, f_2 are composable morphisms then we say F is a *covariant functor* if

$$F_{\rm mor}(f_1 \circ f_2) = F_{\rm mor}(f_1) \circ F_{\rm mor}(f_2) \tag{17.9}$$

and we say that F is a *contravariant functor* if

$$F_{\rm mor}(f_1 \circ f_2) = F_{\rm mor}(f_2) \circ F_{\rm mor}(f_1) \tag{17.10}$$

In both cases we also require [233]

$$F_{\rm mor}({\rm Id}_X) = {\rm Id}_{F(X)} \tag{17.11}$$

We usually drop the subscript on F since it is clear what is meant from context.

Exercise

Using the alternative definition of a category in terms of data $p_{0,1} : X_1 \to X_0$ define the notion of a functor writing out the relevant commutative diagrams.

Exercise *Opposite Category*

If $\mathcal{C}$ is a category then the *opposite category* $\mathcal{C}^{\rm opp}$ is defined by just reversing all arrows. More formally: The set of objects is the same and

$$\hom_{\mathcal{C}^{\rm opp}}(X, Y) := \hom_{\mathcal{C}}(Y, X) \tag{17.12}$$

so for every morphism $f \in \hom_{\mathcal{C}}(Y, X)$ we associate $f^{\rm opp} \in \hom_{\mathcal{C}^{\rm opp}}(X, Y)$ such that

$$f_1 \circ_{\mathcal{C}^{\rm opp}} f_2 = (f_2 \circ_{\mathcal{C}} f_1)^{\rm opp} \tag{17.13}$$

a.) Show that if $F : \mathcal{C} \to \mathcal{D}$ is a contravariant functor then one can define in a natural way a covariant functor $F : \mathcal{C}^{\rm opp} \to \mathcal{D}$.

b.) Show that if $F : \mathcal{C} \to \mathcal{D}$ is a covariant functor then we can naturally define another covariant functor $F^{\rm opp} : \mathcal{C}^{\rm opp} \to \mathcal{D}^{\rm opp}$

Example 1: Every category has a canonical functor to itself, called the identity functor $Id_{\mathcal{C}}$.

Example 2: There is an obvious functor, the *forgetful functor* from **GROUP** to **SET**. This idea extends to many other situations where we "forget" some mathematical structure and map to a category of more primitive objects.

[233] Although we do have $F_{\rm mor}({\rm Id}_X) \circ F_{\rm mor}(f) = F_{\rm mor}(f)$ for all $f \in \hom(Y, X)$ and $F_{\rm mor}(f) \circ F_{\rm mor}({\rm Id}_X) = F_{\rm mor}(f)$ for all $f \in \hom(X, Y)$ this is not the same as the statement that $F_{\rm mor}({\rm Id}_X) \circ \phi = \phi$ for all $\phi \in \hom(F(Y), F(X))$, so we need to impose this extra axiom.

Example 3: Since **AB** is a subcategory of **GROUP** there is an obvious functor $\mathcal{F}$: **AB** $\to$ **GROUP**.

Example 4: In an exercise below you are asked to show that the abelianization of a group defines a functor $\mathcal{G}$: **GROUP** $\to$ **AB**.

Example 5: Fix a group G. Then in the notes above we have on several occasions used the functor

$$F_G : \mathbf{SET} \to \mathbf{GROUP} \tag{17.14}$$

by observing that if X is a set, then $F_G(X) = Maps[X \to G]$ is a group. Check this is a contravariant functor: If $f : X_1 \to X_2$ is a map of sets then

$$F_G(X_1) \xleftarrow{F_G(f)} F_G(X_2) \tag{17.15}$$

The map $F_G(f)$ is usually denoted f^* and is known as the *pull-back*. To be quite explicit: If Ψ is a map of $X_2 \to G$ then $f^*(\Psi) := \Psi \circ f$ is a map $X_1 \to G$.

This functor is used in the construction of certain *nonlinear sigma models* which are quantum field theories where the target space is a group G. The viewpoint that we are studying the representation theory of an infinite-dimensional group of maps to G has been extremely successful in a particular case of the *Wess-Zumino-Witten* model, a certain two dimensional quantum field theory that enjoys conformal invariance (and more).

Example 6: Now let us return to the category **G-SET**. Now fix any set Y. Then in the notes above we have on several occasions used the functor

$$F_{G,Y} : \mathbf{G\text{-}SET} \to \mathbf{G\text{-}SET} \tag{17.16}$$

by observing that if X is a G-set, then $F_Y(X) = Maps[X \to Y]$ is also a G-set. To check this is a contravariant functor we write:

$$\begin{aligned} [g \cdot (f^*\Psi)](x_1) &= (f^*\Psi)(g^{-1} \cdot x_1) \\ &= \Psi(f(g^{-1} \cdot x_1)) \\ &= \Psi(g^{-1} \cdot (f(x_1))) \\ &= (g \cdot \Psi)(f(x_1)) \\ &= (f^*(g \cdot \Psi))(x_1) \end{aligned} \tag{17.17}$$

and hence $\Psi \to g \cdot \Psi$ is a morphism of G-sets.

This functor is ubiquitous in quantum field theory: If a spacetime enjoys some symmetry (for example rotational or Poincaré symmetry) then the same group will act on the space of fields defined on that spacetime.

Example 7: Fix a nonnegative integer n and a group G. Then the group cohomology we discussed above (take the trivial twisting $\omega_g = \mathrm{Id}_A$ for all g) defines a covariant functor

$$H^n(G, \bullet) : \mathbf{AB} \to \mathbf{AB} \tag{17.18}$$

To check this is really a functor we need to observe the following: If $\varphi : A_1 \to A_2$ is a homomorphism of Abelian groups then there is an induced homomorphim, usually denoted

$$\varphi_* : H^n(G, A_1) \to H^n(G, A_2) \tag{17.19}$$

You have to check that $\mathrm{Id}_* = \mathrm{Id}$ and

$$(\varphi_1 \circ \varphi_2)_* = (\varphi_1)_* \circ (\varphi_2)_* \tag{17.20}$$

Strictly speaking we should denote φ_* by $H^n(G, \varphi)$, but this is too fastidious for the present author.

Example 8: Fix a nonnegative integer n and any group A. Then the group cohomology we discussed above (take the trivial twisting $\omega_g = \mathrm{Id}_A$ for all g) defines a contravariant functor

$$H^n(\bullet, A) : \mathbf{GROUP} \to \mathbf{AB} \tag{17.21}$$

To check this is really a functor we need to observe the following: If $\varphi : G_1 \to G_2$ is a homomorphism of Abelian groups then there is an induced homomorphim, usually denoted φ^*

$$\varphi^* : H^n(G_2, A) \to H^n(G_1, A) \tag{17.22}$$

Example 9: *Topological Field Theory.* The very definition of topological field theory is that it is a functor from a bordism category of manifolds to the category of vector spaces and linear transformations. For much more about this one can consult a number of papers. Two online resources are

http://www.physics.rutgers.edu/∼gmoore/695Fall2015/TopologicalFieldTheory.pdf
https://www.ma.utexas.edu/users/dafr/bordism.pdf

Note that in example 2 there is no obvious functor going the reverse direction. When there are functors both ways between two categories we might ask whether they might be, in some sense, "the same." But saying precisely what is meant by "the same" requires some care.

Definition If $\mathcal{C}_1$ and $\mathcal{C}_2$ are categories and $F_1 : \mathcal{C}_1 \to \mathcal{C}_2$ and $F_2 : \mathcal{C}_1 \to \mathcal{C}_2$ are two functors then a *natural transformation* $\tau : F_1 \to F_2$ is a rule which, for every $X \in Obj(\mathcal{C}_1)$ assigns an arrow $\tau_X : F_1(X) \to F_2(X)$ so that, for all $X, Y \in Obj(\mathcal{C}_1)$ and all $f \in \hom(X, Y)$,

$$\tau_Y \circ F_1(f) = F_2(f) \circ \tau_X \tag{17.23}$$

Or, in terms of diagrams.

$$\begin{array}{ccc} F_1(X) & \xrightarrow{F_1(f)} & F_1(Y) \\ \Big\downarrow{\scriptstyle \tau_X} & & \Big\downarrow{\scriptstyle \tau_Y} \\ F_2(X) & \xrightarrow{F_2(f)} & F_2(Y) \end{array} \tag{17.24}$$

Example 1: *The evaluation map.* Here is another tautological construction which nevertheless can be useful. Let S be any set and define a functor

$$F_S : \mathbf{SET} \to \mathbf{SET} \tag{17.25}$$

by saying that on objects we have

$$F_S(X) := Map[S \to X] \times S \tag{17.26}$$

and if $\varphi : X_1 \to X_2$ is a map of sets then

$$F_S(\varphi) : Map[S \to X_1] \times S \to Map[S \to X_2] \times S \tag{17.27}$$

is defined by $F_S(\varphi) : (f, s) \mapsto (\varphi \circ f, s)$. Then we claim there is a natural transformation to the identity functor. For every set X we have

$$\tau_X : F_S(X) = Map[S \to X] \times S \to \mathrm{Id}(X) = X \tag{17.28}$$

It is defined by $\tau_X(f, s) := f(s)$. This is known as the "evaluation map." Then we need to check

$$\begin{array}{ccc} F_S(X) & \xrightarrow{\tau_X} & X \\ \downarrow{\scriptstyle F_S(\varphi)} & & \downarrow{\scriptstyle \varphi} \\ F_S(Y) & \xrightarrow{\tau_Y} & Y \end{array} \tag{17.29}$$

commutes. If you work it out, it is just a tautology.

Example 2: *The determinant.* [234] Let **COMMRING** be the category of commutative rings with morphisms the ring morphisms. (So, $\varphi : R_1 \to R_2$ is a homomorphism of Abelian groups and moreover $\varphi(r \cdot s) = \varphi(r) \cdot \varphi(s)$.) Let us consider two functors

$$\mathbf{COMMRING} \to \mathbf{GROUP} \tag{17.30}$$

The first functor F_1 maps a ring R to the multiplicative group $U(R)$ of multiplicatively invertible elements. This is often called the group of units in R. If φ is a morphism of rings and $r \in U(R_1)$ then $\varphi(r) \in U(R_2)$ and the map $\varphi_* : U(R_1) \to U(R_2)$ defined by

$$\varphi_* : r \mapsto \varphi(r) \tag{17.31}$$

is a group homomorphism. So F_1 is a functor. The second functor F_2 maps a ring R to the matrix group $GL(n, R)$ of $n \times n$ matrices such that there exists an inverse matrix with values in R. Again, if $\varphi : R_1 \to R_2$ is a morphism then applying φ to each matrix element defines a group homomorphism $\varphi_* : GL(n, R_1) \to GL(n, R_2)$. Now consider the determinant of a matrix $g \in GL(n, R)$. The usual formula

$$\det(g) := \sum_{\sigma \in S_n} \epsilon(\sigma) g_{1,\sigma(1)} \cdot g_{2,\sigma(2)} \cdots g_{n,\sigma(n)} \tag{17.32}$$

[234] This example uses some terms from linear algebra which can be found in the "User's Manual," Chapter 2 below.

makes perfect sense for $g \in GL(n,R)$. Moreover,

$$\det(g_1 g_2) = \det(g_1)\det(g_2) \tag{17.33}$$

Now we claim that the determinant defines a natural transformation $\tau : F_1 \to F_2$. For each object $R \in Ob(\mathbf{COMMRING})$ we assign the morphism

$$\tau_R : GL(n,R) \to U(R) \tag{17.34}$$

defined by $\tau_R(g) := \det(g)$. Thanks to (17.33) this is indeed a morphism in the category **GROUP**, that is, it is a group homomorphism. Moreover, it satisfies the required commutative diagram because if $\varphi : R_1 \to R_2$ is a morphism of rings then

$$\varphi_*(\det(g)) = \det(\varphi_*(g)). \tag{17.35}$$

Example 3: *Natural transformations in cohomology theory.* Cohomology groups provide natural examples of functors, as we have stressed above. There are a number of interesting natural transformations between these different cohomology-group functors.

♣Can we explain an elementary example with group cohomology as developed so far??? ♣

Definition Two categories are said to be *equivalent* if there are functors $F : \mathcal{C}_1 \to \mathcal{C}_2$ and $G : \mathcal{C}_2 \to \mathcal{C}_1$ together with isomorphisms (via natural transformations) $FG \cong Id_{\mathcal{C}_2}$ and $GF \cong Id_{\mathcal{C}_1}$. (Note that FG and $Id_{\mathcal{C}_2}$ are both objects in the category of functors $\mathrm{FUNCT}(\mathcal{C}_2, \mathcal{C}_2)$ so it makes sense to say that they are isomorphic.)

♣Should explain example showing category of finite-dimensional vector spaces over a field is equivalent to the catetgory of nonnegative integers. ♣

Many important theorems in mathematics can be given an elegant and concise formulation by saying that two seemingly different categories are in fact equivalent. Here is a (very selective) list: [235]

Example 1: Consider the category with one object for each nonnegative integer n and the morphism space $GL(n,\kappa)$ of invertible $n \times n$ matrices over the field κ. These categories are equivalent. That is one way of saying that the only invariant of a finite-dimensional vector space is its dimension.

Example 2: The basic relation between Lie groups and Lie algebras the statement that the functor which takes a Lie group G to its tangent space at the identity, $T_1 G$ is an equivalence of the category of connected and simply-connected Lie groups with the category of finite-dimensional Lie algebras. One of the nontrivial theorems in the theory is the existence of a functor from the category of finite-dimensional Lie algebras to the category of connected and simply-connected Lie groups. Intuitively, it is given by exponentiating the elements of the Lie algebra.

Example 3: Covering space theory is about an equivalence of categories. On the one hand we have the category of coverings of a pointed space (X, x_0) and on the other hand

[235] I thank G. Segal for a nice discussion that helped prepare this list.

the category of topological spaces with an action of the group $\pi_1(X, x_0)$. Closely related to this, Galois theory can be viewed as an equivalence of categories.

Example 4: The category of unital commutative C^*-algebras is equivalent to the category of compact Hausdorff topological spaces. This is known as Gelfand's theorem.

Example 5: Similar to the previous example, an important point in algebraic geometry is that there is an equivalence of categories of commutative algebras over a field κ (with no nilpotent elements) and the category of affine algebraic varieties.

Example 6: Pontryagin duality is a nontrivial self-equivalence of the category of locally compact abelian groups (and continuous homomorphisms) with itself.

Example 7: A generalization of Pontryagin duality is Tannaka-Krein duality between the category of compact groups and a certain category of linear tensor categories. (The idea is that, given an abstract tensor category satisfying certain conditions one can construct a group, and if that tensor category is the category of representations of a compact group, one recovers that group.)

Example 8: The Riemann-Hilbert correspondence can be viewed as an equivalence of categories of flat connections (a.k.a. linear differential equations, a.k.a. D-modules) with their monodromy representations.

♣This needs a lot more explanation. ♣

In physics, the statement of "dualities" between different physical theories can sometimes be formulated precisely as an equivalence of categories. One important example of this is mirror symmetry, which asserts an equivalence of (A_∞)-) categories of the derived category of holomorphic bundles on X and the Fukaya category of Lagrangians on $X^\vee$. But more generally, nontrivial duality symmetries in string theory and field theory have a strong flavor of an equivalence of categories.

Exercise *Playing with natural transformations*

a.) Given two categories $\mathcal{C}_1, \mathcal{C}_2$ show that the natural transformations allow one to define a category FUNCT($\mathcal{C}_1, \mathcal{C}_2$) whose objects are functors from $\mathcal{C}_1$ to $\mathcal{C}_2$ and whose morphisms are natural transformations. For this reason natural transformations are often called "morphisms of functors."

b.) Write out the meaning of a natural transformation of the identity functor $Id_{\mathcal{C}}$ to itself. Show that $End(Id_{\mathcal{C}})$, the set of all natural transformations of the identity functor to itself is a monoid.

Exercise *Freyd's theorem*

A "practical" way to tell if two categories are equivalent is the following:

By definition, a *fully faithful functor* is a functor $F : \mathcal{C}_1 \to \mathcal{C}_2$ where $F_{\rm mor}$ is a bijection on all the hom-sets. That is, for all $X, Y \in Obj(\mathcal{C}_1)$ the map

$$F_{\rm mor} : \hom(X, Y) \to \hom(F_{\rm obj}(X), F_{\rm obj}(Y)) \tag{17.36}$$

is a bijection.

Show that $\mathcal{C}_1$ is equivalent to $\mathcal{C}_2$ iff there is a fully faithful functor $F : \mathcal{C}_1 \to \mathcal{C}_2$ so that any object $\alpha \in Obj(\mathcal{C}_2)$ is isomorphic to an object of the form $F(X)$ for some $X \in Obj(\mathcal{C}_1)$.

Exercise

As we noted above, there is a functor **AB** $\to$ **GROUP** just given by inclusion.

a.) Show that the abelianization map $G \to G/[G, G]$ defines a functor **GROUP** $\to$ **AB**.

b.) Show that the existence of nontrivial perfect groups, such as A_5, implies that this functor cannot be an equivalence of categories.

In addition to the very abstract view of categories we have just sketched, very concrete objects, like groups, manifolds, and orbifolds can profitably be viewed as categories.

One may always picture a category with the objects constituting points and the morphisms directed arrows between the points as shown in Figure 44.

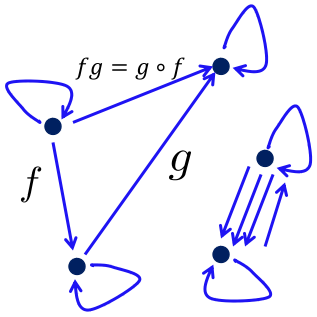


Figure 44: Pictorial illustration of a category. The objects are the black dots. The arrows are shown, and one must give a rule for composing each arrow and identifying with one of the other arrows. For example, given the arrows denoted f and g it follows that there must be an arrow of the type denoted $f \circ g$. Note that every object x has at least one arrow, the identity arrow in $Hom(x, x)$.

As an extreme example of this let us consider a category with only *one object*, but we allow the possibility that there are several morphisms. For such a category let us look carefully at the structure on morphisms $f \in Mor(\mathcal{C})$. We know that there is a binary operation, with an identity 1 which is associative.

But this is just the definition of a monoid!

If we have in addition inverses then we get a group. Hence:

Definition A *group* is a category with one object, all of whose morphisms are invertible.

To see that this is equivalent to our previous notion of a group we associate to each morphism a group element. Composition of morphisms is the group operation. The invertibility of morphisms is the existence of inverses.

We will briefly describe an important and far-reaching generalization of a group afforded by this viewpoint. Then we will show that this viewpoint leads to a nice geometrical construction making the formulae of group cohomology a little bit more intuitive.

****

CONSTRUCT EXERCISE HERE EXAMINING HOW CONCEPTS OF FUNCTORS AND NATURAL TRANSFORMATIONS TRANSLATE INTO GROUP THEORY LANGUAGE WHEN SPECIALIZED TO THE CATEGORIES CORRESPONDING TO GROUPS

****

17.1 Groupoids

Definition A *groupoid* is a category all of whose morphisms are invertible.

Note that for any object x in a groupoid, $\hom(x,x)$ is a group. It is called the *automorphism group* of the object x.

Example 1. Any equivalence relation on a set X defines a groupoid. The objects are the elements of X. The set $\mathrm{Hom}(a,b)$ has one element if $a \sim b$ and is empty otherwise. The composition law on morphisms then means that $a \sim b$ with $b \sim c$ implies $a \sim c$. Clearly, every morphism is invertible.

Example 2. Consider time evolution in quantum mechanics with a time-dependent Hamiltonian. There is no sense to time evolution $U(t)$. Rather one must speak of unitary evolution $U(t_1,t_2)$ such that $U(t_1,t_2)U(t_2,t_3) = U(t_1,t_3)$. Given a solution of the Schrodinger equation $\Psi(t)$ we may consider the state vectors $\Psi(t)$ as objects and $U(t_1,t_2)$ as morphisms. In this way a solution of the Schrodinger equation defines a groupoid. ♣Clarify this remark. ♣

Example 3. Let X be a topological space. The fundamental groupoid $\pi_{\leq 1}(X)$ is the category whose objects are points $x \in X$, and whose morphisms are homotopy classes of paths $f : x \to x'$. These compose in a natural way. Note that the automorphism group of a point $x \in X$, namely, $\hom(x,x)$ is the fundamental group of X based at x, $\pi_1(X,x)$.

Example 4. Gauge theory: Objects = connections on a principal bundle. Morphisms = gauge transformations. This is the right point of view for thinking about some more

exotic (abelian) gauge theories of higher degree forms which arise in supergravity and string theories.

Example 5. In the theory of string theory orbifolds and orientifolds spacetime must be considered to be a groupoid. Suppose we have a right action of G on a set X, so we have a map

$$\Phi : X \times G \to X \tag{17.37}$$

such that

$$\Phi(\Phi(x, g_1), g_2) = \Phi(x, g_1 g_2) \tag{17.38}$$

$$\Phi(x, 1_G) = x \tag{17.39}$$

for all $x \in X$ and $g_1, g_2 \in G$. We can just write $\Phi(x, g) := x \cdot g$ for short. We can then form the category $X//G$ with

$$\begin{aligned} Ob(X//G) &= X \\ Mor(X//G) &= X \times G \end{aligned} \tag{17.40}$$

We should think of a morphism as an arrow, labeled by g, connecting the point x to the point $x \cdot g$. The target and source maps are:

♣FIGURE NEEDED HERE! ♣

$$p_0((x, g)) := x \cdot g \qquad\qquad p_1((x, g)) := x \tag{17.41}$$

The composition of morphisms is defined by

$$(xg_1, g_2) \circ (x, g_1) := (x, g_1 g_2) \tag{17.42}$$

or, in the other notation (better suited to a right-action):

$$(x, g_1)(xg_1, g_2) := (x, g_1 g_2) \tag{17.43}$$

Note that $(x, 1_G) \in \hom(x, x)$ is the identity morphism, and the composition of morphisms makes sense because we have a group action. Also note that $pt//G$ where G has the trivial action on a point realizes the group G as a category, as sketched above.

Example 6. In the theory of string theory orbifolds and orientifolds spacetime must be considered to be a groupoid. (This is closely related to the previous example.)

Exercise

For a group G let us define a groupoid denoted $G//G$ (for reasons explained later) whose objects are group elements $Obj(G//G) = G$ and whose morphisms are arrows defined by

$$g_1 \xrightarrow{h} g_2 \tag{17.44}$$

iff $g_2 = h^{-1} g_1 h$. This is the groupoid of principal G-bundles on the circle.

Draw the groupoid corresponding to S_3.

Exercise *The Quotient Groupoid*

a.) Show that whenever G acts on a set X one can canonically define a groupoid: The objects are the points $x \in X$. The morphisms are pairs (g, x), to be thought of as arrows $x \xrightarrow{g} g \cdot x$. Thus, $X_0 = X$ and $X_1 = G \times X$.

b.) What is the automorphism group of an object $x \in X$.

This groupoid is commonly denoted as $X /\!/ G$.

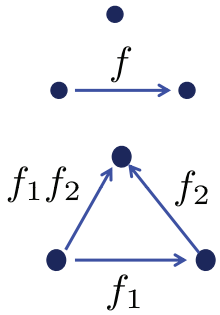


Figure 45: Elementary $0, 1, 2$ simplices in the simplicial space $|\mathcal{C}|$ of a category

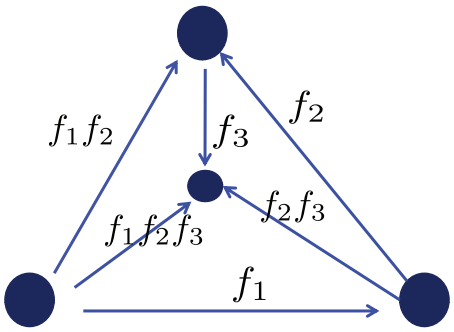


Figure 46: An elementary 3-simplex in the simplicial space $|\mathcal{C}|$ of a category

17.2 The topology behind group cohomology

Now, let us show that this point of view on the definition of a group can lead to a very nontrivial and beautiful structure associated with a group.

An interesting construction that applies to any category is its associated simplicial space $|\mathcal{C}|$.

This is a space made by gluing together simplices [236] whose simplices are:

- 0-simplices = objects
- 1-simplices = $\Delta_1(f)$ associated to each morphism $f : x_0 \to x_1 \in X_1$.
- 2-simplices: $\Delta(f_1, f_2)$ associated composable morphisms

$$(f_1, f_2) \in X_2 := \{(f_1, f_2) \in X_1 \times X_1 | p_0(f_1) = p_1(f_2)\} \tag{17.45}$$

- 3-simplices: $\Delta(f_1, f_2, f_3)$ associated to 3 composable morphisms, i.e. elements of:

$$X_3 = \{(f_1, f_2, f_3) \in X_1 \times X_1 \times X_1 | p_0(f_i) = p_1(f_{i+1}), i = 1, 2\} \tag{17.46}$$

- and so on. There are infinitely many simplices of arbitarily high dimension because we can keep composing morphisms as long as we like.

And so on. See Figures 45 and 46. The figures make clear how these simplices are glued together:

$$\partial\Delta_1(f) = x_1 - x_0 \tag{17.47}$$

$$\partial\Delta_2(f_1, f_2) = \Delta_1(f_1) - \Delta_1(f_1 f_2) + \Delta_1(f_2) \tag{17.48}$$

and for Figure 46 view this as looking down on a tetrahedron. Give the 2-simplices of Figure 45 the counterclockwise orientation and the boundary of the 3-simplex the induced orientation from the outwards normal. Then we have

$$\partial\Delta(f_1, f_2, f_3) = \Delta(f_2, f_3) - \Delta(f_1 f_2, f_3) + \Delta(f_1, f_2 f_3) - \Delta(f_1, f_2) \tag{17.49}$$

Note that on the three upper faces of Figure 46 the induced orientation is the ccw orientation for $\Delta(f_1, f_2 f_3)$ and $\Delta(f_2, f_3)$, but with the cw orientation for $\Delta(f_1 f_2, f_3)$. On the bottom fact the inward orientation is ccw and hence the outward orientation is $-\Delta(f_1, f_2)$.

Clearly, we can keep composing morphisms so the space $|\mathcal{C}|$ has simplices of arbitrarily high dimension, that is, it is an infinite-dimensional space.

Let look more closely at this space for the case of a group, regarded as a category with one object. Then in the above pictures we identify all the vertices with a single vertex.

For each group element g we have a one-simplex $\Delta_1(g)$ beginning and ending at this vertex.

For each ordered pair (g_1, g_2) we have an oriented 2-simplex $\Delta(g_1, g_2)$, etc. We simply replace $f_i \to g_i$ in the above formulae, with g_i now interpreted as elements of G:

$$\partial\Delta(g) = 0 \tag{17.50}$$

$$\partial\Delta(g_1, g_2) = \Delta_1(g_1) + \Delta_1(g_2) - \Delta_1(g_1 g_2) \tag{17.51}$$

[236] Technically, it is a *simplicial space*.

$$\partial\Delta(g_1,g_2,g_3) = \Delta(g_2,g_3) - \Delta(g_1g_2,g_3) + \Delta(g_1,g_2g_3) - \Delta(g_1,g_2) \tag{17.52}$$

See Figure 46.

Let us construct this topological space a bit more formally:

We begin by defining $n+1$ maps from $G^n \to G^{n-1}$ for $n \geq 1$ given by

$$\begin{aligned} d^0(g_1,\ldots,g_n) &= (g_2,\ldots,g_n) \\ d^1(g_1,\ldots,g_n) &= (g_1g_2,g_3,\ldots,g_n) \\ d^2(g_1,\ldots,g_n) &= (g_1,g_2g_3,g_4,\ldots,g_n) \\ &\cdots\cdots \\ &\cdots\cdots \\ d^{n-1}(g_1,\ldots,g_n) &= (g_1,\ldots,g_{n-1}g_n) \\ d^n(g_1,\ldots,g_n) &= (g_1,\ldots,g_{n-1}) \end{aligned} \tag{17.53}$$

On the other hand, we can view an n-simplex Δ_n as

$$\Delta_n := \{(t_0,t_1,\ldots,t_n) | t_i \geq 0 \qquad \& \qquad \sum_{i=0}^{n} t_i = 1\} \tag{17.54}$$

Now, there are also $(n+1)$ *face maps* which map an $(n-1)$-simplex Δ_{n-1} into one of the $(n+1)$ faces of the n-simplex Δ_n:

$$\begin{aligned} d_0(t_0,\ldots,t_{n-1}) &= (0,t_0,\ldots,t_{n-1}) \\ d_1(t_0,\ldots,t_{n-1}) &= (t_0,0,t_1,\ldots,t_{n-1}) \\ &\cdots\cdots \\ &\cdots\cdots \\ d_n(t_0,\ldots,t_{n-1}) &= (t_0,\ldots,t_{n-1},0) \end{aligned} \tag{17.55}$$

d_i embeds the $(n-1)$ simplex into the face $t_i = 0$ which is opposite the ith vertex $t_i = 1$ of Δ_n.

Now we identify [237]

$$(\amalg_{n=0}^{\infty} \Delta_n \times G^n) \,/\sim$$

via

$$(d_i(\vec{t}),\vec{g}) \sim (\vec{t},d^i(\vec{g})). \tag{17.56}$$

The space we have constructed this way has a homotopy type denoted BG. This homotopy type is known as the *classifying space of the group* G. It can be characterized as the homotopy type of a topological space which is both contractible and admits a free G-action.

Note that for all $g \in G$, $\partial\Delta_1(g) = 0$, so for each group element there is a closed loop. On the other hand

$$\Delta_1(1_G) = \partial(\Delta_2(1_G,1_G)) \tag{17.57}$$

[237] This means we take the set of equivalence classes and impose the weakest topology on the set of equivalence classes so that the projection map is continuous.

so $\Delta_1(1_G)$ is a contractible loop. But all other loops are noncontractible. (Show this!) Therefore:

$$\pi_1(BG, *) \cong G \tag{17.58}$$

Moreover, if G is a finite group one can show that all the higher homotopy groups of BG are contractible. So then BG is what is known as an *Eilenberg MacLane space* $K(G, 1)$.

Even for the simplest nontrivial group $G = \mathbb{Z}/2\mathbb{Z}$ the construction is quite nontrivial and BG has the homotopy type of $\mathbb{R}P^\infty$.

Now, an n-cochain in $C^n(G, \mathbb{Z})$ (here we take $A = \mathbb{Z}$ for simplicity) is simply an assignment of an integer for each n-simplex in BG. Then the coboundary and boundary maps are related by

$$\langle d\phi_n, \Delta \rangle = \langle \phi_n, \partial\Delta \rangle \tag{17.59}$$

and from the above formulae we recover, rather beautifully, the formula for the coboundary in group cohomology.

Remarks:

1. When we defined group cohomology we also used homogeneous cochains. This is based on defining G as a groupoid from its left action and considering the mapping of groupoids $G//G \to pt//G$. ♣Explain more here? ♣

2. A Lie group is a manifold and hence has its own cohomology groups as a manifold, $H^n(G; \mathbb{Z})$. There is a relation between these: There is a group homomorphism

$$H^{n+1}_{\text{group cohomology}}(G; \mathbb{Z}) \to H^n_{\text{topological space cohomology}}(G; \mathbb{Z}) \tag{17.60}$$

3. One can show that $H^n(BG; \mathbb{Z})$ is always a finite abelian group if G is a finite group. [GIVE REFERENCE].

4. The above construction of BG is already somewhat nontrivial even for the trivial group $G = \{1_G\}$. Indeed, following it through for the 2-cell, we need to identify the three vertices of a triangle to one vertex, and the three edges to a single edge, embedded as a closed circle. If you do this by first identifying two edges and then try to identify the third edge you will see why it is called the "dunce's cap." It is true, but hard to visualize, that this is a contractible space. Things only get worse as we go to higher dimensions. A better construction, due to Milnor, is to construct what is known as a "simplicial set," and then collapse all degenerate simplices to a point. This gives a nicer realization of BG, but one which is homotopy equivalent to the one we described above. For the trivial category with one object and one morphism one just gets a topological space consisting of a single point. [238]

5. The "space" BG is really only defined up to homotopy equivalence. For some G there are very nice realizations as infinite-dimensional homogeneous spaces. This is useful for defining things like "universal connections." For example, one model for

[238] I thank G. Segal for helpful remarks on this issue.

$B\mathbb{Z}$ is as the humble circle $\mathbb{R}/\mathbb{Z} = S^1$. This generalizes to lattices $B\mathbb{Z}^d = T^d$, the d-dimensional torus. On the other hand $B\mathbb{Z}_2$ must be infinite-dimensional but it can be realized as $\mathbb{RP}^\infty$, the quotient of the unit sphere in a real infinite-dimensional separable Hilbert space by the antipodal map. Similarly, $BU(1)$ is $\mathbb{CP}^\infty$, realized as the quotient of the unit sphere in a complex infinite-dimensional separable Hilbert space by scaling vectors by a phase: $\psi \to e^{i\theta}\psi$.

18. Lattice Gauge Theory

As an application of some of the general concepts of group theory we discuss briefly lattice gauge theory.

Lattice gauge theory can be defined on any graph: There is a set of unoriented edges $\bar{\mathcal{E}}$. Each edge can be given either orientation and we denote the set of oriented edges by $\mathcal{E}$. The set of vertices is denoted $\mathcal{V}$ and source and target maps that tells us the vertex at the beginning and end of each oriented edge:

$$s : \mathcal{E} \to \mathcal{V} \qquad\qquad t : \mathcal{E} \to \mathcal{V} \tag{18.1}$$

We will view the union of edges $\bar{\mathcal{E}}$ (i.e. forgetting the orientation) as a topological space and denote it as Γ.

The original idea of Ken Wilson was that we could formulate Yang-Mills theory on a "lattice approximation to Euclidean spacetime" which we visualize as a cubic lattice in $\mathbb{R}^d$ for some d. Then, the heuristic idea is, that as the bond lengths are taken to zero we get a good approximation to a field theory in the continuum. Making this idea precise is highly nontrivial! For example, just one of the many issues that arise is that important symmetries such as Euclidean or Poincaré symmetries of the continuum models we wish to understand are broken, in this formulation, to crystallographic symmetries.

18.1 Some Simple Preliminary Computations

A rather trivial part of the idea is to notice the following: Suppose we have a field theory on $\mathbb{R}^d$ of fields

$$\phi : \mathbb{R}^d \to \mathcal{T} \tag{18.2}$$

where $\mathcal{T}$ is some "target space." Then if we consider the embedded hypercubic lattice:

$$\Lambda_a := \{a(n_1, \ldots, n_d) \in \mathbb{R}^d | n_i \in \mathbb{Z}\} \tag{18.3}$$

and we restrict ϕ to Λ_a then at neighboring vertices the value of ϕ will converge as $a \to 0$:

$$\lim_{a\to 0} \phi(\vec{x}_0 + a\hat{e}_\mu) = \phi(\vec{x}_0) \tag{18.4}$$

where $\hat{e}_\mu$, $\mu = 1, \ldots, d$ is a unit vector in the μth direction. Moreover, if $\phi : \mathbb{R}^d \to \mathcal{T}$ is differentiable and $\mathcal{T}$ is a linear space then

$$\lim_{a\to 0} a^{-1}(\phi(\vec{x}_0 + a\hat{e}_\mu) - \phi(\vec{x}_0)) = \partial_\mu \phi(\vec{x}_0) \tag{18.5}$$

and so on.

In lattice field theory we attempt to go the other way: We assume that we have fields defined on a sequence of lattices $\Lambda_a \subset \mathbb{R}^d$ and try to take an $a \to 0$ limit to define a continuum field theory.

Here is a simple paradigm to keep in mind: [239] Consider the one-dimensional lattice $\mathbb{Z}$, but it is embedded in the real line so that bond-length is a, so $\Lambda_a = \{an | n \in \mathbb{Z}\} \subset \mathbb{R}$. Our degrees of freedom will be a real number $\phi_\ell(n)$ at each lattice site $n \in \mathbb{Z}$, and it will evolve in time to give a motion $\phi_\ell(n,t)$ according to the action:

$$S = \int_{\mathbb{R}} dt \sum_{n \in \mathbb{Z}} \left(\frac{m}{2} \dot{\phi}_\ell(n,t)^2 - \frac{k}{2} \left(\phi_\ell(n,t) - \phi_\ell(n+1,t) \right)^2 \right) \tag{18.6}$$

We can think of this as a system of particles of mass m fixed at the vertices of Λ_a with neighboring particles connected by a spring with spring constant k. For the action to have proper units, $\phi_\ell(n,t)$ should have dimensions of length, suggesting it measures the displacement of the particle in some orthogonal direction to the real line. The equations of motion are of course:

$$m \frac{d^2}{dt^2} \phi_\ell(n,t) = k(\phi_\ell(n+1,t) - 2\phi_\ell(n,t) + \phi_\ell(n-1,t)) \tag{18.7}$$

Now we wish to take the $a \to 0$ limit. We assume that there is some differentiable function $\phi_{cont}(x,t)$ such that

$$\phi_{cont}(x,t)|_{x=an} = \phi_\ell(n,t) \tag{18.8}$$

so by Taylor expansion

$$\phi_\ell(n+1,t) - 2\phi_\ell(n,t) + \phi_\ell(n-1,t) = a^2 \frac{d^2}{dx^2} \phi_{cont}|_{x=an} + \mathcal{O}(a^3) \tag{18.9}$$

Now suppose we scale the parameters of the Lagrangian so that

$$m = aT \qquad k = \frac{v^2 T}{a} \tag{18.10}$$

then, if the limits really exist, the continuum function $\phi_{cont}(x,t)$ must satisfy the wave equation:

$$\frac{d^2}{dt^2} \phi_{cont} - v^2 \frac{d^2}{dx^2} \phi_{cont} = 0 \tag{18.11}$$

whose general solution is

$$\Phi_{left}(x+vt) + \Phi_{right}(x-vt) \tag{18.12}$$

The general solution is described by arbitrary wavepackets traveling to the left and right along the real line. (We took $v > 0$ here.) We can also see this at the level of the Lagrangian since if $\phi_\ell(n,t)$ is well-approximated by a continuum function $\phi_{cont}(x,t)$ then

$$S \to T \int_{\mathbb{R}} dt \int_{\mathbb{R}} dx \left[\frac{1}{2} \left(\frac{d}{dt} \phi_{cont} \right)^2 - \frac{v^2}{2} \left(\frac{d}{dx} \phi_{cont} \right)^2 \right] + \mathcal{O}(a) \tag{18.13}$$

Remarks

[239] Here we will just latticize the spatial dimension of a $1+1$ dimensional field theory. In the rest of the section we latticize spacetime with Euclidean signature.

1. In the lattice theory there will certainly be sequences of field configurations $\phi_{lattice}(n,t)$ that have no good continuum limit. The idea is that these are unimportant to the physics because they have huge actions whose contributions to the path integral is unimportant in the continuum limit.

2. Keeping in mind the interpretation of $\phi_{cont}(x,t)$ as a height in a direction orthogonal to the real axis, we see that we are describing a string of tension T.

18.2 Gauge Group And Gauge Field

In lattice gauge theory we choose a group G - known as the *gauge group*. For the moment it can be any group. The dynamical degree of freedom is a *gauge field*, or more precisely, the dynamical object is the *gauge equivalence class* or isomorphism class of the gauge field. This will be defined below.

In mathematics, a gauge field is called a *connection*.

To give the definition of a connection let $\mathcal{P}$ be the set of all connected open paths in Γ. For example, we can think of it as the set of continuous maps $\gamma : [0,1] \to \Gamma$. Since we are working on a graph you can also think of a path γ as a sequence of edges $e_1, e_2, \ldots, e_k$ such that

$$t(e_i) = s(e_{i+1}) \qquad 1 \leq i \leq k-1 \tag{18.14}$$

(We also allow for the trivial path $\gamma_v(t) = v$ for some fixed vertex v which has no edges.) However, the former definition is superior because it generalizes to connections on other topological spaces.

Now, by definition, a connection is just a map

$$\mathbb{U} : \mathcal{P} \to G, \tag{18.15}$$

which satisfies the composition law: If we concatenate two paths γ_1 and γ_2 to make a path $\gamma_1 \circ \gamma_2$, so that the concatenated path begins at $\gamma_1(0)$ and ends at $\gamma_2(1)$ and such that $\gamma_1(1) = \gamma_2(0)$, that is, the end of γ_1 is the beginning of γ_2, then we must have:

$$\mathbb{U}(\gamma_1 \circ \gamma_2) = \mathbb{U}(\gamma_1)\mathbb{U}(\gamma_2) \tag{18.16}$$

If our path is the trivial path then

$$\mathbb{U}(\gamma_v) = 1_G \tag{18.17}$$

and if $\gamma^{-1}(t) = \gamma(1-t)$ is the path run backwards then

$$\mathbb{U}(\gamma^{-1}) = (\mathbb{U}(\gamma))^{-1} \tag{18.18}$$

♣Are these really independent conditions? ♣

Note that if the path γ is made by concatenating edges $e_1, e_2, \ldots, e_k$ then

$$\mathbb{U}(\gamma) = \mathbb{U}(e_1)\mathbb{U}(e_2)\cdots\mathbb{U}(e_k) \tag{18.19}$$

so, really, in lattice gauge theory it suffices to know the $\mathbb{U}(e)$ for the edges. If e^{-1} is the edge e with the opposite orientation then

$$\mathbb{U}(e^{-1}) = \mathbb{U}(e)^{-1} \tag{18.20}$$

We will denote the space of all connections by $\mathcal{A}(\Gamma)$.

Remark: *Background heuristics*: For those who know something about gauge fields in field theory we should think of $\mathbb{U}(e)$ as the parallel transport (in some trivialization of our principal bundle) along the edge e. From these parallel transports along edges we can recover the components of the gauge field. To explain more let us assume for simplicity that $G = U(N)$ is a unitary group, or some matrix subgroup of $U(N)$.

Recall some elementary ideas from the theory of Lie groups: If α is any anti-Hermitian matrix then $\exp[\alpha]$ is a unitary matrix. Moreover, if α is "small" then $\exp[\alpha]$ is close to the identity. Conversely, if U is "close" to the identity then it can be uniquely written in the form $U = \exp[\alpha]$ for a "small" anti-Hermitian matrix α. Put more formally: The tangent space to $U(N)$ at the identity is the (real!) vector space of $N \times N$ anti-Hermitian matrices. (This vector space is a real Lie algebra, because the commutator of anti-Hermitian matrices is an anti-Hermitian matrix.) Moreover, the exponential map gives a good coordinate chart in some neighborhood of the identity of the topological group $U(N)$.

The poor man's way of understanding the relation between Lie algebras and Lie groups is to use the very useful Baker-Campbell-Hausdorf formula: If A, B are $n \times n$ matrices then the formula gives an expression for an $n \times n$ matrix C so that

$$e^A e^B = e^C \tag{18.21}$$

The formula is a (very explicit) infinite set of terms all expressed in terms of multiple commutators. The first few terms are:

$$\boxed{C = A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \frac{1}{12}[B,[B,A]] + \frac{1}{24}[A,[B,[A,B]]] + \cdots} \tag{18.22}$$

The series is convergent as long as A, B are small enough (technically, such that the characteristic values of $\mathrm{Ad}(A)$ and $\mathrm{Ad}(B)$ are less than 2π in magnitude). See Chapter 8 for a full explanation. Note in particular that if we expand in small parameters ϵ_1, ϵ_2 then

$$e^{\epsilon_1 A} e^{\epsilon_2 B} e^{-\epsilon_1 A} e^{-\epsilon_2 B} = e^{\epsilon_1 \epsilon_2 [A,B] + \cdots} \tag{18.23}$$

Now, returning to lattice gauge theory: In the usual picture of "approximating" Euclidean $\mathbb{R}^d$ by $\mathbb{Z}^d$ with bond-length a we can write a fundamental edge $e_\mu(\vec{n})$ as the straight line in $\mathbb{R}^d$ from $\vec{n}$ to $\vec{n} + a\hat{e}_\mu$. If a is small and we have some suitable continuity then $\mathbb{U}(e_\mu(\vec{n}))$ will be near the identity and we can write:

$$\mathbb{U}(e_\mu(\vec{n})) = \exp[a A_\mu^{lattice}(\vec{n})] \tag{18.24}$$

for some anti-Hermitian matrix $A_\mu(a\vec{n})$. In lattice gauge theory, the connections with a good continuum limit are those such that there is a locally defined 1-form valued in $N \times N$ anti-Hermitian matrices $A_\mu^{cont}(\vec{x})dx^\mu$ so that $A_\mu^{cont}(a\vec{n}) = A_\mu^{lattice}(\vec{n})$.

Now, the gauge field $\mathbb{U}$ has redundant information in it. The reason it is useful to include this redundant information is that many aspects of locality become much clearer

when working with $\mathcal{A}(\Gamma)$ as we will see when trying to write actions below. The redundant information is reflected in a *gauge transformation* which is simply a map

$$f : \mathcal{V} \to G \tag{18.25}$$

The idea is that if γ is a path then the gauge fields $\mathbb{U}$ and $\mathbb{U}'$ related by the rule

$$\mathbb{U}'(\gamma) = f(s(\gamma))\mathbb{U}(\gamma)f(t(\gamma))^{-1} \tag{18.26}$$

are deemed to be gauge equivalent, i.e. isomorphic. We denote the set of gauge transformations by $\mathcal{G}(\Gamma)$. Note that, being a function space whose target is a group, this set is a group in a natural way. It is called *the group of gauge transformations.* [240] The group of gauge transformations $\mathcal{G}(\Gamma)$ acts on $\mathcal{A}(\Gamma)$. The moduli space of gauge inequivalent fields is the set of equivalence classes: $\mathcal{A}(\Gamma)/\mathcal{G}(\Gamma)$. Mathematicians would call these isomorphism classes of connections.

It might seem like there is no content here. Can't we always choose $f(s(\gamma))$ to set $\mathbb{U}'(\gamma)$ to 1? Yes, in general, except when $s(\gamma) = t(\gamma)$, that is, when γ is a closed loop based at a vertex, say v_0. For such closed loops we are stuck, all we can do by gauge transformations is conjugate:

$$\mathbb{U}'(\gamma) = g\mathbb{U}(\gamma)g^{-1} \tag{18.27}$$

where g is the gauge transformation at v_0. Moreover, if we start the closed loop at another vertex on the loop then the parallel transport is again in the same conjugacy class. Thus there is gauge invariant information associated to a loop γ: The conjugacy class of the $\mathbb{U}(\gamma)$. That is: The *holonomy function*:

$$\mathrm{Hol}_{\mathbb{U}} : L\Gamma \to \mathrm{Conj}(G) \tag{18.28}$$

that maps the loops in Γ to the conjugacy class:

$$\mathrm{Hol}_{\mathbb{U}} : \gamma \mapsto C(\mathbb{U}(\gamma)) \tag{18.29}$$

is gauge invariant: If $\mathbb{U}' \sim \mathbb{U}$ are gauge equivalent then

$$\mathrm{Hol}_{\mathbb{U}'} = \mathrm{Hol}_{\mathbb{U}} \tag{18.30}$$

In fact, one can show that $\mathrm{Hol}_{\mathbb{U}}$ is a complete invariant, meaning that we have the converse: If $\mathrm{Hol}_{\mathbb{U}'} = \mathrm{Hol}_{\mathbb{U}}$ then $\mathbb{U}'$ is gauge equivalent to $\mathbb{U}$. Put informally:

> *The gauge invariant information in a gauge field, or connection, is encoded in the set of conjugacy classes associated to the closed loops in* Γ.

Exercise

[240] AND IS NOT TO BE CONFUSED WITH THE gauge group G!!!!

Show that if γ is a closed loop beginning and ending at v_0 and if v_1 is another vertex on the path γ then if γ' describes the "same" loop but starting at v_1 then $\mathbb{U}(\gamma)$ and $\mathbb{U}(\gamma')$ are in the same conjugacy class in G.

Exercise

Consider a graph Γ which forms a star: There is one central vertex, and r "legs" each consisting or N_i edges radiating outward, where $i = 1, \ldots, r$.

a.) Show explicitly that any gauge field can be gauged to $\mathbb{U} = 1$.

b.) What is the unbroken subgroup of the group of gauge transformations? (That is, what is the automorphism group of the gauge field $\mathbb{U} = 1$?)

Exercise

Consider a d-dimensional hypercubic lattice with periodic boundary conditions, so that we are "approximating a torus" which is a product of "circles" of length Na.

What is the maximal number of edges so that we can set $\mathbb{U}(e) = 1$?

18.3 Defining A Partition Function

Next, to do physics, we need to define a gauge invariant action. At the most general level this is simply a function $F : \mathcal{A}(\Gamma)/\mathcal{G}(\Gamma) \to \mathbb{C}$ so that we can define a "partition function":

$$Z = \sum_{[\mathbb{U}] \in \mathcal{A}(\Gamma)/\mathcal{G}(\Gamma)} F([\mathbb{U}]) \tag{18.31}$$

If Γ is finite and G is finite this sum is just a finite sum. If Γ is finite and G is a finite-dimensional Lie group then $\mathcal{A}(\Gamma)/\mathcal{G}(\Gamma)$ is a finite-dimensional topological space and the "sum" needs to be interpreted as some kind of integral. Since a connection on Γ is completely determined by its values on the elementary edges (for a single orientation) we can, noncanonically, identity the space of all connections as

$$\mathcal{A}(\Gamma) \cong G^{|\bar{\mathcal{E}}|}. \tag{18.32}$$

Similarly

$$\mathcal{G}(\Gamma) \cong G^{|\mathcal{V}|} \tag{18.33}$$

Now we need a way of integrating over the group. If G is a finite group and $F : G \to \mathbb{C}$ is a function then

$$\int_G F d\mu := \frac{1}{|G|} \sum_{g \in G} F(g) \tag{18.34}$$

This basic idea can be generalized to Lie groups. A Lie group is a manifold and we define a measure on it $d\mu$. (If G is a simple Lie group then there is a canonical choice of measure up to an overall scale.) As a simple example, coonsider $U(1) = \{e^{i\theta}\}$ then the integration is

$$\int_0^{2\pi} F(e^{i\theta}) \frac{d\theta}{2\pi} \tag{18.35}$$

In all cases, the crucial property of the group integration is that, for all h we have

$$\int_G F(gh) d\mu(g) = \int_G F(hg) d\mu(g) = \int_G F(g) d\mu(g) \tag{18.36}$$

This property defines what is called a *left-right-invariant measure*. It is also known as the *Haar measure*.

In general the Haar measure is only defined up to an overall scale. In the above examples we chose the normalization so that the volume of the group is 1.

Now, choosing a left-right-invariant measure we can define:

$$Z = \frac{1}{\mathrm{vol}\,(\mathcal{G}(\Gamma))} \int_{\mathcal{A}(\Gamma)} \hat{F}(\mathbb{U}) d\mu_{\mathcal{A}(\Gamma)} \tag{18.37}$$

where $\hat{F}$ is a lifting of F to a $\mathcal{G}(\Gamma)$-invariant function on $\mathcal{A}(\Gamma)$ and $d\mu_{\mathcal{G}(\Gamma)}$ is the Haar measure on $G^{|\bar{\mathcal{E}}|}$ induced by a choice of Haar measure on G. It is gauge invariant because the Haar measure is left- and right- invariant.

If we want to impose locality then it is natural to have $\hat{F}(\mathbb{U})$ depend only on the local gauge invariant data. This motivates us to consider "small" loops and consider a *class function*.

In general, a class function on a group G is a function $F : G \to \mathbb{C}$ such that $F(hgh^{-1}) = F(g)$ for all $h \in G$. We should clearly take $\hat{F}$ to be some kind of class function. A natural source of class functions are traces in representations, for if $\rho : G \to GL(N, \mathbb{C})$ is a matrix representation then $\chi(g) := \mathrm{Tr}\rho(g)$ is a class function by cyclicity of the trace. (This class function is called the *character of the representation*.)

The smallest closed loops we can make are the "plaquettes." For $\Lambda_a \subset \mathbb{R}^d$ these would be labeled by a pair of directions μ, ν with $\mu \neq \nu$ and would be the closed loop

$$a\vec{n} \to a\vec{n} + a\hat{e}_\mu \to a\vec{n} + a\hat{e}_\mu + a\hat{e}_\nu \to a\vec{n} + a\hat{e}_\nu \to a\vec{n} \tag{18.38}$$

Let us denote this plaquette as $p_{\mu\nu}(\vec{n})$.

♣FIGURE NEEDED HERE! ♣

Given a class function $F : G \to \mathbb{C}$ we can form a partition function by taking

$$\hat{F}(\mathbb{U}) := e^{-S(\mathbb{U})} := e^{-\sum_p S(p)} \tag{18.39}$$

where we have summed over all plaquettes in the exponential to make this look more like a discrete approximation to a field theory path integral, and the action $S(p)$ of a plaquette p is some class function applied to $\mathbb{U}(p)$. If G is a continuous group then we need to interpret the sum over $\mathcal{A}(\Gamma)/\mathcal{G}(\Gamma)$ as some kind of integral, as discussed above.

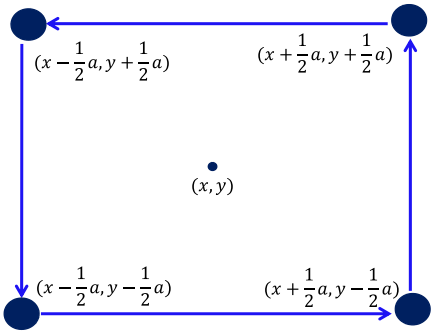


Figure 47: A small plaquette, centered on a surface element in a tangent plane with coordinates (x, y) and centered on a point with coordinates (x, y). The holonomy around the plaquette, to leading order in an expansion in small values of bond-length a is governed by the curvature tensor evaluated on that area element.

Remark: *More background heuristics*: For those who know something about gauge fields in field theory we should think of the parallel transport $\mathbb{U}(p)$ around a plaquette p as defining the components of the curvature on a small area element $dx^\mu \wedge dx^\nu$ at some point $\vec{x}_0 = a\vec{n}$ (in some framing). Indeed, using the idea that

$$\mathbb{U}(e_\mu(\vec{n})) \sim \exp[aA_\mu^{cont}|_{\vec{x}=a\vec{n}}] \tag{18.40}$$

we can try to take a "limit" where $a \to 0$. The plaquette $p_{\mu\nu}(\vec{n})$ is two-dimensional so, temporarily choosing coordinates so that $\mu = 1$ and $\nu = 2$ we can write the plaquette gauge group element as

$$e^{aA_1(x,y-\frac{1}{2}a)} e^{aA_2(x+\frac{1}{2}a,y)} e^{-aA_1(x,y+\frac{1}{2}a)} e^{-aA_2(x-\frac{1}{2}a,y)} \tag{18.41}$$

See Figure 47. Now, using the BCH formula [241] we define the *fieldstrength of the gauge field* or, equivalently, the *curvature of the connection* by

$$\mathbb{U}(p_{\mu\nu}(\vec{n})) = \exp[a^2 F_{\mu\nu} + \mathcal{O}(a^4)] \tag{18.44}$$

Here in the continuum we would have the relation:

$$F_{\mu\nu}(\vec{x}) = \partial_\mu A_\nu(\vec{x}) - \partial_\nu A_\mu(\vec{x}) + [A_\mu(\vec{x}), A_\nu(\vec{x})] \tag{18.45}$$

[241] Warning: If you are not careful the algebra can be extremely cumbersome here! Taylor expansion in to order a^2 gives:

$$e^{aA_1 - \frac{a^2}{2}\partial_2 A_1} e^{aA_2 + \frac{a^2}{2}\partial_1 A_2} e^{-aA_1 - \frac{a^2}{2}\partial_2 A_1} e^{-aA_2 + \frac{a^2}{2}\partial_1 A_2} \tag{18.42}$$

We only need to keep the first commutator term in the BCH formula if we are working to order a^2 so we get

$$e^{a^2(\partial_1 A_2 - \partial_2 A_1 + [A_1, A_2]) + \mathcal{O}(a^3)} \tag{18.43}$$

A standard action used in lattice gauge theory in the literature is constructed as follows:

First, choose a finite-dimensional unitary representation of G, that is, a group homomorphism

$$\rho : G \to U(r) \tag{18.46}$$

Next, define the action for a plaquette to be

$$S(p) = K(r - \mathrm{Re}[\mathrm{Tr}\rho(\mathbb{U}(p))]) \tag{18.47}$$

for some constant K. Note that the trivial gauge field has action $S(p) = 0$. Moreover, every unitary matrix can be diagonalized, by the spectral theorem, with eigenvalues $e^{\mathrm{i}\theta_i(p)}$, $i = 1, \ldots, r$ and then

$$S(p) = K\sum_{i=1}^{r}(1 - \cos\theta_i(p)) = 2K\sum_{i=1}\sin^2(\theta_i(p)/2) \tag{18.48}$$

is clearly positive definite for $K > 0$. This is good for unitarity (or its Euclidean counterpart - "reflection positivity.")

Remarks:

1. *Correlation Functions*: The typical physical quantities we might want to compute are expectation values of products of gauge invariant operators. In view of our discussion of gauge equivalence classes of gauge fields above one very natural way to make such gauge invariant operators is via *Wilson loop operators*. For these one chooses a matrix representation $R : G \to GL(N, \mathbb{C})$ of G (totally unrelated to the choice we made in defining the action) and a particular loop γ and defines:

   $$W(R, \gamma)(\mathbb{U}) := \mathrm{Tr}_{\mathbb{C}^N} R(\mathbb{U}(\gamma)) \tag{18.49}$$

   So, $W(R, \gamma)$ should be regarded as a gauge invariant function

   $$W(R, \gamma) : \mathcal{A}(\Gamma) \to \mathbb{C} \tag{18.50}$$

   and therefore we can consider the expectation values:

   $$\langle\prod_i W(R_i, \gamma_i)\rangle := \frac{\int_{\mathcal{A}(\Gamma)}\prod_i W(R_i, \gamma_i)e^{-S(\mathbb{U})}d\mu_{\mathcal{A}(\Gamma)}}{\int_{\mathcal{A}(\Gamma)}e^{-S(\mathbb{U})}d\mu_{\mathcal{A}(\Gamma)}} \tag{18.51}$$

2. *Yet more background heuristics*: For those who know something about gauge fields in field theory we can begin to recognize something like the Yang-Mills action if we use (18.44) and write

   $$S(p) = K\sum_{p_{\mu\nu}(\vec{n})}(r - \mathrm{Re}[\mathrm{Tr}\rho(\mathbb{U}(p_{\mu\nu}(\vec{n})))]) \to -\frac{1}{2}Ka^4\sum_{\vec{n}\in\mathbb{Z}^d}\sum_{\mu\neq\nu}\mathrm{Tr}\rho(F_{\mu\nu}(a\vec{n}))^2 \tag{18.52}$$

   ♣There is a bit of a cheat here since you did not work out the plaquette to order a^4. ♣

The heuristic limit (18.52) is to be compared with the Yang-Mills action

$$S_{YM} = -\frac{1}{2g_0^2}\int_X d^d x \sqrt{\det g} g^{\mu\lambda} g^{\nu\rho} \mathrm{Tr} F_{\mu\nu} F_{\lambda\rho} \tag{18.53}$$

where here we wrote it in Euclidean signature on a Riemannian manifold M. The trace is in some suitable representation and the normalization of the trace can be absorbed in a rescaling of the coupling constant g_0. If we use the representation $\rho : G \to U(r)$ then

$$\frac{1}{g_0^2} = K a^{4-d} \qquad \Rightarrow \qquad K = \frac{a^{d-4}}{g_0^2} \tag{18.54}$$

The constant K must be dimensionless so that $d = 4$ dimensions is selected as special. For $d = 4$ the Yang-Mills coupling g_0^2 is dimensionless. It has dimensions of length to a positive power for $d > 4$ and length to a negative power for $d < 4$. To take the continuum limit we should hold g_0^2 fixed and scale K as above as $a \to 0$.

3. *Very important subtlety in the case $d = 4$* Actually, if one attempts to take the limit more carefully, the situation becomes more complicated in $d = 4$, because in quantum mechanics there are important effects known as *vacuum fluctuations*. What is expected to happen (based on continuum field theory) is that, if we replace K by $g^{-2}(a)$ and allow a-dependence then we can get a good limit of, say, correlation functions of Wilson loop vev's if we scale $g^2(a)$ so that

$$\frac{8\pi^2}{g^2(a_1)} = \frac{8\pi^2}{g^2(a_2)} + \beta \log \frac{a_1}{a_2} + \mathcal{O}(g^2(a_2)) \tag{18.55}$$

where there are higher order terms in the RHS in an expansion in $g^2(a_2)$. Here β is a constant, depending on the gauge group G and other fields in the theory. For $G = SU(n)$ we have the renowned result of D. Gross and F. Wilczek, and of D. Politzer that

$$\beta = -\frac{11}{3} n \tag{18.56}$$

As long as $\beta < 0$ we see that $g^2(a_2) \to 0$ as $a_2 \to 0$. This is known as *asymptotic freedom*. It has the good property that as we attempt to take $a_2 \to 0$ the higher order terms on the RHS are at least formally going to zero.

4. One can therefore ask, to what extent is this continuum limit rigorously defined and how rigorously has (18.55) been established from the lattice gauge theory approach. My impression is that it is still open. Two textbooks on this subject are:

   1. C. Itzykson and J.-M. Drouffe, *Statistical Field Theory*, Cambridge
   2. M. Creutz, *Quarks, gluons, and lattices*, Cambridge

5. *Phases and confinement.* Many crucial physical properties can be deduced from Wilson loop vev's. In Yang-Mills theory a crucial question is whether, for large planar loops γ $\langle W(R,\gamma)\rangle$ decays like $\exp[-T Area(\gamma)]$ or $\exp[-\mu Perimeter(\gamma)]$. If it decays like the area one can argue that quarks will be confined. For a nice explanation see S. Coleman, *Aspects Of Symmetry*, for a crystal clear explanation.

6. *Including quarks and QCD.* The beta function is further modified if there are "matter fields" coupling to the gauge fields. If we introduce n_f Dirac fermions in the fundamental representation of $SU(n)$ then (18.56) is modified to:

$$\beta = -\left(\frac{11}{3}n - \frac{2}{3}n_f\right) \tag{18.57}$$

The theory of the strong nuclear force between quarks and gluons is based on $n = 3$ and $n_f = 6$. Actually, there is a strong hierarchy of quark masses so for low energy questions $n_f = 2$ (for "up" and "down" quarks) is more relevant.

7. There are very special situations in which $\beta = 0$ and in fact all the higher terms on the RHS of the "renormalization group equation" (18.55) vanish. These lead to scale-invariant theories, and in good cases, to conformal field theories. In the modern viewpoint on field theory, these conformal field theories are the basic building blocks of all quantum field theories.

18.4 Hamiltonian Formulation

EXPLAIN HILBERT SPACE FOR 1+1 CASE IS $L^2(G)$.

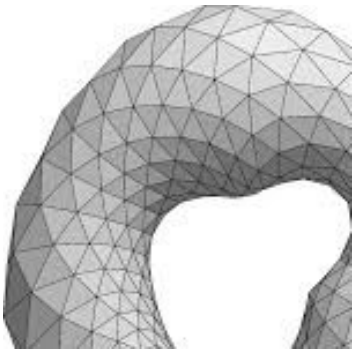


Figure 48: A triangulated surface. Figure from Wikipedia.

18.5 Topological Gauge Theory

A very popular subject in discussions of topological phases of matter is a set of models known as "topological gauge theories." In general, topological field theories are special

classes of field theories that are independent of distances in spacetime. They focus on the topological aspects of physics. A formal mathematical definition is that it is a functor from some bordism category to, say, the category $\mathbf{Vect}_\kappa$.

If G is a finite group and we are working on a smooth manifold then there can be no curvature tensor, so all gauge fields are "flat." They can still be nontrivial since $\mathbb{U}(\gamma)$ can still be nontrivial for homotopically nontrivial loops. The simplest example would be $0+1$ dimensional Yang-Mills theory on a circle. If the action is literally zero then the partition function is just

$$Z = \frac{1}{|G|}\sum_{g\in G} 1 \tag{18.58}$$

Recalling our discussion of the class equation we recognize that the partition function can be written as:

$$Z = \sum_{c.c.} \frac{1}{|Z(g)|} \tag{18.59}$$

where we sum over conjugacy classes in the group and weight each class by one over the order of the centralizer of some (any) representative of that class. This second form of the sum can be interpreted as a sum over the isomorphism classes of principal G-bundles over the circle, weighted by one over the automorphism group of the bundle.

For those who know something about gauge theory note that this illustrates a very general principle: *In the partition function of a gauge theory we sum over all the isomorphism classes of bundles with connection: We weight the bundle with connection by a gauge invariant functional divided by the order of the automorphism group of the bundle with connection.*

It is also worth remarking that, quite generally in field theory, the partition function on a manifold of the form $X \times S^1$ can be interpreted as a trace in a Hilbert space. With proper boundary conditions for the "fields" around S^1 we simply have

$$Z(X\times S^1) = \mathrm{Tr}_{\mathcal{H}(X)} e^{-\beta H} \tag{18.60}$$

where β is the length of the circle. In a topological theory the Hamiltonian $H = 0$, so we just get the dimension of the Hilbert space associated to the spatial slice X. In the case of Yang-Mills in $0+1$ dimensions we see that the Hilbert space associated to a point is just $\mathcal{H} = \mathbb{C}$.

In lattice models of topological gauge theories in higher dimensions we insert the gauge-invariant function

$$\prod_p \delta(\mathbb{U}(p)) \tag{18.61}$$

where $\delta(g)$ is the Dirac delta function relative to the measure $d\mu(g)$ we chose on G, and is concentrated at $g = 1_G$. Here we take the product over all plaquettes that are meant to be "filled in" in the continuum limit. That means that the parallel transport around "small" loops defined by plaquettes will be trivial. This does not mean that the gauge field is trivial! For example if we consider a triangulation of a compact surface or higher dimensional manifold with nontrivial fundamental group then there can be nontrivial holonomy around

homotopically nontrivial loops. In general, a connection, or gauge field, such that $\mathbb{U}(\gamma) = 1$ for homotopically nontrivial loops (this is equivalent to the vanishing of the curvature 2-form $F_{\mu\nu}$) is known as a *flat connection* or *flat gauge field*. In topological gauge theories we sum over (isomorphism classes of) flat connections.

Note that (18.61) is just part of the definition of a topological gauge theory. We want to do this so that physical quantities only depend on topological aspects of the theory. In standard Yang-Mills theory $\langle W(R,\gamma)\rangle$ will depend on lots of details of γ. Indeed, one definition of the curvature is how $W(R,\gamma)$ responds to small deformations of γ. In topological gauge theories we want

$$\langle \prod_i W(R_i,\gamma_i)\rangle \tag{18.62}$$

to be independent of (nonintersecting!) γ_i under homotopy. Therefore, our measure should be concentrated on flat gauge fields, at least in some heuristic sense. In lattice topological gauge theory we do this by hand.

Remark: In general, flat gauge fields for a group G on a manifold M are classified, up to gauge equivalence by the conjugacy classes of homomorphisms $\mathrm{Hom}(\pi_1(M,x_0),G)$.

For a flat gauge field, the standard Wilson action we discussed above will simply vanish. We can get a wider class of models by using group cocycles. This was pointed out in the paper

R. Dijkgraaf and E. Witten, "Topological Gauge Theory And Group Cohomology," Commun.Math.Phys. 129 (1990) 39.

and topological gauge theories that make use of group cocycles for the action are now known as *Dijkgraaf-Witten models*.

For simplicity we now take our group G to be a finite group. Let us start with a two-dimensional model. We can view Γ as a triangulation of an oriented surface M as in Figure 48. We want a local action, so let us restrict to a flat gauge field on a triangle as in Figure 45. We want to assign the local "Boltzman weight." It will be a function:

$$W : G \times G \to \mathbb{C}^* \tag{18.63}$$

(If we wish to match to some popular physical theories we might take it to be $U(1)$-valued. The distinction will not matter for anything we discuss here.) Now referring to Figure 45 we assign the weight

$$W(g_1,g_2) \tag{18.64}$$

to this triangle. But now we have to decide if we are to use this, or $W(g_2,(g_1g_2)^{-1})$ or $W((g_1g_2)^{-1},g_1)$. In general these complex numbers will not be equal to each other. So we number the vertices $1,\ldots,|\mathcal{V}|$ and then for any triangle T we start with the vertices with the two smallest numbers. Call this $W(T)$. This will define an orientation that might or might not agree with that on the surface M. Let $\epsilon(T) = +1$ if it agrees and $\epsilon(T) = -1$ if

it does not. Then the Boltzman weight for a flat gauge field configuration $\mathbb{U}$ on the entire surface is defined to be

$$W(\mathbb{U}) := \prod_T W(T)^{\epsilon(T)} \tag{18.65}$$

Now, if this weight is to be at all physically meaningful we definitely want the dependence on all sorts of choices to drop out.

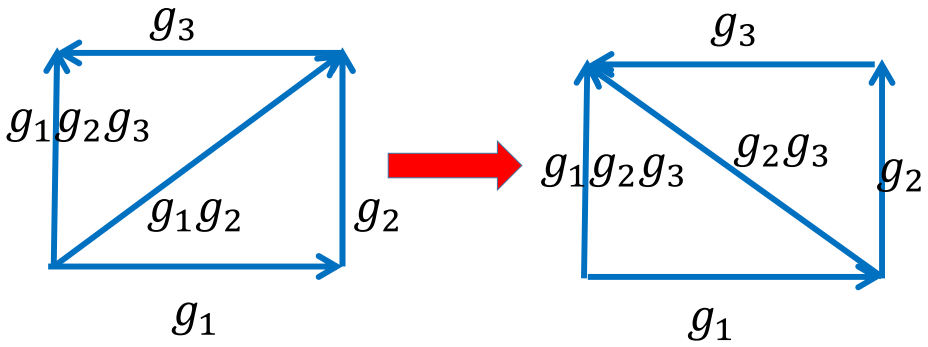


Figure 49: A local change of triangulation of type I.

Now, one thing we definitely want to have is independence of the choice of triangulation. A theorem of combinatorial topology states that any two triangulations can be related by a sequence of local changes of type I and type II illustrated in Figure 49 and 50, respectively. We see that the invariance of the action under type I requires:

$$W(g_1, g_2)W(g_1g_2, g_3) = W(g_1, g_2g_3)W(g_2, g_3) \tag{18.66}$$

and this is the condition that W should be a 2-cocycle. Similarly, the change of type II doesn't matter provided

$$W(g_1, g_2) = W(g_1, g_2g_3)W(g_2, g_3)W(g_1g_2, g_3)^{-1} \tag{18.67}$$

which is again guaranteed by the cocycle equation! This strongly suggests we can get a good theory by using a 2-cocycle, and that is indeed the case. But we need to check some things first:

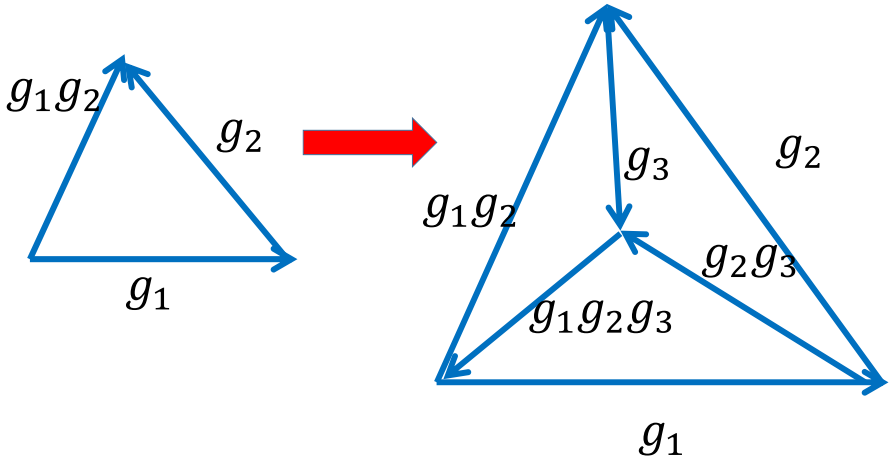


Figure 50: A local change of triangulation of type II.

1. The dependence on the labeling of the vertices drops out using an argument based on topology we haven't covered. This can be found in the Dijkgraaf-Witten paper. Similarly, if W is changed by a coboundary then we modify

$$W(g_1, g_2) \to W(g_1, g_2)\frac{t(g_1)t(g_2)}{t(g_1g_2)} \tag{18.68}$$

   that is, we modify the weight by a factor based on a product around the edges. When multiplying the contributions of the individual triangles to get the total weight (18.65) the edge factors will cancel out from the two triangles sharing a common edge.

2. The action is not obviously gauge invariant, since it is certainly not true in general that $W(g_1, g_2)$ is equal to

$$W(h(v_1)^{-1}g_1h(v_2), h(v_2)^{-1}g_2h(v_3)) \tag{18.69}$$

   for all group elements $h(v_1), h(v_2), h(v_3) \in G$. The argument that, nevertheless, the total action (18.65) is invariant is given (for the $d = 3$ case) in the Dijkgraaf-Witten paper around their equation (6.29).

   ♣Cop out. Give a better argument. Explain that Chern-Simons actions change by boundary terms and it is too much to hope for exact local gauge invariance. ♣

3. The idea above generalizes to define a topological gauge theory on oriented manifolds in d-dimensions for any d, where one uses a d-cocycle on G with values in $\mathbb{C}^*$ (or $U(1)$). These topological gauge theories are known as "Dijkgraaf-Witten theories." The Boltzmann weight W represents a topological term in the action that exists and is nontrivial even for flat gauge fields.

4. The invariance under the change of type II in Figure 50, which can be generalized to all dimensions is particularly interesting. It means that the action is an "exact renormalization group invariant" in the sense reminiscent of block spin renormalization. [242] This fits in harmoniously with the alleged the metric-independence of the topological gauge theory.

5. The case $d = 3$ is of special interest, and was the main focus of the original Dijkgraaf-Witten paper. In this case we have constructed a "lattice Chern-Simons invariant," and the theory with a cocycle $[W] \in H^3(BG, U(1)) = H^3_{\text{groupcohomology}}(G, U(1))$ is a Chern-Simons theory for gauge group G. In the case of G finite one can show that $H^3(BG, U(1)) \cong H^4(BG; \mathbb{Z})$. In general the level of a Chern-Simons theory is valued in $H^4(BG; \mathbb{Z})$ for all compact Lie groups G.

ALSO DISCUSS HAMILTONIAN VIEWPOINT! Check out D. Harlow and H. Ooguri, Appendix F of https://arxiv.org/pdf/1810.05338.pdf

19. Example: Symmetry Protected Phases Of Matter In $1+1$ Dimensions

[242] The idea of block spin renormalization, invented by Leo Kadanoff, is that we impose some small lattice spacing a as a UV cutoff and try to describe an effective theory at ever larger distances. So, we block spins together in some way, define an effective spin, and then an effective action

$$e^{-S_{eff}} := \sum_{fixed-effective-spins} e^{-S(spins)} \tag{18.70}$$

The hope is that at long distances, with ever larger blocks, the "relevant" parts of S_{eff} converge to a useful infrared field theory description.

Preprint typeset in JHEP style - HYPER VERSION

Chapter 2: Linear Algebra User's Manual

Gregory W. Moore

Abstract: An overview of some of the finer points of linear algebra usually omitted in physics courses. May 3, 2021

-TOC- **Contents**

1. Introduction **5**

2. Basic Definitions Of Algebraic Structures: Rings, Fields, Modules, Vector Spaces, And Algebras **6**
2.1 Rings 6
2.2 Fields 7
2.2.1 Finite Fields 8
2.3 Modules 8
2.4 Vector Spaces 9
2.5 Algebras 10

3. Linear Transformations **14**

4. Basis And Dimension **16**
4.1 Linear Independence 16
4.2 Free Modules 16
4.3 Vector Spaces 17
4.4 Linear Operators And Matrices 20
4.5 Determinant And Trace 23

5. New Vector Spaces from Old Ones **24**
5.1 Direct sum 24
5.2 Quotient Space 28
5.3 Tensor Product 30
5.4 Dual Space 34

6. Tensor spaces **38**
6.1 Totally Symmetric And Antisymmetric Tensors 39
6.2 Algebraic structures associated with tensors 44
6.2.1 An Approach To Noncommutative Geometry 47

7. Kernel, Image, and Cokernel **47**
7.1 The index of a linear operator 50

8. A Taste of Homological Algebra **51**
8.1 The Euler-Poincaré principle 54
8.2 Chain maps and chain homotopies 55
8.3 Exact sequences of complexes 56
8.4 Left- and right-exactness 56

9. Relations Between Real, Complex, And Quaternionic Vector Spaces **59**
9.1 Complex structure on a real vector space 59
9.2 Real Structure On A Complex Vector Space 64
9.2.1 Complex Conjugate Of A Complex Vector Space 66
9.2.2 Complexification 67
9.3 The Quaternions 69
9.4 Quaternionic Structure On A Real Vector Space 79
9.5 Quaternionic Structure On Complex Vector Space 79
9.5.1 Complex Structure On Quaternionic Vector Space 81
9.5.2 Summary 81
9.6 Spaces Of Real, Complex, Quaternionic Structures 81

10. Some Canonical Forms For a Matrix Under Conjugation **85**
10.1 What is a canonical form? 85
10.2 Rank 86
10.3 Eigenvalues and Eigenvectors 87
10.4 Jordan Canonical Form 89
10.4.1 Proof of the Jordan canonical form theorem 94
10.5 The stabilizer of a Jordan canonical form 96
10.5.1 Simultaneous diagonalization 98

11. Sesquilinear forms and (anti)-Hermitian forms **100**

12. Inner product spaces, normed linear spaces, and bounded operators **101**
12.1 Inner product spaces 101
12.2 Normed linear spaces 103
12.3 Bounded linear operators 104
12.4 Constructions with inner product spaces 105

13. Hilbert space **106**

14. Banach space **109**

15. Projection operators and orthogonal decomposition **110**

16. Unitary, Hermitian, and normal operators **113**

17. The spectral theorem: Finite Dimensions **116**
17.1 Normal and Unitary matrices 118
17.2 Singular value decomposition and Schmidt decomposition 118
17.2.1 Bidiagonalization 118
17.2.2 Application: The Cabbibo-Kobayashi-Maskawa matrix, or, how bidiagonalization can win you the Nobel Prize 119
17.2.3 Singular value decomposition 121
17.2.4 Schmidt decomposition 122

18. Operators on Hilbert space **123**
18.1 Lies my teacher told me 123
18.1.1 Lie 1: The trace is cyclic: 123
18.1.2 Lie 2: Hermitian operators have real eigenvalues 123
18.1.3 Lie 3: Hermitian operators exponentiate to form one-parameter groups of unitary operators 124
18.2 Hellinger-Toeplitz theorem 124
18.3 Spectrum and resolvent 126
18.4 Spectral theorem for bounded self-adjoint operators 131
18.5 Defining the adjoint of an unbounded operator 136
18.6 Spectral Theorem for unbounded self-adjoint operators 138
18.7 Commuting self-adjoint operators 139
18.8 Stone's theorem 140
18.9 Traceclass operators 141

19. The Dirac-von Neumann axioms of quantum mechanics **144**

20. Canonical Forms of Antisymmetric, Symmetric, and Orthogonal matrices **151**
20.1 Pairings and bilinear forms 151
20.1.1 Perfect pairings 151
20.1.2 Vector spaces 152
20.1.3 Choosing a basis 153
20.2 Canonical forms for symmetric matrices 153
20.3 Orthogonal matrices: The real spectral theorem 156
20.4 Canonical forms for antisymmetric matrices 157
20.5 Automorphism Groups of Bilinear and Sesquilinear Forms 158

21. Other canonical forms: Upper triangular, polar, reduced echelon **160**
21.1 General upper triangular decomposition 160
21.2 Gram-Schmidt procedure 160
21.2.1 Orthogonal polynomials 161
21.3 Polar decomposition 163
21.4 Reduced Echelon form 165

22. Families of Matrices **165**
22.1 Families of projection operators: The theory of vector bundles 165
22.2 Codimension of the space of coinciding eigenvalues 169
22.2.1 Families of complex matrices: Codimension of coinciding characteristic values 170
22.2.2 Orbits 171
22.2.3 Local model near $\mathcal{S}_{\text{sing}}$ 172
22.2.4 Families of Hermitian operators 173
22.3 Canonical form of a family in a first order neighborhood 175

22.4 Families of operators and spectral covers 176
22.5 Families of matrices and differential equations 181
22.5.1 The WKB expansion 183
22.5.2 Monodromy representation and Hilbert's 21st problem 186
22.5.3 Stokes' phenomenon 186

23. $\mathbb{Z}_2$-graded, or super-, linear algebra 191
23.1 Super vector spaces 191
23.2 Linear transformations between supervector spaces 195
23.3 Superalgebras 197
23.4 Modules over superalgebras 203
23.5 Free modules and the super-General Linear Group 207
23.6 The Supertrace 209
23.7 The Berezinian of a linear transformation 210
23.8 Bilinear forms 214
23.9 Star-structures and super-Hilbert spaces 215
23.9.1 SuperUnitary Group 218
23.10 Functions on superspace and supermanifolds 219
23.10.1 Philosophical background 219
23.10.2 The model superspace $\mathcal{R}^{p|q}$ 221
23.10.3 Superdomains 222
23.10.4 A few words about sheaves 223
23.10.5 Definition of supermanifolds 225
23.10.6 Supervector fields and super-differential forms 228
23.11 Integration over a superdomain 231
23.12 Gaussian Integrals 235
23.12.1 Reminder on bosonic Gaussian integrals 235
23.12.2 Gaussian integral on a fermionic point: Pfaffians 235
23.12.3 Gaussian integral on $\mathcal{R}^{p|q}$ 240
23.12.4 Supersymmetric Cancelations 241
23.13 References 242

24. Determinant Lines, Pfaffian Lines, Berezinian Lines, and anomalies 243
24.1 The determinant and determinant line of a linear operator in finite dimensions 243
24.2 Determinant line of a vector space and of a complex 245
24.3 Abstract defining properties of determinants 247
24.4 Pfaffian Line 247
24.5 Determinants and determinant lines in infinite dimensions 249
24.5.1 Determinants 249
24.5.2 Fredholom Operators 250
24.5.3 The determinant line for a family of Fredholm operators 251
24.5.4 The Quillen norm 252
24.5.5 References 253

24.6 Berezinian of a free module 253
24.7 Brief Comments on fermionic path integrals and anomalies 254
24.7.1 General Considerations 254
24.7.2 Determinant of the one-dimensional Dirac operator 255
24.7.3 A supersymmetric quantum mechanics 257
24.7.4 Real Fermions in one dimension coupled to an orthogonal gauge field 258
24.7.5 The global anomaly when M is not spin 259
24.7.6 References 260

25. Quadratic Forms And Lattices 260
25.1 Definition 261
25.2 Embedded Lattices 263
25.3 Some Invariants of Lattices 268
25.3.1 The characteristic vector 274
25.3.2 The Gauss-Milgram relation 274
25.4 Self-dual lattices 276
25.4.1 Some classification results 279
25.5 Embeddings of lattices: The Nikulin theorem 283
25.6 References 283

26. Positive definite Quadratic forms 283

27. Quivers and their representations 283

---

1. Introduction

Linear algebra is of course very important in many areas of physics. Among them:

1. Tensor analysis - used in classical mechanics and general relativity.
2. The very formulation of quantum mechanics is based on linear algebra: The states in a physical system are described by "rays" in a projective Hilbert space, and physical observables are identified with Hermitian linear operators on Hilbert space.
3. The realization of symmetry in quantum mechanics is through representation theory of groups which relies heavily on linear algebra.

For this reason linear algebra is often taught in physics courses. The problem is that it is often mis-taught. Therefore we are going to make a quick review of basic notions stressing some points not usually emphasized in physics courses.

We also want to review the basic canonical forms into which various types matrices can be put. These are very useful when discussing various aspects of matrix groups.

For more information useful references are Herstein, Jacobsen, Lang, Eisenbud, *Commutative Algebra* Springer GTM 150, Atiyah and MacDonald, *Introduction to Commutative Algebra*. For an excellent terse summary of homological algebra consult S.I. Gelfand and Yu. I. Manin, *Homological Algebra*.

We will only touch briefly on some aspects of functional analysis - which is crucial to quantum mechanics. The standard reference for physicists is:

Reed and Simon, *Methods of Modern Mathematical Physics*, especially, vol. I.

2. Basic Definitions Of Algebraic Structures: Rings, Fields, Modules, Vector Spaces, And Algebras

2.1 Rings

In the previous chapter we talked about groups. We now overlay some extra structure on an abelian group R, with operation $+$ and identity 0, to define what is called a *ring*. The new structure is a *second* binary operation $(a, b) \to a \cdot b \in R$ on elements $a, b \in R$. We demand that this operation be associative, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, and that it is compatible with the pre-existing additive group law. To be precise, the two operations $+$ and $\cdot$ are compatible in the sense that there is a distributive law:

$$a \cdot (b + c) = a \cdot b + a \cdot c \tag{2.1}$$

$$(a + b) \cdot c = a \cdot c + b \cdot c \tag{2.2}$$

Remarks

1. A ring with a multiplicative unit 1_R such that $a \cdot 1_R = 1_R \cdot a = a$ is called a *unital ring* or a *ring with unit*. One needs to be careful about this because many authors will simply assume that "ring" means a "ring with unit." [1]

2. If $a \cdot b = b \cdot a$ then R is a *commutative ring*.

3. If R is any ring we can then form another ring, $M_n(R)$, the ring of $n \times n$ matrices with matrix elements in R. Even if R is a commutative ring, the ring $M_n(R)$ will be noncommutative in general if $n > 1$.

Example 1: A good example of a ring is $R = \mathbb{Z}$ with $+$ and $\cdot$ being the usual notions of addition and multiplication. Note that R is just a monoid, not a group with respect to $\cdot$.

Example 2: Another good example is $R = \mathbb{Z}/n\mathbb{Z}$ again with $+$ and $\cdot$ inherited from the usual addition and multiplication on $\mathbb{Z}$. As we have discussed many times, $\mathbb{Z}/n\mathbb{Z}$ as an Abelian group with $+$ is isomorphic to the group of nth roots of unity where the Abelian

[1] Some authors use the term "rng" – pronounced "rung" - for a ring possibly without a unit. We will not do that. Similarly, one can define a notion called a "rig" - which is a ring without negatives. That is, it is an abelian monoid with the operation $+$ and a compatible multiplication $\cdot$.

group law is multiplication of complex numbers. Note that the ring structure is not so natural in the latter picture. If we considered the nth roots of unity as isomorphic to $\mathbb{Z}/n\mathbb{Z}$ as a ring the multiplication law would be:

$$e^{2\pi \mathrm{i}\frac{k_1}{n}} \cdot e^{2\pi \mathrm{i}\frac{k_2}{n}} = e^{2\pi \mathrm{i}\frac{k_1 k_2}{n}} \tag{2.3}$$

It is well-defined and perfectly sensible. But it is not ordinary multiplication of complex numbers!

Example 3: Let $U \subset \mathbb{C}$ be an open set in the complex plane and consider $\mathcal{O}(U)$, the set of all holomorphic functions on U. This is a ring with the obvious addition and multiplication of holomorphic functions. Note that we will not have inverses for the multiplication law because some holomorphic functions on U will have zeroes in U.

2.2 Fields

Definition: A commutative ring R such that $R^* = R - \{0\}$ is also an abelian group with respect to $\cdot$ is called a *field*.

Two examples of fields which we have used again and again are $\mathbb{R}$ and $\mathbb{C}$.

Some important examples of rings which are not fields are

1. $\mathbb{Z}$.
2. $\mathbb{Z}/N\mathbb{Z}$, when N is not prime.
3. If R is any ring then we can form the ring of polynomials with coefficients in R, denoted $R[x]$. Iterating this we obtain polynomial rings in several variables $R[x_1, \ldots, x_n]$. Similarly, we can consider a ring of power series in x.
4. Let U be an open subset of the complex plane (or of $\mathbb{C}^n$) then we can consider the ring $\mathcal{O}(U)$ of holomorphic functions on U.

Some important examples of fields closely related to the above examples:

1. $\mathbb{Q}$
2. $\mathbb{Z}/N\mathbb{Z}$, when $N = p$ is prime.
3. For the ring of polynomials $R[x]$ we can consider the associated field of fractions $p(x)/q(x)$ where $q(x)$ is nonzero. This is an example of "localization." ♣Check. R has to be a PID? ♣
4. Let U be an open subset of the complex plane (or of $\mathbb{C}^n$) then we can consider the field $\mathcal{M}(U)$ of meromorphic functions on U.

2.2.1 Finite Fields

A beautiful theorem in algebra (see, e.g. the book by Jacobsen) states that the finite fields must have order $q = p^k$ which is a prime power. Moreover, up to isomorphism, the field is unique and it is variously denoted as $\mathbb{F}_q$ or $GF(q)$. For $k = 1$, i.e. $q = p$ we can identify $\mathbb{F}_p$ with $\mathbb{Z}/p\mathbb{Z}$.

For $k > 1$ the field $\mathbb{F}_{p^k}$ is *not* to be confused with the ring $\mathbb{Z}/p^k\mathbb{Z}$. For example $R = \mathbb{Z}/4\mathbb{Z}$ is not a field with the usual ring multiplication. For example $2 \in R^*$ and $2 \cdot 2 = 0\text{mod}4$. One way to represent the field $\mathbb{F}_4$ is as a set

$$\mathbb{F}_4 = \{0, 1, \omega, \bar{\omega}\} \tag{2.4}$$

with the relations

$$\begin{aligned}
0 + x &= x \qquad \forall x \in \mathbb{F}_4 \\
x + x &= 0 \qquad \forall x \in \mathbb{F}_4 \\
1 + \omega &= \bar{\omega} \\
1 + \bar{\omega} &= \omega \\
\omega + \bar{\omega} &= 1 \\
0 \cdot x &= 0 \qquad \forall x \in \mathbb{F}_4 \\
1 \cdot x &= x \qquad \forall x \in \mathbb{F}_4 \\
\omega \cdot \bar{\omega} &= 1 \\
\omega \cdot \omega &= \bar{\omega} \\
\bar{\omega} \cdot \bar{\omega} &= \omega
\end{aligned} \tag{2.5}$$

Note that although $\omega^3 = \bar{\omega}^3 = 1$ you cannot identity ω with a complex number the third root of unity: This field is not a subfield of $\mathbb{C}$.

The field $\mathbb{F}_q$ can be identified with an "extension" field of $\mathbb{F}_p$ where the polynomial equation $X^q - X = 0$ has q roots.

1. Finite fields are sometimes used to define special groups with interesting properties. For example, it makes sense to speak of $SL(n, \mathbb{F}_q)$ and these finite groups have very beautiful properties.

2. Finite fields are often used in the theory of classical and quantum error-correcting codes.

2.3 Modules

Definition A *module* over a ring R is a set M with a multiplication $R \times M \to M$ such that for all $r, s \in R$ and $v, w \in M$:

1. M is an abelian group wrt $+$, called "vector addition."

2. $r(v+w) = rv + rw$

3. $r(sv) = (rs)v$

4. $(r+s)v = rv + sv$

Axioms 2,3,4 simply say that all the various operations on R and M are compatible in the natural way.

Remarks:

1. If the ring has a multiplicative unit 1_R then we require $1_R \cdot v = v$.

2. If the ring is noncommutative then one should distinguish between left and right modules. Above we have written the axioms for a left-module. For a right-module we have $(v \cdot r) \cdot s = v \cdot (rs)$.

3. There is an important generalization known as a *bimodule* over two rings R_1, R_2. A bimodule M is simultaneously a left R_1-module and a right R_2-module. A good example is the set of $n \times m$ matrices over a ring R, which is a bimodule over $R_1 = M_n(R)$ and $R_2 = M_m(R)$.

4. Any ring is a bimodule over itself. For a positive integer n we define the module R^n of n-tuples of elements of R by componentwise addition and multiplication in the obvious way. This is an R-bimodule and also an $M_n(R)$ bimodule.

5. In quantum field theory if we divide up a spatial domain into two parts with a codimension one subspace then the states localized near the division is a bimodule for operators localized on the left and the right of the partition.

Examples:

1. Any Abelian group is a $\mathbb{Z}$ module, and any $\mathbb{Z}$-module is just an Abelian group.

2. Meromorphic functions with a pole at some point z_0 in the complex plane with order $\leq n$. These form an important example of a module over the ring of holomorphic functions.

2.4 Vector Spaces

Recall that a *field* is a commutative ring such that $R^* = R - \{0\}$ is also an abelian group.

Let κ be a field. Then, by definition a *vector space over* κ is simply a κ-module. Written out in full, this means:

♣In the rest of the notes we need to change notation for a general field from k to κ since k is often also used as an integer or a momentum. ♣

Definition . V is a vector space over a field κ if for every $\alpha \in \kappa, v \in V$ there is an element $\alpha v \in V$ such that

1. V is an abelian group under $+$

2. $\alpha(v+w) = \alpha v + \alpha w$

3. $\alpha(\beta v) = (\alpha\beta)v$

4. $(\alpha+\beta)v = \alpha v + \beta v$

5. $1v = v$

for all $\alpha, \beta \in \kappa, v, w \in V$

For us, the field κ will almost always be $\kappa = \mathbb{R}$ or $\kappa = \mathbb{C}$. In addition to the well-worn examples of $\mathbb{R}^n$ and $\mathbb{C}^n$ two other examples are

1. Recall example 2.9 of Chapter 1: If X is any set then the power set $\mathcal{P}(X)$ is an Abelian group with $Y_1 + Y_2 := (Y_1 - Y_2) \cup (Y_2 - Y_1)$. As we noted there, $2Y = \emptyset$. So, $\mathcal{P}(X)$ is actually a vector space over the field $\mathbb{F}_2$.

2.5 Algebras

♣Take material on ideals below and move it here. ♣

So far we have taken abelian groups and added binary operations $R \times R \to R$ to define a ring and $R \times M \to M$ to define a module. It remains to consider the case $M \times M \to M$. In this case, the module is known as an *algebra*.

Everything that follows can also be defined for modules over a ring but we will state the definitions for a vector space over a field.

♣Rewrite, and do everything for modules over a ring?? ♣

Definition An *algebra* over a field κ is a vector space A over κ with a notion of multiplication of two vectors

$$A \times A \to A \tag{2.6}$$

denoted:

$$a_1, a_2 \in A \to a_1 \odot a_2 \in A \tag{2.7}$$

which has a ring structure compatible with the scalar multiplication by the field. Concretely, this means we have axioms:

i.) $(a_1 + a_2) \odot a_3 = a_1 \odot a_3 + a_2 \odot a_3$

ii.) $a_1 \odot (a_2 + a_3) = a_1 \odot a_2 + a_1 \odot a_3$

iii.) $\alpha(a_1 \odot a_2) = (\alpha a_1) \odot a_2 = a_1 \odot (\alpha a_2), \qquad \forall \alpha \in \kappa.$

The algebra is *unital*, i.e., it has a unit, if $\exists 1_A \in A$ (*not* to be confused with the multiplicative unit $1 \in \kappa$ of the ground field) such that:

iv.) $1_A \odot a = a \odot 1_A = a$

In the case of rings we assumed associativity of the product. It turns out that this is too restrictive when working with algebras. If, in addition, the product of vectors satisfies:

$$(a_1 \odot a_2) \odot a_3 = a_1 \odot (a_2 \odot a_3) \tag{2.8}$$

for all $a_1, a_2, a_3 \in A$ then A is called an *associative algebra*.

Remark: We have used the heavy notation $\odot$ to denote the product of vectors in an algebra to stress that it is a new structure imposed on a vector space. But when working with algebras people will generally just write $a_1 a_2$ for the product. One should be careful here as it can (and will) happen that a given vector space can admit more than one interesting algebra product structure.

Example 1 $M_n(\kappa)$ is a vector space over κ of dimension n^2. It is also an associative algebra because matrix multiplication defines an algebraic structure of multiplication of the "vectors" in $M_n(\kappa)$.

Example 2 More generally, if A is a vector space over κ then $\mathrm{End}(A)$ is an associative algebra. (See next section for the definition of this notation.)

In general, a *nonassociative algebra* means a not-necessarily associative algebra. In any algebra we can introduce the *associator*

$$[a_1, a_2, a_3] := (a_1 \cdot a_2) \cdot a_3 - a_1 \cdot (a_2 \cdot a_3) \tag{2.9}$$ eq:associator

Note that it is trilinear. There are important examples of non-associative algebras such as Lie algebras and the octonions.

Definition A *Lie algebra* over a field κ is an algebra A over κ where the multiplication of vectors $a_1, a_2 \in A$, satisfies in addition the two conditions:

1. $\forall a_1, a_2 \in A$:
$$a_2 \odot a_1 = -a_1 \odot a_2 \tag{2.10}$$ eq:LieDef-1
2. $\forall a_1, a_2, a_3 \in A$:
$$((a_1 \odot a_2) \odot a_3) + ((a_3 \odot a_1) \odot a_2) + ((a_2 \odot a_3) \odot a_1) = 0 \tag{2.11}$$ eq:LieDef-2

This is known as the *Jacobi relation.*

Now, tradition demands that the product on a Lie algebra be denoted not as $a_1 \odot a_2$ but rather as $[a_1, a_2]$ where it is usually referred to as the *bracket.* So then the two defining conditions (2.10) and (2.11) are written as:

1. $\forall a_1, a_2 \in A$:
$$[a_2, a_1] = -[a_1, a_2] \tag{2.12}$$
2. $\forall a_1, a_2, a_3 \in A$:
$$[[a_1, a_2], a_3] + [[a_3, a_1], a_2] + [[a_2, a_3], a_1] = 0 \tag{2.13}$$

Remarks:

1. Note that we call $[a_1, a_2]$ the bracket and not the commutator. It might well not be possible to write $[a_1, a_2] = a_1 \odot a_2 - a_2 \odot a_1$ where $\odot$ is some other multiplication structure defined within A. Rather $[\cdot, \cdot] : A \times A \to A$ is just an abstract product satisfying the two rules (2.10) and (2.11). Let us give two examples to illustrate the point:

   - Note that the vector space $A \subset M_n(\kappa)$ of anti-symmetric matrices is not closed under normal matrix multiplication: If a_1 and a_2 are antisymmetric matrices then $(a_1 a_2)^{tr} = a_2^{tr} a_1^{tr} = (-a_2)(-a_1) = a_2 a_1$ and in general this is not $-a_1 a_2$. So, it is not an algebra under normal matrix multiplication. But if we define the bracket using normal matrix multiplication

   $$[a_1, a_2] = a_1 a_2 - a_2 a_1 \tag{2.14}$$

   where on the RHS $a_1 a_2$ means matrix multiplication. Since $[a_1, a_2]$ is an anti-symmetric matrix the product is closed within A. The Jacobi relation is then inherited from the associativity of matrix multiplication. This Lie algebra is sometimes denoted $\mathfrak{o}(n, \kappa)$. It is the Lie algebra of an orthogonal group.

   - Consider first order differential operators of $\mathcal{C}^\infty$ functions on the line. These will be written as $D = f(x)\frac{d}{dx} + g(x)$ for smooth functions $f(x), g(x)$. Then the ordinary composition of two differential operators $D_1 \circ D_2$ is a second order differential operator. Nevertheless if we take the difference:

   $$\begin{aligned}[D_1, D_2] &= (f_1(x)\frac{d}{dx} + g_1(x))(f_2(x)\frac{d}{dx} + g_2(x)) - (f_2(x)\frac{d}{dx} + g_2(x))(f_1(x)\frac{d}{dx} + g_1(x)) \\ &= \left(f_1 f_2' - f_2 f_1'\right)(x)\frac{d}{dx} + (f_1 g_2' - f_2 g_1')(x)\end{aligned} \tag{2.15}$$

   we get a first order differential operator. It is obviously anti-symmetric and one can check the Jacobi relation.

   - In both these examples we embed the Lie algebra into a larger associative algebra where the bracket can be written as $[a_1, a_2] = a_1 \odot a_2 - a_2 \odot a_1$ with $\odot$ the algebra product that only closes within the larger algebra.

2. A Lie algebra is in general a nonassociative algebra. Indeed, using the Jacobi relation we can compute the associator as:

   $$[a_1, a_2, a_3] = [[a_1, a_2], a_3] - [a_1, [a_2, a_3]] = [[a_1, a_2], a_3] + [[a_2, a_3], a_1] = -[[a_3, a_1], a_2] \tag{2.16}$$

   and the RHS is, in general, nonzero.

3. Note that the vector space of $n \times n$ matrices over κ, that is, $M_n(\kappa)$ has two interesting algebra structures: One is matrix multiplication. It is associative. The other is a

Lie algebra structure where the bracket is defined by the usual commutator. It is nonassociative. It is sometimes denoted $\mathfrak{gl}(n,\kappa)$, and such a notation would definitely imply a Lie algebra structure.

Exercise *Opposite Algebra*

If A is an algebra we can always define another algebra $A^{\rm opp}$ with the product

$$a_1 \odot^{\rm opp} a_2 := a_2 \odot a_1 \tag{2.17}$$

a.) Show that $\odot^{\rm opp}$ indeed defines the structure of an algebra on the set A.

b.) Consider the algebra $M_n(\kappa)$ where κ is a field. Is it isomorphic to its opposite algebra?

c.) Give an example of an algebra not isomorphic to its opposite algebra. ♣Need to provide an answer here. ♣

Exercise *Structure constants*

In general, if $\{v_i\}$ is a basis for the algebra then the *structure constants* are defined by

$$v_i \cdot v_j = \sum_k c_{ij}^k v_k \tag{2.18}$$

a.) Write out a basis and structure constants for the algebra $M_n(k)$.

Exercise

a.) If A is an algebra, then it is a module over itself, via the left-regular representation (LRR). $a \to L(a)$ where

$$L(a) \cdot b := ab \tag{2.19}$$

Show that if we choose a basis a_i then the structure constants

$$a_i a_j = c_{ij}{}^k a_k \tag{2.20}$$

define the matrix elements of the LRR:

$$\left(L(a_i)\right)_j{}^k = c_{ij}^k \tag{2.21}$$

An algebra is said to be *semisimple* if these operators are diagonalizable.

b.) If A is an algebra, then it is a bimodule over $A \otimes A^o$ where A^o is the opposite algebra.

---

3. Linear Transformations

Definition

a.) A *linear transformation* or *linear operator* between two R modules is a map $T : M_1 \to M_2$ which is a group homomorphism with respect to $+$:

$$T(m + m') = T(m) + T(m') \tag{3.1}$$

and moreover such that $T(r \cdot m) = r \cdot T(m)$ for all $r \in R$, $m \in M_1$.

b.) T is an *isomorphism* if it is one-one and onto.

c.) The set of all linear transformations $T : M_1 \to M_2$ is denoted $\mathrm{Hom}(M_1, M_2)$, or $\mathrm{Hom}_R(M_1, M_2)$ when we wish to emphasize the underlying ring R.

There are some algebraic structures on spaces of linear transformations we should immediately take note of:

1. $\mathrm{Hom}_R(M_1, M_2)$ is an abelian group where the group operation is addition of linear operators: $T_1 + T_2$.

2. Moreover $\mathrm{Hom}_R(M_1, M_2)$ is an R-module provided that R is a commutative ring.

3. In particular, if V_1, V_2 are vector spaces over a field k then $\mathrm{Hom}(V_1, V_2)$ is itself a vector space over k.

4. If M is a module over a ring R then sometimes the notation

$$\mathrm{End}_R(M) := \mathrm{Hom}_R(M, M) \tag{3.2}$$

is used. In this case composition of linear transformations $T_1 \circ T_2$ defines a binary operation on $\mathrm{End}_R(M)$, and if R is commutative this is itself a ring because

$$T_1 \circ (T_2 + T_3) = T_1 \circ T_2 + T_1 \circ T_3 \tag{3.3}$$

and so forth.

5. In general if M is a module over a commutative ring R then $\mathrm{End}_R(M)$ is not a group wrt $\circ$, since inverses don't always exist. However we may define:

Definition The set of *invertible* linear transformations of M, denoted $GL(M, R)$, is a group. If we have a vector space over a field k we generally write $GL(V)$.

Example: For $R = \mathbb{Z}$ and $M = \mathbb{Z} \oplus \mathbb{Z}$, the group of invertible transformations is isomorphic to $GL(2, \mathbb{Z})$. [2]

[2] Warning: This is NOT the same as 2×2 matrices over $\mathbb{Z}$ with nonzero determinant!

A *representation* of an algebra A is a vector space V and a morphism of algebras $T : A \to \mathrm{End}(V)$. This means that

$$\begin{aligned} T(\alpha_1 a_1 + \alpha_2 a_2) &= \alpha_1 T(a_1) + \alpha_2 T(a_2) \\ T(a_1 \odot a_2) &= T(a_1) \odot T(a_2) \end{aligned} \tag{3.4}$$

eq:repalgebra

Remarks

1. We must be careful here about the algebra product $\odot$ being used since, as noted above, there are two interesting algebra structures on $\mathrm{End}(V)$ given by composition and by commutator. If we speak of a morphism of algebras what is usually meant by $\odot$ on the RHS of (3.4) is composition of linear transformations. However, if we are speaking of a representation of Lie algebras then we mean the commutator. So, for a Lie algebra a representation would satisfy

$$T([a_1, a_2]) = T(a_1) \circ T(a_2) - T(a_2) \circ T(a_1) \tag{3.5}$$

2. If we consider the algebra $M_n(\kappa)$ with matrix multiplication as the algebra product then a theorem states that the general representation is a direct sum (See Section **** below) of the fundamental, or defining representation $V_{fund} = \kappa^{\oplus n}$. That is, the general representation is

$$V_{fund} \oplus \cdots \oplus V_{fund} \tag{3.6}$$

   If we have m summands then $T(a)$ would be a block diagonal matrix with a on the diagonal m times. This leads to a concept called "Morita equivalence" of algebras: Technically, two algebras A, B are "Morita equivalent" if their categories of representations are equivalent categories. In practical terms often it just means that $A = M_n(B)$ or vice versa. [3]

3. On the other hand, if we consider $M_n(\kappa)$ as a Lie algebra then the representation theory is much richer, and will be discussed in Chapter **** below.

Exercise

Let R be any ring. Show that if $\mathcal{M}$ is an $R[x]$-module then we can associate to it an R-module M together with a linear transformation $T : M \to M$.

b.) Conversely, show that if we are given an R-module M together with a linear transformation T then we can construct uniquely an $R[x]$ module $\mathcal{M}$.

Thus, $R[x]$-modules are in one-one correspondence with pair (M, T) where M is an R-module and $T \in \mathrm{End}_R(M)$.

---

[3] For proofs of these statements see, for example Drozd and Kirichenko, *Finite Dimensional Algebras*, Springer, or Appendix A of my lecture notes at http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf

4. Basis And Dimension

4.1 Linear Independence

Definition . Let M be a module over a ring R.

1. If $S \subset M$ is any subset of M the *linear span of* S is the set of finite linear combinations of vectors drawn from S:

$$L(S) := Span(S) := \{\sum r_i v_i : r_i \in R, v_i \in S\} \tag{4.1}$$ eq:linespan

$L(S)$ is the smallest submodule of M containing S. We also call S a *generating set* of $L(S)$.

2. A set of vectors $S \subset M$ is said to be linearly independent if for any finite sum of vectors in $L(S)$:

$$\sum_s \alpha_s v_s = 0 \qquad \Rightarrow \qquad \alpha_s = 0 \tag{4.2}$$

3. A linearly independent generating set S for a module M is called a *basis for* M. We will often denote a basis by a symbol like $\mathcal{B}$.

Remarks:

1. A basis $\mathcal{B}$ need not be a finite set. However, all sums above are finite sums. In particular, when we say that $\mathcal{B}$ generates V this means that every vector $m \in M$ can be written (uniquely) as a finite linear combination of vectors in $\mathcal{B}$.

2. To appreciate the need for restriction to a finite sums in the definitions above consider the vector space $\mathbb{R}^\infty$ of infinite tuples of real numbers $(x_1, x_2, \dots)$. (Equivalently, the vector space of all functions $f : \mathbb{Z}_+ \to \mathbb{R}$.) Infinite sums like

$$(1,1,1,\dots) - (2,2,2,\dots) + (3,3,3,\dots) - (4,4,4,\dots) + \cdots \tag{4.3}$$

   are clearly ill-defined.

3. For a finite set $S = \{v_1, \dots, v_n\}$ we will also write

$$L(S) := \langle v_1, \dots, v_n \rangle \tag{4.4}$$

4.2 Free Modules

A module is called a *free module* if it has a basis. If the basis is finite the free module is isomorphic to R^n for some positive integer n.

Not all modules are free modules, e.g.

1. $\mathbb{Z}/n\mathbb{Z}$ is not a free $\mathbb{Z}$-module. Exercise: Explain why.

2. Fix a set of points $\{z_1, \ldots, z_k\}$ in the complex plane and a set of integers $n_i \in \mathbb{Z}$ associated with those points. The set of holomorphic on $\mathbb{C} - \{z_1, \ldots, z_k\}$ which have convergent Laurent expansions of the form

$$f(z) = \frac{a^i_{-n_i}}{(z - z_i)^{n_i}} + \frac{a^i_{-(n_i-1)}}{(z - z_i)^{n_i-1}} + \cdots \tag{4.5}$$

in the neighborhood of $z = z_i$, for all $i = 1, \ldots, k$ is a module over the ring of holomorphic functions, but it is not a free module.

isVectorSpaces

4.3 Vector Spaces

One big simplification when working with vector spaces rather than modules is that they are always free modules. We should stress that this is not obvious! The statement is false for general modules over a ring, as we have seen above, and the proof requires the use of Zorn's lemma (which is equivalent to the axiom of choice).

Theorem 4.3.1:

a.) Every nonzero vector space V has a basis.

b.) Given any linearly independent set of vectors $S \subset V$ there is a basis $\mathcal{B}$ for V with $S \subset \mathcal{B}$.

Proof: Consider the collection $\mathcal{L}$ of linearly independent subsets $S \subset V$. If $V \neq 0$ then this collection is nonempty. Moreover, for every ascending chain of elements in $\mathcal{L}$:

$$S_1 \subset S_2 \subset \cdots \tag{4.6}$$

the union $\cup_i S_i$ is a set of linearly independent vectors and is hence in $\mathcal{L}$. We can then invoke Zorn's lemma to assert that there exists a maximal element $\mathcal{B} \subset \mathcal{L}$. That is, it is a linearly independent set of vectors not properly contained in any other element of $\mathcal{L}$.

We claim that $\mathcal{B}$ is a basis. To see this, consider the linear span $L(\mathcal{B}) \subset V$. If $L(\mathcal{B})$ is a proper subset of V there is a vector $v_* \in V - L(\mathcal{B})$. But then we claim that $\mathcal{B} \cup \{v_*\}$ is a linearly independent set of vectors. The reason is that if

$$\alpha_* v_* + \sum_{w \in \mathcal{B}} \beta_w w = 0 \tag{4.7}$$

(remember: all but finitely many $\beta_w = 0$ here) then if $\alpha_* = 0$ we must have $\beta_w = 0$ because $\mathcal{B}$ is a linearly independent set. But if $\alpha_* \neq 0$ then we can divide by it. (It is exactly at this point that we use the fact that we are working with a vector space over a field κ rather than a general modular over a ring R!!) Then we would have

$$v_* = -\sum_{w \in \mathcal{B}} \frac{\beta_w}{\alpha_*} w \tag{4.8}$$

but this contradicts the hypothesis that $v_* \notin L(\mathcal{B})$. Thus we conclude that $L(\mathcal{B}) = V$ and hence $\mathcal{B}$ is a basis.

To prove part (b) apply Zorn's lemma to the set of linearly independent sets containing a fixed linearly independent set S. ♠

Theorem 4.3.2: Let V be a vector space over a field κ. Then any two bases for V have the same cardinality.
Proof: See Lang, *Algebra*, ch. 3 Sec. 5. Again the proof explicitly uses the fact that you can divide by nonzero scalars.

By this theorem we know that if V has a finite basis $\{v_1, \ldots, v_n\}$ then any other basis has n elements. (The basic idea is to observe that for any linearly independent set of m elements we must have $m \leq n$, so two bases must have the same cardinality.)

We call this basis-invariant integer n the *dimension of* V:

$$n := \dim_\kappa V \tag{4.9}$$

the dimension of V. If there is no finite basis then V is infinite-dimensional.

Remarks

1. Note well that the notion of dimension refers to the ground field. If $\kappa_1 \subset \kappa_2$ then the notion of dimension over κ_1 and κ_2 will be different. For example, any vector space over $\kappa = \mathbb{C}$ is, *a fortiori* also a vector space over $\kappa = \mathbb{R}$. Let us call it $V_{\mathbb{R}}$. It is the same set, but now the vector space structure on this Abelian group is just defined by the action of real scalars. Then we will see that:

$$\dim_{\mathbb{R}} V = 2\dim_{\mathbb{C}} V \tag{4.10}$$

   We will come back to this important point in Section 9.

2. Any two finite dimensional vector spaces of the same dimension are isomorphic. However, it is in general not true that two infinite-dimensional vector spaces are isomorphic. However the above theorem states that if they have bases $\{v_i\}_{i\in I}$ and $\{w_\alpha\}_{\alpha\in I'}$ with a one-one map $I \to I'$ then they are isomorphic.

3. The only invariant of a finite dimensional vector space is its dimension. One way to say this is the following: Let **VECT** be the category of finite-dimensional vector spaces and linear transformations. Define another category **vect** whose objects are the nonnegative integers $n = 0, 1, 2, \ldots$ and whose morphisms $\hom(n, m)$ are $m \times n$ matrices, with composition of morphisms given by matrix multiplication. (If n or m is zero there is a unique morphism with the properties of a zero matrix.) We claim that **VECT** and **vect** are equivalent categories. It is a good exercise to prove this.

4. Something which can be stated or proved without reference to a particular basis is often referred to as *natural* or *canonical* in mathematics. (We will use these terms interchangeably.) More generally, these terms imply that a mathematical construction

does not make use of any extraneous information. Often in linear algebra, making a choice of basis is just such an extraneous piece of data. One of the cultural differences between physicists and mathematicians is that mathematicians often avoid making choices and strive for naturality. This can be a very good thing as it oftentimes happens that expressing a construction in a basis-dependent fashion obscures the underlying conceptual simplicity. On the other hand, insisting on not using a basis can sometimes lead to obscurity. We will try to strike a balance.

5. One of the many good reasons to insist on natural constructions is that these will work well when we consider continuous families of vectors spaces (that is, when we consider vector bundles). Statements which are basis-dependent will tend not to have analogs for vector bundles, whereas natural constructions easily generalize to vector bundles.

   For those to whom "vector bundle" is a new concept a good, nontrivial, and ubiquitous example is the following: [4] Consider the family of projection operators $P_\pm(\hat{x}) : \mathbb{C}^2 \to \mathbb{C}^2$ labeled by a point $\hat{x}$ in the unit sphere in three dimensions: $\hat{x} \in S^2 \subset \mathbb{R}^3$. We take them to be

$$P_\pm(\hat{x}) = \frac{1}{2}(1 \pm \hat{x} \cdot \vec{\sigma}) \tag{4.11}$$

   The images $L_{\pm,\hat{x}}$ of $P_\pm(\hat{x})$ are one-dimensional subspaces of $\mathbb{C}^2$. So, explicitly:

$$L_{\pm,\hat{x}} := \{P_\pm v | v \in \mathbb{C}^2\} \tag{4.12}$$

eq:SpinLines

   For those who know about spin, if we think of $\mathbb{C}^2$ as a Qbit consisting of a spin-half particle then $L_{\pm,\hat{x}}$ is the line in which the particle spins along $\hat{x}$ (for the + case) and along $-\hat{x}$ (for the − case).

   This is a good example of a "family of vector spaces." More generally, if we have a family of projection operators $P(s)$ acting on some fixed vector space V and depending on some control parameters s valued in some manifold then we have a family of vector spaces

$$E_s = P(s)[V] = \mathrm{Im}(P(s)) \tag{4.13}$$

   parametrized by that manifold. If the family of projection operators depends "continuously" on s, (note that you need a topology on the space of projectors to make mathematical sense of that) then our family of vector spaces is a *vector bundle*. In fact, a theorem in bundle theory states that every vector bundle $\mathcal{E}$ over a manifold M is isomorphic to

$$\mathcal{E} = \{(m, \psi) | \psi \in \mathrm{Im}(P(m))\} \tag{4.14}$$

   where $P(m)$ is a continuously varying family of projection operators on a fixed vector space $\mathbb{C}^N$ for some N, that is a continuous map from M into the space of projection operators on $\mathbb{C}^N$.

[4] Some terms such as "projection operator" are only described below, so the reader might wish to return to this - important! - remark later.

Returning to (4.12), since they are one-dimensional subspaces we can certainly say that, for every $\hat{x} \in S^2$ there are isomorphisms

$$\psi_{\pm}(\hat{x}) : L_{\pm,\hat{x}} \to \mathbb{C} \tag{4.15}$$

However, methods of topology can be used to prove rigorously that there is no continuous family of such isomorphisms. Morally speaking, if there had been a natural family of isomorphisms one would have expected it to be continuous.

6. When V is infinite-dimensional there are different notions of what is meant by a "basis." The notion we have defined above is known as a *Hamel basis*. In a Hamel basis we define "generating" and "linear independence" using finite sums. In a vector space there is no *a priori* notion of infinite sums of vectors, because defining such infinite sums requires a notion of convergence. If V has more structure, for example, if it is a Banach space, or a Hilbert space (see below) then there is a notion of convergence and we can speak of other notions of basis where we allow convergent infinite sums. These include *Schauder basis, Haar basis, ...* In the most important case of a Hilbert space, an *orthonormal basis* is a maximal orthonormal set. Every Hilbert space has an orthonormal basis, and one can write every vector in Hilbert space as an infinite (convergent!) linear combination of orthonormal vectors. See

   K. E. Smith, http://www.math.lsa.umich.edu/∼kesmith/infinite.pdf

   for a nice discussion of the issues involved.

4.4 Linear Operators And Matrices

Let V, W be finite dimensional vector spaces over κ. Given a linear operator $T : V \to W$ and ordered bases $\{v_1, \ldots, v_m\}$ for V and $\{w_1, \ldots w_n\}$ for W, we may associate a matrix $M \in Mat_{n\times m}(k)$ to T:

$$Tv_i = \sum_{s=1}^{n} M_{si} w_s \qquad\qquad i = 1, \ldots, m \tag{4.16}$$

eq:mtrx

Note! A matrix depends on a *choice* of ordered basis. The same linear transformation can look very different in different bases. A particularly interesting example is,

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \sim \begin{pmatrix} x & y \\ z & -x \end{pmatrix} \tag{4.17}$$

whenever $x^2 + yz = 0$ and $(x, y, z) \neq 0$.

In general if we change bases

$$\begin{aligned} \tilde{w}_s &= \sum_{t=1}^{n} (g_2)_{ts} w_t \\ \tilde{v}_j &= \sum_{j=1}^{m} (g_1)_{ji} v_j \end{aligned} \tag{4.18}$$

then with respect to the new bases the same linear transformation is expressed by the new matrix:

$$\tilde{M} = g_2^{-1} M g_1. \tag{4.19}$$

With the choice of indices in (4.16) *composition* of linear transformations corresponds to *matrix multiplication*. If $T_1 : V_1 \to V_2$ and $T_2 : V_2 \to V_3$ and we choose ordered bases

$$\begin{aligned} &\{v_i\}_{i=1,\ldots,d_1} \\ &\{w_s\}_{s=1,\ldots,d_2} \\ &\{u_x\}_{x=1,\ldots,d_3} \end{aligned} \tag{4.20}$$

then

$$(T_2 \circ T_1) v_i = \sum_{x=1}^{d_3} (M_2 M_1)_{xi} u_x \qquad\qquad i = 1, \ldots, d_1 \tag{4.21}$$

where

$$(M_2 M_1)_{xi} = \sum_{s=1}^{d_2} (M_2)_{xs} (M_1)_{si} \tag{4.22}$$

Remarks

1. Left- vs. Right conventions: One could compose linear transformations as $T_1 T_2$ and then all the indices would be transposed... ♣explain more clearly ♣

2. *Change of coordinates and active vs. passive transformations.* Given a *choice* of basis $\{v_i\}$ of an n-dimensional vector space V there is a canonically determined system of *coordinates* on V defined by

$$v = \sum_i v_i x^i = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \tag{4.23}$$

If we apply a linear transformation $T : V \to V$ then we can think of the transformation in two ways:

A.) We can say that when we move the vector $v \mapsto T(v)$ the coordinates of the vector change according to

$$T(v) = \sum_i T(v_i) x^i = \sum_i v_i \tilde{x}^i \tag{4.24}$$

with

$$\begin{pmatrix} \tilde{x}^1 \\ \vdots \\ \tilde{x}^n \end{pmatrix} = \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \cdots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \tag{4.25}$$

B.) On the other hand, we could say that the transformation defines a new basis $\tilde{v}_i = T(v_i)$ with

$$\begin{pmatrix} \tilde{v}_1 & \cdots & \tilde{v}_n \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ldots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{pmatrix} \tag{4.26}$$

This leads to the *passive* viewpoint: We could describe linear transformations by saying that they are just changing basis. In this description the vector does not change, but only its coordinate description changes. In the passive view the new coordinates of the *same* vector are gotten from the old by

$$v = \sum_i \tilde{v}_i y^i = \sum_i v_i x^i \tag{4.27}$$

and hence the change of coordinates is:

$$\vec{y} = M^{-1}\vec{x} \tag{4.28}$$

Exercise
Let V_1 and V_2 be finite dimensional vector spaces over κ. Show that

$$\dim_\kappa \mathrm{Hom}(V_1, V_2) = (\dim_\kappa V_1)(\dim_\kappa V_2) \tag{4.29}$$

Exercise
If $n = \dim V < \infty$ then $GL(V)$ is isomorphic to the group $GL(n, \kappa)$ of invertible $n \times n$ matrices over κ, but it is not canonically isomorphic.

Exercise
A one-dimensional vector space L (also sometimes referred to as a *line*) is isomorphic to, but not canonically isomorphic to, the one-dimensional vector space κ.

Show that, nevertheless, the one-dimensional vector space $\mathrm{Hom}(L, L)$ is indeed canonically isomorphic to the vector space κ.

4.5 Determinant And Trace

♣Perhaps just move the trace and determinant to the sections below where we can give natural definitions. ♣

Let V be finite-dimensional. Two important quantities associated with a linear transformation $T : V \to V$ are the trace and determinant.

To define the trace we choose any ordered basis for V so we can define a matrix M_{ij} relative to that basis. Then we can define:

$$\mathrm{tr}\,(T) := \mathrm{tr}\,(M) := \sum_{i=1}^{n} M_{ii} \tag{4.30}$$

eq:trace-def

Then we note that if we change basis we have $M \to g^{-1}Mg$ for $g \in GL(n,\kappa)$ and the above expression remains invariant thanks to cyclicity of the trace. This is a good example where it is simplest to choose a basis and define the quantity, even though it is canonically associated to the linear transformation.

For the determinant of $T : V \to V$ we can choose an ordered basis as before and define:

$$\det(T) := \det(M) := \frac{1}{n!} \sum_{\sigma \in S_n} \epsilon_{\sigma(1)\cdots\sigma(n)} M_{1\sigma(1)} \cdots M_{n\sigma(n)} \tag{4.31}$$

eq:det-def1

One can show that $\det(M_1 M_2) = \det(M_1)\det(M_2)$ and therefore under $M \to g^{-1}Mg$ the determinant is unchanged and therefore $\det(T)$ is a natural quantity.

The determinant provides an example where there is a perfectly good natural definition that never makes use of a choice of basis. One does need the anti-symmetric product of vector spaces and the notion of an orientation. See below. it makes use of

Remarks:

1. Note well that the above definitions only apply to a linear transformation of a vector space to itself: That is: $T : V \to V$. If $T : V \to W$ is a linear transformation between different vector spaces - *even if they have the same dimension!* - then there is no natural notion of trace. There is a generalization of the notion of determinant. See Section §24 below.

2. The notion of trace and determinant can be extended to certain special linear operators on infinite-dimensional vector spaces. For example, bounded linear operators are traceclss if [***** SEE SECTION **** BELOW]. For Fredholm operators on Hilbert space there is a notion of Fredholm determinant.

Exercise

Prove that the expressions on the RHS of (**??**) are basis independent so that the equations make sense.

Exercise *Standard identities*
Prove:
1.) $\mathrm{tr}\,(M+N) = \mathrm{tr}\,(M) + \mathrm{tr}\,(N)$
2.) $\mathrm{tr}\,(MN) = \mathrm{tr}\,(NM)$
3.) $\mathrm{tr}\,(SMS^{-1}) = \mathrm{tr}\,M$, $S \in GL(n,k)$.
4.) $\det(MN) = \det M \det N$
5.) $\det(SMS^{-1}) = \det M$
What can you say about $\det(M+N)$?

Exercise *Laplace expansion in complementary minors*
Show that the determinant of an $n \times n$ matrix $A = (a_{ij})$ can be written as

$$\det A = \sum_H \epsilon^{HK} b_H c_K \tag{4.32}$$

Here we fix an integer p, H runs over disjoint subsets of order p, $H = \{h_1, \ldots, h_p\}$ and K is the complementary set so that $H \amalg K = \{1, \ldots, n\}$. Then we set

$$b_H := \det(a_{i,h_j})_{1 \le i,j \le p} \tag{4.33}$$

$$c_L = \det(a_{j,k_l})_{p+1 \le j \le n, 1 \le l \le q} \tag{4.34}$$

$p + q = n$ and

$$\epsilon^{HK} = \mathrm{sign} \begin{pmatrix} 1 & 2 & \cdots & p & p+1 & \cdots & n \\ h_1 & h_2 & \cdots & h_p & k_1 & \cdots & k_q \end{pmatrix} \tag{4.35}$$

---

5. New Vector Spaces from Old Ones

subsec:dirsum

5.1 Direct sum

Given two modules V, W over a ring we can form the direct sum. As a set we have:

$$V \oplus W := \{(v,w) : v \in V, w \in W\} \tag{5.1}$$ eq:ds

while the module structure is defined by:

$$\alpha(v_1, w_1) + \beta(v_2, w_2) = (\alpha v_1 + \beta v_2, \alpha w_1 + \beta w_2) \tag{5.2}$$ eq:dsii

valid for all $\alpha, \beta \in R, v_1, v_2 \in V, w_1, w_2 \in W$. We sometimes denote (v,w) by $v \oplus w$. In particular, if the ring is a field these constructions apply to vector spaces.

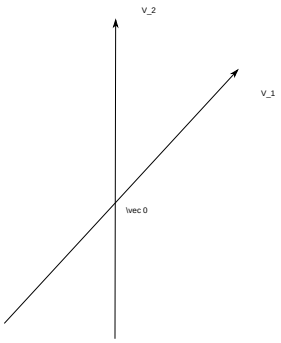


Figure 1: $\mathbb{R}^2$ is the direct sum of two one-dimensional subspaces V_1 and V_2. fig:drctsum

If V, W are finite dimensional vector spaces then:

$$\dim(V \oplus W) = \dim V + \dim W \tag{5.3}$$

Example 5.1.1:

$$\mathbb{R}^n \oplus \mathbb{R}^m \cong \mathbb{R}^{n+m} \tag{5.4}$$

Similarly for operators: With $T_1 : V_1 \to W_1, T_2 : V_2 \to W_2$ we define

$$(T_1 \oplus T_2)(v \oplus w) := T_1(v) \oplus T_2(w) \tag{5.5}$$

for $v \in V_1$ and $w \in W_1$.

Suppose we choose ordered bases:

1. $\{v_1^{(1)}, \ldots, v_{n_1}^{(1)}\}$ for V_1
2. $\{v_1^{(2)}, \ldots, v_{n_2}^{(2)}\}$ for V_2
3. $\{w_1^{(1)}, \ldots, w_{m_1}^{(1)}\}$ for W_1
4. $\{w_1^{(2)}, \ldots, w_{m_2}^{(2)}\}$ for W_2

Then, we have matrix representations M_1 and M_2 of T_1 and T_2, respectively. Among the various bases for $V_1 \oplus V_2$ and $W_1 \oplus W_2$ it is natural to choose the ordered bases:

1. $\{v_1^{(1)} \oplus 0, \ldots, v_{n_1}^{(1)} \oplus 0, 0 \oplus v_1^{(2)}, \ldots, 0 \oplus v_{n_2}^{(2)}\}$ for $V_1 \oplus V_2$
2. $\{w_1^{(1)} \oplus 0, \ldots, w_{m_1}^{(1)} \oplus 0, 0 \oplus w_1^{(2)}, \ldots, 0 \oplus w_{m_2}^{(2)}\}$ for $W_1 \oplus W_2$

With respect to these ordered bases the matrix of $T_1 \oplus T_2$ will be block diagonal:

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \tag{5.6}$$ eq:DirectSumMa

But there are, of course, other choices of bases one could make.

Remarks:

1. *Internal and external direct sum.* What we have defined above is sometimes known as *external direct sum.* If $V_1, V_2 \subset V$ are linear subspaces of V, then $V_1 + V_2$ makes sense as a linear subspace of V. If $V_1 + V_2 = V$ and in addition $v_1 + v_2 = 0$ implies $v_1 = 0$ and $v_2 = 0$, that is, if $V_1 \cap V_2 = \{0\}$ then we say that V is the *internal direct sum of* V_1 *and* V_2. In this case every vector $v \in V$ has a canonical decomposition $v = v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$ and hence there is a canonical isomorphism $V \cong V_1 \oplus V_2$. Therefore, we will in general not distinguish carefully between internal and external direct sum. Note well, however, that it can very well happen that $V_1 + V_2 = V$ and yet $V_1 \cap V_2$ is a nonzero subspace. In this case V is most definitely *not* a direct sum! As an extreme example, note that $V + V = V$ but $V \oplus V$ is not isomorphic to V unless V is the 0-vector space.

2. *Subtracting vector spaces in the Grothendieck group.* Since we can "add" vector spaces with the direct sum it is a natural question whether we can also "subtract" vector spaces. There is indeed such a notion, but one must treat it with care. The Grothendieck group can be defined for any monoid. It can then be applied to the monoid of vector spaces. If M is a commutative monoid so we have an additive operator $m_1 + m_2$ and a 0, but no inverses then we can formally introduce inverses as follows: We consider the set of pairs (m_1, m_2) with the equivalence relation that $(m_1, m_2) \sim (m_3, m_4)$ if

$$m_1 + m_4 = m_2 + m_3 \tag{5.7}$$

Morally speaking, the equivalence class of the pair $[(m_1, m_2)]$ can be thought of as the difference $m_1 - m_2$ and we can now add

$$[(m_1, m_2)] + [(m_1', m_2')] = [(m_1 + m_1', m_2 + m_2')] \tag{5.8}$$

From the definition we easily see $[(m, m)] = [(0, 0)]$. Similarly, it is easy to see that $[(m_2, m_1)] + [(m_1, m_2)] = [(m_1 + m_2, m_1 + m_2)] = [(0, 0)]$. (Note we used the commutative property of the monoid addition.) Therefore $[(m_2, m_1)]$ is the inverse of $[(m_1, m_2)]$. Therefore, the set of equivalence classes $[(m_1, m_2)]$ is an Abelian group associated to the monoid M. It is usually denoted $K(M)$. For example, if we apply this to the monoid of nonnegative integers then we recover the Abelian group of all integers.

This idea can be generalized to suitable categories. One setting is that of an *additive category* $\mathcal{C}$. What this means is that the morphism spaces $\hom(x_1, x_2)$ are Abelian groups and the composition law is bi-additive, meaning the composition of morhphisms

$$\hom(x, y) \times hom(y, z) \to \hom(x, z) \tag{5.9}$$

is "bi-additive" i.e. distributive: $(f + g) \circ h = f \circ h + g \circ h$. In an additive category we also have a zero object 0 that has the property that $\hom(0, x)$ and $\hom(x, 0)$ are the trivial Abelian group. Finally, there is a notion of direct sum of objects $x_1 \oplus x_2$, that is, given two objects we can produce a new object $x_1 \oplus x_2$ together with distinguished

♣Need to check: Do you need an abelian category to define $K(\mathcal{C})$. More generally we can define $K(\mathcal{C})$ for an exact cagetory. Is there a problem with just using additive categories? Definition here seems to make sense. ♣

morphisms $\iota_1 \in \hom(x_1, x_1 \oplus x_2)$ and $\iota_2 \in \hom(x_2, x_1 \oplus x_2)$ so that

$$\hom(z, x_1) \times \hom(z, x_2) \to \hom(z, x_1 \oplus x_2) \tag{5.10}$$

defined by $(f, g) \mapsto \iota_1 \circ f + \iota_2 \circ f$ is an isomorphism of Abelian groups.

The category **VECT** of finite-dimensional vector spaces and linear transformations or the category $\mathrm{Mod}(R)$ of modules over a ring R are good examples of additive categories. In such a category we can define the *Grothendieck group* $K(\mathcal{C})$. It is the abelian group of pairs of objects $(x, y) \in Obj(\mathcal{C}) \times Obj(\mathcal{C})$ subject to the relation

$$(w, x) \sim (y, z) \tag{5.11}$$

if there exists an object u so that the object $w \oplus z \oplus u$ is isomorphic to $y \oplus x \oplus u$. In the case of **VECT** we can regard $[(V_1, V_2)] \in K(\mathbf{VECT})$ as the formal difference $V_1 - V_2$. It is then a good exercise to show that, as Abelian groups $K(\mathbf{VECT}) \cong \mathbb{Z}$. (This is again essentially the statement that the only invariant of a finite-dimensional vector space is its dimension.)

When we apply this construction to continuous families of vector spaces parametrized by topological spaces we obtain a very nontrivial mathematical subject known as *K-theory*. In particular, if M is a manifold and $\pi_1 : \mathcal{E}_1 \to M$ and $\pi_2 : \mathcal{E}_2 \to M$ are two vector bundles it is possible to define $\mathcal{E}_1 \oplus \mathcal{E}_2$. Each fiber is the direct sum, and the fibers vary continuously. So the set $\mathbf{VECT}(M)$ of isomorphism classes of vector bundles over M is a monoid. The corresponding Grothendieck group, denoted just by $K(M)$, is an important Abelian group associated to M known as the K-theory of M. This Abelian group is a topological invariant of M. This is the beginning of a very nontrivial and beautiful subject in mathematics known as K-theory.

In physics one way such *virtual vector spaces* arise is in $\mathbb{Z}_2$-graded or super-linear algebra. See Section §23 below. Given a supervector space $V^0 \oplus V^1$ it is natural to associate to it the virtual vector space $V^0 - V^1$ (but you cannot go the other way - why not?). Some important constructions only depend on the virtual vector space (or, more generally, the virtual vector bundle, when working with families).

♣This exercise is too easy, and it is already answered in text above. ♣

Exercise

Construct an example of a vector space V and proper subspaces $V_1 \subset V$ and $V_2 \subset V$ such that

$$V_1 + V_2 = V \tag{5.12}$$

but $V_1 \cap V_2$ is not the zero vector space. Rather, it is a vector space of positive dimension. In this case V is not the internal direct sum of V_1 and V_2.

Exercise
1. $\mathrm{Tr}(T_1 \oplus T_2) = \mathrm{Tr}(T_1) + \mathrm{Tr}(T_2)$
2. $\det(T_1 \oplus T_2) = \det(T_1)\det(T_2)$

Exercise
a.) Show that there are natural isomorphisms

$$V \oplus W \cong W \oplus V \tag{5.13}$$

$$(U \oplus V) \oplus W \cong U \oplus (V \oplus W) \tag{5.14}$$

b.) Suppose I is some set, not necessarily ordered, and we have a family of vector spaces V_i indexed by $i \in I$. One can give a definition of the vector space:

$$\oplus_{i \in I} V_i \tag{5.15}$$

but to do so in general one should use the restricted product so that an element is a collection $\{v_i\}$ of vectors $v_i \in V_i$ where in the Cartesion product all but finitely many v_i are zero. Define a vector space structure on this set.

Exercise
How would the matrix in (5.6) change if we used bases:

1. $\{v_1^{(1)} \oplus 0, \ldots, v_{n_1}^{(1)} \oplus 0, 0 \oplus v_1^{(2)}, \ldots, 0 \oplus v_{n_2}^{(2)}\}$ for $V_1 \oplus V_2$
2. $\{0 \oplus w_1^{(2)}, \ldots, 0 \oplus w_{m_2}^{(2)}, w_1^{(1)} \oplus 0, \ldots, w_{m_1}^{(1)} \oplus 0\}$ for $W_1 \oplus W_2$

5.2 Quotient Space

$W \subset V$ is a vector subspace then

$$V/W \tag{5.16}$$

is the space of equivalence classes $[v]$ where $v_1 \sim v_2$ if $v_1 - v_2 \in W$. This is the quotient of abelian groups. It becomes a vector space when we add the rule

$$\alpha(w + W) := \alpha w + W. \tag{5.17}$$

Claim: V/W is also a vector space. If V, W are finite dimensional then

$$\dim(V/W) = \dim V - \dim W \tag{5.18}$$

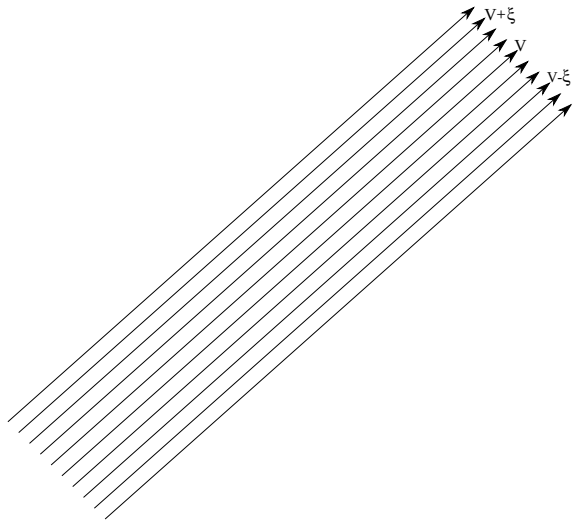


Figure 2: Vectors in the quotient space $\mathbb{R}^2/V$, $[v] = V + v$. The quotient space is the moduli space of lines parallel to V. fig:quotspace

We define a *complementary subspace* [5] to W to be another subspace $W' \subset V$ so that V is the internal direct sum of W and W'. Recall that this means that every $v \in V$ can be *uniquely* written as $v = w + w'$ with $w \in W$ and $w' \in W'$ so that $V \cong W \oplus W'$. It follows from Theorem 4.3.1 that a complementary subspace to W always exists and moreover there is an isomorphism:

$$V/W \cong W' \tag{5.19}$$ eq:vwwprime

Note that given $W \subset V$ there is a canonical vector space V/W, but there is no unique choice of W' so the isomorphism (5.19) cannot be natural.

Warning! One should not confuse (as is often done) V/W with a subspace of V. If V is an inner product space (see section *** below) then there is a notion of $W^\perp \subset V$. If V is a complete inner product space with respect to a positive definite inner product (i.e., if V is a Hilbert space) then the orthogonal projection theorem below shows that $V \cong W \oplus W^\perp$. But without extra structure, such as an inner product there is no canonical transverse space to W.

Exercise *Practice with quotient spaces*

a.) $V = \mathbb{R}^2$, $W = \{(\alpha_1 t, \alpha_2 t) : t \in \mathbb{R}\}$. If $\alpha_1 \neq 0$ then we can identify $V/W \cong \mathbb{R}$ via $s \to (0, s) + W$. Show that the inverse transformation is given by $(v_1, v_2) \to s$ where $s = (\alpha_1 v_2 - \alpha_2 v_1)/\alpha_1$. What happens if $\alpha_1 = 0$?

b.) If $V = \mathbb{R}^n$ and $W \cong \mathbb{R}^m$, $m < n$ with $W = \{(x_1, \ldots, x_m, 0, \ldots, 0)\}$, when is $v_1 + W = v_2 + W$?

---

[5] Please note, it is not a "complimentary subspace." A "complimentary subspace" might praise your appearance, or accompany snacks on an airplane flight.

Exercise

Suppose $T : V_1 \to V_2$ is a linear transformation and $W_1 \subset V_1$ and $W_2 \subset V_2$ are linear subspaces. Under what conditions does T descend to a linear transformation

$$\tilde{T} : V_1/W_1 \to V_2/W_2 \qquad ? \tag{5.20}$$

The precise meaning of "descend to" is that $\tilde{T}$ fits in the commutative diagram

$$\begin{array}{ccc} V_1 & \xrightarrow{T} & V_2 \\ \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi_2} \\ V_1/W_1 & \xrightarrow{\tilde{T}} & V_2/W_2 \end{array} \tag{5.21}$$

5.3 Tensor Product

The (technically) natural definition is a little sophisticated, (see, e.g., Lang's book *Algebra*, and remark 3 below), but for practical purposes we can describe it in terms of bases:

Let $\{v_i\}$ be a basis for V, and $\{w_s\}$ be a basis for W. Then $V \otimes W$ is the vector space spanned by $v_i \otimes w_s$ subject to the rules:

$$\begin{aligned} (\alpha v + \alpha' v') \otimes w &= \alpha(v \otimes w) + \alpha'(v' \otimes w) \\ v \otimes (\alpha w + \alpha' w') &= \alpha(v \otimes w) + \alpha'(v \otimes w') \end{aligned} \tag{5.22}$$

eq:dspp

for all $v, v' \in V$ and $w, w' \in W$ and all scalars α, α'.

If V, W are finite dimensional then so is $V \otimes W$ and

$$\dim(V \otimes W) = (\dim V)(\dim W) \tag{5.23}$$

In particular:

$$\mathbb{R}^n \otimes \mathbb{R}^m \cong \mathbb{R}^{nm} \tag{5.24}$$

We can similarly discuss the tensor product of operators: With $T_1 : V_1 \to W_1, T_2 : V_2 \to W_2$ we define

$$(T_1 \otimes T_2)(v_i \otimes w_j) := (T_1(v_i)) \otimes (T_2(w_j)) \tag{5.25}$$

eq:deftprd

Remarks

1. Examples where the tensor product occurs in physics are in quantum systems with several independent degrees of freedom. In general two distinct systems with Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ have statespace $\mathcal{H}_1 \otimes \mathcal{H}_2$. For example, consider a system of N spin $\frac{1}{2}$ particles. The Hilbert space is

$$\mathcal{H} = \overbrace{\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}^{\text{N times}} \qquad \dim_{\mathbb{C}} \mathcal{H} = 2^N \tag{5.26}$$

2. In Quantum Field Theory the implementation of this idea for the Hilbert space associated with a spatial manifold, when the spatial manifold is divided in two parts by a domain wall, is highly nontrivial due to UV divergences.

3. In spacetimes with nontrivial causal structure the laws of quantum mechanics become difficult to understand. If two systems are separated by a horizon, should we take the tensor product of the Hilbert spaces of these systems? Questions like this quickly lead to a lot of interesting puzzles.

4. *Defining Tensor Products Of Modules Over Rings*

   The tensor product of modules over a ring R can always be defined, even if the modules are not free and do not admit a basis. If we have no available basis then the above low-brow approach will not work. One needs to use a somewhat more abstract definition in terms of a "universal property."

   Consider any *bilinear* mapping $f : V \times W \to U$ for any R-module U. Then the characterizing (or "universal") property of a tensor product is that this map f "factors uniquely through a map from the tensor product." That means, there is

   (a) A bilinear map $\pi : V \times W \to V \otimes_R W$

   (b) A unique linear map $f' : V \otimes_R W \to U$ such that

$$f = f' \circ \pi \tag{5.27}$$

   In terms of commutative diagrams

$$\begin{array}{ccc} V \times W & \xrightarrow{\pi} & V \otimes_R W \\ & \searrow^{f} & \downarrow{\scriptstyle f'} \\ & & U \end{array} \tag{5.28}$$

   One then proves that, if a module $V \otimes_R W$ satisfying this property exists then it is unique up to unique isomorphism. Hence this property is in fact a defining property of the tensor product: This is the "natural" definition one finds in math books.

   The above property defines the tensor product, but does not prove that such a thing exists. To construct the tensor product one considers the free R-module generated by objects $v \times w$ where $v \in V, w \in W$ and takes the quotient by the submodule generated by all vectors of the form:

$$(v_1 + v_2) \times w - v_1 \times w - v_2 \times w \tag{5.29}$$

$$v \times (w_1 + w_2) - v \times w_1 - v \times w_2 \tag{5.30}$$

$$\alpha(v \times w) - (\alpha v) \times w \tag{5.31}$$

$$\alpha(v \times w) - v \times (\alpha w) \tag{5.32}$$

The projection of a vector $v \times w$ in this (incredibly huge!) module into the quotient module are denoted by $v \otimes w$.

An important aspect of this natural definition is that it allows us to define the tensor product $\otimes_{i \in I} V_i$ of a family of vector spaces labeled by a not necessarily ordered (but finite) set I.

Exercise

Given any three vector spaces U, V, W over a field κ show that there are natural isomorphisms:

a.) $V \otimes W \cong W \otimes V$

b.) $(V \otimes W) \otimes U \cong V \otimes (W \otimes U)$

c.) $U \otimes (V \oplus W) \cong U \otimes V \oplus U \otimes W$

d.) If $T_1 \in \mathrm{End}(V_1)$ and $T_2 \in \mathrm{End}(V_2)$ then under the isomorphism $V_1 \otimes V_2 \cong V_2 \otimes V_1$ the linear transformation $T_1 \otimes T_2$ is mapped to $T_2 \otimes T_1$.

e.) Show that $T_1 \otimes 1$ commutes with $1 \otimes T_2$.

Exercise *Practice with the $\otimes$ product*

1. $\mathrm{Tr}(T_1 \otimes T_2) = \mathrm{Tr}(T_1) \cdot \mathrm{Tr}(T_2)$
2. $\det(T_1 \otimes T_2) = (\det T_1)^{\dim V_2} \cdot (\det T_2)^{\dim V_1}$

Exercise *Matrices For Tensor Products Of Operators*

Let V, W be vector spaces over a field κ (or, more generally, free modules over a ring R). Suppose that a linear transformation $T^{(1)} : V \to V$ has matrix $A_1 \in M_n(R)$, with matrix elements $(A_1)_{ij}$, $1 \le i, j \le n$ with respect to an ordered basis $\{v_1, \ldots, v_n\}$ and $T^{(2)} : W \to W$ has matrix $A_2 \in M_m(R)$ with matrix elements $(A_2)_{ab}$, $1 \le a, b \le m$ with respect to an ordered basis $\{w_1, \ldots, w_m\}$.

a.) Show that if we use the ("lexicographically") ordered basis

$$v_1 \otimes w_1, v_1 \otimes w_2, \ldots, v_1 \otimes w_m, v_2 \otimes w_1, \ldots, v_2 \otimes w_m, \ldots, v_n \otimes w_1, \ldots, v_n \otimes w_m \tag{5.33}$$

Then the matrix for the operator $T^{(1)} \otimes T^{(2)}$ may be obtained from the following rule: Take the $n \times n$ matrix for $T^{(1)}$. Replace each of the matrix elements $(A_1)_{ij}$ by the $m \times m$ matrix

$$(A_1)_{ij} \to ((A_1)_{ij}(A_2)_{ab})_{1 \le a,b \le m} = \begin{pmatrix} (A_1)_{ij}(A_2)_{11} & \cdots & (A_1)_{ij}(A_2)_{1m} \\ \vdots & \vdots & \vdots \\ (A_1)_{ij}(A_2)_{m1} & \cdots & (A_1)_{ij}(A_2)_{mm} \end{pmatrix} \tag{5.34}$$

The result is an element of the ring

$$M_n(M_m(R)) \cong M_{nm}(R) \tag{5.35}$$

b.) Show that if we instead use the ordered basis

$$v_1 \otimes w_1, v_2 \otimes w_1, \ldots, v_n \otimes w_1, v_1 \otimes w_2, \ldots, v_n \otimes w_2, \ldots, v_1 \otimes w_m, \ldots, v_n \otimes w_m \tag{5.36}$$

Then to compute the matrix for the same linear transformation, $T^{(1)} \otimes T^{(2)}$, we would instead start with the matrix $A^{(2)}$ and replace each matrix element by inserting that matrix element times the matrix $A^{(1)}$. The net result is an element of The result is an element of the ring

$$M_m(M_n(R)) \cong M_{mn}(R) \tag{5.37}$$

Since $M_{nm}(R) = M_{mn}(R)$ we can compare to the expression in (a) and it will in general be different.

c.) Using the two conventions of parts (a) and (b) compute

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \otimes \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \tag{5.38}$$

as a 4×4 matrix.

Exercise *Tensor product of modules*

The above universal definition also applies to tensor products of modules over a ring R:

$$M \otimes_R N \tag{5.39}$$

where it can be very important to specify the ring R. In particular we have the very crucial relation that $\forall m \in M, r \in R, n \in N$:

$$m \cdot r \otimes n = m \otimes r \cdot n \tag{5.40}$$

where we have written M as a right-module and N as a left module so that the expression works even for noncommutative rings. [6]

These tensor products have some features which might be surprising if one is only familiar with the vector space example.

a.) $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Z}_n = 0$

b.) $\mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Z}_m = \mathbb{Z}_{(n,m)}$

In particular $\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathbb{Z}_q$ is the 0-module if p, q are relatively prime.

---

[6] In general a left-module for a ring R is naturally isomorphic to a right-module for the opposite ring R^{opp}. When R is commutative R and R^{opp} are also naturally isomorphic.

5.4 Dual Space

Consider the *linear functionals* $Hom(V, \kappa)$. This is a vector space known as the *dual space*.

This vector space is often denoted $\check{V}$, or by $V^\vee$, or by V^*. Since V^* is sometimes used for the complex conjugate of a complex vector space we will generally use the more neutral symbol $V^\vee$

♣Eliminate the use of V^*. This is too confusing. ♣

One can prove that $\dim V^\vee = \dim V$ so if V is finite dimensional then V and $V^\vee$ must be isomorphic. However, there is no natural isomorphism between them. If we choose a basis $\{v_i\}$ for V then we can define linear functionals ℓ_i by the requirement that

$$\ell_i(v_j) = \delta_{ij} \qquad \forall v_j \tag{5.41}$$

and then we extend by linearity to compute ℓ_i evaluated on linear combinations of v_j. The linear functionals ℓ_i form a basis for $V^\vee$ and it is called *dual basis* for $V^\vee$ with respect to the v_i. Sometimes we will call the dual basis $\hat{v}^i$ or $v_i^\vee$.

Remarks:

1. It is important to stress that there is no *natural* isomorphism between V and $V^\vee$. Once one chooses a basis v_i for V then there is indeed a naturally associated dual basis $\ell_i = \hat{v}^i = v_i^\vee$ for $V^\vee$ and then both vector spaces are isomorphic to $\kappa^{\dim V}$. The lack of a natural isomorphism means that when we consider vector spaces in families, or add further structure, then it can very well be that V and $V^\vee$ become nonisomorphic. For example, if $\pi : E \to M$ is a vector bundle over a manifold then there is a canonically associated vector bundle $\pi^\vee : E^\vee \to M$, known as the *dual bundle*, whose fiber $E_m^\vee$ above a point $m \in M$ is the dual space of the fiber E_m of E above m. That is:
$$(E^\vee)_m := (E_m)^\vee := \mathrm{Hom}(E_m, \kappa) \tag{5.42}$$
In general $E^\vee$ and E are nonisomorphic vector bundles. As a simple example, let us return to the two rank one complex line bundles $\pi_\pm : L_\pm \to S^2$ defined by the family of projection operators $P_\pm = \frac{1}{2}(1 \pm \hat{x}\cdot\vec{\sigma})$. The dual bundle to L_+ is not isomorphic to L_+. So that this means is that, even though there is indeed a family of isomorphisms
$$\psi(\hat{x}) : L_{+,\vec{x}} \cong L_{+,\vec{x}}^\vee \tag{5.43}$$
(because both sides are one dimensional vector spaces!) there is in fact no continuous family of isomorphisms. One can prove this using the following facts: The topological type of a complex line bundle over S^2 is completely classified by its *first Chern class*, which, in this case, is just an integer. It turns out that for a line bundle $c_1(L^\vee) = -c_1(L)$. Now $c_1(L_\pm 1) = \pm 1$, so, in fact $L_+^\vee \cong L_-$.

   We remark that Dirac's famous paper of 1931 showed (in modern terms) that the wavefunction of an electron confined to a two-dimensional sphere surrounding a magnetic monopole of charge m is actually not a complex-valued function on the sphere but rather a *section of a line bundle*. This means that for every $\hat{x} \in S^2$ we have $\psi(\hat{x}) \in (L_m)_{\hat{x}}$ for a complex line bundle $L_m \to S^2$. The magnetic charge m is the same as $c_1(L_m)$. In particular $L_\pm$ are the line bundles where an electron wavefunction is valued in the presence of a magnetic monopole of charge ± 1.

2. *Notation.* There is also a notion of a complex conjugate of a complex vector space which should not be confused with $V^\vee$ (the latter is defined for a vector space over any field κ). We will denote the complex conjugate, defined in Section §9 below, by $\overline{V}$. Similarly, the dual operator $T^\vee$ below is not to be confused with Hermitian adjoint. The latter is only defined for inner product spaces, and will be denoted $T^\dagger$. Note, however, that for complex numbers we will occasionally use z^* for the complex conjugate.

♣For consistency you really should use $\bar{z}$ for the complex conjugate of a complex number z everywhere. ♣

3. If
$$T : V \to W \tag{5.44}$$
is a linear transformation between two vector spaces then there is a canonical dual linear transformation
$$T^\vee : W^\vee \to V^\vee \tag{5.45}$$
To define it, suppose $\ell \in W^\vee$. Then we define $T^\vee(\ell)$ by saying how it acts on a vector $v \in V$. The formula is:
$$(T^\vee(\ell))(v) := \ell(T(v)) \tag{5.46}$$

4. Note especially that dualization "reverses arrows": If
$$V \xrightarrow{T} W \tag{5.47}$$
then
$$V^\vee \xleftarrow{T^\vee} W^\vee \tag{5.48}$$
This is a general principle: If there is a commutative diagram of linear transformations and vector spaces then dualization reverses all arrows.

♣Need exercise relating $(A_1 \otimes A_2)^{tr}$ to $A_1^{tr} \otimes A_2^{tr}$. ♣

Exercise
If V is finite dimensional show that
$$(V^\vee)^\vee \cong V \tag{5.49}$$
and in fact, this is a natural isomorphism (you do not need to choose a basis to define it).

For this reason the dual pairing between V and $V^\vee$ is often written as $\langle \ell, v \rangle$ to emphasize the symmetry between the two factors.

Exercise *Matrix of* $T^\vee$

a.) Suppose $T : V \to W$ is a linear transformation, and we choose a basis $\{v_i\}$ for V and $\{w_s\}$ for W, so that the matrix of the linear transformation is M_{si}.

Show that the matrix of $T^\vee : W^\vee \to V^\vee$ with respect to the dual bases to $\{v_i\}$ and $\{w_s\}$ is the transpose:

$$(M^{tr})_{is} := M_{si} \tag{5.50}$$

b.) Suppose that v_i and v_i' are two bases for a vector space V and are related by $v_i' = \sum_j S_{ji} v_j$. Show that the corresponding dual bases ℓ_i and ℓ_i' are related by $\ell_i' = \sum_j \hat{S}_{ji}\ell_j$ where

$$\hat{S} = S^{tr,-1} \tag{5.51}$$

c.) For those who know about vector bundles: If a bundle $\pi : E \to M$ has a coordinate chart with $\mathcal{U}_\alpha \subset M$ and transition functions $g^E_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to GL(n,\kappa)$ show that the dual bundle has transition functions $g^{E^\vee}_{\alpha\beta} = (g^E_{\alpha\beta})^{-1,tr}$.

Exercise

Prove that there are canonical isomorphisms: [7]

$$\mathrm{Hom}(V,W) \cong V^\vee \otimes W. \tag{5.52}$$ eq:HomVW-iso

$$\mathrm{Hom}(V,W)^\vee \cong \mathrm{Hom}(W,V) \tag{5.53}$$

$$\mathrm{Hom}(V,W) \cong \mathrm{Hom}(W^\vee, V^\vee) \tag{5.54}$$

Although the isomorphism (5.52) is a natural isomorphism it is useful to say what it means in terms of bases: If

$$Tv_i = \sum_s M_{si} w_s \tag{5.55}$$

then the corresponding element in $V^\vee \otimes W$ is

$$T = \sum_{i,s} M_{si} v_i^\vee \otimes w_s \tag{5.56}$$

Exercise *A Useful Canonical Isomorphism*

Suppose that L is a one-dimensional space over κ. The field κ can itself be regarded as a one-dimensional vector space over κ, but of course the isomorphism of L with κ is not canonical. There is no canonical way to associate a nonzero vector $v \in L$ to, say, the element $1 \in \kappa$.

[7] *Answer*: The second and third isomorphisms follow easily from the first. To establish the first note that an element of $V^\vee \otimes W$ certainly determines a linear transformation $V \to W$ by contraction of $V^\vee$ with V. Then, at least for finite dimensional vector spaces, you can just compare dimensions to check that this is an isomorphism.

Show, in two ways, that, nevertheless, the one-dimensional vector space

$$\mathrm{Hom}_\kappa(L, L) \tag{5.57}$$

is indeed canonically isomorphic with κ. [8]

Exercise *Natural definition of the trace*

a.) Show that for any vector space V over κ there is a natural linear operator

$$ev : V^\vee \otimes V \to \kappa \tag{5.58}$$

b.) Show that if V is finite-dimensional then there is a natural linear operator [9]

$$\mathbf{1} : \kappa \to V^\vee \otimes V \tag{5.60}$$

c.) The composition $ev \circ \mathbf{1}$ defines an element of $\mathrm{Hom}_\kappa(\kappa, \kappa)$ but $\mathrm{Hom}_\kappa(\kappa, \kappa)$ is naturally isomorphic to κ. Therefore, $ev \circ \mathbf{1}$ can naturally be identified with an element of κ. Show that it is just $\dim_\kappa V$.

d.) If $T : V \to V$ and V is finite dimensional, consider the composition of linear transformations:

$$\kappa \xrightarrow{\mathbf{1}} V^\vee \otimes V \xrightarrow{1 \otimes T} V^\vee \otimes V \xrightarrow{ev} \kappa \tag{5.61}$$

This defines an element of $\mathrm{Hom}_\kappa(\kappa, \kappa)$ and is therefore, naturally, an element of κ. Show that this is just the trace of T.

Exercise

If U is any vector space and $W \subset U$ is a linear subspace then we can define

$$W^\perp := \{\ell | \ell(w) = 0 \quad \forall w \in W\} \subset U^\vee. \tag{5.62}$$

[8] *Answer*: One way to do this is to note that if $T \in \mathrm{Hom}(L, L)$ then it must be of the form $T(v) = \alpha v$ for some scalar $\alpha \in \kappa$. So we send $T \to \alpha$. A second way to think about this is that $\mathrm{Hom}(L, L) \cong L^\vee \otimes L$. Now, the one-dimensional space $L^\vee \otimes L$ does have a canonical basis vector: Choose any nonzero vector $v \in L$ and define $\ell_v \in L^\vee$ to be the linear transformation defined by $\ell_v(v) = 1$. Then $\ell_v \otimes v \in L^\vee \otimes L$ is in fact independent of v. So we have a canonical isomorphism $\mathbf{1} : \kappa \to L^\vee \otimes L$ defined by $\mathbf{1}(1) = \ell_v \otimes v$.

[9] *Answer*:If one insists on complete naturality then the way to define this is to note that there is a unique operator $\mathbf{1}$ such that

$$V \xrightarrow{Id \otimes \mathbf{1}} V \otimes V^\vee \otimes V \xrightarrow{ev \otimes Id} V \tag{5.59}$$

is the identity matrix. If $\{v_i\}$ is a basis then we can say that $\mathbf{1}(1) = \sum_i v_i^\vee \otimes v_i$. You can check that, although we chose a basis $\sum_i v_i^\vee \otimes v_i$ does not depend on the choice of basis, and hence in that sense it is still natural.

Show that
a.) $(W^\perp)^\perp = W$.
b.) $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$.
c.) There is a canonical isomorphism $(U/W)^\vee \cong W^\perp$.

Exercise *Flags*

♣Some good exercises can be extracted from the snakes paper. ♣

6. Tensor spaces

Given a vector space V we can form

$$V^{\otimes n} \equiv V \otimes V \otimes \cdots \otimes V \tag{6.1}$$

Elements of this vector space are called *tensors of rank n over V*.

Actually, we could also consider the dual space $V^\vee$ and consider more generally

$$V^{\otimes n} \otimes (V^\vee)^{\otimes m} \tag{6.2}$$

eq:mxdtens

Up to isomorphism, the order of the factors does not matter.

Elements of (6.2) are called *mixed tensors of type* (n, m). For example

$$\mathrm{End}(V) \cong V^\vee \otimes V \tag{6.3}$$

are mixed tensors of type $(1, 1)$.

Now, if we choose an ordered basis $\{v_i\}$ for V then we have a canonical dual basis for $V^\vee$ given by $\{\hat{v}^i\}$ with

$$\hat{v}^i(v_j) = \delta^i{}_j \tag{6.4}$$

Notice, we have introduced a convention of upper and lower indices which is very convenient when working with mixed tensors.

A typical mixed tensor can be expanded using the basis into its components:

$$T = \sum_{i_1, i_2, \ldots, i_n, j_1, \ldots, j_m} T^{i_1, i_2, \ldots, i_n}{}_{j_1, \ldots, j_m} v_{i_1} \otimes \cdots \otimes v_{i_n} \otimes \hat{v}^{j_1} \otimes \cdots \otimes \hat{v}^{j_m} \tag{6.5}$$

We will henceforth often assume the summation convention where repeated indices are automatically summed.

Recall that if we make a change of basis

$$w_i = g^j{}_i v_j \tag{6.6}$$

(sum over j understood) then the dual bases are related by

$$\hat{w}^i = \hat{g}_j{}^i \hat{w}^j \tag{6.7}$$

where the matrices are related by

$$\hat{g} = g^{tr,-1} \tag{6.8}$$

The fact that g has an upper and lower index makes good sense since it also defines an element of $\mathrm{End}(V)$ and hence the matrix elements are components of a tensor of type $(1,1)$.

Under change of basis the (passive) change of components is given by

$$(T')^{i'_1,\ldots,i'_n}{}_{j'_1,\ldots,j'_m} = g^{i'_1}{}_{k_1}\cdots g^{i'_n}{}_{k_n}\hat{g}_{j'_1}{}^{\ell_1}\cdots\hat{g}_{j'_m}{}^{\ell_m}T^{k_1,\ldots,k_n}{}_{\ell_1,\ldots,\ell_m} \tag{6.9}$$

This is the standard transformation law for tensors.

♣Comment on "covariant" and "contravariant" indices. ♣

6.1 Totally Symmetric And Antisymmetric Tensors

There are two subspaces of $V^{\otimes n}$ which are of particular importance. Note that we have a homomorphism $\rho : S_n \to GL(V^{\otimes n})$ defined by

$$\rho(\sigma) : u_1 \otimes \cdots \otimes u_n \to u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(n)} \tag{6.10}$$

on any vector in $V^{\otimes n}$ of the form $u_1 \otimes \cdots \otimes u_n$ Then we extend by linearity to all vectors in the vector space. Note that with this definition

$$\rho(\sigma_1) \circ \rho(\sigma_2) = \rho(\sigma_1\sigma_2) \tag{6.11}$$

Thus, $V^{\otimes n}$ is a *representation* of S_n in a natural way. This representation is *reducible*, meaning that there are invariant proper subspaces. (See Chapter four below for a systematic treatment.)

Two particularly important proper subspaces are:

$S^n(V)$: These are the *totally symmetric tensors*, i.e. the vectors invariant under S_n. This is the subspace where $T(\sigma) = 1$ for all $\sigma \in S_n$.

$\Lambda^n(V)$: antisymmetric tensors, transform into $v \to \pm v$ depending on whether the permutation is even or odd. These are also called n-forms. This is the subspace of vectors on which $T(\sigma)$ acts by ± 1 given by $T(\sigma) = \epsilon(\sigma)$.

Remarks

1. A basis for $\Lambda^n V$ can be defined using the extremely important wedge product construction. If v, w are two vectors we define

$$v \wedge w := \frac{1}{2!}(v \otimes w - w \otimes v) \tag{6.12}$$

♣Note: Assumes κ is characteristic zero. Should say what happens for modules more generally. ♣

If $v_1, \ldots, v_n$ are any n vectors we define

$$v_1 \wedge \cdots \wedge v_n := \frac{1}{n!}\sum_{\sigma \in S_n} \epsilon(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)} \tag{6.13}$$

Note that

$$v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(n)} = \mathrm{sign}(\sigma) v_1 \wedge \cdots \wedge v_n \tag{6.14}$$

Thus, if we choose an ordered basis w_i for V then $\Lambda^n V$ is spanned by the vectors:

$$w_{i_1} \wedge \cdots \wedge w_{i_n} \qquad\qquad i_1 < i_2 < \cdots < i_n \tag{6.15}$$

and thus if $\dim V = d$ then $\Lambda^n V$ has dimension $\binom{d}{n}$. In particular, $\Lambda^d(V)$ is one-dimensional.

2. In quantum mechanics if V is the Hilbert space of a single particle the Hilbert space of n identical particles is a subspace of $V^{\otimes n}$. In three space dimensions the only particles that appear in nature are bosons and fermions, which are states in $S^n(V)$ and $\Lambda^n(V)$, respectively. In this way, given a space V of one-particle states - interpreted as the span of creation operators $\{a_j^\dagger\}$ we form the bosonic Fock space

$$S^\bullet V = \mathbb{C} \oplus \oplus_{\ell=1}^\infty S^\ell V \tag{6.16}$$

when the $a_j^\dagger$'s commute or the Fermionic Fock space

$$\Lambda^\bullet V = \mathbb{C} \oplus \oplus_{\ell=1}^\infty \Lambda^\ell V \tag{6.17}$$

when they anticommute. (The latter terminates, if V is finite-dimensional).

3. In chapter 4 we explain the beautiful Schur-Weyl duality theorem: If V Is the fundamental representation of $GL(d,\kappa)$ then $V^{\otimes n}$ is a representation of $S_n \times GL(d,\kappa)$. That is, the S_n and $GL(d,\kappa)$ representations commute. We can decompose this representation into irreps of S_n. These irreps are labeled by Young diagrams with n boxes. Then, as a representation of S_n, we have isotypical decomposition:

$$V^{\otimes n} \cong \oplus_{Y \in \mathcal{Y}_n} D(Y) \otimes R(Y) \tag{6.18}$$

where $R(Y)$ is the irrep of S_n associated to Y and $D(Y)$, which is a representation of $GL(d,\kappa)$ turns out to be an irreducible representation of $GL(d,\kappa)$. Moreover, considering all integers n we obtain ALL the finite-dimensional representations of $GL(d,\kappa)$. These statements also hold true if we replace $GL(d,\mathbb{C})$ by $U(d)$.

Exercise *Decomposing* $V^{\otimes 2}$

a.) Show that $\Lambda^2(V)$ is given by linear combinations of vectors of the form $x \otimes y - y \otimes x$, and $S^2(V)$ is given by vectors of the form $x \otimes y + y \otimes x$.

Show that [10]

$$V^{\otimes 2} \cong S^2(V) \oplus \Lambda^2(V) \tag{6.19}$$

b.) If V_1, V_2 are two vector spaces show that

$$S^2(V_1 \oplus V_2) \cong S^2(V_1) \oplus S^2(V_2) \oplus V_1 \otimes V_2 \tag{6.20}$$

$$\Lambda^2(V_1 \oplus V_2) \cong \Lambda^2(V_1) \oplus \Lambda^2(V_2) \oplus V_1 \otimes V_2 \tag{6.21}$$

♣Incorporate some nice generalizations of this from CSS project with Jeff. ♣

Exercise *Counting dimensions*

Suppose V is finite dimensional of dimension $\dim V = d$. It is of interest to count the dimensions of $\Lambda^n(V)$ and $S^n(V)$. Think of $\Lambda^* V$ and $S^* V$ as Fock spaces of fermions and bosons with oscillators spanning the vector space V.

a.) Show that

$$\sum_{n\geq 0} q^n \dim\Lambda^n(V) = (1+q)^d \tag{6.22}$$

and therefore

$$\dim\Lambda^n(V) = \binom{d}{n} \tag{6.23}$$

b.) Show that

$$\sum_{n\geq 0} q^n \dim S^n(V) = \frac{1}{(1-q)^d} \tag{6.24}$$

and therefore

$$\dim S^n(V) = \binom{n+d-1}{n} \tag{6.25}$$

Remark The formula for the number of nth-rank symmetric tensors in d dimensions involves *multichoosing*. In combinatorics we denote

$$\left(\binom{d}{n}\right) := \binom{n+d-1}{n} = \binom{n+d-1}{d-1} \tag{6.26}$$

eq:d-multichoo

and say "d multichoose n." This is the number of ways of counting n objects from a set of d elements where repetition is allowed and order does not matter. One standard proof is the "method of stars and bars." Suppose $T_{i_1,\ldots,i_n}$ is a totally symmetric rank n tensor. For a fixed set of indices $i_1,\ldots,i_n$, there will be a certain number of $1's$, $2's$, etc. So for a completely symmetric tensor the value of $T_{i_1,\ldots,i_n}$ is the value of T with the indices in nondecreasing order: $T_{1,\ldots 1,2,\ldots,2,\ldots,d,\ldots,d}$. (What we wrote is a little misleading since the number of, say, 2's might actually be zero.) How many distinct values can there be? We can imagine $n+d-1$ slots, and we are to choose $d-1$ of these slots as the position of a bar. Then between the bars we insert 1's, then 2's, etc., reading from left to right. Thus, the position of the bars completely determines the number of 1's, 2's, etc. and this is how we are labeling the independent components $T_{1,\ldots 1,2,\ldots,2,\ldots,d,\ldots,d}$. Thus, the independent components of a symmetric rank n tensor in d dimensions are in 1-1 correspondence with a choice of $(d-1)$ slots out of a total of $n+d-1$ slots. This is the binomial coefficient (6.26).

[10] *Answer*: $\Lambda^2(V)$ is the span of vectors of the form $x\otimes y - y\otimes x$, while $S^2(V)$ is the span of vectors of the form $x\otimes y + y\otimes x$. But we can write any $x\otimes y$ as a sum of two such vectors in an obvious way.

Expanding out the binomial coefficients a bit it is interesting to note that there is a pleasing symmetry between the antisymmetric and symmetric cases:

$$\dim\Lambda^n(V) = \binom{d}{n} = \frac{d(d-1)(d-2)\cdots(d-(n-1))}{n!} \tag{6.27}$$

$$\dim S^n(V) = \left(\binom{d}{n}\right) = \frac{d(d+1)(d+2)\cdots(d+(n-1))}{n!} \tag{6.28}$$

♣Does this mean that bosons in negative numbers of dimensions are fermions? ♣

Exercise

a.) Let $\dim V = 2$. Show that

$$\dim S^n(V) = n+1 \tag{6.29}$$

b.) Let $d = \dim V$. Show that

$$S^3(V) \oplus \Lambda^3(V) \tag{6.30}$$

is a subspace of $V^{\otimes 3}$ of codimension $\frac{2}{3}d(d^2-1)$ and hence has positive codimension for $d > 1$.

Thus, there are more nontrivial representations of the symmetric group to account for. These will be discussed in Chapter 4.

Exercise *Linear Transformations*

a.) If $T : V \to W$ is a linear transformation show that there are natural linear transformations:

$$\Lambda^k T : \Lambda^k V \to \Lambda^k W \tag{6.31}$$

$$S^k T : S^k V \to S^k W \tag{6.32}$$

See chapter **** below where this is applied to define determinants and pfaffians in a basis-independent way.

Exercise *Generating functions for symmetric powers*

Let V be a vector space over a field κ. It is sometimes useful to think of it as a representation of a group G. Note that if we have $\mathbb{R}_+$ acting on V by scaling $t : v \mapsto \lambda v$ for $t > 0$ then $\mathrm{Sym}^j(V)$ has the action of t^j. We can form:

$$\mathrm{Sym}_t^\bullet(V) := \oplus_{j=0}^\infty t^j \mathrm{Sym}^j(V) \tag{6.33}$$

where we interpret $\mathrm{Sym}^0(V) = \kappa$. Note that $W \otimes \kappa = W$ for any vector space over κ (or representation of G, if κ is the trivial representation). So κ functions as an identity, and in particular is an invertible vector space. The parameter t is put in as a bookkeeping convenience. It keeps track of the scaling weight.

a.) Show that

$$\mathrm{Sym}_t^\bullet(V_1 \oplus V_2) = \mathrm{Sym}_t^\bullet(V_1) \otimes \mathrm{Sym}_t^\bullet(V_2) \tag{6.34}$$

b.) Using this show that:

$$\mathrm{Sym}^3(V_1 \oplus V_2) = \mathrm{Sym}^3(V_1) + \mathrm{Sym}^2(V_1)V_2 + V_1\mathrm{Sym}^2(V_2) + \mathrm{Sym}^3(V_2) \tag{6.35}$$ eq:Sym3True

c.) Show that it makes sense to say for virtual spaces (or representations)

$$\mathrm{Sym}_t^\bullet(V_1 \ominus V_2) = \frac{\mathrm{Sym}_t^\bullet(V_1)}{\mathrm{Sym}_t^\bullet(V_2)} \tag{6.36}$$ eq:Diff

d.) Since κ is an invertible vector space, this makes sense by formally expanding the denominator as a power series. For example, show that:

$$\mathrm{Sym}^2(V_1 \ominus V_2) = \mathrm{Sym}^2(V_1) - V_1V_2 + V_2^2 - 2\mathrm{Sym}^2(V_2) \tag{6.37}$$ eq:Sym2Virt

$$\mathrm{Sym}^3(V_1 \ominus V_2) = \mathrm{Sym}^3(V_1) - \mathrm{Sym}^2(V_1)V_2 - V_1\mathrm{Sym}^2(V_2) + V_1V_2^2 + 2\mathrm{Sym}^2(V_2)V_2 - V_2^3 - \mathrm{Sym}^3(V_2) \tag{6.38}$$ eq:Sym3Virt

$$\mathrm{Sym}_t^\bullet(V_1 \ominus \kappa) = \frac{\mathrm{Sym}_t^\bullet(V_1)}{\mathrm{Sym}_t^\bullet(\kappa)} = \frac{\mathrm{Sym}_t^\bullet(V_1)}{1 + t + t^2 + \cdots} = (1-t)\mathrm{Sym}_t^\bullet(V_1) \tag{6.39}$$ eq:Diff

$$\mathrm{Sym}^{n+1}(V - \mathbf{1}) = \mathrm{Sym}^{n+1}(V) - \mathrm{Sym}^n(V) \tag{6.40}$$ eq:SimpleDiff

---

Exercise

Suppose a group G acts on a finite-dimensional vector space V and the action of $\rho(g)$ can be diagonalized so that $V \cong \oplus_i L_i$ with $\rho(g)$ acting as g_i on L_i. Show that the characters of the induced representation on $\Lambda^k(V)$ and $S^k(V)$ can be computed from

$$\mathrm{ch}_{\Lambda_t^* V}\rho(g) = \prod_i (1 + tg_i) \tag{6.41}$$

$$\mathrm{ch}_{S_t^* V}\rho(g) = \prod_i \frac{1}{1 - tg_i} \tag{6.42}$$

:TensorAlgebra

6.2 Algebraic structures associated with tensors

There are a number of important algebraic structures associated with tensor spaces.

Definition For a vector space V over κ the *tensor algebra* TV is the $\mathbb{Z}$-graded algebra over κ with underlying vector space:

$$T^\bullet V := \kappa \oplus V \oplus V^{\otimes 2} \oplus V^{\otimes 3} \oplus \cdots \tag{6.43}$$

and multiplication defined by using the tensor product to define the multiplication

$$V^{\otimes k} \times V^{\otimes \ell} \to V^{\otimes (k+\ell)} \tag{6.44}$$

and then extending by linearity.

Remarks

1. In concrete terms the algebra multiplication is defined by the very natural formula:

$$(v_1 \otimes \cdots \otimes v_k) \cdot (w_1 \otimes \cdots \otimes w_\ell) := v_1 \otimes \cdots \otimes v_k \otimes w_1 \otimes \cdots \otimes w_\ell \tag{6.45}$$

2. Note that we can define $V^{\otimes 0} := \kappa$ and $V^{\otimes n} = \{0\}$ when n is a negative integer making

$$T^\bullet V = \oplus_{\ell \in \mathbb{Z}} V^{\otimes \ell} \tag{6.46}$$

into a $\mathbb{Z}$-graded algebra. The vectors V are then regarded as having degree = 1.

Several quotients of the tensor algebra are of great importance in mathematics and physics.

♣This discussion of ideals really belongs in the section on algebras. Then you can give standard examples such as subalgebra of $\mathbb{Z}$ of integers divisible by n. etc. ♣

Quite generally, if A is an algebra and $B \subset A$ is a subalgebra we can ask if the vector space A/B admits an algebra structure. The only natural definition would be

$$(a + B) \odot (a' + B) := a \odot a' + B \tag{6.47}$$

However, there is a problem with this definition! It is not necessarily well-defined. The problem is very similar to trying to define a group structure on the set of cosets G/H of a subgroup $H \subset G$. Just as in that case, we can only give a well-defined product when $B \subset A$ satisfies a suitable condition. In the present case we need to know that, for all $a, a' \in A$ and for all $b, b' \in B$ then there must be a $b'' \in B$ so that

$$(a + b) \odot (a' + b') = a \odot a' + b'' \tag{6.48}$$

Taking various special cases this implies that for all $b \in B$ and all $a \in A$ we must have

$$a \odot b \in B \qquad \& \qquad b \odot a \in B \tag{6.49}$$

Such a subalgebra $B \subset A$ is known as a (left- and right-) *ideal*.

If we have a subset $S \subset A$ then $I(S)$, the ideal generated by S, also denoted simply (S) is the smallest ideal in A that contains S. It exists because the intersection of two ideals that contain S is an ideal that contains S and the set of ideals that contain S is nonempty: A itself is an ideal.

1. *Symmetric algebra*: This is the quotient of $T^\bullet V$ by the ideal generated

$$S = \{v \otimes w - w \otimes v | v, w \in V\} \tag{6.50}$$

It is denoted $S^\bullet V$, and as a vector space is

$$S^\bullet V = \oplus_{\ell=0}^{\infty} S^\ell V \tag{6.51}$$

If we denote the symmetrization of an elementary tensor simply by

$$v_1 \cdots v_k := \frac{1}{k!} \sum_{\sigma \in S_k} v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)} \tag{6.52}$$

then the product is simply

$$(v_1 \cdots v_k) \cdot (w_1 \cdots w_\ell) = v_1 \cdots v_k w_1 \cdots w_\ell \tag{6.53}$$

It is also the free commutative algebra generated by V. Even when V is finite dimensional this is an infinite-dimensional algebra.

2. *Exterior algebra*: This is the quotient of $T^\bullet V$ by the ideal $I(S)$ generated by

$$S = \{v \otimes w + w \otimes v | v, w \in V\} \tag{6.54}$$

It is denoted $\Lambda^\bullet V$, and as a vector space is

$$\Lambda^\bullet V = \oplus_{\ell=0}^{\infty} \Lambda^\ell V \tag{6.55}$$

with product given by exterior product of forms

$$(v_1 \wedge \cdots \wedge v_k) \cdot (w_1 \wedge \cdots \wedge w_\ell) := v_1 \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge w_\ell \tag{6.56}$$

When V is finite dimensional this algebra is finite dimensional.

3. *Clifford algebra*: Let Q be a symmetric quadratic form on V, and we assume that V is a vector space over a field of characteristic not equal to two. Then the Clifford algebra $C\ell(Q)$ is the algebra defined by $TV/I(S)$ where $I(S)$ is the ideal generated by

$$S = \{v \otimes w + w \otimes v - 2Q(v, w)1_\kappa | v, w \in V\} \tag{6.57}$$

This is an extremely important algebra in physics. For the moment let us just note the following elementary points: If we choose a basis $\{e_i\}$ for V and $Q(e_i, e_j) = Q_{ij}$ then we can think of $C\ell(Q)$ as the algebra over κ generated by the e_i subject to the relations:

$$e_i e_j + e_j e_i = Q_{ij} \tag{6.58}$$

If Q can be diagonalized then we can choose a basis and we write

$$e_i e_j + e_j e_i = q_i \delta_{ij} \tag{6.59}$$

Strictly speaking, when one speaks of a "Clifford algebra" it is understood that Q is nondegenerate so the q_i are all nonzero. The above construction makes sense for any symmetric quadratic form. If $Q = 0$ we obtain what is known as the *Grassmann algebra*. If Q is degenerate but nonzero one can separate the two cases and take a tensor product since we have a canonical isomorphism [11]

$$C\ell(Q_1 \oplus Q_2) \cong C\ell(Q_1)\widehat{\otimes}C\ell(Q_2) \tag{6.60}$$

To go further it depends quite a bit one what field we are working with. If $\kappa = \mathbb{C}$ and the $q_i \neq 0$ we can change basis so that

$$e_i e_j + e_j e_i = 2\delta_{ij} \tag{6.61}$$

This is, of course, the familiar algebra of "gamma matrices" and the Clifford algebras are intimately related with spinors and spin representations, and a choice of gamma matrices is a choice of representation of $C\ell(Q)$. If $\kappa = \mathbb{R}$ the best we can do is choose a basis so that

$$e_i e_j + e_j e_i = 2\eta_i \delta_{ij} \tag{6.62}$$

where $\eta_i \in \{\pm 1\}$. We note that if V is finite-dimensional then, as a vector space

$$C\ell(Q) \cong \oplus_{j=0}^{d} \Lambda^j V \tag{6.63}$$

but this isomorphism is completely false as algebras. (Why?) Clifford algebras are discussed in great detail in Chapter 10. For much more information about this crucial case see

1. The classic book: C. Chevalley, *The algebraic theory of spinors*

2. The chapters by P. Deligne in the AMS books on Strings and QFT for mathematicians.

3. Any number of online lecture notes. Including my own:

http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf (Chapter 13)

http://www.physics.rutgers.edu/∼gmoore/PiTP-LecturesA.pdf (Section 2.3)

4. *Universal enveloping algebra.* Let V be a Lie algebra. (See Chapter 8(?) below.) Then $U(V)$, the universal enveloping algebra is TV/I where I is the ideal generated by $v \otimes w - w \otimes v - [v, w]$.

Remarks

1. Explain coalgebra structures.

2. Explain about A_∞ and L_∞ algebra.

3. Koszul duality for quadratic algebras

[11] To do this right we need to regard the Clifford algebra as a $\mathbb{Z}_2$-graded, or super-algebra. Then we must take the graded tensor product. See the references at the end of this remark for further explanation.

6.2.1 An Approach To Noncommutative Geometry

One useful way of thinking about the symmetric algebra is that $S^\bullet V^\vee$ is the algebra of polynomial functions on V. Note that there is a natural evaluation map

$$S^k(V^\vee) \times V \to \kappa \tag{6.64}$$

defining a polynomial function on V. To make this quite explicit choose a basis $\{v_i\}$ for V so that there is canonically a dual basis $\{\hat{v}^i\}$ for $V^\vee$. Then an element of $S^k(V^\vee)$ is given by a totally symmetric tensor $T_{i_1\cdots i_k}\hat{v}^{i_1}\cdots\hat{v}^{i_k}$, and, when evaluated on a general element $x^i v_i$ of V we get the number

$$T_{i_1\cdots i_k}x^{i_1}\cdots x^{i_k} \tag{6.65}$$

so the algebraic structure is just the multiplication of polynomials.

Now, quite generally, *derivation* of an algebra A is a linear map $D : A \to A$ which obeys the Leibniz rule

$$D(ab) = D(a)b + aD(b) \tag{6.66}$$

In differential geometry, derivations arise naturally from vector fields and indeed, the general derivation of the symmetric algebra $S^\bullet(V^\vee)$ is given by a vector field

$$D = \sum_i f^i(x)\frac{\partial}{\partial x^i} \tag{6.67}$$

Now these remarks give an entree into the subject of *noncommutative geometry*: We can also speak of derivations of the tensor algebra $T^\bullet V$. Given the geometrical interpretation of $S^\bullet V$ it is natural to consider these as functions on a *noncommutative manifold*. We could, for example, introduce formal variables x^i which do not commute and still consider functions of these. Then, vector fields on this noncommutative manifold would simply be derivations of the algebra. In general, if noncommutative geometry the name of the game is to replace geometrical concepts with equivalent algebraic concepts using commutative rings (or fields) and then generalize the algebraic concept to noncommutative rings.

Remarks:

1. A very mild form of noncommutative geometry is known as *supergeometry*. We discuss that subject in detail in Section §23 below. For now, note that our discussion of derivations and vector fields has a nice extension to the exterior algebras.

7. Kernel, Image, and Cokernel

A linear transformation between vector spaces (or R-modules)

$$T : V \to W \tag{7.1}$$

is a homomorphism of abelian groups and so, as noted before, there are automatically three canonical vector spaces:

$$\ker T := \{v : Tv = 0\} \subset V. \tag{7.2}$$

$$\mathrm{im}T := \{w : \exists v \qquad w = Tv\} \subset W. \tag{7.3}$$

and

$$\mathrm{im}T \cong V/\ker T \tag{7.4}$$

One defines exact sequences and short exact sequences exactly as for groups.

In linear algebra, since the target W is abelian we can make one new construction not available for general groups. We define the *cokernel* of T to be the vector space:

$$\mathrm{cok}T := W/\mathrm{im}T. \tag{7.5}$$

Remarks:

1. If V, W are inner product spaces then $\mathrm{cok}T \cong \ker T^\dagger$. (See definition and exercise below.)

♣Should say more here, and explain how cokernels are relevant. How to construct generator from parity check and vice versa using orthogonal vectors. ♣

2. Classical Error-Correcting Codes. In the theory of classical error correcting codes a linear $[n, k]$ code is a k-dimensional subspace $\mathcal{C} \subset \mathbb{F}_2^n$. The codewords have n bits and there are k independent bits with the potential to encode 2^k messages. As an example: The most elementary way to make sure you are not misunderstood is to repeat yourself. The most elementary way to make sure you are not misunderstood is to repeat yourself. The most elementary way to make sure you are not misunderstood is to repeat yourself. The repetition code on one bit is an $[r, 1]$ code where r is the number of repetitions. It sends the bit 0 to $(0, \ldots, 0)$ and 1 to $(1, \ldots, 1)$. In general, the $[n, k]$ code can be thought of as being determined by a linear map:

$$G : \mathbb{F}_2^k \to \mathbb{F}_2^n \tag{7.6}$$

Choosing a standard basis for $\mathbb{F}_2^k$ and $\mathbb{F}_2^n$ the code is determined by a matrix $G \in \mathrm{Mat}_{n\times k}(\mathbb{F}_2)$ called the generator matrix. It is often useful to think of $\mathcal{C}$ as the kernel of a linear transformation

$$H : \mathbb{F}_2^n \to \mathbb{F}_2^{n-k} \tag{7.7}$$

so that the sequence

$$0 \longrightarrow \mathbb{F}_2^k \xrightarrow{G} \mathbb{F}_2^n \xrightarrow{H} \mathbb{F}_2^{n-k} \longrightarrow 0 \tag{7.8}$$

is exact. Again, choosing a standard basis we have a matrix $H \in \mathrm{Mat}_{(n-k)\times n}(\mathbb{F}_2)$ such that H is of rank $(n-k)$ and $HG = 0$. The matrix H is known as the *parity check matrix*. It is quite useful because if there is an error in transmission and the intended message $m \in \mathbb{F}_2^k$ is in fact $y' = G(m) + e$, where e is an error then we can sometimes diagonose the error and even correct it. To diagnose the error we apply the parity check operator so that $Hy' = He$. If $Hy' \neq 0$ we can be sure that an error has occured, so that Hy' is called the *error syndrome*. [12] In good cases we can

[12] Of course, it is possible that and error $e \in \mathcal{C}$ has occurred. Part of the goal of constructing good error-correcting codes is to make the vectors in $\mathcal{C}$ "extremely sparse" so that it is very unlikely to have $e \in \mathcal{C}$.

even find e and subtract it to get $y' - e = G(m)$ and then deduce m. For example, if the codewords are "far apart" and e is "small" then there will be a unique vector $y \in \mathcal{C}$ which is closest to y'. Here the proper notion of distance is the *Hamming distance*. If we know the generator matrix G then to construct H we need not just a complementary subspace to $\mathrm{Im}(G)$ but the orthogonal one. If H is of the form

$$H = \begin{pmatrix} A & I_{n-k} \end{pmatrix} \tag{7.9}$$

then

$$G = \begin{pmatrix} I_k \\ -A \end{pmatrix} \tag{7.10}$$

(this is valid over any field, such as $\mathbb{F}_q$). For this, and much more, consult [13]

Exercise

a.) If

$$0 \to V_1 \to V_2 \to V_3 \to 0 \tag{7.11}$$

is a short exact sequence of vector spaces, then $V_3 \cong V_2/V_1$.

b.) Show that a splitting of this sequence is equivalent to a choice of complementary subspace to V_1 in V_2.

Exercise *The massless vectorfield propagator*

Let $k_\mu, \mu = 1, \ldots, d$ be a nonvanishing vector in $\mathbb{R}^d$ with its canonical Euclidean metric.

a.) Compute the rank and kernel of

$$M_{\mu\nu}(k) := k_\mu k_\nu - \delta_{\mu\nu}\vec{k}^2 \tag{7.12}$$

In gauge theory, when we have restricted to the space of planewaves $A_\mu(k)e^{\mathrm{i}k\cdot x}$ the kernel of $M_{\mu\nu}(k)$ represent the gauge modes. The quotient represents the gauge invariant information. For computations it is preferable to work with a subspace rather than a quotient, so we choose a gauge to invert the propagator. Of course, the complementary space is not unique. That is why there is a choice of gauge.

b.) Compute an inverse on the orthogonal complement (in the Euclidean metric) of the kernel.

---

[13] Nielsen and Chuang, *Quantum Computation and Quantum Information*, Cambridge, section 10.4.1

Exercise
Show that if $T : V \to W$ is *any* linear operator then there is an exact sequence

$$0 \to \ker T \to V \stackrel{T}{\to} W \to \mathrm{cok} T \to 0 \tag{7.13}$$

7.1 The index of a linear operator

If $T : V \to W$ is any linear operator such that $\ker T$ and $\mathrm{cok} T$ are finite dimensional we define the *index of the operator* T to be:

$$\mathrm{Ind}(T) := \dim\mathrm{cok} T - \dim\ker T \tag{7.14}$$

For V and W finite dimensional vector spaces you can easily show that

$$\mathrm{Ind} T = \dim W - \dim V \tag{7.15}$$

Notice that, from the LHS, we see that it does not depend on the details of T!

As an example, consider the *family* of linear operators $T_\lambda : v \to \lambda v$ for $\lambda \in \mathbb{C}$. Note that for $\lambda \neq 0$

$$\ker T = \{0\} \qquad \mathrm{cok} T = V/V \cong \{0\} \tag{7.16}$$

but for $\lambda = 0$

$$\ker T = \mathrm{cok} T = V \tag{7.17}$$

Both the kernel and cokernel change, but the index remains *invariant*.

As another example consider the family of operators $T_\lambda : \mathbb{C}^2 \to \mathbb{C}^2$ given in the standard basis by:

$$T_\lambda = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \tag{7.18}$$ eq:fmly

Now, for $\lambda \neq 0$

$$\ker T_\lambda = \{0\} \qquad \mathrm{cok} T_\lambda = \mathbb{C}^2/\mathbb{C}^2 \cong \{0\} \tag{7.19}$$

but for $\lambda = 0$

$$\ker T_0 = \mathbb{C} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathrm{cok} T_0 = \mathbb{C}^2/\mathbb{C} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{7.20}$$

and again the index is invariant.

In infinite dimensions one must be more careful. Clearly, one cannot define the index to be $\dim W - \dim V$. However, there is a special class of operators on infinite-dimensional Hilbert spaces for which $\dim\mathrm{cok}(T)$ and $\dim\ker(T)$ are in fact finite-dimensional. These, (with some extra technical constraints) are known as *Fredholm operators*. The index of a Fredholm operator is then defined as $\dim\mathrm{cok}(T) - \dim\ker(T)$ and it turns out to be a very interesting object. One important theorem in index theory is then that if T_λ is a *continuous* family of Fredholm operators (where one defines a topology on the space of Fredholm operators using the norm topology from operator algebra theory) then the index is continuous, and hence, being an integer, is independent of λ.

The index of an operator plays an essential role in modern mathematical physics. To get an idea of why it is important, notice that in the finite-dimensional case, the RHS of the above formula for the index does not refer to the details of T, and yet provides some information on the number of zero eigenvalues of T and $T^\dagger$. One of the great achievements of 20th century mathematics is the *Atiyah-Singer index theorem* and its many variants. A special case of the theorem gives the index for important geometrical operators such as the Dirac operator coupled to metrics and gauge fields defined on general (compact) manifolds. The zero-modes of these operators often have great physical significance, but solving the Dirac equation explicitly is completely out of the question. The index theorem expresses the index in terms of certain topological invariants, and these topological invariants are often readily computable.

Exercise

Compute the index of the family of operators:

$$T_u = \begin{pmatrix} u & u & u^2 \\ \sin(u) & \sin(u) & \sin^2(u) \end{pmatrix} \tag{7.21}$$

Find special values of u where the kernel and cokernel jump.

Exercise

Consider the usual harmonic oscillator operators $a, a^\dagger$ acting on a separable Hilbert space. (See below.) Compute the index of a and $a^\dagger$.

Consider the families of operators $T = \lambda a$ and $T = \lambda a^\dagger$ for $\lambda \in \mathbb{C}$. Is the index invariant?

References on the Atiyah-Singer theorem:

For physicists: Eguchi, Gilkey, Hanson, Physics Reports; Nakahara.

For mathematicians: Michelson and Lawson, *Spin Geometry*

8. A Taste of Homological Algebra

****

NEED SOME INTRO. COHOMOLOGY AND HOMOLOGY MEASURE OBSTRUCTIONS. SOME HOMOLOGICAL ALGEBRA HAS APPEARED IN PHYSICS IN THE PAST 20 YEARS.

There are many, many, textbooks on homological algebra. It is also generally covered in textbooks on algebraic topology. Among the many, one of the clearest and easiest to read is

J.W. Vick, *homology theory: an introduction to algebraic topology*, Academic Press

****

A module M over a ring R is said to be $\mathbb{Z}$-*graded* if it is a direct sum over modules

$$M = \oplus_{n\in\mathbb{Z}} M^n \tag{8.1}$$

Physicists should think of the grading as a charge of some sort in some kind of "space of quantum states."

If the ring is $\mathbb{Z}$-graded then also

$$R = \oplus_{n\in\mathbb{Z}} R^n \tag{8.2}$$

and moreover $R^n \cdot R^m \subset R^{n+m}$. In this case it is understood that the ring action on the module is also $\mathbb{Z}$-graded so

$$R^n \times M^{n'} \to M^{n+n'} \tag{8.3}$$

****

EXAMPLES:

1. One good example are the differential forms on a manifold. The ring can be taken to be the ring of smooth functions on a manifold X and $M^n = \Omega^n(X)$ are the smooth differential forms on X.

****

Definition Let M_1, M_2 be two $\mathbb{Z}$-graded R-modules. We say that an R-module homomorphism $\Phi : M_1 \to M_2$ is *of degree* k if

$$\Phi : M_1^n \to M_2^{n+k} \tag{8.4}$$

for all $n \in \mathbb{Z}$.

Put differently, we can introduce an operator $\mathcal{C}_M$ which acts by multiplication by n on M^n. In physical examples, it might be some kind of conserved charge on spaces of states. For example, n might be "fermion number" of some quantum states. Then $\Phi : M_1 \to M_2$ is of degree k if

$$C_{M_2}\Phi - \Phi\mathcal{C}_{M_1} = k\Phi \tag{8.5}$$

****

EXAMPLES

****

A *cochain complex* is a $\mathbb{Z}$-graded module together with a linear operator usually called $d : M \to M$ or $Q : M \to M$ where $Q^2 = 0$ and Q is "of degree one." What this means is that Q increases the degree by one, so Q take M^n into M^{n+1}. A degree one operator d that squares to 0 is often called a *differential*.

A cochain complex is usually indicated as a sequence of modules M^n and linear maps d^n:

$$\cdots \longrightarrow M^{n-1} \xrightarrow{d^{n-1}} M^n \xrightarrow{d^n} M^{n+1} \longrightarrow \cdots \tag{8.6}$$

with the crucial property that

$$\forall n \in \mathbb{Z} \qquad d^n d^{n-1} = 0 \tag{8.7}$$

The sequence might be infinite or finite on either side. If it is finite on both sides then $M^n = 0$ for all but finitely many n.

It is important to stress that the cochain complex is *not* an exact sequence. Indeed, we can characterize how much it differs from being an exact sequence by introducing its *cohomology*:

$$H(M, d) := \ker d / \mathrm{im} d \tag{8.8}$$

This definition makes sense, because $d^2 = 0$. Because the complex is $\mathbb{Z}$-graded, and d has degree one the cohomology is also $\mathbb{Z}$-graded and we can define

$$H^n(M, d) := \ker d^n / \mathrm{im} d^{n-1} \tag{8.9}$$

Thus, the cohomology of a cochain complex is a $\mathbb{Z}$-graded Abelian group.

Similarly, there is a notion of a *chain complex* and its corresponding *homology*. Now we have a $\mathbb{Z}$-graded module M over R and now we have an operator, again called the differential and denoted ∂ of degree -1 which squares to zero $\partial^2 = 0$. The only difference from the cochain complex is that the degree is -1 and not $+1$. Denoting the components of M of degree n by M_n we now have

$$\cdots \longleftarrow M_{n-1} \xleftarrow[\partial_n]{} M_n \xleftarrow[\partial_{n+1}]{} M_{n+1} \longleftarrow \cdots \tag{8.10}$$

and again $\partial_n \partial_{n+1} = 0$. The *homology of the chain complex* is the Abelian group $\ker\partial / \mathrm{im}\partial$. Again it is $\mathbb{Z}$-graded, and the component of degree n is

$$H_n(M_\bullet, \partial) = \ker\partial_n / \mathrm{im}\partial_{n+1} \tag{8.11}$$

Given a cochain complex (M^n, d) one can always take the dual complex to obtain a corresponding *chain complex*. We define $M_n := \mathrm{Hom}(M^n, R)$ and $\partial := d^\vee$. Similarly, given a chain complex we can produce a cochain complex. We discuss more about the duality between chain and cochain complexes in section *** below.

Examples

1. Continuing with our example of differential forms. $d : \Omega^k(X) \to \Omega^{k+1}(X)$ is the exterior derivative.

2. To an Abelian group A we can associate (nonuniquely) a cochain complex **** so that A is isomorphic to the homology of this complex

3. Algebraic topology: singular (co-)homology.

4. DeRham cohomology

5. CW homology.

6. Quiver representations.

7. Supersymmetric quantum mechanics and Morse theory.

8. Derived categories in theories of D-branes

8.1 The Euler-Poincaré principle

Suppose we have a finite cochain complex of vector spaces which we will assume begins at degree zero. Thus

$$0 \longrightarrow V^0 \xrightarrow{d^0} V^1 \xrightarrow{d^1} \cdots \xrightarrow{d^{n-1}} V^n \longrightarrow 0 \tag{8.12}$$

with $d^{j+1}d^j = 0$.

Let us compare the dimensions of the spaces in the cochain complex to the dimensions of the cohomology groups. We do this by introducing a *Poincare polynomial*

$$P_{V^\bullet}(t) = \sum_{j=0}^{n} t^j \mathrm{dim} V^j \tag{8.13}$$

$$P_{H^\bullet}(t) = \sum_{j=0}^{n} t^j \mathrm{dim} H^j \tag{8.14}$$

Now we claim that

$$P_{V^\bullet}(t) - P_{H^\bullet}(t) = (1+t)W(t) \tag{8.15}$$

eq:MorseIn

where $W(t)$ is a polynomial in t with nonnegative integer coefficients.

Proof: Let h^j be the dimension of the cohomology. Then by choosing representatives and complementary spaces we can find subspaces $W^j \subset V^j$ of dimension w^j so that

♣Need to explain more about these w^j. ♣

$$\begin{aligned}
\mathrm{dim} V^0 &= h^0 + w^0 \\
\mathrm{dim} V^1 &= h^1 + w^0 + w^1 \\
\mathrm{dim} V^2 &= h^2 + w^1 + w^2 \\
&\vdots \quad \vdots \\
\mathrm{dim} V^n &= h^n + w^{n-1}
\end{aligned} \tag{8.16}$$

so $W(t) = w^0 + w^1 t + \cdots + w^{n-1}t^{n-1}$.

Putting $t = -1$ we have the beautiful *Euler-Poincaré principle*:

$$\sum (-1)^i \mathrm{dim} V^i = \sum (-1)^i \mathrm{dim} H^i \tag{8.17}$$

eq:EulerChar

This common integer is called the *Euler characteristic* of the complex.

Remark: These observations have their origin in topology and geometry. Equation (8.17) is related, among many other things, to the fact that one can compute topological invariants of topological spaces by counting simplices in a simplicial decomposition, or in other ways associated to (co)chain complexes. One way of getting such complexes, that is

closely related to physics is via *Morse theory*. In Morse theory, you put a manifold on a table as in figure

[NEED A FIGURE: RIEMANN SURFACE ON A TABLE]

If one considers a generic function $h : M \to \mathbb{R}$ on a compact manifold one can look at the *critical points*. These are points $p \in M$ where $dh(p) = 0$. If all the critical points are nondegenerate the one can speak of the number of independent directions along which h decreases. This is called the Morse index. It turns out that the free vector space generated by the critical points defines a complex - the Morse-Smale-Witten complex. Then equation (8.15) implies the *Morse inequalities*: The number N_j of critical points with j downward directions is always bigger than the Better number b_j: We have $N_j \geq b_j$. Nevertheless $\sum_j (-1)^j N_j = \sum_j (-1)^j b_j = \chi(M)$, the Euler character of M. In Witten's interpretation, the generators of the complex are the approximate ground states of a quantum system, while b_j measures the true number of ground states (of Fermion number j). The inequality $N_j \geq b_j$ has a very simple physical interpretation: Some approximate groundstates are lifted due to instanton effects. [14]

8.2 Chain maps and chain homotopies

Suppose (C, ∂_C) and (D, ∂_D) are two cochain complexes. A *chain map* is a homomorphism of R-modules, $f : C \to D$, of degree zero that commutes with the differentials. That is, the diagram:

$$\begin{array}{ccc} C & \xrightarrow{\partial_C} & C \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f} \\ D & \xrightarrow{\partial_D} & D \end{array} \tag{8.18}$$

commutes. Even more explicitly: "degree zero" means $f : C_n \to D_n$ is a homomorphism of R-modules, for every n, that is, if preserves the degree and moreover $f\partial_{C,n} = \partial_{D,n} f$.

This defines a "morphism of complexes"

Given a chain map, one automatically has a homomorphism of Abelian groups

$$f_* : H_*(C, \partial_C) \to H_*(D, \partial_D) \tag{8.19}$$

defined by

$$f_*([c]) := [f(c)]. \tag{8.20}$$

eq:InducedHomo

Of course, there are entirely analogous definitions for cochain complexes.

****

PULLBACK AND PUSHFORWARD

****

[14] The connection of Morse theory to supersymmetric quantum mechanics was discovered by Witten in a classic paper, "Supersymmetry and Morse theory." Witten's interpretation has had enormous impact on both physics and mathematics subsequently. For a recent exposition of the main ideas, together with some extra details not spelled out in the original paper the reader might wish to consult chapter 10 of https://arxiv.org/pdf/1506.04087.pdf.

Exercise
Show that (8.20) is well-defined.

Definition Suppose that $f_1, f_2 : C \to D$ are chain maps. We say that an R-module homomorphism $T : C \to D$ of degree -1 so that

$$d_D T - T d_C = f_1 - f_2 \tag{8.21}$$

is a *chain homotopy between f_1 and f_2*

In diagrams we have

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & C^{n-1} & \xrightarrow{d_C^{n-1}} & C^n & \xrightarrow{d_C^n} & C^{n+1} & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle f_i} & \swarrow{\scriptstyle T} & \downarrow{\scriptstyle f_i} & \swarrow{\scriptstyle T} & \downarrow{\scriptstyle f_i} & \swarrow{\scriptstyle T} & \downarrow{\scriptstyle f_i} \\ \cdots & \longrightarrow & D^{n-1} & \xrightarrow[d_D^{n-1}]{} & D^n & \xrightarrow[d_D^n]{} & D^{n+1} & \longrightarrow & \cdots \end{array} \tag{8.22}$$

A key fact is that if there exists a chain homotopy between f_1 and f_2 then, on the homology, we have $(f_1)_* = (f_2)_*$. After all

$$\begin{aligned} (f_1)_*([c]) &:= [f_1(c)] \\ &= [f_2(c) + d_D T(c) - T(d_C(c))] \\ &= [f_2(c)] \end{aligned} \tag{8.23}$$

where we pass to the third line because $d_C(c) = 0$ and $[x + d_D(y)] = [x]$ for any x and y.

An important special case of the above is where $C = D$ and $f_1 = Id$, or indeed any invertible map. and $f_2 = 0$. Then, it follows that the cohomology vanishes. This is a very useful technique for proving that obstructions vanish.

GIVE EXAMPLE.

REMARKS: NEED TO EXPLAIN WHAT THIS HAS TO DO WITH HOMOTOPY.

EXAMPLES:

8.3 Exact sequences of complexes

Define SES of complexes. Then give the LES and describe the connecting homomorphism.

8.4 Left- and right-exactness

♣Probably should generalize to R-modules. Then need to talk about resolutions.... ♣

Suppose [15] that A, B, C are Abelian groups in an exact sequence:

$$0 \longrightarrow A \xrightarrow{\iota} B \xrightarrow{\pi} C \longrightarrow 0 \tag{8.24}$$

and let G be any Abelian group then we can ask what happens to the exact sequence when we apply the functors $A \mapsto \mathrm{Hom}(A, G)$ and $A \mapsto A \otimes G$. In general the exactness of the sequence is only partially preserved.

[15] We are following a nice brief discussion in Bott and Tu, *Differential Forms in Algebraic Topology*, Springer GTM 82

We claim that

$$0 \longrightarrow \mathrm{Hom}(C,G) \xrightarrow{\pi^*} \mathrm{Hom}(B,G) \xrightarrow{\iota^*} \mathrm{Hom}(A,G) \tag{8.25}$$

eq:HomSeq

Here, if $f : A \to B$ is a group homomorphism we can define $f^* : \mathrm{Hom}(B,G) \to \mathrm{Hom}(A,G)$ by $f^*(\phi) := \phi \circ f$. Note carefully that the map ι^* is not necessarily surjective so we are missing an arrow $\to 0$ at the end of (8.25). In the jargon we say that *the functor* $A \to \mathrm{Hom}(A,G)$ *is left-exact.*

Similarly, we claim that:

$$A \otimes G \xrightarrow{\iota\otimes 1} B \otimes G \xrightarrow{\pi\otimes 1} C \otimes G \longrightarrow 0 \tag{8.26}$$

eq:TensSeq

is exact, but again, note that we do not necessarily have injectivity of $\iota \otimes 1$. In the jargon we say that *the functor* $A \to A \otimes G$ *is right-exact.*

Ext and *Tor* are functors that measure the failure of the exactness of the above sequences.

Let C be any Abelian group. Then it has a free resolution:

♣Maybe for Tor we should give injective resolution? ♣

$$0 \longrightarrow R \xrightarrow{\iota} F \xrightarrow{\pi} C \longrightarrow 0 \tag{8.27}$$

where F is a free Abelian group. We just choose a set of generators for C and let F be the free Abelian group on those generators. Then [see Jacobson, sec. 3.6] any subgroup of a free Abelian group is free so R is a free Abelian group as well. Now we apply the above discussion. We define

$$\mathrm{Ext}(C,G) := \mathrm{cok}(\iota^*) = \mathrm{Hom}(R,G)/\mathrm{im}\iota^* \tag{8.28}$$

$$\mathrm{Tor}(C,G) := \ker(\iota \otimes 1) \tag{8.29}$$

Here is how this works in one of the simplest cases. We take $C = \mathbb{Z}/m\mathbb{Z}$. So we can take $R = \mathbb{Z}$ and $F = \mathbb{Z}$ but $\iota : R \to F$ is multiplication by m, that is $\iota(x) = mx$. Now let $G = \mathbb{Z}$. Then $\mathrm{Hom}(C,G) = 0$ and $\mathrm{Hom}(F,G) = \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) \cong \mathbb{Z}$ and $\mathrm{Hom}(R,G) = \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) \cong \mathbb{Z}$. Now, what is ι^* ? It turns out to be multiplication by m once again, so the cokernel is $\mathbb{Z}/m\mathbb{Z}$, so

$$\mathrm{Ext}(\mathbb{Z}_m,\mathbb{Z}) \cong \mathbb{Z}_m \tag{8.30}$$

In this way, with a little patience, you can prove

$$\begin{aligned} \mathrm{Ext}(\mathbb{Z}/m\mathbb{Z},\mathbb{Z}) &\cong \mathbb{Z}/m\mathbb{Z} \\ \mathrm{Ext}(\mathbb{Z},G) &= 0 \\ \mathrm{Ext}(\mathbb{Z}/m\mathbb{Z},\mathbb{Z}/n\mathbb{Z}) &\cong \mathbb{Z}/gcd(m,n)\mathbb{Z} \\ \mathrm{Ext}(\mathbb{Z}/n\mathbb{Z},G) &\cong G/nG \end{aligned} \tag{8.31}$$

****

Explain how Ext(A,G) classifies extensions. Hence the name.

Incorporate remarks from Moore-Segal 2.7.1; Bredon pp. 271-280.

****

Let us consider a simple example of how Tor can arise. Again, we take $C = \mathbb{Z}/m\mathbb{Z}$, so $R = \mathbb{Z}$ and $F = \mathbb{Z}$ and $\iota : R \to F$ is multiplication by m, that is $\iota(x) = mx$. Now taking $G = \mathbb{Z}/n\mathbb{Z}$, equation (8.26) becomes

$$\mathbb{Z} \otimes \mathbb{Z}_n \xrightarrow{\times m \otimes 1} \mathbb{Z} \otimes \mathbb{Z}_n \xrightarrow{\pi \otimes 1} \mathbb{Z}_m \otimes \mathbb{Z}_n \longrightarrow 0 \tag{8.32}$$

eq:Simple-Tor-

So we need to work out the kernel of the homomorphism $\mathbb{Z}_n \to \mathbb{Z}_n$ given by $\phi(x) = mx$. That is, we should solve for x in

$$mx = 0 \bmod n \tag{8.33}$$

If $n = g \cdot p$ and $m = g \cdot q$ with $(p, q) = 1$ then this is equivalent to

$$qx = 0 \text{mod} p \tag{8.34}$$

but q is invertible modulo p so $x = 0 \text{mod} p$ so the kernel is

$$\{p + n\mathbb{Z}, 2p + n\mathbb{Z}, \ldots, p \cdot \frac{n}{p} + n\mathbb{Z}\} \tag{8.35}$$

and so $\ker(\iota \otimes 1) \cong \mathbb{Z}/g\mathbb{Z}$ where $g = gcd(m, n)$, and

$$\text{Tor}(\mathbb{Z}_n, \mathbb{Z}_m) \cong \mathbb{Z}_{(n,m)} \tag{8.36}$$

One can show that for Tor we have:

$$\begin{aligned} \text{Tor}(A, G) &\cong \text{Tor}(G, A) \\ \text{Tor}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}) &= 0 \\ \text{Tor}(\mathbb{Z}, \mathbb{Z}/m\mathbb{Z}) &= 0 \\ \text{Tor}(\mathbb{Z}, \mathbb{Z}) &= 0 \\ \text{Tor}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) &\cong \mathbb{Z}/gcd(m, n)\mathbb{Z} \end{aligned} \tag{8.37}$$

Exercise $A \to \text{Hom}(A, G)$ *is contravariant and left-exact*
Prove (8.25). [16]

Exercise $A \to A \otimes G$ *is covariant and right-exact*

[16] *Answer* First, note that if $\pi^*(\phi) = 0$ then $\phi(\pi(b)) = 0$ for every $b \in B$. But every element $c \in C$ is of the form $\pi(b)$ for some b therefore for every $c \in C$ we have $\phi(c) = 0$ therefore $\phi = 0$. Therefore π^* is injective. Next, if $\iota^*(\phi) = 0$ for $\phi \in \text{Hom}(B, G)$ then $\phi(\iota(a)) = 0$ for every $a \in A$. Then we can choose a section (not a splitting) $s : C \to B$. Note that we can define $\tilde{\phi} \in \text{Hom}(C, G)$ by $\tilde{\phi}(c) = \phi(s(c))$. Of course, two sections will differ by $s(c) = s'(c) + \iota(f(c))$ for some function $f : C \to A$, but since $\phi(\iota(f(c)) = 0$ the ambiguity does not matter and hence $\phi = \pi^*(\tilde{\phi})$.

Prove equation (8.26). [17]

Exercise *Divisible Groups*

Definition: A group G is said to be *divisible* if for every $g \in G$ and every nonzero integer n there is a g' with $ng' = g$. Sometimes, such groups are said to be *injective*.

Thus, for example, $\mathbb{R}$ and $U(1)$ are divisible, but $\mathbb{Z}$ is not.

♣Need to write out answers for this. ♣

a.) Show that if

$$0 \to G \to I \to J \to 0 \tag{8.38}$$

with I, J divisible groups then for a finite abelian group A:

$$0 \to Hom(A,G) \to Hom(A,I) \to Hom(A,J) \to Ext(A,G) \to 0 \tag{8.39}$$

b.) Apply part (a) to $0 \to \mathbb{Z} \to \mathbb{R} \to U(1) \to 0$ to obtain $\mathrm{Ext}(A,\mathbb{Z}) \cong \mathrm{Hom}(A, U(1))$, the Pontryagin dual of A.

c.) Show that if G is divisible then $Ext(A,G) = 0$ for all abelian groups A.

---

9. Relations Between Real, Complex, And Quaternionic Vector Spaces

ec:RealComplex

Physical quantities are usually expressed in terms of real numbers. Thus, for example, we think of the space of electric and magnetic fields in terms of real numbers. In quantum field theory more generally one often works with real vector spaces of fields. On the other hand, quantum mechanics urges us to use the complex numbers. One could formulate quantum mechanics using only the real numbers, but it would be terribly awkward to do so. Quantum mechanics teaches us that complex vector spaces are a fundamental part of reality. In physics and mathematics it is often important to have a firm grasp of the relation between complex and real structures on vector spaces, and this section explains that relation in excruciating detail.

9.1 Complex structure on a real vector space

Definition Let V be a real vector space. A *complex structure* on V is an $\mathbb{R}$-linear map $I : V \to V$ such that $I^2 = -1$.

Choose a squareroot of -1 and denote it by i. If V is a real vector space with a complex structure I, then we can define an associated complex vector space, which we will

[17] *Answer*: Choose a section $s : C \to B$ for π. Then any element $\sum_i c_i \otimes g_i \in C \otimes G$ is in the image of $\sum_i s(c_i) \otimes g_i \in B \otimes G$. Now suppose that $(\pi \otimes 1)(\sum_i b_i \otimes g_i) = \sum_i \pi(b_i) \otimes g_i = 0$. Then, on the one hand $s(\pi(b_i)) - b_i = \iota(a_i)$ for some a_i and hence $\sum_i s(\pi(b_i)) \otimes g_i = \sum_i \iota(a_i) \otimes g_i + \sum_i b_i \otimes g_i$. But $\sum_i s(\pi(b_i)) \otimes g_i = (s \otimes 1)(\sum_i \pi(b_i) \otimes g_i) = 0$. Therefore $\sum_i b_i \otimes g_i = -\sum_i \iota(a_i) \otimes g_i = (\iota \otimes 1)(-\sum_i a_i \otimes g_i)$.

denote by (V, I). We take (V, I) to be identical with V, *as sets*, but define the scalar multiplication of a complex number $z \in \mathbb{C}$ on a vector v by

$$z \cdot v := \alpha \cdot v + I(\beta \cdot v) = \alpha \cdot v + \beta \cdot I(v) \tag{9.1}$$

where $z = \alpha + \mathrm{i}\beta$ with $\alpha, \beta \in \mathbb{R}$, and we are stressing scalar multiplication on vectors by putting in a $\cdot$. We will usually omit the $\cdot$.

Remark: If V_1 and V_2 are real vector spaces with complex structures I_1 and I_2 then a complex linear map

$$T : (V_1, I_1) \to (V_2, I_2) \tag{9.2}$$

is a real linear map $T : V_1 \to V_2$ such that

$$T \circ I_1 = I_2 \circ T \ . \tag{9.3}$$ eq:C-linear-co

We now come to an important point:

A finite-dimensional real vector space admits a complex structure iff it is even dimensional.

To prove this, note that for any nonzero vector $v \in V$, the vectors v and $I(v)$ are linearly independent over $\mathbb{R}$. To prove this, suppose that on the contrary there are nonzero scalars $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha v + \beta I(v) = 0 \tag{9.4}$$ eq:ab-1

Then applying I to this equation and using $I^2 = -1$ we get

$$\beta v - \alpha I(v) = 0 \tag{9.5}$$ eq:ab-2

Multiply (9.4) by α and (9.5) by β and add the equations to get

$$(\alpha^2 + \beta^2) v = 0 \tag{9.6}$$

Since $v \neq 0$ and α, β are real we learn that $\alpha = \beta = 0$. It follows that if V is finite dimensional then its dimension must be even and there is always a real basis for V in which I takes the form

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{9.7}$$

where the blocks are $n \times n$ for an integer $n = \frac{1}{2}\dim_{\mathbb{R}} V$. Conversely, if V is even dimensional consider the above matrix with respect to any basis.

Note that, if I is a complex structure on V, then so is SIS^{-1} for any invertible real linear map $S : V \to V$. In fact, all the complex structures are related to the standard one above by a change of basis:

Lemma If I is any $2n \times 2n$ *real* matrix which squares to -1_{2n} then there is $S \in GL(2n, \mathbb{R})$ such that

$$SIS^{-1} = I_0 := \begin{pmatrix} 0 & -1_n \\ 1_n & 0 \end{pmatrix} \tag{9.8}$$ eq:CanonCS

We leave the proof as an exercise below.

Note carefully that, while v and $I(v)$ are linearly independent in the real vector space V, they are *linearly dependent* in the complex vector space (V, I) since

$$\mathrm{i} \cdot v + (-1) \cdot I(v) = 0 \tag{9.9}$$

Indeed, our lemma also shows that if V is finite dimensional and has a complex structure then the dimension of the complex vector space (V, I) is:

$$\dim_{\mathbb{C}}(V, I) = \frac{1}{2}\dim_{\mathbb{R}} V \tag{9.10}$$ eq:half-dim

Example Consider the real vector space $V = \mathbb{R}^2$. Let us choose

$$I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \ . \tag{9.11}$$

Then multiplication of the complex scalar $z = x + iy$, with $x, y \in \mathbb{R}$ on a vector $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \in \mathbb{R}^2$ can be defined by:

$$(x + iy) \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} := \begin{pmatrix} a_1 x - a_2 y \\ a_1 y + a_2 x \end{pmatrix} \tag{9.12}$$ eq:cplxstrone

By equation (9.10) this must be a one-complex dimensional vector space, so it should be isomorphic to $\mathbb{C}$ as a complex vector space. Indeed this is the case. Define $\Psi : (V, I) \to \mathbb{C}$ by

$$\Psi : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \mapsto a_1 + i a_2 \tag{9.13}$$

Then one can check (exercise!) that this is an isomorphism of complex vector spaces. The main point to check is that $\Psi \circ I = \mathrm{i}\Psi$ (this is the condition (9.3) above).

Quite generally, if I is a complex structure then so is $\tilde{I} = -I$. So what happens if we take our complex structure to be instead:

$$\tilde{I} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad ? \tag{9.14}$$

Now the rule for multiplication by a complex number in $(V, \tilde{I})$ is

$$(x + iy) \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} := \begin{pmatrix} a_1 x + a_2 y \\ -a_1 y + a_2 x \end{pmatrix} \tag{9.15}$$ eq:cplxstronep

One can check that $\tilde{\Psi} : (V, \tilde{I}) \to \mathbb{C}$ defined by

$$\tilde{\Psi} : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \mapsto a_1 - ia_2 \tag{9.16}$$

is also an isomorphism of complex vector spaces. (Check carefully that $\tilde{\Psi}(z\vec{a}) = z\tilde{\Psi}(\vec{a})$.)

Now, up to isomorphism there is only one one-dimensional complex vector space, so there must be an isomorphism of (V, I) with $(V, \tilde{I})$ as complex vector spaces. We now describe it carefully: Note that if we introduce the real linear operator

$$C := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{9.17}$$

then $C^2 = 1$ and

$$CIC^{-1} = CIC = -I \tag{9.18}$$

Note that C does not define a $\mathbb{C}$-linear transformation $(V, I) \to (V, I)$. Rather, if we define $\mathcal{C} : (V, I) \to (V, \tilde{I})$ by

$$\mathcal{C} : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \to \begin{pmatrix} a_1 \\ -a_2 \end{pmatrix} \tag{9.19}$$

then you can check that

$$\mathcal{C} \circ I = \tilde{I} \circ \mathcal{C} \tag{9.20}$$

meaning that $\mathcal{C}$ is an isomorphism of the two complex vector spaces. (Recall equation (9.3).)

What can we say about the set of all complex structures on $\mathbb{R}^2$? We have already seen below that there is a transitive action of $GL(n, \mathbb{R})$ on the space of complex structures by (matrix) conjugation. A useful theorem (see remark below) says that every $S \in GL(2, \mathbb{R})$ can be written as:

$$S = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\xi \sin(\theta) & \xi \cos(\theta) \end{pmatrix} \tag{9.21}$$

eq:KAN-GL2

with $x \in \mathbb{R}$, $\lambda_1, \lambda_2 > 0$, $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ and, finally, $\xi \in \{\pm 1\}$ is a sign that coincides with the sign of $\det(S)$ and tells us which of the two connected components of $GL(2, \mathbb{R})$ S sits in. We then compute that

$$\begin{aligned} SI_0S^{-1} &= \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \xi\lambda_1/\lambda_2 \\ -\xi\lambda_2/\lambda_1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -x\alpha & \alpha^{-1} + x^2\alpha \\ -\alpha & x\alpha \end{pmatrix} \end{aligned} \tag{9.22}$$

(where $\alpha = \xi\lambda_2/\lambda_1$ so $\alpha \in \mathbb{R}^*$ and $x \in \mathbb{R}$). The most general complex structure is uniquely written in this form, and so the space of complex structures is a two-dimensional manifold with two connected components.

In section 9.6 below we describe the generalization of this result to the space of complex structures on any finite dimensional real vector space.

Remarks:

1. The decomposition (9.21) is often very useful. Indeed, it generalizes to $GL(n,\mathbb{R})$, where it says we can write any $g \in GL(n,\mathbb{R})$ as $g = nak$ where $k \in O(n)$, a is diagonal with positive entries, and $n = 1 + t$ where t is strictly upper triangular. This is just a statement of the Gram-Schmidt process. See section 21.2 below. This kind of decomposition generalizes to all algebraic Lie groups, and it is this more general statement that constitutes the "KAN theorem."

2. It is often useful to define complex structures *compatible with a metric*. "Compatible" means that $\| I(v) \|^2 = \| v \|^2$. In other words, I is an orthogonal transformation. For the example of $\mathbb{R}^2$ we take the standard Euclidean norm:

$$\| \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \|^2 := a_1^2 + a_2^2 \tag{9.23}$$

   Then, $I \in O(2)$. One easily checks, by explicit multiplication, that if S is an orthogonal transformation then

$$SI_0S^{-1} = SI_0S^{tr} = (\det S)I_0 \tag{9.24}$$

   so there are precisely two complex structures compatible with the Euclidean norm: I_0 and $-I_0$.

Exercise *Canonical form for a complex structure*
Prove equation (9.8) above. [18]

[18] *Answer*: Proceed by induction. If $V \neq 0$ then there is a nonzero vector $v \in V$ and hence $\langle v, I(v)\rangle \subset V$ is a nonzero two-dimensional subspace. If $\dim_{\mathbb{R}} V = 2$ we are done. If $\dim_{\mathbb{R}} V = 2n+2$ and we assume the result is true up to dimension $2n$ then we consider the quotient space

$$V/\langle v, I(v)\rangle \tag{9.25}$$

and we note that I descends to a map $\tilde{I}$ on this quotient space and by the induction hypothesis there is a basis $[w_1], \ldots, [w_{2n}]$ for $V/\langle v, I(v)\rangle$ such that

$$\begin{aligned} \tilde{I}([w_i]) &= [w_{n+i}] \\ \tilde{I}([w_{n+i}]) &= -[w_i] \qquad i = 1, \ldots, n \end{aligned} \tag{9.26}$$

Now, choosing specific representatives w_i we know there are scalars $\alpha_1^i, \ldots, \beta_2^i$ such that

$$\begin{aligned} \tilde{I}(w_i) &= w_{n+i} + \alpha_1^i v + \beta_1^i I(v) \\ \tilde{I}([w_{n+i}]) &= -w_i + \alpha_2^i v + \beta_2^i I(v) \qquad i = 1, \ldots, n \end{aligned} \tag{9.27}$$

and consistency of these equations with $I^2 = -1$ implies $\alpha_2^i = \beta_1^i$ and $\beta_2^i = -\alpha_1^i$. Now check that $\tilde{w}_i := w_i + \alpha_1^i I(v)$ and $\tilde{w}_{n+i} = w_{n+i} + \beta_1^i I(v)$ is the suitable basis in which I takes the desired form.

9.2 Real Structure On A Complex Vector Space

Given a complex vector space V can we produce a real vector space? Of course, by restriction of scalars, if V is complex, then it is also a real vector space, which we can call $V_{\mathbb{R}}$. V and $V_{\mathbb{R}}$ are the same as *sets* but in $V_{\mathbb{R}}$ the vectors v and iv, are linearly independent (they are clearly not linearly independent in V!). Thus:

$$\dim_{\mathbb{R}} V_{\mathbb{R}} = 2\dim_{\mathbb{C}} V. \tag{9.28}$$

There is another way we can get real vector spaces out of complex vector spaces. A *real structure* on a complex vector V space produces a different real vector space of half the real dimension of $V_{\mathbb{R}}$, that is, a vector space of real dimension equal to the complex dimension of V.

Definition Let V_1, V_2 be complex vector spaces. An *antilinear* map $\mathcal{T} : V_2 \to V_2$ is a map that satisfies:

1. $\mathcal{T}(v + v') = \mathcal{T}(v) + \mathcal{T}(v')$,
2. $\mathcal{T}(\alpha v) = \alpha^* \mathcal{T}(v)$ where $\alpha \in \mathbb{C}$ and $v \in V_1$.

Note that $\mathcal{T}$ is a linear map between the underlying real vector spaces $(V_1)_{\mathbb{R}}$ and $(V_2)_{\mathbb{R}}$.

Remark: Antilinear operators famously appear in quantum mechanics when dealing with symmetries which reverse the orientation of time. In condensed matter physics they also appear as "particle-hole symmetry operators." (This is a little confusing since in relativistic particle physics the charge conjugation operator is complex linear.)

Definition Suppose V is a complex vector space. Then a *real structure* on V is an antilinear map $\mathcal{C} : V \to V$ such that $\mathcal{C}^2 = +1$.

If $\mathcal{C}$ is a real structure on a complex vector space V then we can define *real vectors* to be those such that

$$\mathcal{C}(v) = v \tag{9.29}$$

Let us call the set of such real vectors V_+. This set is a real vector space, but it is not a complex vector space, because $\mathcal{C}$ is antilinear. Indeed, if $\mathcal{C}(v) = +v$ then $\mathcal{C}(iv) = -iv$. If we let V_- be the *imaginary vectors*, for which $\mathcal{C}(v) = -v$ then we claim

$$V_{\mathbb{R}} = V_+ \oplus V_- \tag{9.30}$$

The proof is simply the isomorphism

$$v \mapsto \left(\frac{v + \mathcal{C}(v)}{2}\right) \oplus \left(\frac{v - \mathcal{C}(v)}{2}\right) \tag{9.31}$$

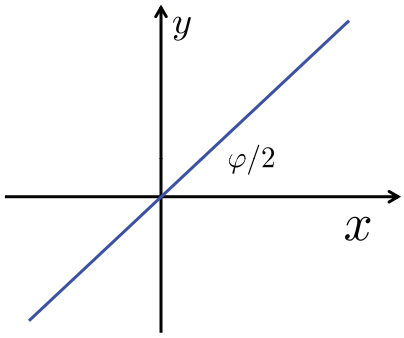


Figure 3: The real structure $\mathcal{C}$ has fixed vectors given by the blue line. This is a real vector space determined by the real structure $\mathcal{C}$.

fig:REALVECTOR

Moreover multiplication by i defines an isomorphism of real vector spaces: $V_+ \cong V_-$. Thus we have

$$\dim_{\mathbb{R}} V_+ = \dim_{\mathbb{C}} V \tag{9.32}$$

Example Take $V = \mathbb{C}$, and let $\varphi \in \mathbb{R}/2\pi\mathbb{Z}$ and define:

$$\mathcal{C} : x + iy \to e^{i\varphi}(x - iy) \tag{9.33}$$

The fixed vectors under $\mathcal{C}$ consist of the real line at angle $\varphi/2$ to the x-axis as shown in Figure 3. Note that the lines are not oriented so the ambiguity in the sign of $e^{i\varphi/2}$ does not matter.

Once again, note that there can be many distinct real structures on a given complex vector space. In our case, the space of distinct real structures is the space of real unoriented lines in $\mathbb{R}^2$ and this is known as the $\mathbb{RP}^1$. In general, the space of real structures is discussed in Section 9.6 below.

In general, a complex vector space V equipped with a basis (over $\mathbb{C}$) has a canonically associated real structure. Indeed suppose $\{v_i\}$ is a basis for V, then we can define the real structure:

$$\mathcal{C}(\sum_i z_i v_i) = \sum_i \bar{z}_i v_i \tag{9.34}$$

and thus

$$V_+ = \{\sum a_i v_i | a_i \in \mathbb{R}\} \tag{9.35}$$

Exercise *Antilinear maps from the real point of view*

Suppose W is a real vector space with complex structure I giving us a complex vector space (W, I).

Show that an antilinear map $\mathcal{T} : (W, I) \to (W, I)$ is the same thing as a real linear transformation $T : W \to W$ such that

$$TI + IT = 0 \tag{9.36}$$

That is, an anti-linear map, from the real point of view is just an $\mathbb{R}$-linear map T that anticommutes with I.

9.2.1 Complex Conjugate Of A Complex Vector Space

There is another viewpoint on what a real structure is which can be very useful. If V is a complex vector space then we can, canonically, define another complex vector space $\bar{V}$. We begin by declaring $\bar{V}$ to be the same *set*. Thus, for every vector $v \in V$, the same vector, regarded as an element of $\bar{V}$ is simply written $\bar{v}$. However, $\bar{V}$ is different from V *as a complex vector space* because we alter the vector space structure by altering the rule for scalar multiplication by $\alpha \in \mathbb{C}$:

$$\alpha \cdot \bar{v} := \overline{\alpha^* \cdot v} \tag{9.37}$$

eq:conjvs

where α^* is the complex conjugate in $\mathbb{C}$.

Of course $\overline{\bar{V}} = V$.

Note that, given any $\mathbb{C}$-linear map $T : V \to W$ between complex vector spaces there is, canonically, a $\mathbb{C}$-linear map

$$\bar{T} : \bar{V} \to \bar{W} \tag{9.38}$$

defined by

$$\bar{T}(\bar{v}) := \overline{T(v)} \tag{9.39}$$

With the notion of $\bar{V}$ we can give an alternative definition of an *anti-linear map*: An anti-linear map $\mathcal{T} : V \to V$ is the same as a $\mathbb{C}$-linear map $T : V \to \bar{V}$, related by

$$\mathcal{T}(v) = \overline{T(v)} \tag{9.40}$$

Similarly, we can give an alternative definition of a real structure on a complex vector space V as a $\mathbb{C}$- *linear* map

$$C : V \to \bar{V} \tag{9.41}$$

such that $C\bar{C} = 1$ and $\bar{C}C = 1$, where $\bar{C} : \bar{V} \to V$ is canonically determined by C as above. In order to relate this to the previous viewpoint note that $\mathcal{C} : v \mapsto \bar{C}(\bar{v})$ is an antilinear transformation $V \to V$ which squares to 1.

Exercise

A linear transformation $T : V \to W$ between two complex vector spaces with real structures C_V and C_W commutes with the real structures if the diagram

$$\begin{array}{ccc} V & \stackrel{T}{\to} & W \\ \downarrow C_V & & \downarrow C_W \\ \bar{V} & \stackrel{\bar{T}}{\to} & \bar{W} \end{array} \tag{9.42}$$

commutes.

Show that in this situation T defines an $\mathbb{R}$-linear transformation on the underlying real vector spaces: $T_+ : V_+ \to W_+$.

Exercise *Complex conjugate from the real point of view*

a.) Show that every complex vector space is isomorphic to a complex vector space of the form (W, I) where W is a real vector space and I is a complex structure on W.

b.) Suppose W is a real vector space with complex structure I so that we can form the complex vector space (W, I). Show that

$$\overline{(W, I)} = (W, -I) \tag{9.43}$$

9.2.2 Complexification

If V is a real vector space then we can define its *complexification* $V_{\mathbb{C}}$ by putting a complex structure on $V \oplus V$. This is simply the real linear transformation

$$I : (v_1, v_2) \mapsto (-v_2, v_1) \tag{9.44}$$

eq:cplx-def-1

and clearly $I^2 = -1$. This complex vector space $(V \oplus V, I)$ is known as the *complexification of* V. Another way to define the complexification of V is to take

$$V_{\mathbb{C}} := V \otimes_{\mathbb{R}} \mathbb{C}. \tag{9.45}$$

eq:cplx-def-2

Note that we are taking a tensor product of vector spaces over $\mathbb{R}$ to get a real vector space. Since $\mathbb{C}$ is two-dimensional as a real vector space $V_{\mathbb{C}}$ has twice the (real) dimension of V. But now $V \otimes_{\mathbb{R}} \mathbb{C}$ has a natural action of the complex numbers:

$$z \cdot (v \otimes z') := v \otimes zz' \tag{9.46}$$

making $V_{\mathbb{C}}$ into a complex vector space. In an exercise below you show that the two definitions of complexification we have just given are in fact equivalent.

Note that

$$\dim_{\mathbb{C}} V_{\mathbb{C}} = \dim_{\mathbb{R}} V \tag{9.47}$$

Note that $V_{\mathbb{C}}$ has a canonical real structure. Indeed

$$\overline{V_{\mathbb{C}}} = V \otimes_{\mathbb{R}} \overline{\mathbb{C}} \tag{9.48}$$

and we can define $C : V_{\mathbb{C}} \to \overline{V_{\mathbb{C}}}$ by setting

$$C : v \otimes 1 \mapsto v \otimes \bar{1} \tag{9.49}$$

and extending by $\mathbb{C}$-linearity. Thus

$$\begin{aligned} C(v\otimes z) &= C(z\cdot(v\otimes 1)) && \text{def of } V_{\mathbb{C}} \\ &= z\cdot C((v\otimes 1)) && \mathbb{C}-\text{linear extension} \\ &= z\cdot(v\otimes \bar{1}) \\ &= v\otimes \overline{z^*} && \text{definition of scalar action on} \bar{V}_{\mathbb{C}} \end{aligned} \tag{9.50}$$

Finally, it is interesting to ask what happens when one begins with a *complex* vector space V and then complexifies the underlying real space $V_{\mathbb{R}}$. If V is complex then we claim there is an isomorphism of complex vector spaces:

$$(V_{\mathbb{R}})_{\mathbb{C}} \cong V\oplus \bar{V} \tag{9.51}$$

eq:complxi

Proof: The vector space $(V_{\mathbb{R}})_{\mathbb{C}}$ is, by definition, generated by the space of pairs (v_1, v_2), $v_i \in V_{\mathbb{R}}$ with complex structure defined by $I : (v_1, v_2) \to (-v_2, v_1)$. Now we map:

$$\psi : (v_1, v_2) \mapsto (v_1 + iv_2)\oplus(v_1 - iv_2) \tag{9.52}$$

and compute

$$(x + Iy)\cdot(v_1, v_2) = (xv_1 - yv_2, xv_2 + yv_1) \tag{9.53}$$

so

$$\psi : z\cdot(v_1, v_2) \mapsto (x + iy)\cdot(v_1 + iv_2)\oplus(x - iy)\cdot(v_1 - iv_2) = z\cdot v + \bar{z}\cdot\bar{v} \tag{9.54}$$

Another way to look at (9.51) is as follows. Suppose the complex vector space V is of the form (W, I) with W a real vector space and I a complex structure on W. Now $(W, I)_{\mathbb{R}} \cong W$. Now consider $V_{\mathbb{R}}\otimes_{\mathbb{R}}\mathbb{C} = W\otimes_{\mathbb{R}}\otimes|IC$. There are now two ways of multiplying by a complex number $z = x + iy$: We can multiply the second factor $\mathbb{C}$ by z or we could operate on the first factor with $x+Iy$. We can decompose our space $V\otimes_{\mathbb{R}}\mathbb{C}$ into eigenspaces where $I = +i$ and $I = -i$ using the projection operators

$$P_{\pm} = \frac{1}{2}(1 \mp I\otimes i) \tag{9.55}$$

the images of these projection operators have complex structures, and are isomorphic to V and $\bar{V}$ as complex vector spaces, respectively:

$$((W, I)_{\mathbb{R}})_{\mathbb{C}} \cong (W, I)\oplus\overline{(W, I)} \tag{9.56}$$

Exercise *Equivalence of two definitions*

Show that the two definitions (9.44) and (9.45) define canonically isomorphic complex vector spaces.

Exercise
Show that

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} = \mathbb{C} \oplus \mathbb{C} \tag{9.57}$$

$$\mathbb{C} \otimes_{\mathbb{C}} \mathbb{C} \cong \mathbb{C} \tag{9.58}$$

as algebras.

Exercise
Suppose V is a complex vector space with a real structure C and that V_+ is the real vector space of fixed points of C.

Show that, as complex vector spaces

$$V \cong V_+ \otimes_{\mathbb{R}} \mathbb{C}. \tag{9.59}$$

9.3 The Quaternions

Definition The *quaternion algebra* $\mathbb{H}$ is the associative algebra over $\mathbb{R}$ with generators $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}$ satisfying the relations

$$\mathfrak{i}^2 = -1 \qquad \mathfrak{j}^2 = -1 \qquad \mathfrak{k}^2 = -1 \tag{9.60}$$ eq:quat-1

$$\mathfrak{k} = \mathfrak{i}\mathfrak{j} \tag{9.61}$$ eq:quat-2

Note that, as a consequence of these relations we have

$$\mathfrak{i}\mathfrak{j} + \mathfrak{j}\mathfrak{i} = \mathfrak{i}\mathfrak{k} + \mathfrak{k}\mathfrak{i} = \mathfrak{j}\mathfrak{k} + \mathfrak{k}\mathfrak{j} = 0 \tag{9.62}$$ eq:quat-3

The quaternions form a four-dimensional algebra over $\mathbb{R}$. As a vector space we can write

$$\mathbb{H} = \mathbb{R}\mathfrak{i} + \mathbb{R}\mathfrak{j} + \mathbb{R}\mathfrak{k} + \mathbb{R} \cong \mathbb{R}^4 \tag{9.63}$$

where the expression between $=$ and $\cong$ is an internal direct sum. The algebra is associative, but noncommutative. It has a rich and colorful history. See the remark below.

Note that if we denote a generic quaternion by

$$q = x_1\mathfrak{i} + x_2\mathfrak{j} + x_3\mathfrak{k} + x_4 \tag{9.64}$$

with $x_1, \ldots, x_4 \in \mathbb{R}$, then we can define the conjugate quaternion by the equation

$$\bar{q} := -x_1\mathfrak{i} - x_2\mathfrak{j} - x_3\mathfrak{k} + x_4 \tag{9.65}$$

Now using the relations we compute

$$q\bar{q} = \bar{q}q = \sum_{\mu=1}^{4} x_\mu x_\mu \in \mathbb{R}_+ \tag{9.66}$$

This defines a norm on the quaternion algebra:

$$\| q \|^2 := q\bar{q} = \bar{q}q \tag{9.67}$$

The norm satisfies an important and slightly nontrivial identity:

$$\| q_1 q_2 \|^2 = \| q_1 \|^2 \| q_2 \|^2 \tag{9.68}$$

See the exercises below.

If follows that the *unit quaternions*, i.e. those with $\| q \| = 1$ form a nonabelian group. (Exercise: Prove this!) In the exercises below you show that this group is isomorphic to $SU(2)$.

One fact about the quaternions that is often quite useful is the following. There is a left- and right-action of the quaternions on itself. If $\mathfrak{q}$ is a quaternion then we can define $L(\mathfrak{q}) : \mathbb{H} \to \mathbb{H}$ by

$$L(\mathfrak{q}) : \mathfrak{q}' \mapsto \mathfrak{q}\mathfrak{q}' \tag{9.69}$$

and similarly there is a right-action

$$R(\mathfrak{q}) : \mathfrak{q}' \mapsto \mathfrak{q}'\mathfrak{q} \tag{9.70}$$

The algebra of operators $L(\mathfrak{q})$ is isomorphic to $\mathbb{H}$ and the algebra of operators $R(\mathfrak{q})$ is isomorphic to $\mathbb{H}^{\rm opp}$, which in turn is isomorphic to $\mathbb{H}$ itself. Now $\mathbb{H}$ is a four-dimensional real vector space and $L(\mathfrak{q})$ and $R(\mathfrak{q})$ are commuting real-linear operators. Therefore there is an inclusion

$$\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}^{\rm opp} \hookrightarrow {\rm End}_{\mathbb{R}}(\mathbb{H}) \cong {\rm End}(\mathbb{R}^4) \tag{9.71}$$

Since $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}^{\rm opp}$ has real dimension 16 this is isomorphism of algebras over $\mathbb{R}$.

Remarks:

1. It is very interesting to compare $\mathbb{C}$ and $\mathbb{H}$.

   - We obtain $\mathbb{C}$ from $\mathbb{R}$ by introducing a squareroot of -1, which we can denote as $\mathfrak{i}$, but we obtain $\mathbb{H}$ from $\mathbb{R}$ by introducing three independent squareroots of -1, which are traditionally denoted as $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}$.
   - $\mathbb{C}$ is associative and commutative, while $\mathbb{H}$ is associative, but noncommutative.
   - There is a conjugation operation $z \to \bar{z}$ and $q \to \bar{q}$ so that $z\bar{z} \in \mathbb{R}_+$ and $q\bar{q} = \bar{q}q \in \mathbb{R}_+$ and is zero iff $z = 0$ or $q = 0$, respectively.
   - The set of all square-roots of -1 in $\mathbb{C}$ is $\{\pm\mathfrak{i}\}$ is a sphere S^0. The set of all square-roots of -1 in $\mathbb{H}$ is a sphere S^2.

- The group of unit norm elements in $\mathbb{C}$ is the Abelian group $U(1)$. The group of unit norm elements in $\mathbb{H}$ is the non-Abelian group $SU(2)$.

2. *A little history*

There is some colorful history associated with the quaternions which mathematicians are fond of relating. My sources for the following are

1. E.T. Bell, *Men of Mathematics* and C.B. Boyer, *A History of Mathematics.*

2. J. Baez, "The Octonions," Bull. Amer. Math. Soc. 39 (2002), 145-205. Also available at http://math.ucr.edu/home/baez/octonions/octonions.html

The second of these is highly readable and informative.

In 1833, W.R. Hamilton presented a paper in which, for the first time, complex numbers were explicitly identified with pairs (x, y) of real numbers. Hamilton stressed that the multiplication law of complex numbers could be written as:

$$(x, y)(u, v) = (xu - vy, vx + yu)$$

and realized that this law could be interpreted in terms of rotations of vectors in the plane.

Hamilton therefore tried to associate ordered triples (x, y, z) of real numbers to vectors in $\mathbb{R}^3$ and sought to discover a multiplication law which "expressed rotation" in $\mathbb{R}^3$. It seems he was trying to look for what we today call a normed division algebra. According to his account, for 10 years he struggled to define a multiplication law – that is, an algebra structure – on ordered 3-tuples of real numbers. I suspect most of his problem was that he didn't know what mathematical structure he was really searching for. This is a situation in which researchers in mathematics and physics often find themselves, and it can greatly confound and stall research. At least one of his stumbling blocks was the assumption that the algebra had to be commutative.

Then finally – so the story goes – on October 16, 1843 he realized quite suddenly during a walk that if he *dropped the commutativity law* then he could write a consistent algebra in four dimensions, that is, he denoted $q = a + b\mathfrak{i} + c\mathfrak{j} + d\mathfrak{k}$ and realized that he should impose

$$\mathfrak{i}^2 = \mathfrak{j}^2 = \mathfrak{k}^2 = \mathfrak{ijk} \tag{9.72}$$

He already knew that $\mathfrak{i}^2 = -1$ was essential, so surely $\mathfrak{j}^2 = \mathfrak{k}^2 = -1$. Then $\mathfrak{i}^2 = \mathfrak{ijk}$ implies $\mathfrak{i} = \mathfrak{jk}$. But for this to square to one we need $\mathfrak{jk} = -\mathfrak{kj}$. Apparently he carved these equations into Brougham Bridge while his wife, and presumably, not the police, stood by. He lost no time, and then went on to announce his discovery to the Royal Irish Academy, the very same day. Hamilton would spend the rest of his life championing the quaternions as something of cosmic significance, basic to the structure of the universe and of foundational importance to physics. That is not the general attitude today: The quaternions fit very nicely into the structure of Clifford algebras, and they are particularly important in some aspects of four-dimensional geometry and the geometry of hyperkahler manifolds. However, in more

general settings the vector analysis notation introduced by J. W. Gibbs at Yale has proved to be much more flexible and congenial to working with mathematical physics in general dimensions.

The day after his great discovery, Hamilton wrote a detailed letter to his friend John T. Graves explaining what he had found. Graves replied on October 26th, and in his letter he said:

" There is still something in the system which gravels me. I have not yet any clear views as to the extent to which we are at liberty arbitrarily to create imaginaries, and to endow them with supernatural properties. ... If with your alchemy you can make three pounds of gold, why should you stop there? "

By Christmass Graves had discovered the octonions, and in January 1844, had gone on to a general theory of "2^m-ions" but stopped upon running into an "unexpected hitch."

Later, again inspired by Hamilton's discovery of the quaternions, Arthur Cayley independently discovered the octonions in 1845.

See Chapter 7 below for a discussion of the octonions.

3. What Hamilton and Graves discovered are special cases of a general construction is known as the *c-double* or Cayley-Dickson process. Given an algebra $\mathcal{A}$ with an involutive anti-homomorphism $j : v \to \bar{v}$ we form the vector space $\mathcal{A}^{(d)} = \mathcal{A} \oplus \mathcal{A}$. The new algebra structure is defined by a choice of a nonzero element $c \in \mathcal{A}$ and the new product is [19]

$$(v_1, v_2) \cdot (w_1, w_2) := (v_1 w_1 + c\overline{w_2} v_2, w_2 v_1 + v_2 \overline{w_1}) \tag{9.73}$$

Since $j(v) = \bar{v}$ is linear over κ, the new algebra product is bilinear, and therefore $\mathcal{A}^{(d)}$ is itself an algebra. The octonions are obtained using $\mathcal{A} = \mathbb{H}$ with $c = -1$. At each stage of the process the algebraic structure becomes more complicated.

Exercise *Due Diligence*
Show that (9.60) and (9.61) imply (9.62). [20]

Exercise *Quaternionic Conjugation And The Opposite Quaternion Algebra*

[19] There is a choice of convention here. The standard text by Schaefer, p.45 uses a different convention. We are following Jacobsen.

[20] *Answer*: Substitute $\mathfrak{k} = \mathfrak{ij}$ into the equation $\mathfrak{k}^2 = -1$. Multiply this equation on the left by $\mathfrak{i}$ and on the right by $\mathfrak{j}$ and use (9.60). A similar manipulation applies to the other equations.

a.) Show that quaternion conjugation is an anti-automorphism, that is:

$$\overline{q_1 q_2} = \overline{q_2}\ \overline{q_1} \tag{9.74}$$

b.) Show that $\mathbb{H}^{\rm opp} \cong \mathbb{H}$.

Exercise *Identities On Sums Of Squares*

a.) If z_1 and z_2 are complex numbers then

$$|z_1|^2 |z_2|^2 = |z_1 z_2|^2 \tag{9.75}$$

Show that this can be interpreted as an identity for sums of squares of <u>real</u> numbers x, y, s, t:

$$(x^2 + y^2)(s^2 + t^2) = (xs - yt)^2 + (ys + xt)^2 \tag{9.76}$$

b.) Show that

$$\| q_1 \|^2 \| q_2 \|^2 = \| q_1 q_2 \|^2 \tag{9.77}$$

and interpret this as an identity of the form

$$\left(\sum_{\mu=1}^{4} (x_\mu)^2\right)\left(\sum_{\mu=1}^{4} (y_\mu)^2\right) = \left(\sum_{\mu=1}^{4} (z_\mu)^2\right) \tag{9.78}$$

where z_μ is of the form

$$z_\mu = \sum_{\nu,\lambda=1}^{4} a_{\mu\nu\lambda} x_\nu y_\lambda \tag{9.79}$$

Find the explicit formula for z_μ.

Exercise *A Matrix Representation Of The Quaternions*

a.) Show that

$$\mathfrak{i} \to -\sqrt{-1}\sigma^1 \qquad \mathfrak{j} \to -\sqrt{-1}\sigma^2 \qquad \mathfrak{k} \to -\sqrt{-1}\sigma^3 \tag{9.80}$$

defines a set of 2×2 complex matrices satisfying the quaternion algebra. Under this mapping a quaternion q is identified with a 2×2 complex matrix

$$q \to \rho(q) = \begin{pmatrix} z & -\bar{w} \\ w & \bar{z} \end{pmatrix} \tag{9.81}$$

eq:2x2-quat

with $z = -\mathrm{i}(x_3 + \mathrm{i}x_4)$ and $w = -\mathrm{i}(x_1 + \mathrm{i}x_2)$.

b.) Show that the set of matrices in (9.81) may be characterized as the set of 2×2 complex matrices A so that

$$A^* = JAJ^{-1} \qquad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{9.82}$$

If we introduce the epsilon symbol $\epsilon^{\alpha\beta}$ which is totally antisymmetric and such that $\epsilon^{12} = 1$ (this is a choice) then we can write the condition as

$$(A_{\alpha\dot{\beta}})^* = \epsilon^{\alpha\gamma}\epsilon^{\dot{\beta}\dot{\delta}}A_{\gamma\dot{\delta}} \tag{9.83}$$

Exercise *Quaternions And Rotations*

a.) Using the fact that $q\bar{q} = \bar{q}q$ is a real scalar show that the set of unit-norm elements in $\mathbb{H}$ is a group. Using the representation (9.81) show that this group is isomorphic to the group $SU(2)$.

b.) Define the *imaginary quaternions* to be the subspace of $\mathbb{H}$ of quaternions such that $\bar{q} = -q$. Denote this as $\Im(\mathbb{H})$. Show that $\Im(\mathbb{H}) \cong \mathbb{R}^3$, and show that $\Im(\mathbb{H})$ is in fact a real Lie algebra with

$$[q_1, q_2] := q_1q_2 - q_2q_1 \tag{9.84}$$

eq:ImagQuatLie

Using (9.81) identify this as the Lie algebra of $SU(2)$: $\Im(\mathbb{H}) \cong \mathfrak{su}(2)$. (Recall that $\mathfrak{su}(2)$ is the real Lie algebra of 2×2 traceless anti-Hermitian matrices.)

c.) Show that $\det(\rho(q)) = q\bar{q} = x_\mu x_\mu$ and use this to define a homomorphism

$$\rho : SU(2) \times SU(2) \to SO(4) \tag{9.85}$$

eq:DoubleCover

d.) Show that we have an exact sequence [21]

$$1 \to \mathbb{Z}_2 \to SU(2) \times SU(2) \to SO(4) \to 1 \tag{9.86}$$

where $\iota : \mathbb{Z}_2 \to SU(2) \times SU(2)$ takes the nontrivial element -1 to $(-1_2, -1_2)$.

e.) Show that under the homomorphism ρ the diagonal subgroup of $SU(2) \times SU(2)$ preserves the scalars and acts as the group $SO(3)$ of rotations of $\Im(\mathbb{H}) \cong \mathbb{R}^3$.

---

[21] *Answer*: We use the homomorphism $\rho : SU(2) \times SU(2) \to SO(4)$ defined by $\rho(q_1, q_2) : q \mapsto q_1 q \overline{q_2}$. To compute the kernel we search for pairs (q_1, q_2) of unit quaternions so that $q_1 q \overline{q_2} = q$ for all $q \in \mathbb{H}$. Applying this to $q = 1$ gives $q_1 = q_2$. Then applying it to $q = \mathrm{i}, \mathrm{j}, \mathrm{k}$ we see that $q_1 = q_2$ must be a real scalar. The only such unit quaternions are $q_1 \in \{\pm 1\}$. To check the image is entirely in $SO(4)$ you can check this for diagonal matrices in $SU(2) \times SU(2)$ and then everything conjugate to these must be in $SO(4)$, so the image is a subgroup of $SO(4)$ of dimension six and therefore must be all of $SO(4)$.

Exercise *Unitary Matrices Over Quaternions And Symplectic Groups*

a.) Show that the algebra $\mathrm{Mat}_n(\mathbb{H})$ of $n \times n$ matrices with quaternionic entries can be identified as the subalgebra of $\mathrm{Mat}_{2n}(\mathbb{C})$ of matrices A such that

$$A^* = JAJ^{-1} \qquad\qquad J = \begin{pmatrix} 0 & -1_n \\ 1_n & 0 \end{pmatrix} \tag{9.87}$$

b.) Show that the unitary group over $\mathbb{H}$:

$$U(n,\mathbb{H}) := \{u \in \mathrm{Mat}_n(\mathbb{H}) | u^\dagger u = 1\} \tag{9.88}$$

is isomorphic to

$$USp(2n) := \{u \in U(2n,\mathbb{C}) | u^* = JuJ^{-1}\} \tag{9.89}$$

To appreciate the notation show that matrices $u \in USp(2n)$ also satisfy

$$u^{tr} J u = J \tag{9.90}$$

which is the defining relation of $Sp(2n,\mathbb{C})$.

Exercise *Complex structures on* $\mathbb{R}^4$

a.) Show that the complex structures on $\mathbb{R}^4$ compatible with the Euclidean metric can be identified as the maps

$$q \mapsto nq \qquad\qquad n^2 = -1 \tag{9.91}$$

OR

$$q \mapsto qn \qquad\qquad n^2 = -1 \tag{9.92}$$

b.) Use this to show that the space of such complex structures is $S^2 \amalg S^2$.

Exercise *Regular Representation*

♣This is too important to be an exercise, and is used heavily later. ♣

Compute the left and right regular representations of $\mathbb{H}$ on itself as follows: Choose a real basis for $\mathbb{H}$ with $v_1 = \mathfrak{i}, v_2 = \mathfrak{j}, v_3 = \mathfrak{k}, v_4 = 1$. Let $L(\mathfrak{q})$ denote left-multiplication by a quaternion $\mathfrak{q}$ and $R(\mathfrak{q})$ right-multiplciation by $\mathfrak{q}$. Then the representation matrices are:

$$L(\mathfrak{q})v_a := \mathfrak{q} \cdot v_a := L(\mathfrak{q})_{ba} v_b \tag{9.93}$$

$$R(\mathfrak{q})v_a := v_a \cdot \mathfrak{q} := R(\mathfrak{q})_{ba} v_b \tag{9.94}$$

a.) Show that:

$$L(\mathfrak{i}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \tag{9.95}$$

$$L(\mathfrak{j}) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{9.96}$$

$$L(\mathfrak{k}) = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{9.97}$$

$$R(\mathfrak{i}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \tag{9.98}$$

$$R(\mathfrak{j}) = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{9.99}$$

$$R(\mathfrak{k}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{9.100}$$

b.) Show that these matrices generate the full 16-dimensional algebra $M_4(\mathbb{R})$. This is the content of the statement that

$$\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}^{\rm opp} \cong {\rm End}(\mathbb{R}^4) \tag{9.101}$$

Exercise *'t Hooft symbols and the regular representation of* $\mathbb{H}$

The famous 't Hooft symbols, introduced by 't Hooft in his work on instantons in gauge theory are defined by

♣CHECK FACTORS OF TWO!!! ♣

$$\alpha^{\pm,i}_{\mu\nu} := \pm\delta_{i\mu}\delta_{\nu 4} \mp \delta_{i\nu}\delta_{\mu 4} + \epsilon_{i\mu\nu} \tag{9.102}$$

where $1 \leq \mu, \nu \leq 4$, $1 \leq i \leq 3$ and $\epsilon_{i\mu\nu}$ is understood to be zero if μ or ν is equal to 4. (Note: Some authors will use the notation $\eta^i_{\mu\nu}$, and some authors will use a different overall normalization.)

a.) Show that

$$\alpha^{+,1} = R(\mathfrak{i}) \qquad \alpha^{+,2} = R(\mathfrak{j}) \qquad \alpha^{+,3} = R(\mathfrak{k}) \tag{9.103}$$

$$\alpha^{-,1} = -L(\mathfrak{i}) \qquad \alpha^{-,2} = -L(\mathfrak{j}) \qquad \alpha^{-,3} = -L(\mathfrak{k}) \tag{9.104}$$

b.) Verify the relations

$$\begin{aligned}[\alpha^{\pm,i},\alpha^{\pm,j}] &= -2\epsilon^{ijk}\alpha^{\pm,k}\\ [\alpha^{\pm,i},\alpha^{\mp,j}] &= 0\end{aligned} \tag{9.105}$$

c.) Let $\mathfrak{so}(4)$ of 4×4 denote the real Lie algebra of real anti-symmetric matrices. It is the Lie algebra of $SO(4)$. Show that it is of dimension 6 and that every element can be uniquely decomposed as $L(q_1)-R(q_2)$ where q_1,q_2 are imaginary quaternions.

d.) Show that the map

$$\Im(\mathbb{H})\oplus\Im(\mathbb{H})\to\mathfrak{so}(4) \tag{9.106}$$

$$q_1\oplus q_2\to L(q_1)-R(q_2) \tag{9.107}$$

defines an isomorphism of Lie algebras

$$\mathfrak{su}(2)\oplus\mathfrak{su}(2)\cong\mathfrak{so}(4) \tag{9.108}$$

eq:so4-isomorp

e.) Using (b) and (9.81) show that the above isomorphism of Lie algebras can be expressed as a mapping of generators

$$\begin{aligned}\frac{\sqrt{-1}}{2}\sigma^i\oplus 0 &\mapsto \frac{1}{2}\alpha^{+,i}\\ 0\oplus\frac{\sqrt{-1}}{2}\sigma^i &\mapsto \frac{1}{2}\alpha^{-,i}\end{aligned} \tag{9.109}$$

f.) Now show that

$$\{\alpha^{\pm,i},\alpha^{\pm,j}\}=-2\delta^{ij} \tag{9.110}$$

and deduce that:

$$\begin{aligned}\alpha^{+,i}\alpha^{+,j} &= -\delta^{ij}-\epsilon^{ijk}\alpha^{+,k}\\ \alpha^{-,i}\alpha^{-,j} &= -\delta^{ij}-\epsilon^{ijk}\alpha^{-,k}\end{aligned} \tag{9.111}$$

g.) Deduce that the inverse isomorphism to (9.108) is

$$T\mapsto\left(-\sqrt{-1}\mathrm{Tr}(\alpha^{+,i}T)\sigma^i\right)\oplus\left(-\sqrt{-1}\mathrm{Tr}(\alpha^{-,i}T)\sigma^i\right) \tag{9.112}$$

Exercise *Quaternions And (Anti-)Self-Duality*

a.) Introduce the 4d epsilon tensor $\epsilon_{\mu\nu\lambda\rho}$ with $\epsilon_{1234}=+1$. Show that the rank-two antisymmetric tensors $\alpha^{+,i}_{\mu\nu}$ for fixed i are *self-dual* and *anti-self-dual* in the sense that

$$\alpha^{+,i}_{\mu\nu}=\frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\alpha^{+,i}_{\lambda\rho} \tag{9.113}$$

$$\alpha^{-,i}_{\mu\nu}=-\frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\alpha^{-,i}_{\lambda\rho} \tag{9.114}$$

b.) On $\mathfrak{so}(4)$, which, as a vector space, can be identified with the space of two-index anti-symmetric tensors define

$$(*T)_{\mu\nu} := \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}T_{\lambda\rho} \tag{9.115}$$

Show that the linear transformation $* : \mathfrak{so}(4) \to \mathfrak{so}(4)$ satisfies $*^2 = 1$. Therefore

$$P^{\pm} = \frac{1}{2}(1 \pm *) \tag{9.116}$$

are projection operators. Interpret the isomorphism (9.108) as the decomposition into self-dual and anti-self-dual tensors.

c.) If T is an antisymmetric tensor with components $T_{\mu\nu}$ then a common notation is $P^{\pm}(T) = T^{\pm}$. Check that

$$\begin{aligned} T^{\pm}_{12} &= \frac{1}{2}(T_{12} \pm T_{34}) = \pm T^{\pm}_{34} \\ T^{\pm}_{13} &= \frac{1}{2}(T_{13} \pm T_{42}) = \pm T^{\pm}_{42} \\ T^{\pm}_{14} &= \frac{1}{2}(T_{14} \pm T_{23}) = \pm T^{\pm}_{23} \end{aligned} \tag{9.117}$$

We remark that the choice $\epsilon_{1234} = +1$, instead of $\epsilon_{1234} = -1$ is a choice of orientation on $\mathbb{R}^4$. A change of orientation exchanges self-dual and anti-self-dual.

d.) Recall the notation that $v_\mu \in \mathbb{H}$ is the natural basis of quaternions $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}, 1$. Show that $v^-_{\mu\nu} = v_\mu \overline{v_\nu} - v_\nu \overline{v_\mu}$ are anti-self-dual and $v^+_{\mu\nu} = \overline{v_\mu} v_\nu - \overline{v_\nu} v_\mu$ are self-dual.

Exercise *More about $SO(4)$ matrices*

Show that if $x_{1,2,3,4}$ are real then

$$L(x_4 + x_1\mathfrak{i} + x_2\mathfrak{j} + x_3\mathfrak{k}) = \begin{pmatrix} x_4 & -x_3 & x_2 & x_1 \\ x_3 & x_4 & -x_1 & x_2 \\ -x_2 & x_1 & x_4 & x_3 \\ -x_1 & -x_2 & -x_3 & x_4 \end{pmatrix} \tag{9.118}$$

eq:L-GenQuat

b.) Show that when $x_\mu x_\mu = 1$ the matrix $L(x_4 + x_1\mathfrak{i} + x_2\mathfrak{j} + x_3\mathfrak{k})$ is an $SO(4)$ rotation.

c.) Similarly

$$R(y_4 + y_1\mathfrak{i} + y_2\mathfrak{j} + y_3\mathfrak{k}) = \begin{pmatrix} y_4 & y_3 & -y_2 & y_1 \\ -y_3 & y_4 & y_1 & y_2 \\ y_2 & -y_1 & y_4 & y_3 \\ -y_1 & -y_2 & -y_3 & y_4 \end{pmatrix} \tag{9.119}$$

eq:R-GenQuat

is an $SO(4)$ rotation when $y_\mu y_\mu = 1$.

d.) Show that the general $SO(4)$ matrix is a product of these.

e.) Show that, in particular that if we identify $\mathfrak{k} = \mathrm{i}\sigma^3$ then

$$\rho(e^{\mathrm{i}\theta\sigma^3}, 1) = \begin{pmatrix} R(\theta) & 0 \\ 0 & R(-\theta) \end{pmatrix} \tag{9.120}$$

$$\rho(1, e^{\mathrm{i}\theta\sigma^3}) = \begin{pmatrix} R(\theta) & 0 \\ 0 & R(\theta) \end{pmatrix} \tag{9.121}$$

where ρ is the homomorphism defined in (9.85) and

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{9.122}$$

9.4 Quaternionic Structure On A Real Vector Space

Definition: A *quaternionic vector space* is a vector space V over $\kappa = \mathbb{R}$ together with three real linear operators $I, J, K \in \mathrm{End}(V)$ satisfying the quaternion relations. In other words, it is a real vector space which is a module for the quaternion algebra.

Example 1: Consider $\mathbb{H}^{\oplus n} \cong \mathbb{R}^{4n}$. Vectors are viewed as n-component column vectors with quaternion entries. Each quaternion is then viewed as a four-component real vector. The operators I, J, K are componentwise left-multiplication by $L(\mathfrak{i}), L(\mathfrak{j}), L(\mathfrak{k})$.

It is possible to put a quaternionic Hermitian structure on a quaternionic vector space and thereby define the quaternionic unitary group. Alternatively, we can define $U(n, \mathbb{H})$ as the group of $n \times n$ matrices over $\mathbb{H}$ such that $uu^\dagger = u^\dagger u = 1$. In order to define the conjugate-transpose matrix we use the quaternionic conjugation $q \to \bar{q}$ defined above.

Exercise *A natural sphere of complex structures*

Show that if V is a quaternionic vector space with complex structures I, J, K then there is a natural sphere of complex structures given by

$$\mathcal{I} = x_1 I + x_2 J + x_3 K \qquad x_1^2 + x_2^2 + x_3^2 = 1 \tag{9.123}$$

9.5 Quaternionic Structure On Complex Vector Space

Just as we can have a complex structure on a real vector space, so we can have a *quaternionic structure* on a complex vector space V. This is a $\mathbb{C}$-anti-linear operator K on V which squares to -1. Once we have $K^2 = -1$ we can combine with the operator I which is just multiplication by $\sqrt{-1}$, to produce $J = KI$ and then we can check the quaternion relations. The underlying real space $V_{\mathbb{R}}$ is then a quaternionic vector space.

Example 2: The canonical example is given by taking a complex vector space V and forming

$$W = V \oplus \bar{V} \tag{9.124}$$

The underlying real vector space $W_{\mathbb{R}}$ has quaternion actions:

$$I : (v_1, \overline{v_2}) \mapsto (\mathrm{i}v_1, \mathrm{i}\overline{v_2}) = (\mathrm{i}v_1, \overline{-\mathrm{i}v_2}) \tag{9.125}$$

$$J : (v_1, \overline{v_2}) \mapsto (-v_2, \overline{v_1}) \tag{9.126}$$

$$K : (v_1, \overline{v_2}) \mapsto (-\mathrm{i}v_2, \overline{-\mathrm{i}v_1}) \tag{9.127}$$

Remarks

1. Tensor Products Of Quaternionic Vector Spaces. Let V_1, V_2 be complex vector spaces with a quaternionic structure. Note that $V_1 \otimes_{\mathbb{C}} V_2$ is not (naturally) a quaternionic vector space, but rather a vector space with a real structure. The reason is that, if K_1, K_2 are the anti-linear operators defining the quaternionic structure on V_1, V_2, respectively, then $K_1 \otimes K_2$ is again anti-linear, but squares to $+1$, not -1. Similarly, if $V_1, \ldots, V_n$ are quaternionic vector spaces then $V_1 \otimes_{\mathbb{C}} V_2 \otimes_{\mathbb{C}} \cdots \otimes_{\mathbb{C}} V_n$ has a natural quaternionic structure if n is odd and a natural real structure if n is even.

2. A representation $\rho : G \to GL(V)$, where V is a real vector space, is said to be a *quaternionic representation* if $\rho(g)$ commute with I, J, K for all g. A good example of a quaternionic representation is the provided by the fundamental representation V of $SU(2)$. Let $V \cong \mathbb{C}^2$ be the defining 2-dimensional representation of $SU(2)$. We can identify

   $$V_{\mathbb{R}} \cong \mathbb{R}^4 \cong \mathbb{H} \tag{9.128}$$

   and then the $SU(2)$ action is the $U(1, \mathbb{H})$ left action on $\mathbb{H}$. This becomes manifest if we identify $\mathbb{H}$ with the space of 2×2 complex matrices:

   $$\begin{pmatrix} z & -\bar{w} \\ w & \bar{z} \end{pmatrix} \tag{9.129}$$

   with $\psi \in V$ identified with

   $$\psi = \begin{pmatrix} z \\ w \end{pmatrix} \tag{9.130}$$

   Now, the action of I, J, K commuting with the $\rho(u)$ is clearly given by $-R(\mathfrak{i})$, $-R(\mathfrak{j})$, and $-R(\mathfrak{k})$, respectively. More generally, the $\mathbb{H}$ action commuting with the $SU(2)$ action on the representation is given by right-multiplication by $\bar{\mathfrak{q}}$ with $\mathfrak{q} \in \mathbb{H}$.

3. It follows that, if V is the fundamental representation of $SU(2)$, then $V^{\otimes n}$ is a real representation for n even and is a quaternionic representation for n odd. As we saw from the Clebsch-Gordon decomposition, $V^{\otimes n} \cong V(n/2) \oplus \cdots$. Thus we conclude that the spin j representation is real for j integral and quaternionic for j half-integral.

4. Kramers' theorem. We can return to Kramers' theorem: If there is a time-reversing symmetry in quantum mechanics and $T^2 = -1$ then T defines a quaternionic structure. But the complex dimension of a quaternionic vector space is necessarily even. This is the Kramers degeneracy.

9.5.1 Complex Structure On Quaternionic Vector Space

♣This subsubsection needs to be written more carefully. ♣

Recall that a quaternionic vector space is a real vector space V with an action of the quaternions. So for every $q \in \mathbb{H}$ we have $T(q) \in \mathrm{End}_{\mathbb{R}}(V)$ such that

$$T(q_1)T(q_2) = T(q_1 q_2) \tag{9.131}$$

In other words, a representation of the real quaternion algebra.

If we think of V as a copy of $\mathbb{H}^n$ with the quaternionic action left-multiplication by q componentwise, so that

$$T(q)\begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} := \begin{pmatrix} qq_1 \\ \vdots \\ qq_n \end{pmatrix} \tag{9.132}$$

then a complex structure would be a left-action by any $GL(n,\mathbb{H})$ conjugate of $T(\mathfrak{i})$. If we wish to preserve the norm, then it is a $U(n,\mathbb{H})$ conjugate of $T(\mathfrak{i})$.

A complex structure then describes an embedding of $\mathbb{C}^n$ into $\mathbb{H}^n$ so that we have an isomorphism of

$$\mathbb{H}^n \cong \mathbb{C}^n \oplus \mathbb{C}^n \tag{9.133}$$

as complex vector spaces.

9.5.2 Summary

To summarize we have described four basic structures we can put on vector spaces:

1. A *complex structure* on a real vector space W is a real linear map $I : W \to W$ with $I^2 = -1$. That is, a representation of the real algebra $\mathbb{C}$.

2. A *real structure* on a complex vector space V is a $\mathbb{C}$-anti-linear map $K : V \to V$ with $K^2 = +1$.

3. A *quaternionic structure* on a complex vector space V is a $\mathbb{C}$-anti-linear map $K : V \to V$ with $K^2 = -1$.

4. A *complex structure on a quaternionic vector space* V is a representation of the real algebra $\mathbb{H}$ with a complex structure commuting with the $\mathbb{H}$-action.

9.6 Spaces Of Real, Complex, Quaternionic Structures

-RCH-Structure

This section makes use of the "stabilizer-orbit theorem." See the beginning of section 3.

♣So, we should move this to chapter three as an application of the Stabilizer-Orbit theorem. ♣

We saw above that if V is a finite-dimensional real vector space with a complex structure then by an appropriate choice of basis we have an isomorphism $V \cong \mathbb{R}^{2n}$, for a suitable integer n, and $I = I_0$. So, choose an isomorphism of V with $\mathbb{R}^{2n}$ and identify with space of

complex structures on V with those on $\mathbb{R}^{2n}$. Then the general complex structure on $\mathbb{R}^{2n}$ is of the form SI_0S^{-1} with $S \in GL(2n,\mathbb{R})$. In other words, there is a transitive action of $GL(2n,\mathbb{R})$ on the space of complex structures. We can then identify the space with a homogeneous space of $GL(2n,\mathbb{R})$ by computing the stabilizer of I_0. Now if

$$gI_0g^{-1} = I_0 \tag{9.134}$$

eq:Stab-I0

for some $g \in GL(2n,\mathbb{R})$ then we can write g in block-diagonal form

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{9.135}$$

and then the condition (9.134) is equivalent to $C = -B$ and $D = A$, so that

$$\mathrm{Stab}_{GL(2n,\mathbb{R})}(I_0) = \{g \in GL(2n,\mathbb{R}) | g = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}\} \tag{9.136}$$

we claim this subgroup of $GL(2n,\mathbb{R})$ is isomorphic to $GL(n,\mathbb{C})$. To see this simply note the above matrix is $A \otimes 1_{2\times 2} + B \otimes (\mathrm{i}\sigma^2)$ and we can diagonalize $\mathrm{i}\sigma^2$. Explicitly, if we introduce the matrix

$$S_0 := -\frac{1}{2\mathrm{i}} \begin{pmatrix} 1_n & 1_n \\ -\mathrm{i}1_n & \mathrm{i}1_n \end{pmatrix} \tag{9.137}$$

then if $g \in \mathrm{Stab}_{GL(2n,\mathbb{R})}(I_0)$ and we write it in block form we have

$$S_0^{-1} g S_0 = \begin{pmatrix} A - \mathrm{i}B & 0 \\ 0 & A + \mathrm{i}B \end{pmatrix} \tag{9.138}$$

thus the determinant is

$$\det(g) = |\det(A + \mathrm{i}B)|^2 \tag{9.139}$$

and since $\det(g) \neq 0$ we know that $\det(A + \mathrm{i}B) \neq 0$. Conversely, if we have a matrix $h \in GL(n,\mathbb{C})$ we can decompose it into its real and imaginary parts $h = A + \mathrm{i}B$ and embed into $GL(2n,\mathbb{R})$ via

$$h \mapsto \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \tag{9.140}$$

(We could change the sign of B in this embedding. The two embeddings differ by the complex conjugation automorphism of $GL(n,\mathbb{C})$.)

We conclude that the space of complex structures is a homogeneous space

$$\boxed{\mathrm{ComplexStr}(\mathbb{R}^{2n}) \cong GL(2n,\mathbb{R})/GL(n,\mathbb{C})} \tag{9.141}$$

We could demand that our complex structures are compatible with the Euclidean metric on $\mathbb{R}^{2n}$. Then the conjugation action by $O(2n)$ is transitive and the above embedding is an embedding of $U(n)$ into $O(2n)$ and

$$\boxed{\mathrm{CompatComplexStr}(\mathbb{R}^{2n}) \cong O(2n)/U(n)} \tag{9.142}$$

We now turn to the real structures on a finite-dimensional complex vector space. We can choose an isomorphism $V \cong \mathbb{C}^n$ and then the general real structure is related to

$$\mathcal{C}_0 : \sum_{i=1}^{n} z_i e_i \to \sum_{i=1}^{n} z_i^* e_i \tag{9.143}$$

by conjugation: $\mathcal{C} = g^{-1}\mathcal{C}_0 g$ with $g \in GL(n,\mathbb{C})$. [22] The stabilizer of $\mathcal{C}_0$ is rather obviously $GL(n,\mathbb{R})$, which sits naturally in $GL(n,\mathbb{C})$ as a subgroup. Thus:

$$\boxed{\text{RealStr}(\mathbb{C}^n) \cong GL(n,\mathbb{C})/GL(n,\mathbb{R})} \tag{9.144}$$

$$\boxed{\text{CompatRealStr}(\mathbb{C}^n) \cong U(n)/O(n)} \tag{9.145}$$

Let us now consider quaternionic structures on a complex vector space. We identify these as anti-linear operators $J : V \to V$ that square to -1 and denote the set by $\text{QuatStr}(V)$. Now, $GL(V)$ acts on $\text{QuatStr}(V)$ by $K \mapsto gKg^{-1}$. Moreover, this action is transitive. so we just need to determine a stabilizer group. Again, we can fix an isomorphism $V \cong \mathbb{C}^{2n}$ and we choose an anti-linear operator that squares to -1. Let us choose:

♣Need to supply argument. ♣

$$J_0 : \begin{pmatrix} v_1 \\ v_2^* \end{pmatrix} \to \begin{pmatrix} -v_2 \\ v_1^* \end{pmatrix} \tag{9.146}$$

where $v_1, v_2 \in \mathbb{C}^n$. Now we compute the stabilizer J_0, that is, the matrices $g \in GL(2n,\mathbb{C})$ such that $gJ_0g^{-1} = J_0$ when acting on vectors in $\mathbb{C}^{2n}$. In terms of matrices this means:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} g \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = g^* \tag{9.147}$$

which works out to:

$$\text{Stab}_{GL(2n,\mathbb{C})}(J_0) = \{g \in GL(2n,\mathbb{C}) | g = \begin{pmatrix} A & B \\ -B^* & A^* \end{pmatrix}\} \tag{9.148}$$

Recall our characterization of $n \times n$ matrices over the quaternions. It follows that this defines an embedding of $GL(n,\mathbb{H}) \to GL(2n,\mathbb{C})$. [23] So

$$\boxed{\text{QuatStr}(\mathbb{C}^{2n}) \cong GL(2n,\mathbb{C})/GL(n,\mathbb{H})} \tag{9.149}$$

Putting the natural Hermitian structure on $\mathbb{C}^{2n}$ we could demand that quaternionic structures are compatible with this Hermitian structure. Then the conjugation action by $U(2n)$ is transitive and the above embedding is an embedding of $USp(2n)$ into $U(2n)$ and

$$\boxed{\text{CompatQuatStr}(\mathbb{C}^{2n}) \cong U(2n)/USp(2n)} \tag{9.150}$$

[22] Proof: Given a real structure the fixed point set is a real subspace, but that space has a basis $\{e_i\}$ and then the real structure is what we wrote above. But all bases are related by $GL(n,\mathbb{C})$.

[23] It is possible, but tricky to define the notion of a determinant of a matrix of quaternions. It is best to think of $GL(n,\mathbb{H})$ as a Lie group with Lie algebra $Mat_n(\mathbb{H})$, or in terms of $2n \times 2n$ matrices over $\mathbb{C}$, the group we have written explicitly.

Finally, in a similar way we find that the space of complex structures on a quaternionic vector space can be identified with

$$\boxed{\text{CompatCmplxStr}(\mathbb{H}^n) \cong USp(2n)/U(n)} \tag{9.151}$$

Remarks

1. *Relation to Cartan involutions.* The above homogeneous spaces have an interesting relation to *Cartan involutions*. A Cartan involution [24] θ on a Lie algebra is a Lie algebra automorphism so that $\theta^2 = 1$. Decomposing into $\pm$ eigenspaces we have

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p} \tag{9.152}$$

where $\mathfrak{k} = \{X \in \mathfrak{g} | \theta(X) = X\}$ and $\mathfrak{p}$ is the -1 eigenspace and we have moreover

$$[\mathfrak{k}, \mathfrak{k}] = \mathfrak{k} \qquad [\mathfrak{k}, \mathfrak{p}] = \mathfrak{p} \qquad [\mathfrak{p}, \mathfrak{p}] = \mathfrak{k} \tag{9.153}$$

At the group level we have an involution $\tau : G \to G$ so that at the identity element $d\tau = \theta$. Then if $K = \text{Fix}(\tau)$ we have a diffeomorphism of G/K with the subset in G of "anti-fixed points":

$$G/K \cong \mathcal{O} := \{g \in G | \tau(g) = g^{-1}\} \tag{9.154}$$

The above structures, when made compatible with natural metrics are nice examples:

Complex structures on real vector spaces: $\mathbb{R}^{2n} \cong \mathbb{C}^n$. Moduli space:

$$O(2n)/U(n) \tag{9.155}$$ eq:ClassCartSp

This comes from $\tau(g) = I_0 g I_0^{-1}$ where I_0 is (9.8).

Real structures on complex vector spaces: $\mathbb{R}^n \hookrightarrow \mathbb{C}^n$. Moduli space

$$U(n)/O(n) \tag{9.156}$$ eq:ClassCartSp

This comes from $\tau(u) = u^*$.

Quaternionic structures on complex vector spaces: $\mathbb{C}^{2n} \cong \mathbb{H}^n$. Moduli space:

$$U(2n)/Sp(n) \tag{9.157}$$ eq:ClassCartSp

Viewing $Sp(n)$ as $USp(2n) := U(2n) \cap Sp(2n; \mathbb{C})$ we can use the involutive automorphism $\tau(g) = I_0^{-1} g^* I_0$ on $U(2n)$. The fixed points in $U(2n)$ are the group elements with $g I_0 g^{tr} = I_0$, but this is the defining equation of $Sp(2n, \mathbb{C})$.

Complex structures on quaternionic vector spaces: $\mathbb{C}^n \hookrightarrow \mathbb{H}^n$. Moduli space:

$$Sp(n)/U(n) \tag{9.158}$$ eq:ClassCartSp

[24] See Chapter **** below for much more detail.

Viewing $Sp(n)$ as unitary $n \times n$ matrices over the quaternions the involution is $\tau(g) = -\mathbf{i}g\mathbf{i}$, i.e. conjugation by the unit matrix times $\mathbf{i}$.

When Cartan classified compact symmetric spaces he found the 10 infinite series of the form $O \times O/O$, $U \times U/U$, $Sp \times Sp/Sp$, $O/O \times O$, $U/U \times U$, $Sp/Sp \times Sp$ and the above for families. In addition there is a finite list of exceptional cases.

2. *The 10-fold way.* In condensed matter physics there is a very beautiful classification of ensembles of Hamiltonians with a given symmetry type known as the 10-fold way. It is closely related to the above families of Cartan symmetric spaces, as discovered by Altland and Zirnbauer. See, for example,

   http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf

   http://www.physics.rutgers.edu/∼gmoore/PiTP-LecturesA.pdf

Exercise *Quaternionic Structures On* $\mathbb{R}^4$

In an exercise above we showed that the space of Euclidean-metric-compatible quaternionic structures on $\mathbb{R}^4$ is $S^2 \amalg S^2$. Explain the relation of this to the coset $O(4)/U(2)$.

---

10. Some Canonical Forms For a Matrix Under Conjugation

10.1 What is a canonical form?

We are going to collect a compendium of theorems on special forms into which matrices can be put.

There are different ways one might wish to put matrices in a "canonical" or standard form.

In general we could consider multiplying our matrix by different linear transformations on the left and the right

$$A \to S_1 A S_2 \tag{10.1}$$

where S_1, S_2 are invertible.

On the other hand, if A is the matrix of a linear transformation $T : V \to V$ then change of bases leads to change of A by conjugation:

$$A \to S A S^{-1} \tag{10.2}$$ eq:congj

for invertible S.

On the other hand, if A is the matrix of a bilinear form on a vector space (see below) then the transformation will be of the form:

$$A \to S A S^{tr} \tag{10.3}$$ eq:trnwsp

and here we can divide the problem into the cases where A is symmetric or antisymmetric.

There is some further important fine print on the canonical form theorems: The theorems can be different depending on whether the matrix elements in

$$\mathbb{C} \supset \mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z} \tag{10.4}$$

Also, we could put restrictions on $S \in GL(n, \mathbb{C})$ and require the matrix elements of S to be in $\mathbb{R}, \mathbb{Q}, \mathbb{Z}$. Finally, we could require S to be unitary or orthogonal. As we put more restrictions on the problem the nature of the canonical forms changes.

Some useful references for some of these theorems:

1. Herstein, sec. 6.10,
2. Hoffman and Kunze, Linear Algebra

10.2 Rank

If $T : V \to W$ is a linear transformation between two vector spaces the dimension of the image is called the *rank* of T. If V and W are finite dimensional complex vector spaces it is the only invariant of T under change of basis on V and W:

Theorem 1 Consider any $n \times m$ matrix over a field κ, $A \in Mat_{n\times m}(\kappa)$. This has a left and right action by $GL(n, \kappa) \times GL(m, \kappa)$. By using this we can always bring A to the canonical form: Let r denote the rank, that is, the dimension of the image space. Then there exist $g_1 \in GL(n, \kappa)$ and $g_2 \in GL(m, \kappa)$ so that $g_1 A g_2^{-1}$ is of the form:

♣left- action and right-action not defined until next chapter. ♣

a.) If $r < n, m$:

$$\begin{pmatrix} 1_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{pmatrix} \tag{10.5}$$

eq:rankinv

b.) If $r = n < m$ then we write the matrix as

$$\begin{pmatrix} 1_n & 0_{n\times(m-n)} \end{pmatrix} \tag{10.6}$$

c.) If $r = m < n$ then we write the matrix as

$$\begin{pmatrix} 1_m \\ 0_{(n-m)\times n} \end{pmatrix} \tag{10.7}$$

d.) If $r = n = m$ then we have the identity matrix.

That is, the only invariant under arbitrary change of basis of domain and range is the rank.

The proof easily follows from the fact that any subspace $V' \subset V$ has a complementary vector space V'' so that $V' \oplus V'' \cong V$.

Exercise

Prove this. Choose an arbitrary basis for κ^n and κ^m and define an operator T using the matrix A. Now construct a new basis, beginning by choosing a basis for $\ker T$.

Exercise
If M is the matrix of a rank 1 operator, in any basis, then it has the form

$$M_{i\alpha} = v_i w_\alpha \qquad i = 1, \ldots, n; \alpha = 1, \ldots, m \tag{10.8}$$

for some vectors v, w.

sec:ee

10.3 Eigenvalues and Eigenvectors

Now let us consider a *square matrix* $A \in Mat_{n\times n}(\kappa)$. Suppose moreover that it is the matrix of a linear transformation $T : V \to V$ expressed in some basis. If we are actually studying the operator T then we no longer have the luxury of using different transformations g_1, g_2 for change of basis on the domain and range. We must use the *same* invertible matrix S, and hence our matrix transforms by conjugation

$$A \to S^{-1}AS \tag{10.9}$$

This changes the classification problem dramatically.

When thinking about this problem it is useful to introduce the basic definition: If $T : V \to V$ is a linear operator and $v \in V$ is a nonzero vector so that

$$Tv = \lambda v \tag{10.10}$$

then λ is called an *eigenvalue* of T and v is called an *eigenvector*. A similar definition holds for a matrix. [25]

Example. The following matrix has two eigenvalues and two eigenvectors:

$$\begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{10.11}$$

$$\begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\lambda \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{10.12}$$

Note that the two equations can be neatly summarized as one by making the eigenvectors columns of a square matrix:

$$\begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix} \tag{10.13}$$

and so the matrix of eigenvectors diagonalizes our operator.

[25] Alternative terminology: characteristic value, characteristic vector.

Generalizing the previous example, if $T \in End(V)$ and V has a basis $\{v_1, \ldots, v_n\}$ of eigenvectors of T with eigenvalues λ_i. Then, wrt that basis, the associated matrix is diagonal:

$$\begin{pmatrix} \lambda_1 & 0 & 0 \cdots & 0 \\ 0 & \lambda_2 & 0 \cdots & 0 \\ \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 \cdots & \lambda_n \end{pmatrix} \tag{10.14}$$ eq:diagmatrix

In general, if A is the matrix of T with respect to some basis (not necessarily a basis of eigenvectors) and if S is a matrix whose columns are n linearly independent eigenvectors then

$$AS = S\mathrm{Diag}\{\lambda_1, \ldots, \lambda_n\} \qquad \Rightarrow \qquad S^{-1}AS = \mathrm{Diag}\{\lambda_1, \ldots, \lambda_n\} \tag{10.15}$$

As we shall see, not every matrix has a basis of eigenvectors. Depending on the field, a matrix might have no eigenvectors at all. A simple example is that over the field $\kappa = \mathbb{R}$ the matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{10.16}$$

has no eigenvectors at all. Thus, the following fact is very useful:

Theorem 10.3.1. If $A \in M_n(\mathbb{C})$ is any complex matrix then it has at least one nonvanishing eigenvector.

In order to prove this theorem it is very useful to introduce the characteristic polynomial:

Definition The *characteristic polynomial* of a matrix $A \in M_n(\kappa)$ is

$$p_A(x) =: \det(x\mathbf{1_n} - A) \tag{10.17}$$ eq:charpoly

Proof of Theorem 10.3.1: The characteristic polynomial $p_A(x)$ has at least one root, call it λ, over the complex numbers. Since, $\det(\lambda\mathbf{1_n} - A) = 0$ the matrix $\lambda\mathbf{1_n} - A$ has a nonzero kernel $K \subset \mathbb{C}^n$. If v is a nonzero vector in K then it is an eigenvector. ♠

So - a natural question is -

Given a matrix A, does it have a basis of eigenvectors? Equivalently, can we diagonalize A via $A \to S^{-1}AS$?

NO! YOU CAN'T DIAGONALIZE EVERY MATRIX!

Definition A matrix $x \in M_n(\mathbb{C})$ is said to be *semisimple* if it can be diagonalized.

Remarks:

1. Note well: The eigenvalues of A are zeroes of the characteristic polynomial.

2. We will discuss Hilbert spaces in Section §13 below. When discussing operators T on Hilbert space one must distinguish eigenvalues of T from the elements of the *spectrum* of T. For a (bounded) operator T on a Hilbert space we define the *resolvent* of T to be the set of complex numbers λ so that $\lambda\mathbf{1} - T$ is 1-1 and onto. The complement of the resolvent is defined to be the *spectrum* of T. In infinite dimensions there are different ways in which the condition of being $1-1$ and invertible can go wrong. The *point spectrum* consists of the eigenvalues, that is, the set of λ so that $\ker(\lambda\mathbf{1} - T)$ is a nontrivial subspace of the Hilbert space. In general it is a proper subset of the spectrum of T. See section 18.3 below for much more detail.

2. Theorem 10.3.1 is definitely false if we replace $\mathbb{C}$ by $\mathbb{R}$. It is also false in infinite dimensions. For example, the Hilbert hotel operator has no eigenvector. To define the Hilbert hotel operator choose an ON basis $\phi_1, \phi_2, \ldots$ and let $S : \phi_i \to \phi_{i+1}$, $i = 1, \ldots$. In terms of harmonic oscillators we can represent S as

$$S = \frac{1}{\sqrt{a^\dagger a}} a^\dagger. \tag{10.18}$$

Exercise
If A is 2×2 show that

$$p_A(x) = x^2 - x\mathrm{tr}\,(A) + \det(A) \tag{10.19}$$

We will explore the generalization later.

Exercise
Show that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{10.20}$$

cannot be diagonalized.

(Note that it does have an eigenvector, of eigenvalue 0.)

10.4 Jordan Canonical Form

Although you cannot diagonalize every matrix, there is a canonical form which is "almost" as good: Every matrix $A \in M_n(\mathbb{C})$ can be brought to *Jordan canonical form* by conjugation by $S \in GL(n, \mathbb{C})$:

$$A \to S^{-1}AS \tag{10.21}$$

We will now explain this

Definition: A $k \times k$ matrix of the form:

$$J_\lambda^{(k)} = \begin{pmatrix} \lambda & 1 & & & 0 \\ 0 & \lambda & 1 & & \cdot \\ 0 & 0 & \lambda & 1 & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & 0 \\ & & & & 1 \\ 0 & \cdot & \cdot & \cdot & 0 \ \lambda \end{pmatrix} \tag{10.22}$$

is called an *elementary Jordan block belonging to* λ. We can also write

$$J_\lambda^{(k)} = \lambda 1 + \sum_{i=1}^{k-1} e_{i,i+1} \tag{10.23}$$

Example: The first few elementary Jordan blocks are:

$$J_\lambda^{(1)} = \lambda, \qquad J_\lambda^{(2)} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \qquad J_\lambda^{(3)} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \cdots \tag{10.24}$$

Exercise *Jordan blocks and nilpotent matrices*
Write the Jordon block as

$$J_\lambda^{(k)} = \lambda \mathbf{1_k} + N^{(k)} \tag{10.25}$$

Show that $(N^{(k)})^\ell \neq 0$ for $\ell < k$ but $(N^{(k)})^k = 0$.

The $J_\lambda^{(k)}$ are the atoms in the world of matrices with complex matrix elements. They cannot be broken into more elementary parts by similarity transformation. For $k > 1$ the Jordan blocks cannot be diagonalized. One easy way to see this is to write:

$$J_\lambda^{(k)} - \lambda \mathbf{1_k} = N^{(k)} \tag{10.26}$$

eq:nilpjord

If $J_\lambda^{(k)}$ could be diagonalized, then so could the LHS of (10.26). However $N^{(k)}$ cannot be diagonalized since $(N^{(k)})^k = 0$, and $N^{(k)} \neq 0$. Another proof uses the characteristic polynomial. The characteristic polynomial of a diagonalizable matrix is $p_A(x) = \prod(x - \lambda_i)$. Now note that the characteristic polynomial of the Jordan matrix is:

$$p_{J^{(k)}}(x) := \det[x1 - J_\lambda^{(k)}] = (x - \lambda)^k \tag{10.27}$$

eq:charpol

Hence if J could be diagonalized it would have to equal $SJ_\lambda^{(k)}S^{-1} = Diag\{\lambda\}$. But then we can invert this to $J_\lambda^{(k)} = S^{-1}Diag\{\lambda\}S = Diag\{\lambda\}$, a contradiction for $k > 1$.

Theorem/Definition: *Every* matrix $A \in M_n(\mathbb{C})$ can be conjugated to Jordan canonical form over the complex numbers. A Jordan canonical form is a matrix of the form:

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ & & \\ & \ddots & \\ & & \\ 0 & & A_s \end{pmatrix} \tag{10.28}$$

where we have blocks A_i corresponding to the *distinct* roots $\lambda_1, \ldots, \lambda_s$ of the characteristic polynomial $p_A(x)$ and each block A_i has the form:

$$A_i = \begin{pmatrix} J_{\lambda_i}^{(k_1^i)} & & & & \\ & J_{\lambda_i}^{(k_2^i)} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & J_{\lambda_i}^{(k_{\ell_i}^i)} \end{pmatrix} \tag{10.29}$$

where $J_{\lambda_i}^{(k_j^i)}$ is the elementary Jordan blocks belonging to λ_i and

$$n = \sum_{i=1}^{s} \sum_{t=1}^{\ell_i} k_t^i \tag{10.30}$$

Proof: We sketch a proof briefly below. See also Herstein section 6.6.

Note that the *characteristic polynomial* looks like

$$\begin{aligned} p(x) \equiv \det[x1 - A] &= \prod_i (x - \lambda_i)^{\kappa_i} \\ \kappa_i &= \sum_{j=1}^{\ell_i} k_j^i \end{aligned} \tag{10.31}$$

eq:charpolp

Remarks

1. Thus, if the roots of the characteristic polynomial are all distinct then the matrix is diagonalizable. This condition is sufficient, but not necessary, since $\lambda \mathbf{1_n}$ is diagonal, but has multiple characteristic values.

2. The above theorem implies that every matrix A can be put in the form:

$$A = A_{ss} + A_{nilp} \tag{10.32}$$

where A_{ss} ("the semisimple part") is diagonalizable and A_{nilp} is nilpotent. Note that if D is diagonal then

$$\mathrm{tr}\, D(N^{(k)})^{\ell} = 0 \tag{10.33}$$

for $\ell > 0$ and hence

$$\mathrm{tr}\, A^n = \mathrm{tr}\, A_{ss}^n. \tag{10.34}$$

Thus, the traces of powers of a matrix do not characterize A uniquely, unless it is diagonalizable.

Exercise *Jordan canonical form for nilpotent operators*

If $T : V \to V$ is a nilpotent linear transformation show that there are vectors $v_1, \ldots, v_\ell$ so that V has a basis of the form:

$$\begin{aligned} \mathcal{B} = \{ & v_1, Tv_1, T^2v_1, \ldots, T^{b_1-1}v_1; \\ & v_2, Tv_2, \ldots, T^{b_2-1}v_2; \\ & \cdots \\ & v_\ell, Tv_\ell, \ldots, T^{b_\ell-1}v_\ell \} \end{aligned} \tag{10.35}$$

eq:JCF

where $\dim V = b_1 + \cdots + b_\ell$ and

$$T^{b_j} v_j = 0 \qquad j = 1, \ldots, \ell \tag{10.36}$$

Exercise *The Cayley-Hamilton theorem*

If $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ is a polynomial in x then we can evaluate it on a matrix $A \in Mat_n(k)$:

$$f(A) := a_0 + a_1 A + \cdots + a_m A^m \in Mat_n(k) \tag{10.37}$$

a.) Show that if $p_A(x)$ is the characteristic polynomial of A then

$$p_A(A) = 0 \tag{10.38}$$

b.) In general, for any matrix A, there is a smallest degree polynomial $m_A(x)$ such that $m_A(x = A) = 0$. This is called *the minimal polynomial of* A. In general $m_A(x)$

divides $p_A(x)$, but might be different from $p_A(x)$. Give an example of a matrix such that $m_A(x) \neq p_A(x)$. [26]

Exercise *A useful identity*

In general, given a power series $f(x) = \sum_{n\geq 0} a_n x^n$ we can evaluate it on a matrix $f(A) = \sum_{n\geq 0} a_n A^n$. In particular, we can define $\exp(A)$ and $\log A$ for a matrix A by the power series expansions of $\exp(x)$ in x and $\log(x)$ in $(1-x)$.

a.) Show that

$$\det\exp A = \exp\mathrm{Tr} A \tag{10.39}$$

Sometimes written in the less accurate form

$$\det M = \exp\mathrm{Tr}\log M \tag{10.40}$$

b.) Suppose M is invertible, and δM is "small" compared to M. Show that

$$\begin{aligned}\sqrt{\det(M+\delta M)} = \sqrt{\det M}\Bigg[1 + \frac{1}{2}\mathrm{Tr}(M^{-1}\delta M) \\ + \frac{1}{8}(\mathrm{Tr}(M^{-1}\delta M))^2 - \frac{1}{4}\mathrm{Tr}(M^{-1}\delta M)^2 + \mathcal{O}((\delta M)^3)\Bigg]\end{aligned} \tag{10.41}$$

Exercise *Jordan canonical form and cohomology*

Recall from Section **** that a chain complex has a degree one map $Q : M \to M$ with $Q^2 = 0$. If M is a complex vector space of dimension $d < \infty$ show that the Jordan form of Q is

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes 1_{d_1} \oplus 0_{d_2} \tag{10.42}$$

where $d = 2d_1 + d_2$. Show that the cohomology is isomorphic to $\mathbb{C}^{d_2}$.

Exercise *Nilpotent* 2×2 *matrices*

A matrix such that $A^m = 0$ for some $m > 0$ is called *nilpotent*.

a.) Show that any 2×2 nilpotent matrix must satisfy the equation:

$$A^2 = 0 \tag{10.43}$$

[26] *Answer*: The simplest example is just $A = \lambda \mathbf{1}_k$ for $k > 1$. For a more nontrivial example consider $A = J^{(2)}_\lambda \oplus J^{(2)}_\lambda$.

b.) Show that any 2×2 matrix solving $A^2 = 0$ is of the form:

$$A = \begin{pmatrix} x & y \\ z & -x \end{pmatrix} \tag{10.44}$$

where

$$x^2 + yz = 0. \tag{10.45}$$ eq:aone

The solutions to Equation (10.45) form a complex variety known as the A_1 singularity. It is a simple example of a (singular, noncompact) Calabi-Yau variety.

c.) If A is a 2×2 matrix do $\mathrm{tr}\, A, \det A$ determine its conjugacy class?

Exercise *Nilpotent matrices*

A matrix such that $A^m = 0$ for some $m > 0$ is called *nilpotent*.

a.) Characterize the matrices for which $p_A(x) = x^n$.

b.) Characterize the $n \times n$ matrices for which $N^2 = 0$.

c.) Characterize the $n \times n$ matrices for $N^k = 0$ for some integer k.

Exercise *Flat connections on a punctured sphere*

The moduli space of *flat* $GL(2,\mathbb{C})$ *connections on the three punctured sphere* is equivalent to the set of pairs (M_1, M_2) of two matrices in $GL(2,\mathbb{C})$ up to simultaneous conjugation. Give an explicit description of this space.

10.4.1 Proof of the Jordan canonical form theorem

We include the proof here for completeness.

♣Should say exactly where you use the ground field being $\mathbb{C}$. ♣

Part 1: Decompose the space according to the different characteristic roots:

Let $A \in Mat_n(\mathbb{C})$ be a matrix acting on $\mathbb{C}^n$. Let $m_A(x)$ be the minimal polynomial of A. This is the polynomial of least degree such that $m_A(A) = 0$. Suppose $m_A(x) = q_1(x)q_2(x)$, where q_1, q_2 are relatively prime polynomials. Then define

$$\begin{aligned} V_1 &:= \{v : q_1(A)v = 0\} \\ V_2 &:= \{v : q_2(A)v = 0\} \end{aligned} \tag{10.46}$$ eq:jordproof

We claim $\mathbb{C}^n = V_1 \oplus V_2$, and A acts block-diagonally in this decomposition.

To see this, note that there exist polynomials $r_1(x), r_2(x)$ such that

$$q_1(x)r_1(x) + q_2(x)r_2(x) = 1 \tag{10.47}$$ eq:relprime

If $u \in V_1 \cap V_2$ then, applying (10.47) with $x = A$ to u we get $u = 0$. Thus

$$V_1 \cap V_2 = \{0\} \tag{10.48}$$ eq:inter

Now, apply (10.47) to any vector u to get:

$$\begin{aligned} u &= q_1(A)(r_1(A)u) + q_2(A)(r_2(A)u) \\ &= u_1 \oplus u_2 \end{aligned} \tag{10.49}$$

Since $q_2(A)u_1 = m_A(A)(r_1(A)u) = 0$ we learn that $u_1 \in V_2$, and similarly $u_2 \in V_1$. Thus, any vector u is in the sum $V_1 + V_2$. Combined with (10.48) this means

$$\mathbb{C}^n = V_1 \oplus V_2. \tag{10.50}$$

Finally, V_1, V_2 are invariant under A. Thus, A acts block diagonally on $V_1 \oplus V_2$.

Now, factoring $m_A(x) = \prod_i (x - \lambda_i)^{\rho_i}$ into distinct roots we obtain a block decomposition of A on $\mathbb{C}^n = \oplus_i V_i$ where $(x - \lambda_i)^{\rho_i}$ is the minimal polynomial of A on V_i. Consider A restricted to V_i. We can subtract $\lambda_i 1$, which is invariant under all changes of basis to assume that $A_i^{\rho_i} = 0$.

Part 2: Thus, the proof of Jordan decomposition is reduced to the Jordan decomposition of matrices M on $\mathbb{C}^n$ such that the minimal polynomial is $M^\rho = 0$. [27]

We will approach this by showing using induction on $\dim V$ that a nilpotent operator $T : V \to V$ has a basis of the form (10.35). The initial step is easily established for $\dim V = 1$ (or $\dim V = 2$). Now for the inductive step note that if T is nilpotent and nonzero then $T(V) \subset V$ is a *proper subspace*. After all, if $T(V) = V$ then applying T successively we would obtain a contradiction. Then, by the inductive hypothesis there must be a basis for $T(V)$ of the form given in (10.35).

Now, let us consider the kernel of T. This contains the *linearly independent* vectors $T^{b_1}v_1, \ldots, T^{b_\ell}v_\ell$. We can complete this to a basis for $\ker T$ with some vectors $w_1, \ldots, w_m$. Now, since $v_i \in T(V)$ there must be vectors $u_i \in V$ with $T(u_i) = v_i$. Then, we claim,

$$\begin{aligned} \mathcal{B} = \{&u_1, Tu_1, T^2v_1, \ldots, T^{b_1}u_1; \\ &u_2, Tu_2, \ldots, T^{b_2}u_2; \\ &\cdots \\ &u_\ell, Tu_\ell, \ldots, T^{b_\ell}u_\ell; \\ &w_1, \ldots, w_m\} \end{aligned} \tag{10.51}$$ eq:JCF-2

is a basis for V. Of course, we have $T^{b_j+1}u_j = 0$ and $Tw_i = 0$. First, these vectors are

[27] Here we follow a very elegant proof given by M. Wildon at http://www.ma.rhul.ac.uk/ uvah099/Maths/JNFfinal.pdf.

linearly independent: Suppose there were a relation of the form

$$\begin{aligned} 0 = & \kappa_1^0 u_1 + \kappa_1^1 T u_1 + \cdots + \kappa_1^{b_1} T^{b_1} u_1 \\ & + \kappa_2^0 u_2 + \kappa_2^1 T u_2 + \cdots + \kappa_2^{b_2} T^{b_2} u_2 \\ & + \cdots \\ & + \kappa_\ell^0 u_\ell + \kappa_\ell^1 T u_\ell + \cdots + \kappa_\ell^{b_\ell} T^{b_\ell} u_\ell + \\ & + \xi_1 w_1 + \cdots + \xi_m w_m \end{aligned} \tag{10.52}$$

Apply T to this equation and use linear independence of the basis for $T(V)$, then use linear independence of the basis for $\ker T$. Finally, we can see that (10.51) is complete because

$$\begin{aligned} \dim V &= \dim\ker T + \dim\mathrm{im} T \\ &= (\ell + m) + (b_1 + \cdots + b_\ell) \\ &= \sum_{j=1}^{\ell} (b_j + 1) + m \end{aligned} \tag{10.53}$$

This completes the argument. ♠

10.5 The stabilizer of a Jordan canonical form

Given a matrix $x \in M_n(\mathbb{C})$ it is often important to understand what matrices will commute with it.

For example, if $x \in GL(n, \mathbb{C})$ we might wish to know the stabilizer of the element under the action of conjugation:

$$Z(x) := \{ g : g^{-1} x g = x \} \subset GL(n, \mathbb{C}) \subset M_n(\mathbb{C}) \tag{10.54}$$

In order to find the commutant of x it suffices to consider the commutant of its Jordan canonical form. Then the following theorem becomes useful:

Lemma Suppose k, ℓ are positive integers and A is a $k \times \ell$ matrix so that

$$J_{\lambda_1}^{(k)} A = A J_{\lambda_2}^{(\ell)} \tag{10.55}$$ eq:CommuteJ

Then

1. If $\lambda_1 \neq \lambda_2$ then $A = 0$.
2. If $\lambda_1 = \lambda_2 = \lambda$ and $k = \ell$ then A is of the form

$$A^{(k)}(\alpha) = \alpha_0 \cdot 1 + \alpha_1 J_\lambda^{(k)} + \alpha_2 (J_\lambda^{(k)})^2 + \cdots + \alpha_{k-1} (J_\lambda^{(k)})^{k-1} \tag{10.56}$$ eq:A-alpha

   for some set of complex numbers $\alpha_0, \ldots, \alpha_{k-1}$.
3. If $\lambda_1 = \lambda_2 = \lambda$ and $k < \ell$ then A is of the form

$$\begin{pmatrix} 0 & A^{(k)}(\alpha) \end{pmatrix} \tag{10.57}$$

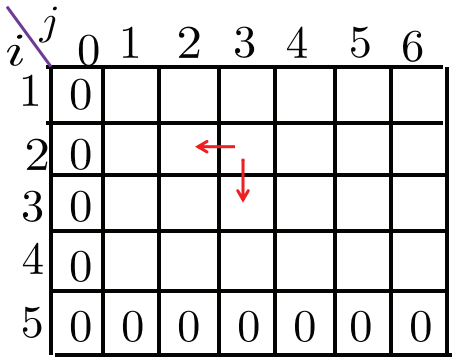


Figure 4: The commutation relation implies that entries in a box are related to those to the left and underneath in this enhanced matrix, as indicated by the arrows. fig:JORDANCOMM

4. If $\lambda_1 = \lambda_2 = \lambda$ and $k > \ell$ then A is of the form

$$\begin{pmatrix} A^{(\ell)}(\alpha) \\ 0 \end{pmatrix} \tag{10.58}$$

Proof: Write

$$A = \sum_{i=1}^{k} \sum_{j=1}^{\ell} A_{ij} e_{ij} \tag{10.59}$$

Then (10.55) is equivalent to

$$(\lambda_2 - \lambda_1) A = \sum_{i=1}^{k-1} \sum_{j=1}^{\ell} A_{i+1,j} e_{ij} - \sum_{i=1}^{k} \sum_{j=2}^{\ell} A_{i,j-1} e_{ij} \tag{10.60}$$ eq:CommuteJ2

Now enhance the matrix A_{ij} to $\hat{A}_{ij}$ by adding a row $i = k+1$ and a column $j = 0$ so that

$$\hat{A}_{ij} = \begin{cases} A_{ij} & 1 \le i \le k, 1 \le j \le \ell \\ 0 & i = k+1 \quad \text{or} \quad j = 0 \end{cases} \tag{10.61}$$

so (10.60) becomes

$$(\lambda_2 - \lambda_1) \hat{A}_{ij} = \hat{A}_{i+1,j} - \hat{A}_{i,j-1} \qquad 1 \le i \le k, 1 \le j \le \ell \tag{10.62}$$

Now consider Figure 4. If $\lambda_1 \neq \lambda_2$ then we use the identity in the $i = k, j = 1$ box to conclude that $A_{k,1} = 0$. Then we successively use the identity going up the $j = 1$ column to find that $A_{i,1} = 0$ for all i. Then we start again at the bottom of the $j = 2$ column and work up. In this way we find $A = 0$. If $\lambda_1 = \lambda_2$ the identity just tells us that two entries along a diagonal have to be the same. Thus all diagonals with one of the zeros on the edge must be zero. The other diagonals can be arbitrary. But this is precisely what a matrix of type $A^{(k)}(\alpha)$ looks like. ♠

Remark: Note that the matrix $A^{(k)}(\alpha)$ in (10.56) above can be rewritten in the form

$$A^{(k)} = \beta_0 + \beta_1 N^{(k)} + \beta_2 (N^{(k)})^2 + \cdots + \beta_{k-1}(N^{(k)})^{k-1} \tag{10.63}$$

which has the form, for example for $k = 5$:

$$A^{(5)} = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ 0 & \beta_0 & \beta_1 & \beta_2 & \beta_3 \\ 0 & 0 & \beta_0 & \beta_1 & \beta_2 \\ 0 & 0 & 0 & \beta_0 & \beta_1 \\ 0 & 0 & 0 & 0 & \beta_0 \end{pmatrix} \tag{10.64}$$

It is clear from this form that the matrix is invertible iff $\beta_0 \neq 0$.

A consequence of this Lemma is that for any $x \in M_n(\mathbb{C})$ the subgroup $Z(x) \subset GL(n, \mathbb{C})$ must have complex dimension at least n. Some terminology (which is common in the theory of noncompact Lie groups) that one might encounter is useful to introduce here in its simplest manifestation:

Definition x is said to be *regular* if the dimension of $Z(x)$ is precisely n. That is, a regular element has the smallest possible centralizer.

Exercise *Regular and semisimple*

Show that x is regular and semisimple iff *all the roots of the characteristic polynomial are distinct.* We will use this term frequently in following sections.

10.5.1 Simultaneous diagonalization

A second application of the above Lemma is simultaneous diagonalization:

Theorem: Two diagonalizable matrices which commute $[A_1, A_2] = 0$ are simultaneously diagonalizable.

Proof: If we first diagonalize A_1 then we get diagonal blocks corresponding to the distinct eigenvalues λ_i:

$$A_1 = \begin{pmatrix} \lambda_1 \mathbf{1}_{\mathbf{r_1} \times \mathbf{r_1}} & 0 & \cdots \\ 0 & \lambda_2 \mathbf{1}_{\mathbf{r_2} \times \mathbf{r_2}} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \tag{10.65}$$

We have now "broken the gauge symmetry" to $GL(r_1, \mathbb{C}) \times GL(r_2, \mathbb{C}) \times \cdots$. Moreover, since the λ_i are distinct, a special case of our lemma above says that A_2 must have a block-diagonal form:

$$A_2 = \begin{pmatrix} A_2^{(1)} & 0 & \cdots \\ 0 & A_2^{(2)} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \tag{10.66}$$

and we can now diagonalize each of the blocks without spoiling the diagonalization of A_1. ♠

In quantum mechanics this theorem has an important physical interpretation: Commuting Hermitian operators are observables whose eigenvalues can be simultaneously measured.

Exercise *Hilbert scheme of points*

Warning: This is a little more advanced than what is generally assumed here.

Suppose two $N \times N$ *complex* matrices X_1, X_2 commute

$$[X_1, X_2] = 0 \tag{10.67}$$ eq:hilscha

What can we say about the pair up to simultaneous conjugation? If they can be simultaneously diagonalized then we may write:

$$SX_iS^{-1} = Diag\{z_i^{(1)}, \dots, z_i^{(N)}\} \qquad i = 1, 2 \tag{10.68}$$ eq:hilschb

and hence we can associate to (10.67) N points $(z_1^{(k)}, z_2^{(k)}) \in \mathbb{C}^2$, $k = 1, 2, \dots, N$. In fact, because of conjugation, they are N unordered points. Thus the set of diagonalizable X_i satisfying (10.67) is the symmetric product $\mathrm{Sym}^N(\mathbb{C}^2)$.

In general, we can only put X_1, X_2 into Jordan canonical form. Thus, the "moduli space" of conjugacy classes of pairs (X_1, X_2) of simultaneously diagonalizable matrices is more complicated. This is still not a good space. To get a good space we consider only the conjugacy classes of "stable triples." These are triples (X_1, X_2, v) where $[X_1, X_2] = 0$ and $v \in \mathbb{C}^N$ is a vector such that $X_1^{n_1} X_2^{n_2} v$ span $\mathbb{C}^N$. In this case we get a very interesting smooth variety known as the *Hilbert scheme of N points on $\mathbb{C}^2$*.

a.) Show that $Hilb^N(\mathbb{C}^2)$ can be identified with the set of ideals $I \subset \mathbb{C}[z_1, z_2]$ such that $\mathbb{C}[z_1, z_2]/I$ is a vector space of dimension N.

Hint: Given an ideal of codimension N, observe that multiplication by z_1, z_2 on $\mathbb{C}[z_1, z_2]/I$ define two commuting linear operators. Conversely, given (X_1, X_2, v) define $\phi : \mathbb{C}[z_1, z_2] \to \mathbb{C}^N$ by $\phi(f) := f(X_1, X_2)v$.

b.) Write matrices corresponding to the ideal

$$I = (z_1^N, z_2 - (a_1 z_1 + \cdots + a_{N-1} z_1^{N-1})) \tag{10.69}$$

The point is, by allowing nontrivial Jordan form, the singular space $\mathrm{Sym}^N(\mathbb{C}^2)$ is resolved to the nonsingular space $\mathrm{Hilb}^N(\mathbb{C}^2)$.

Exercise *ADHM equations*

Find the general solution of the 2×2 ADHM equations:

$$\begin{aligned} &[X_1, X_1^\dagger] + [X_2, X_2^\dagger] + ii^\dagger - j^\dagger j = \zeta_R \\ &[X_1, X_2] + ij = 0 \end{aligned} \tag{10.70}$$ eq:adhm

modulo $U(2)$ transformations.

---

11. Sesquilinear forms and (anti)-Hermitian forms

sec:sips

Definition 11.1 A *sesquilinear form* on a complex vector space V is a function

$$h : V \times V \to \mathbb{C} \tag{11.1}$$

which is anti-linear in the first variable and linear in the second. That is: for all $v_i, w_i \in V$ and $\alpha_i \in \mathbb{C}$:

$$\begin{aligned} h(v, \alpha_1 w_1 + \alpha_2 w_2) &= \alpha_1 h(v, w_1) + \alpha_2 h(v, w_2) \\ h(\alpha_1 v_1 + \alpha_2 v_2, w) &= \alpha_1^* h(v_1, w) + \alpha_2^* h(v_2, w) \end{aligned} \tag{11.2}$$

Note that h defines a $\mathbb{C}$-linear map $\bar{V} \times V \to \mathbb{C}$ and hence (by the universal property) factors through a unique $\mathbb{C}$-linear map

$$\tilde{h} : \bar{V} \otimes V \to \mathbb{C} \tag{11.3}$$

Conversely, such a map defines a sesquilinear form. *Thus, the space of all sesquilinear forms, which is itself a complex vector space, is isomorphic to* $(\bar{V} \otimes V)^*$. We write:

$$\mathrm{Sesq}(V) \cong (\bar{V} \otimes V)^* \tag{11.4}$$

Now note that there are canonical maps

$$\mathrm{Sesq}(V) \otimes V \to \bar{V}^* \tag{11.5}$$ eq:sesqmap-1

$$\mathrm{Sesq}(V) \otimes \bar{V} \to V^* \tag{11.6}$$ eq:sesqmap-2

given by the contraction $V^* \otimes V \to \mathbb{C}$ and $\bar{V}^* \otimes \bar{V} \to \mathbb{C}$, respectively. Written out more explicitly, what equation (11.6) means is that if we are given a sesquilinear form h and an element $\bar{w} \in \bar{V}$ we get a corresponding element $\ell_{h,\bar{w}} \in V^*$ given by

$$\ell_{h,\bar{w}}(v) := h(w, v) \tag{11.7}$$ eq:sesquisom-3

and similarly, for (11.5), given an element $w \in V$ and an h we get an element $\tilde{\ell}_{h,w} \in (\bar{V})^*$ given by

$$\tilde{\ell}_{h,w}(\bar{v}) := h(v, w) \tag{11.8}$$

Definition 11.2

1. A sesquilinear form is said to be *nondegenerate* if for all nonvanishing $v \in V$ there is some $w \in V$ such that $h(v,w) \neq 0$ and for all nonvanishing $w \in V$ there is some v with $h(v,w) \neq 0$.

2. An *Hermitian form* on a complex vector space V is a sesquilinear form such that for all $v, w \in V$:

$$h(v,w) = (h(w,v))^* \tag{11.9}$$

3. If $h(v,w) = -(h(w,v))^*$ then h is called *skew-Hermitian* or *anti-Hermitian.*

Remarks

1. If h is nondegenerate then (11.5) and (11.6) define *isomorphisms* $V \cong \bar{V}^*$ and $\bar{V} \cong V^*$, respectively. In general there is a canonical anti-linear isomorphism $V \cong \bar{V}$ and hence also a canonical antilinear isomorphism $V^* \cong (\bar{V})^*$. However, as we have stressed, there is no canonical isomorphism $V \cong V^*$. What the above definitions imply is that such an isomorphism is provided by a nondegenerate sesquilinear form.

2. In particular, an anti-linear isomorphism $V \cong V^*$ is provided by a nondegenerate Hermitian form. This is used in quantum mechanics in the Dirac bra-cket formalism where the anti-linear isomorphism $V \to V^*$ is denoted

$$|\psi\rangle \to \langle\psi| \tag{11.10}$$

So say this more precisely in the above language: If $v \in V$ is denoted $|\psi\rangle$ and $w \in V$ is denoted $|\chi\rangle$ then, *given an underlying nondegenerate sesquilinear form* we can write

$$h(v,w) = \ell_{h,\bar{v}}(w) = \langle\psi|\chi\rangle \tag{11.11}$$

If the underlying sesquilinear form is hermitian then $\langle\psi|\psi\rangle$ will be real. In fact, since probabilities are associated with such expressions we want it to be positive. This leads us to inner product spaces.

Exercise

Show that the most general Hermitian form on a complex vector space, expressed wrt a basis e_i is

$$h(\sum z_i e_i, \sum w_j e_j) = \sum_{i,j} z_i^* h_{ij} w_j \tag{11.12}$$

where $(h_{ij})^* = h_{ji}$. That is, h_{ij} is an Hermitian matrix.

---

12. Inner product spaces, normed linear spaces, and bounded operators

12.1 Inner product spaces

Definition 11.3 An *inner product space* is a vector space V over a field k ($k = \mathbb{R}$ or $k = \mathbb{C}$ here) with a positive Hermitian form. That is we have a k-valued function

$$(\cdot,\cdot) : V \times V \to k \tag{12.1}$$

satisfying the four axioms:

i.) $(x, y+z) = (x,y) + (x,z)$

ii.) $(x, \alpha y) = \alpha(x,y)$ $\alpha \in k$

iii.) $(x,y) = (y,x)^*$

iv.) $\forall x$, the *norm of* x:

$$\| x \|^2 := (x,x) \geq 0 \tag{12.2}$$

and moreover $(x,x) = 0 \leftrightarrow x = 0$.

Axioms (i),(ii),(iii) say we have a symmetric quadratic form for $k = \mathbb{R}$ and an Hermitian form for $k = \mathbb{C}$. The fourth axiom tells us that the form is not only nondegenerate, but "positive." [28] In quantum mechanics, we usually deal with such inner products because of the probability interpretation of the values (ψ, ψ). Probabilities should be nonnegative.

Example 1: $\mathbb{C}^n$ with

$$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} (x^i)^* y^i \tag{12.3}$$

Example 2: $\mathbb{R}^n$, here $k = \mathbb{R}$ and

$$(\vec{x}, \vec{y}) = \vec{x} \cdot \vec{y} = \sum x^i y^i \tag{12.4}$$

Example 3: $\mathcal{C}[a,b]$= the set of complex-valued continuous functions on the interval $[a,b]$.

$$(f,g) := \int_a^b f(x)^* g(x) dx \tag{12.5}$$

eq:ContFun

Definition 11.4: A set of vectors $\{x_i\}$ in an inner product space is called *orthonormal* [29] if $(x_i, x_j) = \delta_{ij}$.

Theorem 11.1: If $\{x_i\}_{i=1,\ldots,N}$ is an ON set then

$$\| x \|^2 = \sum_{i=1}^{N} |(x, x_i)|^2 + \| x - \sum_{i=1}^{N} (x_i, x) x_i \|^2 \tag{12.6}$$

Proof: Note that $\sum (x_i, x) x_i$ and $x - \sum (x_i, x) x_i$ are orthogonal. ♠

Theorem 11.2: (Bessel's inequality)

$$\| x \|^2 \geq \sum_{i=1}^{N} |(x, x_i)|^2 \tag{12.7}$$

[28] The term "positive" is standard usage, although "nonnegative" would be more logical.

[29] We often abbreviate to ON

Proof: Immediate from the previous theorem.

Theorem 11.3: (Schwarz inequality)

$$\| x \| \cdot \| y \| \geq |(x, y)| \tag{12.8}$$

Proof: It is true for $y = 0$. If $y \neq 0$ then note that $\{y/ \| y \|\}$ is an ON set. Apply Bessel's inequality ♠

12.2 Normed linear spaces

A closely related notion to an inner product space is

Definition: A *normed linear space* or *normed vector space* is a vector space V (over $k = \mathbb{R}$ or $k = \mathbb{C}$) with a function $\| \cdot \| : V \to \mathbb{R}$ such that

i.) $\| v \| \geq 0$, $\forall v \in V$

ii.) $\| v \| = 0$ iff $v = 0$

iii.) $\| \alpha v \| = |\alpha| \, \| v \|$

iv.) $\| v + w \| \leq \| v \| + \| w \|$

An inner product space is canonically a normed linear space because we can define

$$\| v \| = +\sqrt{(v, v)} \tag{12.9}$$

and verify all the above properties. However, the converse is not necessarily true. See the exercise below. The canonical example of normed linear spaces which are not inner product spaces are the bounded operators on an infinite-dimensional Hilbert space. See below.

Exercise *Another proof of the Schwarz inequality*
Note that $\| x - \lambda y \|^2 \geq 0$. Minimize wrt λ.

Exercise *Polarization identity and the parallelogram theorem*

a.) Given an inner product space V, prove that the inner product can be recovered from the norm using the polarization identity:

$$(x, y) = \frac{1}{4}[(\| x + y \|^2 - \| x - y \|^2) - i(\| x + iy \|^2 - \| x - iy \|^2)] \tag{12.10}$$

which is also sometimes written as

$$4(y, x) = \sum_{k=0}^{4} i^k (x + i^k y, x + i^k y) \tag{12.11}$$

b.) Prove that conversely a normed linear space is an inner product space if and only if the norm satisfies the parallelogram law:

$$\| x+y \|^2 + \| x-y \|^2 = 2 \| x \|^2 + 2 \| y \|^2 \tag{12.12}$$

Warning: This is hard. Start by proving additive linearity in y.

c.) Give an example of a normed linear space which is *not* an inner product space.

12.3 Bounded linear operators

Definition: A *bounded linear transformation* of a normed linear space $(V_1, \| \cdot \|_1)$ to $(V_2, \| \cdot \|_2)$ is a linear map $T : V_1 \to V_2$ such that $\exists C \geq 0$ with

$$\| Tv \|_2 \leq C \| v \|_1 \tag{12.13}$$ eq:bdd

for all $v \in V_1$. In this case, we define the norm of the operator

$$\| T \| = \sup_{\|v\|_1=1} \| Tv \|_2 = \inf C \tag{12.14}$$ eq:bddi

Theorem: For a linear operator between two normed vector spaces

$$T : (V_1, \| \cdot \|_1) \to (V_2, \| \cdot \|_2) \tag{12.15}$$

the following three statements are equivalent:

1. T is continuous at $x = 0$.
2. T is continuous at every $x \in V_1$.
3. T is bounded

Proof: The linearity of T shows that it is continuous at one point iff it is continous everywhere. If T is bounded by C then for any $\epsilon > 0$ we can choose $\delta < \epsilon/C$ and then $\| x \| < \delta$ implies $\| T(x) \| < \epsilon$. Conversely, if T is continuous at $x = 0$ then choose any $\epsilon > 0$ you like. We know that there is a δ so that $\| x \| < \delta$ implies $\| T(x) \| < \epsilon$. But this means that for any $x \neq 0$ we can write

$$\| T(x) \| = \frac{\| x \|}{\delta} \| T\left(\delta \frac{x}{\| x \|}\right) \| < \frac{\epsilon}{\delta} \| x \| \tag{12.16}$$

and hence T is bounded ♠

12.4 Constructions with inner product spaces

A natural question at this point is how the constructions we described for general vector spaces generalize to inner product spaces and normed vector spaces.

1. *Direct sum.* This is straightforward: If V_1 and V_2 are inner product spaces then so is $V_1 \oplus V_2$ where we define
$$(x_1 \oplus y_1, x_1 \oplus y_2)_{V_1 \oplus V_2} := (x_1, x_2)_{V_1} + (y_1, y_2)_{V_2} \tag{12.17}$$

2. *Tensor product* One can extend the tensor product to primitive vectors in the obvious way
$$(x_1 \otimes y_1, x_2 \otimes y_2)_{V_1 \otimes V_2} := (x_1, x_2)_{V_1}(y_1, y_2)_{V_2} \tag{12.18}$$
and then extending by linearity, so that
$$(x_1 \otimes y_1, x_2 \otimes y_2 + x_3 \otimes y_3)_{V_1 \otimes V_2} := (x_1 \otimes y_1, x_2 \otimes y_2)_{V_1 \otimes V_2} + (x_1 \otimes y_1, x_3 \otimes y_3)_{V_1 \otimes V_2} \tag{12.19}$$

3. For $\mathrm{Hom}(V_1, V_2)$ and dual spaces see Section §14.

What about quotient space? If $W \subset V$ is a subset of an inner product space then W of course becomes an inner product space. Can we make V/W an inner product space? Clearly this is problematic since the obvious attempted definition
$$([v_1], [v_2]) \stackrel{?}{=} (v_1, v_2)_V \tag{12.20}$$
would only be well-defined if if
$$(v_1 + w_1, v_2 + w_2) \stackrel{?}{=} (v_1, v_2) \tag{12.21}$$
for all $w_1, w_2 \in W$ and $v_1, v_2 \in V$. This is clearly impossible!

Although we cannot put an inner product on V/W one might ask if there is a canonical complementary space inner product space to W. There is a natural candidate, the *orthogonal complement*, defined by:
$$W^\perp := \{y \in V | \forall x \in W, (x, y) = 0\} \tag{12.22}$$
So the question is, do we have
$$V \stackrel{?}{=} W \oplus W^\perp \tag{12.23}$$
Note that we certainly have $W \cap W^\perp = \{0\}$ by positivity of the inner product. So the question is whether $V = W + W^\perp$. We will see that this is indeed always true, when V is finite dimensional, in Section §15. In infinite-dimensions we must be more careful as the following example shows:

Example: Let $V = \mathcal{C}[0, 1]$ be the inner product space of continuous complex-valued functions on the unit interval with inner product (12.5). Let $W \subset V$ be the subspace of functions which *vanish* on $[\frac{1}{2}, 1]$:
$$W = \{f \in \mathcal{C}[0, 1] | \qquad f(x) = 0 \qquad \frac{1}{2} \leq x \leq 1\} \tag{12.24}$$

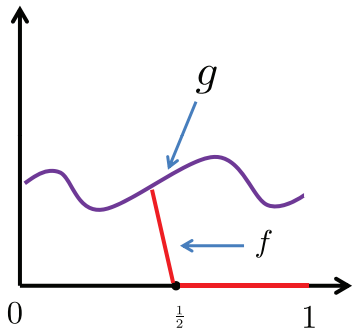


Figure 5: The function $g(x)$ must be orthogonal to all functions $f(x)$ in W. We can use the functions $f(x)$ shown here to see that $g(x)$ must vanish for $0 \leq x \leq \frac{1}{2}$. fig:Continuous

What is $W^\perp$? Any function $g \in W^\perp$ must be orthogonal to the function $f \in W$ which agrees with g for $x \in [0, \frac{1}{2} - \epsilon]$ and then continuously interpolates to $f(\frac{1}{2}) = 0$. See Figure 5. This implies that $g(x) = 0$ for $x < \frac{1}{2}$, but since g is continuous we must have $g(\frac{1}{2}) = 0$. Thus

$$W^\perp = \{g \in \mathcal{C}[0,1] | \qquad g(x) = 0 \qquad 0 \leq x \leq \frac{1}{2}\} \tag{12.25}$$

Now, clearly, $W + W^\perp$ is a *proper* subset of V since it cannot contain the simple function $h(x) = 1$ for all x. We will see that by making the inner product space *complete* we can eliminate such pathology.

13. Hilbert space

c:HilbertSpace

In order to do the analysis required for quantum mechanics one often has to take limits. It is quite important that these limits exist. The notion of inner product space is too flexible.

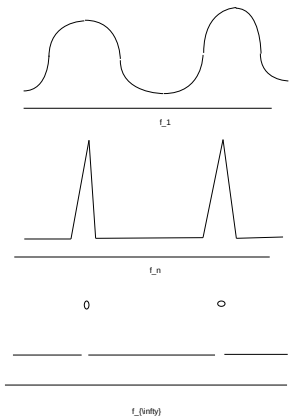


Figure 6: A sequence of continuous functions approaches a normalizable, but discontinuous function. fig:discfunc

For example, $\mathcal{C}[a,b]$ is an inner product space, but it is certainly possible to take a sequence of continuous functions $\{f_n\}$ such that $\| f_n - f_m \| \to 0$ for large n, m but f_n has no limiting continuous function as in 6. That's bad.

Definition: A *complete* inner product space is called a *Hilbert space*

Complete means: every Cauchy sequence converges to a point in the space.

Example 1: $\mathbb{R}^n$ and $\mathbb{C}^n$ are real and complex Hilbert spaces, respectively.

Counter-Example 2: $\mathcal{C}[a,b]$, the continuous complex-valued functions on $[a,b]$ is *not* a Hilbert space.

Example 3: Define

$$L^2[a,b] \equiv \left\{ f : [a,b] \to \mathbb{C} : |f|^2 \text{is measurable} \quad \text{and} \quad \int_a^b |f(x)|^2 dx < \infty \right\} \tag{13.1}$$

It is not obvious, but it is true, that this defines a Hilbert space. To make it precise we need to introduce "measurable functions." For a discussion of this see Reed and Simon. This is the guise in which Hilbert spaces arise most often in quantum mechanics.

Example 4:

$$\ell^2 := \{ \{x_n\}_{n=1}^{\infty} : \sum_{n=1}^{\infty} |x_n|^2 < \infty \} \tag{13.2}$$

♣Should give the example of Bargmann Hilbert space $Hol(\mathbb{C}, e^{-|z|^2} d^2 z)$ since this is a very nice representation for the HO algebra. ♣

Definition Two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are said to be *isomorphic* if there is an inner-product-preserving 1-1 linear transformation *onto* $\mathcal{H}_2$,

$$U : \mathcal{H}_1 \to \mathcal{H}_2 \tag{13.3}$$

such that

$$\forall x, y \in \mathcal{H}_1 : (Ux, Uy)_{\mathcal{H}_2} = (x, y)_{\mathcal{H}_1} \tag{13.4}$$

Such an operator U is also known as a *unitary transformation.*

Remark: In particular, U is norm preserving, that is, it is an *isometry.* By the polarization identity we could simply define U to be an isometry.

What are the invariants of a Hilbert space? By a general set-theoretic argument it is easy to see that every Hilbert space $\mathcal{H}$ has an ON basis. (For a proof see Reed & Simon Theorem II.5). The Bessel and Schwarz inequalities apply. If the basis is countable then $\mathcal{H}$ is called *separable.*

Theorem: Let $\mathcal{H}$ be a separable Hilbert space. Let S be an ON basis. Then
a.) If $|S| = N < \infty$, then $\mathcal{H} \cong \mathbb{C}^N$
b.) If $|S| = \infty$ then $\mathcal{H} \cong \ell^2$

Proof: (Case b): Let $\{y_n\}$ be a complete ON system. Then $U : x \to \{(y_n, x)\}$ is a unitary isomorphism with ℓ_2 ♠

Example 5 Consider $L^2[0, 2\pi]$. Then

$$\{\phi_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}\} \tag{13.5}$$

is a complete ON system. This statement summarizes the Fourier decomposition:

$$f(x) = \sum c_n \phi_n(x) \tag{13.6}$$

Example 6 $L^2(\mathbb{R})$. Note that the standard plane waves $e^{ik\cdot x}$ are *not* officially elements of $L^2(\mathbb{R})$. Nevertheless, there are plenty of ON bases, e.g. take the Hermite functions $H_n(x)e^{-x^2}$ (or any complete set of eigenfunctions of a Schrödinger operator whose potential goes to ∞ at $x \to \pm\infty$) and this shows that $L^2(\mathbb{R})$ is a separable Hilbert space. Indeed, all elementary quantum mechanics courses show that there is a highest weight representation of the Heisenberg algebra so that we have an ON basis $|n\rangle$, $n = 0, 1, \ldots$ and

$$a|n\rangle = \sqrt{n}|n-1\rangle \tag{13.7}$$

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \tag{13.8}$$

The mapping of normalized Hermite functions to $|n\rangle$ gives an isometry of $L^2(\mathbb{R})$ with ℓ^2.

Remarks

1. *Dual space* Our various linear algebra operations can be carried over to Hilbert space provided one uses some care. The direct sum is straightforward. In order for the dual space $\mathcal{H}^*$ to be a *Hilbert space* we must use the *continuous* linear functions to define $\mathcal{H}^*$. Therefore, these are the *bounded* operators in the vector space of all linear operators $\mathrm{Hom}(\mathcal{H}, \mathbb{C})$. This space is again a Hilbert space, and the isomorphism $\mathcal{H} \cong \mathcal{H}^*$ is provided by the *Riesz representation theorem*, which says that every bounded linear functional $\ell : \mathcal{H} \to \mathbb{C}$ is of the form

   $$\ell(v) = (w, v) \tag{13.9}$$

   for some unique $w \in \mathcal{H}$. This is the Hilbert space analog of the isomorphism (11.7).

2. *Tensor product.* We have indicated how to define the tensor product of inner product spaces. However, for Hilbert spaces completeness becomes an issue. See Reed+Simon Sec. II.4 for details. Of course, the tensor product of Hilbert spaces is very important in forming Fock spaces.

3. *"All Hilbert spaces are the same."* One sometimes encounters this slogan. What this means is that all separable Hilbert spaces are isomorphic as Hilbert spaces. Nevertheless Hilbert spaces of states appear in very different physical situations.

E.g. the Hilbert space of QCD on a lattice is a separable Hilbert space, so is it "the same" as the Hilbert space of a one-dimensional harmonic oscillator? While there is indeed an isomorphism between the two, it is not a physically natural one. The physics is determined by the kinds of operators that we consider acting on the Hilbert space. Very different (algebras) of operators are considered in the harmonic oscillator and in QCD. The isomorphism in question would take a physically natural operator in one context to a wierd one in a different context. So, in the next sections we will study in more detail linear operators on vector spaces.

Exercise *An application of the Schwarz inequality to quantum mechanics*
Consider the quantum mechanics of a particle on a line. Show that

$$(\langle\psi|x^2|\psi\rangle)^2 \leq \langle\psi|x^4|\psi\rangle \tag{13.10}$$

by applying the Schwarz inequality to $\psi(x)$ and $x^2\psi(x)$.

Exercise *The uncertainty principle*
Apply the Schwarz inequality to $\psi_1 = x\psi(x)$ and $\psi_2 = k\hat{\psi}(k)$ (where $\hat{\psi}(k)$ is the Fourier transform) by using the Plancherel theorem. Deduce the *uncertainty principle*:

$$\langle\psi|x^2|\psi\rangle\langle\psi|p^2|\psi\rangle \geq \frac{1}{4} \tag{13.11}$$

with minimal uncertainty (saturating the inequality) for the Gaussian wavepackets:

$$\psi(x) = Ae^{-Bx^2} \tag{13.12}$$

---

14. Banach space

ec:BanachSpace

The analog of a Hilbert space for normed linear spaces is called a Banach space:
Definition: A complete normed linear space is called a *Banach space*.

All Hilbert spaces are Banach spaces, but the converse is not true. A key example is the set of bounded operators on Hilbert space:

Theorem $\mathcal{L}(\mathcal{H})$ is a Banach space with the operator norm.

Sketch of proof: $\mathcal{L}(\mathcal{H})$ is a complex vector space and the operator norm makes it a normed linear space (the axioms are easily checked). If $\{A_n\}_{n=1}^{\infty}$ is a Cauchy sequence of operators in the operator norm then for all x, $A_n x$ is a Cauchy sequence of vectors, and hence we can define $Ax = \lim_{n\to\infty} A_n x$, and it is not difficult to show that A is a bounded linear operator and $A_n \to A$ in the operator norm. See Reed-Simon Theorem III.2 for details ♠

Remarks:

1. In fact, the proof of the above theorem proves more: If V_1, V_2 are normed linear spaces and V_2 is complete, that is, if V_2 is a Banach space, then $\mathcal{L}(V_1, V_2)$ is a Banach space, with the operator norm.

♣So, state the theorem in this generality and let $\mathcal{L}(\mathcal{H})$ be a corollary. ♣

2. There are several different notions of convergence of a sequence of operators A_n between normed linear spaces $V_1 \to V_2$, and consequently several different topologies on the space of linear operators. The operator norm defines one topology. Another is the "strong topology" whereby $A_n \to A$ if for all $x \in V_1$, $\lim_{n\to\infty} \| A_n x - Ax \| = 0$. This is different from the norm topology because the rate at which $\| A_n x - Ax \|$ goes to zero might depend in an important way on x. There is also an intermediate "compact-open" topology which can also be useful.

♣Explain more here. ♣

♣Also point out that $C[0,1]$ with the sup norm *is a complete normed linear space.* So it is a Banach space, but not an inner product space. ♣

15. Projection operators and orthogonal decomposition

ctionOperators

As we discussed above, if $W \subset V$ be a subspace of an inner product space then we can define the *orthogonal subspace* by:

$$W^\perp = \{y : (x, y) = 0 \quad \forall x \in W\} \tag{15.1}$$

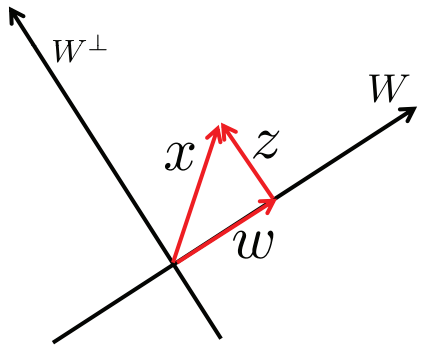


Figure 7: Vectors used in the projection theorem.

fig:Projectior

Theorem(*The projection theorem*) Suppose V is a Hilbert space and $W \subset V$ is a *closed* subspace. Then:

$$V \cong W \oplus W^\perp \tag{15.2}$$

that is, every vector $x \in V$ can be *uniquely* decomposed as $x = z + w$ with $w \in W, z \in W^\perp$.

Proof: We follow the proof in Reed-Simon:

The first point to establish (and the point which fails if we use an inner product space which is not a Hilbert space) is that given any $x \in V$ there is a unique vector $w \in W$ which is closest to x. Let $d := \inf_{w\in W} \| x - w \| \geq 0$. There must be a sequence $w_n \in W$ with

$\lim_{n\to\infty} \| x - w_n \| = d$. But then we can write:

$$\begin{aligned} \| w_n - w_m \|^2 =& \| (w_n - x) + (x - w_m) \|^2 \\ &= 2 \| x - w_n \|^2 + 2 \| x - w_m \|^2 - \| 2x - (w_n + w_m) \|^2 \\ &= 2 \| x - w_n \|^2 + 2 \| x - w_m \|^2 - 4 \| x - \frac{1}{2}(w_n + w_m) \|^2 \\ &\leq 2 \| x - w_n \|^2 + 2 \| x - w_m \|^2 - 4d^2 \end{aligned} \tag{15.3}$$

where in line 2 we used the parallelogram identity. Now we know that the limit of the RHS is zero, therefore, for all $\epsilon > 0$ there is an N so that for $n, m > N$ the RHS is less than ϵ. Therefore $\{w_n\}$ is a Cauchy sequence and, since W is closed, $\lim w_n = w \in W$ exists and minimizes the distance.

Now denote $z := x - w$, where w is the distance minimizing $w \in W$ we have just found. We need to prove that z is in $W^\perp$. Let $d = \| x - w \|$. Then for all $t \in \mathbb{R}, y \in W$:

$$d^2 \leq \| x - (w + ty) \|^2 = \| z - ty \|^2 = d^2 - 2tRe(z, y) + t^2 \| y \|^2 \tag{15.4}$$

This must hold for all t and therefore,

$$Re(z, y) = 0 \tag{15.5}$$

for all y. If we have a real vector space then $(z, y) = 0$ for all $y \in W \Rightarrow z \in W^\perp$, so we are done. If we have a complex vector space we replace $t \to it$ above and prove

$$Im(z, y) = 0 \tag{15.6}$$

Therefore, z is orthogonal to every vector $y \in W$. ♠

Remarks:

1. The theorem definitely *fails* if we drop the positive definite condition on $(\cdot, \cdot)$. Consider $\mathbb{R}^2$ with "inner product" defined on basis vectors e, f by

$$(e, e) = (f, f) = 0 \qquad (e, f) = (f, e) = 1 \tag{15.7}$$

The product is nondegenerate, but if $W = \mathbb{R}e$ then $W^\perp = W$.

In general, for *any* vector space V, not necessarily an inner product space, we can define a *projection operator* to be an operator $P \in End(V)$ such that $P^2 = P$. It then follows from trivial algebraic manipulation that if we define $Q = 1 - P$ we have

1. $Q^2 = Q$ and $QP = PQ = 0$
2. $1 = P + Q$

Moreover we claim that

$$V = \ker P \oplus \mathrm{im} P \tag{15.8}$$

Proof: First, for any linear transformation $\ker T \cap \mathrm{im} T = \{0\}$, so that applies to $T = P$. Next we can write, for any vector v, $v = Pv + (1 - P)v$, and we note that $(1 - P)v \in \ker P$. ♠

In finite dimensions, P can be brought to Jordan form and the equation $P^2 = P$ then shows that it is diagonal with diagonal eigenvalues 0 and 1.

Now, if we are in a *Hilbert space* V and $W \subset V$ is a closed subspace the projection theorem says that every $x \in V$ has a unique decomposition $x = w + w^\perp$ in $W \oplus W^\perp$. Therefore, if we define $P(x) = w$ we have a projection operator. Clearly, Q is the projector to $W^\perp$. We now note that these projectors satisfy the following relation

$$(x_1, Px_2) = (w_1 + w_1^\perp, w_2) = (w_1, w_2) \tag{15.9}$$

$$(Px_1, x_2) = (w_1, w_2 + w_2^\perp) = (w_1, w_2) \tag{15.10}$$

and so for all $x_1, x_2 \in \mathcal{H}$, we have $(Px_1, x_2) = (x_1, Px_2)$. As we will see below, this means that P (and likewise Q) is *self-adjoint*. In general, a self-adjoint projection operator is also known as an *orthogonal projection operator*.

Conversely, given a self-adjoint projection operator P we have an orthogonal decomposition $V = W \oplus W^\perp$. Therefore, *there is a 1-1 correspondence between closed subspaces of V and self-adjoint projection operators.*

As an application we prove

Theorem [Riesz representation theorem]: If $\ell : \mathcal{H} \to \mathbb{C}$ is a bounded linear functional then there exists a unique $y \in \mathcal{H}$ so that $\ell(x) = (y, x)$. Moreover $\| \ell \| = \| y \|$ so that $\mathcal{H}^* \cong \mathcal{H}$.

Proof: If $\ell = 0$ then we take $y = 0$. If $\ell \neq 0$ then there is some vector not in the kernel. Now, because ℓ is continuous $\ker\ell$ is a closed subspace of $\mathcal{H}$ and hence there is an orthogonal projection P to $(\ker\ell)^\perp$. Now, the equation $\ell(x) = 0$ is one (complex) equation and hence the kernel should have complex codimension one. That is, P is a projector to a one-dimensional subspace. Therefore we can choose any nonzero $x_0 \in (\ker\ell)^\perp$ and then

$$P(x) = \frac{(x_0, x)}{(x_0, x_0)} x_0 \tag{15.11}$$

is the projector to the orthogonal complement to $\ker\ell$. If $Q = 1 - P$ then Q is the orthogonal projector to $\ker\ell$ and hence

$$\ell(x) = \ell(P(x) + Q(x)) = \ell(P(x)) = \ell(x_0) \frac{(x_0, x)}{(x_0, x_0)} = (y, x) \tag{15.12}$$

where $y = \frac{\ell(x_0)^*}{(x_0, x_0)} x_0$. Given this representation one easily checks $\| \ell \| = \| y \|$ ♠

Remarks Thus, there is a 1-1 correspondence between projection operators and closed linear subspaces. The operator norm defines a topology on the space of bounded operators, and then the *space of projection operators* is a very interesting manifold, known as the Grassmannian. The Grassmannian has several disconnected components, labelled by the rank of P. Each component has very intricate and interesting topology. Grassmannians of projection operators on Hilbert space are very important in the theory of fiber bundles. In physics they arise naturally in the quantization of free fields in curved spacetime and more recently have played a role both in mathematical properties of the S-matrix of N=4

supersymmetric Yang-Mills theory as well as in the classification of topological phases of matter in condensed matter theory.

Exercise

Suppose that Γ is an operator such that $\Gamma^2 = 1$. Show that

$$\Pi_\pm = \frac{1}{2}(1 \pm \Gamma) \tag{15.13}$$

are projection operators, that is, show that: Show that

$$\Pi_+^2 = \Pi_+ \qquad \Pi_-^2 = \Pi_- \qquad \Pi_+\Pi_- = \Pi_-\Pi_+ = 0 \tag{15.14}$$

---

16. Unitary, Hermitian, and normal operators

Let V, W be finite dimensional inner product spaces. It is important in this section that we are working with finite dimensional vector spaces, otherwise the theory of operators is much more involved. See Reed and Simon and below.

Given

$$T : V \to W \tag{16.1}$$

we can *define* the adjoint operator

$$T^\dagger : W \to V \tag{16.2}$$ eq:defadji

by:

$$\forall x \in W, y \in V \qquad (T^\dagger x, y) := (x, Ty) \tag{16.3}$$

Here we are using the property that the inner product is a nondegenerate form: To define the vector $T^\dagger x$ it suffices to know its inner product with all vectors y. (In fact, knowing the inner product on a basis will do.)

If $T : V \to V$ and $\{v_i\}$ is an ordered *orthonormal* basis, then it follows that the matrices wrt this basis satisfy:

$$(T^\dagger)_{ij} = (T_{ji})^* \tag{16.4}$$ eq:daggmat

It also follows immediately from the definition that

$$\ker T^\dagger = (\mathrm{im} T)^\perp \tag{16.5}$$

Thus, we have the basic picture of an operator between inner product spaces given in Figure 8

Definition:

If V is an inner product space and $T : V \to V$ is linear then:

1. $T^\dagger = T$ defines an *Hermitian* or *self-adjoint* operator.

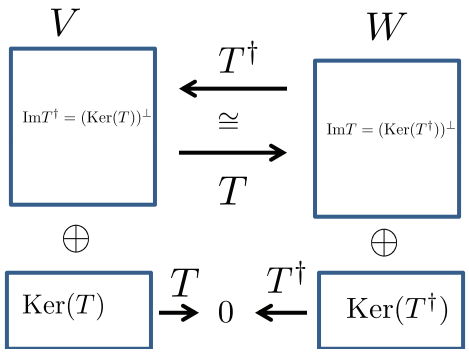


Figure 8: Orthogonal decomposition of domain and range associated to an operator T between inner product spaces.

fig:INNPRODSPA

2. $TT^\dagger = T^\dagger T = 1$ defines a *unitary operator*.
3. $TT^\dagger = T^\dagger T$ defines a *normal operator*.

Put differently, a unitary operator on an inner product space is a norm preserving linear operator (a.k.a. an *isometry*) i.e.

$$\forall v \in V \qquad \| T(v) \|^2 = \| v \|^2 \tag{16.6}$$

It is worthwhile expressing these conditions in terms of the matrices of the operators wrt an ordered ON basis v_i of V. Using (16.4) we see that with respect to an ON basis:

Hermitian matrices H_{ij} satisfy: $H_{ij} = H^*_{ji}$.

Unitary matrices U_{ij} satisfy: $\sum_k U_{ik} U^*_{jk} = \delta_{ij}$

Remarks

1. Recalling that $\mathrm{cok}T := W/\mathrm{im}T$ we see that

$$\mathrm{cok}T \cong \mathrm{ker}T^\dagger \tag{16.7}$$

2. For a finite dimensional Hilbert space $\mathcal{L}(\mathcal{H})$ is an inner product space with natural inner product [30]

$$(A_1, A_2) = \mathrm{Tr} A_1^\dagger A_2 \tag{16.8}$$

but this will clearly not work in infinite dimensions since, for example, the unit operator is bounded but has no trace. In finite dimensions this shows that if V_1, V_2 are inner product spaces then so is $\mathrm{Hom}(V_1, V_2)$ in a natural way. In particular, V^* is an inner product space.

[30] Warning, the norm $\| A \|$ defined by this inner product is *not* the same as the operator norm!

Exercise

a.) If $\mathrm{Tr}T^\dagger T = 0$, show that $T = 0$.

Exercise

Let $\{v_i\}, \{\hat{v}_i\}$ be any two ON bases for a vector space V. Show that the operator defined by $U : v_i \to \hat{v}_i$ is unitary.

Exercise *Unitary vs. Orthogonal*

Show that if $U \in M_n(\mathbb{C})$ is both unitary and real then it is an orthogonal matrix.

Exercise *Eigenvalues of hermitian and unitary operators*

a.) The eigenvalues λ of Hermitian operators are real $\lambda^* = \lambda$. This mathematical fact is compatible with the postulate of quantum mechanics that states that observables are represented by Hermitian operators.

b.) The eigenvalues of unitary operators are phases: $|\lambda| = 1$.

Exercise *The Cayley transform*

The Mobius transformation $z \to w = \frac{1+iz}{1-iz}$ maps the upper half of the complex plane to the unit disk. This transform has an interesting matrix generalization:

a.) Show that if H is Hermitian then

$$U = (1 + iH)(1 - iH)^{-1} \tag{16.9}$$

is a unitary operator.

b.) Show that if U is a unitary operator then $H = i(1 - U)(1 + U)^{-1}$ is Hermitian provided $(1 + U)$ is invertible.

Exercise

a.) Show that if H is Hermitian and α is real then $U = \exp[i\alpha H]$ is unitary.

b.) Show that if H is an Hermitian operator and $\| H \| \leq 1$ then

$$U_\pm = H \pm i\sqrt{1 - H^2} \tag{16.10}$$

are unitary operators.

c.) Show that any matrix can be written as the sum of four unitary matrices.

---

17. The spectral theorem: Finite Dimensions

A nontrivial and central fact in math and physics is:

The (finite dimensional) spectral theorem: Let V be a finite dimensional inner product space over $\mathbb{C}$ and let $H : V \to V$ be an Hermitian operator. Then:

a.) If $\lambda_1, \ldots, \lambda_k$ are the *distinct* eigenvalues of H then there are mutually orthogonal Hermitian projection operators $P_{\lambda_1}, \ldots, P_{\lambda_k}$ such that:

$$H = \sum_i \lambda_i P_{\lambda_i} = \lambda_1 P_{\lambda_1} \oplus \lambda_2 P_{\lambda_2} \oplus \cdots \lambda_k P_{\lambda_k} \tag{17.1}$$

eq:spctrldecom

$$1 = P_{\lambda_1} + \cdots + P_{\lambda_k} \tag{17.2}$$

eq:sprtrlii

b.) There exists an ON basis of V of eigenvectors of H.

Idea of the proof:

We proceed by induction on $\dim V$. The case $\dim V = 1$ is clear.

Now suppose $\dim V = n > 1$. By theorem 10.3.1 of section 10.3 H has one nonvanishing eigenvector v. Consider $W = L(\{v\})$, the span of v. The orthogonal space $W^\perp$ is also an inner product space. Moreover H takes $W^\perp$ to $W^\perp$: If $x \in W^\perp$ then

$$(Hx, v) = (x, Hv) = \lambda(x, v) = 0 \tag{17.3}$$

so $Hx \in W^\perp$. Further, the restriction of H to $W^\perp$ is Hermitian. Since $\dim W^\perp = \dim V - 1$ by the inductive hypothesis there is an ON basis for H on $W^\perp$ and hence there is one for V. ♠

In terms of matrices, any Hermitian matrix is unitarily diagonalizable:

$$H^\dagger = H \Rightarrow \exists U, \quad UU^\dagger = 1 \quad s.t. : UHU^\dagger = Diag\{E_1, \ldots, E_n\}, E_i \in \mathbb{R} \tag{17.4}$$

Exercise

Show that the orthogonal projection operators P_i for H can be written as polynomials in H. Namely, we can take

$$P_{\lambda_i} = \prod_{j \neq i} \left(\frac{H - \lambda_j}{\lambda_i - \lambda_j} \right) \tag{17.5}$$

Exercise

Let $f(x)$ be a convergent power series and let H be Hermitian. Then

$$f(H) = \sum_i f(\lambda_i) P_{\lambda_i} \tag{17.6}$$

Exercise

Let X_n be the symmetric $n \times n$ matrix whose elements are all zero except for the diagonal above and below the principal diagonal. These matrix elements are 1. That is:

$$X_n = \sum_{i=1}^{n-1} e_{i,i+1} + \sum_{i=2}^{n} e_{i,i-1} \tag{17.7}$$

where $e_{i,j}$ is the matrix with 1 in row i and column j, and zero otherwise.

Find the eigenvalues and eigenvectors of X_n.

Exercise

Let H be a positive definite Hermitian matrix. Evaluate

$$\lim_{N\to\infty} \frac{\log(v, H^N v)}{N} \tag{17.8}$$

Exercise *The Seesaw Mechanism*

Consider the matrix

$$H = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{17.9}$$ eq:mtrxp

a.) Assuming H is Hermitian find the eigenvalues and eigenvectors of H.

An important discovery in particle physics of the past few years is that neutrinos have nonzero masses. One idea in neutrino physics is the seesaw mechanism - there are two "flavors" of neutrinos with a mass matrix

$$\begin{pmatrix} 0 & m \\ m^* & M \end{pmatrix} \tag{17.10}$$ eq:seesaw

where M is real. The absolute values of the eigenvalues give the masses of the two neutrinos.

c.) Find the eigenvalues of (17.10), and give a simplified expression for the eigenvalues in the limit where $|m| \ll |M|$.

d.) Suppose it is known that $|m| \cong 1\text{TeV} = 10^3\text{GeV}$, and a neutrino mass of $1eV$ is measured experimentally. What is the value of the large scale M?

e.) For what values of m, M does the kernel of (17.10) jump in dimension? Verify the constancy of the index.

17.1 Normal and Unitary matrices

Together with simultaneous diagonalization we can extend the set of unitarily diagonalizable matrices. Recall that a complex $n \times n$ matrix A is called *normal* if $AA^\dagger = A^\dagger A$,

Theorem:

1. Every normal matrix is diagonalizable by a *unitary matrix*.

2. Every normal operator T on a finite-dimensional inner product space has a spectral decomposition

$$T = \sum_{i=1}^{k} \mu_i P_{\mu_i} \tag{17.11}$$

where $\mu_i \in \mathbb{C}$ and $\{P_{\mu_i}\}$ are mutually orthogonal projection operators summing to the identity.

Proof: Note that we can decompose any matrix as

$$A = H + K \tag{17.12}$$

with $H^\dagger = H$ and $K^\dagger = -K$, antihermitian. Thus, iK is hermitian. If A is normal then $[H, K] = 0$. Thus we have two commuting hermitian matrices, which can be simultaneously diagonalized. ♠

As an immediate corollary we have

Theorem The eigenvectors of a unitary operator on an inner product space V form a basis for V. Every unitary operator on a finite dimensional inner product space is unitarily diagonalizable.

Proof: $U^\dagger U = 1 = UU^\dagger$ so U is a normal matrix. ♠

17.2 Singular value decomposition and Schmidt decomposition

17.2.1 Bidiagonalization

Theorem Any matrix $A \in M_n(\mathbb{C})$ can be *bidiagonalized* by unitary matrices. That is, there always exist unitary matrices $U, V \in U(n)$ such that

$$UAV^\dagger = \Lambda \tag{17.13}$$

is diagonal with nonnegative entries

Proof: First diagonalize $AA^\dagger$ and $A^\dagger A$ by U, V, so $UAA^\dagger U^\dagger = D_1$ and $VA^\dagger AV^\dagger = D_2$. Then note that $\mathrm{Tr}D_1^\ell = \mathrm{Tr}D_2^\ell$ for all positive integers ℓ. Therefore, up to a permutation, D_1, D_2 are the same diagonal matrices, but a permutation is obtained by conjugation with a unitary matrix, so we may assume $D_1 = D_2$. Then it follows that $UAV^\dagger$ is normal, and hance can be unitarily diagonalized. Since we have separate phase degrees of freedom for U and V we can rotate away the phases on the diagonal to make the diagonal entries nonnegative. ♠.

Remarks:

1. By suitable choice of U and V the diagonal elements of Λ can be arranged in monotonic order. The set of these values are called the *singular values* of A.

2. This theorem is very useful when investigating moduli spaces of vacua in supersymmetric gauge theories.

♣Should give some examples... ♣

Exercise
Find a unitary bidiagonalization of the Jordan block $J_\lambda^{(2)}$.

17.2.2 Application: The Cabbibo-Kobayashi-Maskawa matrix, or, how bidiagonalization can win you the Nobel Prize

This subsubsection assumes some knowledge of relativistic field theory.

One example where bidiagonalization is important occurs in the theory of the "Kobayashi-Maskawa matrix" describing the mixing of different quarks. The $SU(3) \times SU(2) \times U(1)$ standard model has quark fields U_{Ri}, D_{Ri} neutral under $SU(2)$ and

$$\psi_{iL} = \begin{pmatrix} U_{Li} \\ D_{Li} \end{pmatrix} \tag{17.14}$$

forming a doublet under the gauge group $SU(2)$. Here i is a *flavor index* running over the families of quarks. In nature, at observable energies, i runs from 1 to 3. $SU(3)$ color and spin indices are suppressed in this discussion.

In the Standard Model there is also a doublet of scalars fields, the Higgs field:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \qquad \tilde{\phi} = \begin{pmatrix} (\phi^0)^* \\ \phi^- \end{pmatrix} \tag{17.15}$$

where $\phi^- = -(\phi^+)^*$ and $\phi^0, \phi^+ \in \mathbb{C}$. Both $\phi, \tilde{\phi}$ transform as $SU(2)$ doublets.

The Yukawa terms coupling the Higgs to the quarks lead to two "quark mass terms":

$$-g_{jk}\bar{D}_{jR}\phi^\dagger \psi_{kL} + hc \tag{17.16}$$

and

$$-\tilde{g}_{jk}\bar{U}_{jR}\tilde{\phi}^{\dagger}\psi_{kL} + hc \tag{17.17}$$

The matrices g_{jk} and $\tilde{g}_{jk}$ are assumed to be generic, and must be fixed by experiment.

For energetic reasons the scalar field ϕ_0 is nonzero (and constant) in nature. ("Develops a vacuum expectation value"). At low energies, $\phi^{\pm} = 0$ and hence these terms in the Lagrangian simplify to give mass terms to the quarks:

$$\sum_{i,j=1}^{3} \bar{U}_{Ri} M^U_{ij} U_{Lj} \tag{17.18}$$

and

$$\sum_{i,j=1}^{3} \bar{D}_{Ri} M^D_{ij} D_{Lj} \tag{17.19}$$

Here i, j run over the quark flavors, $U, \bar{U}, D, \bar{D}$ are quark wavefunctions. M^U, M^D are arbitrary complex matrices. We would like to go to a "mass eigenbasis" by bi-diagonalizing them with positive entries. The positive entries are identified with quark masses.

By bidiagonalization we know that

$$M^D_{ij} = (V_1^{\dagger} m^D V_2)_{ij} \tag{17.20}$$

$$M^U_{ij} = (V_3^{\dagger} m^U V_4)_{ij} \tag{17.21}$$

where m^D, m^U are diagonal matrices with real nonnegative eigenvalues and V_s, $s = 1, 2, 3, 4$ are four 3×3 unitary matrices. It is important that the V_s, are unitary because we want to use them to redefine our quark fields without changing the kinetic energy terms.

How much physical information is in the unitary matrices $V_1, \ldots, V_4$? We would like to rotate away the unitary matrices by a field redefinition of the quark fields. The rest of the Lagrangian looks like (again suppressing $SU(3)$ color and hence the gluon fields):

$$\begin{aligned} &\bar{U}_{jR}(\gamma \cdot \partial + \frac{2}{3}\gamma \cdot B)U_{jR} + \bar{D}_{jR}(\gamma \cdot \partial - \frac{1}{3}\gamma \cdot B)D_{jR} \\ &+ \bar{\psi}_{jL}(\gamma \cdot \partial + \frac{1}{6}\gamma \cdot B + \gamma \cdot A)\psi_{jL} \end{aligned} \tag{17.22}$$

where we just sum over the flavor index and the operator $\gamma \cdot \partial + q\gamma \cdot B$ with $q \in \mathbb{R}$ is diagonal for our purposes. Crucially, it is diagonal in the flavor space. The $SU(2)$ gauge field A_μ has off-diagonal components $W^{\pm}_\mu$:

$$A_\mu = \begin{pmatrix} A^3_\mu & W^+_\mu \\ W^-_\mu & -A^3_\mu \end{pmatrix} \tag{17.23}$$

leading, among other things, to the charged current interaction

$$W^+_\mu \bar{U}_{Li} \gamma^\mu D_{Li} \tag{17.24}$$

eq:cc-int

Clearly, we can rotate away V_1, V_3 by a field redefinition of D_{jR}, U_{jR} (taking the case of three flavors and giving the quarks their conventional names):

$$\begin{aligned} \left(\bar{d}\ \bar{s}\ \bar{b}\right)_R &= \bar{D}_R V_1^\dagger \\ \left(\bar{u}\ \bar{c}\ \bar{t}\right)_R &= \bar{U}_R V_3^\dagger \end{aligned} \tag{17.25}$$

However, we also need to rotate U_{jL} and D_{jL} to a mass eigenbasis by *different* matrices V_2 and V_4:

$$\begin{aligned} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L &= V_2 D_L \\ \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L &= V_4 U_L \end{aligned} \tag{17.26}$$

Therefore the charged current interaction (17.24) when expressed in terms of mass eigenstate fields is not diagonal in "flavor space." Rather, when we rotate to the mass basis the unitary matrix $S = V_4 V_2^\dagger$ enters in the charged current

$$\left(\bar{u}\ \bar{c}\ \bar{t}\right)_L \gamma^\mu S \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{17.27}$$

where u, c, t, d, s, b are mass eigenstate fields.

The unitary matrix S is called the Kobayashi-Maskawa matrix. It is still not physically meaningful because that by using further diagonal phase redefinitions that S only depends on 4 physical parameters, instead of the 9 parameters in an arbitrary unitary matrix.

Much effort in current research in experimental particle physics is devoted to measuring the matrix elements experimentally.

Reference: H. Georgi, *Weak Interactions and Modern Particle Theory.*

Exercise

Repeat the above discussion for N quark flavors. How many physically meaningful parameters are there in the weak interaction currents?

17.2.3 Singular value decomposition

The singular value decomposition applies to any matrix $A \in M_{m\times n}(\mathbb{C})$ and generalizes the bidiagonalization of square matrices. It has a wide variety of applications.

Theorem Suppose that $A \in M_{m\times n}(\mathbb{C})$, which we can take, WLOG so that $m \leq n$. Then there exist unitary matrices $U \in U(m)$ and $V \in U(n)$ so that

$$UAV = \begin{pmatrix} \Lambda_{m\times m} & 0_{m\times(n-m)} \end{pmatrix} \tag{17.28}$$ eq:sing-val-de

where Λ is a diagonal matrix with *nonnegative* entries (known as the *singular values* of A).

Proof: We have already proven the case with $m = n$ so assume that $m < n$. Enhance A to

$$\hat{A} = \begin{pmatrix} A \\ 0_{(n-m)\times n} \end{pmatrix} \tag{17.29}$$

Then bidiagonalization gives $\hat{U}, \hat{V} \in U(n)$ so that $\hat{U}\hat{A}\hat{V}$ is diagonal. Note that

$$\hat{U}\hat{A}\hat{A}^\dagger\hat{U} = \begin{pmatrix} D & 0_{m\times(n-m)} \\ 0_{(n-m)\times m} & 0_{(n-m)\times(n-m)} \end{pmatrix} \tag{17.30}$$

and hence if we break up $\hat{U}$ into blocks

$$\hat{U} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \tag{17.31}$$

then $U = U_{11}$ is in fact a unitary matrix in $U(m)$. But then

$$UA\hat{V} = \begin{pmatrix} \Lambda_{m\times m} & 0_{m\times(n-m)} \end{pmatrix} \tag{17.32}$$

and we can WLOG take Λ to have nonnegative entires ♠

Remarks

1. The eigenvalues of D are known as the *singular values* of A, the the decomposition (17.28) is known as the *singular value decomposition* of A.

2. The singular value decomposition can be rephrased as follows: If $T : V_1 \to V_2$ is a linear map between finite dimensional inner product spaces V_1 and V_2 then there exist ON sets $\{u_n\} \subset V_1$ and $\{w_n\} \subset V_2$ (not necessarily complete) and positive numbers λ_n so that
$$T = \sum_{n=1}^{N} \lambda_n (u_n, \cdot) w_n \tag{17.33}$$ eq:SVP-DEC-2

17.2.4 Schmidt decomposition

Let us consider a tensor product of two finite-dimensional inner product spaces $V_1 \otimes V_2$. We will assume, WLOG that $\dim V_1 \leq \dim V_2$. Then a vector $v \in V_1 \otimes V_2$ is called *separable* or *primitive* if it is of the form $v = v_1 \otimes v_2$ where $v_1 \in V_1$ and $v_2 \in V_2$. In quantum information theory vectors which are not separable are called *entangled*. The reader might want to keep in mind that V_1, V_2 are finite-dimensional Hilbert spaces.

The general vector in $V_1 \otimes V_2$ is a linear combination of separable vectors. Schmidt decomposition is a canonical form of this linear combination:

Theorem Given an arbitary vector $v \in V_1 \otimes V_2$ there exist ordered ON bases $\{u_i\}_{i=1}^{\dim V_1}$ for V_1 and $\{w_a\}_{a=1}^{\dim V_2}$ for V_2, so that

$$v = \sum_{i=1}^{\dim V_1} \lambda_i u_i \otimes w_i \tag{17.34}$$

with $\lambda_i \geq 0$.

Proof: Choose arbitrary ON bases $\{\tilde{u}_i\}$ for V_1 and $\{\tilde{w}_a\}$ for V_2. Then we can expand

$$v = \sum_{i=1}^{\dim V_1} \sum_{a=1}^{\dim V_2} A_{ia} \tilde{u}_i \otimes \tilde{w}_a \tag{17.35}$$

where A is a complex $m \times n$ matrix. Now from the singular value decomposition we can write $A = UDV$, where

$$D = \begin{pmatrix} \Lambda & 0 \end{pmatrix} \tag{17.36}$$

and Λ is an $m \times m$ diagonal matrix with nonnegative entries. Now use U, V to change basis. ♠

Remark: The rank of A is known as the *Schmidt number*. If it is larger than one and v is a quantum state in a bipartite system then the state is entangled.

18. Operators on Hilbert space

18.1 Lies my teacher told me

c:LieMyTeacher

18.1.1 Lie 1: The trace is cyclic:

ubsec:FirstLie

But wait! If

$$\mathrm{Tr}AB = \mathrm{Tr}BA \tag{18.1}$$

then

$$\mathrm{Tr}[A, B] = 0 \tag{18.2}$$

and if we consider $q\psi(x) = x\psi(x)$ and $p\psi(x) = -i\frac{d}{dx}\psi(x)$ on $L^2(\mathbb{R})$ then $[p, q] = -i1$ and $\mathrm{Tr}(1) = \infty$, not zero!

18.1.2 Lie 2: Hermitian operators have real eigenvalues

osec:SecondLie

But wait! Let us consider $A = q^n p + pq^n$ acting on $L^2(\mathbb{R})$ with $n > 1$. Since p, q are Hermitian this is surely Hermitian. But then

$$A\psi = \lambda\psi \tag{18.3}$$

is a simple first order differential equation whose general solution is easily found to be

$$\psi(x) = \kappa x^{-n/2} \exp[-\frac{i\lambda}{2(n-1)} x^{1-n}] \tag{18.4}$$

where κ is a constant. Then,

$$|\psi(x)|^2 = |\kappa|^2 |x|^{-n} \exp[\frac{\mathrm{Im}\lambda}{(n-1)} x^{1-n}] \tag{18.5}$$

This will be integrable for $x \to \pm\infty$ for $n > 1$ and it will be integrable at $x \to 0$ for $\mathrm{Im}\lambda < 0$ and n odd. Thus it would appear that the spectrum of $q^n p + pq^n$ is the entire lower half-plane!

18.1.3 Lie 3: Hermitian operators exponentiate to form one-parameter groups of unitary operators

ubsec:ThirdLie

But wait! Let us consider p on $L^2[0,1]$. Then $\exp[iap] = \exp[a\frac{d}{dx}]$ is the translation operator

$$(\exp[a\frac{d}{dx}]\psi)(x) = \psi(x+a) \tag{18.6}$$

But this can translate a wavefunction with support on the interval right off the interval! How can such an operator be unitary?!

18.2 Hellinger-Toeplitz theorem

sec:HT-Theorem

One theorem which points the way to the resolution of the above problems is the Hellinger-Toeplitz theorem. First, we begin with a definition:

Definition: A *symmetric everywhere-defined* linear operator, T, on $\mathcal{H}$ is an operator such that

$$(x, Ty) = (Tx, y) \qquad \forall x, y \in \mathcal{H} \tag{18.7}$$

We can also call this an everywhere defined self-adjoint operator. Then one might find the Hellinger-Toeplitz theorem a bit surprising:

Theorem: A symmetric everywhere-defined linear operator must be bounded.

In order to prove this one must use another theorem called the "closed graph theorem." This is one of three closely related theorems from the theory of operators between Banach spaces, the other two being the "bounded inverse theorem" and the "open mapping theorem."

Theorem[Bounded Inverse Theorem]: If $T : \mathcal{B}_1 \to \mathcal{B}_2$ is a bounded 1-1 operator from one Banach space onto (i.e. surjectively) another then $T^{-1} : \mathcal{B}_2 \to \mathcal{B}_1$ is bounded (hence continuous).

Proof: See Reed-Simon.

An immediate consequence of this is the

Theorem[Closed Graph Theorem]: If $T : \mathcal{B}_1 \to \mathcal{B}_2$ is a linear operator then the *graph of* T, defined by

$$\Gamma(T) := \{x \oplus Tx | x \in \mathcal{B}_1\} \subset \mathcal{B}_1 \oplus \mathcal{B}_2 \tag{18.8}$$

is a closed subspace of $\mathcal{B}_1 \oplus \mathcal{B}_2$ iff T is bounded.

Proof: This follows immediately from the bounded inverse theorem: Note that $\Gamma(T)$ is a normed linear space. Therefore, if it is closed it is itself a Banach space. Next, the map $\Gamma(T) \to \mathcal{B}_1$ given by $x \oplus Tx \to x$ is 1-1 bounded and continuous. Therefore the inverse $x \to x \oplus Tx$ is bounded which implies T is bounded. Conversely, if T is bounded then it is continuous so $\Gamma(T)$ is closed. ♠

Given the closed graph theorem the proof of the Hellinger-Toeplitz theorem is quite elegant. We need only prove that the graph of an everywhere-defined symmetric operator $T : \mathcal{H} \to \mathcal{H}$ is closed. Suppose then that $x_n \oplus Tx_n$ is a Cauchy sequence in $\Gamma(T) \subset \mathcal{H} \oplus \mathcal{H}$. Since $\Gamma(T) \subset \mathcal{H} \oplus \mathcal{H}$ and since $\mathcal{H} \oplus \mathcal{H}$ is a Hilbert space we know that

$$\lim_{n\to\infty} x_n \oplus Tx_n = x \oplus y \tag{18.9}$$

for some $x \oplus y \in \mathcal{H} \oplus \mathcal{H}$. Then if $y = Tx$ it will follow that $\Gamma(T)$ is closed. Now we note that for all $z \in \mathcal{H}$:

$$\begin{aligned}
(z, y) &= \lim_{n\to\infty} (z, Tx_n) && \text{def. of } y \\
&= \lim_{n\to\infty} (Tz, x_n) && T \quad \text{is symmetric} \\
&= (Tz, x) && \text{def. of } x \\
&= (z, Tx) && T \quad \text{is symmetric}
\end{aligned} \tag{18.10}$$

Therefore $y = Tx$, so $\Gamma(T)$ is closed, so T is bounded ♠

Now, many of the operators of interest in quantum mechanics are clearly unbounded, for example, the multiplication operator q on $L^2(\mathbb{R})$ satisfies

$$\| q\psi \|^2 = \int_{\mathbb{R}} x^2 |\psi(x)|^2 dx \tag{18.11}$$

Clearly there are wavefunctions with $\| \psi \| = 1$ but with support at arbitrarily large x, so q is unbounded. On the other hand it is equally obvious that q is symmetric. There is no contradiction with the HT theorem because of course it is not everywhere defined. Indeed, suppose $\psi(x)$ is a smooth square-integrable function decaying like $x^{-1-\epsilon}$ at large $|x|$ for some ϵ. For $0 < \epsilon \leq \frac{1}{2}$ the wavefunction $x\psi(x)$ is not square-integrable. Similar remarks apply to standard operators such as p and Schrödinger operators. These operators are only *partially defined*, that is, they are only defined on a linear subspace $W \subset \mathcal{H}$. We return to this theme in Section §18.6.

Exercise *Puzzle to resolve*
Consider $\mathcal{H} = \ell_2$ and define $T \in \mathcal{L}(\mathcal{H})$ by

$$T(e_n) = t_n e_n \tag{18.12}$$

where $\{t_n\}$ is a sequence of nonzero complex numbers with $\sum |t_n|^2 < \infty$.

a.) Show that T is an injective bounded operator.

b.) It would seem that this diagonal matrix has an obvious inverse

$$T^{-1}(e_n) = \frac{1}{t_n} e_n \tag{18.13}$$

On the other hand, such an operator is obviously unbounded! Why doesn't this contradict the Bounded Inverse Theorem?

---

ctrumResolvent

18.3 Spectrum and resolvent

Given a bounded operator $T \in \mathcal{L}(\mathcal{H})$ we partition the complex plane into two sets:

Definition:

1. The *resolvent set* or *regular set* of T is the subset $\rho(T) \subset \mathbb{C}$ of complex numbers so that $\lambda 1 - T$ is a bijective, i.e. a $1-1$ onto transformation $\mathcal{H} \to \mathcal{H}$.
2. The *spectrum of* T is the complement: $\sigma(T) := \mathbb{C} - \rho(T)$.

Now there are exactly three mutually exclusive ways the condition that $(\lambda 1 - T)$ is 1-1 can go wrong, and this leads to the decomposition of the spectrum:

Definition: The spectrum $\sigma(T)$ can be decomposed into three disjoint sets:

$$\sigma(T) = \sigma_{\text{point}}(T) \cup \sigma_{\text{res}}(T) \cup \sigma_{\text{cont}}(T) \tag{18.14}$$

1. If $\ker(\lambda 1 - T) \neq \{0\}$, that is, there is an eigenvector of T in $\mathcal{H}$ with eigenvalue λ, then λ is in the *point spectrum*.
2. If $\ker(\lambda 1 - T) = \{0\}$ but $\text{Im}(\lambda 1 - T)$ is not dense, then λ is in the *residual spectrum*.
3. If $\ker(\lambda 1 - T) = \{0\}$ and $\text{Im}(\lambda 1 - T)$ is dense but not all of $\mathcal{H}$, then λ is in the *continuous spectrum*.

Note that by the bounded inverse theorem, if $\lambda \in \rho(T)$ then the inverse, known as the *resolvent*

$$R_\lambda := (\lambda 1 - T)^{-1}, \tag{18.15}$$

is bounded. Now, for any bounded operator T we can try to expand

$$R_\lambda := \frac{1}{\lambda 1 - T} = \frac{1}{\lambda}\left(1 + \sum_{n=1}^{\infty} \left(\frac{T}{\lambda}\right)^n\right) \tag{18.16}$$

eq:Res-Exp

This converges in the norm topology for $|\lambda| > \| T \|$. One can check that then the series is an inverse to $\lambda 1 - T$ and is bounded. It follows that $\sigma(T)$ is inside the closed unit disk of radius $\| T \|$.

Similarly the condition that λ be in $\rho(T)$ is an *open condition*: If $\lambda \in \rho(T)$ then so is every complex number in some neighborhood of λ. Therefore, $\sigma(T)$ is a closed subset of $\mathbb{C}$. More formally, we can prove:

Theorem: The resolvent set $\rho(T)$ is open (hence $\sigma(T)$ is closed) and in fact

$$\| R_\lambda \| \leq \frac{1}{\mathrm{dist}(\lambda, \sigma(T))} \tag{18.17}$$ eq:NormDist

Proof: Suppose $\lambda_0 \in \rho(T)$. Consider the formal expansion

$$\begin{aligned} R_\lambda(T) &= \frac{1}{\lambda - T} = \frac{1}{\lambda - \lambda_0 + \lambda_0 - T} \\ &= \frac{1}{\lambda_0 - T} \frac{1}{1 - \frac{\lambda_0 - \lambda}{\lambda_0 - T}} \\ &= R_{\lambda_0}(T) \left[1 + \sum_{n=1}^{\infty} (\lambda_0 - \lambda)^n (R_{\lambda_0}(T))^n \right] \end{aligned} \tag{18.18}$$

Therefore, if $\| R_{\lambda_0}(T) \| \, |\lambda - \lambda_0| < 1$ then the series converges in the norm topology. Once we know it converges the formal properties will in fact be true properties and hence the series represents $R_\lambda(T)$, which will be bounded. Hence $\lambda \in \rho(T)$ for those values ♠

Remarks:

1. The proof shows that the map $\lambda \to R_\lambda(T)$ from $\rho(T)$ to $\mathcal{L}(\mathcal{H})$ is holomorphic.

Definition For any everywhere defined bounded operator T we define the *adjoint of* T, denoted $T^\dagger$ exactly as in the finite-dimensional case:

$$(Tx, y) = (x, T^\dagger y) \qquad \forall x, y \in \mathcal{H} \tag{18.19}$$

Remark If $\lambda \in \sigma_{\mathrm{res}}(T)$ then $\lambda^* \in \sigma_{\mathrm{point}}(T^\dagger)$. To see this, suppose that $\mathrm{im}(\lambda 1 - T)$ is not dense. Then there is some nonzero vector y not in the closed subspace $\overline{\mathrm{im}(\lambda 1 - T)}$ and by the projection theorem we can take it to be orthogonal to $\overline{\mathrm{im}(\lambda 1 - T)}$. That means

$$(y, (\lambda 1 - T)x) = ((\lambda^* 1 - T^\dagger)y, x) = 0 \tag{18.20}$$

for all x, which means that y is an eigenvector of $T^\dagger$ with eigenvalue λ^*. Therefore

$$\sigma_{\mathrm{res}}(T)^* \subset \sigma_{\mathrm{point}}(T^\dagger) \tag{18.21}$$

Example Let us return to the shift operator, or Hilbert hotel operator $S \in \mathcal{L}(\mathcal{H})$ for $\mathcal{H} = \ell^2$:

♣Notation clash: x_i are complex numbers, not a sequence of vectors in Hilbert space. ♣

$$S : (x_1, x_2, \dots) \mapsto (0, x_1, x_2, \dots) \tag{18.22}$$

In terms of harmonic oscillators $S = \frac{1}{\sqrt{a^\dagger a}} a^\dagger$. This is bounded and everywhere defined and one easily computes the adjoint, which just shifts to the left:

$$S^\dagger : (x_1, x_2, \dots) \mapsto (x_2, x_3 \dots) \tag{18.23}$$

Or, if one prefers, $S^\dagger = \frac{1}{\sqrt{a^\dagger a + 1}} a$.

Applying the remark of (18.16) to our case, where one easily checks $\| S \| = \| S^\dagger \| = 1$, we conclude that both $\sigma(S)$ and $\sigma(S^\dagger)$ are contained in the closed unit disk D.

Now, one easily shows that if $|\lambda| < 1$ then

$$\Psi_\lambda := (1, \lambda, \lambda^2, \dots) \tag{18.24}$$

is in ℓ^2 and

$$S^\dagger \Psi_\lambda = \lambda \Psi_\lambda \tag{18.25}$$

Since the spectrum must be closed it must be that $\sigma(S^\dagger)$ is the entire closed unit disk D.

On the other hand, let us consider the solutions to

$$y = (\lambda 1 - S)x \tag{18.26}$$

In terms of the components this leads to the equations

$$\begin{aligned} y_1 &= \lambda x_1 \\ y_2 &= \lambda x_2 - x_1 \\ &\vdots \quad \vdots \\ y_n &= \lambda x_n - x_{n-1} \\ &\vdots \quad \vdots \end{aligned} \tag{18.27}$$

which is easily solved, at least formally, to give

$$x_n = \lambda^{-1} y_n + \lambda^{-2} y_{n-1} + \cdots + \lambda^{-n} y_1 \tag{18.28}$$

It immediately follows that if $y = 0$ then $x = 0$, that is, the kernel of $(\lambda 1 - S)$ is $\{0\}$ and hence there is no point spectrum for S. Moreover, if $|y_1| = 1$ while $y_2 = y_3 = \cdots = 0$ then $x_n = \lambda^{-n} y_1$ is clearly not normalizable for $|\lambda| \leq 1$. Now let ξ_n be a sequence of positive numbers monotonically decreasing to zero so that $\sum \xi_n^2 |\lambda|^{-2n}$ does not converge. Then using the triangle inequality one easily checks that

$$|x_n| > \xi_n |\lambda^{-n} y_1| \tag{18.29}$$

and hence $\{x_n\}$ cannot be normalizable for $|y_n| < (\xi_{n-1} - \xi_n)|\lambda^{1-n} y_1|$. Therefore $\mathrm{im}(\lambda 1 - S)$ does not intersect the open ball defined by $|y_n| < (\xi_{n-1} - \xi_n)|\lambda^{1-n} y_1|$. and hence is not

dense for $|\lambda| \leq 1$. Thus $\sigma(S)$ is the closed unit disk, the point spectrum is zero and the residual spectrum is D.

Finally, we need to consider the nature of the spectrum of $S^\dagger$ when $|\lambda| = 1$. Then Ψ_λ is not in the Hilbert space so there is no point spectrum. On the other hand, if $\mathrm{im}(\lambda 1 - S^\dagger)$ were not dense then there would be a y not in its closure and by the projection theorem we can take it to be orthogonal to $\mathrm{im}(\lambda 1 - S^\dagger)$. But this means $(y, (\lambda - S^\dagger)x) = 0$ for all x which means $(\lambda^* 1 - S)y = 0$, but we know that S has no eigenvectors. Thus $|\lambda| = 1$ is in the spectrum of $S^\dagger$ but is *neither* in the point nor the residual spectrum! We conclude that $\mathrm{im}(\lambda 1 - S^\dagger)$ is dense, but not equal to $\mathcal{H}$. To exhibit a vector outside the image we can try to solve $y = (\lambda 1 - S^\dagger)x$. The formal solution is

$$\begin{aligned} x_1 &= \lambda^{-1} y_1 + \lambda^{-2} y_2 + \lambda^{-3} y_3 + \cdots \\ x_2 &= \lambda^{-1} y_2 + \lambda^{-2} y_3 + \lambda^{-3} y_4 + \cdots \\ x_3 &= \lambda^{-1} y_3 + \lambda^{-2} y_4 + \lambda^{-3} y_5 + \cdots \\ &\vdots \quad \vdots \end{aligned} \tag{18.30}$$

So, if we take $y_n = \lambda^n / n$ then $y \in \ell^2$ but is not in the image.

In summary we have the table:

Operator	Spectrum	Point Spectrum	Residual Spectrum	Continuous Spectrum
S	D	$\emptyset$	D	$\emptyset$
$S^\dagger$	D	Interior(D)	$\emptyset$	$\lvert\lambda\rvert = 1$

Theorem Suppose $T : \mathcal{H} \to \mathcal{H}$ is an everywhere defined self-adjoint operator. Then

1. The spectrum $\sigma(T) \subset \mathbb{R}$ is a subset of the real numbers.
2. The residual spectrum $\sigma_{\mathrm{res}}(T) = \emptyset$.

Proof: The usual proof from elementary quantum mechanics courses shows that the point spectrum is real: If $Tx = \lambda x$ then

$$\lambda(x, x) = (x, Tx) = (Tx, x) = \lambda^*(x, x) \tag{18.31}$$

so $\lambda \in \mathbb{R}$.

Now we show the residual spectrum is empty. We remarked above that $\sigma_{\rm res}(T)^* \subset \sigma_{\rm point}(T^\dagger)$. But if $T^\dagger = T$ then if $\lambda \in \sigma_{\rm res}(T)$ we must have $\lambda^* \in \sigma_{\rm point}(T)$ and hence λ is real, but this is impossible since the point and residual spectra are disjoint.

Now, let λ and μ be any real numbers and compute

$$\| (T-(\lambda+i\mu))x \|^2 = (x,(T-\lambda+i\mu)(T-\lambda-i\mu)x) = (x,((T-\lambda)^2+\mu^2)x) = \| (T-\lambda)x \|^2 + \mu^2 \| x \|^2 \tag{18.32}$$

eq:smeng

This shows that if $\mu \neq 0$ then $T-(\lambda+i\mu)$ is invertible. If we let $x = (T-(\lambda+i\mu))^{-1}y$ then (18.32) implies that $\mu^2 \| x \|^2 \leq \| y \|^2$. But this means that

$$\frac{\| (T-(\lambda+i\mu))^{-1}y \|}{\| y \|} = \frac{\| x \|}{\| y \|} \leq \frac{1}{|\mu|} \tag{18.33}$$

and hence $(T-(\lambda+i\mu)$ has a bounded inverse for $\mu \neq 0$. Therefore, $\lambda + i\mu \in \rho(T)$ for $\mu \neq 0$ and hence $\sigma(T) \subset \mathbb{R}$. ♠

A useful criterion for telling when $\lambda \in \sigma(T)$ is the following:

Definition: A *Weyl sequence* is [31] a sequence of vectors $z_n \in D(T)$ such that $\| z_n \| = 1$ and $\| (\lambda - T)z_n \| \to 0$.

Theorem: [Weyl criterion]

a.) If T has a Weyl sequence then $\lambda \in \sigma(T)$.

b.) If λ is on the boundary of $\rho(T)$ then T has a Weyl sequence.

Proof:

a.) If there is a Weyl sequence and $\lambda \in \rho(T)$ then

$$1 = \| z_n \| = \| R_\lambda(T)(\lambda - T)z_n \| \leq \| R_\lambda(T) \| \| (\lambda - T)z_n \| \to 0 \tag{18.34}$$

which is impossible. Therefore $\lambda \in \sigma(T)$.

b.) Suppose $\lambda \in \overline{\rho(T)} - \rho(T)$, then there is a sequence of complex numbers $\{\lambda_n\}$ with $\lambda_n \in \rho(T)$ and $\lambda_n \to \lambda$ such that $\mathrm{dist}(\lambda_n, \sigma(T)) \to 0$. Therefore, by (18.17) we know $\| R_{\lambda_n}(T) \| \nearrow +\infty$ and therefore there are vectors y_n so that

$$\frac{\| R_{\lambda_n}(T)y_n \|}{\| y_n \|} \nearrow +\infty \tag{18.35}$$

Now set $z_n = R_{\lambda_n}(T)y_n$. These will be nonzero (since $R_{\lambda_n}(T)$ is invertible) and hence we can normalize y_n so that $\| z_n \| = 1$. But then

$$\begin{aligned} \| (\lambda - T)z_n \| &= \| (\lambda - \lambda_n)z_n + (\lambda_n - T)z_n \| \\ &= \| (\lambda - \lambda_n)z_n + y_n \| \\ &\leq |\lambda - \lambda_n| + \| y_n \| \to 0 \end{aligned} \tag{18.36}$$

[31] We state it so that it applies to unbounded operators with dense domain $D(T)$. See below. For bounded operators take $D(T) = \mathcal{H}$.

so $\{z_n\}$ is a Weyl sequence. ♠

Example Consider the operator on $L^2[a,b]$ given by the position operator $q\psi(x) = x\psi(x)$. Clearly this is a bounded operator $\| q \| \leq b$. It is everywhere defined and symmetric, hence, it is self-adjoint. It does not have eigenvectors [32]. For any $x_0 \in (a,b)$ we can take good approximations to the Dirac delta function:

$$\psi_{\epsilon,x_0} = \left(\frac{2}{\pi}\right)^{1/4} \frac{1}{\epsilon^{1/2}} e^{-(x-x_0)^2/\epsilon^2} \tag{18.37}$$

and, on the real line $\| (q-x_0)^2\psi_{\epsilon,x_0} \|^2 = \epsilon^2/\sqrt{2}$, so $(q-x_0)^{-1}$ could hardly be a bounded operator. Thus $\sigma(q) = [a,b]$ and the spectrum is entirely continuous spectrum.

Exercise *The C^* identity*
Show that

$$\| T^\dagger T \| = \| T \|^2 \tag{18.38}$$

Remark: In general, a Banach algebra [33] which has an anti-linear involution so that $(ab)^* = b^*a^*$ and which satisfies $\| a^*a \| = \| a \|^2$ is known as a *C^*-algebra*. There is a rather large literature on the subject. It can be shown that every C^* algebra is a †-closed subalgebra of the algebra of bounded operators on Hilbert space.

Exercise
Show that $S^\dagger S = 1$ but $SS^\dagger$ is not one, but rather is a projection operator. That means that S is an example of a *partial isometry*.

♣Say more? ♣

♣Give the example of the Harper operator $U + U^* + \lambda(V + V^*)$ whose spectrum is a Cantor set. ♣

18.4 Spectral theorem for bounded self-adjoint operators

Now we would like to explain the statement (but not the proof) of the spectral theorem for self-adjoint operators on Hilbert space - a major theorem of von Neumann.

We begin with an everywhere-defined self-adjoint operator $T \in \mathcal{L}(\mathcal{H})$. As we have seen, T is bounded and $\sigma(T) \subset \mathbb{R}$ is a disjoint union of the point and continuous spectrum.

The spectral theorem says - roughly - that in an appropriate basis T is just a multiplication operator, like $\psi(x) \to x\psi(x)$. Roughly, for each $\lambda \in \sigma(T)$ we choose eigenvectors

[32] The "position eigenstates" of elementary quantum mechanics are distributions, and are not vectors in the Hilbert space.

[33] A *Banach algebra* is an algebra which is also a complete normed vector space and which satisfies $\| ab \| \leq \| a \| \cdot \| b \|$, an inequality which is easily verified for $\mathcal{L}(\mathcal{H})$. So $\mathcal{L}(\mathcal{H})$ is an example of a Banach algebra.

$|\psi_{\lambda,i}\rangle$ where i indicates possible degeneracy of the eigenvalue and then we aim to write something like

$$T \sim \int_{\sigma(T)} \lambda \left(\sum_i |\psi_{\lambda,i}\rangle\langle\psi_{\lambda,i}| \right) d\mu_T(\lambda) \tag{18.39}$$

with some measure $\mu_T(\lambda)$ on the spectrum. Clearly, unless we have a discrete point spectrum with finite dimensional eigenspaces this representation is at best heuristic. In that latter case

$$d\mu_T(\lambda) = \sum_n \delta(\lambda - \lambda_n) d\lambda \tag{18.40}$$

where we sum over the distinct eigenvalues λ_n.

In order to give a precise and general formulation of the spectral theorem von Neumann introduced the notion of a *projection-valued measure*, which we will now define. First we need:

Definition: The (Borel) measurable subsets of the real line $\mathbb{R}$ is the smallest collection $\mathcal{B}(\mathbb{R})$ of subsets of the real line such that

1. All intervals $(a, b) \in \mathcal{B}(\mathbb{R})$.
2. $\mathcal{B}(\mathbb{R})$ is closed under complement: If $E \in \mathcal{B}(\mathbb{R})$ then $\mathbb{R} - E \in \mathcal{B}(\mathbb{R})$.
3. $\mathcal{B}(\mathbb{R})$ is closed under countable union.

Remarks:

1. These axioms imply that $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ and $\emptyset \in \mathcal{B}(\mathbb{R})$.
2. The good thing about this collection of subsets of $\mathbb{R}$ is that one can define a "good" notion of "size" or measure $\mu(E)$ of an element $E \in \mathcal{B}(\mathbb{R})$ such that $\mu((a, b)) = b - a$ and μ is additive on disjoint unions. It turns out that trying to define such a measure μ for arbitrary subsets of $\mathbb{R}$ leads to paradoxes and pathologies.
3. We say a property holds "almost everywhere" if the set where it fails to hold is of measure zero.

Definition: A *projection-valued measure* is a map

$$P : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(\mathcal{H}) \tag{18.41}$$

such that

1. $P(E)$ is an orthogonal projection operator for all $E \in \mathcal{B}(\mathbb{R})$.
2. $P(\emptyset) = 0$ and $P(\mathbb{R}) = 1$.

3. If $E = \amalg_{i=1}^{\infty}$ is a countable *disjoint* union of sets $E_i \in \mathcal{B}(\mathbb{R})$ then

$$P(E) = \text{s} - \lim_{n\to\infty} \sum_{i=1}^{n} P(E_i) \tag{18.42}$$

where the convergence is in the strong topology.

Remarks

1. The meaning of convergence in the strong topology is that a sequence of operators $T_n \to T$ if, for all $x \in \mathcal{H}$, $\| T_n x - Tx \| \to 0$.

2. Given a PVM and a nonzero vector $x \in \mathcal{H}$ there is a corresponding ordinary measure P_x on $\mathbb{R}$. We define it on a measurable set E by

$$P_x(E) = \frac{(x, P(E)x)}{(x,x)} \tag{18.43}$$

♣No need to divide by (x,x). You can just use a positive measure not normalized to one. ♣

This is a measure because, as is easily verified: $P_x(\emptyset) = 0$, $P_x(\mathbb{R}) = 1$, $P_x(E) \geq 0$, and, if $E = \amalg_{i=1}^{\infty}$ is a countable *disjoint* union of sets $E_i \in \mathcal{B}(\mathbb{R})$ then

$$P_x(E) = \sum_{i=1}^{\infty} P_x(E_i) \tag{18.44}$$

3. It will be convenient below to use the notation

$$P_x(\lambda) := P_x((-\infty, \lambda]) \tag{18.45}$$

eq:Plambda

This will be a measureable function and the corresponding measure $dP_x(\lambda)$ has the property that

$$P_x(E) = \int_E dP_x(\lambda) \tag{18.46}$$

Theorem[Spectral Theorem for bounded operators] If T is an everywhere-defined self-adjoint operator on $\mathcal{H}$ then

1. There is a PVM P_T so that for all $x \in \mathcal{H}$, we have

$$\frac{(x, Tx)}{(x,x)} = \int_{\mathbb{R}} \lambda dP_{T,x}(\lambda) \tag{18.47}$$

eq:DiagMtx

where $P_{T,x}(\lambda)$ is the measurable function associated to P_T via (18.45).

2. If f is a (measurable) function on $\mathbb{R}$ then $f(T)$ makes sense and for all $x \in \mathcal{H}$, we have

$$\frac{(x, f(T)x)}{(x,x)} = \int_{\mathbb{R}} f(\lambda) dP_{T,x}(\lambda) \tag{18.48}$$

eq:DiagMtx-f

Rough idea of the proof: The basic idea of the proof is to look at the algebra of operators generated by T. This is a commutative algebra. For example, it contains all the polynomials

in T. If we take some kind of closure then it will contain all *continuous functions of* T. This statement is known as the "continuous functional calculus." Now, this continuous algebra is identified with the continuous functions on the compact set $X = \sigma(T)$. Moreover, given any $x \in \mathcal{H}$ we have a linear map $\Lambda_T : C(X) \to \mathbb{R}$ given by

$$f \mapsto \Lambda_T(f) := (x, f(T)x) \tag{18.49}$$

Moreover, the map is positive, meaning that if f is positive then $\Lambda_T(f) \geq 0$. Then a general theorem - known as the Riesz-Markov theorem - says that any positive linear functional $C(X) \to \mathbb{R}$ where X is Hausdorff is of the form $f \mapsto \int_X f d\mu$. Therefore, given T and x there is a corresponding measure $\mu_{T,x}$ and we have

$$(x, f(T)x) = \int_{\sigma(T)} f d\mu_{T,x} \tag{18.50}$$

Now, using this equation one extends the continuous functional calculus to the Borel functional calculus - namely, now we make sense of operators $g(T)$ where g is not necessarily continuous, but at least measurable. In particular, if g is a characteristic function it is discontinuous, but $g(T)$ will be a projection operator. ♠

Remarks:

1. Note that (18.47) is enough to determine all the matrix elements of T. This equation determines (x, Tx) for all x and then we can use the polarization identity:

$$\begin{aligned}(x, Ty) = \frac{1}{4}\Big[&((x+y), T(x+y)) - ((x-y), T(x-y)) \\ &+ i((x-iy), T(x-iy)) - i((x+iy), T(x+iy))\Big]\end{aligned} \tag{18.51}$$

   which can also be written

$$4(y, Tx) = \sum_{k=0}^{3} i^k (x + i^k y, T(x + i^k y)) \tag{18.52}$$

   Note that on a real Hilbert space we cannot multiply by i, but then $(y, Tx) = (Tx, y) = (x, Ty)$ for self-adjoint T so that it suffices to work just with $x \pm y$ in the corresponding polarization identity.

2. Equation (18.48) is meant to capture the idea that in the block-diagonalized basis provided by P_T the operator T is diagonalized.

3. It follows from the definition of a PVM that if $\| x \| = 1$

$$(x, P(E)x) = \int_{\mathbb{R}} \chi_E(\lambda) dP_x(\lambda) = \int_E dP_x(\lambda) \tag{18.53}$$

   where $\chi_E(\lambda)$ is the characteristic function of the set E.

4. Using the previous remark Using this we can see that, as expected, for a self-adjoint operator T the PVM P_T has support on the spectrum $\sigma(T)$ in the sense that:

   $\lambda \in \sigma(T)$ *iff* $P_T((\lambda-\epsilon,\lambda+\epsilon)) \neq 0$ *for all* $\epsilon > 0$.

   To prove this suppose first that $P_T((\lambda_0-\epsilon,\lambda_0+\epsilon)) \neq 0$ for all $\epsilon > 0$. Then take the sets $E_n = (\lambda_0 - \frac{1}{n}, \lambda_0 + \frac{1}{n})$. Since $P_T(E_n)$ is nonzero we can take a sequence of nonzero vectors z_n in the image of $P_T(E_n)$ and normalize them to $\| z_n \| = 1$. Then

$$\begin{aligned} \| (T-\lambda_0)z_n \|^2 &= \| (T-\lambda_0)P_T(E_n)z_n \|^2 \\ &= \int_{\mathbb{R}} (\lambda-\lambda_0)^2 \chi_{E_n}(\lambda) dP_{T,z_n}(\lambda) \leq \frac{1}{n^2} \end{aligned} \tag{18.54}$$

   so we have a Weyl sequence and hence $\lambda \in \sigma(T)$. Conversely, suppose that $P_T((\lambda_0 - \epsilon, \lambda_0+\epsilon)) = 0$ for some $\epsilon > 0$. For such an ϵ define the function:

$$f_\epsilon(\lambda) := \begin{cases} 0 & |\lambda-\lambda_0| < \epsilon \\ \frac{1}{\lambda_0-\lambda} & |\lambda-\lambda_0| \geq \epsilon \end{cases} \tag{18.55}$$

   Then

$$\begin{aligned} (\lambda_0 - T)f_\epsilon(T) &= (\lambda_0 - T)\int_{|\lambda_0-\lambda|\geq\epsilon} \frac{1}{\lambda_0-\lambda} dP_T(\lambda) \\ &= \int_{|\lambda_0-\lambda|\geq\epsilon} \frac{\lambda_0-\lambda}{\lambda_0-\lambda} dP_T(\lambda) \\ &= \int_{|\lambda_0-\lambda|\geq\epsilon} dP_T(\lambda) = 1 \end{aligned} \tag{18.56}$$

   Similarly, $f(T)(\lambda_0 - T) = 1_{D(T)}$. Therefore, $\lambda_0 - T$ is a bijection of $D(T)$ with $\mathcal{H}$ and hence $\lambda \in \rho(T)$. ♠

Example: Suppose T has a finite pure point spectrum $\{\lambda_n\}_{n=1}^N$ with eigenspaces V_{λ_n}. Then define

$$\delta_\lambda(E) = \int_E \delta(\lambda'-\lambda)d\lambda' = \begin{cases} 1 & \lambda \in E \\ 0 & \lambda \notin E \end{cases} \tag{18.57}$$

Then the projection value measure of T is

$$P_T(E) = \sum_{\lambda_n \in E} P_{V_{\lambda_n}} = \sum_{n=1}^N \delta_{\lambda_n}(E) P_{V_{\lambda_n}} \tag{18.58}$$

In particular, if $T = Id_{\mathcal{H}}$ is the unit operator then

$$P_T(E) = \begin{cases} Id_{\mathcal{H}} & 1 \in E \\ 0 & 1 \notin E \end{cases} \tag{18.59}$$

Exercise
Show that if P is a PVM then [34]

$$P(E_1)P(E_2) = P(E_1 \cap E_2) \tag{18.60}$$

18.5 Defining the adjoint of an unbounded operator

Recall from the Hellinger-Toeplitz theorem that an unbounded operator on an infinite-dimensional Hilbert space cannot be everywhere defined and self-adjoint. On the other hand, as we explained, physics requires us to work with unbounded self-adjoint operators.

Therefore, we should consider *partially defined* linear operators. That is, linear operators T from a proper subspace $D(T) \subset \mathcal{H}$ to $\mathcal{H}$. Giving the domain $D(T)$ of the operator is an essential piece of data in defining the operator.

Definition:
a.) The *graph of* T is the subset

$$\Gamma(T) := \{x \oplus Tx | x \in D(T)\} \subset \mathcal{H} \oplus \mathcal{H} \tag{18.61}$$

b.) T is *closed* if $\Gamma(T)$ is a closed subspace of $\mathcal{H} \oplus \mathcal{H}$.

c.) An operator T_2 is an *extension of an operator* T_1 if $\Gamma(T_1) \subset \Gamma(T_2)$. That is, $D(T_1) \subset D(T_2)$ and when T_2 is restricted to $D(T_1)$ it agrees with T_1. This is usually denoted $T_1 \subset T_2$.

Definition: If $D(T) \subset \mathcal{H}$ is dense then we define the subset $D(T^\dagger) \subset \mathcal{H}$ to be the set of $y \in \mathcal{H}$ so that there exists a $z \in \mathcal{H}$ such that for all $x \in D(T)$,

$$(Tx, y) = (x, z). \tag{18.62}$$

If $y \in D(T^\dagger)$ then z is unique (since $D(T)$ is dense) and we denote

$$z = T^\dagger y \tag{18.63}$$

This defines a linear operator $T^\dagger$ with domain $D(T^\dagger)$ called the *adjoint of* T.

Remark: One way of characterizing $D(T^\dagger)$ is that $y \in D(T^\dagger)$ iff $x \mapsto (y, Tx)$ extends to a to a *bounded* linear operator. Then z exists, by the Riesz representation theorem.

Definition: A densely defined operator T is

[34] *Answer*: First show that if $E_1 \cap E_2 = \emptyset$ then $P(E_1)P(E_2) = P(E_2)P(E_1) = 0$. Do that by using the PVM axiom to see that $P(E_1 \cup E_2) = P(E_1) + P(E_2)$. Square this equation to conclude that $\{P(E_1), P(E_2)\} = 0$. But now, multiply this equation on the left and then on the right by $P(E_1)$ to show that $[P(E_1), P(E_2)] = 0$. Next, write $P(E_1) = P(E_1 - E_1 \cap E_2) + P(E_1 \cap E_2)$ and $P(E_2) = P(E_2 - E_1 \cap E_2) + P(E_1 \cap E_2)$ and multiply.

a.) *Symmetric* if $T \subset T^\dagger$.
b.) *Self-adjoint* if $T = T^\dagger$.

Remarks

1. Let us unpack this definition a bit. An operator T is symmetric iff

$$(x, Ty) = (Tx, y) \tag{18.64}$$

for all $x, y \in D(T)$. However, $x \to (Tx, y)$ might be bounded for a larger class of vectors $y \notin D(T)$. When T is self-adjoint this does not happen and $D(T^\dagger) = D(T)$.

2. Unfortunately, different authors use the term "Hermitian operator" in ways which are inequivalent for unbounded operators. Some authors (such as Reed and Simon) use the term to refer to symmetric operators while other authors (such as Takhtadjan) use the term to refer to self-adjoint operators. So we will use only "symmetric" and "self-adjoint" and reserve the term "Hermitian" for the finite-dimensional case, where no confusion can arise.

Example: Let us use the "momentum" $p = -i\frac{d}{dx}$ to define an operator with a dense domain $D(T)$ within $L^2[0,1]$. The derivative is clearly not defined on all L^2 functions. Now, a function $f : [0,1] \to \mathbb{C}$ is "absolutely continuous" if $f'(x)$ exists almost everywhere and $|f'(x)|$ is integrable. In particular, the fundamental theorem of calculus holds

$$f(b) - f(a) = \int_a^b f'(x)dx \tag{18.65}$$

In this case $f'(x)$ is indeed in $L^2[0,1]$.

We begin with an operator T defined by the domain:

$$D(T) = \{f | f \in AC[0,1], f(0) = f(1) = 0\} \tag{18.66}$$

then elementary integration by parts shows that $T = -i\frac{d}{dx}$ is a symmetric operator. An elaborate argument in Reed-Simon VIII.2, p. 257 shows that with this definition of T we have:

$$D(T^\dagger) = \{f | f \in AC[0,1]\} \tag{18.67}$$

and in Reed-Simon vol. 2, Section X.1, p. 141 it is shown that there is a one-parameter family of self-adjoint extensions $T_\alpha = T_\alpha^\dagger$ labeled by a phase α:

$$D(T_\alpha) = \{f | f \in AC[0,1], f(0) = \alpha f(1)\} \tag{18.68}$$

It easy to appreciate this even without all the heavy machinery of defining self-adjoint extensions of symmetric operators. Formally proving that the operator is symmetric requires that the boundary terms in the integration by parts vanishes, that is:

$$(T\psi_1, \psi_2) = (\psi_1, T\psi_2) \tag{18.69}$$

implies

$$\psi_1^*(1)\psi_2(1) - \psi_1^*(0)\psi_2(0) = 0 \tag{18.70}$$

If $\psi_2 \in D(T)$ then this will be satisfied for $\psi_1 \in D(T^\dagger)$ because both terms separately vanish. We can attempt to extend this definition to a larger domain. If we try to let both $\psi_1, \psi_2 \in D(T^\dagger)$ the condition will fail. The intermediate choice is to choose a phase α and require $\psi(1) = \alpha\psi(0)$.

Exercise
Suppose $T_1 \subset T_2$ are densely defined operators. Show that $T_2^\dagger \subset T_1^\dagger$.

---

ec:UnboundedSA

18.6 Spectral Theorem for unbounded self-adjoint operators

Having introduced the notion of projection valued measures we are now in a position to state the spectral theorem for (possibly unbounded) self-adjoint operators on Hilbert space:

The Spectral Theorem: There is a 1-1 correspondence between self-adjoint operators T on a Hilbert space $\mathcal{H}$ and projection valued measures such that:

a.) Given a PVM P a corresponding self-adjoint operator T_P can be defined by the diagonal matrix elements:

$$\frac{(x, T_P x)}{(x,x)} = \int_{\mathbb{R}} \lambda dP_x(\lambda) \tag{18.71}$$

eq:Gen-ST-i

with domain

$$D(T_P) = \{x \in \mathcal{H} | \int_{\mathbb{R}} \lambda^2 dP_x(\lambda) < \infty\} \tag{18.72}$$

eq:Gen-ST-ii

b.) Conversely, given T there is a corresponding PVM P_T such that (18.71) and (18.72) hold.

c.) Moreover, given a self-adjoint operator T if f is any (Borel measurable) function then there is an operator $f(T)$ with domain

$$D(f(T)) = \{x \in \mathcal{H} | \int_{\mathbb{R}} |f(\lambda)|^2 dP_{T,x}(\lambda) < \infty\} \tag{18.73}$$

eq:Gen-ST-iii

such that

$$\frac{(x, f(T)x)}{(x,x)} = \int_{\mathbb{R}} f(\lambda) dP_{T,x}(\lambda) \tag{18.74}$$

eq:Gen-ST-iv

Example 1. If T has pure point spectrum $\{\lambda_n\}$ with closed eigenspaces V_n then

$$T = \sum_n \lambda_n P_{V_n} \tag{18.75}$$

$$P_T(E) = \sum_{\lambda_n \in E} P_{V_n} \tag{18.76}$$

where P_{V_n} are the orthogonal projections to the subspaces V_n.

Example 2. Take $\mathcal{H} = L^2(\mathbb{R})$ and define $T = q$ by $q\psi(x) = x\psi(x)$ with a domain given by those wavefunctions with falloff at least as fast as $|x|^{-\alpha}$ at infinity, with $\alpha > 3/2$. Then

$$(P_T(E)\psi)(x) = \begin{cases} \psi(x) & x \in E \\ 0 & x \notin E \end{cases} \tag{18.77}$$

Example 3. Take $\mathcal{H} = L^2(\mathbb{R})$ and define $T = p$ by $p\psi(x) = -i\frac{d}{dx}\psi(x)$ with a domain given by absolutely continuous wavefunctions with L^2 integrable derivative. [35] Now, if we make a Fourier transform then this is again a multiplication operator, so now

$$(P_T(E)\psi)(x) = \int_{\mathbb{R}} dy \int_E \frac{dk}{2\pi} e^{ik(x-y)}\psi(y) \tag{18.78}$$

Remarks

1. This is a major theorem and proofs can be found in many places. We mention just Chapters VII and VIII of Reed-Simon, and Chapter 3 of G. Teschl, *Mathematical Methods in Quantum Mechanics*.

2. The resolvent set and spectrum of a *closed* but possibly unbounded operator are defined along the same lines as in the bounded case: $\lambda \in \rho(T)$ if $\lambda - T$ is a bijection of $D(T)$ onto $\mathcal{H}$. It follows that the resolvent $R_\lambda = (\lambda - T)^{-1}$ is a bounded operator $\mathcal{H} \to D(T)$. The spectrum is the complement of the resolvent set as before, and, as before, if T is self-adjoint $\sigma(T)$ is a subset of $\mathbb{R}$.

3. One can show that if T is a bounded self-adjoint operator then $\sigma(T)$ is a bounded subset of $\mathbb{R}$.

4. What is going on with our example $q^n p + pq^n$ in Section §18.1.2 above? At least a partial answer is that the most obvious domain on which one can prove the operator is symmetric (and hence has real point spectrum) is the set of wavefunctions so that $q^n p\psi$ is L^2. These must fall off as $\psi(x) \sim |x|^{-\alpha}$ for $\alpha > n - 1/2$ and the putative eigenfunctions exhibited above lie outside that domain.

18.7 Commuting self-adjoint operators

Once we start admitting partially defined operators on Hilbert space we have stepped onto a slippery slope. If T_1 and T_2 are only defined on $D(T_1)$ and $D(T_2)$ respectively then $T_1 + T_2$ is only defined on $D(T_1) \cap D(T_2)$ and $T_1 \circ T_2$ is similarly only defined on

$$D(T_1 \circ T_2) = \{x | x \in D(T_2) \qquad \text{and} \qquad T_2(x) \in D(T_1)\} \tag{18.79}$$

[35] This is an example of something called a *Sobolev space*.

The problem is that these subspaces might be small, or even the zero vector space.

Example 1 Take any $y \notin D(T_1)$ and let $T_2(x) = (z, x)y$ for some z. Then $T_1 \circ T_2$ is only defined on the zero vector.

Example 2: Let $\{x_n\}$ be an ON basis for $\mathcal{H}$ and let $\{y_n\}$ be another ON basis so that each y_n is an infinite linear combination of the x_m's and vice versa. Then let $D(T_1)$ be the set of finite linear combinations of the x_n's and $D(T_2)$ be the set of finite linear combinations of the y_n's. Then $D(T_1)$ and $D(T_2)$ are dense and $D(T_1) \cap D(T_2) = \{0\}$. In order to produce an example of two such ON bases consider $\mathcal{H} = \ell^2(\mathbb{C})$ and take the Cayley transform of $T = \lambda(S + S^\dagger)$ where λ is real and of magnitude $|\lambda| < \frac{1}{2}$. Then

$$U = (1 + iT)(1 - iT)^{-1} \tag{18.80}$$

is a well-defined unitary operator which takes the standard basis $e_n = (0, \dots, 0, 1, 0 \dots)$ of ℓ^2 to a new basis f_n all of which are infinite linear combinations of the e_n.

Definition: Two self-adjoint operators T_1 and T_2 are said to *commute* if their PVM's commute, that is, for all $E_1, E_2 \in \mathcal{B}(\mathbb{R})$

$$P_{T_1}(E_1)P_{T_2}(E_2) = P_{T_2}(E_2)P_{T_1}(E_1) \tag{18.81}$$

When T_1, T_2 are bounded the spectral theorem shows that this reduces to the usual notion that $[T_1, T_2] = 0$.

18.8 Stone's theorem

Part of the proof of the spectral theorem involves showing that if $x \to f(x)$ is a continuous function, or more generally a measurable function, then if T is self-adjoint the operator $f(T)$ is densely defined and makes sense. In particular, if this is applied to the exponential function $f(x) = \exp[ix]$ one obtains an operator with domain all of $\mathcal{H}$ (by (18.73)) which is in fact a bounded operator. All the good formal properties that we expect of this operator are in fact true:

Theorem: If T is a self-adjoint operator then the family of operators $U(t) = \exp[itT]$ satisfies

1. $U(t)U(s) = U(t + s)$
2. $t \to U(t)$ is continuous in the strong operator topology.
3. The limit $\lim_{t\to 0}(U(t)x - x)/t$ exists iff $x \in D(T)$ in which case the limit is equal to $iT(x)$

♣Should explain that $t \to U(t)$ is continuous in the norm topology iff T is bounded. ♣

Stone's theorem, is a converse statement: First, we define a *strongly continuous one parameter group* is a homomorphism from $\mathbb{R}$ to the group of unitary operators on Hilbert space which is continuous in the strong topology, that is:

1. $U(t)U(s) = U(t + s)$

2. For each $x \in H$, $\lim_{t_1 \to t_2} U(t_1)x = U(t_2)x$.

Theorem [Stone's theorem]: If $U(t)$ is a strongly continuous one-parameter group of unitary operators on $\mathcal{H}$ then there is a self-adjoint operator T on $\mathcal{H}$ such that $U(t) = \exp[itT]$.

Remarks

1. If T is bounded then we can simply define

$$U(t) = \sum_{n=0}^{\infty} \frac{(itT)^n}{n!} \tag{18.82}$$

   This converges in the operator norm. In particular $t \to U(t)$ is continuous in the operator norm. However, such a definition will not work if T is an unbounded operator.

2. The proof is in Reed-Simon Theorem VIII.8, p.266.

3. Let us return to our third lie of §18.1.3. We cannot exponentiate $T = -id/dx$ on $L^2[0,1]$ simply because it is unbounded and not self-adjoint with the domains $D(T)$ and $D(T^\dagger)$ given above. Since T is not defined on functions with nonzero values at $x = 0, 1$ it is not surprising that we cannot define the translation of a wavepacket past that point. If we take one of the self-adjoint extensions T_α then we are working with twisted boundary conditions on the circle. Now is is quite sensible to be able to translate by an arbitrary amount around the circle.

4. What about the translation operator on the half-line? Our naive discussion above should make it clear that in this case p is not even *essentially self-adjoint*. [36] So there is no self-adjoint extension of p acting on the Sobelev space with $\psi(x) = 0$ at $x = 0$.

5. Stone's theorem can also be proven for continuity in the compact-open topology.

18.9 Traceclass operators

We would like to define the trace of an operator but, as the first lie in Section 18.1.1 shows, we must use some care.

For simplicity we will restrict attention in this section to *bounded operators*.

Definition: An operator is called *positive* if [37]

$$(x, Tx) \geq 0 \qquad \forall x \in \mathcal{H} \tag{18.83}$$

Three immediate and easy properties of positive bounded operators are:

Theorem: If T is a positive bounded operator on a complex Hilbert space $\mathcal{H}$ then

[36] If the closure of the graph $\overline{\Gamma(T)}$ is the graph of an operator we call that operator $\bar{T}$: $\overline{\Gamma(T)} = \Gamma(\bar{T})$, and we say that T is closeable. A symmetric operator T is essentially self-adjoint if $\bar{T}$ is self-adjoint.

[37] A more accurate term would be *nonnegative*. But this is standard terminology.

1. T is self-adjoint.

2. $|(x,Ty)|^2 \leq (x,Tx)(y,Ty)$ for all $x,y \in \mathcal{H}$.

3. T has a unique positive square-root $S \in \mathcal{L}(\mathcal{H})$, i.e. S is positive and $S^2 = T$.

Proof:

1. Note that $(x,Tx) = \overline{(x,Tx)} = (Tx,x)$. Now, if $\mathcal{H}$ is a complex Hilbert space we can then use the polarization identity to prove $(x,Ty) = (Ty,x)$ and hence T is self-adjoint. The statement fails for real Hilbert spaces.

2. Consider the inequalities $0 \leq (x+\lambda e^{i\theta}y, T(x+\lambda e^{i\theta}y))$, where λ is real and we choose a suitable phase $e^{i\theta}$ and require the discriminant of the resulting quadratic polynomial in λ to be nonpositive.

3. Follows immediately from the spectral theorem. ♠

Theorem/Definition: If T is a positive operator on a separable Hilbert space we define the trace of T by

$$\mathrm{Tr}(T) := \sum_{n=1}^{\infty}(u_n, Tu_n) \tag{18.84}$$

eq:TraceDef-1

where $\{u_n\}$ is an ON basis for $\mathcal{H}$. This sum, (which might be infinite) does not depend on the ON basis.

Proof: The sum in (18.84) is a sum of nonnegative terms and hence the partial sums are strictly increasing. They either diverge to infinity or have a limit. We use the square-root property, namely $T = S^2$ with S self-adjoint and positive to check independence of basis. Let $\{v_m\}$ be any other ON basis:

$$\begin{aligned}\mathrm{Tr}(T) &= \sum_{n=1}^{\infty}(u_n,Tu_n) = \sum_{n=1}^{\infty} \parallel Su_n \parallel^2 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}|(Su_n,v_m)|^2 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}|(u_n,Sv_m)|^2 \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}|(u_n,Sv_m)|^2 = \sum_{m=1}^{\infty} \parallel Sv_m \parallel^2 = \sum_{m=1}^{\infty}(v_m,Tv_m)\end{aligned} \tag{18.85}$$

The exchange of infinite sums in going to the second line is valid because all terms are nonnegative ♠

We now use the squareroot lemma and the above theorem to define the traceclass ideal $\mathcal{I}_1$:

Definition: The *traceclass operators* $\mathcal{I}_1 \subset \mathcal{L}(\mathcal{H})$ are those operators so that $|T| := \sqrt{T^\dagger T}$ has a finite trace: $\mathrm{Tr}(|T|) < \infty$.

With this definition there is a satisfactory notion of trace:

Theorem: [Properties of traceclass operators]

1. $\mathcal{I}_1 \subset \mathcal{L}(\mathcal{H})$ is a $*$-closed *ideal*: This means: if $T_1, T_2 \in \mathcal{I}_1$ then $T_1 + T_2 \in \mathcal{I}_1$ and if $T_3 \in \mathcal{L}(\mathcal{H})$ then $T_1 T_3 \in \mathcal{I}_1$ and $T_3 T_1 \in \mathcal{I}_1$, and, finally, T is traceclass iff $T^\dagger$ is.

2. If $T \in \mathcal{I}_1$ the trace, defined by

$$\mathrm{Tr}(T) := \sum_{n=1}^{\infty} (u_n, T u_n) \tag{18.86}$$

   where $\{u_n\}$ is any ON basis, is independent of the choice of ON basis and defines a linear functional $\mathcal{I}_1 \to \mathbb{C}$.

3. If $T_1 \in \mathcal{I}_1$ and $T_2 \in \mathcal{L}(\mathcal{H})$ then the trace is cyclic:

$$\mathrm{Tr}(T_1 T_2) = \mathrm{Tr}(T_2 T_1) \tag{18.87}$$

Proofs: The proofs are straightforward but longwinded. See Reed-Simon, Section VI.6.

Finally, we mention one more commonly used class of operators intermediate between traceclass and bounded operators:

Definition: The *compact operators* on $\mathcal{H}$, denoted $\mathcal{K}(\mathcal{H})$ is the norm-closure of the operators of finite rank. [38]

Thanks to the singular value decomposition, the canonical form of a compact operator follows immediately: There are ON sets $\{u_n\}$ and $\{w_m\}$ in $\mathcal{H}$ (not necessarily complete) and positive numbers λ_n so that

$$T = \sum_{n=1}^{\infty} \lambda_n (u_n, \cdot) w_n \tag{18.88}$$

where the convergence of the infinite sum is in the operator norm. Hence the only possible accumulation point of the λ_n is zero. For a compact self-adjoint operator there is a complete ON basis $\{u_n\}$ with

$$T = \sum_{n=1}^{\infty} \lambda_n (u_n, \cdot) u_n \tag{18.89}$$

where λ_n are real and $\lim_{n\to\infty} \lambda_n = 0$. This is called the Hilbert-Schmidt theorem.

Next, we have $\mathcal{I}_1 \subset \mathcal{K}(\mathcal{H})$. This follows because if T is traceclass so is $T^\dagger T$, but this implies that for any ON basis $\{u_n\}$ consider the linear span L_N of $\{u_n\}_{n=1}^N$. Then if $y \in L_N^\perp$

[38] This definition is fine for operators on Hilbert space but will not work for operators on Banach space. In that case one must use a different criterion, equivalent to the above for Hilbert spaces. See Reed-Simon Section VI.5.

is nonzero we can normalize it to $\| y \| = 1$ and since $L_N \cup \{y\}$ can be completed to an ON basis

$$\| Ty \|^2 + \sum_{n=1}^{N} \| Tu_n \|^2 \leq \mathrm{Tr} T^\dagger T < \infty \tag{18.90}$$

so

$$\| Ty \|^2 \leq \mathrm{Tr} T^\dagger T - \sum_{n=1}^{N} \| Tu_n \|^2 \tag{18.91}$$

Since $\mathrm{Tr} T^\dagger T < \infty$ the RHS goes to zero for $N \to \infty$. This means that

$$T = \sum_{n=1}^{\infty} (u_n, \cdot) T u_n \tag{18.92}$$

with converges in the operator norm because

$$\begin{aligned} \| T - \sum_{n=1}^{N} (u_n, \cdot) T u_n \| &= \sup_{y \neq 0} \frac{\| Ty - \sum_{n=1}^{N} (u_n, y) T u_n \|}{\| y \|} \\ &= \sup_{y \neq 0} \frac{\| T \left(y - \sum_{n=1}^{N} (u_n, y) u_n \right) \|}{\| y \|} \\ &= \sup_{y \neq 0} \frac{\| T y^\perp \|}{\sqrt{\| y^\perp \|^2 + \| y^\parallel \|^2}} \end{aligned} \tag{18.93}$$

where $y^\perp$ is the orthogonal projection to $L_N^\perp$.

In particular, for a positive traceclass operator T there is an ON basis with

$$T = \sum_{n=1}^{\infty} \lambda_n (u_n, \cdot) u_n \tag{18.94}$$

eq:Schmidt-Dec

where $\lambda_n \geq 0$, and $\mathrm{Tr}(T) = \sum_{n=1}^{\infty} \lambda_n$. This theorem is important when we discuss physical states and density matrices in quantum mechanics in Section §19 below.

Exercise

Give an example of a positive operator on a real Hilbert space which is not self-adjoint.[39]

---

19. The Dirac-von Neumann axioms of quantum mechanics

sec:DvN-Axioms

The Dirac-von Neumann axioms attempt to make mathematically precise statements associated to the physical description of quantum systems.

[39] *Answer*: Consider $1 + J^{(2)}$ on $\mathbb{R}^2$.

1. *"Space of states"*: To a "physical system" we assign a complex separable $\mathbb{Z}_2$-graded Hilbert space $\mathcal{H}$ known as the "space of states."

2. *Physical observables*: The set of physical quantities which are "observable" in this system is in 1-1 correspondence with the set $\mathcal{O}$ of self-adjoint operators on $\mathcal{H}$.

3. *Physical states*: The set of physical states of the quantum system is in 1-1 correspondence with the set $\mathcal{S}$ of positive (hence self-adjoint) trace-class operators ρ with $\mathrm{Tr}(\rho) = 1$.

4. *Physical Measurement*: Physical measurements of an observable $T \in \mathcal{O}$ when the system is in a state $\rho \in \mathcal{S}$ are governed by a probability distribution $P_{T,\rho}$ on the real line of possible outcomes. The probability of measuring the value in a set $E \in \mathcal{B}(\mathbb{R})$ is defined by
$$P_{T,\rho}(E) = \mathrm{Tr} P_T(E)\rho \tag{19.1}$$ eq:Born-VN
where P_T is the projection-valued measure of the self-adjoint operator T.

5. *Symmetries.* To state the postulate we need a definition:

   Definition An *automorphism* of a quantum system to be a pair of bijective maps $\beta_1 : \mathcal{O} \to \mathcal{O}$ and $\beta_2 : \mathcal{S} \to \mathcal{S}$ where β_1 is real linear on $\mathcal{O}$ such that (β_1, β_2) preserves probability measures:
   $$P_{\beta_1(T),\beta_2(\rho)} = P_{T,\rho} \tag{19.2}$$
   The automorphisms form a group QuantAut.

   Now, the symmetry axiom posits that if a physical system has a group G of symmetries then there is a homomorphism $\rho : G \to \mathrm{QuantAut}$.

   ♣Need to put some continuity properties on ρ. For example, for continuous groups we probably want ρ to be continuous in the compact-open topology. ♣

6. *Time evolution.* If the physical system has a well-defined notion of time, then evolution of the system in time is governed by a strongly continuous groupoid of unitary operators [40] and
$$\rho(t_2) = U(t_1, t_2)^{-1}\rho(t_1)U(t_1, t_2) \tag{19.3}$$ eq:TimeEvolve

7. *Collapse of the wavefunction* If a measurement of a physical observable corresponding to a self-adjoint operator T is made on a state ρ then the state changes discontinuously according to the result of the measurement. If T has pure point spectrum $\{\lambda_n\}_n$ with eigenspaces V_n then when T is measured the state changes discontinuously to
$$\rho \to \hat{\rho} = \sum_n P_{V_n}\rho P_{V_n} \tag{19.4}$$ eq:StateCollap
When T has a continuous spectrum an analogous, but more complicated, formula holds which takes into account the resolution of the measuring apparatus measuring a continuous spectrum.

[40] By this we simply mean that $(t_1, t_2) \to U(t_1, t_2)$ is continuous in the strong operator topology and $U(t_1, t_2)U(t_2, t_3) = U(t_1, t_3)$.

8. *Combination of systems.* If two physical systems, represented by Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are "combined" (for example, they are allowed to interact or are otherwise considered to be part of one system) then the combined system is described by the $\mathbb{Z}_2$-graded tensor product $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$.

♣This doesn't have much content. Need to say something about observables. ♣

Remarks

1. First, let us stress that the terms "physical system," "physical observables," and "symmetries of a physical system," are not *a priori* defined mathematical terms, although we do hope that they are meaningful terms to the reader. The point of the first few axioms is indeed to identify concrete mathematical objects to associate with these physical notions.

2. *"Space of states"*: The $\mathbb{Z}_2$-grading is required since we want to incorporate fermions. See below for $\mathbb{Z}_2$-graded linear algebra. Some authors have toyed with the idea of using real or quaternionic Hilbert spaces but despite much effort nobody has given a compelling case for using either. I have not seen any strong arguments advanced for why the Hilbert space should be separable. But the theory of self-adjoint operators is likely to be a good deal more complicated for non-separable Hilbert spaces.

3. *Pure states vs. mixed states.* The "physical states" referred to in axiom 3 are often called *density matrices* in the physics literature. By the Schmidt decomposition of positive traceclass operators (18.94) we can write

$$\rho = \sum \rho_n |n\rangle\langle n| \tag{19.5}$$

where $\rho_n \geq 0$ with $\sum \rho_n = 1$ and $|n\rangle$ is an ON basis for $\mathcal{H}$. Note that the set $\mathcal{S}$ is a convex set: If $\rho_1, \rho_2 \in \mathcal{S}$ then so is $t\rho_1 + (1-t)\rho_2$ for $t \in [0,1]$. For any convex set there is a notion of the *extremal points*. These are the points which are not of the form $t\rho_1 + (1-t)\rho_2$ with $0 < t < 1$ for any value of $\rho_1 \neq \rho_2$. In the convex set of physical states the extremal points correspond to projection operators onto one-dimensional subspaces

$$\rho_\ell = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} \qquad\qquad \psi \in \ell \tag{19.6}$$

The space of these extremal points are called the *pure states*. States which are not pure states are called *mixed states*. The pure states are equivalently the one-dimensional projection operators and hence are in 1-1 correspondence with the *space of lines in* $\mathcal{H}$. The space of lines in $\mathcal{H}$ is known as the *projective Hilbert space* $\mathbb{P}\mathcal{H}$. Any nonzero vector $\psi \in \mathcal{H}$ determines a line $\ell = \{z\psi | z \in \mathbb{C}\}$ so $\mathbb{P}\mathcal{H}$ is often thought of as vectors up to scale, and we can identify $\mathbb{P}\mathcal{H} = (\mathcal{H} - \{0\})/\mathbb{C}^*$. Thus, calling $\mathcal{H}$ a "space of states" is a misnomer for two reasons. First, vectors in $\mathcal{H}$ can only be used to define pure states rather than general states. Second, different vectors, namely those in the same line define the same state, so a pure state is an equivalence class of vectors in $\mathcal{H}$.

4. *Born-von Neumann formula.* Equation (19.1) goes back to the Born interpretation of the absolute square of the wavefunction as a probability density. Perhaps we should call it the Born-von Neumann formula. To recover Born's interpretation, if $\rho = |\psi\rangle\langle\psi|$ is a pure state defined by a normalized wavefunction $\psi(x) \in L^2(\mathbb{R})$ of a quantum particle on $\mathbb{R}$ and $T = q$ is the position operator then the probability of finding the particle in a measurable subset $E \in \mathcal{B}(\mathbb{R})$ is

$$P_{q,\rho}(E) = \mathrm{Tr} P_q(E)\rho = \int_E |\psi(x)|^2 dx \tag{19.7}$$

5. *Heisenberg Uncertainty Principle.* Based on its physical and historical importance one might have thought that the Heisenberg uncertainty principle would be a fundamental axiom, but in fact it is a consequence of the above. To be more precise, for a bounded self-adjoint operator $T \in \mathcal{O}$ the average value of T in state ρ is

$$\langle T\rangle_\rho := \mathrm{Tr}(T\rho) \tag{19.8}$$

We then define the *variance* or mean deviation $\sigma_{T,\rho}$ by

$$\sigma^2_{T,\rho} := \mathrm{Tr}(T - \langle T\rangle)^2\rho \tag{19.9}$$

Then if T_1 and T_2 are bounded self-adjoint operators we have, for any real number λ and phase $e^{i\theta}$,

$$0 \leq \mathrm{Tr}\left((T_1 + e^{i\theta}\lambda T_2)^\dagger (T_1 + e^{i\theta}\lambda T_2)\rho\right) \tag{19.10}$$

(provided $T_1 + e^{i\theta}\lambda T_2$ has a dense domain) and since the discriminant of the quadratic polynomial in λ is nonpositive we must have

$$\mathrm{Tr}(T_2^2\rho)\mathrm{Tr}(T_1^2\rho) \geq \frac{1}{4}\left(\langle (e^{i\theta}T_1T_2 + e^{-i\theta}T_2T_1)\rangle_\rho\right)^2 \tag{19.11}$$

Note that $(e^{i\theta}T_1T_2 + e^{-i\theta}T_2T_1)$ is (at least formally) self-adjoint and hence the quantity on the RHS is nonnegative. We can replace $T \to T - \langle T\rangle$ in the above and we deduce the general Heisenberg uncertainty relation: For all $e^{i\theta}$ we have the inequality:

$$\sigma^2_{T_1,\rho}\sigma^2_{T_2,\rho} \geq \frac{1}{4}\left(\langle (e^{i\theta}T_1T_2 + e^{-i\theta}T_2T_1)\rangle_\rho - 2\cos\theta\langle T_1\rangle_\rho\langle T_2\rangle_\rho\right)^2 \tag{19.12}$$

If we specialize to $\theta = \pi/2$ we get the Heisenberg uncertainty relation as usually stated:

$$\sigma^2_{T_1,\rho}\sigma^2_{T_2,\rho} \geq \frac{1}{4}\left(\langle i[T_1,T_2]\rangle_\rho\right)^2 \tag{19.13}$$

Actually, this does not quite accurately reflect the real uncertainty in *successive* measurements of noncommutative observables because the first measurement alters the state. For a recent discussion see [41]

6. The data of the first four axioms are completely general and are not specific to any physical system. The next two axioms rely on properties specific to a physical system.

[41] J. Distler and S. Paban, "On Uncertainties of Successive Measurements," arXiv:1211.4169.

7. *Symmetries* The meaning of β_1 being linear on $\mathcal{O}$ is that if $T_1, T_2 \in \mathcal{O}$ and $D(T_1) \cap D(T_2)$ is a dense domain such that $\alpha_1 T_1 + \alpha_2 T_2$, with α_1, α_2 real has a unique self-adjoint extension then $\beta_1(\alpha_1 T_1 + \alpha_2 T_2) = \alpha_1 \beta_1(T_1) + \alpha_2 \beta_1(T_2)$. A consequence of the symmetry axiom is that β_2 is affine linear on states:

$$\beta_2(t\rho_1 + (1-t)\rho_2) = t\beta_2(\rho_1) + (1-t)\beta_2(\rho_2) \tag{19.14}$$

eq:Aff-Lin

The argument for this is that (β_1, β_2) must preserve expectation values $\langle T \rangle_\rho$. However, positive self-adjoint operators of trace one are themselves observables and we have $\langle \rho_1 \rangle_{\rho_2} = \langle \rho_2 \rangle_{\rho_1}$, so the restriction of β_1 to $\mathcal{S}$ must agree with β_2. Now apply linearity of β_1 on the self-adjoint operators. From (19.14) it follows [42] that β must take extreme states to extreme states, and hence β_2 induces a map $\beta : \mathbb{P}\mathcal{H} \to \mathbb{P}\mathcal{H}$. Now, define the overlap of two lines $\ell_1, \ell_2 \in \mathbb{P}\mathcal{H}$ by

$$\mathcal{P}(\ell_1, \ell_2) = \frac{|\langle \psi_1 | \psi_2 \rangle|^2}{\| \psi_1 \|^2 \| \psi_2 \|^2} \tag{19.15}$$

where $\psi_1 \in \ell_1$ and $\psi_2 \in \ell_2$ are nonzero vectors. Preservation of probabilities implies that

$$\mathcal{P}(\beta(\ell_1), \beta(\ell_2)) = \mathcal{P}(\ell_1, \ell_2) \tag{19.16}$$

eq:PreservePro

So we can think of the group QuantAut as the group of maps $\mathbb{P}\mathcal{H} \to \mathbb{P}\mathcal{H}$ which satisfy (19.16). Note that if $T : \mathcal{H} \to \mathcal{H}$ is linear or anti-linear and preserves norms $\| T\psi \| = \| \psi \|$ then T descends to a map $\bar{T} := \mathbb{P}\mathcal{H} \to \mathbb{P}\mathcal{H}$ satisfying (19.16). Now *Wigner's theorem*, proved in Chapter *** below asserts that *every* map $\beta \in$ QuantAut is of this form. More precisely, there is an exact sequence:

$$1 \to U(1) \to \mathrm{Aut}_{\mathbb{R}}(\mathcal{H}) \to \mathrm{QuantAut} \to 1 \tag{19.17}$$

eq:WignerTheor

where $\mathrm{Aut}_{\mathbb{R}}(\mathcal{H})$ is the group of unitary *and* anti-unitary transformations of $\mathcal{H}$ and $U(1)$ acts on $\mathcal{H}$ by scalar multiplication. See Chapter *** below for more detail.

♣Need to comment on "Virasoro symmetry" in 2d CFT and spontaneously broken symmetries in QFT. ♣

8. *Schrödinger equation.* Suppose that the system has time-translation invariance. Then we can use the symmetry axiom and the dynamics axiom to conclude that $U(t_1, t_2) = U(t_2 - t_1)$ is a strongly-continuous one parameter group of unitary operators. Then by Stone's theorem there is a self-adjoint generator known as the *Hamiltonian* H and usually normalized by

$$U(t) = \exp\left[-\frac{i}{\hbar} t H\right] \tag{19.18}$$

where $\hbar$ is Planck's constant, so H has units of energy. Then if $\rho(t)$ is a pure state and it can be described by a differentiable family of vectors $|\psi(t)\rangle$ in $\mathcal{H}$ then these vectors should satisfy the Schrödinger equation

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle \tag{19.19}$$

More generally, if $t \to H(t)$ is a family of self-adjoint operators then $U(t_1, t_2) = P\exp - \frac{i}{\hbar} \int_{t_1}^{t_2} H(t')dt'$.

♣Need to say in what sense it is continuous ♣

[42] For some interesting discussion of related considerations see B. Simon, "Quantum Dynamics: From Automorphism to Hamiltonian."

9. *Collapse of the wavefunction.* We have given the result where a measurement is performed but the value of the measurement is not recorded. We can also speak of conditional probabilities. If the measured value of T is λ_i then

$$\rho \to \rho' = \frac{P_{V_i}\rho P_{V_i}}{\mathrm{Tr} P_{V_i}\rho} \tag{19.20}$$

eq:MeausreE1

This rule should be thought of in terms of conditional probability. When we know something happened we should renormalize our probability measures to account for this. This is related to "Bayesian inference" and "Bayesian updating." The denominator in (19.20) is required so that ρ' has trace one. (Note that the denominator is nonzero because by assumption the value λ_i was measured.)

Of course this axiom is quite notorious and generates a lot of controversy. Briefly, there is a school of thought which denies that there is any such discontinuous change in the physical state. Rather, one should consider the quantum mechanics of the full system of measuring apparatus together with the measured system. All time evolution is smooth unitary evolution (19.3). If the measuring apparatus is macroscopic then the semiclassical limit of quantum mechanics leads to classical probability laws governing the measuring apparatus and one can derive the *appearance* of the collapse of the wavefunction. This viewpoint relies on *phase decoherence* of nearly degenerate states in a large Hilbert space of states describing a fixed value of a classical observable. According to this viewpoint Axiom 7 should *not* be an axiom. Rather, it is an effective description of "what really happens." For references see papers of W. Zurek. [43]

10. *Simultaneous measurement.* If T_1 and T_2 commute then they can be "simultaneously measured." What this means is that if we measure T_1 then the change (19.4) of the physical state does not alter the probability of the subsequent measurement of T_2:

$$\begin{aligned} P_{T_2,\hat{\rho}}(E) &= \mathrm{Tr} P_{T_2}(E) \sum_n P_{V_n}\rho P_{V_n} \\ &= \sum_n \mathrm{Tr} P_{T_2}(E) P_{V_n}\rho P_{V_n} \\ &= \sum_n \mathrm{Tr} P_{V_n} P_{T_2}(E)\rho \\ &= \mathrm{Tr}(\sum_n P_{V_n}) P_{T_2}(E)\rho \\ &= \mathrm{Tr} P_{T_2}(E)\rho = P_{T_2,\rho}(E) \end{aligned} \tag{19.21}$$

Although sometimes stated as an axiom this is really a consequence of what was said above. (And we certainly don't want any notion of simultaneity to be any part of the fundamental axioms of quantum mechanics!)

[43] Short expositions are in T. Banks, "Locality and the classical limit of quantum mechanics," arXiv:0809.3764 [quant-ph]; "The interpretation of quantum mechanics," http://blogs.discovermagazine.com/cosmicvariance/files/2011/11/banks-qmblog.pdf We also recommend S. Coleman's classic colloquium, "Quantum Mechanics: In Your Face." http://media.physics.harvard.edu/video/?id=SidneyColeman QMIYF

11. The fact that we work with $\mathbb{Z}_2$-graded Hilbert spaces and take a $\mathbb{Z}_2$-graded tensor product can have important consequences. An example arises in the currently fashionable topic of Majorana fermions in condensed matter theory.

12. *Relation to classical mechanics.* There is a formulation of classical mechanics which is closely parallel to the above formulation of quantum mechanics. In order to discuss it one should focus on C^*-algebras. In quantum mechanics one can consider the C^* algebra of bounded operators $\mathcal{L}(\mathcal{H})$ or its subalgebra of compact operators $\mathcal{K}(\mathcal{H})$. The positive elements are self-adjoint, and hence observables. If $\mathcal{M}$ is a phase space, i.e. a symplectic manifold, then the analogous C^* algebra is $C_0(\mathcal{M})$, the commutative C^* algebra of complex-valued continuous functions $f : \mathcal{M} \to \mathbb{C}$ vanishing at infinity[44] with $\| f \| = \sup_{x\in\mathcal{M}} |f(x)|$. The observables are the real-valued functions and the positive functions are necessarily observables. Now, in general, one defines

    Definition A *state* on a C^* algebra $\mathfrak{A}$ is a linear map $\omega : \mathfrak{A} \to \mathbb{C}$ which is positive, i.e., $\omega(A) \geq 0$ if $A \geq 0$, and of norm 1.

    Then there are two relevant theorems:

    Theorem 1: If $\mathfrak{A} = \mathcal{K}(\mathcal{H})$ then the space of states in the sense of C^*-algebra theory is in fact the set of positive traceclass operators of trace 1, i.e., the set of density matrices of quantum mechanics.

    Theorem 2: If X is any Hausdorff space and $\mathfrak{A} = C_0(X)$ then the space of states is the space of probability measures on X.

    Therefore, in the formulation of classical mechanics one defines the observables $\mathcal{O}_{\rm class}$ to be $f \in C_0(\mathcal{M}; \mathbb{R})$ and the states $\mathcal{S}_{\rm class}$ to be probability measures $d\mu$ on $\mathcal{M}$. Then, given an observable f and a state $d\mu$ we get a probability measure on $\mathbb{R}$, which, when evaluated on a Borel set $E \in \mathcal{B}(\mathbb{R})$ is

    $$P_{f,d\mu}(E) := \int_{f^{-1}(E)} d\mu \tag{19.22}$$

    The expectation value of f is $\int_X f d\mu$ and if $d\mu$ is a Dirac measure at some point $x \in \mathcal{M}$ then there is no variance, $\langle f^2 \rangle_{d\mu} = \langle f \rangle^2_{d\mu}$. Finally, since $\mathcal{M}$ is symplectic there is a canonical Liouville measure $d\mu_{\rm Liouville} = \frac{\omega^n}{n!}$ where ω is the symplectic form and given a state $d\mu$ we can define $d\mu(x) = \rho(x) d\mu_{\rm Liouville}$. Then the classical analog of the Schrödinger equation is the Liouville equation

    $$\frac{d\rho(x;t)}{dt} = -\{H, \rho\} \tag{19.23}$$

    This is a good formalism for describing semiclassical limits and coherent states.

13. Of course, our treatment does not begin to do justice to the physics of quantum mechanics. Showing how the above axioms really lead to a description of Nature requires an entire course on quantum mechanics. We are just giving the bare bones axiomatic framework.

[44] This means that for all $\epsilon > 0$ the set $\{x \in \mathcal{M} || f(x)| \geq \epsilon\}$ is compact.

References: There is a large literature on attempts to axiomatize quantum mechanics. The first few chapters of Dirac's book is the first and most important example of such an attempt. Then in his famous 1932 book *The Mathematical Foundations of Quantum Mechanics* J. von Neumann tried to put Dirac's axioms on a solid mathematical footing, introducing major advances in mathematics (such as the theory of self-adjoint operators) along the way. For an interesting literature list and commentary on this topic see the Notes to Section VIII.11 in Reed and Simon. We are generally following here the very nice treatment of L. Takhtadjan, *Quantum Mechanics for Mathematicians*, GTM 95 which in turn is motivated by the approach of G. Mackey, *The Mathematical Foundations of Quantum Mechanics*, although we differ in some important details.

Exercise

Show that (19.1) is in fact a probability measure on the real line.

Exercise *States of the two-state system*

Consider the finite-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2$. Show that the physical states can be parametrized as:

$$\rho = \frac{1}{2}(1 + \vec{x} \cdot \vec{\sigma}) \tag{19.24}$$

where $\vec{x} \in \mathbb{R}^3$ and $\vec{x}^2 \leq 1$. Note that this is a convex set and the set of extremal points is the sphere $S^2 \cong \mathbb{C}P^1 = \mathbb{P}\mathcal{H}$.

Exercise *Von Neumann entropy*

The von Neumann entropy of a state ρ is defined to be $S(\rho) := -\mathrm{Tr}(\rho \log \rho)$.

Suppose that $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is a product system. For a state ρ define $\rho_B \in \mathcal{L}\mathcal{H}_B$ by $\rho_B = \mathrm{Tr}_{\mathcal{H}_A}(\rho)$ and similarly for ρ_A.

Show that if ρ is a pure state then $S(\rho_A) = S(\rho_B)$.

---

20. Canonical Forms of Antisymmetric, Symmetric, and Orthogonal matrices

anFormSymmOrth

20.1 Pairings and bilinear forms

20.1.1 Perfect pairings

Definition. Suppose M_1, M_2, M_3 are R-modules for a ring R.

1.) Then a *M_3-valued pairing* is a bilinear map

$$b : M_1 \times M_2 \to M_3 \tag{20.1}$$

2.) It is said to be *nondegenerate* if the induced maps

$$L_b : M_1 \to \mathrm{Hom}_R(M_2, M_3) \qquad m_1 \mapsto b(m_1, \cdot) \tag{20.2}$$

and

$$R_b : M_2 \to \mathrm{Hom}_R(M_1, M_3) \qquad m_2 \mapsto b(\cdot, m_2) \tag{20.3}$$

are injective.

3.) It is said to be a *perfect pairing* if these maps are injective and surjective, i.e. L_b and R_b define isomorphisms.

Examples

1. $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined by $b(x, y) = kxy$ is nondegenerate for $k \neq 0$ but only a perfect pairing of $\mathbb{Z}$-modules (i.e. abelian groups) for $k = \pm 1$.
2. $\mathbb{Z} \times U(1) \to U(1)$ defined by $b(n, e^{i\theta}) = e^{in\theta}$ is a perfect pairing of $\mathbb{Z}$-modules (i.e. abelian groups) but $b(n, e^{i\theta}) = e^{ikn\theta}$ for k an integer of absolute value > 1 is not a perfect pairing.
3. $\mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ given by $b(r, s) = rs \mathrm{mod} n$ is a perfect pairing of $\mathbb{Z}$-modules (i.e. abelian groups).

20.1.2 Vector spaces

Now we specialize to a pairing of vector spaces over a field κ with $M_3 = \kappa$. Then a pairing is called a bilinear form:

Definition. A bilinear form on a vector space is a map:

$$\langle \cdot, \cdot \rangle : V \times V \to \kappa \tag{20.4}$$

which is *linear* in both variables.

It is called a *symmetric quadratic form* if:

$$\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle \qquad \forall v_1, v_2 \in V \tag{20.5}$$

eq:symmt

and *antisymmetric quadratic form* if:

$$\langle v_1, v_2 \rangle = -\langle v_2, v_1 \rangle \qquad \forall v_1, v_2 \in V \tag{20.6}$$

eq:symmtp

and it is *alternating* if: $\langle v, v \rangle = 0$ for all $v \in V$.

Remark: Note we are using angle brackets for bilinear forms and round brackets for sesquilinear forms. Some authors have the reverse convention!

Exercise

a.) Use the universal property of the tensor product to show that the space of bilinear forms on a vector space over κ is

$$\mathrm{Bil}(V) \cong (V \otimes V)^* = V^* \otimes V^*. \tag{20.7}$$

b.) Thus, a bilinear form induces two maps $V \to V^*$ (by contracting with the first or second factor). Show that if the bilinear form is nondegenerate then this provides two isomorphisms of V with V^*.

Exercise

a.) Show that if the field κ is not of characteristic 2 then a form is alternating iff it is antisymmetric.

b.) Show that alternating and antisymmetric are not equivalent for a vector space over the field $\mathbb{F}_2$. [45]

20.1.3 Choosing a basis

If we choose a an ordered basis $\{v_i\}$ for V then a quadratic form is given by a matrix

$$Q_{ij} = \langle v_i, v_j \rangle \tag{20.8}$$

Under a change of basis

$$w_i = \sum_j S_{ji} v_j \tag{20.9}$$

the matrix changes by

$$\tilde{Q}_{ij} = \langle w_i, w_j \rangle = (S^{tr} Q S)_{ij} \tag{20.10}$$

Note that the symmetry, or anti-symmetry of Q_{ij} is thus preserved by arbitrary change of basis. (This is not true under similarity transformations $Q \to SQS^{-1}$.)

Remark: The above definitions apply to any module M over a ring R. We will use the more general notion when discussing abstract integral lattices.

20.2 Canonical forms for symmetric matrices

Theorem If $Q \in M_n(\kappa)$ is symmetric, (and κ is any field of characteristic $\neq 2$) then there is a nonsingular matrix $S \in GL(n, \kappa)$ such that $S^{tr} Q S$ is diagonal:

$$S^{tr} Q S = Diag\{\lambda_1, \ldots, \lambda_n\} \tag{20.11}$$ eq:brdg

[45] *Answer*: Alternating implies antisymmetric, but not vice versa.

Proof:[46]

Suppose we have a quadratic form Q. If $Q = 0$ we are done. If $Q \neq 0$ then there exists a v such that $Q(v, v) \neq 0$, because

$$2Q(v, w) = Q(v + w, v + w) - Q(v, v) - Q(w, w) \tag{20.12}$$

(Note: Here we have used symmetry and the invertibility of 2.)

Now, we proceed inductively. Suppose we can find $v_1, \ldots, v_k$ so that $Q(v_i, v_j) = \lambda_i \delta_{ij}$ with $\lambda_i \neq 0$, and define V_k to be the span of $\{v_1, \ldots, v_k\}$.

Let

$$V_k^{\perp} := \{w | Q(w, v) = 0 \qquad \forall v \in V_k\} \tag{20.13}$$

We claim $V = V_k \oplus V_k^{\perp}$. First note that if $u = \sum a_i v_i \in V_k \cap V_k^{\perp}$ then $Q(u, v_i) = 0$ implies $a_i = 0$ since $\lambda_i \neq 0$. Moreover, for any vector $u \in V$

$$u^{\perp} = u - \sum_i Q(u, v_i) \lambda_i^{-1} v_i \in V_k^{\perp} \tag{20.14}$$

and therefore u is in $V_k + V_k^{\perp}$.

Now consider the restriction of Q to $V_k^{\perp}$. If this restriction is 0 we are done. If the restriction is not zero, then there exists a $v_{k+1} \in V_k^{\perp}$ with $Q(v_{k+1}, v_{k+1}) = \lambda_{k+1} \neq 0$, and we proceed as before. On a finite dimensional space the procedure must terminate ♠.

Remark: The above theorem definitely fails for a field of characteristic 2. For example the symmetric quadratic form

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{20.15}$$

on $\mathbb{F}_2 \oplus \mathbb{F}_2$ cannot be diagonalized.

Returning to fields of characteristic $\neq 2$, the diagonal form above still leaves the possibility to make further transformations by which we might simplify the quadratic form. Now by using a further diagonal matrix D we can bring it to the form:

$$(SD)^{tr} Q(SD) = Diag\{\mu_1^2 \lambda_1, \ldots, \mu_n^2 \lambda_n\} \tag{20.16}$$

eq:diagtrmn

Now, at this point, the choice of field κ becomes very important.

Suppose the field $\kappa = \mathbb{C}$. Then note that by a further transformation of the form (20.16) we can always bring A to the form

$$Q = \begin{pmatrix} 1_r & 0 \\ 0 & 0 \end{pmatrix} \tag{20.17}$$

However, over $k = \mathbb{R}$ there are further invariants:

[46] Taken from Jacobsen, Theorem 6.5.

Theorem: [Sylvester's law]. For any real symmetric matrix A there is an invertible real matrix S so that

$$SQS^{\rm tr} = Diag\{1^p, (-1)^q, 0^n\} \qquad S \in GL(n, \mathbb{R}) \tag{20.18}$$

Proof: Now, λ_i, μ_i in (20.16) must both be real. Using real μ_i we can set $\mu_i^2 \lambda_i = \pm 1, 0$. ♠

Remarks

1. The point of the above theorem is that, since μ_i are real one *cannot* change the sign of the eigenvalue λ_i. The rank of A is $p+q$. The **signature** is (p, q, n) (sometimes people use $p-q$). If $n = 0$ *A is nondegenerate.* If $n = q = 0$ *A is positive definite*

2. If $\kappa = \mathbb{Q}$ there are yet further invariants, since not every positive rational number is the square of a rational number, so the invariants are in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$.

Finally, we note that the transormations $A \to SAS^{-1}$ and $A \to SAS^{tr}$ do not interact very well in the following sense: Suppose you know a complex matrix A is symmetric. Does that give you any useful information about its diagonalizability or its Jordan form under $A \to SAS^{-1}$ with $S \in GL(n, \mathbb{C})$? The answer is *no!*:

Theorem An arbitrary complex matrix is similar to a complex symmetric matrix.

Idea of proof: Since there is an S with SAS^{-1} in Jordan form it suffices to show that $J_\lambda^{(k)}$ is similar to a complex symmetric matrix. Write

$$J_\lambda^{(k)} = \lambda 1 + \frac{1}{2}(N + N^{tr}) + \frac{1}{2}(N - N^{tr}) \tag{20.19}$$

One diagonalizes $(N - N^{tr})$ by a unitary matrix U such that $U(N + N^{tr})U^{-1}$ remains symmetric. ♠

Exercise

a.) Show that the complex symmetric matrix

$$\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \tag{20.20}$$

has a zero trace and determinant and find its Jordan form.

b.) Show that there is no nonsingular matrix S such that

$$S \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} S^{tr} \tag{20.21}$$

is diagonal.

20.3 Orthogonal matrices: The real spectral theorem

♣Clarify that now you are adding the data of an inner product space together with a bilinear form! ♣

Sometimes we are interested only in making transformations for S an orthogonal matrix.

Theorem [Real Finite Dimensional Spectral Theorem]. If A is a real symmetric matrix then it can be diagonalized by an orthogonal transformation:

$$A^{tr} = A, A \in M_n(\mathbb{R}) \to \exists S \in O(n,\mathbb{R}): \qquad SAS^{tr} = Diag\{\lambda_1, \ldots, \lambda_n\} \tag{20.22}$$

Proof: The proof is similar to that for the finite dimensional spectral theorem over $\mathbb{C}$. If we work over the complex numbers, then we have at least one characteristic vector $Av = \lambda v$. Since λ is real and A hermitian, $Av^* = \lambda v^*$, so $A(v + v^*) = \lambda(v + v^*)$. Thus, A in fact has a real eigenvector. Now take the orthogonal complement to $(v + v^*)$ and use induction. ♠

Example: In mechanics we use this theorem to define moments of inertia

As an application we have the analog of the theorem that unitary matrices can be unitarily diagonalized:

Theorem Every real orthogonal matrix O can be brought to the form:

$$SOS^{tr} = Diag\{+1^r, -1^q, R(\theta_i)\} \qquad S \in O(n,\mathbb{R}) \tag{20.23}$$

by an orthogonal transformation. Here

$$R(\theta_i) = \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix} \qquad \theta_i \neq n\pi \tag{20.24}$$

Proof: Consider $T = O + O^{-1} = O + O^{tr}$ on $\mathbb{R}^n$. This is real symmetric, so by the spectral theorem (over $\mathbb{R}$) there is an orthogonal basis in which $\mathbb{R}^n = \oplus_i V_i$ where T has eigenvalue λ_i on the subspace V_i and λ_i are the distinct eigenvalues of T. They are all real. Note that for all vectors $v \in V_i$ we have

$$(O^2 - \lambda_i O + 1)v = 0 \tag{20.25}$$ eq:QuadO1

so O restricted to V_i satisfies $O^2 - \lambda_i O + 1 = 0$. Therefore, if $v \in V_i$ then the vector space $W = Span\{v, Ov\}$ is preserved by O. Moreover, O preserves the decomposition $W \oplus W^{\perp}$. Therefore by induction we need only analyze the cases where W is 1 and 2-dimensional. If W is one dimensional then $Ov = \mu v$ and we easily see that $\mu^2 = 1$ so $\mu = \pm 1$. Suppose W is two-dimensional and O acting on W satisfies

$$O^2 - \lambda O + 1 = 0 \tag{20.26}$$ eq:QuadO

Now, by complexification we know that O is unitary and hence it is diagonalizable (over $\mathbb{C}$) and its eigenvalues must be phases $\{e^{i\theta_1}, e^{i\theta_2}\}$. On the other hand (20.26) will be true after diagonalization (over $\mathbb{C}$) so $\lambda = e^{i\theta} + e^{-i\theta} = 2\cos\theta$ for both angles $\theta = \theta_1$

and $\theta = \theta_2$. Now, on this two-dimensional space we have $O^{tr} = O^{-1}$ as 2×2 matrices. Therefore det$O$ is ± 1 and moreover

$$O = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow O^{tr} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \qquad \text{and} \qquad O^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{20.27}$$

Thus, $O^{tr} = O^{-1}$ implies $a = d$ and $b = -c$ if det$O = +1$ and it implies $a = -d$ and $b = c$ if det$O = -1$. In the first case we go back to equation (20.26) and solve for a, b to find $O = R(\pm\theta)$. In the second case we find

$$O = R(\theta)P \tag{20.28}$$

with

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{20.29}$$

Next we observe that $P^2 = 1$ and $PR(\phi)P = R(-\phi)$ so we may then write

$$O = R(\phi)PR(\phi)^{tr} \tag{20.30}$$

with $2\phi = \theta$, and hence in the second case we can transform O to P which is of the canonical type given in the theorem. ♠

20.4 Canonical forms for antisymmetric matrices

Theorem Let κ be any field. If $A \in M_n(\kappa)$ is antisymmetric there exists $S \in GL(n, \kappa)$ that brings A to the canonical form:

$$SAS^{tr} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus 0_{n-r} \tag{20.31}$$

Proof: The proof is very similar to the case of symmetric forms. Let us suppose that A is the matrix with respect to an antisymmetric quadratic form Q on a vector space V. If $Q = 0$ we are done. If $Q \neq 0$ then there must be linearly independent vectors u, v with $Q(u, v) = q \neq 0$. Define $u_1 = u$ and $v_1 = q^{-1}v$. Now Q has the required canonical form with respect to the ordered basis $\{u_1, v_1\}$.

Now we proceed by induction. Suppose we have constructed linearly independent vectors $(u_1, v_1, \cdots, u_k, v_k)$ such that $Q(u_i, v_j) = \delta_{ij}$, and $Q(u_i, u_j) = Q(v_i, v_j) = 0$. Let $V_k = Span\{u_1, v_1, \cdots, u_k, v_k\}$. Then again $V = V_k \oplus V_k^{\perp}$ where again we define

$$V_k^{\perp} := \{w | Q(w, v) = 0 \qquad \forall v \in V_k\} \tag{20.32}$$

It is easy to see that $V_k \cap V_k^{\perp}$ and if $w \in V$ is any vector then

$$w^{\perp} = w + \sum_i (Q(w, u_i)v_i - Q(w, v_i)u_i) \in V_k^{\perp} \tag{20.33}$$

so $V = V_k + V_k^\perp$. Restricting Q to $V_k^\perp$ we proceed as above. ♠

As usual, if we put a restriction on the change of basis we get a richer classification:

Theorem Every *real antisymmetric* matrix $A^{tr} = -A$ can be skew-diagonalized by $S \in O(n, \mathbb{R})$, that is, A can be brought to the form:

$$SAS^{tr} = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 & \cdots \\ -\lambda_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \lambda_2 & \cdots \\ 0 & 0 & -\lambda_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{20.34}$$

The λ_i are called the *skew eigenvalues*. Note that, without a choice of orientation they are only defined up to sign.

Idea of proof: There are two ways to prove this. One way is to use the strategies above by using induction on the dimension. Alternatively, we can view A as an operator on $\mathbb{R}^n$ and observe that iA is an Hermitian operator on $\mathbb{C}^n$ so there is a basis of ON eigenvectors of iA. But

$$iAv = \lambda v \tag{20.35}$$

with λ real implies

$$\begin{aligned} Av &= -i\lambda v \\ Av^* &= i\lambda v^* \end{aligned} \tag{20.36}$$

So if an ON basis on $\mathbb{C}^n$ then we get an orthogonal basis on $\mathbb{R}^n$ consisting of $u_1 = v + v^*$ and $u_2 = i(v - v^*)$ and

$$\begin{aligned} Au_1 &= \lambda u_2 \\ Au_2 &= -\lambda u_1 \end{aligned} \tag{20.37}$$

Letting S be the matrix given by the columns of these vectors we have $S^{tr}AS$ is in the above block-diagonal form ♠

20.5 Automorphism Groups of Bilinear and Sesquilinear Forms

Given a bilinear form, or a sesquilinear form Q on a vector space V, the *automorphism group of the form* is the subgroup of operators $T \in GL(V)$ such that

$$Q(Tv, Tw) = Q(v, w) \tag{20.38}$$ eq:autgroup

for all $v, w \in V$.

In special cases these groups have special names:

1. *Q is a nondegenerate symmetric bilinear form on V*: Then the group of automorphisms is denoted $O(Q)$ and it is called the orthogonal group of the form. If V is a complex vector space of dimension n and we choose a basis with $Q = 1$ then we

define the matrix group $O(n,\mathbb{C})$ as the group of $n \times n$ complex invertible matrices S such that $S^{tr}S = 1$. If V is a real vector space and Q has signature $(-1)^p, (+1)^q$ then we can choose a basis in which the matrix form of Q is

$$\eta := \begin{pmatrix} -1_p & 0 \\ 0 & +1_q \end{pmatrix} \tag{20.39}$$

and the resulting matrix group, denoted $O(p,q;\mathbb{R})$, is the group of invertible matrices so that

$$S^{tr}\eta S = \eta \tag{20.40}$$

2. *Q is a nondegenerate anti-symmetric bilinear form on V*: In this case the group of automorphisms is called the *symplectic group* $Sp(Q)$. If V is finite dimensional and we are working over any field κ we can choose a basis for V in which Q has matrix form

$$J = \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix} \tag{20.41}$$

and then the resulting matrix group, which is denoted $Sp(2n;\kappa)$ for field κ, is the set of invertible matrices with matrix elements in κ such that

$$S^{tr}JS = J \tag{20.42}$$

3. *Sesquilinear forms.* It also makes sense to talk about the automorphism group of a sesquilinear form on a complex inner product space. If there is an ON basis $\{e_i\}$ in which the matrix $h(e_i, e_j)$ is of the form

$$h_{ij} = \begin{pmatrix} 1_p & 0 \\ 0 & -1_q \end{pmatrix} \tag{20.43}$$

and then the resulting matrix group, which is denoted $U(p,q)$ or $U(p,q;\mathbb{C})$, is the set of invertible matrices in $GL(n,\mathbb{C})$ so that

$$U^\dagger h U = h \tag{20.44}$$

Remarks

1. When working over rings and not fields we might not be able to bring Q to a simple standard form like the above, nevertheless, $Aut(Q)$ remains a well-defined group.

2. We will look at these groups in much more detail in our chapter on a Survey of Matrix Groups

21. Other canonical forms: Upper triangular, polar, reduced echelon

21.1 General upper triangular decomposition

Theorem 13 Any complex matrix can be written as $A = UT$, where U is unitary and T is upper triangular.

This can be proved by successively applying reflections to the matrix A. I.e. we define

$$R_{ij}(v) = \delta_{ij} - 2\frac{v_i v_j}{v \cdot v} \tag{21.1}$$

This is a reflection in the hyperplane $v^\perp$ and hence an orthogonal transformation. Consider the vector A_{k1}. If it is zero there is nothing to do. If it is nonzero then this vector, together with e_1 span a 2-dimensional plane. We can reflect in a line in this plane to make A_{k1} parallel to e_1. Now consider everything in $e_1^\perp$. Then A_{k2} is a vector which forms a 2-dimensional plane with e_2. We can repeat the process. In this way one can choose vectors $v_1, \ldots, v_n$ so that $R(v_n) \cdots R(v_1)A$ is upper triangular.

For this and similar algorithms G.H. Golub and C.F. Van Loan, *Matrix Computations*.

21.2 Gram-Schmidt procedure

c:Gram-Schmidt

In the case where A is nonsingular the above theorem can be sharpened. In this case the upper triangular decomposition is closely related to the Gram-Schmidt procedure. Recall that the Gram-Schmidt procedure is the following:

Let $\{u_i\}$ be a set of linearly independent vectors. The GS procedure assigns to this an ON set of vectors $\{v_i\}$ with the same linear span:

The procedure:

a.) Let $w_1 = u_1$, define $v_1 = w_1 / \parallel w_1 \parallel$.

b.) Let $w_2 = u_2 - (v_1, u_1)v_1$, define $v_2 = w_2 / \parallel w_2 \parallel$.

c.) Let $w_n = u_n - \sum_{k=1}^{n-1}(v_k, u_n)v_k$, define $v_n = w_n / \parallel w_n \parallel$.

Theorem 14 Any nonsingular matrix $A \in GL(n, \mathbb{C})$ can be uniquely written as

$$A = UT \tag{21.2}$$

where U is unitary and T is upper triangular, with positive real diagonal entries. Any nonsingular matrix $A \in GL(n, \mathbb{R})$ can be uniquely written as

$$A = OT \tag{21.3}$$

where O is orthogonal and T is upper triangular with positive real diagonal entries.

Proof: Note that in the Gram-Schmidt procedure the bases are related by

$$v_i = \sum T_{ji} u_j \tag{21.4}$$

eq:gmi

where T is an invertible *upper triangular* matrix. Now, let A_{ij} be any nonsingular matrix. Choose any ON basis $\tilde{v}_i$ for $\mathbb{C}^n$ and define:

$$u_j := \sum_{i=1}^{n} A_{ij}\tilde{v}_i \tag{21.5}$$ eq:gmii

This is another basis for the vector space. Then applying the GS procedure to the system $\{u_j\}$ we get an ON set of vectors v_i satisfying (21.4). Therefore,

$$v_j = \sum_{j,k} A_{kj}T_{ji}\tilde{v}_k \tag{21.6}$$ eq:gmiii

Since v_i and $\tilde{v}_i$ are two ON bases, they are related by a unitary transformation, therefore

$$U_{ki} = \sum_{j} A_{kj}T_{ji} \tag{21.7}$$ eq:gmiv

is unitary. Since T is invertible the theorem follows ♠.

Exercise *Gram-Schmidt at a glance*

Let $u_1, \ldots, u_n$ be n linearly independent vectors. Show that the result of the Gram-Schmidt procedure is summarized in the single formula:

$$v_n = \frac{(-1)^{n-1}}{\sqrt{D_{n-1}D_n}}\det\begin{pmatrix} u_1 & \cdots & u_n \\ (u_1,u_1) & \cdots & (u_n,u_1) \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ (u_1,u_{n-1}) & \cdots & (u_n,u_{n-1}) \end{pmatrix} \tag{21.8}$$ eq:gssci

$$D_n = \det\begin{pmatrix} (u_1,u_1) & \cdots & (u_n,u_1) \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ (u_1,u_n) & \cdots & (u_n,u_n) \end{pmatrix}$$

21.2.1 Orthogonal polynomials

Let $w(x)$ be a nonnegative function on $[a,b]$ such that

$$\int_a^b x^N w(x)dx < \infty \qquad N \geq 0 \tag{21.9}$$ eq:opi

We can define a Hilbert space by considering the complex-valued functions on $[a,b]$ such that

$$\int_a^b |f(x)|^2 w(x) dx < \infty \tag{21.10}$$ eq:opii

Let us call this space $L^2\big([a,b], d\mu\big)$, the L^2 functions wrt the measure $d\mu(x) = w(x)dx$. Applying the Gram-Schmidt procedure to the system of functions

$$\{1, x, x^2, \dots\} \tag{21.11}$$

leads to a system of *orthogonal polynomials* in terms of which we may expand any smooth function.

Example. Legendre polynomials. Choose $[a,b] = [-1,1]$, $w(x) = 1$. We obtain:

$$\begin{aligned} u_0 = 1 \qquad &\to \qquad \phi_0 = \frac{1}{\sqrt{2}} \\ u_1 = x \qquad &\to \qquad \phi_1 = \sqrt{\frac{3}{2}} x \\ u_2 = x^2 \qquad &\to \qquad \phi_2 = \sqrt{\frac{5}{2}} \frac{3x^2 - 1}{2} \end{aligned} \tag{21.12}$$ eq:legpolys

In general

$$\phi_n(x) = \sqrt{\frac{2n+1}{2}} P_n(x) \tag{21.13}$$

where $P_n(x)$ are the Legendre polynomials. We will meet them (more conceptually) later.

Exercise *Systems of orthogonal poynomials*
Work out the first few for
1. Tchebyshev I: $[-1,1]$, $w(x) = (1-x^2)^{-1/2}$
2. Tchebyshev II: $[-1,1]$, $w(x) = (1-x^2)^{+1/2}$
3. Laguerre: $[0,\infty)$, $w(x) = x^k e^{-x}$
4. Hermite: $(-\infty,\infty)$, $w(x) = e^{-x^2}$

For tables and much information, see Abramowitz-Stegun.

Remarks Orthogonal polynomials have many uses:

1. Special functions, special solutions to differential equations.

2. The general theory of orthogonal polynomials has proven to be of great utility in investigations of large N matrix integrals.

3. See B. Simon, *Orthogonal Polynomials on the Unit Circle*, Parts 1,2 for much more about orthogonal polynomials.

21.3 Polar decomposition

There is an analog for matrices of *polar decompositions*, generalizing the representation of complex numbers by phase and magnitude: $z = re^{i\theta}$. Here is the matrix analog:

♣This section should include the important case of complex ips and comment on the generalization to Hilbert space with partial isometries. ♣

Theorem Any matrix $A \in M_n(\mathbb{C})$ can be written as

$$A = UP \tag{21.14}$$

or

$$A = P'U \tag{21.15}$$

where P, P' are positive semidefinite and U is unitary. Moreover, the decomposition is unique if A is nonsingular.

Proof: The proof is a straightforward application of the singular value decomposition. Recall that we can write

$$A = U\Lambda V \tag{21.16}$$

where Λ is diagonal with nonnegative entries and U and V are unitary. Therefore we can write

$$A = (U\Lambda U^{-1}) \cdot (UV) = (UV) \cdot (V^{-1}\Lambda V) \tag{21.17}$$

Now note that both $(U\Lambda U^{-1})$ and $(V^{-1}\Lambda V)$ are positive semidefinite, and if A is nonsingular, positive definite. ♠

Remarks:

1. Taking the determinant recovers the polar decomposition of the determinant: $\det A = re^{i\theta}$ with $r = \det P$ and $e^{i\theta} = \det U$.
2. Note that $A^\dagger A = P^2$ so we could define P as the positive squareroot $P = \sqrt{A^\dagger A}$. This gives another approach to proving the theorem.

The version of the theorem over the real numbers is:

Theorem. Any invertible real $n \times n$ matrix A has a unique factorization as

$$A = PO \tag{21.18}$$

where P is a positive-definite symmetric matrix and O is orthogonal.

Proof: Consider AA^{tr}. This matrix is symmetric and defines a positive definite symmetric form. Since such forms can be diagonalized we know that there is a squareroot.

Let $P := (AA^{tr})^{1/2}$. There is a unique positive definite square root. Now check that $O := P^{-1}A$ is orthogonal. ♠

A related theorem is

Theorem. Any nonsingular matrix $A \in Mat_n(\mathbb{C})$ can be decomposed as:

$$A = SO \tag{21.19}$$

where S is complex symmetric and O is complex orthogonal.

Proof: Gantmacher, p.7.

Finally, we consider the generalization to operators on Hilbert space. Here there is an important new phenomenon. We can see it by considering the shift operator S on ℓ^2:

$$S : (x_1, x_2, x_3, \dots) \to (0, x_1, x_2, x_3, \dots) \tag{21.20}$$

Recall that $S^\dagger$ is the shift to the left, so $S^\dagger S = 1$, but $SS^\dagger$ is *not* one, rather it is $1 - |0\rangle\langle 0|$, in harmonic oscillator language.

Definition A *partial isometry* $V : \mathcal{H} \to \mathcal{H}$ is an operator so that

$$VV^\dagger V = V \tag{21.21}$$

The shift operator above is a good example of a partial isometry.

In general, note that $V^\dagger V = P_i$ and $VV^\dagger = P_f$ are both projection operators. Now we claim that

$$1 - V^\dagger V \tag{21.22}$$

is the orthogonal projection to $\ker(V)$. It is clear that if $\psi \in \ker(V)$ then $(1 - V^\dagger V)\psi = \psi$ and conversely $V(1 - V^\dagger V)\psi = 0$. Therefore, $V^\dagger V$ is the orthogonal projector to $(\ker V)^\perp$. Similarly, $VV^\dagger$ is the orthogonal projector to $\mathrm{im}(V)$.

$(\ker V)^\perp$ with orthogonal projector $V^\dagger V$ is called the *initial subspace*

$\mathrm{im}(V)$ with orthgonal projector $VV^\dagger$ is called the *final subspace.*

Note that V is an isometry when restricted to $V : (\ker V)^\perp \to \mathrm{im}(V)$, hence the name "partial isometry."

Theorem If T is a bounded operator on Hilbert space there is a partial isometry V so that

$$T = V\sqrt{T^\dagger T} = V|T| \tag{21.23}$$

and V is uniquely determined by $\ker V = \ker T$.

For the proof, see Reed-Simon Theorem VI.10 and for the unbounded operator version Theorem VIII.32. Note that for compact operators it follows from the singular value decomposition, just as in the finite dimensional case.

Remark: In string field theory and noncommutative field theory partial isometries play an important role. In SFT they are used to construct solutions to the string field equations. In noncommutative field theory they are used to construct "noncommutative solitons."

21.4 Reduced Echelon form

Theorem 17. Any matrix $A \in GL(n, \mathbb{C})$ can be factorized as

$$A = N\Pi B \tag{21.24}$$

where N is upper-triangular with $1's$ on the diagonal, Π is a permutation matrix, and B is upper-triangular.

Proof: See Carter, Segal, and MacDonald, *Lectures on Lie Groups and Lie Algebras*, p. 65.

Remarks

1. When we work over nonalgebraically closed fields we sometimes can only put matrices into *rational canonical form*. See Herstein, sec. 6.7 for this.

22. Families of Matrices

In many problems in mathematics and physics one considers continuous, differentiable, or holomorphic families of linear operators. When studying such families one is led to interesting geometrical constructions. In this section we illustrate a few of the phenomena which arise when considering linear algebra in families.

22.1 Families of projection operators: The theory of vector bundles

Let us consider two simple examples of families of projection operators.

Let $\theta \sim \theta + 2\pi$ be a coordinate on the circle. Fix $V = \mathbb{R}^2$, and consider the operator

$$\Gamma(\theta) = \cos\theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \sin\theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{22.1}$$

Note that $\Gamma^2 = 1$ and accordingly we can define two projection operators

$$P_{\pm}(\theta) = \frac{1}{2}(1 \pm \Gamma(\theta)) \tag{22.2}$$

Let us consider the eigenspaces as a function of θ. For each θ the image of $P_+(\theta)$ is a real line in $\mathbb{R}^2$. So, let us consider the set

$$\mathcal{L}_+ := \{(e^{i\theta}, v) | P_+(\theta)v = v\} \subset S^1 \times \mathbb{R}^2 \tag{22.3}$$

At fixed θ we can certainly choose a basis vector, i.e. an eigenvector, in the line given by the image of $P_+(\theta)$. What happens if we try to make a *continuous* choice of such a basis vector as a function of θ? The most natural choice would be the family of eigenvectors:

$$\begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \tag{22.4}$$ eq:Mobius-1

Now, θ is identified with $\theta + 2\pi$, and the projection operator only depends on θ modulo 2π. However, (22.4) is *not globally well-defined*! If we shift $\theta \to \theta + 2\pi$ then the eigenvector changes by a minus sign.

But we stress again that even though there is no globally well-defined continuous choice of eigenvector the real line given by the image of $P_+(\theta)$ is well-defined. For example, we can check:

$$\begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \mathbb{R} = - \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \mathbb{R} \subset \mathbb{R}^2 \tag{22.5}$$ eq:moblinone

The family of real lines over the circle define what is called a *real line bundle*. Another example of a real line bundle is $S^1 \times \mathbb{R}$ which is, topologically, the cylinder. However, our family is clearly different from the cylinder. One can prove that it is impossible to find a *continuous* choice of basis for all values of θ. Indeed, one can picture this real line bundle as the Mobius strip, which makes its topological nontriviality intuitively obvious.

Example 2 In close analogy to the previous example consider

$$\Gamma(\hat{x}) = \hat{x} \cdot \vec{\sigma} \tag{22.6}$$

for $\hat{x} \in S^2$, the unit sphere $\hat{x}^2 = 1$. Once again $(\Gamma(\hat{x}))^2 = 1$. Consider the projection operators

$$P_\pm(\hat{x}) = \frac{1}{2}(1 \pm \hat{x} \cdot \vec{\sigma}) \tag{22.7}$$

On $S^2 \times \mathbb{C}^2$ the eigenspaces of $P_\pm$ define a complex line for each point $\hat{x} \in S^2$. If we let $\mathcal{L}_+$ denote the total space of the line bundle, that is

$$\mathcal{L}_+ = \{(\hat{x}, v) | \hat{x} \cdot \vec{\sigma} v = +v\} \subset S^2 \times \mathbb{C}^2 \tag{22.8}$$

Then you can convince yourself that this is NOT the trivial complex line bundle $\mathcal{L}_+ \neq S^2 \times \mathbb{C}$. Using standard polar coordinates for the sphere, away from the south pole we can take the line to be spanned by

$$e_+ = \begin{pmatrix} \cos \frac{1}{2}\theta \\ e^{i\phi} \sin \frac{1}{2}\theta \end{pmatrix} \tag{22.9}$$

while away from the north pole the eigenline is spanned by

$$e_- = \begin{pmatrix} e^{-i\phi} \cos \frac{1}{2}\theta \\ \sin \frac{1}{2}\theta \end{pmatrix} \tag{22.10}$$

But note that, just as in our previous example, there is no continuous choice of nonzero vector spanning the eigenline for all points on the sphere.

What we are describing here is a nontrivial *line bundle* with transition function $e_+ = e^{i\phi}e_-$ on the sphere minus two points. This particular line bundle is called the Hopf line bundle, and of great importance in mathematical physics.

A generalization of this construction using the quaternions produces a nontrivial rank two complex vector bundle over S^4 known as the instanton bundle.

These two examples have a magnificent generalization to the theory of vector bundles. We will just summarize some facts. A full explanation would take a different Course.

Definition: A (complex or real) *vector bundle* E over a topological space X is the space of points

$$E := \{(x, v) : P(x)v = v\} \subset X \times \mathcal{H} \tag{22.11}$$

where $P(x)$ is a continuous family of orthogonal finite rank projection operators in a (complex or real) separable infinite-dimensional Hilbert space $\mathcal{H}$.

This is not the standard definition of a vector bundle, but it is equivalent to the usual definition. Note that there is an obvious map

$$\pi : E \to X \tag{22.12}$$

given by $\pi(x, v) = x$. The fibers of this map are vector spaces of dimension n:

$$E_x := \pi^{-1}(x) = \{(x, v) | v \in \mathrm{im}P(x)\} \tag{22.13}$$

carries a natural structure of a vector space:

$$\alpha(x, v) + \beta(x, w) := (x, \alpha v + \beta w) \tag{22.14}$$

so we are just doing linear algebra in families. We define a *section* of E to be a continuous map $s : X \to E$ such that $\pi(s(x)) = x$. It is also possible to talk about tensors on X with values in E. In particular $\Omega^1(X; E)$ denotes the sections of the space of 1-forms on X with values in E.

Note that, in our definition, a vector bundle is the same thing as a continuous family of projection operators on Hilbert space. So a vector bundle is the same thing as a *continuous* map

$$P : X \to Gr_n(\mathcal{H}) \tag{22.15}$$

where $Gr_n(\mathcal{H})$ is the Grassmannian of rank n projection operators in the norm topology.

Definition: Two vector bundles E_1, E_2 of rank n are *isomorphic* if there is a homotopy of the corresponding projection operators P_1, P_2. Therefore, the isomorphism classes of vector bundles is the same as the set of homotopy classes $[X, Gr_n(\mathcal{H})]$.

This viewpoint is also useful for defining connections. If $\psi : X \to \mathcal{H}$ is a continuous map into Hilbert space then

$$s(x) := (x, P(x)\psi(x) \tag{22.16}$$

is a section of E. Every section of E can be represented in this way. Then a *projected connection* on E is the map $\nabla^{\rm proj} : \Gamma(E) \to \Omega^1(X; E)$ given by $P \circ d \circ P$ where d is the exterior differential and we have written $\circ$ to emphasize that we are considering the composition of three operators. With local coordinates x^μ we can write:

$$\nabla^{\rm proj} : (x, s(x)) \to dx^\mu(x, P(x)\frac{\partial}{\partial x^\mu}(P(x)s(x))) \tag{22.17}$$

For those who know about curvature, the curvature of this connection is easily shown to be:

$$F = PdPdPP \tag{22.18}$$

and representatives of the characteristic classes $\mathrm{ch}_n(E)$ (thought of as DeRham cohomology classes) are then defined by the differential forms

$$\omega_n(E) = \frac{1}{n!(2\pi i)^n}\mathrm{Tr}(PdPdP)^n \tag{22.19}$$

Remarks

1. In physics the projected connection is often referred to as the *Berry connection*, and it is related to the quantum adiabatic theorem. The formula for the curvature is important in, for example, applications to condensed matter physics. In a typical application there is a family of Hamiltonians with a gap in the spectrum and P is the projector to eigenspaces below that gap. For example in the quantum Hall effect with P the projector to the lowest Landau level the first Chern class is given by

   $$\omega_1 = \frac{1}{2\pi i}\mathrm{Tr}PdPdP \tag{22.20}$$

   which turns out to be related to the Kubo formula, as first noted by TKNN.

2. A beautiful theorem of Narasimhan-Ramanan shows that *any* connection on a vector bundle can be regarded as the pull-back of a projected connection. See [47]

3.

Exercise *The Bott's projector and Dirac's monopole*
In a subsection below we will make use of the *Bott projector*:

$$P(z, \bar{z}) = \frac{1}{1 + |z|^2}\begin{pmatrix} 1 & \bar{z} \\ z & |z|^2 \end{pmatrix} \tag{22.21}$$

[47] D. Quillen, "Superconnection character forms and the Cayley transform," Topology, **27**, (1988) 211; J. Dupont, R. Hain, and S. Zucker, "Regulators and characteristic classes of flat bundles," arXiv:alg-geom/9202023.

a.) Check that this is a projection operator. [48]

b.) Show that the projector has a well-defined limit for $z \to \infty$.

c.) Show that under stereographic projection $S^2 \to \mathbb{R}^2 \cong \mathbb{C}$

$$z = \frac{x^1 + ix^2}{1 + x^3} \qquad \frac{1 + x^3}{2} = \frac{1}{1 + |z|^2} \tag{22.22}$$

we have

$$P_+(\hat{x}) = P(z, \bar{z}) \tag{22.23}$$

$$P_-(\hat{x}) = Q(z, \bar{z}) = \frac{1}{1 + |z|^2} \begin{pmatrix} |z|^2 & -\bar{z} \\ -z & 1 \end{pmatrix} \tag{22.24}$$

d.) Regarding the projector $P_+(\hat{x})$ as a projection operator on $\mathbb{R}^3 - \{0\}$, show that the projected connection is the same as the famous Dirac magnetic monopole connection by computing

♣Need to check these two equations ♣

$$\nabla^{\text{proj}} e_+ = \frac{1}{2}(1 - \cos\theta) d\phi e_+ \tag{22.25}$$

$$\nabla^{\text{proj}} e_- = -\frac{1}{2}(1 + \cos\theta) d\phi e_- \tag{22.26}$$

e.) Show that for the Bott projector [49]

$$\text{Tr}(PdPdP) = -\text{Tr}(QdQdQ) = -\frac{dzd\bar{z}}{(1 + |z|^2)^2} \tag{22.27}$$

and show that $\int_{\mathbb{C}} \text{ch}_1(\mathcal{L}_\pm) = \pm 1$.

Exercise

Show that any vector bundle E has a complementary bundle $E^\perp$ so that $E \oplus E^\perp = X \times \mathbb{C}^N$, where $\oplus$ is a family version of the direct sum of vector spaces.

22.2 Codimension of the space of coinciding eigenvalues

♣This sub-section is a little out of order because it uses group actions and homogeneous spaces which are only covered in Chapter 3. But it fits very naturally in this Section. ♣

We now consider more general families of linear operators $T(s) : V \to V$ where V is finite-dimensional and s is a set of parameters. Mathematically, we have a space $\mathcal{S}$ and a map $T : \mathcal{S} \to \text{End}(V)$ and we can put various conditions on that map: Continuous, differentiable, analytic,... Physically, $\mathcal{S}$ is often a set of "control parameters," for example coupling constants, masses, or other "external" parameters which can be varied.

We no longer assume the $T(s)$ are projection operators.

[48] This is obvious if you note that the second column is just $\bar{z}$ times the first column. But it is also good to do the matrix multiplication.

[49] *Hint*: In order to avoid a lot of algebra write $P = n\tilde{P}$ with $n = (1 + r^2)^{-1}$, $r^2 = |z|^2$ so that dn is proportional to dr^2.

22.2.1 Families of complex matrices: Codimension of coinciding characteristic values

A very important question that often arises is: What is the subset of points in $\mathcal{S}$ where T changes in some important way. Here is a sharp version of that question:

What is the set of points $\mathcal{S}_{\rm sing} \subset \mathcal{S}$ where characteristic values of $T(s)$ coincide?

In equations:

$$\mathcal{S}_{\rm sing} = \{s \in \mathcal{S} | p_{T(s)}(x) \quad \text{has multiple roots}\} \tag{22.28}$$

where $p_T(x) = \det(x1 - T)$ is the characteristic polynomial.

We can only give useful general rules for *generic* families. We first argue that it suffices to look at the space $\mathrm{End}(V) \cong M_n(\mathbb{C})$ itself. Let $\mathcal{D} \subset \mathrm{End}(V)$ be the sublocus where the characteristic polynomial has multiple eigenvalues:

$$\mathcal{D} := \{T | p_T(x) \quad \text{has multiple roots}\} \tag{22.29}$$

If $s \to T(s)$ is generically $1-1$ and $\mathcal{F} \subset \mathrm{End}(V)$ is the image of the family then

$$\mathcal{S}_{\rm sing} = T^{-1}(\mathcal{D} \cap \mathcal{F}) \tag{22.30}$$

Now, the codimension of $\mathcal{S}_{\rm sing}$ in $\mathcal{S}$ is the same as:

$$\mathrm{cod}(\mathcal{D} \cap \mathcal{F}) - \mathrm{cod}(\mathcal{F}) \tag{22.31}$$

See Figure 9.

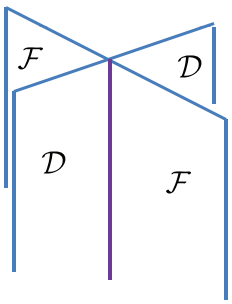


Figure 9: For generic families if there are d transverse directions to $\mathcal{D}$ in $M_n(\mathbb{C})$ and there are f transverse dimensions to $\mathcal{F}$ in $M_n(\mathbb{C})$ then there will be $d+f$ transverse dimensions to $\mathcal{D} \cap \mathcal{F}$ in $M_n(\mathbb{C})$ and d transverse dimensions to $\mathcal{S}_{\rm sing}$ in $\mathcal{S}$. fig:CODIMENSIO

On the other hand, for generic subspaces, the *codimensions* add for intersections:

$$\mathrm{cod}(\mathcal{D} \cap \mathcal{F}) = \mathrm{cod}(\mathcal{D}) + \mathrm{cod}(\mathcal{F}) \tag{22.32}$$

Therefore, if $s \to T(s)$ is a generic 1-1 family then the codimension of the set where characteristic values coincide in $\mathcal{S}$ is the same as the codimension of $\mathcal{D}$ in $M_n(\mathbb{C})$.

In general $\mathcal{D} \subset \mathrm{End}(V)$ can be exhibited as an algebraic variety. This follows from Exercises **** and **** at the end of chapter 3 where we show that the subspace is defined by the resultant of two polynomials in x:

$$\mathrm{Res}(p_T(x), p_T'(x)) = 0 \tag{22.33}$$

The resultant is a polynomial in the coefficients of $p_T(x)$ which in turn can be expressed in terms of polynomials in the matrix elements of T with respect to an ordered basis.

Example: If $n = 2$ then

$$\begin{aligned} \det(x1 - T) &= x^2 - \mathrm{Tr}(T)x + \det(T) \\ &= x^2 + a_1 x + a_0 \\ a_1 &= -\mathrm{Tr}(T) \\ a_0 &= \frac{1}{2}(\mathrm{Tr}(T))^2 - \mathrm{Tr}(T^2) \end{aligned} \tag{22.34}$$

The subspace in the space of all matrices where two characteristic roots coincide is clearly

$$a_1^2 - 4a_0 = 4\mathrm{Tr}(T^2) = (\mathrm{Tr}(T))^2 = 0 \tag{22.35}$$

which is an algebraic equation on the matrix elements.

The complex codimension of the solutions to one algebraic equation in $M_n(\mathbb{C})$ is one, and therefore we have the general rule:

The general element of a generic family of complex matrices will be diagonalizable, and the sublocus where at least two characteristic roots coincide will be real codimension two.

**** FIGURE OF $\mathcal{D}$ WITH TRANSVERSE SPACE IDENTIFIED WITH $\mathbb{C}$ ****

Note that it follows that, in a generic family the generic operator $T(s)$ will be diagonalizable, and will only fail to be diagonalizable on a complex codimension one subvariety of $\mathcal{S}$.

22.2.2 Orbits

It is interesting to view the above rule in a different way by counting dimensions of orbits. Using techniques discussed in chapters below the space $DIAG^*$ of diagonalizable matrices with distinct eigenvalues is fibered over

$$(\mathbb{C}^n - \Delta)/S_n \tag{22.36}$$

where Δ is the subspace where any two entries coincide. The fibration is the map of T to the unordered set of eigenvalues. The fiber is the set of matrices with a given set of eigenvalues and this is a homogeneous space, hence the fiber is

$$GL(n)/GL(1)^n \tag{22.37}$$

We can therefore compute the dimension of $DIAG^*$. The base is n-dimensional and the fiber is $n^2 - n$ dimensional so the total is n^2-dimensional, in accordance with the idea that $DIAG^*$ is the complement of a positive codimension subvariety.

However, let us now consider $DIAG$, the set of all *diagonalizable* matrices in $M_n(\mathbb{C})$, possibly with coinciding eigenvalues. Then of course $DIAG$ is still of full dimension n^2, but the subspace $\mathcal{D} \cap DIAG$ where two eigenvalues coincide is in fact complex codimension three!

Example: The generic 2×2 complex matrix is diagonalizable. However, the space of diagonalizable 2×2 matrices with coinciding eigenvalues is the space of matrices $\lambda 1$ and is one complex dimensional. Clearly this has codimension three!

More generally, if we consider the space of diagonalizable matrices and look at the subspace where two eigenvalues coincide, but all the others are distinct then we have a fibration over $(\mathbb{C}^{n-1} - \Delta)$ with fiber

$$GL(n)/(GL(2)^n \times GL(1)^{n-2}) \tag{22.38}$$

of dimension $n^2 - (n - 2 + 4) = n^2 - n - 2$. The base is of dimension $(n - 1)$ so the total space is of dimension $n^2 - 3$ and hence complex codimension 3 in $Mat_n(\mathbb{C})$.

The above discussion might seem puzzling since we also argued that the space of matrices where characteristic roots of the $p_A(x)$ coincide is complex codimension one, not three. Of course, the difference is accounted form by considering matrices with nontrivial Jordan form. The orbit of

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \tag{22.39}$$

is two-complex-dimensional. Together with the parameter λ this makes a complex codimension one subvariety of $M_n(\mathbb{C})$.

More generally

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \oplus_i \lambda_i 1 \tag{22.40}$$

with distinct λ_i has a stabilizer of dimension $2 + (n - 2) = n$ so the space of such matrices is $(n^2 - n) + (n - 2) + 1 = n^2 - 1$ dimensional.

22.2.3 Local model near $\mathcal{S}_{\rm sing}$

It is of some interest to make a good model of how matrices degenerate when approaching a *generic* point in $\mathcal{S}_{\rm sing}$.

Suppose $s \to s_* \in \mathcal{S}_{\rm sing}$ and two distinct roots of the characteristic polynomial say, $\lambda_1(s), \lambda_2(s)$, have a common limit $\lambda(s_*)$. We can choose a family of projection operators

$P(s)$ of rank 2, whose range is the two-dimensional subspace spanned by the eigenvectors $v_1(s)$ and $v_2(s)$ so that

$$T(s) = P_s t(s) P_s + Q_s T(s) Q_s \tag{22.41}$$

and such that $Q_s T(s) Q_s$ has a limit with distinct eigenvalues on the image of Q_s. The operator $t(s)$ is an operator on a two-dimensional subspace and we may choose some fixed generic ordered basis and write:

$$t(s) = \lambda(s)1 + \begin{pmatrix} z(s) & x(s) - iy(s) \\ x(s) + iy(s) & -z(s) \end{pmatrix} \tag{22.42}$$

where λ, x, y, z are all *complex*. $\lambda(s)$ is some smooth function going to $\lambda(s_*)$ while

$$x(s_*)^2 + y(s_*)^2 + z(s_*)^2 = 0 \tag{22.43}$$

is some generic point on the nilpotent cone of 2×2 nilpotent matrices. That generic point will have $z(s_*) \neq 0$ and hence $x(s_*) \pm iy(s_*) \neq 0$.

***** FIGURE OF A DOUBLE-CONE WITH A PATH ENDING ON THE CONE AT A POINT s_* NOT AT THE TIP OF THE CONE ****

Therefore we can consider the smooth matrix

$$S(s) = \begin{pmatrix} z(s) & 1 \\ x(s) + iy(s) & 0 \end{pmatrix} \tag{22.44}$$

which will be invertible in some neighborhood of s_*. Now a small computation shows that

$$S^{-1} t(s) S = \lambda(s)1 + \begin{pmatrix} 0 & 1 \\ w & 0 \end{pmatrix} \tag{22.45}$$

$$w = x^2 + y^2 + z^2 \tag{22.46}$$

Therefore we can take w to be a coordinate in the normal bundle to $\mathcal{D} \subset M_2(\mathbb{C})$ and the generic family of degenerating matrices has (generically) a nonsingular family of bases where the operator can be modeled as

$$T(w) = P \begin{pmatrix} 0 & 1 \\ w & 0 \end{pmatrix} P + T^{\perp} \qquad w \in \mathbb{C} \tag{22.47}$$

22.2.4 Families of Hermitian operators

If we impose further conditions then the rule for the codimension can again change.

An important example arises if we consider families of *Hermitian* matrices. In this case, the codimension of the subvariety where two eigenvalues coincide is *real* codimension 3.

Let us prove this for a family of 2×2 matrices. Our family of matrices is

$$\begin{pmatrix} d_1(s) & z(s) \\ \bar{z}(s) & d_2(s) \end{pmatrix} \tag{22.48}$$

where d_1, d_2 are real and z is complex.

The eigenvalues coincide when the discriminant of the characteristic polynomial vanishes. This is the condition

$$b^2 - 4ac = (d_1 + d_2)^2 - 4(d_1 d_2 - |z|^2) = (d_1 - d_2)^2 + 4|z|^2 = 0 \tag{22.49}$$

Thus, $d_1 = d_2$ and $z = 0$ is the subvariety. Moreover, in the neighborhood of this locus the family is modeled on

$$\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix} + \vec{x} \cdot \vec{\sigma} \tag{22.50}$$

For the general case the subspace of Hermitian matrices where exactly two eigenvalues coincide and are otherwise distinct is a fibration over

$$(\mathbb{R}^{(n-2)} - \Delta)/S_{n-2} \times \mathbb{R} \tag{22.51}$$

(the fibration being given by the map to the unordered set of eigenvalues) with fiber:

$$U(n)/(U(2) \times U(1)^{n-2}) \tag{22.52}$$

♣Again, this discussion is out of place. We have not discussed quotients yet. ♣

The fiber has real dimension $n^2 - (4 + (n-2)) = n^2 - n - 2$ and the base has real dimension $n - 1$ so the total dimension is $(n^2 - n - 2) + (n - 1) = n^2 - 3$. So the codimension is 3.

**** FIGURE OF $\mathcal{D}$ AS LINE WITH TRANSVERSE SPACE A PLANE, BUT LABELED AS $\mathbb{R}^3$ *****

Near to the level crossing the universal form of the matrix is

$$T(\vec{x}) = P_{\vec{x}} \left((\lambda + \vec{a} \cdot \vec{x}) 1_{2\times 2} + \vec{x} \cdot \vec{\sigma} + \mathcal{O}(x^2) \right) P_{\vec{x}} + T^{\perp}(\vec{x}) \tag{22.53}$$

where $\vec{x}$ is a local parameter normal to the codimension three subspace, $P_{\vec{x}}$ is a smooth family of rank 2 projectors with a smooth limit at $\vec{x} = 0$, $\lambda \in \mathbb{R}$, and $\vec{a} \in \mathbb{R}^3$.

Example In solid state physics the energy levels in bands cross at *points* in the Brillouin zone. (See Chapter 4 below.)

Exercise

Show that the subspace of Hermitian matrices where exactly k eigenvalues coincide and are otherwise distinct is of real codimension $k^2 - 1$ in the space of all Hermitian matrices.

The way in which these subspaces fit together is quite intricate. See

V. Arnold, "Remarks on Eigenvalues and Eigenvectors of Hermitian Matrices. Berry Phase, Adiabatic Connections and Quantum Hall Effect," Selecta Mathematica, Vol. 1, No. 1 (1995)

References: For some further discussion see Ref: von Neumann + Wigner, Avron-Seiler, Ann. Phys. 110(1978)85; B. Simon, "Holonomy, the Quantum Adiabatic Theorem, and Berry's Phase," Phys. Rev. Lett. **51**(1983)2167

22.3 Canonical form of a family in a first order neighborhood

In this section we consider families in a slightly different way: Can we put families into canonical form by conjugation?

Suppose we have a family of complex matrices $T(s)$ varying continuously with some control parameter $s \in S$ and we allow ourselves to make similarity transformations $g(s)$,

$$T(s) \to g(s)T(s)g(s)^{-1} \tag{22.54}$$

where $g(s)$ must vary continuously with s.

We know that if we work at a *fixed* value of s then we can put $T(s)$ into Jordan canonical form. We can ask – can we put an arbitrary family of matrices into some canonical form?

For example, if $T(s_0)$ is diagonal for some s_0, can we choose $g(s)$ for s near s_0 so that $T(s)$ is diagonal in the neighborhood of s_0?

This is a hard question in general. Let us consider what we can say to first order in perturbations around s_0. For simplicity, suppose S is a neighborhood around 0 in the complex plane, so s_0 is zero. Write

$$T(s) = T_0 + s\delta M + \mathcal{O}(s^2) \tag{22.55}$$

and let us assume that T_0 has been put into some canonical form, such as Jordan canonical form. Now write $g(s) = 1 + s\delta g + \mathcal{O}(s^2)$. Then

$$g(s)T(s)g(s)^{-1} = T_0 + s\delta M + s[\delta g, T_0] + \mathcal{O}(s^2) \tag{22.56}$$

eq:firstpert

For any matrix $m \in M_n(\mathbb{C})$ let us introduce the operator

$$Ad(m) : M_n(\mathbb{C}) \to M_n(\mathbb{C}) \tag{22.57}$$

$$Ad(m)(X) := [m, X] \tag{22.58}$$

What we learn from (22.56) is that we can "conjugate away" anything in the image of $Ad(T_0)$.

That is, *the space of nontrivial perturbations is the cokernel of the operator* $Ad(T_0)$.

Let us suppose that $T_0 = Diag\{\lambda_1, \dots, \lambda_n\}$ is diagonal with *distinct* eigenvalues. Then if $\delta g \in M_n(\mathbb{C})$

$$(\mathrm{Ad}(T_0)\delta g)_{ij} = (\lambda_i - \lambda_j)\delta g_{ij} \tag{22.59}$$

Thus, the range is the space of off-diagonal matrices. We can always conjugate away any off-diagonal element of $(\delta M)_{ij}$, but we cannot conjugate away the diagonal elements. The cokernel can be represented by the diagonal matrices.

Thus, near a matrix with distinct eigenvalues we can, at least to first order in perturbation theory, take the perturbed matrix to be diagonal.

If some eigenvalues coincide then we might not be able to conjugate away some off-diagonal elements.

Moreover, as we have seen above, it is perfectly possible to have a family of matrices degenerate from diagonalizable to nontrivial Jordan form.

Exercise

Compute the cokernel of the operator

$$Ad(J_\lambda^{(k)}) : M_k(\mathbb{C}) \to M_k(\mathbb{C}) \tag{22.60}$$

and find a natural subspace complementary to $\mathrm{im}(Ad(J_\lambda^{(k)}))$ in $M_k(\mathbb{C})$.

Answer: As we showed in equation (10.55) above the kernel of $Ad(J_\lambda^{(k)})$ consists of matrices of the form

$$A = a_1\mathbf{1_k} + a_2(e_{1,2} + e_{2,3} + \cdots + e_{k-1,k}) + a_3(e_{1,3} + e_{2,4} + \cdots + e_{k-2,k}) + \cdots + a_k e_{1,k} \tag{22.61}$$

That is, it is a general polynomial in $J_\lambda^{(k)}$. It is therefore a k-dimensional space. By the index, we know that the cokernel is therefore k-dimensional. Therefore, there is a k-dimensional space of nontrivial perturbations.

Example: By direct computation we find for $k = 2$ the general perturbation is equivalent to

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \delta_1 & \delta_2 \end{pmatrix} \tag{22.62}$$

and in general the matrices

$$J_\lambda^{(k)} + \sum_{i=1}^{k} \delta_i e_{ki} \tag{22.63}$$

for $\delta_1, \ldots, \delta_k$ free parameters give a set of representatives of the cokernel.

b.) Show that

$$\det\left(J_\lambda^{(k)} + \sum_{i=1}^{k} \delta_i e_{ki}\right) = \lambda^k + \delta_k\lambda^{k-1} - \delta_{k-1}\lambda^{k-2} \pm \cdots + (-1)^{k-1}\delta_1 \tag{22.64}$$

See R. Gilmore, *Catastrophe Theory*, ch. 14 for further details. There are many applications of the above result.

22.4 Families of operators and spectral covers

In this subsection we describe the spectral cover construction which allows us to translate the data of a family of operators into a purely geometrical object.

Thus, suppose we have a generic family $T(s)$ of complex $n \times n$ matrices over $s \in \mathcal{S}$. As we saw above, generically $T(s)$ is regular semisimple so we let

$$\mathcal{S}^* = \mathcal{S} - \mathcal{S}_{\mathrm{sing}} \tag{22.65}$$

and for $s \in \mathcal{S}^*$ we have

$$T(s) = \sum_{i=1}^{n} \lambda_i(s) P_{\lambda_i}(s) \tag{22.66}$$

where $P_{\lambda_i}(s)$ are orthogonal projection operators and the $\lambda_i(s)$ are *distinct*. Note that the sum on the RHS makes sense without any choice of ordering of the λ_i.

Now consider the space

$$\widetilde{\mathcal{S}} := \{(s,\lambda) | p_{T(s)}(\lambda) = 0\} \subset \mathcal{S} \times \mathbb{C} \tag{22.67}$$

That is, the fiber of the map $\widetilde{\mathcal{S}} \to \mathcal{S}$ is the space of λ's which are characteristic values of $T(s)$ at $s \in \mathcal{S}$.

Over $\mathcal{S}^*$ all the eigenvalues are distinct so

$$\widetilde{\mathcal{S}}^* \to \mathcal{S}^* \tag{22.68}$$

is a smooth n-fold cover. In general since $\pi_1(\mathcal{S}^*) \neq 0$ it will be a nontrivial cover, meaning that there is only locally, but not globally a well-defined ordering of the eigenvalues $\lambda_1(s), \dots, \lambda_n(s)$. In general only the unordered set $\{\lambda_1(s), \dots, \lambda_n(s)\}$ is well-defined over $\mathcal{S}^*$. See equation (22.75) et. seq. below for the case $n = 2$.

Now, note that there is a well-defined map to the space of rank one projectors:

$$P : \widetilde{\mathcal{S}}^* \to Gr_1(\mathbb{C}^n) \tag{22.69}$$

Namely, if $(s,\lambda) \in \widetilde{\mathcal{S}}^*$ then λ is an eigenvalue of $T(s)$ and $P(s,\lambda)$ is the projector to the eigenline $L_\lambda \subset V = \mathbb{C}^n$ spanned by the eigenvector with eigenvalue λ.

In general, a vector bundle whose fibers form a dimension one vector space is called a *line bundle*. Moreover, if we look at the fibers over the different sheets of the covering space $\widetilde{\mathcal{S}}^* \to \mathcal{S}^*$ we get an s-dependent decomposition of V into a sum of lines:

$$V = \oplus_{i=1}^n L_{\lambda_i(s)} \tag{22.70}$$

Now the RHS does depend on the ordering of the $\lambda_i(s)$, but up to isomorphism it does not depend on an ordering. The situation for $n = 2$ is sketched schematically in Figure 10

Therefore, there is a 1-1 correspondence between the family of operators $T(s)$ parametrized by $s \in \mathcal{S}^*$ and complex line bundles over the n-fold covering space $\widetilde{\mathcal{S}}^* \to \mathcal{S}^*$.

Now, let us ask what happens when we try to consider the full family over $\mathcal{S}$. Certainly $\widetilde{\mathcal{S}} \to \mathcal{S}$ still makes sense, but now, over $\mathcal{S}_{\text{sing}}$ some characteristic values will coincide and the covering is an n-fold *branched covering*.

Let us recall the meaning of the term *branched covering*. In general, a branched covering is a map of pairs $\pi : (Y, R) \to (X, B)$ where $R \subset Y$ and $B \subset X$ are of real codimension two. R is called the *ramification locus* and B is called the *branch locus*. The map $\pi : Y - R \to X - B$ is a regular covering and if it is an n-fold covering we say the branched covering is an n-fold branched cover. On the other hand, near any point $b \in B$ there is a neighborhood U of b and local coordinates

$$(x_1, \dots, x_{d-2}; w) \in \mathbb{R}^{d-2} \times \mathbb{C}, \tag{22.71}$$

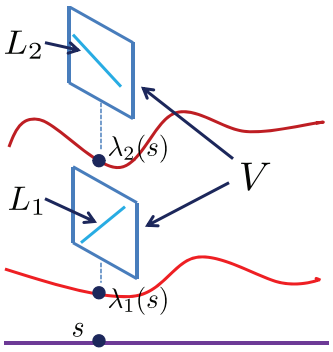


Figure 10: We consider a family of two-by-two matrices $T(s)$ with two distinct eigenvalues $\lambda_1(s)$ and $\lambda_2(s)$. These define a two-fold covering of $\mathcal{S}$, the spectral cover. Moreover, the eigenlines associated to the two sheets give a decomposition of $V = \mathbb{C}^2$ into a sum of lines which varies with s.

fig:TWOEIGENLI

where $\dim_{\mathbb{R}} X = d$ and $w = 0$ describes the branch locus $B \cap U$. More importantly, $\pi^{-1}(U) = \amalg_\alpha \widetilde{U}_\alpha$ is a disjoint union of neighborhoods in Y of points $r_\alpha \in R$ with local coordinates

$$(x_1, \ldots, x_{d-2}; \xi_\alpha) \in \mathbb{R}^{d-2} \times \mathbb{C} \tag{22.72}$$

so that the map $\pi_\alpha : \widetilde{U}_\alpha \to U$ (where π_α is just the restriction of π) is just given by

$$\pi_\alpha : (x_1, \ldots, x_{d-2}; \xi_\alpha) \to (x_1, \ldots, x_{d-2}; \xi_\alpha^{e_\alpha}) \tag{22.73}$$

where e_α are positive integers called *ramification indices*.

In plain English: For any $b \in B$ there are several points r_α in the preimage of π above b and near any r_α the map π looks like a mapping of unit disks in the complex plane $\xi \to w = \xi^e$. Note that for an n-fold covering

$$\sum_\alpha e_\alpha = n \tag{22.74}$$

The case where exactly one ramification index is $e = 2$ and all the others are equal to one is called a *simple branch point*.

♣Put in figure of disks covering disks. ♣

Now, we have argued that the interesting part of $T(s)$ near a generic point $s_* \in \mathcal{S}_{\text{sing}}$ is of the form

$$T(w) = \begin{pmatrix} 0 & 1 \\ w & 0 \end{pmatrix} \tag{22.75}$$

eq:SimpleDegFa

For this family the characteristic polynomials are clearly

$$p_{T(w)}(x) = x^2 - w \tag{22.76}$$

and the set of characteristic roots is the unordered set $\{+\sqrt{w}, -\sqrt{w}\}$. This is just the unordered set $\{+\xi, -\xi\}$ labeling the two sheets of the 2-fold branched cover. By taking appropriate real slices we may picture the situation as in Figure 11.

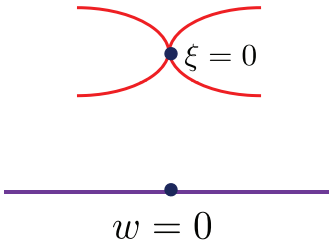


Figure 11: A particular real representation of a complex 2-fold branched cover of a disk over a disk. The horizontal axis in the base is the real line of the complex w plane. The verticle axis corresponds to the real axis of the complex ξ plane on the right and the purely imaginary axis of the complex ξ plane on the left. fig:BRANCHCOVE

The monodromy of the eigenvalues is nicely illustrated by considering the closed path

$$w(t) = w_0 e^{2\pi i t} \qquad 0 \leq t \leq 1 \tag{22.77}$$ eq:closedcurve

If we choose at $t = 0$ a particular squareroot ξ_0 of w_0 then this closed path lifts to

$$\xi(t) = \xi_0 e^{\pi i t} \qquad 0 \leq t \leq 1 \tag{22.78}$$ eq:liftedcurve

and the value at $t = 1$ is the *other* squareroot $-\xi_0$ of w_0.

Now let us consider the eigenlines at w. If $w \neq 0$ (i.e. $s \in \mathcal{S}^*$) then there are two distinct eigenlines which are the span of the eigenvectors

$$v_\pm = \begin{pmatrix} 1 \\ \pm\xi \end{pmatrix} \tag{22.79}$$

The eigenlines are just the images of the two Bott projectors:

$$P_\epsilon(\xi) = \frac{1}{1+|\xi|^2} \begin{pmatrix} 1 & \epsilon\xi^* \\ \epsilon\xi & |\xi|^2 \end{pmatrix} \tag{22.80}$$

where $\epsilon = \pm 1$. Along the lifted curve (22.78) P_+ evolves into P_- and so the two eigenlines get exchanged under monodromy.

Note well that $P_\epsilon(\xi)$ has a nice smooth limit as $\xi \to 0$. Therefore the two eigenlines smoothly degenerate to a single line over the single point $\xi = 0$.

In the holomorphic case one can use the fact that holomorphic functions of many variables cannot have complex codimension two singularities (Hartog's theorem) to conclude that:

Therefore, at least for generic holomorphic families over a complex space $\mathcal{S}$ there is a 1-1 correspondence between the families of $n \times n$ matrices $T(s)$ and holomorphic line bundles over n-fold branched covers $\widetilde{\mathcal{S}} \to \mathcal{S}$.

Remarks

1. *Higgs bundles.* The method of spectral covers is very important in certain aspects of Yang-Mills theory. In particular, a version of the Yang-Mills equations on $C \times \mathbb{R}^2$, where C is a Riemann surface lead to differential equations on a connection on C known as the *Hitchin equations*. It turns out that solutions to these are equivalent to holomorphic families of operators on C with the technical difference that the operators $T(s)$, with $s \in C$, (known as *Higgs fields*) are valued in $(1,0)$-forms on C. In this case the spectral cover technique converts a difficult analytic problem such as the solution of Yang-Mills equations into a purely geometrical problem: Describing holomorphic line bundles on n-fold coverings in T^*C.

Exercise

Describe the spectral cover for the family

$$T(z) = \begin{pmatrix} 0 & 1 \\ z^2 - E & 0 \end{pmatrix} \tag{22.81}$$

over the complex plane $z \in \mathbb{C}$.

Exercise

Suppose that $T(z)$ is a two-dimensional holomorphic matrix. Show that the spectral cover is a hyperelliptic curve and write an equation for it in terms of the matrix elements of $T(z)$.

Exercise *Hitchin map*

The characteristic polynomial defines a natural map $h : \mathrm{End}(V) \to \mathbb{C}^n$, known in this context as the Hitchin map:

$$\begin{aligned} h : T &\mapsto (a_1(T), \ldots, a_n(T)) \\ \det(x\mathbf{1_n} - T) &= x^n + a_1 x^{n-1} + \cdots + a_n \end{aligned} \tag{22.82}$$

eq:hitchin

Define the universal spectral cover over $\mathbb{C}^n$ to be

$$\widetilde{\mathbb{C}^n} := \{(a_1, \ldots, a_n; t) | t^n + a_1 t^{n-1} + \cdots + a_n = 0\} \tag{22.83}$$

eq:univers

Show that $\bar{S}$ is a fiber product of S with $\widetilde{\mathbb{C}^n}$. [50]

22.5 Families of matrices and differential equations

Another way families of matrices arise is in the theory of differential equations.

As motivation, let us consider the Schrödinger equation:

$$\left(-\hbar^2 \frac{d^2}{dx^2} + V(x)\right)\psi(x) = E\psi(x) \tag{22.84}$$

where we have rescaled V and E by $2m$. We keep $\hbar$ because we are going to discuss the WKB approximation. We could scale it to one and then some other (dimensionless!) physical parameter must play the role of a small parameter.

We can write this as the equation

$$\hbar^2 \frac{d^2}{dx^2}\psi(x) = v(x)\psi(x) \tag{22.85}$$

with $v(x) = V(x) - E$. Moreover, if $v(x)$ is the restriction to the real line of a meromorphic function $v(z)$ on the complex plane we can write an ODE on the complex plane:

$$\hbar^2 \partial_z^2 \psi = v(z)\psi \tag{22.86}$$

This can be converted to a 2×2 matrix equation:

$$\hbar \frac{\partial}{\partial z}\psi = \begin{pmatrix} 0 & 1 \\ v(z) & 0 \end{pmatrix}\psi \tag{22.87}$$

eq:mtrx-schrod

where

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{22.88}$$

is a column vector.

Given this reformulation it is natural to consider more generally the matrix differential equation

$$\hbar \frac{\partial}{\partial z}\psi = A(z)\psi \tag{22.89}$$

eq:norderf

where $A(z)$ is some meromorphic matrix function of z.

For example, the Airy differential equation corresponds to

$$A(z) = \begin{pmatrix} 0 & 1 \\ z & 0 \end{pmatrix} \tag{22.90}$$

while the harmonic oscillator corresponds to

$$A(z) = \begin{pmatrix} 0 & 1 \\ z^2 - E & 0 \end{pmatrix} \tag{22.91}$$

eq:HO-MATRIX

[50] *Answer*: T defines a map $p_T : \mathcal{S} \to \mathbb{C}^n$ by taking the coefficients of the characteristic polynomial. You can then pull back the covering $\widetilde{\mathbb{C}^n} \to \mathbb{C}^n$ along T to get $\widetilde{\mathcal{S}}$.

For the theory we are developing it is interesting to generalize further and let $A(z)$ be a meromorphic $n \times n$ matrix. This subsumes the theory of nth order linear ODE's, but the more general setting is important in applications and leads to greater flexibility.

Now if $A(z)$ is nonsingular in a simply connected region $\mathcal{R}$ then there is an n-dimensional space of solutions of (22.89) in $\mathcal{R}$. If $\psi^{(1)}, \ldots, \psi^{(n)}$ are n linearly independent solutions they can all be put together to make an $n \times n$ matrix solution

$$\Psi = \begin{pmatrix} \psi^{(1)} & \cdots & \psi^{(n)} \end{pmatrix} = \begin{pmatrix} \psi_1^{(1)} & \cdots & \psi_1^{(n)} \\ \vdots & & \vdots \\ \psi_n^{(1)} & \cdots & \psi_n^{(n)} \end{pmatrix} \tag{22.92}$$

The solutions $\psi^{(i)}$ are linearly independent, iff Ψ is invertible.

The best way to think about solutions of linear ODE's is in fact to look for invertible matrix solutions to

$$\hbar \frac{\partial}{\partial z} \Psi = A(z) \Psi \tag{22.93}$$

eq:bestway

Note that if C is a constant matrix, then $\Psi \to \Psi C$ is another solution. We can think of this freedom as corresponding to the choices of initial conditions specifying the independent solutions $\psi^{(i)}$.

What happens if we multiply from the left by a constant matrix? The equation is not preserved in general since in general C will not commute with $A(z)$. This is not a bug, but a feature, and indeed it is useful to consider more generally making a redefinition

$$\Psi(z) = g(z) \tilde{\Psi}(z) \tag{22.94}$$

where $g(z)$ is an invertible $n \times n$ matrix function of z. Then we change the equation to:

$$\hbar \frac{\partial}{\partial z} \tilde{\Psi} = A^g(z) \tilde{\Psi} \tag{22.95}$$

where

$$A^g = g^{-1} A g - \hbar g^{-1} \partial_z g \tag{22.96}$$

is the "gauge-transform" of A.

Thus we learn, that when working with families of matrices, the proper notion of equivalence might not be that of conjugation, but of gauge transformation. Which is the proper notion depends on the problem under consideration.

From this point of view, solving (22.93) can be interpreted as finding the gauge transformation Ψ that gauges A to 0.

What can we say about the existence of solutions? Let us think of z as valued on the complex plane. Locally, if $A(z)$ does not have singularities then we can always solve the

equation with the path ordered exponential. We choose a path $z(t)$ starting from $z_{\rm in}$

$$\begin{aligned}\Psi(z(t)) &= \mathrm{Pexp}\left[\frac{1}{\hbar}\int_{z_{\rm in}}^{z(t)} A(w)dw\right]\Psi(z_{\rm in}) \\ &:= \left(1+\sum_{n=1}^{\infty}\hbar^{-n}\int_{z_{\rm in}}^{z} dz_1\int_{z_{\rm in}}^{z_1} dz_2\cdots\int_{z_{\rm in}}^{z_{n-1}} dz_n A(z_1)\cdots A(z_n)\right)\Psi(z_{\rm in}) \\ &:= \left(1+\sum_{n=1}^{\infty}\frac{1}{\hbar^n n!}\int_0^t dt_1\frac{dz(t_1)}{dt_1}\int_0^t dt_2\frac{dz(t_2)}{dt_2}\cdots\int_0^t dt_n\frac{dz(t_n)}{dt_n}P[A(z(t_1))\cdots A(z(t_n))]\right)\Psi(z_{\rm in})\end{aligned} \tag{22.97}$$

eq:Pexp

where all the integrations are along the path: $\int dz... = \int dt\frac{dz}{dt}...$ and $P[...]$ is the time-ordered product with later times on the left.

If $A(z)$ is meromorphic then the path ordered exponential exists provided the path $z(t)$ does not go through any singularities. This is clear from the third line of (22.97) since $A(z(t))$ will be bounded on the interval $[0,t]$. The second line makes clear that $\Psi(z)$ is locally holomorphic. However, the solution *does* depend on the homotopy class of the path in $\mathbb{C}$ minus the singularities. We will return to this below.

22.5.1 The WKB expansion

Sometimes the path-ordered exponential is prohibitively difficult to evaluate and we have to make due with approximate solutions. One way to get such approximate solutions is the WKB method. Note that if we could gauge transform $A(z)$ to be *diagonal* then the equation is easily solved, because in the 1×1 case:

$$\hbar\partial_z\psi = a(z)\psi \Rightarrow \psi(z) = \exp\left(\frac{1}{\hbar}\int_{z_0}^{z} a(w)dw\right)\psi(z_0) \tag{22.98}$$

This observation motivates us to find a gauge transformation which implements the diagonalization order by order in $\hbar$. To this end we introduce the ansatz:

$$\Psi_{wkb} = S(z,\hbar)e^{\frac{1}{\hbar}\Delta(z,\hbar)} \tag{22.99}$$

eq:WKB-ansatz

where $\Delta(z,\hbar)$ is *diagonal*, and we assume there are series expansions

$$S(z,\hbar) = S_0(z) + \hbar S_1(z) + \hbar^2 S_2(z) + \cdots \tag{22.100}$$

$$\Delta(z,\hbar) = \Delta_0(z) + \hbar\Delta_1(z) + \hbar^2\Delta_2(z) + \cdots \tag{22.101}$$

Now, to determine these series we substitute (22.99) into equation (22.93) and bring the resulting equation to the form

$$(A(z) - \hbar S' S^{-1})S = S\Delta' \tag{22.102}$$

eq:wkbi

where $\Delta' = \frac{\partial\Delta}{\partial z}$ and $S' = \frac{\partial S}{\partial z}$. Now we look at this equation order by order in $\hbar$.

At zeroth order we get

$$A(z)S_0 = S_0\Delta_0' \tag{22.103}$$

Thus, $S_0(z)$ must diagonalize $A(z)$ and Δ_0' is the diagonal matrix of eigenvalues of $A(z)$. Of course, $A(z)$ might fail to be diagonalizable at certain places. We will return to this.

Now write (22.102) as

$$A(z)S - \hbar S' = S\Delta' \tag{22.104}$$ eq:wkbii

Now substitute $A(z) = S_0\Delta_0' S_0^{-1}$. Next, make a choice of diagonalizing matrix S_0 and multiply the equation on the left by S_0^{-1} to get

$$\Delta_0' S_0^{-1} S - \hbar S_0^{-1} S' = S_0^{-1} S \Delta' \tag{22.105}$$ eq:wkbiii

Equation (22.105) is the best form in which to substitute the series in $\hbar$. At order $\hbar^n$, $n > 0$ we get

$$\Delta_0' S_0^{-1} S_n - S_0^{-1} S_{n-1}' = \sum_{i=0}^{n} S_0^{-1} S_i \Delta_{n-i}' \tag{22.106}$$ eq:wkbiv

separating out the $i = 0$ and $i = n$ terms from the RHS the equation is easily rearranged to give

$$\begin{aligned} &[\Delta_0', S_0^{-1}S_1] - \Delta_1' = S_0^{-1}S_0' \\ &[\Delta_0', S_0^{-1}S_n] - \Delta_n' = S_0^{-1}S_{n-1}' + \sum_{i=1}^{n-1} S_0^{-1} S_i \Delta_{n-i}' \qquad\qquad n > 1 \end{aligned} \tag{22.107}$$ eq:wkbivp

In an inductive procedure, every term on the RHS is known. On the left-hand side Δ_n' is *diagonal*, so we take it to be the diagonal component of the RHS.

As long as the eigenvalues of Δ_0' are *distinct* $Ad(\Delta_0')$ maps ONTO the space of off-diagonal matrices, and hence we can solve for $S_0^{-1}S_n$. In this way we generate the WKB series.

Remarks

1. The WKB procedure will fail when $A(z)$ has nontrivial Jordan form. This happens when the characteristic polynomial has multiple roots. These are the branch points of the Riemann surface defined by the characteristic equation.

2. Indeed, returning to the matrix $A(z)$ corresponding to the Schrödinger equation (22.87). The characteristic equation $\lambda^2 - v(z) = 0$ has branchpoints at zeroes $v(z_0)$. Recalling that $v(x) = V(x) - E$ these are just the turning points of the usual mechanics problem. In the neighborhood of such points one can write an exact solution in terms of Airy functions, and then match to the WKB solution to produce a good approximate solution.

3. Note that - so long as $A(z)$ is diagonalizable with distinct eigenvalues - the above procedure only determines $S_0(z)$ up to right-multiplication by a diagonal matrix $D_0(z)$. However, the choice of diagonal matrix then enters in the equation determining Δ_1 and S_1: $\Delta_1' = -Diag(S_0^{-1}S_0')$. Thus, if we change our choice

$$S_0(z) \to \tilde{S}_0(z) = S_0(z)D_0(z) \tag{22.108}$$

then we have:

$$\tilde{\Delta}_1(z) = -\mathrm{Diag}(\tilde{S}_0^{-1}\tilde{S}_0') = -\mathrm{Diag}(S_0^{-1}S_0') - \mathrm{Diag}(D_0^{-1}D_0') \tag{22.109}$$

and when substituting into (22.99) the change of S_0 is canceled by the change of Δ_1. Similarly, S_n is only determined up to the addition of a matrix of the form S_0D_n where D_n is an arbitrary diagonal matrix function of z. However, at the next stage in the procedure D_n will affect Δ_{n+1} in such a way that the full series $S\exp[\frac{1}{\hbar}\Delta]$ is unchanged.

4. In general, the WKB series is only an *asymptotic series*.

Exercise
Write the general n^{th} order linear ODE in matrix form.

Exercise
Show that for the matrix

$$A(z) = \begin{pmatrix} 0 & 1 \\ v(z) & 0 \end{pmatrix} \tag{22.110}$$

We can must have $\Delta_0'(z) = \sqrt{v(z)}\sigma^3$ and we can choose S_0 to be

$$S_0 = \begin{pmatrix} v^{-1/4} & v^{-1/4} \\ v^{1/4} & -v^{1/4} \end{pmatrix} \tag{22.111}$$

Show that with this choice of $S_0(z)$ we have

$$[\Delta_0', S_0^{-1}S_1] - \Delta_1' = -\frac{1}{4}\sigma^1\frac{d}{dz}\log v(z) \tag{22.112}$$

and therefore

$$\Delta(z,\hbar) = \sigma_3\int_{z_0}^{z}\sqrt{v(z')}dz' + \mathcal{O}(\hbar^2) \tag{22.113}$$

22.5.2 Monodromy representation and Hilbert's 21st problem

Consider again the matrix equation (22.93).

If $A(z)$ is holomorphic near z_0 then so is the solution $\Psi(z)$. On the other hand, there will be interesting behavior when $A(z)$ has singularities.

Definition A *regular singular point* z_* is a point where $A(z)$ has Laurent expansion of the form

$$A(z) = \frac{A_{-1}}{z - z_*} + \cdots \tag{22.114}$$

with A_{-1} regular semisimple.

We have

Theorem [Fuchs' theorem]: Near a RSP there exist convergent series solutions in a disk around $z = z_*$. and if $A_{-1} = \mathrm{Diag}\{\lambda_1, \ldots, \lambda_n\}$ then

$$\Psi(z) = \mathrm{Diag}\{(z - z_*)^{\lambda_1}, \ldots, (z - z_*)^{\lambda_n}\} \left(1 + \Psi_1(z - z_*) + \Psi_2(z - z_*)^2 + \cdots\right) \tag{22.115}$$

where $\Psi_1, \Psi_2, \ldots$ are constant matrices, is a *convergent series* in some neighborhood of z_*

Note that in general the solution will have monodromy around $z = z_*$: Analytic continuation around a counterclockwise oriented simple closed curve around $z = 0$ gives

$$\Psi(z_* + (z - z_*)e^{2\pi i}) = \mathrm{Diag}\{e^{2\pi i\lambda_1}, \ldots, e^{2\pi i\lambda_n}\}\Psi(z) \tag{22.116}$$

If $A(z)$ only has regular singular points at, say, $p_1, \ldots, p_s$ then analytic continuation defines a *representation*

$$\rho : \pi_1(\mathbb{C} - \{p_1, \ldots, p_s\}, z_0) \to GL(n, \mathbb{C}) \tag{22.117}$$

eq:MonRep

known as the *monodromy representation of the differential equation.*

Remark: Riemann was the first to investigate this problem, completely solving the case of $n = 2$ with three regular singular points. In his famous address to the International Congress of Mathematicians in Paris in 1900 D. Hilbert presented a list of 23 problems for the mathematics of the 20th century. The 21st problem was, roughly, in modern terms:

Given an irreducible n-dimensional representation (22.117)*, find a differential equation for which it is a monodromy representation.*

This problem has a complicated history, with claimed solutions and counterexamples. We note that there is a very physical approach to the problem using free fermion conformal field theory correlation functions which was pursued by the Kyoto school of Miwa, Jimbo, et. al. It is also the first example of what is called the "Riemann-Hilbert correspondence," which plays an important role in algebraic geometry.

♣Explain more clearly if the problem has been solved or not! ♣

22.5.3 Stokes' phenomenon

A subject closely related to the WKB analysis is Stokes' phenomenon. We give a brief account here.

Definition An *irregular singular point* is a singular point of the form

$$A(z) = \frac{A_{-n}}{z^n} + \cdots \tag{22.118}$$

with $n > 1$.

Let us consider the simplest kind of ISP, which we put at $z = 0$:

$$A(z) = \frac{R}{z^2} + \frac{A_{-1}}{z} + \cdots \tag{22.119}$$

with $R = \mathrm{Diag}\{r_1, \ldots, r_n\}$. Then the series method will produce a *formal* solution

$$\Psi_f = \left(1 + \Psi_1 z + \Psi_2 z^2 + \cdots\right) e^{-R/z} \tag{22.120}$$

The big difference will now be that the series is only *asymptotic* for $z \to 0$.

♣HERE GOES DEFINITION OF ASYMPTOTIC SERIES. STRESS DEPENDENCE ON CHOICE OF RAY. ♣

Example: Consider

$$\frac{d}{dz}\Psi = \left(\frac{r\sigma^3}{z^2} + \frac{s\sigma^1}{z}\right)\Psi \tag{22.121}$$

Substituting

$$\Psi_f = U e^{-r\sigma^3/z} \tag{22.122}$$

with $U = 1 + zU_1 + z^2 U_2 + \cdots$ we find

$$\frac{dU}{dz} = \mathrm{Ad}(\frac{r\sigma^3}{z^2})U + \frac{s\sigma^1}{z}U \tag{22.123}$$

Writing this out we get the equations

$$\begin{aligned} [r\sigma^3, U_1] &= -s\sigma^1 \\ [\sigma^3, U_{n+1}] &= \frac{(n - s\sigma^1)}{r} U_n \end{aligned} \tag{22.124}$$

and the factor of n on the RHS in the second line shows that coefficients in U_n are going to grow like $n!$ and hence the series will only be asymptotic.

Indeed, in this case the formal series is easily shown to be

$$U_n = -\left(\frac{-1}{2r}\right)^n \frac{\prod_{j=1}^{n-1}(j^2 - s^2)}{n!}(s + n\sigma^1)(\sigma^3)^n \tag{22.125}$$

eq:FormSerEx

so the prefactor grows like $n!$.

Definition The rays $(r_i - r_j)\mathbb{R}_+$ starting at $z = 0$ are known as *Stokes rays* and the open regions between these rays are *Stokes sectors*. See Figure 12.

Now one can prove

Theorem: Let ρ be a ray which is *not* a Stokes ray, and let $\mathbb{H}_\rho$ be the half-plane containing ρ as in Figure 13. Then there is a unique solution Φ_ρ which is asymptotic to the formal solution as $z \to 0$ along any ray in $\mathbb{H}_\rho$:

$$\Phi_\rho e^{R/z} \to 1 \tag{22.126}$$

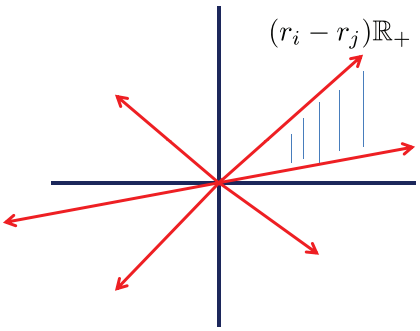


Figure 12: Stokes sectors fig:STOKESECTO

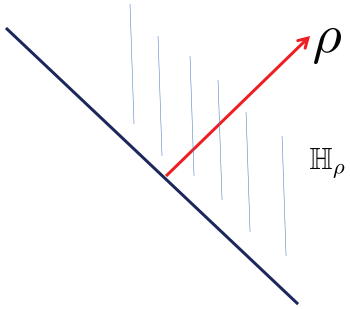


Figure 13: There is a true solution to the differential equation asymptotic to the formal solution in the half-plane $\mathbb{H}_\rho$. fig:HALFPLANE

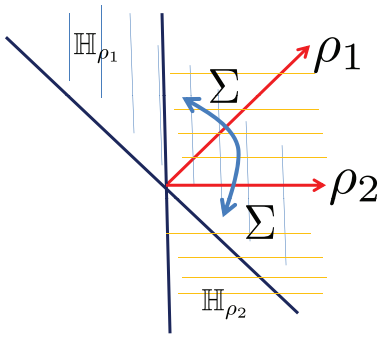


Figure 14: There is a true solution to the differential equation asymptotic to the formal solution in the half-plane $\mathbb{H}_\rho$. fig:HPOVERLAP

where $z \to 0$ along any ray in $\mathbb{H}_\rho$, and hence

$$\lim_{z\to 0} z^{-n} \left(\Phi_\rho(z) e^{R/z} - (1 + z\Psi_1 + \cdots + z^n \Psi_n) \right) = 0 \tag{22.127}$$

It is very important here that the limit is taken along a ray in $\mathbb{H}_\rho$, otherwise the statement will be false. Indeed, in general one *cannot* find a formal solution in a *larger* domain which is asymptotic to the formal solution! This is one version of *Stokes phenomenon*.

Now, consider two rays ρ_1, ρ_2, neither of which is a Stokes ray. The half-planes overlap as in Figure 14. Then, by uniqueness of the solution to the differential equation we know that

$$\Phi_{\rho_1} = \Phi_{\rho_2} S_\Sigma \qquad \text{on} \qquad \mathbb{H}_{\rho_1} \cap \mathbb{H}_{\rho_2} \tag{22.128}$$

where S_Σ is a matrix which is constant as a function of z. (It might well depend on other parameters in the differential equation.) Moreover, $S_\Sigma = 1$ if there is no Stokes' ray in Σ, but $S_\Sigma \neq 1$ if there are Stokes' rays in Σ. If there is precisely one Stokes ray ℓ in Σ then we set $S_\Sigma = S_\ell$ and call S_ℓ the *Stokes factor* for ℓ.

♣That sentence is really a theorem. ♣

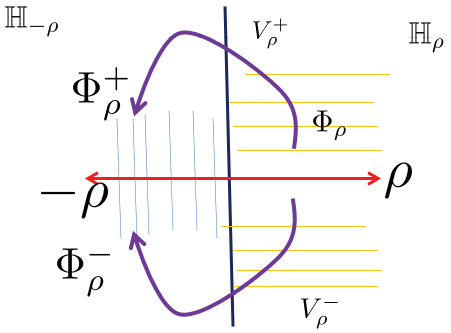


Figure 15: There is a true solution to the differential equation asymptotic to the formal solution in the half-plane $\mathbb{H}_\rho$.

fig:TWOAC

Now we can describe an analog of monodromy for ISP's: Choose rays $\pm\rho$ which are not Stokes rays. Starting with Φ_ρ in $\mathbb{H}_\rho$ there are two analytic continuations to $\mathbb{H}_{-\rho}$, as shown in Figure 15. Call these two analytic continuations $\Phi_\rho^\pm$ then we have:

Theorem:

$$\Phi_\rho^+ = \Phi_{-\rho} S_+ \qquad \text{in} \qquad \mathbb{H}_{-\rho} \tag{22.129}$$

$$\Phi_\rho^- = \Phi_{-\rho} S_- \qquad \text{in} \qquad \mathbb{H}_{-\rho} \tag{22.130}$$

are given by

$$S_+ =: \prod_{\ell \in V_+(\rho)}^{ccw} S_\ell : \tag{22.131}$$

$$S_- =: \prod_{\ell \in V_-(\rho)}^{cw} S_\ell : \tag{22.132}$$

where the products are ordered so successive rays are counterclockwise or clockwise. These are called Stokes matrices, and serve as the analogs of monodromy matrices in the irregular singular point case.

Remarks

1. There is a generalization of this story to higher order poles. If

$$A(z) = \frac{R}{z^{\ell+1}} + \cdots \tag{22.133}$$

with R regular semisimple then the formal series solution has the form

$$\Psi_f = U(z)e^{Q(z)} \tag{22.134}$$

$$U(z) = 1 + zU_1 + z^2U_2 + \cdots \tag{22.135}$$

$$Q(z) = \frac{Q_\ell}{z^\ell} + \cdots + \frac{Q_1}{z} + Q_0 \log z \tag{22.136}$$

Moreover, a true solution asymptotic to the formal solution will generally only exist in angular sectors of angular width $|\Delta\theta| < \frac{\pi}{\ell}$. For more details about this see Coddington and Levinson, or Hille's book on ODEs.

2. There is a great deal more to be said about the kind of groups the Stokes matrices live in and their use in parametrizing flat connections and their applications to Yang-Mills theory.

Exercise

a.) Derive the formal series (22.125).

b.) Show that the equation can be reduced to a second order ODE with one irregular singular point and one regular singular point.

Answer:

a.) Write

$$U_n = w_n 1 + x_n\sigma^1 + y_n\sigma^2 + z_n\sigma^3 \tag{22.137}$$

so that

$$\begin{aligned} 2ix_{n+1}\sigma^2 - 2iy_{n+1}\sigma^1 = &\frac{n}{r}(x_n\sigma^1 + y_n\sigma^2 + z_n\sigma^3) \\ &- \frac{s}{r}(x_n + iy_n\sigma^3 - iz_n\sigma^2) \end{aligned} \tag{22.138}$$

Deduce that $nw_n = sx_n$ and $nz_n = isy_n$ and

$$x_{n+1} = \frac{1}{2ir}\frac{n^2 - s^2}{n}y_n \qquad y_{n+1} = -\frac{1}{2ir}\frac{n^2 - s^2}{n}x_n \tag{22.139}$$

The induction starts with

$$U_1 = \frac{s}{2ir}\sigma^2 + \frac{s^2}{2r}\sigma^3 \tag{22.140}$$

so $x_{2n+1} = y_{2n} = 0$. The rest is simple induction.

b.) Write out the differential equation on a two-component column vector and eliminate ψ_2 to get:

$$\psi'' + \frac{2ir - sz}{z(ir - sz)}\psi' - \frac{r^2 + s^2z^2}{z^4}\psi = 0 \tag{22.141}$$

♣Transform this to the confluent hg equation. ♣

Exercise *Harmonic Oscillator*

Consider the matrix (22.91) corresponding to the harmonic oscillator differential equation. Using the known asymptotics of the parabolic cylinder functions work out the Stokes sectors and Stokes factors for this equation.

Exercise *Three singular points*

Suppose that a second order differential equation on the extended complex plane has three singular points with monodromy given by regular semisimple elements conjugate to

$$\begin{pmatrix} \mu_i & 0 \\ 0 & \mu_i^{-1} \end{pmatrix} \qquad i = 1, 2, 3 \tag{22.142}$$

Show that for generic $\mu_i \in \mathbb{C}$ the equation $M_1 M_2 M_3 = 1$ can be solved, up to simultaneous conjugation of the M_i, in terms of the μ_i, and write M_2, M_3 in a basis where M_1 is diagonal.

In fancy mathematical terms: The moduli space of flat $SL(2,\mathbb{C})$ connections with fixed conjugacy class of regular semisimple monodromy around three points on $\mathbb{C}P^1 - \{p_1, p_2, p_3\}$ has no moduli.

♣YOU SHOULD GIVE AN EXAMPLE OF HOW THIS GIVES REFLECTION/TRANSMISSIO COEFFICIENTS IN SOME SCATTERING PROBLEMS. ♣

---

23. $\mathbb{Z}_2$-graded, or super-, linear algebra

rLinearAlgebra

In this section "super" is merely a synonym for "$\mathbb{Z}_2$-graded." Super linear algebra is extremely useful in studying supersymmetry and supersymmetric quantum theories, but its applications are much broader than that and the name is thus a little unfortunate.

Superlinear algebra is very similar to linear algebra, but there are some crucial differences, which we highlight in this section. *It's all about signs.*

We are going to be a little bit pendantic and long-winded in this section because the subject is apt to cause confusion.

23.1 Super vector spaces

It is often useful to add the structure of a $\mathbb{Z}_2$-*grading* to a vector space. A $\mathbb{Z}_2$-graded vector space over a field κ is a vector space over κ which, moreover, is written as a direct sum

$$V = V^0 \oplus V^1. \tag{23.1}$$ eq:zeet

The vector spaces V^0, V^1 are called the even and the odd subspaces, respectively. We may think of these as eigenspaces of a "parity operator" P_V which satisfies $P_V^2 = 1$ and is $+1$ on V^0 and -1 on V^1. If V^0 and V^1 are finite dimensional, of dimensions m, n respectively we say the super-vector space has graded-dimension or superdimension $(m|n)$.

A vector $v \in V$ is called *homogeneous* if it is an eigenvector of P_V. If $v \in V^0$ it is called *even* and if $v \in V^1$ it is called *odd*. We may define a *degree* or *parity* of homogeneous vectors by setting $\deg(v) = \bar{0}$ if v is even and $\deg(v) = \bar{1}$ if v is odd. Here we regard $\bar{0}, \bar{1}$ in the additive abelian group $\mathbb{Z}/2\mathbb{Z} = \{\bar{0}, \bar{1}\}$. Note that if v, v' are homogeneous vectors of the same degree then

$$\deg(\alpha v + \beta v') = \deg(v) = \deg(v') \tag{23.2}$$ eq:dnge

for all $\alpha, \beta \in \kappa$. We can also say that $P_V v = (-1)^{\deg(v)} v$ acting on homogeneous vectors. For brevity we will also use the notation $|v| := \deg(v)$. Note that $\deg(v)$ is *not defined* for general vectors in V.

Mathematicians define the *category of super vector spaces* so that a *morphism* from $V \to W$ is a linear transformation which *preserves grading*. We will denote the space of morphisms from V to W by $\underline{\mathrm{Hom}}(V, W)$. These are just the ungraded linear transformations of ungraded vector spaces, $T : V \to W$, which commute with the parity operator $TP_V = P_W T$.

So far, there is no big difference from, say, a $\mathbb{Z}$-graded vector space. However, important differences arise when we consider *tensor products*.

So far we defined a category of supervector spaces, and now we will make it into a *tensor category*. (See definition below.)

The tensor product of two $\mathbb{Z}_2$ graded spaces V and W is $V \otimes W$ as vector spaces over κ, but the $\mathbb{Z}_2$-grading is defined by the rule:

$$\begin{aligned} (V \otimes W)^0 &:= V^0 \otimes W^0 \oplus V^1 \otimes W^1 \\ (V \otimes W)^1 &:= V^1 \otimes W^0 \oplus V^0 \otimes W^1 \end{aligned} \tag{23.3}$$ eq:tsnp

Thus, under tensor product the degree is additive on homogeneous vectors:

$$\deg(v \otimes w) = \deg(v) + \deg(w) \tag{23.4}$$ eq:tnesvct

If κ is any field we let $\kappa^{p|q}$ denote the supervector space:

$$\kappa^{p|q} = \underbrace{\kappa^p}_{\text{even}} \oplus \underbrace{\kappa^q}_{\text{odd}} \tag{23.5}$$

Thus, for examples:

$$\mathbb{R}^{n_e|n_o} \otimes \mathbb{R}^{n'_e|n'_o} \cong \mathbb{R}^{n_e n'_e + n_o n'_o | n_e n'_o + n_o n'_e} \tag{23.6}$$

and in particular:

$$\mathbb{R}^{1|1} \otimes \mathbb{R}^{1|1} = \mathbb{R}^{2|2} \tag{23.7}$$

$$\mathbb{R}^{2|2} \otimes \mathbb{R}^{2|2} = \mathbb{R}^{8|8} \tag{23.8}$$

$$\mathbb{R}^{8|8} \otimes \mathbb{R}^{8|8} = \mathbb{R}^{128|128} \tag{23.9}$$

Now, in fact we have a *braided tensor category*:

In ordinary linear algebra there is an isomorphism of tensor products

$$c_{V,W} : V \otimes W \to W \otimes V \tag{23.10}$$ eq:BrdIso

given by $c_{V,W} : v \otimes w \mapsto w \otimes v$. In the super-commutative world there is also an isomorphism (23.10) defined by taking

$$c_{V,W} : v \otimes w \to (-1)^{|v|\cdot|w|} w \otimes v \tag{23.11}$$ eq:SuperBraid

on homogeneous objects, and extending by linearity.

Let us pause to make two remarks:

1. Note that in (23.11) we are now viewing $\mathbb{Z}/2\mathbb{Z}$ as a ring, not just as an abelian group. Do not confuse $\deg v + \deg w$ with $\deg v \deg w$! In computer science language $\deg v + \deg w$ corresponds to *XOR*, while $\deg v \deg w$ corresponds to *AND*.

2. It is useful to make a general rule: In equations where the degree appears it is understood that all quantities are homogeneous. Then we extend the formula to general elements by linearity. Equation (23.11) is our first example of a general rule: In the supercommutative world, commuting any object of homogeneous degree A with an object of homogeneous degree B results in an "extra" sign $(-1)^{AB}$. This is sometimes called the "*Koszul sign rule*."

With this rule the tensor product of a collection $\{V_i\}_{i\in I}$ of supervectorspaces

$$V_{i_1} \otimes V_{i_2} \otimes \cdots \otimes V_{i_n} \tag{23.12}$$ eq:TensSupVect

of supervector spaces is well-defined and *independent of the ordering of the factors*. This is a slightly nontrivial fact. See the remarks below.

We define the $\mathbb{Z}_2$-graded-symmetric and $\mathbb{Z}_2$-graded-antisymmetric products to be the images of the projection operators

$$P = \frac{1}{2}\left(1 \pm c_{V,V}\right) \tag{23.13}$$

Therefore the $\mathbb{Z}_2$-graded-symmetric product of a supervector space is the $\mathbb{Z}_2$-graded vector space with components:

$$\begin{aligned} S^2(V)^0 &\cong S^2(V^0) \oplus \Lambda^2(V^1) \\ S^2(V)^1 &\cong V^0 \otimes V^1 \end{aligned} \tag{23.14}$$

and the $\mathbb{Z}_2$-graded-antisymmetric product is

$$\begin{aligned} \Lambda^2(V)^0 &\cong \Lambda^2(V^0) \oplus S^2(V^1) \\ \Lambda^2(V)^1 &\cong V^0 \otimes V^1 \end{aligned} \tag{23.15}$$

Remarks

1. In this section we are stressing the differences between superlinear algebra and ordinary linear algebra. These differences are due to important signs. If the characteristic of the field κ is 2 then ± 1 are the same. Therefore, in the remainder of this section we assume κ is a field of characteristic different from 2.

2. Since the transformation $c_{V,W}$ is nontrivial in the $\mathbb{Z}_2$-graded case the fact that (23.12) is well-defined is actually slightly nontrivial. To see the issue consider the tensor product $V_1 \otimes V_2 \otimes V_3$ of three super vector spaces. Recall the relation $(12)(23)(12) = (23)(12)(23)$ of the symmetric group. Therefore, we should have "coherent" isomorphisms:

$$(c_{V_2,V_3} \otimes 1)(1 \otimes c_{V_1,V_3})(c_{V_1,V_2} \otimes 1) = (1 \otimes c_{V_1,V_2})(c_{V_1,V_3} \otimes 1)(1 \otimes c_{V_2,V_3}) \qquad (23.16)$$ eq:z2yb

and this is easily checked.

In general a *tensor category* is a category with a bifunctor $\mathcal{C} \times \mathcal{C} \to \mathcal{C}$ denoted $(X, Y) \to X \otimes Y$ with an associativity isomorphism $F_{X,Y,Z} : (X \otimes Y) \otimes Z \cong X \otimes (Y \otimes Z)$ satisfying the pentagon coherence relation. A braiding is an isomorphism $c_{X,Y} : X \otimes Y \to Y \otimes X$. The associativity and braiding isomorphisms must satisfy "coherence equations." The category of supervector spaces is perhaps the simplest example of a braided tensor category going beyond the category of vector spaces.

3. Note well that $S^2(V)$ as a supervector space does *not* even have the same dimension as $S^2(V)$ in the ungraded sense! Moreover, if V has a nonzero odd-dimensional summand then $\Lambda^n(V)$ does not vanish no matter how large n is.

4. With this notion of symmetric product we can nicely unify the bosonic and fermionic Fock spaces. If we have a system with bosonic and fermionic oscillators then there is a natural supervector space V spanned by the bosonic and fermionic creation operators, where the bosonic oscillators are even and the fermionic oscillators are odd. Then the $\mathbb{Z}_2$-graded, or super-Fock space $S^\bullet(V)$ naturally gives the full Fock space of the free boson-fermion system. That is, we have the isomorphism of ungraded vector spaces:

$$\underbrace{S^\bullet V}_{\text{graded symmetrization}} = \underbrace{S^\bullet V^0 \otimes \Lambda^\bullet V^1}_{\text{ungraded tensor product of vector spaces}} \qquad (23.17)$$

Exercise
a.) Show that $c_{V,W} c_{W,V} = 1$.
b.) Check (23.16).

Exercise *Reversal of parity*

a.) Introduce an operation which switches the parity of a supervector space: $(\Pi V)^0 = V^1$ and $(\Pi V)^1 = V^0$. Show that Π defines a functor of the category of supervector spaces to itself which squares to one.

b.) In the category of finite-dimensional supervector spaces when are V and ΠV isomorphic? [51]

c.) Show that one can identify ΠV as the functor defined by tensoring V with the canonical odd one-dimensional vector space $\kappa^{0|1}$.

23.2 Linear transformations between supervector spaces

If the ground field κ is taken to have degree 0 then the dual space $V^\vee$ in the category of supervector spaces consists of the morphisms $V \to \kappa^{1|0}$. Note that $V^\vee$ inherits a natural $\mathbb{Z}_2$ grading:

$$\begin{aligned} (V^\vee)^0 &:= (V^0)^\vee \\ (V^\vee)^1 &:= (V^1)^\vee \end{aligned} \tag{23.18}$$

eq:duals

Thus, we can say that $(V^\vee)^\epsilon$ are the linear functionals $V \to \kappa$ which vanish on $V^{1+\epsilon}$.

Taking our cue from the natural isomorphism in the ungraded theory:

$$\mathrm{Hom}(V, W) \cong V^\vee \otimes W \tag{23.19}$$

we use the same definition so that the space of linear transformations between two $\mathbb{Z}_2$-graded spaces becomes $\mathbb{Z}_2$ graded. We also write $\mathrm{End}(V) = \mathrm{Hom}(V, V)$.

In particular, a linear transformation is an even linear transformation between two $\mathbb{Z}_2$-graded spaces iff $T : V^0 \to W^0$ and $V^1 \to W^1$, and it is odd iff $T : V^0 \to W^1$ and $V^1 \to W^0$. Put differently:

$$\begin{aligned} \mathrm{Hom}(V, W)^0 &\cong \mathrm{Hom}(V^0, W^0) \oplus \mathrm{Hom}(V^1, W^1) \\ \mathrm{Hom}(V, W)^1 &\cong \mathrm{Hom}(V^0, W^1) \oplus \mathrm{Hom}(V^1, W^0) \end{aligned} \tag{23.20}$$

The general linear transformation is neither even nor odd.

If we choose a basis for V made of vectors of homogeneous degree and order it so that the even degree vectors come first then with respect to such a basis even transformations have block diagonal form

$$T = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \tag{23.21}$$

eq:matev

while odd transformations have block diagonal form

[51] *Answer*: An isomorphism is a degree-preserving isomorphism of vector spaces. Therefore if V has graded dimension $(m|n)$ then ΠV has graded dimension $(n|m)$ so they are isomorphic in the category of supervector spaces iff $n = m$.

$$T = \begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix} \tag{23.22}$$ eq:matevp

Remarks

1. Note well! There is a difference between $\mathrm{Hom}(V,W)$ and $\underline{\mathrm{Hom}}(V,W)$. The latter is the space of morphisms from V to W in the category of supervector spaces. They consist of just the even linear transformations: [52]

$$\underline{\mathrm{Hom}}(V,W) = \mathrm{Hom}(V,W)^0 \tag{23.23}$$

2. If $T : V \to W$ and $T' : V' \to W'$ are linear operators on super-vector-spaces then we can define the $\mathbb{Z}_2$ graded tensor product $T \otimes T'$. Note that $\deg(T \otimes T') = \deg(T) + \deg(T')$, and on homogeneous vectors we have

$$(T \otimes T')(v \otimes v') = (-1)^{\deg(T')\deg(v)} T(v) \otimes T'(v') \tag{23.24}$$ eq:tensortmsn

As in the ungraded case, $\mathrm{End}(V)$ is a ring, but now it is a $\mathbb{Z}_2$-graded ring under composition: $T_1 T_2 := T_1 \circ T_2$. That is if $T_1, T_2 \in \mathrm{End}(V)$ are homogeneous then $\deg(T_1 T_2) = \deg(T_1) + \deg(T_2)$, as one can easily check using the above block matrices. These operators are said to *graded-commute*, or *supercommute* if

$$T_1 T_2 = (-1)^{\deg T_1 \deg T_2} T_2 T_1 \tag{23.25}$$ eq:KRule

Remark: Now what shall we take for the definition of $GL(\kappa^{p|q})$? This should be the group of automorphisms of the object $\kappa^{p|q}$ in the category of super-vector-spaces. These must be even invertible maps and so

$$GL(\kappa^{p|q}) \cong GL(p;\kappa) \times GL(q;\kappa). \tag{23.26}$$ eq:kpq-iso

Some readers will be puzzled by equation (23.26). The Lie algebra of this group (see Chapter 8) is

$$\mathfrak{gl}(p;\kappa) \oplus \mathfrak{gl}(q;\kappa) \tag{23.27}$$

and is *not* the standard super Lie algebra $\mathfrak{gl}(p|q;\kappa)$. (See Chapter 12).

Another indication that there is something funny going on is that the naive definition of $GL(\kappa^{p|q})$, namely that it is the subset of $\mathrm{End}(\kappa^{p|q})$ of invertible linear transformations, will run into problems with (23.25). For example consider $\kappa^{1|1}$. Then, choosing a basis (say 1) for κ we get a basis of homogeneous vectors on $\kappa^{1|1}$. Then the operator

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{23.28}$$

[52] Warning! Some authors use the opposite convention for distinguishing hom in the category of supervector spaces from "internal hom."

is an odd element with $T^2 = 1$. On the other hand, we might have expected T to super-commute with itself, but then the sign rule (23.25) implies [53] that if it super-commutes with itself then $T^2 = 0$, but this is not the case.

We will define a more general group with the correct super Lie algebra, but to do so we need to discuss the notion of supermodules over a superalgebra.

Exercise

Show that if $T : V \to W$ is a linear transformation between two super-vector spaces then

a.) T is even iff $TP_V = P_W T$

b.) T is odd iff $TP_V = -P_W T$.

23.3 Superalgebras

The set of linear transformations $\mathrm{End}(V)$ of a supervector space is an example of a superalgebra. In general we have:

Definition

a.) A *superalgebra* $\mathcal{A}$ is a supervector space over a field κ together with a morphism

$$\mathcal{A} \otimes \mathcal{A} \to \mathcal{A} \tag{23.29}$$

of supervector spaces. We denote the product as $a \otimes a' \mapsto aa'$. Note this implies that

$$\deg(aa') = \deg(a) + \deg(a'). \tag{23.30}$$

We assume our superalgebras to be unital so there is a $1_{\mathcal{A}}$ with $1_{\mathcal{A}} a = a 1_{\mathcal{A}} = a$. Henceforth we simply write 1 for $1_{\mathcal{A}}$.

b.) The superalgebra is *associative* if $(aa')a'' = a(a'a'')$.

c.) Two elements a, a' in a superalgebra are said to *graded-commute*, or *super-commute* provided

$$aa' = (-1)^{|a||a'|} a'a \tag{23.31}$$

If every pair of elements a, a' in a superalgebra graded-commmute then the superalgebra is called *graded-commutative* or *supercommutative*.

d.) The *supercenter*, or $\mathbb{Z}_2$*-graded center* of an algebra, denoted $Z_s(\mathcal{A})$, is the subsuperalgebra of $\mathcal{A}$ such that all homogeneous elements $a \in Z_s(\mathcal{A})$ satisfy

$$ab = (-1)^{|a||b|} ba \tag{23.32}$$

for all homogeneous $b \in \mathcal{A}$.

Example 1: *Matrix superalgebras.* If V is a supervector space then $\mathrm{End}(V)$ as described above is a matrix superalgebra. One can show that the supercenter is isomorphic to κ, consisting of the transformations $v \to \alpha v$, for $\alpha \in \kappa$.

[53] so long as the characteristic of κ is not equal to two

Example 2: *Grassmann algebras.* The Grassmann algebra of an ordinary vector space W is just the exterior algebra of W considered as a $\mathbb{Z}_2$-graded algebra. We will denote it as Grass$[W]$.

In plain English, we take vectors in W to be odd and use them to generate a superalgebra with the rule that

$$w_1 w_2 + w_2 w_1 = 0 \tag{23.33}$$

for all w_1, w_2. In particular (provided the characteristic of κ is not two) we have $w^2 = 0$ for all w.

Thus, if we choose basis vectors $\theta^1, \dots, \theta^n$ for W then we can view Grass(W) as the quotient of the supercommutative polynomial superalgebra $\kappa[\theta^1, \dots, \theta^n]/I$ where the relations in I are:

$$\theta^i \theta^j + \theta^j \theta^i = 0 \qquad\qquad (\theta^i)^2 = 0 \tag{23.34}$$

The typical element then is

$$a = x + x_i \theta^i + \frac{1}{2!} x_{ij} \theta^i \theta^j + \cdots + \frac{1}{n!} x_{i_1,\dots,i_n} \theta^{i_1} \cdots \theta^{i_n} \tag{23.35}$$

The coefficients $x_{i_1,\dots,i_m}$ are m^{th}-rank totally antisymmetric tensors in $\kappa^{\otimes m}$.

We will sometimes also use the notation Grass$[\theta^1, \dots, \theta^n]$.

Definition Let $\mathcal{A}$ and $\mathcal{B}$ be two superalgebras. The *graded tensor product* $\mathcal{A} \widehat{\otimes} \mathcal{B}$ is the superalgebra which is the graded tensor product as a vector space and the multiplication of homogeneous elements satisfies

$$(a_1 \otimes b_1) \cdot (a_2 \otimes b_2) = (-1)^{|b_1||a_2|} (a_1 a_2) \otimes (b_1 b_2) \tag{23.36}$$

eq:GradedTensc

Remarks

1. Every $\mathbb{Z}_2$-graded algebra is also an ungraded algebra: We just forget the grading. However this can lead to some confusions:

2. An algebra can be $\mathbb{Z}_2$-graded-commutative and not ungraded-commutative: The Grassmann algebras are an example of that. We can also have algebras which are ungraded commutative but not $\mathbb{Z}_2$-graded commutative. The Clifford algebras $C\ell_{\pm 1}$ described below provide examples of that.

3. The $\mathbb{Z}_2$-graded-center of an algebra can be different from the center of an algebra as an ungraded algebra. Again, the Clifford algebras $C\ell_{\pm 1}$ described below provide examples.

4. One implication of (23.36) is that when writing matrix representations of graded algebras we do *not* get a matrix representation of the graded tensor product just by taking the tensor product of the matrix representations.

Example 3: *The real Clifford algebras $C\ell_{r+,s-}$.* Clifford algebras are defined for a general quadratic form Q on a vector space V over κ. We will study the Clifford algebras extensively in Chapter 10(??). Nevertheless, a few comments here nicely illustrate some important general points. If we take the case of a real vector space $\mathbb{R}^d$ with quadratic form

$$Q = \begin{pmatrix} +1_r & 0 \\ 0 & -1_s \end{pmatrix} \tag{23.37}$$

Then we get the real Clifford algebras $C\ell_{r+,s-}$. They can also be defined as the $\mathbb{Z}_2$ graded algebra over $\mathbb{R}$ generated by *odd* elements e_i with relations

$$\{e_i, e_j\} = 2Q_{ij} \tag{23.38}$$

Note that since $e_i^2 = \pm 1$ the algebra only admits a $\mathbb{Z}_2$ grading and moreover it is *not supercommutative*, because an odd element squares to zero in a supercommutative algebra.

It is instructive to look at some small values of r, s. Consider $C\ell_{-1}$. This has a single generator e with relation $e^2 = -1$. Therefore

$$C\ell_{-1} = \mathbb{R} \oplus \mathbb{R}e \tag{23.39}$$

as a vector space. The multiplication is

$$(a \oplus be)(c \oplus de) = (ac - bd) \oplus (bc + ad)e \tag{23.40}$$

so $C\ell_{-1}$ is isomorphic to the complex numbers $\mathbb{C}$ as an *ungraded algebra*, although not as a graded algebra. Similarly, $C\ell_{+1}$ is

$$C\ell_{+1} = \mathbb{R} \oplus \mathbb{R}e \tag{23.41}$$

as a vector space with multiplication:

$$(a \oplus be)(c \oplus de) = (ac + bd) \oplus (bc + ad)e. \tag{23.42}$$

As an ungraded algebra this is sometimes known as the "double numbers."

Note that both $C\ell_{-1}$ and $C\ell_{+1}$ are *commutative* as ungraded algebras but *noncommutative* as superalgebras. Thus the centers of these as ungraded algebras are $C\ell_{\pm 1}$ but the supercenter of $C\ell_{\pm 1}$ as graded algebras are $Z_s(C\ell_{\pm 1}) \cong \mathbb{R}$. In fact, for Q nondegenerate it can be shown that

$$Z_s(C\ell(Q)) \cong \mathbb{R} \tag{23.43}$$

We can also look at graded tensor products. First, note that for $n > 0$:

$$C\ell_n \cong \underbrace{C\ell_1 \widehat{\otimes} \cdots \widehat{\otimes} C\ell_1}_{\text{n times}} \tag{23.44}$$

$$C\ell_{-n} \cong \underbrace{C\ell_{-1} \widehat{\otimes} \cdots \widehat{\otimes} C\ell_{-1}}_{\text{n times}} \tag{23.45}$$

More generally we have

$$C\ell_{r+,s-} = \underbrace{C\ell_1 \widehat{\otimes} \cdots \widehat{\otimes} C\ell_1}_{\text{r times}} \widehat{\otimes} \underbrace{C\ell_{-1} \widehat{\otimes} \cdots \widehat{\otimes} C\ell_{-1}}_{\text{s times}} \tag{23.46}$$

We can similarly discuss the complex Clifford algebras $\mathbb{C}\ell_n$. Note that over the complex numbers if $e^2 = +1$ then $(ie)^2 = -1$ so we do not need to account for the signature, and WLOG we can just consider $\mathbb{C}\ell_n$ for $n \geq 0$. In particular, let $D \cong \mathbb{C}\ell_1$. Note that D is *not* a matrix superalgebra since it's dimension as an ordinary complex vector space, namely 2, is not a perfect square.

Definition A super-algebra over κ is *central simple* if, after extension of scalars to an algebraic closure $\bar{\kappa}$ it is isomorphic to a matrix super algebra $\mathrm{End}(V)$ or to $\mathrm{End}(V)\widehat{\otimes}D$.

This is the definition one finds in Section 3.3 of Deligne's *Notes on Spinors*. In particular, it is shown in Chapter 10, with this definition, that the Clifford algebras over $\mathbb{R}$ and $\mathbb{C}$ are central simple.

Exercise *The opposite algebra*

a.) For any ungraded algebra A we can define the *opposite algebra* A^{opp} by the rule

$$a \cdot^{\mathrm{opp}} b := ba \tag{23.47}$$

Show that A^{opp} is still an algebra.

b.) Show that $A \otimes A^{\mathrm{opp}} \cong \mathrm{End}(A)$.

c.) For any superalgebra A we can define the *opposite superalgebra* A^{opp} by the rule

$$a \cdot^{\mathrm{opp}} b := (-1)^{|a||b|} ba \tag{23.48}$$

Show that A^{opp} is still an superalgebra.

d.) Show that A is supercommutative iff $A = A^{\mathrm{opp}}$.

e.) Show that $A\widehat{\otimes}A^{\mathrm{opp}} \cong \mathrm{End}(A)$ as superalgebras.

f.) Show that if $\mathcal{A} = C\ell_{r+,s-}$ then $\mathcal{A}^{\mathrm{opp}} = C\ell_{s+,r-}$.

Exercise *Super Ideals*

An ideal I in a superalgebra is an ideal in the usual sense: For all $a \in \mathcal{A}$ and $b \in I$ we have $ab \in I$ (left ideal) or $ba \in I$ (right ideal), or both (two-sided ideal). The ideal is *homogeneous* if I is the direct sum of $I^0 = I \cap \mathcal{A}^0$ and $I^1 = I \cap \mathcal{A}^1$. (Explain why this is a nontrivial condition!)

a.) Show that the ideal $\mathcal{I}^{\rm odd}$ generated by all odd elements in $\mathcal{A}$ is homogeneous and given by

$$\mathcal{I}^{\rm odd} = (\mathcal{A}^1)^2 \oplus \mathcal{A}^1 \tag{23.49}$$

b.) Show that

$$\mathcal{A}/\mathcal{I}^{\rm odd} \cong \mathcal{A}^0/((\mathcal{A}^1)^2) \tag{23.50}$$

c.) Another definition of *central simple* is that there are no nontrivial homogeneous two-sided ideals. Show that this is equivalent to the definition above.

d.) Describe an explicit basis for the ideal generated by all odd elements in the Grassmann algebra $\kappa[\theta_1, \ldots, \theta_n]$.

e.) Give an example of a supercommutative algebra which is *not* a Grassmann algebra. [54]

Exercise *Invertibility lemma*

Let $\mathcal{A}$ be a *supercommutative* superalgebra and let $\mathcal{I}^{\rm odd} = (\mathcal{A}^1)$ be the ideal generated by odd elements. Let π be the projection

$$\pi : \mathcal{A} \to \mathcal{A}_{\rm red} = \mathcal{A}/\mathcal{I}^{\rm odd} \cong \mathcal{A}^0/((\mathcal{A}^1)^2) \tag{23.51}$$

a.) Show that a is invertible iff $\pi(a)$ is invertible. [55]

b.) Show that in a Grassmann algebra the map π is the same as reduction modulo nilpotents, or, more concretely, just putting the θ^i's to zero.

Exercise *Supercommutators and super Lie algebras*

The *graded commutator* or *supercommutator* of even elements in a superalgebra is

$$[a, b] := ab - (-1)^{|a||b|}ba \tag{23.52}$$

Since the expression $ab - ba$ still makes sense this notation can cause confusion so one must exercise caution when reading.

Show that the graded commutator satisfies:

1. $[\cdot, \cdot]$ is linear in both entries.

[54] *Answer*: Hint: Consider the possiblity that there are even nilpotent elements in $\mathcal{A}^0$ which are not the square of odd elements. Or consider functions on an algebraic supermanifold.

[55] *Answer*: One direction is trivial. If $\pi(a)$ is invertible then, since π is onto, there is an element $b \in \mathcal{A}$ with $1 = \pi(a)\pi(b) = \pi(ab)$. Therefore it suffices to show that if $\pi(a) = 1$ then a is invertible. But if $\pi(a) = 1$ then there is a finite set of odd elements ξ_i and elements $c_i \in \mathcal{A}$ so that $a = 1 - \nu$ with $\nu = \sum_{i=1}^n c_i\xi_i$. Note that $\nu^{n+1} = 0$ (by supercommutativity and the pigeonhole principle) so that $a^{-1} = 1 + \nu + \cdots + \nu^n$.

2. $[b,a] = (-1)^{1+|a||b|}[a,b]$

3. The *super Jacobi identity*:

$$(-1)^{x_1x_3}[X_1,[X_2,X_3]] + (-1)^{x_2x_1}[X_2,[X_3,X_1]] + (-1)^{x_3x_2}[X_3,[X_1,X_2]] = 0 \quad (23.53)$$

where $x_i = \deg(X_i)$.

These two conditions are abstracted from the properties of super-commutators to define super Lie algebras in Chapter 12 below. Briefly: We define a super vector space $\mathfrak{g}$ to be a *super Lie algebra* if there is an (abstract) map $[\cdot,\cdot] : \mathfrak{g}\times\mathfrak{g} \to \kappa$ which satisfies the conditions 1,2,3 above.

Exercise *Super-Derivations*

Definition: A *derivation* of a superalgebra is a homogeneous linear map $D : A \to A$ such that

$$D(ab) = D(a)b + (-1)^{|D||a|}aD(b) \quad (23.54)$$

a.) Show that the supercommutator of two superderivations is a superderivation.

b.) Show that the odd derivations of the Grassmann algebra are of the form

$$\sum_i f^i \frac{\partial}{\partial\theta^i} \quad (23.55)$$

where f^i are even.

Exercise *Multiplying Clifford algebras*

a.) Show that the real Clifford algebras are of dimension $\dim_{\mathbb{R}} C\ell_n = 2^{|n|}$, for any $n \in \mathbb{Z}$.

b.) Show that if n, m are integers with the same sign then $C\ell_n \widehat{\otimes} C\ell_m \cong C\ell_{n+m}$. Show that if n, m are *any* integers, then

$$C\ell_n \widehat{\otimes} C\ell_m \cong C\ell_{n+m} \widehat{\otimes} M \quad (23.56)$$

where M is a matrix superalgebra.

23.4 Modules over superalgebras

Definition A *super-module M over a super-algebra $\mathcal{A}$* (where $\mathcal{A}$ is itself a superalgebra over a field κ) is a supervector space M over κ together with a κ-linear map $\mathcal{A} \times M \to M$ defining a left-action or a right-action. That is, it is a left-module if, denoting the map by $L : \mathcal{A} \times M \to M$ we have

$$L(a, L(b, m)) = L(ab, m) \tag{23.57}$$

and it is a right-module if, denoting the map by $R : \mathcal{A} \times M \to M$ we have

$$R(a, R(b, m)) = R(ba, m) \tag{23.58}$$

In either case:

$$\deg(R(a, m)) = \deg(L(a, m)) = \deg(a) + \deg(m) \tag{23.59}$$

The notations $L(a, m)$ and $R(a, m)$ are somewhat cumbersome and instead we write $L(a, m) = am$ and $R(a, m) = ma$ so that $(ab)m = a(bm)$ and $m(ab) = (ma)b$. We also sometimes refer to a super-module over a super-algebra $\mathcal{A}$ just as a *representation of* $\mathcal{A}$.

Definition A *linear transformation* between two super-modules M, N over $\mathcal{A}$ is a κ-linear transformation of supervector spaces such that if T is homogeneous and M is a left $\mathcal{A}$-module then $T(am) = (-1)^{|T||a|}aT(m)$ while if M is a right $\mathcal{A}$-module then $T(ma) = T(m)a$. We denote the space of such linear transformations by $\mathrm{Hom}_{\mathcal{A}}(M, N)$. If N is a left $\mathcal{A}$-module then $\mathrm{Hom}_{\mathcal{A}}(M, N)$ is a left $\mathcal{A}$-module with $(a \cdot T)(m) := a \cdot (T(m))$. If N is a right $\mathcal{A}$-module then $\mathrm{Hom}_{\mathcal{A}}(M, N)$ is a right $\mathcal{A}$-module with $(T \cdot a)(m) := (-1)^{|a||m|}T(m)a$. When $M = N$ we denote the module of linear transformations by $\mathrm{End}_{\mathcal{A}}(M)$.

Example 1- continued *Matrix superalgebras.* In the ungraded world a matrix algebra $\mathrm{End}(V)$ for a finite dimensional vector space, say, over $\mathbb{C}$, has a unique irreducible representation, up to isomorphism. This is just the space V itself. A rather tricky point is that if V is a supervector space $V = \mathbb{C}^{p|q}$ then V and ΠV are *inequivalent* representations of $\mathrm{End}(V)$. One way to see this is that if η is a generator of $\Pi = \mathbb{C}^{0|1}$ then $T(\eta v) = (-1)^{|T|}\eta T(v)$ is a priori a different module. In terms of matrices

$$\begin{pmatrix} D & -C \\ -B & A \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{23.60}$$

So the LHS gives a representation of the matrix superalgebra, but it is not related by an automorphism $GL(\mathbb{C}^{p|q})$. The even subalgebra $\mathrm{End}(\mathbb{C}^p) \oplus \mathrm{End}(\mathbb{C}^q)$ has a unique faithful representation $\mathbb{C}^p \oplus \mathbb{C}^q$ and hence the matrix superalgebra $\mathrm{End}(\mathbb{C}^{p|q})$ has exactly two irreducible modules.

Example 3- continued *Clifford Modules.* A good example of supermodules over a superalgebra are the $\mathbb{Z}_2$-graded modules for the $\mathbb{Z}_2$-graded Clifford algebras.

Already for $C\ell_0 \cong \mathbb{R}$ there is a difference between graded and ungraded modules. There is a unique irreducible ungraded module, namely $\mathbb{R}$ acting on itself. But there are two inequivalent graded modules, $\mathbb{R}^{1|0}$ and $\mathbb{R}^{0|1}$.

Let us also discuss the representations of $C\ell_{\pm 1}$. As an ungraded algebra $C\ell_{+1} \cong \mathbb{R} \oplus \mathbb{R}$ because we can introduce projection operators $P_\pm = \frac{1}{2}(1 \pm e)$, so

$$C\ell_{+1} \cong \mathbb{R}P_+ \oplus \mathbb{R}P_- \qquad \text{ungraded!} \tag{23.61}$$

Therefore, there are *two* inequivalent ungraded irreducible representations with carrier space $\mathbb{R}$ and $\rho(e) = \pm 1$. However, as a graded algebra there is a unique irreducible representation, $\mathbb{R}^{1|1}$ with

$$\rho(e) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{23.62}$$

since e is odd and squares to 1.

Similarly, $C\ell_{-1}$ as an *ungraded* algebra is isomorphic to $\mathbb{C}$ and has a unique ungraded irreducible representation: $\mathbb{C}$ acts on itself. (As representations of a real algebra $\rho(e) = \pm i$ are equivalent.) However, as a graded algebra there a unique irreducible representation, $\mathbb{R}^{1|1}$ with

$$\rho(e) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{23.63}$$

Now, $C\ell_{1,-1}$ has two irreducible graded representations $\mathbb{R}^{1|1}_\pm$ with

$$\rho(e_1) = \pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} := \epsilon \tag{23.64}$$

Note that these are both odd, they anticommute, and they square to ± 1, respectively. Moreover, they generate all linear transformations on $\mathbb{R}^{1|1}$. Therefore, $C\ell_{1,-1}$ is a supermatrix algebra:

$$C\ell_{1,-1} \cong \mathrm{End}(\mathbb{R}^{1|1}) \tag{23.65}$$

It is interesting to compare this with $C\ell_{+2}$. Now, as an *ungraded algebra* we have a representation

$$\rho(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{23.66}$$

eq:epsdef

since these matrices anticommute and both square to $+1$. These generate the full matrix algebra $M_2(\mathbb{R})$ as an ungraded algebra. However, if we try to use these operators on $\mathbb{R}^{1|1}$ this is *not* a representation of $C\ell_{+2}$ as a graded algebra because $\rho(e_2)$ is not odd.

In fact, $C\ell_{+2}$ is *not* equivalent to a matrix superalgebra. In Chapter 10(??) we prove the beautiful periodicity theorem (closely related to Bott periodicity):

Theorem $C\ell_{r+,s-}$ is equivalent to a supermatrix algebra iff $(r - s) = 0 \mathrm{mod} 8$.

There is a unique irreducible representation of $C\ell_{+2}$ as a superalgebra. The carrier space is the $(2|2)$-dimensional space $\mathbb{R}^{2|2}$ and is given - up to similarity - by

$$\rho(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \tag{23.67}$$

It is true that $\mathbb{R}^{2|2} = \mathbb{R}^{1|1} \widehat{\otimes} \mathbb{R}^{1|1}$. But the tensor product of matrix representations does *not* give a matrix representation of the graded tensor product.

If we work with complex Clifford algebras the story is slightly different. $\mathbb{C}\ell_1$ as an ungraded algebra is $\mathbb{C} \oplus \mathbb{C}$ and has two inequivalent ungraded representations. As a graded algebra it has a unique irreducible graded representation $\mathbb{C}^{1|1}$; we could take, for example $\rho(e) = \sigma^1$. Then $\mathbb{C}\ell_2$ as an ungraded algebra is the matrix algebra $M_2(\mathbb{C})$ and as a graded algebra is a matrix superalgebra $\mathrm{End}(\mathbb{C}^{1|1})$. As a matrix superalgebra it actually has *two* inequivalent graded representations, both of which have carrier space $\mathbb{C}^{1|1}$. We could take, for example, $\rho(e_1) = \sigma^1$ and $\rho(e_2) = \pm\sigma^2$. One way to see these are inequivalent is to note that the volume form $\rho(e_1 e_2)$ restricted to the even subspace is a different scalar in the two cases.

We will discuss much more about Clifford modules in Chapter 10, for now, we summarize the discussion here in the following table:

Clifford Algebra	Ungraded algebra	Graded algebra	Ungraded irreps	Graded irreps
$C\ell_{-1}$	$\mathbb{C}$	$\mathbb{R}[e], e^2 = -1$	$\mathbb{C}$	$\mathbb{R}^{1\|1}, \rho(e) = \epsilon$
$C\ell_0$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}^{1\|0}, \mathbb{R}^{0\|1}$
$C\ell_{+1}$	$\mathbb{R} \oplus \mathbb{R}$	$\mathbb{R}[e], e^2 = 1$	$\mathbb{R}_\pm, \rho(e) = \pm 1$	$\mathbb{R}^{1\|1}, \rho(e) = \sigma^1$
$C\ell_{+2}$	$M_2(\mathbb{R})$	$C\ell_{+2}$	$\mathbb{R}^2$	$\mathbb{R}^{2\|2}$
$C\ell_{+1,-1}$	$M_2(\mathbb{R})$	$\mathrm{End}(\mathbb{R}^{1\|1})$	$\mathbb{R}^2$	$\mathbb{R}^{1\|1}_\pm$
$\mathbb{C}\ell_{+1}$	$\mathbb{C} \oplus \mathbb{C}$	$\mathbb{C}\ell_{+1}$	$\mathbb{C}_\pm, \rho(e) = \pm 1$	$\mathbb{C}^{1\|1}$
$\mathbb{C}\ell_{+2}$	$M_2(\mathbb{C})$	$\mathrm{End}(\mathbb{C}^{1\|1})$	$\mathbb{C}^2$	$\mathbb{C}^{1\|1}_\pm$

Remark: In condensed matter physics a *Majorana fermion* is a *real* operator γ which squares to one. If there are several γ_i they anticommute. Therefore, the Majorana fermions generate a *real Clifford algebra* within the set of observables of a physical system admitting Majorana fermions. If we have two sets of Majorana fermions then we expect their combined system to be a tensor product. Here we see that only the graded tensor product

will produce the expected rule for the physical observables. This is one reason why it is important to take a graded tensor product in the amalgamation axiom in the Dirac-von Neuman axioms.

What about the Hilbert spaces representing states of a Majorana fermion? If we view these as representations of an ungraded algebra then we encounter a famous paradox. (For the moment, take the Hilbert spaces to be real.) $C\ell_{+2}$ as an ungraded algebra has irreducible representation $\mathbb{R}^2$. On the other hand, this is a system of *two* Majorana fermions γ_1 and γ_2 so we expect that each Majorana fermion has a Hilbert space $\mathcal{H}_1$ and $\mathcal{H}_2$ and moreover these are isomorphic, so $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ implies that $\dim_{\mathbb{R}}\mathcal{H}_1 = \dim_{\mathbb{R}}\mathcal{H}_2 = \sqrt{2}$. This is nonsense! If we view the Hilbert space representations as complexifications of real representations then the paradox evaporates with the use of the graded tensor products: As irreducible representations we have:

$$\mathbb{R}^{2|2} = \mathbb{R}^{1|1}\widehat{\otimes}\mathbb{R}^{1|1} \tag{23.68}$$

The situation is a bit more tricky if we use complex graded representations. The paradox returns if we insist on using an irreducible representation of $\mathbb{C}\ell_2$, both in the graded and ungraded cases. However, in the graded case we can say that the 7th DvN axiom is satisfied if the physical (graded) representation is

$$\mathbb{C}^{1|1}\widehat{\otimes}\mathbb{C}^{1|1} \cong \mathbb{C}^{1|1}_{+} \oplus \mathbb{C}^{1|1}_{-} \tag{23.69}$$

Exercise *Tensor product of modules*

Let $\mathcal{A}$ and $\mathcal{B}$ be superalgebras with modules M and N, respectively. Show that the rule

$$(a \otimes b) \cdot (m \otimes n) := (-1)^{|b||m|}(am) \otimes (bn) \tag{23.70}$$

does indeed define $M \otimes N$ as an $\mathcal{A}\widehat{\otimes}\mathcal{B}$ module. Be careful with the signs!

Exercise *Left modules vs. right modules*

Suppose $\mathcal{A}$ is supercommutative.

a.) Show that if $(a, m) \to L(a, m)$ is a left-module then the new product $R : \mathcal{A} \times M \to M$ defined by

$$R(a, m) := (-1)^{|a||m|}L(a, m) \tag{23.71}$$

defines M as a *right-module*, that is,

$$R(a_1, R(a_2, m)) = R(a_2 a_1, m) \tag{23.72}$$

b.) Similarly, show that if M is a right-module then it can be canonically considered also to be a left-module.

Because of this we will sometimes write the module multiplication on the left or the right, depending on which order is more convenient to keep the signs down.

Exercise *Representations of Clifford algebras*
Show that

$$\rho(e_1) = \begin{pmatrix} 0 & \sigma^1 \\ \sigma^1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 0 & \sigma^3 \\ \sigma^3 & 0 \end{pmatrix} \tag{23.73}$$

is a graded representation of $C\ell_{+2}$ on $\mathbb{R}^{2|2}$. Show that it is equivalent to the one given above.

23.5 Free modules and the super-General Linear Group

Now let $\mathcal{A}$ be supercommutative. Then we can define a *free right $\mathcal{A}$-module*, $\mathcal{A}^{p|q}$ to be

$$\mathcal{A}^{p|q} = \mathcal{A}^{\oplus p} \oplus (\Pi\mathcal{A})^{\oplus q} \tag{23.74}$$

as a supervector space with the obvious right $\mathcal{A}$-module action.

Since it is a free module we can choose a basis. Set $n = p + q$ and choose a basis e_i, $1 \leq i \leq n$ with e_i even for $1 \leq i \leq p$ and odd for $p + 1 \leq i \leq p + q = n$. Then we can identify

$$\mathcal{A}^{p|q} \cong e_1\mathcal{A} \oplus \cdots \oplus e_n\mathcal{A} \tag{23.75}$$

We define the degree of $e_i a$ to be $\deg(e_i) + \deg(a)$ so that the even part of $\mathcal{A}^{p|q}$ is

$$(\mathcal{A}^{p|q})^0 = e_1\mathcal{A}^0 \oplus \cdots \oplus e_p\mathcal{A}^0 \oplus e_{p+1}\mathcal{A}^1 \oplus \cdots \oplus e_n\mathcal{A}^1 \tag{23.76}$$

Definition: If $\mathcal{A}$ is supercommutative we define $GL(\mathcal{A}^{p|q})$ to be the group of *automorphisms* of $\mathcal{A}^{p|q}$. Recalling that morphisms in the category of supervector spaces are parity-preserving this may be identified with the group of invertible even elements in $\mathrm{End}_{\mathcal{A}}(\mathcal{A}^{p|q})$.

We stress that even though $GL(\mathcal{A}^{p|q})$ is called a supergroup it is actually an honest group. However, it is not an honest manifold, but actually a supermanifold.

♣Mathematicians take a more categorical approach to defining supergroups. Make sure this is compatible. ♣

It is useful to give a matrix description of these groups. We represent the general element m of $\mathcal{A}^{p|q}$ by $m = e_i x^i$, with $x^i \in \mathcal{A}$. Then the general module map $T : \mathcal{A}^{p|q} \to \mathcal{A}^{r|s}$ is determined by its action on basis vectors:

$$T(e_j) = \tilde{e}_\alpha X^\alpha{}_j \qquad X^\alpha{}_j \in \mathcal{A} \tag{23.77}$$

where $\tilde{e}_\alpha$, $\alpha = 1, \ldots, r + s$, are the generators of $\mathcal{A}^{r|s}$.

We say the matrix X with matrix elements $X^{\alpha}{}_{j}$ (which are elements of $\mathcal{A}$) where rows and columns have a parity assigned is an $(r|s) \times (p|q)$ *supermatrix*.

If we choose a basis for $\mathcal{A}^{p|q}$ then we may represent an element $m \in \mathcal{A}^{p|q}$ by a column vector

$$\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \tag{23.78}$$

then the (active) transformation T is given by a matrix multiplication from the left with block form:

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{23.79}$$ eq:blockform

The supermatrix representing the composition of transformations $T_1 \circ T_2$ is the ordinary matrix product of X_1 and X_2.

When T is an even transformation

$$A \in M_{r\times p}(\mathcal{A}^0) \qquad B \in M_{r\times q}(\mathcal{A}^1) \tag{23.80}$$

$$C \in M_{s\times p}(\mathcal{A}^1) \qquad D \in M_{s\times q}(\mathcal{A}^0) \tag{23.81}$$

or, more informally, X is of the form:

$$\begin{pmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{pmatrix} \tag{23.82}$$ eq:even-blockf

When T is an odd transformation

$$A \in M_{r\times p}(\mathcal{A}^1) \qquad B \in M_{r\times q}(\mathcal{A}^0) \tag{23.83}$$

$$C \in M_{s\times p}(\mathcal{A}^0) \qquad D \in M_{s\times q}(\mathcal{A}^1) \tag{23.84}$$

or, more informally, X is of the form:

$$\begin{pmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{pmatrix} \tag{23.85}$$

Example 1: $\mathcal{A} = \kappa$. Then there are no odd elements in $\mathcal{A}$ and we have invertible morphisms. This group of automorphisms of $\kappa^{p|q}$ is isomorphic to $GL(p;\kappa) \times GL(q;\kappa)$ and in a homogeneous basis will have block diagonal form

$$\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \tag{23.86}$$

with A, D invertible.

Example 2: $\mathcal{A} = \kappa[\theta^1, \ldots, \theta^r]$ is a Grassmann algebra, then $GL(\mathcal{A}^{p|q})$ consists of matrices (23.79) which are even, i.e. of the form (23.82), with A, D invertible, which is the same as A, D being invertible modulo θ^i. (See Section §23.7 below.) For example

$$\begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \tag{23.87}$$

$$\begin{pmatrix} 1+\theta_1\theta_2 & \theta_1 \\ \theta_2 & 1-\theta_1\theta_2 \end{pmatrix} \tag{23.88}$$

are examples of such general linear transformations. In fact they are both of the form $\exp(Y)$ for an even supermatrix Y. (Find it!)

Remarks

1. Note a tricky point: If $T : \mathcal{A}^{p|q} \to \mathcal{A}^{r|s}$ is a linear transformation and we have chosen bases as above so that T is represented by a supermatrix X then the supermatrix representing aT is *not* $aX^\alpha{}_j$, rather it is the supermatrix:

$$\begin{pmatrix} a1_{r\times r} & 0 \\ 0 & (-1)^{|a|}a1_{s\times s} \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{23.89}$$

   Similarly, the matrix representing Ta is *not* $X^\alpha{}_j a$, rather it is the supermatrix:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a1_{p\times p} & 0 \\ 0 & (-1)^{|a|}a1_{q\times q} \end{pmatrix} \tag{23.90}$$

2. In Chapter 8 (?) we describe the relation between Lie groups and Lie algebras. Informally this is just given by the exponential map and Lie algebra elements exponentiate to form one-parameter subgroups $g(t) = \exp(tA)$ of G. The same reasoning applies to $GL(\mathcal{A}^{p|q})$ and the super Lie algebra $gl(\mathcal{A}^{p|q})$ is - as a supervector space - the same as $\mathrm{End}(\mathcal{A}^{p|q})$.

23.6 The Supertrace

There are analogs of the trace and determinant for elements of $\mathrm{End}(\mathcal{A}^{p|q})$ with $\mathcal{A}$ supercommutative.

For $X \in \mathrm{End}(\mathcal{A}^{p|q})$ we define the supertrace on homogeneous elements by

$$\mathrm{STr}(X) = \mathrm{STr}(\begin{pmatrix} A & B \\ C & D \end{pmatrix}) := \begin{cases} \mathrm{tr}\,(A) - \mathrm{tr}\,(D) & X \text{ even} \\ \mathrm{tr}\,(A) + \mathrm{tr}\,(D) & X \text{ odd} \end{cases} \tag{23.91}$$

that is

$$\mathrm{STr}(X) = \mathrm{tr}\,(A) - (-1)^{|X|}\mathrm{tr}\,(D) \in \mathcal{A} \tag{23.92}$$

The supertrace satisfies $\mathrm{STr}(X+Y) = \mathrm{STr}(X) + \mathrm{STr}(Y)$ so we can extend it to all of $\mathrm{End}(\mathcal{A}^{p|q})$ by linearity.

Now one can easily check (do the exercise!!) that the supertrace satisfies the properties:

1. $\mathrm{STr}(XY) = (-1)^{|X||Y|}\mathrm{STr}(YX)$ and therefore the supertrace of a graded commutator vanishes. Note that the signs in the definition of the supertrace are crucial for this to be true.

2. $\mathrm{STr}(aX) = a\mathrm{STr}(X)$.

3. If g is even and invertible then $\mathrm{STr}(g^{-1}Xg) = \mathrm{STr}(X)$. This follows from the cyclicity property we just stated. Therefore, the supertrace is basis independent for a free module and hence is an intrinsic property of the linear transformation.

Remark: In the case where $\mathcal{A} = \kappa$ and we have a linear transformation on a supervector space we can say the supertrace of $T \in \mathrm{End}V$ is

$$\mathrm{STr}T := \mathrm{Tr}(P_V T) \tag{23.93}$$ eq:strce

In supersymmetric field theories P_V is often denoted $(-1)^F$, where F is a fermion number and the supertrace becomes $\mathrm{Tr}(-1)^F T$. These traces are very important in obtaining exact results in supersymmetric field theories.

Exercise

a.) Show that in general

$$\mathrm{STr}(T_1T_2) \neq \mathrm{STr}(T_2T_1) \tag{23.94}$$

b.) Check that if T_1, T_2 are homogeneous then

$$\mathrm{STr}(T_1T_2) = (-1)^{\mathrm{deg}T_1 \cdot \mathrm{deg}T_2}\mathrm{STr}(T_2T_1) \tag{23.95}$$

---

sec:Berezinian

23.7 The Berezinian of a linear transformation

Let $\mathcal{A}$ be supercommutative and consider the free $\mathcal{A}$ module $\mathcal{A}^{p|q}$.

While the determinant of a matrix can be defined for any matrix, the the *superdeterminant* or *Berezinian* can only be defined for elements of $GL(\mathcal{A}^{p|q})$. If $X \in \mathrm{End}(\mathcal{A}^{p|q})$ is even and invertible then the value of $\mathrm{Ber}(X)$ lies in the subalgebra of invertible elements of $\mathcal{A}^0$, which we can consider to be $GL(\mathcal{A}^{1|0})$.

The conditions which characterize the Berezinian are :

1. When X can be written as an exponential of a matrix $X = \mathrm{exp}Y$, with $Y \in \mathrm{End}(\mathcal{A}^{p|q})$ we must have

$$\mathrm{Ber}(X) = \mathrm{Ber}(\mathrm{exp}Y) := \mathrm{exp}(\mathrm{STr}\ Y) \tag{23.96}$$ eq:sdettr

2. The Berezinian is multiplicative:

$$\mathrm{Ber}(X_1X_2) = \mathrm{Ber}(X_1)\mathrm{Ber}(X_2) \tag{23.97}$$

eq:MultBer

Note that the two properties (23.96) and (23.97) are compatible thanks to the Baker-Campbell-Hausdorff formula. (See Chapter 8 below.)

We can use these properties to give a formula for the Berezinian of a matrix once we know the key result

♣Actually, need to show the supertrace condition is well-defined. ♣

Lemma Let $\mathcal{A}$ be supercommutative and $\pi : \mathcal{A} \to \mathcal{A}_{\rm red} = \mathcal{A}/\mathcal{I}^{\rm odd}$. This defines a map

$$\pi : \mathrm{End}(\mathcal{A}^{p|q}) \to \mathrm{End}(\mathcal{A}_{\rm red}^{p|q}) \tag{23.98}$$

by applying π to the matrix elements. Then: the supermatrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{23.99}$$

1. Is invertible iff $\pi(A)$ and $\pi(D)$ are invertible.
2. Is in the image of the exponential map iff $\pi(A)$ and $\pi(D)$ are.

Proof: The proof follows closely that of the invertibility lemma above. Note that since X is even then $\pi(X)$ is block diagonal so $\pi(X)$ is invertible iff $\pi(A)$ and $\pi(D)$ are invertible. If $\pi(X)$ is invertible then, since π is onto there is a $Y \in \mathrm{End}(\mathcal{A}^{p|q})$ so that $1 = \pi(X)\pi(Y) = \pi(XY)$. Then $XY = 1 - Z$ for some Z such that $\pi(Z) = 0$. All the matrix elements of Z are nilpotent so there is an N so that $Z^{N+1} = 0$. Then $Y(1 + Z + \cdots + Z^N)$ is the inverse of X.

By the same token if $\pi(X) = \exp(\alpha)$ then we can lift α to $\tilde{\alpha} \in \mathrm{End}(\mathcal{A}^{p|q})$ and $\pi(X\exp[-\tilde{\alpha}]) = 1$ so $X\exp[-\tilde{\alpha}] = 1 - Z$ where Z is nilpotent. Therefore $1 - Z = \exp[z]$ is well-defined because the series for $\log(1 - Z)$ terminates. ♠.

Now – assuming that a Berezinian function actually exists – we can give a formula for what it must be. From the first condition we know that when $\pi(A), \pi(D)$ are in the image of the exponential map then

$$\mathrm{Ber}\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} = \frac{\det A}{\det D} \tag{23.100}$$

Note that the entries of A and D are all even so in writing out the usual definition of determinant there is no issue of ordering. Together with multiplicativity and the fact that the exponential map is onto for $GL(n, \kappa)$ this determines the formula for all block diagonal matrices.

Moreover, upper triangular matrices are in the image of the exponential once again because all the matrix elements of B are nilpotent so that

$$\log\begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} = -\sum_{k=1}^{\infty} \frac{(-1)^k}{k}\begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}^k \tag{23.101}$$

terminates and is a well-defined series. Moreover it is clear that it has supertrace = 0, and therefore

$$\mathrm{Ber}(\begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix}) = 1 \tag{23.102}$$

$$\mathrm{Ber}(\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}) = 1 \tag{23.103}$$

Now for general invertible X we can write

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & BD^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ D^{-1}C & 1 \end{pmatrix} \tag{23.104}$$

and hence multiplicativity implies

$$\mathrm{Ber}(X) = \frac{\det(A - BD^{-1}C)}{\det D} = \frac{\det A}{\det D}\det(1 - A^{-1}BD^{-1}C) \tag{23.105}$$

eq:SDETFORM

There is one more point to settle here. We have shown that the two properties (23.96) and (23.97) uniquely determine the Berezinian of a supermatrix and even determine a formula for it. But, strictly speaking, we have not yet shown that the Berezinian actually exists, because we have not shown that the formula (23.105) is indeed multiplicative. A brute force approach to verifying this would be very complicated.

A better way to proceed is the following. We want to prove that $\mathrm{Ber}(gh) = \mathrm{Ber}(g)\mathrm{Ber}(h)$ for any two group elements g, h. Let us consider the subgroups G^+, G^0, G^- of upper triangular, block diagonal, and lower triangular matrices:

$$G^+ = \{X : X = \begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix}\} \tag{23.106}$$

$$G^0 = \{X : X = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}\} \tag{23.107}$$

$$G^- = \{X : X = \begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}\} \tag{23.108}$$

Any group element can be written as $g = g^+g^0g^-$ where $g^{\pm,0} \in G^{\pm,0}$. So now we need to consider $gh = g^+g^0g^-h^+h^0h^-$. It would be very complicated to rewrite this again as a Gauss decomposition. On the other hand, it is completely straightforward to check multiplicativity of the formula for products of the form g^+k, g^0k, kg^0, and kg^- for any k and $g^{\pm,0} \in G^{\pm,0}$. For example, to check multiplicativity for g^+k we write

$$\begin{pmatrix} 1 & B' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A + B'C & B + B'D \\ C & D \end{pmatrix} \tag{23.109}$$

and now we simply note that

$$\det((A + B'C) - (B + B'D)D^{-1}C) = \det(A - BD^{-1}C) \tag{23.110}$$

The other cases are similarly straightforward. (Check them!!) Therefore, to check multiplicativity we need only check that multiplicativity for products of the form $g^- h^+$. Therefore we need only show

$$\mathrm{Ber}\begin{pmatrix} 1 & B \\ C & 1+CB \end{pmatrix} = 1 \tag{23.111}$$

eq:check-ber-m

because

$$\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}\begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & B \\ C & 1+CB \end{pmatrix} \tag{23.112}$$

This is not completely obvious from (23.105). Nevertheless, it is easily shown: Note that the matrix in (23.111) is in the image of the exponential map. But it is trivial from the relation to the supertrace that if $g = \exp(Y)$ then $\mathrm{Ber}(g^{-1}) = (\mathrm{Ber}(g))^{-1}$. On the other hand,

$$\begin{pmatrix} 1 & B \\ C & 1+CB \end{pmatrix}^{-1} = \begin{pmatrix} 1+BC & -B \\ -C & 1 \end{pmatrix} \tag{23.113}$$

and applying the formula (23.105) to the RHS trivially gives one.

Finally, we remark that from the multiplicativity property it follows that $\mathrm{Ber}(g^{-1}Xg) = \mathrm{Ber}(X)$ and hence the Berezinian is invariant under change of basis. Therefore, it is intrinsically defined for an even invertible map $T \in \mathrm{End}(\mathcal{A}^{p|q})$.

Exercise

Let $\alpha, \beta \in \mathbb{C}^*$. Evaluate

$$\mathrm{Ber}\begin{pmatrix} \alpha & \theta_1 \\ \theta_2 & \beta \end{pmatrix} \tag{23.114}$$

Using both of the expressions above.

Exercise

Show that two alternative formulae for the Berezinian are

$$\mathrm{Ber}(X) = \frac{\det A}{\det(D - CA^{-1}B)} = \frac{\det A}{\det D}(\det(1 - D^{-1}CA^{-1}B))^{-1} \tag{23.115}$$

eq:SDET-ALT

Note that the equality of (23.105) and (23.115) follows because

$$\det(1 - A^{-1}BD^{-1}C) = (\det(1 - D^{-1}CA^{-1}B))^{-1} \tag{23.116}$$

and this in turn is easily established because both matrices are $1 + Nilpotent$ and hence in the image of the exponential map and since $A^{-1}B$ and $D^{-1}C$ are both odd we have

$$\mathrm{STr}(A^{-1}BD^{-1}C)^k = -\mathrm{STr}(D^{-1}CA^{-1}B)^k \tag{23.117}$$

This in turn gives another easy proof of multiplicativity, once one has reduced it to (23.111), which follows immediately from (23.115).

♣This is a better proof. Put this in the text and make the other an exercise. ♣

Exercise

Let $\alpha, \beta \in \mathbb{C}^*$. Evaluate

$$\mathrm{Ber}\begin{pmatrix} \alpha & \theta_1 \\ \theta_2 & \beta \end{pmatrix} \tag{23.118}$$

Using both of the expressions above.

23.8 Bilinear forms

Bilinear forms on super vector spaces $\mathfrak{b} : V \otimes V \to \kappa$ are defined as in the ungraded case: $\mathfrak{b}$ is a bilinear morphism of supervector spaces.

It follows that $\mathfrak{b}(x, y) = 0$ if x and y in V are homogeneous and of opposite parity.

We can identify the set of bilinear forms with $V^\vee \otimes V^\vee$. We can then apply super-symmetrization and super-antisymmetriziation.

Thus, symmetric bilinear forms have a very important extra sign.

$$\mathfrak{b}(x, y) = (-1)^{|x||y|}\mathfrak{b}(y, x) \tag{23.119}$$

This means $\mathfrak{b}$ is symmetric when restricted to $V^0 \times V^0$ and antisymmetric when restricted to $V^1 \times V^1$.

Similarly, antisymmetric bilinear forms have the reverse situation:

$$\mathfrak{b}(x, y) = (-1)^{1+|x||y|}\mathfrak{b}(y, x) \tag{23.120}$$

This means $\mathfrak{b}$ is anti-symmetric when restricted to $V^0 \times V^0$ and symmetric when restricted to $V^1 \times V^1$.

The definition of a *nondegenerate* form is the same as before. A form is nondegenerate iff its restrictions to $V^0 \times V^0$ and $V^1 \times V^1$ are nondegenerate. Therefore, applying the canonical forms of symmetric and antisymmetric matrices we discussed in Section §20 above we know that if $\kappa = \mathbb{R}$ and $\mathfrak{b}$ is a nondegenerate $\mathbb{Z}_2$-graded symmetric form then there is a basis where its matrix looks like

$$Q = \begin{pmatrix} 1_r & 0 & 0 & 0 \\ 0 & -1_s & 0 & 0 \\ 0 & 0 & 0 & -1_m \\ 0 & 0 & 1_m & 0 \end{pmatrix} \tag{23.121}$$

eq:OSPQ

The automorphisms of the bilinear form are the even invertible morphisms $g : V \to V$ such that

$$\mathfrak{b}(gv, gw) = \mathfrak{b}(v, w) \tag{23.122}$$

for all $v, w \in V$. This is just the group $O(r,s) \times Sp(2m;\mathbb{R})$.

As with the general linear group, to define more interesting automorphism groups of a bilinear form we need to consider bilinear forms on the free modules $\mathcal{A}^{p|q}$ over a supercommutative superalgebra $\mathcal{A}$.

Definition: Let $\mathcal{A}$ be a superalgebra. A bilinear form on a (left) $\mathcal{A}$-module M is a morphism of supervector spaces

$$M \otimes M \to \mathcal{A} \tag{23.123}$$

such that

$$\mathfrak{b}(am, m') = a\mathfrak{b}(m, m') \qquad \mathfrak{b}(m, am') = (-1)^{|a||m|} a\mathfrak{b}(m, m') \tag{23.124}$$

Now, if we apply this to the free module $\mathcal{A}^{p|q}$ over a supercommutative algebra $\mathcal{A}$ (which can be considered to be either a left or right $\mathcal{A}$-module then we have simply

$$a\mathfrak{b}(m, m') = \mathfrak{b}(am, m') \qquad \mathfrak{b}(ma, m') = \mathfrak{b}(m, am') \qquad \mathfrak{b}(m, m'a) = \mathfrak{b}(m, m')a \tag{23.125}$$

The automorphism group of $\mathfrak{b}$ is the group of $g \in \mathrm{End}(\mathcal{A}^{p|q})$ which are even and invertible and for which

$$\mathfrak{b}(gm, gm') = \mathfrak{b}(m, m') \tag{23.126}$$

for all $m, m' \in \mathcal{A}^{p|q}$. In the case where $\mathfrak{b}$ is a nondegenerate $\mathbb{Z}_2$-graded symmetric on $\mathcal{A}^{p|q}$ we define an interesting generalization of both the orthogonal and symplectic groups which plays an important role in physics. We could denote it $OSp_{\mathcal{A}}(\mathcal{A}^{p|q})$.

Using ideas from Chapter 8, discussed for the super-case in Chapter 12, we can use this discussion to derive the superLie algebra in the case where we specialize to a Grassmann algebra $\mathcal{A} = \mathbb{R}[\theta^1, \ldots, \theta^{q'}]/I$ so that $\mathcal{A}_{\mathrm{red}} = \mathbb{R}$. If $\mathfrak{b}$ is nondegenerate then on the reduced module $\mathbb{R}^{p|q}$ (where q and q' are not related) it can be brought to the form (23.121), so $p = r + s$ and $q = 2m$. Writing $g(t) = e^{tA}$ and differentiating wrt t at $t = 0$ we derive the Lie algebra of the supergroup. It is the subset of $\mathrm{End}(\mathcal{A}^{p|q})$ such that

$$\mathfrak{b}(Am, m') + \mathfrak{b}(m, Am') = 0 \tag{23.127}$$

If we finally reduce this equation mod nilpotents we obtain an equation on $\mathrm{End}(\mathbb{R}^{p|q})$. That defines a Lie algebra over $\mathbb{R}$ which is usually denoted $\mathrm{osp}(r,s|2m;\mathbb{R})$.

♣notation conflict m ♣

23.9 Star-structures and super-Hilbert spaces

There are at least *three* notions of a real structure on a complex superalgebra which one will encounter in the literature:

1. It is a $\mathbb{C}$-antilinear involutive *automorphism* $a \mapsto a^\bigstar$. Hence $\deg(a^\bigstar) = \deg(a)$ and $(ab)^\bigstar = a^\bigstar b^\bigstar$.

2. It is a $\mathbb{C}$-antilinear involutive *anti-automorphism*. Thus $\deg(a^*) = \deg(a)$ but

$$(ab)^* = (-1)^{|a||b|} b^* a^* \tag{23.128}$$

3. It is a $\mathbb{C}$-antilinear involutive *anti-automorphism*. Thus $\deg(a^\star) = \deg(a)$ but

$$(ab)^\star = b^\star a^\star \tag{23.129}$$

If $\mathcal{A}$ is a supercommutative complex superalgebra then structures 1 and 2 coincide: $a \to a^\bigstar$ is the same as $a \to a^*$. See remarks below for the relation of 2 and 3.

Definition A sesquilinear form h on a complex supervector space $\mathcal{H}$ is a map $h : \mathcal{H}\times\mathcal{H} \to \mathbb{C}$ such that

1. It is even, so that $h(v, w) = 0$ if v and w have opposite parity
2. It is $\mathbb{C}$-linear in the second variable and $\mathbb{C}$-antilinear in the first variable
3. An *Hermitian form* on a supervector space is a sesquilinear form which moreover satisfies the symmetry property:

$$(h(v, w))^* = (-1)^{|v||w|}h(w, v) \tag{23.130}$$

4. If in addition for all nonzero $v \in \mathcal{H}^0$

$$h(v, v) > 0 \tag{23.131}$$

while for all nonzero $v \in \mathcal{H}^1$

$$i^{-1}h(v, v) > 0, \tag{23.132}$$

then $\mathcal{H}$ endowed with the form h is a *super-Hilbert space*.

For bounded operators we define the adjoint of a homogeneous linear operator $T : \mathcal{H} \to \mathcal{H}$ by

$$h(T^*v, w) = (-1)^{|T||v|}h(v, Tw) \tag{23.133}$$

The spectral theorem is essentially the same as in the ungraded case with one strange modification. For even Hermitian operators the spectrum is real. However, for odd Hermitian operators the point spectrum sits in a real subspace of the complex plane which is *not* the real line! If T is odd then an eigenvector v such that $Tv = \lambda v$ must have even and odd parts $v = v_e + v_o$. Then the eigenvalue equation becomes

$$\begin{aligned} Tv_e &= \lambda v_o \\ Tv_o &= \lambda v_e \end{aligned} \tag{23.134}$$

Now the usual proof that the point spectrum is real is modified to:

$$\begin{aligned} \lambda^* h(v_o, v_o) &= h(\lambda v_o, v_o) = h(Tv_e, v_o) = h(v_e, Tv_o) = \lambda h(v_e, v_e) \\ \lambda^* h(v_e, v_e) &= h(\lambda v_e, v_e) = h(Tv_o, v_e) = -h(v_o, Tv_e) = -\lambda h(v_o, v_o) \end{aligned} \tag{23.135}$$

These two equations have the same content: Since $v \neq 0$ and we are in a superHilbert space it must be that

$$h(v_e, v_e) = i^{-1}h(v_o, v_o) > 0 \tag{23.136}$$

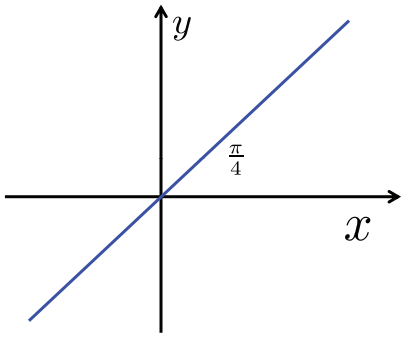


Figure 16: When the Koszul rule is consistently implemented odd super-Hermitian operators have a spectrum which lies along the line through the origin which runs through $1+i$. fig:SUPERHERMI

and therefore the phase of λ is determined. It lies on the line passing through $e^{i\pi/4} = (1+i)/\sqrt{2}$ in the complex plane, as shown in Figure 16

Example: An example of a natural super-Hilbert space is the Hilbert space of L^2-spinors on an even-dimensional manifold with $(-1)^F$ given by the chirality operator. An odd self-adjoint operator which will have nonhomogeneous eigenvectors is the Dirac operator on an even-dimensional manifold. One usually thinks of the eigenvalues as real for this operator and that is indeed the case if we use the star-structure $\star$, number 3 above. See the exercise below.

Remarks

1. In general star-structures 2 and 3 above are actually closely related. Indeed, given a structure $a \to a^*$ of type 2 we can define a structure of type 3 by defining either

$$a^\star = \begin{cases} a^* & |a| = 0 \\ ia^* & |a| = 1 \end{cases} \tag{23.137}$$

   or

$$a^\star = \begin{cases} a^* & |a| = 0 \\ -ia^* & |a| = 1 \end{cases} \tag{23.138}$$ eq:relstar

   It is very unfortunate that in most of the physics literature the definition of a star structure is that used in item 3 above. For example a typical formula used in manipulations in superspace is

$$\overline{\theta_1\theta_2} = \bar\theta_2\bar\theta_1 \tag{23.139}$$

   and the fermion kinetic energy

$$\int dt i\bar\psi \frac{d}{dt}\psi \tag{23.140}$$

is only "real" with the third convention. The rationale for this convention, especially for fermionic fields, is that they will eventually be quantized as operators on a Hilbert space. Physicists find it much more natural to have a standard Hilbert space structure, even if it is $\mathbb{Z}_2$-graded. On the other hand, item 2 implements the Koszul rule consistently and makes the analogy to classical physics as close as possible. So, for example, the fermionic kinetic term is

$$\int dt \bar{\psi} \frac{d}{dt} \psi \tag{23.141}$$

and is "manifestly real."

Fortunately, as we have just noted one convention can be converted to the other, but the difference will, for example, show up as factors of i in comparing supersymmetric Lagrangians in the different conventions, as the above examples show.

Exercise

a.) Show that a super-Hermitian form h on a super-Hilbert space can be used to define an ordinary Hilbert space structure on $\mathcal{H}$ by taking $\mathcal{H}^0 \perp \mathcal{H}^1$ and taking

$$\begin{aligned} (v,w) &:= h(v,w) \qquad\qquad v,w \in \mathcal{H}^0 \\ (v,w) &:= i^{-1} h(v,w) \qquad\quad v,w \in \mathcal{H}^1 \end{aligned} \tag{23.142}$$

eq:UnbradedHer

b.) Show that if T is an operator on a super-Hilbert-space then the super-adjoint T^* and the ordinary adjoint $T^\dagger$, the latter defined with respect to (23.142), are related by

$$T^* = \begin{cases} T^\dagger & |T| = 0 \\ iT^\dagger & |T| = 1 \end{cases} \tag{23.143}$$

c.) Show that $T \to T^\dagger$ is a star-structure on the superalgebra of operators on superspace which is of type 3 above.

d.) Show that if T is an odd self-adjoint operator with respect to $*$ then $e^{-i\pi/4}T$ is an odd self-adjoint operator with respect to $\dagger$. In particular $e^{-i\pi/4}T$ has a point spectrum in the real line.

e.) More generally, show that if a is odd and real with respect to $*$ then $e^{-i\pi/4}a$ is real with respect to $\star$ defined by (23.138).

23.9.1 SuperUnitary Group

Let us return to a general finite-dimensional Hermitian form on a complex supervectorspace. Restricted to V^0 it can be brought to the form $\mathrm{Diag}\{+1_r, -1_s\}$ while restricted to the odd subspace it can be brought to the form $\mathrm{Diag}\{+1_t, -1_u\}$. The automorphism group of (V,h) is therefore $U(r,s) \times U(t,u)$. If we consider instead a free module $\mathcal{A}^{n_e,n_o}$

over a supercommutative algebra $\mathcal{A}$ (where $\mathcal{A}$ is a vector space over $\kappa = \mathbb{C}$) we can still define an Hermitian form $h : \mathcal{A}^{n_e,n_o} \times \mathcal{A}^{n_e,n_o} \to \mathcal{A}$. If $\mathcal{A}_{\rm red} = \mathbb{C}$ and h is of the above type with $n_e = r + s$ and $n_o = t + u$ then the automorphism group of h is $U_{\mathcal{A}}(r, s|p, q)$. If we derive the Lie algebra and reduce modulo nilpotents we then obtain the super Lie algebra $\mathrm{u}(r, s|p, q; \mathbb{C})$ which is the subset of $\mathrm{End}(\mathbb{C}^{n_e|n_o})$

$$h(Av, v') + (-1)^{|A||v|}h(v, Av') = 0 \tag{23.144}$$

i.e $\mathrm{u}(r, s|p, q)$ is the real super Lie algebra of super-anti-unitary operators. We will say much more about this in Chapter 12.

Exercise *Fixed points*
Let $\eta_{r,s}$, and $\eta_{t,u}$ be diagonal matrices...
Show that $\mathrm{u}(r, s|p, q)$ the the set of fixed points of the antilinear involution ...
♣ FILL IN ♣

23.10 Functions on superspace and supermanifolds

23.10.1 Philosophical background

Sometimes one can approach the subjects of topology and geometry through algebra and analysis. Two famous examples of this are

1. Algebraic geometry: The geometry of vanishing loci of systems of polynomials can be translated into purely algebraic questions about commutative algebra.

2. Gelfand's Theorem on commutative C^*-algebras

We now explain a little bit about Gelfand's theorem:

There is a 1-1 correspondence between Hausdorff topological spaces and commutative C^*-algebras.

If X is a Hausdorff topological space then we can form $C_0(X)$, which is the space of all *continuous* complex valued functions $f : X \to \mathbb{C}$ which "vanish at infinity." What this means is that for all $\epsilon > 0$ the set of $x \in X$ so that $|f(x)| \geq \epsilon$ is a compact set. This is a C^*-algebra with involution $f \mapsto f^*$ where $f^*(x) := (f(x))^*$ and the norm is

$$\| f \| := \sup_{x \in X} |f(x)| \tag{23.145}$$

Then there is a 1-1 correspondence between isomorphism classes of topological spaces and isomorphism classes of commutative C^*-algebras.

The way one goes from a commutative C^* algebra $\mathcal{A}$ to a topological space is that one defines $\Delta(\mathcal{A})$ to be the set of - any of

a.) The C^*-algebra morphisms $\chi : \mathcal{A} \to \mathbb{C}$.

b.) The maximal ideals

c.) The irreducible representations.

For a commutative C^* algebra the three notions are equivalent. The space $\Delta(\mathcal{A})$ carries a natural topology since there is a norm on linear maps $\mathcal{A} \to \mathbb{C}$ of Banach spaces. It turns out that $\Delta(\mathcal{A})$ is a Hausdorff space. Gelfand's theorem then says that $C_0(\Delta(\mathcal{A}))$ is in fact isomorphic as a C^* algebra to $\mathcal{A}$, while $\Delta(C_0(X))$ is homeomorphic as a topological space to X.

The correspondence is very natural if we interpret a,b,c in terms of $\mathcal{A} = C_0(X)$. then, given a point $x \in X$ we have

a.) The morphism $\chi_x : f \mapsto f(x)$

b.) The maximal ideal $\mathfrak{m}_x = \ker(\chi_x) = \{f \in C_0(X) | f(x) = 0\}$

c.) The representations $\rho_x(f) = f(x)$.

Remarks:

1. As a simple and very important example of how geometry is transformed into algebra, a continuous map of topological spaces $f : X \to Y$ is in 1-1 correspondence with a C^*-algebra homomorphism $\varphi_f : C_0(Y) \to C_0(X)$, given by the "pullback": $\varphi_f(g) := g \circ f$. We are going to exploit this idea over and over in the following pages.

2. Notice that if X is just a finite disjoint union of n points then $C_0(X) \cong \mathbb{C} \oplus \cdots \oplus \mathbb{C}$ is finite-dimensional, and if X has positive dimension then $C_0(X)$ is infinite-dimensional.

3. Now, on a vector space like $\mathbb{R}^n$ the symmetric algebra $S^\bullet(\mathbb{R}^n)$ can be interpreted as the algebra of polynomial functions on $\mathbb{R}^n$. These are dense (Stone-Weierstrass theorem) in the algebra of continuous functions $C_0(\mathbb{R}^n)$.

Algebraic geometry enhances the scope of geometry by considering more general commutative rings. Perhaps the simplest example is the "thickened point." (The technical term is "connected zero dimensional nonreduced scheme of length 2.") The "thickened point" is defined by saying that its algebra of functions is the commutative algebra $D = \mathbb{C}[\eta]/(\eta^2)$. As a vector space it is $\mathbb{C} \oplus \mathbb{C}\eta$ and the algebra structure is defined by $\eta^2 = 0$. This is an example of an algebra of functions on a "thickened point." How do we study the "thickened point" ? Let us look at maps of this "point" into affine spaces such at $\mathbb{C}^n$. Using the philosophy motivated by the mathematics mentioned above a "map from the thickened point into $\mathbb{C}^n$" is the same thing as an algebra homomorphism $\varphi : \mathbb{C}[t^1, \ldots, t^n] \to D$ where we recall that $\mathbb{C}[t^1, \ldots, t^n]$ is just the algebra of polynomials. Such a homomorphism must be of the form

$$P \mapsto \varphi(P) = \varphi_1(P) + \varphi_2(P)\eta \tag{23.146}$$

Since this is an algebra homomorphism φ_1, φ_2 are linear functionals on the algebra of polynomials and moreover $\varphi(PQ) = \varphi(P)\varphi(Q)$ implies that

$$\begin{aligned} \varphi_1(PQ) &= \varphi_1(P)\varphi_1(Q) \\ \varphi_2(PQ) &= \varphi_1(P)\varphi_2(Q) + \varphi_1(Q)\varphi_2(P) \end{aligned} \tag{23.147}$$

The first equation tells us that φ_1 is just evaluation of the polynomial at a point $\vec{t}_0 = (t_0^1, \ldots, t_0^n)$. The second is then precisely the algebraic way to define a vector field at that point! Thus

$$\varphi_2(P) := \sum_{i=1}^{n} v^i \frac{\partial}{\partial t^i} P|_{\vec{t}_0} \tag{23.148}$$

Therefore, *for every "map from the thickened point to $\mathbb{C}^n$" we associate the data of a point $\vec{t}_0 \in \mathbb{C}^n$ and a vector field at that point.* This amply justifies the term "thickened point." An obvious generalization is to consider instead the commutative algebra $D_N = \mathbb{C}[\eta]/(\eta^N)$. These give different "thickened points." Technically, this is the ring of functions on a "connected zero dimensional nonreduced scheme of length N" which we will just call a "thickened point of order $N-1$." In this case a map into $\mathbb{C}^n$ is characterized by a suitable linear functional on the set of Taylor expansion coefficients of f around some point $\vec{t}_0$.

♣?? check this last sentence. ♣

Noncommutative "geometry" develops this idea by starting with *any* (C^*-) algebra $\mathcal{A}$, not necessarily commutative, and interpreting $\mathcal{A}$ as the "algebra of functions" on some mythical "noncommutative space" and proceeding to study geometrical questions translated into algebraic questions about $\mathcal{A}$. So, for example, if $\mathcal{A} = M_n(\mathbb{C})$ is the algebra of $n \times n$ matrices then there is only one maximal ideal, and the only algebra homomorphism to $M_n(\mathbb{C}) \to \mathbb{C}$ is $\phi(M) = 0$, so $M_n(\mathbb{C})$ is the set of functions on a "space" which is a kind of "nonabelian thickened point."

23.10.2 The model superspace $\mathcal{R}^{p|q}$

Supergeometry is a generalization of algebraic geometry and a specialization of general noncommutative geometry where the algebras we use are supercommutative.

A superpoint has a real or complex algebra of functions given by a Grassmann algebra $\mathrm{Grass}[\theta^1, \ldots, \theta^q]$, depending on whether κ is $\mathbb{R}$ or $\mathbb{C}$, respectively.

Note that this algebra is just $S^\bullet(\mathbb{R}^{0|q})$ where we use the $\mathbb{Z}_2$-graded symmetric algebra. We can say there are q odd coordinates and we are considering polynomial functions of these coordinates.

This motivates the definition of the superspace $\mathcal{R}^{p|q}$ as the "space" whose super-algebra of polynomial functions is

$$S^\bullet(\mathbb{R}^{p|q}) \tag{23.149}$$

where we take the $\mathbb{Z}_2$-graded symmetric algebra. As a $\mathbb{Z}_2$-graded vector space this algebra is just

$$S^\bullet(\mathbb{R}^{p|0}) \widehat{\otimes} \Lambda^\bullet(\mathbb{R}^q) \tag{23.150}$$

(Here we view $\Lambda^\bullet(\mathbb{R}^q) = \Lambda^{\mathrm{ev}}(\mathbb{R}^q) \oplus \Lambda^{\mathrm{odd}}(\mathbb{R}^q)$ as a $\mathbb{Z}_2$-graded vector space.)

Given a choice of basis $\{\theta^1, \ldots, \theta^q\}$ of $\mathbb{R}^{0|q}$ a general super-polynomial on $\mathcal{R}^{p|q}$ can be written as

$$\Phi = \phi_0 + \phi_i \theta^i + \frac{1}{2!}\phi_{i_1 i_2}\theta^{i_1}\theta^{i_2} + \cdots + \frac{1}{n!}\phi_{i_1 \cdots i_q}\theta^{i_1}\cdots\theta^{i_q} \tag{23.151}$$

where the $\phi_{i_1,\ldots,i_m}$ are *even*, totally antisymmetric in $i_1, \ldots, i_m$, and for fixed $i_1, \ldots, i_m$ are polynomials on $\mathbb{R}^{p|0}$.

Given an *ordered* basis $\{\theta^1, \ldots, \theta^q\}$ of $\mathbb{R}^{0|q}$ we can furthermore introduce a *multi-index* $I = (i_1 < i_2 < \cdots < i_k)$ where we say I has length k, and we write $|I| = k$. We denote $I = 0$ for the empty multi-index. Then we can write

$$\Phi = \sum_I \phi_I \theta^I \tag{23.152}$$

where the ϕ_I are ordinary even polynomials on $\mathbb{R}^p$.

Similarly, we can extend these expressions by allowing the ϕ_I to be *smooth* (not just polynomial) functions on $\mathcal{R}^p$ and then we define the algebra of smooth functions on $\mathcal{R}^{p|q}$ to be the commutative superalgebra

$$\begin{aligned} \mathcal{C}^\infty(\mathcal{R}^{p|q}) &:= \mathcal{C}^\infty(\mathbb{R}^p)\widehat{\otimes} S^\bullet(\mathbb{R}^{0|q}) \\ &= \mathcal{C}^\infty(\mathbb{R}^p)[\theta^1, \ldots, \theta^q]/(\theta^i\theta^j + \theta^j\theta^i = 0) \end{aligned} \tag{23.153}$$

eq:DefineRpq

An element of $\mathcal{C}^\infty(\mathcal{R}^{p|q})$ was called by Wess and Zumino a "superfield." The idea is that we have a "function" of (x, θ) where $x = (x^1, \ldots, x^p)$ and "Taylor expansion" in the odd coordinates must terminate so

♣Actually, parities can get a little confusing if we consider things like $\mathcal{W}_\alpha$ in SYM. ♣

$$\Phi(x, \theta) = \sum_I \phi_I(x)\theta^I \tag{23.154}$$

where $\phi_I(x)$ are smooth functions of x. Trying to take this too literally can lead to confusing questions. *What is a "point" in a superspace? Can we localize a function at the coordinate* $\frac{1}{2}\theta$ *instead of* θ*? What is the "value" of a function at a point on superspace?* One way of answering such questions is explained in the remark below about the "functor of points," but often physicists just proceed with well-defined rules and get well-defined results at the end, and leave the philosophy to the mathematicians.

23.10.3 Superdomains

The official mathematical definition of a supermanifold, given below, makes use of the idea of sheaves. To motivate that we first define a *superdomain* $\mathcal{U}^{p|q}$ to be a "space" whose superalgebra of functions is analogous to (23.155):

$$\begin{aligned} \mathcal{C}^\infty(\mathcal{U}^{p|q}) &:= \mathcal{C}^\infty(U)\widehat{\otimes} S^\bullet(\mathbb{R}^{0|q}) \\ &= \mathcal{C}^\infty(U)[\theta^1, \ldots, \theta^q]/(\theta^i\theta^j + \theta^j\theta^i = 0) \end{aligned} \tag{23.155}$$

eq:DefineRpq

where $U \subset \mathbb{R}^p$ is any open set. Denote $\mathcal{O}^{p|q}(U) := \mathcal{C}^\infty(\mathcal{U}^{p|q})$. When $V \subset U$ there is a well-defined morphism of superalgebras

$$r_{U\to V} : \mathcal{O}^{p|q}(U) \to \mathcal{O}^{p|q}(V), \tag{23.156}$$

given simply by restricting from U to V the smooth functions ϕ_I on U. These morphisms are called, naturally enough, the *restriction morphisms*. Note that they are actually morphsims of superalgebras. It is often useful to denote

$$r_{U\to V}(\Phi) := \Phi|_V. \tag{23.157}$$

The restriction morphisms satisfy the following list of fairly evident properties:

1. $r_{U\to U} = Identity$.
2. $(\Phi|_V)|_W = \Phi|_W$ when $W \subset V \subset U$.
3. Suppose $U = \cup_\alpha U_\alpha$ is a union of open sets and $\Phi_1, \Phi_2 \in \mathcal{O}^{p|q}(U)$. Then if $(\Phi_1)|_{U_\alpha} = (\Phi_2)|_{U_\alpha}$ for all α we can conclude that $\Phi_1 = \Phi_2$.
4. Suppose $U = \cup_\alpha U_\alpha$ is a union of open sets and Φ_α is a collection of elements $\Phi_\alpha \in \mathcal{O}^{p|q}(U_\alpha)$. Then if, for all α, β,

$$(\Phi_\alpha)|_{U_\alpha\cap U_\beta} = (\Phi_\beta)|_{U_\alpha\cap U_\beta} \tag{23.158}$$

then we can conclude that there exists a $\Phi \in \mathcal{O}^{p|q}(U)$ such that $(\Phi)|_{U_\alpha} = \Phi_\alpha$.

23.10.4 A few words about sheaves

The properties we listed above for functions on superdomains are actually a special case of a defining list of axioms for a more general notion of a *sheaf*. Since this has been appearing in recent years in physics we briefly describe the more general concept.

Definition

a.) A *presheaf* $\mathcal{F}$ on a topological space X is an association of a set $\mathcal{F}(U)$ to every open set [56] $U \subset X$ such that there is a coherent system of restriction maps. That is, whenever $V \subset U$ there is a map $r_{U\to V}$ so that

$$r_{U,U} = \text{Identity} \qquad r_{V\to W}\circ r_{U\to V} = r_{U\to W} \qquad W \subset V \subset U \tag{23.159}$$

b.) Elements $f \in \mathcal{F}(U)$ are called *sections over* U. If $V \subset U$ we denote $r_{U\to V}(f) := f|_V$.

c.) A *sheaf* $\mathcal{F}$ on a topological space is a presheaf which moreover satisfies the two additional properties when $U = \cup_\alpha U_\alpha$ is a union of open sets:

1. If $f, g \in \mathcal{F}(U)$ and for all α, $f|_{U_\alpha} = g|_{U_\alpha}$, then $f = g$.

2. If for all α we are given $f_\alpha \in \mathcal{F}(U_\alpha)$ such that for all α, β we have $(f_\alpha)|_{U_\alpha\cap U_\beta} = (f_\beta)|_{U_\alpha\cap U_\beta}$ then there exists an $f \in \mathcal{F}(U)$ so that $f|_{U_\alpha} = f_\alpha$.

A good example is the sheaf of $\mathbb{C}^\infty$ functions on a smooth manifold. Another good example is the sheaf of holomorphic functions (The extension axiom is analytic continuation.)

[56] Technically $\emptyset$ is an open set. We should define $\mathcal{F}(\emptyset)$ to be the set with one element. If we have a sheaf of groups, then it should be the trivial group. etc.

In many common examples the sets $\mathcal{F}(U)$ in a sheaf carry some algebraic structure. Thus, we assume there is some "target" category $\mathcal{C}$ so that $\mathcal{F}(U)$ are objects in that category. So, if $\mathcal{C} = \mathbf{GROUP}$ then $\mathcal{F}(U)$ is a group for every open set, and we have a sheaf of groups; if $\mathcal{C} = \mathbf{ALGEBRA}$ is the category of algebras over κ then we have a sheaf of algebras, etc.

♣More examples, and some pictures ♣

If $\mathcal{F}$ and $\mathcal{G}$ are two sheaves on a topological spaces X then a *morphism of sheaves* is the data of a morphism (in the category $\mathcal{C}$ where the sheaf is valued) $\phi(U) : \mathcal{F}(U) \to \mathcal{G}(U)$ for every open set $U \subset X$. Note this must be a morphism in whatever target category $\mathcal{C}$ we are using. Thus, if we have a sheaf of groups, then for each U, $\phi(U)$ is a group homomorphism, and if we have a sheaf of algebras $\phi(U)$ is an algebra homomorphism, etc. Moreover, the morphisms must be compatible with restriction maps:

$$\phi(V) \circ r^{\mathcal{F}}_{U \to V} = r^{\mathcal{G}}_{U \to V} \circ \phi(U) \tag{23.160}$$

♣Write out as a commutative diagram. ♣

We can also speak of morphisms between sheaves on different topological spaces X and Y. To do this, we first define the *direct image sheaf*. Given a continuous map $\varphi : X \to Y$ and a sheaf $\mathcal{F}$ on X we can define a new sheaf $\varphi_*(\mathcal{F})$ on Y. By definition if U is an open set of Y then

$$\varphi_*(\mathcal{F})(U) := \mathcal{F}(\varphi^{-1}(U)) \tag{23.161}$$

Now a morphism of sheaves $(X, \mathcal{F}) \to (Y, \mathcal{G})$ can be defined to be a continuous map $\varphi : X \to Y$ together with a morphism of sheaves over Y, $\phi : \mathcal{G} \to \varphi_*(\mathcal{F})$.

♣Explain contravariance here. ♣

Finally, we will need the notion of the *stalk* of a sheaf at a point $\wp$. If you are familiar with directed limits then we can just write

$$\mathcal{F}(\wp) := \lim_{U : p \in U} \mathcal{F}(U) \tag{23.162}$$

What this means is that we look at sections in infinitesimal neighborhoods of $\wp$ and identify these sections if they agree. To be precise, we consider $\amalg_{U:p\in U}\mathcal{F}(U)$ and identify $f_1 \in \mathcal{F}(U_1)$ with $f_2 \in \mathcal{F}(U_2)$ if there is an open set $p \in W \subset U_1 \cap U_2$ such that $f_1|_W = f_2|_W$. So, with this equivalence relation

$$\mathcal{F}(\wp) = \amalg_{U:p\in U}\mathcal{F}(U)/\sim \tag{23.163}$$

Example: Consider the sheaf of holomorphic functions on $\mathbb{C}$. Then the stalk at z_0 can be identified with the set of formal power series expansions at z_0. For the sheaf of $\mathcal{C}^\infty$ functions on a manifold the stalk at $\wp$ is just $\mathbb{R}$.

Finally, with these definitions we can say that the superdomains $\mathcal{U}^{p|q}$ defined above describe a *sheaf of Grassmann algebras* $\mathcal{O}^{p|q}$ with value on an open set $U \subset \mathbb{R}^p$ given by the Grassmann algebra

$$\mathcal{O}^{p|q}(U) = \mathcal{C}^\infty(\mathcal{U}^{p|q}) = \mathcal{C}^\infty(U)\widehat{\otimes}S^\bullet(\mathbb{R}^{0|q}) \tag{23.164}$$

A *super-change of coordinates* is an invertible morphism of the sheaf $\mathcal{O}^{p|q}$ with itself. Concretely it will be given by an expression like

$$\begin{aligned} \tilde{t}^a &= f^a(t^1, \ldots, t^p|\theta^1, \ldots, \theta^q) \qquad & a = 1, \ldots, p \\ \tilde{\theta}^i &= \psi^i(t^1, \ldots, t^p|\theta^1, \ldots, \theta^q) \qquad & i = 1, \ldots, q \end{aligned} \tag{23.165}$$

where f^a are even elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ and ψ^i are odd elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$, respectively.

Exercise *Alternative definition of a presheaf*

Given a topological space X define a natural category whose objects are open sets $U \subset X$ and whose morphisms are inclusions of open sets.

Given any category $\mathcal{C}$ show that a presheaf with values in $\mathcal{C}$ can be defined as a contravariant functor from the category of open sets in X to $\mathcal{C}$.

Exercise

Show that the stalk of $\mathcal{O}^{p|q}$ at a point $\wp \in \mathbb{R}^p$ is the finite-dimensional Grassmann algebra $\Lambda^*(\mathbb{R}^q)$ over $\mathbb{R}$.

23.10.5 Definition of supermanifolds

One definition of a supermanifold is the following:

Definition A supermanifold M of dimension $(p|q)$ is an ordinary manifold $M_{\rm red}$ with a sheaf $\mathcal{F}$ of Grassmann algebras which is locally equivalent to the supermanifold $\mathcal{R}^{p|q}$. That is, near any $p \in M_{\rm red}$ there is a neighborhood $p \in U$ so that the restriction of the sheaf to U is equivalent (isomorphism of sheaves) to a superdomain $\mathcal{U}^{p|q}$.

In this definition $M_{\rm red}$ is the called the "reduced space" or the "body" of the supermanifold. The sheaf $\mathcal{F}$ of Grassmann algebras has a subsheaf $\mathcal{I}^{\rm odd}$ generated by the odd elements and the quotient sheaf $\mathcal{F}/\mathcal{I}^{\rm odd}$ is the sheaf of $\mathcal{C}^\infty$ functions of the reduced space $M_{\rm red}$.

There is a second (equivalent) definition of supermanifolds which strives to make a close parallel to the definition of manifolds in terms of atlases of charts.

We choose a manifold M of dimension p and define a *super-chart* to be a pair $(\mathcal{U}^{p|q}, c)$ where $c : U \to M$ is a homeomorphism. Then a supermanifold will be a collection of supercharts $(\mathcal{U}_\alpha^{p|q}, c_\alpha)$ so that if $c_\alpha(U_\alpha) \cap c_\beta(U_\beta) = \hat{U}_{\alpha\beta}$ is nonempty then there is a

change of coordinates between coordinates $(t^1_\alpha, \ldots, t^p_\alpha | \theta^1_\alpha, \ldots, \theta^q_\alpha)$ on $\mathcal{O}^{p|q}(c_\alpha^{-1}(\hat{U}_{\alpha\beta}))$ and $(t^1_\beta, \ldots, t^p_\beta | \theta^1_\beta, \ldots, \theta^q_\beta)$ on $\mathcal{O}^{p|q}(c_\beta^{-1}(\hat{U}_{\alpha\beta}))$ given by a collection of functions:

$$\begin{aligned} t^a_\alpha &= f^a_{\alpha\beta}(t^1_\beta, \ldots, t^p_\beta | \theta^1_\beta, \ldots, \theta^q_\beta) \qquad & a = 1, \ldots, p \\ \theta^i_\alpha &= \psi^i_{\alpha\beta}(t^1_\beta, \ldots, t^p_\beta | \theta^1_\beta, \ldots, \theta^q_\beta) \qquad & i = 1, \ldots, q \end{aligned} \tag{23.166}$$

where $f^a_{\alpha\beta}$ are even elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ and $\psi^i_{\alpha\beta}$ are odd elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$, respectively. These maps need to be invertible, in an appropriate sense, and they need to satisfy a version of the cocycle identity when there are nonempty triple overlaps $c_\alpha(U_\alpha) \cap c_\beta(U_\beta) \cap c_\gamma(U_\gamma)$. For a more careful discussion see Chapter III of Leites.

Example A good example of a nontrivial supermanifold is super-complex projective space $\mathbb{C}P^{m|n}$. The reduced manifold is just $\mathbb{C}P^m$. Recall that $\mathbb{C}P^m$ is the space of complex lines in $\mathbb{C}^{m+1}$ and can be thought of as the set of nonzero points $(X^0, \ldots, X^m) \in \mathbb{C}^{m+1}$ modulo the scaling action $X^A \to \lambda X^A$. We denote the equivalence class by $[X^0 : \cdots : X^m]$. Informally, we can define $\mathbb{C}P^{m|n}$ as the "set of points" $(X^0, \ldots, X^m | \theta^1, \ldots, \theta^n) \in \mathbb{C}^{m+1|n}$ with $(X^0, \ldots, X^m) \neq 0$ again with identification by scaling

$$(X^0, \ldots, X^m | \theta^1, \ldots, \theta^n) \sim \lambda (X^0, \ldots, X^m | \theta^1, \ldots, \theta^n) \tag{23.167}$$

To make proper sense of this we could define a standard superatlas by choosing the usual atlas on $\mathbb{C}P^m$ defined by the nonvanishing of one of the homogeneous coordinates

$$U_\alpha := \{ [X^0 : X^1 : \cdots : X^m] | X^\alpha \neq 0 \} \qquad \alpha = 0, \ldots, p \tag{23.168}$$

so local coordinates are given by $t^A_\alpha := X^A / X^\alpha$. (Note that $t^\alpha_\alpha = 1$ is not a coordinate, so coordinates are given by $A = 1, \ldots, m$ omitting α.) Then the supermanifold has

$$\mathcal{F}(U_\alpha) = \mathcal{C}^\infty(U_\alpha)[\theta^1_\alpha, \ldots, \theta^n_\alpha] / (\theta^i_\alpha \theta^j_\alpha + \theta^j_\alpha \theta^i_\alpha = 0) \tag{23.169}$$

and on $U_{\alpha\beta}$ we have the change of coordinates

$$\begin{aligned} t^A_\alpha &= \frac{X^\beta}{X^\alpha} t^A_\beta \qquad & A = 0, \ldots, m \\ \theta^i_\alpha &= \frac{X^\beta}{X^\alpha} \theta^i_\beta \qquad & i = 1, \ldots, n \end{aligned} \tag{23.170}$$

(You can put $A = \alpha$ to learn that $\frac{X^\beta}{X^\alpha} = 1/t^\alpha_\beta$ to get the honest formula.)

Now we can go on to produce more nontrivial examples of supermanifolds by choosing homogeneous even polynomials $P(X^A | \theta^i)$ and dividing the sheaf by the ideal generated by these.

For example, see [57] for an interesting discussion of the sub-supermanifold of $\mathbb{C}P^{2|2}$ defined by

$$X_1^2 + X_2^2 + X_3^2 + \theta_1 \theta_2 = 0 \tag{23.171}$$

Remarks

[57] E. Witten, "Notes on Supermanifolds and Integration," arXiv:1209.2199, Section 2.3.1

1. In general there is no reality condition put on the odd generators θ^i. Therefore, it is natural to consider a supermanifold $\mathbb{R}^{p|*q}$ where the ring of functions can be expanded as above but only ϕ_0 is real, and all the other ϕ_I with $|I| > 0$ are complex polynomials. Gluing these together gives a *cs-supermanifold.*

2. *More philosophy: The functor of points.* There is a way of speaking about "points of a supermanifold" which is a generalization of a standard concept in algebraic geometry. We first give the background in algebraic geometry. For simplicity we just work with a ground field $\kappa = \mathbb{C}$. There is a generalization of algebraic varieties known as "schemes." Again we characterize them locally by their algebras of "polynomial functions," but now we are allowed to introduce nilpotents as in the "thickened point" example discussed above. To characterize the "points" on a scheme X we probe it by taking an arbitrary scheme S and consider the set of all morphisms of schemes $\mathrm{Hom}(S, X)$. The set $\mathrm{Hom}(S, X)$ is, roughly speaking, just the homomorphisms from the algebra of functions on X to the algebra of functions on S. In this context the set $\mathrm{Hom}(S, X)$ is called *the set of S-points of X*. Now, the map $X \mapsto \mathrm{Hom}(S, X)$ is (contravariantly) "functorial in S." This means that if $f : S \to S'$ is a morphism of schemes then there is a natural morphism of sets $F_X(f) : \mathrm{Hom}(S', X) \to \mathrm{Hom}(S, X)$. Therefore, given a scheme X, there is a functor F_X from the category of all schemes [58] to the category of sets, $F_X : \mathbf{SCHEME}^{\mathrm{opp}} \to \mathbf{SET}$ defined on objects by

$$F_X : S \mapsto \mathrm{Hom}(S, X) \tag{23.172}$$

This functor is called the *functor of points.* If we let the "probe scheme" S be a point then its algebra of functions is just $\mathbb{C}$ and $F_X(S) = \mathrm{Hom}(S, X)$ is the set of algebra homomorphisms from functions on X to $\mathbb{C}$. That is, indeed the set of points of the underlying topological space X_{red} of X. More generally, if S is an ordinary algebraic manifold then we should regard $F_X(S)$ as a set of points in X parametrized by S. Of course, we could probe the scheme structure of X more deeply by using a more refined probe, such as a nonreduced point of order N described above. In fact, if we use too few probe schemes S we might miss structure of X, therefore mathematicians use the functor from *all* schemes S to sets. Now, a key theorem justifying this approach (known as the *Yoneda theorem*) states that:

Two schemes X and X' are isomorphic as schemes iff there is a natural transformation between the functors F_X and $F_{X'}$.

Now, we can apply all these ideas to supermanifolds with little change. If M is a supermanifold, and $S = \mathbb{R}^{0|0}$ then the set of S-points of M is precisely the set of points of the underlying manifold M_{red}.

We will not go very deeply into supermanifold theory but we do need a notion of vector fields:

[58] actually, the opposite category $\mathbf{SCHEME}^{\mathrm{opp}}$

23.10.6 Supervector fields and super-differential forms

In the ordinary theory of manifolds the space of vector fields on the manifold is in 1-1 correspondence with the derivations of the algebra of functions. The latter concept makes sense for supermanifolds, provided we take $\mathbb{Z}_2$-graded derivations, and is taken to define the super-vector-fields on a supermanifold.

For $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ the space of derivations is a left supermodule for $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ generated by

$$\frac{\partial}{\partial t^1}, \dots, \frac{\partial}{\partial t^p}, \frac{\partial}{\partial \theta^1}, \dots, \frac{\partial}{\partial \theta^q} \tag{23.173}$$

where we have to say whether the odd derivatives act from the left or the right. We will take them to act from the left so, for example, if $q = 2$ then

$$\begin{aligned} \frac{\partial}{\partial \theta^1}\Phi &= \phi_1 + \phi_{12}\theta^2 \\ \frac{\partial}{\partial \theta^2}\Phi &= \phi_2 - \phi_{12}\theta^1 \end{aligned} \tag{23.174}$$

In general, for any $i \in \{1, \dots, q\}$ we can write always expand Φ in the form $\Phi = \sum_{I: i \notin I}(\phi_I \theta^I + \phi_{i,I}\theta^a\theta^I)$ and then

$$\frac{\partial}{\partial \theta^i}\Phi = \sum_{I: i \notin I} \phi_{i,I}\theta^I \tag{23.175}$$

A simple, but important lemma says that the $\mathcal{O}^{p|q}$ module of derivations is free and of dimension $(p|q)$.

Defining differential forms turns out to be surprisingly subtle. The problems are related to how one defines a grading of expressions like $d\theta^i$ and, related to this, how one defines the exterior derivative. There are two (different) ways to do this.

One way to proceed is to consider the "stalk" of the tangent sheaf $T\mathcal{R}^{p|q}$ at a point $\wp$. This is a module for the real Grassmann algebra $\mathrm{Grass}[\theta^1, \dots, \theta^q]$. (That is, the coefficients of $\frac{\partial}{\partial t^a}$ and $\frac{\partial}{\partial \theta^i}$ are functions of θ^i but not of the t^a, because we restricted to a point $\wp$.) The dual module is denoted $\Omega^1\mathcal{R}^{p|q}(\wp)$. There is an even pairing

$$T\mathcal{R}^{p|q}(\wp) \otimes \Omega^1\mathcal{R}^{p|q}(\wp) \to \mathcal{O}^{p|q}(\wp) \tag{23.176}$$

The pairing is denoted $\langle v, \omega \rangle$ and if Φ_1, Φ_2 are superfunctions then

$$\langle \Phi_1 v, \Phi_2 \omega \rangle = (-1)^{|v||\Phi_2|}\Phi_1\Phi_2\langle v, \omega \rangle \tag{23.177}$$

If we have a system of coordinates $(t^a|\theta^i)$ then $\Omega^1\mathcal{R}^{p|q}(\wp)$ is a free module of rank $(p|q)$ generated by symbols dt^a and $d\theta^i$. Thus,

$$\langle \frac{\partial}{\partial t^a}, dt^b \rangle = \delta^b_a \qquad\qquad \langle \frac{\partial}{\partial \theta^i}, d\theta^j \rangle = \delta^i_j \tag{23.178}$$

and so forth.

Now, to define the differential forms at the point $\wp$ we take the exterior algebra of $\Omega^1\mathcal{R}^{p|q}(\wp)$ to define

$$\Omega^\bullet\mathcal{R}^{p|q}(\wp) := \Lambda^\bullet(\Omega^1\mathcal{R}^{p|q}(\wp)) \tag{23.179}$$

where - very importantly - we are using the $\mathbb{Z}_2$-graded antisymmetrization to define $\Lambda^\bullet$. Thus, the generators dt^a are *anti-commuting* (as usual) while the generators $d\theta^i$ are *commuting*. The stalks can be used to define a sheaf $\Omega^\bullet\mathcal{R}^{p|q}$ and the general section in $\Omega^\bullet\mathcal{R}^{p|q}(U)$ is an expression:

$$\sum_{k=0}^{p}\sum_{\ell=0}^{\infty}\omega_{a_1,\ldots,a_k;i_1,\ldots,i_\ell}dt^{a_1}\cdots dt^{a_k}d\theta^{i_1}\cdots d\theta^{i_\ell} \tag{23.180}$$

eq:GenSuperFor

where $\omega_{a_1,\ldots,a_k;i_1,\ldots,i_\ell}$ are elements of $\mathcal{O}^{p|q}(U)$ which are totally antisymmetric in the $a_1,\ldots,a_k$ and totally symmetric in the $i_1,\ldots,i_\ell$.

If we consider dt^a to be odd and $d\theta^i$ to be even then expressions such as (23.180) can be multiplied. Then, finally, we can define an exterior derivative by saying that $d:\mathcal{O}^{p|q}\to\Omega^1\mathcal{R}^{p|q}$ takes $d:t^a\mapsto dt^a$ and $d:\theta^i\mapsto d\theta^i$ and then we impose the super-Leibniz rule

$$d(\omega_1\omega_2)=d\omega_1\omega_2+(-1)^{|\omega_1|}\omega_1 d\omega_2 \tag{23.181}$$

It is still true that $d^2=0$ and we have the Super-Poincaré lemma: If $d\omega=0$ in $\Omega^\bullet\mathcal{R}^{p|q}$ then $\omega=d\eta$.

Remarks

1. We are following the conventions of Witten's paper cited below. For a nice interpretation of the differential forms on a supermanifold in terms of Clifford and Heisenberg modules see Section 3.2. Note that with the above conventions

$$d(\theta^1\theta^2)=\theta^2 d\theta^1-\theta^1 d\theta^2 \tag{23.182}$$

2. However, there is another, equally valid discussion which is the one taken in Deligne-Morgan. The superderivations define a sheaf of super-modules for the sheaf $\mathcal{O}^{p|q}$ and it is denoted by $T\mathcal{R}^{p|q}$. Then the cotangent sheaf, denoted $\Omega^1\mathcal{R}^{p,q}$ is the dual module with an even pairing:

♣dbend [latex command doesn't work unfortunately ♣

$$T\mathcal{R}^{p|q}\otimes\Omega^1\mathcal{R}^{p|q}\to\mathcal{O}^{p|q} \tag{23.183}$$

The pairing is denoted $\langle v,\omega\rangle$ and if Φ_1,Φ_2 are superfunctions then

$$\langle\Phi_1 v,\Phi_2\omega\rangle=(-1)^{|v||\Phi_2|}\Phi_1\Phi_2\langle v,\omega\rangle \tag{23.184}$$

It we have a system of coordinates $(t|\theta)$ then Ω^1 freely generated as an $\mathcal{O}^{p|q}$-module by dt^a and $d\theta^i$.

Now we define a differential $d:\mathcal{O}^{p|q}\to\Omega^1\mathcal{R}^{p|q}$ to be an *even* morphism of sheaves of super-vector spaces by

$$\langle v,df\rangle:=v(f) \tag{23.185}$$

In particular this implies

$$d(\theta^1\theta^2) = -\theta^2 d\theta^1 + \theta^1 d\theta^2 \tag{23.186}$$

The issue here is that $\Omega^\bullet\mathcal{R}^{p|q}$ is really *bigraded* by the group $\mathbb{Z}\oplus\mathbb{Z}_2$. It has "cohomological degree" in $\mathbb{Z}$ coming from the degree of the differential form in addition to "parity." In general, given vector spaces which are $\mathbb{Z}$-graded *and also* $\mathbb{Z}_2$-graded, so that $V = V^0 \oplus V^1$ as a super-vector-space, and V^0 and V^1 are also $\mathbb{Z}$-graded, then there are *two* conventions for defining the commutativity morphism:

$$c_{V,W} : V\otimes W \to W\otimes V \tag{23.187}$$

1. We have the modified Koszul rule: $c_{V,W} : v\otimes w \mapsto (-1)^{(|v|+\deg(v))(|w|+\deg(w))} w\otimes v$, where $\deg(v), \deg(w)$ refer to the integer grading.

2. We have the modified Koszul rule: $c_{V,W} : v\otimes w \mapsto (-1)^{|v||w|+\deg(v)\deg(w)} w\otimes v$.

In convention 1 we have simply taken a homomorphism of the $\mathbb{Z}\oplus\mathbb{Z}_2$ grading to $\mathbb{Z}_2$. In our notes we have adopted the first convention in making d odd. This makes $d\theta^i$ even because we sum the degree of d (which is one) with the degree of θ^i (which is one modulo two) to get zero, modulo two. In the second convention it would still be true that the $d\theta^i$ commute with the $d\theta^j$, but for a different reason. Convention 2 is adopted in Deligne-Morgan, for reasons explained on p.62.

♣Should discuss this more. What are the $\mathbb{Z}_2$-valued bilinear forms on $\mathbb{Z}\oplus\mathbb{Z}_2$? ♣

Exercise

Suppose that M is a manifold and TM is its tangent bundle. Let ΠTM be the supermanifold where the Grassmann algebra is the Grassmann algebra of the sections of TM.

Show that the $\mathbb{C}^\infty$ functions on TM can be identified with the DeRham complex of the bosonic manifold $\Omega^\bullet$ and, under this correspondence, write d as a supervector field on $\mathbb{C}^\infty(\Pi TM)$.

This observation is often used in applications of supersymmetric field theory to topological invariants.

Exercise

a.) Show that the graded commutator of super-derivations is a superderivation.

b.) Consider the odd vector fields $D = \frac{\partial}{\partial\theta} + \theta\frac{\partial}{\partial t}$ and $Q = \frac{\partial}{\partial\theta} - \theta\frac{\partial}{\partial t}$ on $\mathcal{R}^{1|1}$. Compute $[D,D]$, $[Q,Q]$, and $[Q,D]$.

ANOTHER EXERCISE WITH MORE THETAS.

23.11 Integration over a superdomain

In this section we will say something about how to define an integral of superfunctions on the supermanifold $\mathcal{R}^{p|q}$.

As motivation we again take inspiration from the theory of commutative C^*-algebras. A beautiful theorem - the Riesz-Markov theorem - says that if $\mathcal{A}$ is a commutative C^*-algebra and $\Lambda : \mathcal{A} \to \mathbb{C}$ is a linear functional then there is a (complex-valued) measure $d\mu$ on $X = \Delta(\mathcal{A})$ so that this linear functional is just

$$\Lambda(f) = \int_X f d\mu \tag{23.188}$$

(Recall that $f \in \mathcal{A}$ is canonically a function on X, so the expression on the RHS makes sense.)

So, we will view an integral over $\mathcal{R}^{p|q}$ as a linear functional

$$\Lambda : \mathcal{O}^{p|q}(\mathbb{R}^p) \to \mathbb{R} \tag{23.189}$$

To guide us in our discussion there are three criteria we want from our integral:

1. We want integration by parts (Stokes' theorem) to be valid.
2. We want the Fubini theorem to be valid.
3. We want the definition to reduce to the usual Riemannian integration when $q = 0$.

Let us begin with $p = 0$, the fermionic point. For brevity denote $\mathcal{O}^q := \mathcal{O}^{0|q}(pt) = S^\bullet(\mathbb{R}^{0|q})$. The space of linear functionals

$$\mathcal{D}^q := \mathrm{Hom}(\mathcal{O}^q, \mathbb{R}) \tag{23.190}$$

is a real supervector space of dimension $(2^{q-1}|2^{q-1})$. Indeed, given an *ordered* basis $\theta^1, \ldots, \theta^q$ for $\mathbb{R}^{0|q}$ there is a canonical dual basis δ_I for $\mathcal{D}^q$ defined by $\delta_I(\theta^J) = \delta_I{}^J$ where I, J are multi-indices.

On the other hand, $\mathcal{D}^q$ is also a right $\mathcal{O}^q$-module since if Λ is a linear functional and $g \in \mathcal{O}^q$ we can define

$$(\Lambda \cdot g)(f) := \Lambda(gf) \tag{23.191}$$

It is important to distinguish $\mathcal{D}^q$ as a vector space over $\mathbb{R}$ from $\mathcal{D}^q$ as a module over the supercommutative superalgebra $\mathcal{O}^q$. In the latter case, $\mathcal{D}^q$ is free and of dimension $(1|0)$ or $(0|1)$.

For example, suppose $q = 2$ and we choose an ordered basis $\{\theta^1, \theta^2\}$ for $\mathbb{R}^{0|2}$. Then let $\delta = \delta_{12}$. Then

$$\begin{aligned} \delta &= \delta_{12} \\ \delta \cdot \theta^1 &= \delta_2 \\ \delta \cdot \theta^2 &= -\delta_1 \\ \delta \cdot \theta^1\theta^2 &= \delta_0 \end{aligned} \tag{23.192}$$

In general, given an ordered basis, $\delta = \delta_I$, where I is the multi-index $I = 12\dots q$, is a basis vector for $\mathcal{D}^q$ as an $\mathcal{O}^q$-module: Indeed, right-multiplication by elements of $\mathcal{O}^q$ gives a vector space basis over $\mathbb{R}$ as follows:

$$\begin{aligned} \delta &= \delta_{1\dots q} \\ (\delta\cdot\theta^i) &= (-1)^{i-1}\delta_{1\dots\hat{i}\dots q} \\ (\delta\cdot\theta^i\theta^j) &= \pm\delta_{1\dots\hat{i}\dots\hat{j}\dots q} \\ \vdots &\qquad \vdots \end{aligned} \tag{23.193}$$

Moreover, since the scalar 1 is even δ has degree $q\mathrm{mod}2$ and hence, as a right $\mathcal{O}^q$-module, $\mathcal{D}^q$ has parity $q\mathrm{mod}2$, so it is a free module of type $(\mathcal{O}^q)^{1|0}$ or $(\mathcal{O}^q)^{0|1}$, depending on whether q is even or odd, respectively. Of course, if N_q is a nonzero real number then $N_q\delta$ is also a perfectly good generator of $\mathcal{D}^q$ as a $\mathcal{O}^q$-module.

Now we claim that, given an ordered basis $\{\theta^1,\dots,\theta^q\}$ for $\mathbb{R}^{0|q}$ there is a canonical generator for $\mathcal{D}^q$ which we will denote by

$$\Lambda_q = \int [d\theta^1\cdots d\theta^q] \tag{23.194}$$ eq:CanonicalMe

The notation is apt because this functional certainly satisfies the integration-by-parts property:

$$\int [d\theta^1\cdots d\theta^q]\frac{\partial f}{\partial\theta^i} = 0 \tag{23.195}$$

for any i and and f. Thus, criterion 1 above is automatic in our approach.

However, the integration-by-parts property is satisfied by *any* generator of $\mathcal{D}^q$ as an $\mathcal{O}^q$-module, that is, it is satisfied by any nonzero multiple of Λ_q. How should we normalize Λ_q? We can answer this question by appealing to criterion 2. That is, we require an analog of the Fubini theorem. There is a canonical isomorphism $\mathcal{R}^{0|q_1}\times\mathcal{R}^{0|q_2}\cong\mathcal{R}^{0|q_1+q_2}$, that is there are canonical isomorphism $\mathcal{O}^{q_1}\widehat{\otimes}\mathcal{O}^{q_2}\cong\mathcal{O}^{q_1+q_2}$ (simply given by multiplying the polynomials) and hence canonical isomorphisms

$$\mathcal{D}^{q_1}\widehat{\otimes}\mathcal{D}^{q_2}\cong\mathcal{D}^{q_1+q_2} \tag{23.196}$$

given by

$$(\ell_1\widehat{\otimes}\ell_2)(f_1\widehat{\otimes}f_2) = (-1)^{|\ell_2||f_1|}\ell_1(f_1)\ell_2(f_2) \tag{23.197}$$

Now we require that our canonical integrals Λ_q satisfy

$$\begin{array}{ccc} \mathcal{D}^{q_1}\otimes\mathcal{D}^{q_2} & \xrightarrow{\ \cong\ } & \mathcal{D}^{q_1+q_2} \\ & {\scriptstyle\Lambda_{q_1}\otimes\Lambda_{q_2}}\searrow \quad \swarrow{\scriptstyle\Lambda_{q_1+q_2}} & \\ & \mathbb{R} & \end{array} \tag{23.198}$$ eq:Compat-Lamb

Let $\Lambda_q(\theta^1\cdots\theta^q) := N_q$. Then (23.198) implies that

$$\begin{aligned} N_{q_1+q_2} &= \Lambda_{q_1+q_2}(\theta^1\cdots\theta^{q_1+q_2}) \\ &= (\Lambda_{q_1}\otimes\Lambda_{q_2})(\theta^1\cdots\theta^{q_1}\otimes\theta^{q_1+1}\cdots\theta^{q_1+q_2}) \\ &= (-1)^{q_1q_2}N_{q_1}N_{q_2} \end{aligned} \tag{23.199}$$

The general solution to the equation $N_{q_1+q_2} = (-1)^{q_1 q_2} N_{q_1} N_{q_2}$ is

$$N_q = (-1)^{\frac{1}{2}q(q-1)} (N_1)^q \tag{23.200}$$

So this reduces the question to $q = 1$.

It is customary and natural to normalize the integral so that

$$\int [d\theta]\theta = 1 \tag{23.201}$$

That is, $N_1 = 1$. With this normalization, the Berezin integral on $\mathcal{R}^{0|1}$ is the functional:

$$\int [d\theta](a + b\theta) = b \tag{23.202}$$

Now, for $q > 1$, noticing that $\theta^1 \cdots \theta^q = (-1)^{\frac{1}{2}q(q-1)} \theta^q \cdots \theta^1$ we have shown that demanding that the integral satisfy the "Fubini theorem" (as interpreted above) normalizes the canonical measure so that

$$\boxed{\int [d\theta^1 \cdots d\theta^q] \theta^q \cdots \theta^1 = +1} \tag{23.203}$$

Now let us consider $p > 0$. Criterion 3 above tells us that we don't want our integrals to be literally all linear functionals $\mathcal{O}^{p|q}(\mathbb{R}^p) \to \mathbb{R}$. For example, that would include distributions in the bosonic variables. So we have the official definition

Definition: A *density* on the superspace $\mathcal{R}^{p|q}$ is a linear functional $\mathcal{O}^{p|q}(\mathbb{R}^p) \to \mathbb{R}$ of the form

$$\Phi = \sum_I \phi_I \theta^I \mapsto \sum_I \int_{\mathbb{R}^p} [dt^1 \cdots dt^p] d_I(t) \phi_I(t) \tag{23.204}$$

where $[dt^1 \cdots dt^p]$ is the standard Riemann measure associated with a coordinate system $(t^1, \ldots, t^p)$ for $\mathbb{R}^p$ and $d_I(t)$ are some collection of smooth functions. We denote the space of densities on $\mathcal{R}^{p|q}$ by $\mathcal{D}^{p|q}$.

When $p = 0$ this reduces to our previous description, and $\mathcal{D}^{0|q} = \mathcal{D}^q$. Now, analogous to the previous discussion, $\mathcal{D}^{p|q}$ is once again a $\mathcal{O}^{p|q}(\mathbb{R}^p)$-module of rank $(1|0)$ or $(0|1)$, depending on whether $q = 0 \mathrm{mod} 2$ or $q = 1 \mathrm{mod} 2$, respectively. Once again, given an ordered coordinate system $(t^1, \ldots, t^p | \theta^1, \ldots, \theta^q)$ for $\mathcal{R}^{p|q}$ we have a canonically normalized density which we denote

$$\int [dt^1 \cdots dt^p | d\theta^1 \cdots d\theta^q] \tag{23.205}$$

defined by

$$\int [dt^1 \cdots dt^p | d\theta^1 \cdots d\theta^q] \Phi := \int_{\mathbb{R}^p} [dt^1 \cdots dt^p] \left(\int [d\theta^1 \cdots d\theta^q] \Phi \right) \tag{23.206}$$

where $[dt^1 \cdots dt^p]$ is the Riemannian measure. Thus, we first integrate over the odd coordinates and then over the reduced bosonic coordinates.

♣The conventions on indices t^i, θ^a are opposite to those of the previous section. Use consistent conventions. ♣

Finally, let us give the change of variables formula. Suppose $\mu : \mathcal{R}^{p|q} \to \mathcal{R}^{p|q}$ is an invertible morphism. Then we can define new "coordinates:

$$\begin{aligned} \tilde{t}^i &= \mu^*(t^i) \qquad\qquad i = 1, \dots, p \\ \tilde{\theta}^a &= \mu^*(\theta^a) \qquad\qquad a = 1, \dots, q \end{aligned} \tag{23.207}$$

Then, again because $\mathcal{D}^{p|q}$ is one-dimensional as an $\mathcal{O}^{p|q}(\mathbb{R}^p)$-module, we know that there is an even invertible element of $\mathcal{O}^{p|q}(\mathbb{R}^p)$ so that

$$\int [d\tilde{t}^1 \cdots d\tilde{t}^p | d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] \Phi(\tilde{t}|\tilde{\theta}) = \int [dt^1 \cdots dt^p | d\theta^1 \cdots d\theta^q] j(\mu) \mu^* \Phi \tag{23.208}$$

where $\mu^*\Phi$ is a function of $(t|\theta)$ given by $\mu^*\Phi = \Phi(\tilde{t}(t|\theta)|\tilde{\theta}(t|\theta))$.

Some special cases will make the general formula clear:

1. If $\mu : \mathbb{R}^p \to \mathbb{R}^p$ is an ordinary diffeomorphism then it can be lifted to a superdiffeomorphism just by setting $\mu^*(\theta^a) = \theta^a$ and $\mu^*(t^i) = \tilde{t}^i$. Then the standard change-of-variables result says that

$$\begin{aligned} [d\tilde{t}^1 \cdots d\tilde{t}^p] &= [dt^1 \cdots dt^p] \cdot \left| \det \frac{\partial \tilde{t}^i}{\partial t^j} \right| \\ &= [dt^1 \cdots dt^p] \cdot \mathrm{or}(\mu) \cdot \det \frac{\partial \tilde{t}^i}{\partial t^j} \end{aligned} \tag{23.209}$$

where $\mathrm{or}(\mu) = +1$ if μ is orientation preserving and $\mathrm{or}(\mu) = -1$ if it is orientation reversing.

2. On the other hand, if $\mu^*(t^i) = t^i$ and $\mu^*(\theta^a) = D^a{}_b \theta^b$ then

$$\begin{aligned} \tilde{\theta}^q \cdots \tilde{\theta}^1 &= D^q{}_{b_1} \cdots D^1{}_{b_q} \theta^{b_1} \cdots \theta^{b_q} \\ &= \left(\sum_{\sigma \in S_q} \epsilon(\sigma) D^q{}_{\sigma(q)} \cdots D^1{}_{\sigma(1)} \right) \theta^q \cdots \theta^1 \\ &= \det(D^a{}_b) \theta^q \cdots \theta^1 \end{aligned} \tag{23.210}$$

and therefore

$$[d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] = [d\theta^1 \cdots d\theta^q] (\det(D^a{}_b))^{-1} \tag{23.211}$$

For the general formula we consider the Jacobian

$$\mathrm{Jac}(\mu) = \begin{pmatrix} \frac{\partial \tilde{t}^i}{\partial t^j} & \frac{\partial \tilde{t}^i}{\partial \theta^b} \\ \frac{\partial \tilde{\theta}^a}{\partial t^j} & \frac{\partial \tilde{\theta}^a}{\partial \theta^b} \end{pmatrix} \tag{23.212}$$

which we regard as an element of $\mathrm{End}(\Omega^1 \mathcal{R}^{p|q})$. (Recall that $\Omega^1 \mathcal{R}^{p|q}(U)$ is a free module of rank $(p|q)$ over $\mathcal{O}^{p|q}(U)$.) The formula is

$$j(\mu) = \mathrm{or}(\mu_{\mathrm{red}}) \mathrm{Ber}(\mathrm{Jac}(\mu)) \tag{23.213}$$

♣Needs a full proof.
♣

Example: Consider $\mathcal{R}^{1|2}$. Let $\Phi(\tilde{t}|\tilde{\theta}) = h(\tilde{t})$. Change variables by $\tilde{t} = t + \theta^1\theta^2$ and $\tilde{\theta}^i = \theta^i$. Then

$$\begin{aligned} 0 &= \int [d\tilde{t}|d\tilde{\theta}] h(\tilde{t}) \\ &= \int [dt|d\theta] \left(h(t) + h'(t)\theta^1\theta^2 \right) \\ &= -\int dt \frac{\partial h}{\partial t} \end{aligned} \tag{23.214}$$

eq:SuperIntExp

Note that this identity relies on the validity of integration by parts.

Remarks

1. Note well that $d\theta^a$ are commutative objects but $[d\theta^a]$ are anti-commutative objects in the sense that $[d\theta^1 d\theta^2] = -[d\theta^2 d\theta^1]$, and so on.

2. The possible failure of boundary terms to vanish in examples like 23.214 leads to important subtleties in string perturbation theory. On a supermanifold it might not be possible to say, globally, which even variables are "purely bosonic," that is, "free of nilpotents." This is related to the issue of whether the supermanifold is "split" or not. For recent discussions of these problems see Witten, arXiv:1304.2832, 1209.5461.

23.12 Gaussian Integrals

23.12.1 Reminder on bosonic Gaussian integrals

Let Q_{ij} be a symmetric quadratic form with positive definite real part on $\mathbb{R}^p$. Then the Gaussian integral over $\mathcal{R}^{p|0}$ is

$$(2\pi)^{-p/2} \int [dt^1 \cdots dt^p] \exp[-\frac{1}{2} t^i Q_{ij} t^j] = \frac{1}{(\det Q)^{1/2}} \tag{23.215}$$

where we choose the sign of the square root so that $(\det Q)^{1/2}$ is in the positive half-plane, i.e., we choose the principal branch of the logarithm.

One could analytically continue in Q from this result.

23.12.2 Gaussian integral on a fermionic point: Pfaffians

Let us now consider the Gaussian integral over a fermionic point $\mathcal{R}^{0|q}$.

Let A_{ij} be a $q \times q$ antisymmetric matrix. Consider the Gaussian integral:

$$\int [d\theta^1 \cdots d\theta^q] \exp[\frac{1}{2} \theta^a A_{ab} \theta^b] \tag{23.216}$$

Our first observation is that if q is odd then this integral must vanish! To see this, we recall that we can always skew-diagonalize A:

$$SAS^{tr} = \begin{pmatrix} 0 & \lambda_1 \\ -\lambda_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \lambda_2 \\ -\lambda_2 & 0 \end{pmatrix} \oplus \cdots \tag{23.217}$$

By the change-of-variable formula if we change coordinates $\tilde{\theta}^a = S^a{}_b \theta^b$ then the integral is

$$\begin{aligned} \int [d\theta^1 \cdots d\theta^q] \exp[\frac{1}{2}\theta^a A_{ab} \theta^b] &= \det S \int [d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] \exp[\frac{1}{2}\tilde{\theta}^a \tilde{A}_{ab} \tilde{\theta}^b] \\ &= \det S \int [d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] \exp[\lambda_1 \tilde{\theta}^1 \tilde{\theta}^2 + \lambda_2 \tilde{\theta}^3 \tilde{\theta}^4 + \cdots] \end{aligned} \tag{23.218}$$

Now, if q is odd, then in this expression $\tilde{\theta}^q$ does not appear in the exponential. Therefore the integral has a factor of $\int [d\tilde{\theta}^q] = 0$. This is a very simple example of how an "*unpaired fermion zeromode*" leads to the zero of a fermionic Gaussian integral. See remarks below.

Suppose instead that $q = 2m$ is even. Then the integral can be evaluated in terms of the skew eigenvalues as

$$\det S \prod_{i=1}^{m} (-\lambda_i) \tag{23.219}$$

Recall that an antisymmetric matrix can be skew-diagonalized by an orthogonal matrix S. We didn't quite fix which one, because we didn't specify the signs of the λ_i. Therefore, up to sign, the Gaussian integral is just the product of skew eigenvalues.

On the other hand, the integral can also be evaluated as a polynomial in the matrix elements of A. Indeed the *Pfaffian* of the antisymmetric matrix can be defined as:

$$\mathrm{pfaff}(A) := \int [d\theta^1 \cdots d\theta^{2m}] \exp[\frac{1}{2}\theta^a A_{ab} \theta^b]. \tag{23.220}$$

With a little thought one shows that expanding this out leads to

$$\begin{aligned} \mathrm{pfaff} A &= \frac{1}{m! 2^m} \sum_{\sigma \in S_{2m}} \epsilon(\sigma) A_{\sigma(1)\sigma(2)} \cdots A_{\sigma(2m-1)\sigma(2m)} \\ &= A_{12} A_{34} \cdots A_{2m-1,2m} + \cdots \end{aligned} \tag{23.221}$$

eq:pfaffian

This definition of the Pfaffian resembles that of the determinant of a matrix, but note that it is slightly different. Since A is a bilinear form it transforms as $A \to S^{tr} A S$ under change of basis. Therefore, the Pfaffian is slightly basis-dependent:

$$\mathrm{pfaff}(S^{tr} A S) := \det S \ \cdot \ \mathrm{pfaff}(A) \tag{23.222}$$

eq:pfaff-tmn

We can easily prove this using the change-of-variables formula for the Berezin integral. (Do that!)

Now a beautiful property of the Pfaffian is that it is a canonical square-root of the determinant of an antisymmetric matrix.

$$(\mathrm{pfaff} A)^2 = \det A. \tag{23.223}$$

eq:DetSqrt

(In particular, the determinant of an antisymmetric matrix - a complicated polynomial in the matrix elements - has a canonical polynomial square root.)

Using the Berezin integral we will now give a simple proof of (23.223). First, note that if M is *any* $n \times n$ matrix and we have two sets of generators $\theta_\pm^a$, $a = 1, \ldots, n$ of our Grassmann algebra then

$$\boxed{\int [d\theta_-^1 \cdots d\theta_-^n d\theta_+^n \cdots d\theta_+^1] \exp[\theta_+^i M_{ia} \theta_-^a] = \det M} \tag{23.224}$$

eq:DET-M-GRASS

An easy way to prove this is to make a transformation $M \to S^{-1} M S$ to Jordan canonical form. The change of variables of θ_+^a, θ_-^a by $S^{tr,-1}$ and S, respectively, cancel each other out. [59] Assuming that the matrix is diagonalizable we have

$$\int [d\theta_-^1 \cdots d\theta_-^n d\theta_+^n \cdots d\theta_+^1] \exp[\theta_+^1 \theta_-^1 \lambda_1 + \theta_+^2 \theta_-^2 \lambda_2 + \cdots] = \prod_{i=1}^n \lambda_i = \det M \tag{23.225}$$

To check the sign we observe that the following moves always involve moving an even number of θ's past each other. For example,

$$\begin{aligned} \theta_+^1 \theta_-^1 \theta_+^2 \theta_-^2 \theta_+^3 \theta_-^3 \theta_+^4 \theta_-^4 &= \theta_+^1 \theta_+^2 \theta_-^2 \theta_-^1 \theta_+^3 \theta_-^3 \theta_+^4 \theta_-^4 \\ &= \theta_+^1 \theta_+^2 \theta_+^3 \theta_-^3 \theta_-^2 \theta_-^1 \theta_+^4 \theta_-^4 \\ &= \theta_+^1 \theta_+^2 \theta_+^3 \theta_+^4 \theta_-^4 \theta_-^3 \theta_-^2 \theta_-^1 \end{aligned} \tag{23.226}$$

We leave the case when M has nontrivial Jordan form as a (good) exercise.

Now apply this to $M_{ia} \to A_{ij}$ with $n = 2m$ and consider

$$\det A = \int [d\theta_-^1 \cdots d\theta_-^{2m} d\theta_+^{2m} \cdots d\theta_+^1] \exp[\theta_+^i A_{ij} \theta_-^j] \tag{23.227}$$

Change variables to

$$\theta_\pm^i := \frac{1}{\sqrt{2}} (\psi^i \pm \chi^i) \qquad i = 1, \ldots, 2m \tag{23.228}$$

and note that

$$\theta_+^i A_{ij} \theta_-^j = \frac{1}{2} \psi^i A_{ij} \psi^j - \frac{1}{2} \chi^i A_{ij} \chi^j \tag{23.229}$$

To compute the superdeterminant of the change of variables perhaps the simplest way to

[59] The reader might worry about a sign at this point. To allay this fear note that we could rewrite the measure as $\epsilon [d\theta_-^1 \cdots d\theta_-^n d\theta_+^1 \cdots d\theta_+^n]$. With the latter measure the two factors in the Berezinian clearly cancel each other. But then we encounter the same sign ϵ going back to the desired ordering with $[d\tilde{\theta}_+^1 \cdots d\tilde{\theta}_+^n] = \epsilon [d\tilde{\theta}_+^n \cdots d\tilde{\theta}_+^1]$.

proceed is to compute

$$\begin{aligned}
\int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] \theta_+^1 \cdots \theta_+^{2m} \theta_-^{2m} \cdots \theta_-^1 = \\
\frac{1}{2^{2m}} \int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] (\psi^1 + \chi^1) \cdots (\psi^{2m} + \chi^{2m}) (\psi^{2m} - \chi^{2m}) \cdots (\psi^1 - \chi^1) = \\
\int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] (\chi^{2m} \psi^{2m}) (\chi^{2m-1} \psi^{2m-1}) \cdots (\chi^1 \psi^1) = \\
\int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] (\chi^{2m} \cdots \chi^1 \psi^1 \cdots \psi^{2m}) = \\
\int [d\psi^1 \cdots d\psi^{2m}] (\psi^1 \cdots \psi^{2m}) = \\
(-1)^{\frac{1}{2}(2m)(2m-1)} = (-1)^m
\end{aligned} \tag{23.230}$$

from which we conclude that

$$[d\theta_-^1 \cdots d\theta_-^{2m} d\theta_+^{2m} \cdots d\theta_+^1] = (-1)^m [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] \tag{23.231}$$

So our change of variables gives

$$\begin{aligned}
\det A &= (-1)^m \int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] \exp\left[\frac{1}{2}\psi^i A_{ij} \psi^j - \frac{1}{2}\chi^i A_{ij} \chi^j\right] \\
&= (\mathrm{pfaff}\, A)^2
\end{aligned} \tag{23.232}$$

Which concludes the proof of (23.223) ♠

Remarks

1. ♣ Remarks on localization from integral over a fermion zeromode
2. ♣ Remarks on use of Pfaffian in the general definition of Euler characteristic.
3. Why is the transformation (23.222) compatible with (23.223) and the invariance of the determinant under $A \to S^{-1}AS$? The reason is that for $S^{tr} = S^{-1}$ we have S is orthogonal so that $\det S = \pm 1$ and hence $(\det S)^2 = 1$.
4. *Pfaffians in families.* Sometimes the Pfaffian is defined as a squareroot of the determinant $\det A$ of an antisymmetric matrix. This has the disadvantage that the sign of the Pfaffian is not well-defined. In our definition, for a finite-dimensional matrix, there is a canonical Pfaffian. On the other hand, in some problems it is important to make sense of the Pfaffian of an anti-symmetric form on an infinite-dimensional Hilbert space. So, one needs another definition. Since determinants of infinite-dimensional operators can be defined by zeta-function regularization of the product of their eigenvalues one proceeds by defining the Pfaffian from the square-root of the determinant. So, we try to define the Pfaffian as:

$$\mathrm{pfaff}\, A \quad \stackrel{?}{=} \quad \prod_{\lambda > 0} \lambda \tag{23.233}$$

eq:postv

where the product runs over the skew eigenvalues. But this is not a good definition for many purposes because in families skew eigenvalues can smoothly change sign. Consider, for example the family

$$A(\alpha) = \begin{pmatrix} 0 & \cos\alpha \\ -\cos\alpha & 0 \end{pmatrix} \qquad 0 \leq \alpha \leq 2\pi \tag{23.234}$$

eq:smple

Then the Pfaffian according to the above definition (23.233) would not be differentiable at $\alpha = \pi/2$ and $3\pi/2$. Really, the pfaffian is a section of a line bundle, as we explain in Section §24 below.

One approach to pinning down the sign of the Pfaffian is simply to choose a sign at one point of the family, and then follow the skew eigenvalues continuously. With this definition

$$\mathrm{pfaff}(A(\alpha)) = \cos\alpha \tag{23.235}$$

(in agreement with our definition for finite-dimensional forms). This is a perfectly reasonable definition. However, in some problems involving gauge invariance one meets quadratic forms A which should be identified up to gauge transformation. Suppose we identify A up to orthogonal transformations. Then the equivalence class $[A(\alpha)]$ is a closed family of operators for $0 \leq \alpha \leq \pi$. If we take a smooth definition of the Pfaffian of $A(\alpha)$ then we find that it changes sign under $\alpha \to \alpha + \pi$, so in fact, it behaves more like the section of the Mobius band over the circle. We return to this in Section §24.7 below.

Exercise

a.) Show that for $q = 2$ $\det A = (a_{12})^2$

b.) Show that for $q = 4$

$$\mathrm{Pfaff}(A) = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23} \tag{23.236}$$

Check by direct compuation that indeed

$$\det A = (a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23})^2 \tag{23.237}$$

Exercise

a.) Prove equation (23.221).

b.) Explain why we divide the sum by the order of the group of centrally-symmetric shuffles (See Chapter 1, Sections 4 and 5) WB_m.

Exercise

Prove (23.224) by completing the argument for nontrivial Jordan form.

b.) Prove (23.224) by a direct evaluation of the integral without changing variables to obtain the standard expression

$$\det M = \sum_{\sigma \in S_n} \epsilon(\sigma) M_{1\sigma(1)} M_{2\sigma(2)} \cdots M_{n\sigma(n)} \tag{23.238}$$

Exercise

Let $z_1, \ldots, z_{2N}$ be points in the complex plane. Show that

$$\left(\text{Pfaff} \frac{1}{z_i - z_j} \right)^k \prod_{i<j} (z_i - z_j)^\ell \tag{23.239}$$

is a polynomial in z_i of degree $N(2N-1)\ell - kN$, so long as $k \leq \ell$, which transforms under S_{2N} with the sign $\epsilon(\sigma)^{k+\ell}$.

Expressions like this have proven useful in the theory of the fractional quantum Hall effect.

23.12.3 Gaussian integral on $\mathcal{R}^{p|q}$

Now we put these results together and consider the general Gaussian integral on $\mathcal{R}^{p|q}$:

$$(2\pi)^{-p/2} \int_{\mathcal{R}^{p|q}} [dt|d\theta] \exp\left[-\frac{1}{2} t^a Q_{ab} t^b - t^a B_{ai} \theta^i + \frac{1}{2} \theta^i A_{ij} \theta^j \right] \tag{23.240}$$

We can consider the quadratic form to have matrix elements in a general supercommutative ring (but they are constant in t^a, θ^i) so we allow odd off-diagonal terms like B_{ai}.

We can complete the square with the change of variables:

$$\begin{aligned} \tilde{t}^a &= t^a \\ \tilde{\theta}^i &= \theta^i + (A^{-1})_{ij} t^a B_{aj} \end{aligned} \tag{23.241}$$

The change of variables formula gives $[d\tilde{t}|d\tilde{\theta}] = [dt|d\theta]$ and hence we evaluate the integral to get

$$\frac{\text{Pfaff}(A)}{\sqrt{\det(Q - BA^{-1}B^{tr})}} \tag{23.242}$$

This can be written as $(\text{Ber}(\mathcal{Q}))^{-1/2}$ where $\mathcal{Q}$ is the super-quadratic form

$$\mathcal{Q} = \begin{pmatrix} Q & B \\ B^{tr} & A \end{pmatrix} \tag{23.243}$$

but the latter expression is slightly ambiguous since there are two squareroots of $\det A$.

23.12.4 Supersymmetric Cancelations

Suppose a super-quadratic form on $\mathcal{R}^{n|2n}$ is of the special form

$$\mathcal{Q} = \begin{pmatrix} M^{tr}M & 0 & 0 \\ 0 & 0 & M \\ 0 & -M^{tr} & 0 \end{pmatrix} \tag{23.244}$$ eq:TFT-1

where M is nonsingular and $\mathrm{Re}(M^{tr}M) > 0$. Then the Gaussian integral is just

$$\mathrm{sign}(\det M) \tag{23.245}$$ eq:TFT-2

Note that M (reduced modulo nilpotents) might be a complex matrix, and the integral is still sensible so long as $\mathrm{Re}(M^{tr}M) > 0$. Therefore we define

$$\mathrm{sign}(\det M) := \begin{cases} +1 & |\arg(\det M)| < \pi/4 \\ -1 & |\arg(\det M) - \pi| < \pi/4 \end{cases} \tag{23.246}$$

Thus, the result of the Gaussian integral (23.245) is "almost" independent of the details of M. There is a nice "theoretical" explanation of this fact which is a paradigm for arguments in supersymmetric field theory and topological field theory.

So, let us denote, for brevity

$$[d\theta_- d\theta_+] := [d\theta_-^1 \cdots d\theta_-^n d\theta_+^n \cdots d\theta_+^1] \tag{23.247}$$

and we consider the integral

$$I[M] := (2\pi)^{-n/2} \int_{\mathcal{R}^{n|2n}} [dt|d\theta_- d\theta_+] \exp\left[-\frac{1}{2} t^i (M^{tr}M)_{ik} t^k + \theta_+^i M_{ij} \theta_-^j\right] = \mathrm{sign}(\det M) \tag{23.248}$$ eq:TFT-3

It is useful to introduce n additional bosonic coordinates H^i and instead write this as an integral over $\mathcal{R}^{2n|2n}$:

$$I[M] = (2\pi)^{-n} \int_{\mathcal{R}^{2n|2n}} [dt dH|d\theta_- d\theta_+] \exp\left[-\frac{1}{2} H^i H^i + \sqrt{-1} H^i M_{ij} t^j + \theta_+^i M_{ij} \theta_-^j\right] \tag{23.249}$$ eq:TFT-1

Now, introduce the odd vector field

$$Q := \theta_-^k \frac{\partial}{\partial t^k} - \sqrt{-1} H^k \frac{\partial}{\partial \theta_+^k} \tag{23.250}$$

Note that, on the one hand, the "action" can be written as

$$Q(\Psi) = -\frac{1}{2} H^i H^i + \sqrt{-1} H^i M_{ij} t^j + \theta_+^i M_{ij} \theta_-^j \tag{23.251}$$

where

$$\Psi = -\frac{i}{2} \theta_+^k H^k - \theta_+^i M_{ij} t^j \tag{23.252}$$

and on the other hand,

$$Q^2 = \frac{1}{2}[Q,Q]_+ = 0 \tag{23.253}$$

So, now suppose we perturb $M \to M + \delta M$. Then the change in the Gaussian integral can be written as

$$\delta I[M] = \int [dtdH|d\theta_- d\theta_+] Q(\delta\Psi) e^{Q(\Psi)} = \int [dtdH|d\theta_- d\theta_+] Q(\delta\Psi e^{Q(\Psi)}) = 0 \tag{23.254}$$

where the last inequality follows from integration by parts.

The reader might be bothered by this. The answer (23.248) *does* depend a little bit on M. Moreover, why can't we just use the argument to put M to zero? But then, of course, the integral would seem to be $\infty \times 0$ for $M = 0$. If a perturbation makes M singular then we get a factor of $\infty \times 0$ where the ∞ comes from the integral dt and the 0 comes from the fermionic integral. Recall, however, that in the definition of the integral we do the fermionic integral first and therefore $\int [dtdH|d\theta_- d\theta_+] 1 = 0$. Therefore, we could replace the integrand by

$$e^{Q(\Psi)} - 1 = Q\left(\Psi + \frac{1}{2!}\Psi Q(\Psi) + \cdots\right) \tag{23.255}$$

There will be a term which survives the fermionic integral but it is a total derivative in $\frac{\partial}{\partial t^i}$ which does not vanish at infinity in field space. Thus, *singular perturbations can change the value of the integral.*

23.13 References

In preparing this section we have used the following references:

1. P. Deligne and J.W. Morgan, *Notes on Supersymmetry (following Joseph Bernstein* in *Quantum Fields and Strings: A Course for Mathematicians*, Vol. 1, pp.41-96 AMS

2. D. Leites, "Introduction to supermanifolds," 1980 Russ. Math. Surv. 35 1.

3. E. Witten, "Notes on Supermanifolds and Integration," arXiv:1209.2199.

4. www.math.ucla.edu/ vsv/papers/ch3.pdf

5. J. Groeger, Differential Geometry of Supermanifolds, http://www.mathematik.huberlin.de/ groegerj

   Possibly useful, but I haven't seen them yet:

6. V. Varadarajan, *Supersymmetry for Mathematicians: An Introduction*

7. *Super Linear Algebra* by Kandasamy and Smarandache

   For an extremely accessible discussion of the theory of schemes See

8. D. Eisenbud and J. Harris, *The Geometry of Schemes*, Springer GTM

c:DETERMINANTS

24. Determinant Lines, Pfaffian Lines, Berezinian Lines, and anomalies

24.1 The determinant and determinant line of a linear operator in finite dimensions

Recall that a one-dimensional vector space over κ is called a *line*. If L is a line then a linear transformation $T : L \to L$ can canonically be identified with an element of κ. Indeed, choose any basis vector v for L then $T(v) = tv$, with $t \in \kappa$, and the number t does not depend on v. On the other hand, suppose we have two lines L_1, L_2. They are, of course, isomorphic, but not naturally so. In this case if we have a linear transformation $T : L_1 \to L_2$ then there is not canonical way of identifying T with an element of κ because there is no *a priori* way of identifying a choice of basis for L_1 with a choice of basis for L_2. Put differently, $\mathrm{Hom}(L, L) \cong L^\vee \otimes L \cong \kappa$ is a natural isomorphism, but $\mathrm{Hom}(L_1, L_2) \cong L_1^\vee \otimes L_2$ is just another line.

Now suppose that

$$T : V \to W \tag{24.1}$$

is a linear transformation between different vector spaces over κ where the dimension is possibly larger than one. Then there is no canonical notion of the determinant as a number. Choose ordered bases $\{v_i\}$ for V and $\{w_a\}$ for W, and define M_{ai} to be the matrix of T with respect to that basis. Then under this isomorphism $T \in \mathrm{Hom}(V, W) \cong V^* \otimes W$ may be written as

$$T = \sum_{i,a} M_{ai} \hat{v}_i \otimes w_a \tag{24.2}$$

and if $\mathrm{dim}V = \mathrm{dim}W$ one can of course define the number $\mathrm{det}M_{ai}$. However, a change of basis of V by S_1 and of W by S_2 changes the matrix $M \to S_2^{-1} M S_1$ and hence leads to

$$\mathrm{det}M' = \mathrm{det}S_2^{-1} \mathrm{det}M \mathrm{det}S_1 \tag{24.3}$$

in the new basis. *If V, W are not naturally isomorphic, then we cannot naturally assign a number we would want to call* $\mathrm{det}T$.

Nevertheless, there *is* a good mathematical construction of $\mathrm{det}T$ which is natural with respect to the domain and range of T. What we do is consider the 1-dimensional vector spaces $\Lambda^d V$ and $\Lambda^{d'} W$ where $d = \mathrm{dim}V, d' = \mathrm{dim}W$. Then there is a canonically defined linear transformation

$$\mathrm{det}T : \Lambda^d V \to \Lambda^{d'} W \tag{24.4}$$

For $d \neq d'$ it is zero, and for $d = d'$ we can write it by choosing bases v_i, w_a. Denote the dual basis $\hat{v}_i$ so that $T = \sum_{i,j} M_{ai} \hat{v}_i \otimes w_a$. Then

$$\mathrm{det}T := \frac{1}{(d!)^2} \sum_{a_s, i_s} M_{a_1 i_1} \cdots M_{a_d i_d} \hat{v}_{i_1} \wedge \cdots \wedge \hat{v}_{i_d} \otimes w_{a_1} \wedge \cdots \wedge w_{a_d} \tag{24.5}$$

eq:detelement

The important thing about this formula is that, as opposed to the determinant defined as a polynomial in matrix elements, the object (24.5) is *natural* with respect to both V and W. That is, it is independent of basis (even though we chose a basis to write it out,

if we change basis we get the same object. This is not true of the determinant defined as a polynomial in matrix elements.)

While (24.5) is natural it requires interpretation. It is not a number, it is an element of a one-dimensional vector space, i.e. a *line*. This line is called the *determinant line of* T

$$\mathrm{DET}(T) := \Lambda^{\dim V}(V^*) \otimes \Lambda^{\dim W}(W) \tag{24.6}$$

eq:detline

Thus, we have a one-dimensional vector space $\mathrm{DET}(T)$ and an element of that vector space $\det(T) \in \mathrm{DET}(T)$.

This is a nontrivial concept because:

1. Linear operators often come in families T_s. Then $\mathrm{DET}(T)$ becomes a nontrivial line bundle over parameter space.

2. The theory extends to infinite dimensional operators such as the Dirac operator. Indeed, in finite dimensions $\mathrm{DET}(T)$ does not depend on the choice of operator T except through its domain and target. This is no longer true in infinite dimensions.

Remarks

1. When $\dim V \neq \dim W$ the line $\mathrm{DET}(T)$ defined in (24.6) still makes sense, but we must define the element in that line, $\det T \in \mathrm{DET}(T)$ to be $\det T = 0$.

2. When $W = V$, that is, when they *are* canonically isomorphic then $\det M_{ij}$ is basis independent. Indeed, in this case there is a canonical isomorphism

$$\Lambda^d(V) \otimes \Lambda^d(V^\vee) \cong \kappa \tag{24.7}$$

(κ can be $\mathbb{R}$ or $\mathbb{C}$, or any field.)

Example Above we considered an interesting family of one-dimensional vector spaces

$$\mathcal{L}_+ = \{(\hat{x}, v) | \hat{x} \cdot \vec{\sigma} v = +v\} \subset S^2 \times \mathbb{C}^2 \tag{24.8}$$

For each point $\hat{x} \in S^2$ we have a one-dimensional subspace $\mathcal{L}_{+,\hat{x}} \subset \mathbb{C}^2$. Let us let

$$L = S^2 \times \mathbb{C} \tag{24.9}$$

be another family of one-dimensional vector spaces. We can define an operator

$$T_{\hat{x}} : \mathcal{L}_{+,\hat{x}} \to L_{\hat{x}} \tag{24.10}$$

defined by just projecting to the first component of the vector $v \in \mathcal{L}_{+,\hat{x}}$. To find a matrix of $T_{\hat{x}}$ we need to choose a basis. As we discussed we might choose a basis vector

$$e_+ = \begin{pmatrix} \cos \frac{1}{2}\theta \\ e^{i\phi} \sin \frac{1}{2}\theta \end{pmatrix} \tag{24.11}$$

away from the south pole, while away from the north pole we might choose:

$$e_- = \begin{pmatrix} e^{-i\phi} \cos \frac{1}{2}\theta \\ \sin \frac{1}{2}\theta \end{pmatrix} \tag{24.12}$$

We can also simply choose 1 as a basis for $\mathbb{C}$ for the target of $T_{\hat{x}}$.

With respect to this basis we would have a determinant function

$$\det M_{ai} = \cos\frac{1}{2}\theta \qquad 0 \leq \theta < \pi \tag{24.13}$$

$$\det M_{ai} = e^{-i\phi}\cos\frac{1}{2}\theta \qquad 0 < \theta \leq \pi \tag{24.14}$$

Note that the second expression does make sense at the south pole because $\cos\frac{\pi}{2} = 0$. Clearly this does not define a function on S^2. Rather, it defines a section of a line bundle. Moreover, it has exactly one zero.

Exercise

Consider a finite dimensional vector space V of dimension d.

a.) Show that there is a canonical isomorphism $\Lambda^d(V^\vee) \cong (\Lambda^d V)^\vee$

b.) Show that there is a canonical isomorphism

$$\Lambda^d V^\vee \otimes \Lambda^d V \to \kappa \tag{24.15}$$

24.2 Determinant line of a vector space and of a complex

If V is a finite-dimensional vector space of dimension d then the line $\Lambda^d V$ is often called the *determinant line of* V and denoted

$$\mathrm{DET}(V) := \Lambda^{\dim V} V \tag{24.16}$$

Because there is a natural isomorphism

$$\mathrm{DET}(V)^\vee \otimes \mathrm{DET}(V) \cong \kappa \tag{24.17}$$

and since the one-dimensional space κ acts like a multiplicative identity under $\otimes$, we also denote

$$\mathrm{DET}(V)^{-1} := \mathrm{DET}(V)^\vee \tag{24.18}$$

Therefore we will also denote

$$(v_1 \wedge \cdots \wedge v_d)^{-1} := (v_1 \wedge \cdots \wedge v_d)^\vee \tag{24.19}$$

for any choice of basis $\{v_i\}$.

Again, when we consider families of vector spaces, we can get interesting line bundles from this construction:

Example Consider the Grassmannian $Gr(k,n)$ of k-dimensional subspaces of $\mathbb{C}^n$. For each subspace $W \in Gr(k,n)$ we may associate the line $\Lambda^k W$. This gives another nontrivial

family of lines which is naturally a determinant line bundle. A notable use of this Determinant line in physics is that it can be interpreted as the vacuum line in quantization of n free fermions. Then the possible sets of creation operators form the Grassmannian $Gr(n, 2n)$. ♣Say more ♣

Now consider a short exact sequence of vector spaces:

$$0 \to V_1 \to V_2 \xrightarrow{\pi} V_3 \to 0 \tag{24.20}$$

Then we claim there is a canonical isomorphism

$$\mathrm{DET}(V_1) \otimes \mathrm{DET}(V_3) \cong \mathrm{DET}(V_2) \tag{24.21}$$ eq:DET-MULT

To see this, consider any ordered basis $\{v_1, \ldots, v_{d_1}\}$ for V_1 and $\{w_1, \ldots, w_{d_3}\}$ for V_3. Then lift the w_a to vectors $\tilde{w}_a$ in V_2 so that $\pi(\tilde{w}_a) = w_a$. Then

$$\{v_1, \ldots, v_{d_1}, \tilde{w}_1, \ldots, \tilde{w}_{d_3}\} \tag{24.22}$$

is an ordered basis for V_2. Our canonical isomorphism (24.21) is defined by

$$(v_1 \wedge \cdots \wedge v_{d_1}) \otimes (w_1 \wedge \cdots \wedge w_{d_3}) \mapsto v_1 \wedge \cdots \wedge v_{d_1} \wedge \tilde{w}_1 \wedge \cdots \wedge \tilde{w}_{d_3} \tag{24.23}$$

The main point to check here is that the choice of lifts $\tilde{w}_a$ do not matter on the RHS. This is clear since for a different choice $\tilde{w}'_a - \tilde{w}_a \in V_1$ and the difference is annihilated by the first part of the product. Next, under changes of bases both sides transform in the same way, so the isomorphism is basis-independent, and hence natural. Put differently, the element

$$(v_1 \wedge \cdots \wedge v_{d_1})^{-1} \otimes (v_1 \wedge \cdots \wedge v_{d_1} \wedge \tilde{w}_1 \wedge \cdots \wedge \tilde{w}_{d_3}) \otimes (w_1 \wedge \cdots \wedge w_{d_3})^{-1} \tag{24.24}$$

of the line given in the LHS of (24.25) is actually *independent* of the choice of bases $\{v_1, \ldots, v_{d_1}\}$, $\{w_1, \ldots, w_{d_3}\}$, and the lifts $\tilde{w}_a$. So even though we chose bases to exhibit this element it is basis-independent, and hence natural. It gives therefore a natural isomorphism

$$\mathrm{DET}(V_1)^{-1} \otimes \mathrm{DET}(V_2) \otimes \mathrm{DET}(V_3)^{-1} \cong \kappa \tag{24.25}$$ eq:3-Term-Cplx

The same kind of reasoning as we used to prove (24.21) can be used to prove that if

$$0 \to V_1 \to \cdots \to V_i \xrightarrow{T_i} V_{i+1} \to \cdots \to V_n \to 0 \tag{24.26}$$

is an exact sequence, so $\mathrm{im} T_i = \ker T_{i+1}$, then there is a canonical isomorphism

$$\bigotimes_{i \text{ odd}} \mathrm{DET}(V_i)^{-1} \otimes \bigotimes_{i \text{ even}} \mathrm{DET}(V_i) \cong \kappa \tag{24.27}$$ eq:ExactSequer

---

Exercise *Determinant of a complex*

A multiplicative version of the Euler-Poincaré principal is that if

$$0 \to V_1 \to \cdots \to V_i \xrightarrow{d_i} V_{i+1} \to \cdots \to V_n \to 0 \tag{24.28}$$

is a complex $d_{i+1} d_i = 0$ (i.e. not necessarily exact) then there is a natural isomorphism:

$$\bigotimes_{i \text{ odd}} \mathrm{DET}(V_i)^{-1} \otimes \bigotimes_{i \text{ even}} \mathrm{DET}(V_i) \cong \bigotimes_{i \text{ odd}} \mathrm{DET}(H^i)^{-1} \otimes \bigotimes_{i \text{ even}} \mathrm{DET}(H^i) \tag{24.29}$$ eq:DetComplex

---

24.3 Abstract defining properties of determinants

Because we want to speak of determinants and determinant lines in infinite dimensions it can be useful to take a more abstract approach to the determinant and determinant line of T. These can be abstractly characterized by the following three properties:

1. $\det(T) \neq 0$ iff T is invertible.
2. $\mathrm{DET}(T_2 \circ T_1) \cong \mathrm{DET}(T_2) \otimes \mathrm{DET}(T_1)$
3. If T_1, T_2, T_3 map between two short exact sequences

$$\begin{array}{ccccccccc} 0 & \to & E_1 & \to & E_2 & \stackrel{\pi}{\to} & E_3 & \to & 0 \\ & & T_1 \downarrow & & T_2 \downarrow & & T_3 \downarrow & & \\ 0 & \to & F_1 & \to & F_2 & \stackrel{\pi}{\to} & F_3 & \to & 0 \end{array} \tag{24.30}$$

then

$$\mathrm{DET}(T_2) \cong \mathrm{DET}(T_1) \otimes \mathrm{DET}(T_3) \tag{24.31}$$

canonically, and under this isomorphism

$$\det(T_2) \to \det T_1 \otimes \det T_3 \tag{24.32}$$

Exercise

Show that property (3) means, essentially, that T_2 has block upper-triangular form. Show this by considering

$$\begin{array}{ccccccccc} 0 & \to & \mathbb{C}^n & \stackrel{\iota}{\to} & \mathbb{C}^{n+m} & \stackrel{\pi}{\to} & \mathbb{C}^m & \to & 0 \\ & & T_1 \downarrow & & T_2 \downarrow & & T_3 \downarrow & & \\ 0 & \to & \mathbb{C}^n & \stackrel{\iota}{\to} & \mathbb{C}^{n+m} & \stackrel{\pi}{\to} & \mathbb{C}^m & \to & 0 \end{array} \tag{24.33}$$

where

$$\begin{aligned} \iota &: (x_1, \ldots, x_n) \to (x_1, \ldots, x_n, 0, \ldots, 0) \\ \pi &: (y_1, \ldots, y_{n+m}) \to (y_{n+1}, \ldots, y_{n+m}) \end{aligned} \tag{24.34}$$

Show that

$$T_2 = \begin{pmatrix} T_1 & T_{12} \\ 0 & T_2 \end{pmatrix} \tag{24.35}$$

where T_{12} is in general nonzero.

24.4 Pfaffian Line

sec:PfaffLine

Just as for determinants, Pfaffians are properly regarded as sections of line (bundles).

Recall that for an antisymmetric matrix A, $\mathrm{Pfaff}(S^{tr}AS) = (\det S)\mathrm{Pfaff}(A)$. So the Pfaffian is basis-dependent, yet, once again, there is a basis independent notion of a Pfaffian.

Let T be an antisymmetric bilinear form on a finite-dimensional vector space V.

Recall that a bilinear form T on a vector space V can be regarded as an element of

$$T : V \to V^\vee \tag{24.36}$$

Then, from our definition above

$$\mathrm{DET}(T) = (\Lambda^d V^\vee)^{\otimes 2} \tag{24.37}$$

For an antisymmetric bilinear form we define the Pfaffian line of T to be the "squareroot":

$$\mathrm{PFAFF}(T) := \Lambda^d V^\vee \tag{24.38}$$

On the other hand, a bilinear form defines a map $V \otimes V \to \kappa$ and hence is an element of

$$T \in V^\vee \otimes V^\vee \tag{24.39}$$

Moreover, if T is antisymmetric then it defines a 2-form

$$\omega_T \in \Lambda^2 V^\vee \tag{24.40}$$

If $d = 2m$ we define the Pfaffian element to be:

$$\mathrm{pfaff} T := \frac{\omega_T^m}{m!} \in \Lambda^n V^\vee \tag{24.41}$$

Exercise *Comparing Liouville and Riemann volume forms*

a.) Suppose V is a real vector space with a a positive definite symmetric form g. Show that if A_{ij} is the anti-symmetric matrix of T with respect to an orthonormal basis with respect to g, then

$$\frac{\omega_T^m}{m!} = \mathrm{pfaff}(A)\mathrm{vol}\,(g) \tag{24.42}$$

where $\mathrm{vol}\,(g)$ is the volume form of the metric, i.e. $\mathrm{vol}\,(g) = e^1 \wedge \cdots \wedge e^n$ where $\{e^1, \ldots, e^n\}$ is an ordered ON basis.

Note that if we change ordered ON bases then both $\mathrm{pfaff}(A)$ and $\mathrm{vol}\,(g)$ change by $\det S$, where $S \in O(2m)$, so the product is well-defined.

b.) Let M be a symplectic manifold with symplectic form ω. Choose any Riemannian metric g on M. Then let $A(g)$ be the antisymmetric matrix of ω with respect to an ON frame for g. Note that $\det(A(g))$ does not depend on the choice of ON frame, and is hence a globally well-defined function. Show that M is orientable iff $\det(A(g))$ admits a globally-defined square root on M.

24.5 Determinants and determinant lines in infinite dimensions

24.5.1 Determinants

Now let us consider a linear operator $T : \mathcal{H} \to \mathcal{H}$. We will take $\mathcal{H}$ to be an infinite-dimensional separable Hilbert space. [60]

Can we speak meaningfully of det$T$ in this case? Suppose that $T = 1 + \Delta$, where Δ is traceclass. Then it has a set of eigenvalues $\{\delta_k\}_{k=1}^{\infty}$ with

$$\sum_{k=1}^{\infty} |\delta_k| < \infty \tag{24.43}$$

In this case $\sum_{k=1}^{\infty} \log\delta_k$ is a well-defined absolutely convergent series and

$$\det T := \lim_{N\to\infty} \prod_{k=1}^{N} (1 + \delta_k) \tag{24.44}$$

is well-defined, and can be taken to be the determinant of T.

This is a good start, but the class of operators $1 + traceclass$ is far too small for use in physics and mathematics. Another definition, known as the ζ-function determinant, introduced by Ray and Singer [61] can be defined as follows:

Let $T : \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator with discrete point spectrum $\{\lambda_k\}_{k=1}^{\infty}$ assume that

1. No $\lambda_k = 0$ (otherwise $\det T = 0$).
2. $|\lambda_k| \to \infty$ for $k \to \infty$.
3. If we form the series

$$\zeta_T(s) := \sum_k \lambda_k^{-s} := \sum_{\lambda_k>0} \lambda_k^{-s} + \sum_{\lambda_k<0} e^{-i\pi s} |\lambda_k|^{-s} \tag{24.45}$$

eq:zeta-functi

then the spectrum goes to infinity rapidly enough so that $\zeta_T(s)$ converges to an analytic function on the half-plane $\mathrm{Re}(s) > R_0$, for some R_0, and admits an analytic continuation to a holomorphic function of s near $s = 0$.

When these conditions are satisfied we may define

$$\det_\zeta(T) := \exp[-\zeta_T'(0)] \tag{24.46}$$

Remark: A typical example of an operator for which these conditions apply is an elliptic operator on a compact manifold. For example, the Laplacian acting on tensors on a smooth compact manifold, or the Dirac operator on a smooth compact spin manifold are common examples. See Example **** below.

[60] This is not strictly necessary for some definitions and constructions below.

[61] GIVE REFERENCE.

The next natural question is to consider determinant lines for operators $T : \mathcal{H}^1 \to \mathcal{H}^2$ between two "different" Hilbert spaces. Of course, we proved in Section §13 that any two separable Hilbert spaces are isomorphic. However, there is no natural isomorphism, so the question where an expression like det$T$ should be valued is just as valid as in finite dimensions. A good example is the *chiral* Dirac operator on an even dimensional spin manifold. First, we must identify a suitable class of operators where such determinant lines can make sense.

24.5.2 Fredholom Operators

Definition

a.) An operator $T : \mathcal{H}^1 \to \mathcal{H}^2$ between two separable Hilbert spaces is said to be *Fredholm* if ker$T$ and cok$T$ are finite-dimensional.

b.) The *index* of a Fredholm operator is defined to be

$$\mathrm{Ind}(T) := \dim \ker T - \dim \mathrm{cok} T \tag{24.47}$$

Comments and Facts

1. We are generally interested in *unbounded* operators. Then, as we have discussed the domain $D(T)$ is an important part of the definition of T. The above definition is a little sloppy for unbounded operators. For some purposes, such as index theory one can replace T by $T/(1 + T^\dagger T)$ and work with a bounded operator.

2. One often sees Fredholm operators defined with the extra requirement that $\mathrm{im}(T) \subset \mathcal{H}^2$ is a *closed* subspace. In fact, it can be shown (using the closed graph theorem) that if T is bounded and ker$T$ and cok$T$ are finite dimensional then this requirement is satisfied. ♣Give argument. ♣

3. Another definition one finds is that a bounded operator T is Fredholm iff there is an inverse up to compact operators. That is, T is Fredholm iff there is a bounded operator $S : \mathcal{H}^2 \to \mathcal{H}^1$ so that $TS - 1$ and $ST - 1$ are compact operators. (Recall that compact operators are finite-rank operators, or limits in the norm topology of finite-rank operators.) The equivalence of these definitions is known as *Atkinson's theorem*.

4. The space of all bounded Fredholm operators $\mathcal{F}(\mathcal{H}^1, \mathcal{H}^2)$ inherits a topology from the operator norm.

5. If T is Fredholm then there is an $\epsilon > 0$ (depending on T) so that if $\| K \| < \epsilon$ then $T + K$ is Fredholm and $\mathrm{Ind}(T) = \mathrm{Ind}(T + K)$. Therefore, the index is a continuous map:

$$\mathrm{Ind} : \mathcal{F}(\mathcal{H}) \to \mathbb{Z} \tag{24.48}$$

and is hence constant on connected components.

6. In fact, the space of Fredholm operators has infinitely many connected components, in the norm topology, and these are in 1-1 correspondence with the integers and can be labeled by the index.

7. Warning: In the compact-open and strong operator topologies $\mathcal{F}(\mathcal{H})$ is *contractible*. [62]

♣♣: EXPLAIN: Even for unbounded operators (with dense domain of definition) the definition of Fredholm is that the kernel and cokernel are finite dimensional. There is no need to say that the range is closed. But only when the range is closed is there an isomorphism of the kernel of $T^\dagger$ with the cokernel. A good example is d/dx on $L^2([1,\infty)$ (or do we need the half-line?) The kernel is zero because 1 is not normalizable, so the kernel of $T^\dagger$ is also zero. But we can construct a lot of states which are not in the cokernel. For example $1/x^n$ is in the image of d/dx if $n > 3/2$ but not if $n \leq 3/2$ since the preimage would not be L^2-normalizable. So the range is not closed and the cokernel is not isomorphic to the kernel of the adjoint.

24.5.3 The determinant line for a family of Fredholm operators

There are two descriptions of the determinant line:

Construction 1: We define a line bundle DET first over the index zero component $\mathcal{F}(\mathcal{H}^1,\mathcal{H}^2)_0$ whose fiber at $T \in \mathcal{F}(\mathcal{H}^1,\mathcal{H}^2)$ is

$$\mathrm{DET}|_T := \{(S,\lambda)|S:\mathcal{H}^2\to\mathcal{H}^1, S^{-1}T\in\mathcal{I}_1\}/\sim \tag{24.49}$$

where the equivalence relation is

$$(S_1,\lambda_1)\sim(S_2,\lambda_2) \qquad \leftrightarrow \qquad \lambda_2 = \lambda_1 \det(S_2^{-1}S_1) \tag{24.50}$$

where $S_2^{-1}S_1 = 1 + traceclass$ and we use the standard definition for this.

To check this one has to check that $\sim$ really is an equivalence relation. Next, note that $\mathrm{DET}|_T$ is indeed a one-dimensional vector space with vector space structure $z\cdot[(S,\lambda)] := [(S,z\lambda)]$, so any two vectors are proportional: $[(S_1,\lambda_1)] = \xi[(S_2,\lambda_2)]$ with $\xi = \lambda_1\lambda_2^{-1}\det(S_2^{-1}S_1)$.

The next thing to check is that, in the norm topology, the lines are indeed a continuous family of lines.

There is a canonical section of this line: $\det(T) := [(T,1)]$ (in the index zero component).

[62] Here is an argument provided by Graeme Segal: Identify $\mathcal{H}$ with $L^2[0,1]$ and then write $\mathcal{H}\cong\mathcal{H}_t\oplus\mathcal{H}_{1-t}$ where $\mathcal{H}_t$ and $\mathcal{H}_{1-t}$ are the Hilbert spaces on the intervals $[0,t]$ and $[t,1]$, respectively. Then $\mathcal{H}$ is also isomorphic to $\mathcal{H}_t$. For an operator A on $\mathcal{H}$ let A_t be its image under this isomorphism. Then, one can check that $t\mapsto A_t\oplus 1_{\mathcal{H}_{1-t}}$ is continuous in the norm topology, deforms A to 1, and stays in the set of Fredholm operators if A is Fredholm.

One can show that for any T there is a canonical isomorphism

$$\mathrm{DET}|_T \cong \mathrm{DET}(\mathrm{ker}T)^{-1} \otimes \mathrm{DET}(\mathrm{cok}T) \tag{24.51}$$

The reason we can't just use the RHS as a definition of the determinant line is that in families $T(s)$, we can have the spaces $\mathrm{ker}T(s)$ and $\mathrm{cok}T(s)$ jump discontinuously in dimension.

This leads us to consider the second construction:

Construction 2: Let $\mathcal{S}$ be a family of Fredholm operators $T(s)$. For any positive real number a define

$$\mathcal{U}_a = \{T | a \notin \sigma(T^\dagger T)\} \tag{24.52}$$

If $T \in \mathcal{U}_a$ then we can use the spectral decomposition of the Hermitian operator $T^\dagger T$ to split the Hilbert space into the "low energy" and "high energy" modes:

$$\mathcal{H} = \mathcal{H}_{<a} \oplus \mathcal{H}_{>a} \tag{24.53}$$

(i.e. we use the spectral projection operators). Moreover, since T is Fredholm, the "low energy" space $\mathcal{H}_{<a}$ is in fact *finite-dimensional*.

Now notice that we have an exact sequence:

$$0 \to \mathrm{ker}T \to \mathcal{H}^1_{<a} \quad \stackrel{T}{\to} \quad \mathcal{H}^2_{<a} \to \mathrm{cok}T \to 0 \tag{24.54}$$

Now, using the property of determinant lines in exact sequences we conclude that there is a *canonical* isomorphism

$$\mathrm{DET}(\mathrm{ker}T)^{-1} \otimes \mathrm{DET}(\mathrm{cok}T) \cong \mathrm{DET}(\mathcal{H}^1_{<a})^{-1} \otimes \mathrm{DET}(\mathcal{H}^2_{<a}) \tag{24.55}$$

The advantage of using the RHS of this equation is that now we can consider what happens on overlaps $\mathcal{U}_{ab} = \mathcal{U}_a \cap \mathcal{U}_b$, where we can assume WLOG that $a < b$. Then

$$\mathcal{H}_b = \mathcal{H}_a \oplus \mathcal{H}_{a,b} \tag{24.56}$$

where $\mathcal{H}_{a,b}$ is the sum of eigenspaces with eigenvalues between a, b. Note that there is an isomorphism $T_{a,b} : \mathcal{H}^1_{a,b} \to \mathcal{H}^2_{a,b}$ (which is just the restriction of T) and hence $\mathrm{det}T_{a,b}$ is nonzero and a canonical trivialization of

$$\mathrm{DET}(\mathcal{H}^1_{a,b})^{-1} \otimes \mathrm{DET}(\mathcal{H}^2_{a,b}) \tag{24.57}$$

Using these trivializations the determinant line bundles patch together to give a smooth determinant line bundle over the whole family.

♣More detail. Warn about gerbes. ♣

24.5.4 The Quillen norm

In physical applications we generally want our path integrals and correlation functions to be numbers, rather than sections of line bundles. (Sometimes, we just have to live with the latter situation.)

$$\| \, (v^1 \wedge \cdots \wedge v^n)^{-1} \otimes (w^1 \wedge \cdots \wedge w^m) \, \|^2 = \frac{\det(w^a, w^b)}{\det(v^i, v^j)} \mathrm{det}'_\zeta(T^\dagger T) \tag{24.58}$$

♣ SHOW IT PATCHES NICELY ♣

24.5.5 References

1. G. Segal, Stanford Lecture 2, http://www.cgtp.duke.edu/ITP99/segal/

2. D. Freed, On Determinant Line Bundles

3. D. Freed, "Determinants, Torsion, and Strings," Commun. Math. Phys. **107**(1986)483

4. D. Freed and G. Moore, "Setting the Quantum Integrand of M Theory," hep-th/0409135.

24.6 Berezinian of a free module

There is an analog of the determinant line of a vector space also in $\mathbb{Z}_2$-graded linear algebra.

Let $\mathcal{A}$ be a supercommutative superalgebra and consider a free module $M \cong \mathcal{A}^{p|q}$. Then we can define a free module $\mathrm{Ber}(M)$ of rank $(1|0)$ or $(0|1)$ over $\mathcal{A}$ by assigning for every isomorphism

$$M \cong e_1\mathcal{A} \oplus \cdots \oplus e_{p+q}\mathcal{A} \tag{24.59}$$

a basis vector in $\mathrm{Ber}(M)$ denoted by $[e_1, \ldots, e_p | e_{p+1} \ldots, e_{p_q}]$ so that if T is an automorphism of M then $T(e_i) = e_k X^k{}_j$, with the matrix elements $X^k{}_j \in \mathcal{A}$ and we take

$$[Te_1, \ldots, Te_p | Te_{p+1} \ldots, Te_{p_q}] = \mathrm{Ber}(X)[e_1, \ldots, e_p | e_{p+1} \ldots, e_{p_q}] \tag{24.60}$$

To complete the definition we take the parity of $\mathrm{Ber}(M)$ to be

$$\mathrm{Ber}(M) \cong \begin{cases} \mathcal{A}^{1|0} & q = 0 \mathrm{mod} 2 \\ \mathcal{A}^{0|1} & q = 1 \mathrm{mod} 2 \end{cases} \tag{24.61}$$

Now, given an exact sequence

$$0 \to M' \to M \to M'' \to 0 \tag{24.62}$$

with $M' \cong \mathcal{A}^{p'|q'}$, $M'' \cong \mathcal{A}^{p''|q''}$, and $M \cong \mathcal{A}^{p|q}$ then there is a *natural* isomorphism

$$\mathrm{Ber}(M') \otimes \mathrm{Ber}(M'') \to \mathrm{Ber}(M) \tag{24.63}$$

eq:BerLineMult

This isomorphism is defined by choosing a basis $\{e'_1, \ldots, e'_{p'+q'}\}$ for M' and a complementary basis $\{\tilde{e}''_1, \ldots, \tilde{e}''_{p''+q''}\}$ for M which projects to a basis $\{e''_1, \ldots, e''_{p''+q''}\}$ for M''. Then the multiplicativity isomorphism (24.63) is defined by

$$\begin{aligned} [e'_1, \ldots, e'_{p'} | e'_{p'+1}, \ldots e'_{p'+q'}] \otimes [e''_1, \ldots, e''_{p''} | e''_{p''+1}, \ldots, e''_{p''+q''}] \\ \to [e'_1, \ldots, e'_{p'}, \tilde{e}''_1, \ldots, \tilde{e}''_{p''} | e'_{p'+1}, \ldots e'_{p'+q'}, \tilde{e}''_{p''+1}, \ldots, \tilde{e}''_{p''+q''}] \end{aligned} \tag{24.64}$$

Although we have chosen bases to define the isomorphism one can check that under changes of bases the isomorphism remains of the same form, so in this sense it is "natural."

Now in our discussion of integration over a superdomain the densities $\mathcal{D}^{p|q}$ on $\mathcal{R}^{p|q}$ can be recognized to be simply $\mathrm{Ber}(\Omega^1 \mathcal{R}^{p|q})$. This behaves well under coordinate transformations and so defines a sheaf on $\mathcal{M}_{\mathrm{red}}$ on a supermanifold, and so:

The analog of a density for a manifold is a global section of $\mathrm{Ber}(\Omega^1\mathcal{M})$ on a supermanifold $\mathcal{M}$.

In supersymmetric field theories, this is the kind of quantity we should be integrating.

♣Should complete the parallel discussion by Berezinian line for linear transformation between modules etc. ♣

nomalyComments

24.7 Brief Comments on fermionic path integrals and anomalies

24.7.1 General Considerations

Determinant lines are of great importance in quantum field theory, especially in the theory of anomalies.

♣dbend: [latex command not working]: Warning: This section assumes a knowledge of many things we have yet covered above. It is really out of place here. ♣

Typically one has a fermionic field $\psi(x)$ and a Dirac-like-operator and the path integral involves an expression like

$$\int [d\psi d\bar{\psi}] \exp[\int i\bar{\psi} D\psi] \tag{24.65}$$

where $\int [d\psi d\bar{\psi}]$ is formally an infinite-dimensional version of the Berezin integral. At least formally this should be the determinant of D.

However, it is often the case that D is an operator between two different spaces, e.g. on an even-dimensional spin manifold M the chiral Dirac operator is an operator

$$D : L^2(M, S^+ \otimes E) \to L^2(M, S^- \otimes E) \tag{24.66}$$

where $S^{\pm}$ are the chiral spin bundles on M and E is a bundle with connection (in some representation of the gauge group). There is no canonical way of relating bases for these two Hilbert spaces, so $\det D$ must be an element of the determinant line $\mathrm{DET}(D)$.

If we have families of Dirac operators parametrized, say, by gauge fields, then we have a determinant line bundle $\mathrm{DET}(D)$ over that family, and $\det D$ is a section of that line bundle. If D is Fredholm then we can still define

$$\mathrm{DET}(D) = \Lambda^{mx}(\ker D)^{\vee} \otimes \Lambda^{mx}(\mathrm{cok} D) \tag{24.67}$$

♣Some of above paragraph now redundant ♣

The above remarks have implications for the theory of anomalies. In particular the *geometrical theory of anomalies* due to Atiyah and Singer.

In a general Lagrangian quantum field theory the path integral might look like

$$Z \sim \int_{\mathcal{B}} [d\phi] \int [d\psi d\bar{\psi}] \exp[S_{\mathrm{bosonic}}(\phi) + \int \bar{\psi} D_{\phi} \psi + S_{\mathrm{interaction}}] \tag{24.68}$$

where $\mathcal{B}$ is some space of bosonic fields. For example it might consist of the set of maps from a worldvolume $\mathcal{W}$ to a target manifold M, in the case of a nonlinear sigma model, or it might be the set of gauge equivalence classes of gauge fields, or some combination of these ingredients. Note that the Dirac-like operator D_{ϕ} typically depends on ϕ and $S_{\mathrm{interaction}}$ here indicates interactions of higher order in the fermions. Let us suppose that

the interactions between bosons and fermions can be handled perturbatively. If we first integrate over the fermions then we obtain an expression like

$$Z \sim \int_{\mathcal{B}} [d\phi] \exp[S_{\text{bosonic}}(\phi)] \det D_\phi \tag{24.69}$$

where $\det D_\phi$ is a section of a line bundle $DET(D_\phi)$ over $\mathcal{B}$, rather than a $\mathbb{C}^*$-valued function on $\mathcal{B}$.

This expression is meaningless, even at the most formal level, unless the line bundle has been trivialized with a trivial flat connection.

Remarks

1. In physical examples there is a natural connection on $\text{DET}(D_\phi)$ and it turns out that the vanishing of the curvature is the vanishing of the "perturbative anomaly."

2. If the perturbative anomaly is nonzero then the path integral does not even make formal sense. Of course, there are many other demonstrations using techniques of local field theory that the theory is ill-defined. But this is one elegant way to understand that fact.

3. There can be anomaly-canceling mechanisms. One of the most interesting is the *Green-Schwarz mechanism.* In this method one introduces an action $\exp[S_{GS}[\phi]]$ which is actually not a well-defined function from $\mathcal{B} \to \mathbb{C}^*$, but rather is a section of a line bundle over $\mathcal{B}$, where $\mathcal{B}$ is the space of bosonic fields. If the line bundle with connection is dual to that of the fermions so that $\mathcal{L}_{GS} \otimes \text{DET}(D_\phi)$ has a flat trivialization then $e^{S_{GS}[\phi]} \det D_\phi$ can be given a meaning as a well-defined function on $\mathcal{B}$. Then it can - at least formally - be integrated in the functional integral.

4. Even if the perturbative anomalies cancel, i.e. the connection is flat, if the space of bosonic fields is not simply connected then the flat connection on the determinant line can have nontrivial holonomy. This is the "global anomaly."

♣Actually, the general mechanism was known before GS, but GS is a very nice example of the general idea. ♣

In Section §24.7.3 below we will give perhaps the simplest illustration of a global anomaly.

24.7.2 Determinant of the one-dimensional Dirac operator

As a warmup, let us consider the odd-dimensional Dirac operator on the circle. This maps $L^2(S^1, S) \to L^2(S^1, S)$ where S is the spin bundle on the circle. There are two spin structures. The tangent bundle is trivial and hence we can simply think of spinors as complex functions on the circle with periodic or anti-periodic boundary conditions. So, concretely, we identify the Dirac operator coupled to a real $U(1)$ gauge field as

$$D_a = \frac{d}{dt} + ia(t) \tag{24.70}$$

where $t \sim t+1$ is a coordinate on S^1, $a(t) \in \mathbb{R}$ is periodic and identified via gauge transformations, and D acts on complex-valued functions which are periodic or antiperiodic.

The operator D_a is Fredholm and maps $\mathcal{H} \to \mathcal{H}$, $\det D_a$ as a complex number. The first simplification we can make is that we can gauge $a(t)$ to be constant. But we cannot remove the constant by a well-defined gauge transformation since $\oint a(t)dt$ is gauge invariant. Once this is done the eigenfunctions are $e^{2\pi i(n+\epsilon)t}$ where $n \in \mathbb{Z}$ and $\epsilon = \frac{1}{2}$ for AP and $\epsilon = 0$ for P boundary conditions. The eigenvalue is then $2\pi i(n+\epsilon) + ia$. Note that we can account for both boundary conditions by shifting $a \to a \pm 2\pi\epsilon$ so we will temporarily set $\epsilon = 0$ to simplify the formulae.

We could proceed by evaluating the ζ function. However, it is actually easier to proceed formally as follows:

$$\det D_a \stackrel{?}{=} \prod_{n\in\mathbb{Z}} (2\pi i n + ia) \tag{24.71}$$

It is good to rearrange this formal expression by putting together the positive and negative integers and separating out the $n = 0$ term:

$$\begin{aligned}\prod_{n\in\mathbb{Z}} (2\pi i n + ia) &= ia \prod_{n=1}^{\infty} (2\pi i n + ia)(-2\pi i n + ia) \\ &= (ia) \prod_{n=1}^{\infty} (2\pi n)^2 (1 - \frac{a^2}{(2\pi n)^2}) \\ &= \left(\prod_{n=1}^{\infty} (2\pi n)^2\right) \left(ia \prod_{n=1}^{\infty} (1 - \frac{a^2}{(2\pi n)^2})\right)\end{aligned} \tag{24.72}$$

Note that we can separate out the infinite factor $\prod_{n=1}^{\infty}(2\pi n)^2$, and what remains contains all the a-dependence and is in fact a well-defined product so we have:

$$\det(D_a) = (ia) \prod_{n=1}^{\infty} (1 - \frac{a^2}{(2\pi n)^2}) = 2i \sin(a/2) \tag{24.73}$$ eq:DetCF

where we have used the famous formula

$$\prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) = \frac{\sin \pi z}{\pi z} \tag{24.74}$$

This is the result for P-boundary conditions. The result for AP-boundary conditions is obtained by shifting $a \to a + \pi$

Remark: Using ζ-function regularization the "constant" $\left(\prod_{n=1}^{\infty}(2\pi n)^2\right)$ can be argued

to be in fact = 1 as follows:

$$\begin{aligned}\prod_{n=1}^{\infty}(2\pi n)^2 &= \exp[\sum_{n=1}^{\infty}\log(2\pi n)^2] \\ &= \exp[-\frac{d}{ds}|_{s=0}\sum_{n=1}(2\pi n)^{-2s}] \\ &= \exp[-\frac{d}{ds}|_{s=0}(2\pi)^{-2s}\zeta(2s)] \\ &= \exp[2\log(2\pi)\zeta(0) - 2\zeta'(0)] \\ &= 1\end{aligned} \tag{24.75}$$

where in the last line we used the expansion of the Riemann zeta function around $s = 0$:

$$\zeta(s) = -\frac{1}{2} - s\log\sqrt{2\pi} + \mathcal{O}(s^2) \tag{24.76}$$

The first equality above is formal. The second is a definition. The remaining ones are straightforward (and rigorous) manipulations.

-SpinStructure

24.7.3 A supersymmetric quantum mechanics

Now, let us consider a supersymmetric quantum mechanics with target space given by a Riemannian manifold $(M, g_{\mu\nu})$. If x^μ, $\mu = 1, \ldots, n = \dim M$, are local coordinates in a patch of M then there are maps from the one-dimensional worldline $(x^\mu(t), \psi^\mu(t))$. We can think of $(x^\mu(t), \psi^\mu(t))$ as functions on the supermanifold ΠTM and we have a map from $\mathbb{R}^{1|0} \to \Pi TM$. The action is:

$$S = \int dt\big\{g_{\mu\nu}(x(t))\dot{x}^\mu\dot{x}^\nu + i\psi^a[\frac{d}{dt}\delta^{ab} + \dot{x}^\mu(t)\omega_\mu^{ab}(x(t))]\psi^b\big\} \tag{24.77}$$

eq:SQM

Here $g_{\mu\nu}$ is a Riemannian metric on M and $\omega_\mu^{ab}dx^\mu$ is a spin connection. The $a, b = 1, \ldots, n$ refer to tangent space indices.

Let us consider just the theory of the fermions in a fixed bosonic background.

Consider briefly the Hamiltonian quantization of the system. Classically the bosonic field is just a point $x_0 \in M$. The canonical quantization relations on the fermions ψ^a gives a real Clifford algebra:

$$\{\psi^a, \psi^b\} = 1 \tag{24.78}$$

If $n = 2m$ then the irreducible representations are the chiral spin representations of M. Therefore, the wavefunctions of the theory are sections of a spinor bundle over M. Therefore, the Hilbert space is $L^2(M; S)$, the L^2 sections of the spin bundle over M. Therefore, if the theory is sensible, M should be spin. It is interesting to see how that constraint arises just by considering the path integral on the circle.

Let us consider the path integral on the circle, so $t \sim t + 1$. Again, we focus on the fermionic part of the path integral, so fix a loop $x : S^1 \to M$.

The fermionic path integral gives, formally pfaff(D_A) where

$$D_A = i\left(\frac{d}{dt}\delta^{ab} + A^{ab}(t)\right) \tag{24.79}$$

eq:DeeA

where $A^{ab}(t)$ is the real so$(2m)$ gauge field

$$A^{ab}(t) = \dot{x}^\mu(t)\omega_\mu^{ab}(x(t)). \tag{24.80}$$ eq:LoopGF

It is very useful to generalize the problem and consider the operator D_A in (24.79) for an *arbitrary* $SO(2m)$ gauge field $A^{ab}(t)$, with $a, b = 1, \ldots, 2m$. So we consider the Berezin integral

$$Z = \int [d\psi^a(t)] e^{\int_0^1 i\psi^a(t)\left(\frac{d}{dt}\delta^{ab} + A^{ab}(t)\right)\psi^b(t)dt} \tag{24.81}$$

Formally, this is just pfaff(D_A). In this infinite-dimensional setting we have two approaches to defining it:

1. We can consider the formal product of skew eigenvalues and regularize the product, say, using ζ-function regularization. Then we must choose a sign for each skew eigenvalue.

2. We can evaluate the determinant and attempt to take a squareroot.

We will explain (2), leaving (1) as an interesting exercise.

24.7.4 Real Fermions in one dimension coupled to an orthogonal gauge field

So, we want to evaluate $\det(D_A)$, where $D_A : L^2(S^1, \mathbb{R}^{2m}) \to L^2(S^1, \mathbb{R}^{2m})$. (Note we have complexified our fermions by doubling the degrees of freedom to compute the determinant.)

All the gauge invariant information is in $P\exp \oint A(t)dt \in SO(2m)$. By a constant orthogonal transformation the path ordered exponent can be put in a form

$$P\exp \oint A(t)dt = R(\alpha_1) \oplus R(\alpha_2) \oplus \cdots \oplus R(\alpha_m) \tag{24.82}$$ eq:Cartan

and by a single-valued gauge transformation $A^{ab}(t)$ can be gauged to a form which is t-independent. Recall that

$$R(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \tag{24.83}$$

so that the gauge invariant information only depends on $\alpha_i \sim \alpha_i + 2\pi$.

Therefore using gauge transformations we can reduce the problem of evaluating $\det(D_A)$ to the evaluation of the determinant of

$$D_\alpha = \frac{d}{dt} + \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix} \tag{24.84}$$ eq:RealD

We can diagonalize the matrix and hence we get the Dirac operator

$$D_\alpha = \frac{d}{dt} + \begin{pmatrix} i\alpha & 0 \\ 0 & -i\alpha \end{pmatrix} \tag{24.85}$$

Now, using the result (24.73) above we learn that

$$(\text{Pfaff}\, D_\alpha)^2 = 4\sin^2(\alpha/2) = \det(1 - R(\alpha)) \tag{24.86}$$ eq:twotwo

In general, for the antisymmetric operator D_A (24.79) coupled to any $SO(2m)$ gauge field on the circle we have

$$(\mathrm{Pfaff} D_A)^2 = \det(1 - \mathrm{hol}(A)) \tag{24.87}$$

Now we would like to take a square root of the determinant to define the Pfaffian. Let us consider a family of operators parametrized by $g \in SO(2m)$ with $P\exp \oint A dt = g$. Then $\det(1-g)$ is a function on $SO(2m)$ which is conjugation invariant, and hence we can restrict to the Cartan torus (24.82). It is clear that this does not have a well-defined square-root. If we try to take $\prod_{i=1}^m 2\sin(\alpha_i/2)$ then the expression has an ill-defined sign because we identify $\alpha_i \sim \alpha_i + 2\pi$. The expression *does* have a good meaning as a section of a principal $\mathbb{Z}_2$ bundle over $SO(2m)$. Put differently, if we pull back the function to $Spin(2m)$:

$$1 \to \mathbb{Z}_2 \to Spin(2m) \to SO(2m) \to 1 \tag{24.88}$$

then the function $\det(1-\tilde{g})$ (where we take the determinant in the $2m$-dimensional representation) does have a well-defined square-root. To see this it suffices to work again with the Cartan torus since the functions are conjugation invariant and therefore we need really only consider

$$1 \to \mathbb{Z}_2 \to Spin(2) \to SO(2) \to 1 \tag{24.89}$$

The group $Spin(2)$ is the group of even invertible elements

$$r(\beta) := \exp[\beta e_1 e_2/2] = \cos(\beta/2) + \sin(\beta/2) e_1 e_2 \tag{24.90}$$

in the Clifford algebra $C\ell_{+2}$. Here $\beta \sim \beta + 4\pi$. The projection to $SO(2)$ is given by $r(\beta) \mapsto R(\alpha)$ with $\alpha = \beta$, but a full period in α only lifts to a half-period in β. So on $Spin(2m)$ we have

$$\sqrt{\det(1-\tilde{g})} = \prod_{i=1}^m (2\sin\beta_i/2) \tag{24.91}$$

and this expression is well-defined.

Remark: In fact, this expression has a nice interpretation in terms of the characters in the chiral spin representations:

$$\prod_{i=1}^m (2\sin\beta_i/2) = i^{-m}\left(\mathrm{ch}_{S^+}(\tilde{g}) - \mathrm{ch}_{S^-}(\tilde{g})\right) \tag{24.92}$$

24.7.5 The global anomaly when M is not spin

Let us now return to our supersymmetric quantum mechanics above. We have learned that after integrating out the fermions the path integral on the circle is

$$Z(S^1) = \int [dx^\mu(t)] e^{-\int_0^1 dt g_{\mu\nu}(x(t))\dot{x}^\mu \dot{x}^\nu} \sqrt{\det(1 - \mathrm{Hol}(x^*\omega))} \tag{24.93}$$

where now for a given loop in the manifold $x : S^1 \to M$ we are using the holonomy of the orthogonal gauge field (24.80).

The question is: *Can we consistently define the sign of the square root over all of loop space* $LM = Map(S^1 \to M)$*?*

Let us fix a point $x_0 \in M$ and choose a basis for the tangent space at x_0. Then consider based loops $x(t)$ that begin and end at x_0. Then $\mathrm{Hol}(x^*\omega)$ defines a map from based loops $\Omega_{x_0}(M) \to SO(2m)$. If M is a spin manifold then there is a well-defined lift of this map to $Spin(2m)$. Then a well-defined square-root exists.

♣Put in commutative diagram ♣

On the other hand, it can be shown using topology that if M is not spin then there will be a family of based loops: $x^\mu(t;s) = x^\mu(t;s+1)$ so that at $s=0$ we have a constant map to a point $x^\mu(t;0) = x_0^\mu \in M$ and $x^\mu(t;s)$ loops around a nontrivial 2-sphere in M such that

♣Picture of the lasso of the 2-sphere ♣

$$\sqrt{\det(1-\mathrm{Hol}(x^*\omega))}|_{s=0} = -\sqrt{\det(1-\mathrm{Hol}(x^*\omega))}|_{s=1} \tag{24.94}$$

Thus, if M is not spin, the fermionic path integral in the SQM theory (24.77) cannot be consistently defined for all closed paths $x^\mu(t)$ and therefore the path integral does not makes sense. This is an example of a global anomaly.

Remark: Recall the exercise from §24.4. Apply this to the 2-form on LM defined by:

$$\omega = \oint_0^1 dt \delta x^a(t) \left(\frac{d}{dt}\delta^{ab} + \dot{x}^\mu(t)\omega_\mu^{ab}(x(t)) \right) \delta x^b(t) \tag{24.95}$$

One can argue that ω is closed and nondegenerate, hence it is a symplectic form. Then we can interpret the above remarks as the claim that *A manifold M is spin iff LM is orientable.*

24.7.6 References

1. M.F. Atiyah, "Circular Symmetry and the Stationary Phase Approximation," Asterisque
2. Atiyah and Singer, PNAS
3. O. Alvarez, I.M. Singer, and B. Zumino
4. G. Moore and P. Nelson, "Aetiology of sigma model anomalies,"
5. D. Freed, "Determinants, Torsion, and Strings," Commun. Math. Phys. **107**(1986)483
6. D. Freed and G. Moore, "Setting the Quantum Integrand of M Theory," hep-th/0409135.

25. Quadratic Forms And Lattices

Lattices show up in many ways in physics. The study of lattices in 2 and 3 dimensions is quite useful in solid state physics, in part because the types of atoms in a crystal are (by definition of a crystal) invariant under translation by a lattice. Higher dimensional lattices have also played a role in solid state physics, in the context of "quasicrystals" and also in the classification of quantum Hall states.

In this course we will encounter many very symmetric higher dimensional lattices as root lattices and weight lattices of Lie algebras. These encode many important aspects of the representation theory of compact Lie groups.

It turns out that many special lattices, such as even unimodular lattices play a distinguished role in string theory and in conformal field theory through vertex operator constructions. Lattices also play an important role in the study of compactifications of string theory.

Lattices of charges are again of importance in studying duality symmetries in supersymmetric quantum field theory and string theory.

In math lattices are studied for their own sake, as very beautiful objects, but they also have far flung connections to other branches of math such as number theory and error-correcting codes. See Conway & Sloane, *Sphere Packings, Lattices, and Groups* for a comprehensive survey. Finally, they are also important in topology (as intersection forms), especially in the topology of four-manifolds.

25.1 Definition

The word "lattice" means different things to different people. For some people it is a finitely generated free abelian group. In this case, up to isomorphism there is just one invariant: the rank, and any "lattice" in this sense is just $\mathbb{Z}^r$, up to isomorphism. To some people it is a regular array of points in some space (we will regard these as lattices embedded in a space.) In these notes we take a "lattice" to be a finitely generated free Abelian group with the extra data of a nondegenerate symmetric bilinear form: [63]

Definition A *lattice* Λ is a free abelian group equipped with a nondegenerate, symmetric bilinear quadratic form:

$$\langle \cdot, \cdot \rangle : \Lambda \times \Lambda \to R \tag{25.1}$$

where R is a $\mathbb{Z}$-module. Thus:

1. $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle, \qquad \forall v_1, v_2 \in \Lambda.$
2. $\langle nv_1 + mv_2, v_3 \rangle = n\langle v_1, v_3 \rangle + m\langle v_2, v_3 \rangle, \qquad \forall v_1, v_2, v_3 \in \Lambda$, and $n, m \in \mathbb{Z}$.
3. $\langle v, w \rangle = 0$ for all $w \in \Lambda$ implies $v = 0$.

When $R = \mathbb{Z}$ we say we have an *integral lattice*. We will also consider the cases $R = \mathbb{Q}, \mathbb{R}$.

We say that two lattice $(\Lambda_1, \langle \cdot, \cdot \rangle_1)$ and $(\Lambda_2, \langle \cdot, \cdot \rangle_2)$ are *equivalent* if there is a group isomorphism $\phi : \Lambda_1 \to \Lambda_2$ so that $\phi^*(\langle \cdot, \cdot \rangle_2) = \langle \cdot, \cdot \rangle_1$. The *automorphism group* of the lattice is the group of ϕ's which are isomorphisms of the lattice with itself. These can be a finite or infinite discrete groups and can be very interesting.

There is a simple way of thinking about lattices in terms of matrices of integers. A (finitely generated) free abelian group of rank n is isomorphic, as a group, to $\mathbb{Z}^n$. Therefore,

[63] Hence, sometimes the term "quadratic module" is used.

we can choose an ordered integral basis $\{e_i\}_{i=1}^n$ for the lattice (that is, a set of generators for the abelian group) and then define the $n \times n$ *Gram-Matrix*

$$G_{ij} := \langle e_i, e_j \rangle. \tag{25.2}$$

Of course, the basis is not unique, another one is defined by

$$\tilde{e}_i := \sum_j S_{ji} e_j \tag{25.3}$$ eq:chgbss

where, now, the matrix S must be both *invertible* and *integral valued*, that is

$$S \in GL(n, \mathbb{Z}) \tag{25.4}$$ eq:intgrlequiv

Under the change of bases (25.3) the Gram matrix changes to

$$G \to \tilde{G} = S^{tr} G S \tag{25.5}$$ eq:intgrlequiv

So lattices can be thought of as symmetric nondegenerate matrices of integers with an equivalence relation given by (25.5).

Remark. As we will soon begin to see, the classification of lattices is somewhat nontrivial. In fact, it is an extremely subtle and beautiful problem, only partially solved.

By contrast, the classification of integral antisymmetric forms is fairly straightforward. Any such form can be brought by an integral transformation to the shape: [64]

♣Clarify which addition of Lang's book you are referring to. ♣

$$\begin{pmatrix} 0 & d_1 \\ -d_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & d_2 \\ -d_2 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & d_3 \\ -d_3 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & d_k \\ -d_k & 0 \end{pmatrix} \tag{25.6}$$ eq:antisymm

and this form is unique if we require $d_i > 0$ and $d_1 | d_2 | \cdots | d_k$. This is important in the quantization of certain mechanical systems with compact coordinates and momenta. It classifies the integral symplectic forms on a torus, for example.

Exercise

a.) Show that the group of integer invertible matrices whose inverses are also integer matrices form a group. This is the group $GL(n, \mathbb{Z})$.

Note that it is *not* the same as the set of integer matrices which are invertible. For example

$$\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} \in GL(2, \mathbb{Z}) \tag{25.7}$$

but

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \notin GL(2, \mathbb{Z}) \tag{25.8}$$

[64] For a proof see Lang, Algebra, p. 380.

b.) Show that for $S \in GL(n,\mathbb{Z})$, we necessarily have $|\det S| = 1$.

c.) $SL(n,\mathbb{Z})$ is the subgroup of matrices of determinant 1. What is the center of $SL(n,\mathbb{Z})$?

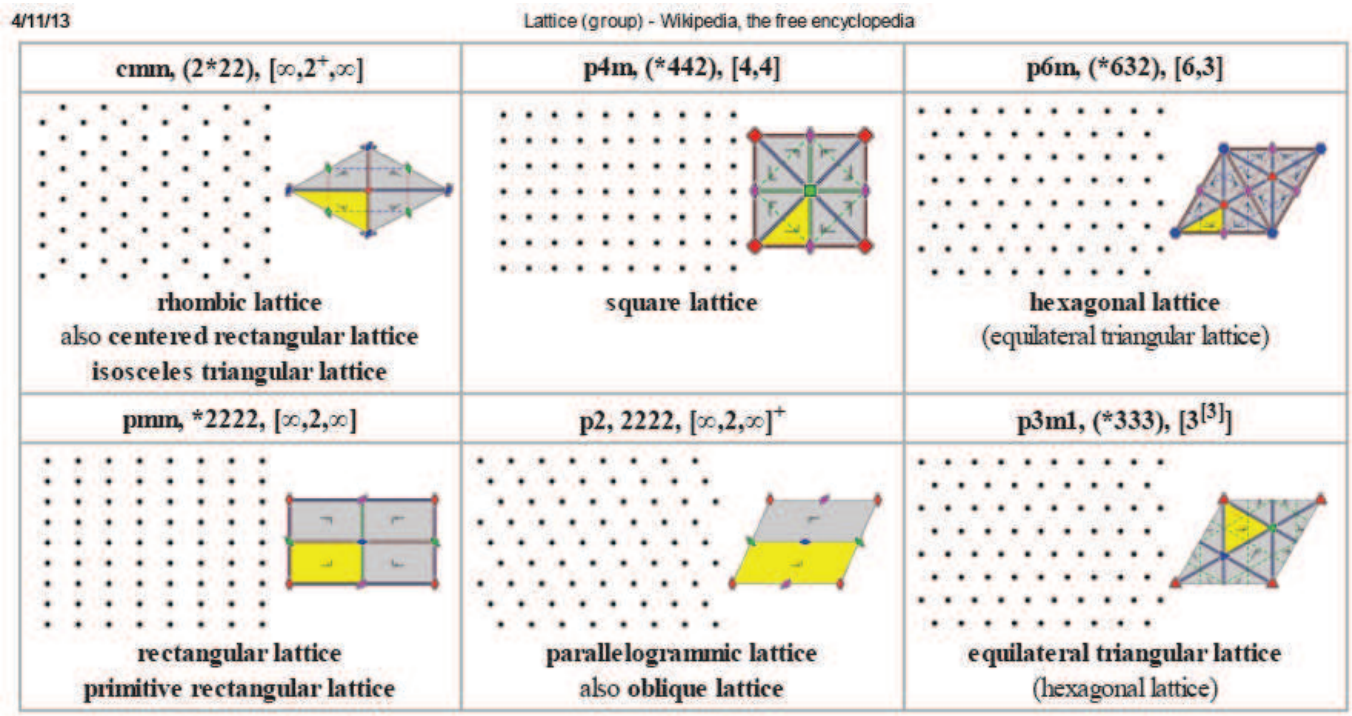


Figure 17: A picture of some important two-dimensional lattices embedded into Euclidean $\mathbb{R}^2$. From Wikipedia. fig:latticeii

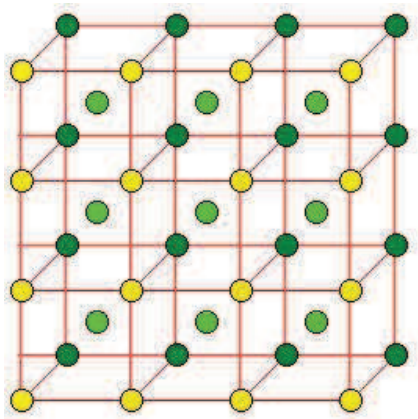


Figure 18: A three-dimensional lattice, known as the body centered cubic lattice. fig:BodyCenter

25.2 Embedded Lattices

Quite often we do not think of lattices in the above abstract way but rather as a discrete

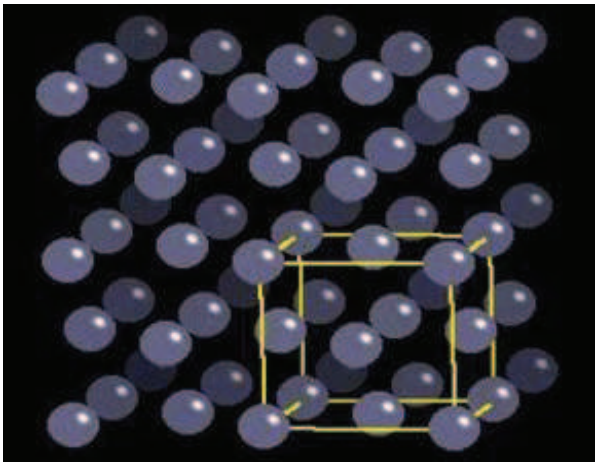


Figure 19: A three-dimensional lattice, known as the face centered cubic lattice. fig:FCC-Lattic

subgroup of $\mathbb{R}^m$. [65] See for example Figure 17, above for some rank 2 lattices in $\mathbb{R}^2$ and Figures 18 and 19 for some embedded lattices in $\mathbb{R}^3$.

To describe an embedded lattice we can consider the generators to be linearly independent vectors $\vec{e}_1, \cdots, \vec{e}_n \in \mathbb{R}^m$. (Necessarily, $m \geq n$). Define

$$\Lambda \equiv \{\sum_{i=1}^{n} \ell_i \vec{e}_i | \ell_i \in \mathbb{Z}\}. \tag{25.9}$$ eq:lattice

As an Abelian group, under vector addition, Λ is isomorphic to $\mathbb{Z}^n$. Moreover, if $\mathbb{R}^m$ is equipped with a symmetric quadratic form (e.g. the Euclidean metric) then the lattice inherits one:

$$\langle \cdot, \cdot \rangle : \Lambda \times \Lambda \to \mathbb{R} \tag{25.10}$$

We simply restrict the quadratic form to the subset $\Lambda \subset \mathbb{R}^m$. (We can also go the other way: The tensor product

$$\Lambda \otimes_{\mathbb{Z}} \mathbb{R} \cong \mathbb{R}^n \tag{25.11}$$

so by extending scalars from $\mathbb{Z}$ to $\mathbb{R}$ a quadratic form on an abstract rank n lattice determines one on $\mathbb{R}^n$.)

If the coordinates of the vectors are $\vec{e}_i = (e_{i1}, \ldots, e_{im})$ (so we view vectors as $1 \times m$ matrices) then we can form an $n \times m$ *generating matrix*

$$M = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ \vdots & \vdots & & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix} \tag{25.12}$$

The lattice is the set of vectors ξM where $\xi \in \mathbb{Z}^n$ is viewed as a $1 \times n$ matrix. If we use the Euclidean metric on $\mathbb{R}^m$ to induce the bilinear form on Λ then the Gram-Matrix is the

[65] By a *discrete subgroup* we mean, heuristically, that there are no accumulation points. Technically, the action on G should be properly discontinuous.

$n \times n$ matrix, $G = MM^{tr}$. Different generating matrices are related by $M \mapsto S^{tr}M$, for $S \in GL(n, \mathbb{Z})$.

Example 1: The most obvious example is $\Lambda = \mathbb{Z}^n \subset \mathbb{R}^n$. For $n = 2$ this is a square lattice, for $n = 3$ it is the simple cubic lattice. For general n we will refer to it as a "hypercubic lattice." The automorphisms will be linear transformations on the vectors, and using the standard basis we can identify them with $n \times n$ matrices. The matrices must be integral matrices to preserve $\mathbb{Z}^n$. But they must also be in in $O(n; \mathbb{R})$ to preserve the quadratic form $M^{tr}M = \mathbf{1}$, that is $S^{tr}\mathbf{1}S = \mathbf{1}$. Since the rows and columns must square to 1 and be orthogonal these are signed permutation matrices. Therefore

$$\mathrm{Aut}(\mathbb{Z}^n) = \mathbb{Z}_2^n \rtimes S_n \tag{25.13}$$ eq:AutZn

where S_n acts by permuting the coordinates $(x_1, \ldots, x_n)$ and $\mathbb{Z}_2^n$ acts by changing signs $x_i \to \epsilon_i x_i$, $\epsilon_i \in \{\pm 1\}$.

Example 2 As a good example of the utility of allowing $m > n$ let us define:

$$A_n := \{(x_0, x_1, \ldots, x_n) \in \mathbb{Z}^{n+1} | \sum_{i=0}^{n} x_i = 0\} \subset \mathbb{R}^{n+1} \tag{25.14}$$ eq:AnDef

A group of automorphisms of the lattice A_n is is rather obvious from (25.14), namely the symmetric group S_{n+1} acts by permutation of the coordinates. Another obvious symmetry is $(x_0, \ldots, x_n) \mapsto (-x_0, \ldots, -x_n)$. These generate the full automorphism group

$$\mathrm{Aut}(A_n) = \mathbb{Z}_2 \times S_{n+1} \qquad n > 1 \tag{25.15}$$

for $n = 1$, $A_1 \cong \sqrt{2}\mathbb{Z} \subset \mathbb{R}$ and the automorphism group is just $\mathbb{Z}_2$.

A nice basis is given by

$$\alpha_i = \vec{e}_i - \vec{e}_{i-1} \qquad i = 1, \ldots, n \tag{25.16}$$ eq:Anbasis

where $\vec{e}_i$, $i = 0, \ldots, n$, are the standard basis vectors in $\mathbb{R}^{n+1}$. The Gram matrix is then the famous Cartan matrix for A_n:

$$G_{ij} = C_{ij} = 2\delta_{i,j} - \delta_{i,j-1} - \delta_{i,j+1} \qquad i, j = 1, \ldots, n \tag{25.17}$$

The corresponding matrix is tridiagonal:

$$C(A_n) = \begin{pmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & \cdots & \cdots & 0 \\ 0 & -1 & 2 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 2 & -1 & 0 \\ 0 & \cdots & \cdots & -1 & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 2 \end{pmatrix} \tag{25.18}$$

Note that we could project two basis vectors of A_2 into a plane $\mathbb{R}^2$ to get

$$\begin{aligned}\alpha_1 &= \sqrt{2}(1,0)\\ \alpha_2 &= \sqrt{2}(-\frac{1}{2},\frac{\sqrt{3}}{2})\end{aligned} \tag{25.19}$$

eq:atwo

and as a check we can use these to compute

♣Give a figure and say how the projection is done. ♣

$$C(A_2) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{25.20}$$

Using these vectors we generate a beautifully symmetric hexagonal lattice in the plane.

♣ FIGURE OF HEXAGONAL LATTICE HERE ♣

Similarly we have

$$C(A_3) = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \tag{25.21}$$

and we could also realize the lattice as the span of vectors in $\mathbb{R}^2$

$$\begin{aligned}\alpha_1 &= \\ \alpha_2 &= \\ \alpha_3 &=\end{aligned} \tag{25.22}$$

eq:atwo

Example 3 Consider the set of points $D_n \subset \mathbb{Z}^n$ defined by

$$D_n := \{(x_1,\dots,x_n) \in \mathbb{Z}^n | x_1 + \cdots + x_n = 0 \mod 2\} \tag{25.23}$$

If we embed into $\mathbb{R}^n$ in the obvious way then we can use the Euclidean metric to induce an integer-valued quadratic form. To get some intuition, let us consider D_3. This is known as the "face-centered-cubic" or fcc lattice. To justify the name note that there is clearly a lattice proportional to the cubic lattice and spanned by $(2,0,0)$, $(0,2,0)$, and $(0,0,2)$. However in each xy, xz, and yz plane the midpoint of each 2×2 square is also a lattice vector. Choosing such midpoint lattice vectors in each plane gives a generating matrix:

$$M = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & -1 & 0 \end{pmatrix} \tag{25.24}$$

and one then computes

$$MM^{tr} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \tag{25.25}$$

so in fact $D_3 \cong A_3$. (This reflects a special isomorphism of simple Lie algebras.)

Example 4 The n-dimensional *bcc lattice*, BCC_n, where *bcc* stands for "body-centered cubic" is the sublattice of $\mathbb{Z}^n$ consisting of $(x_1, \ldots, x_n)$ so that the x_i are either all even or all odd. Note that if all the x_i are odd then adding $\vec{e}$ produces a vector with all x_i even, where $\vec{e} = (1, 1, \ldots, 1) = \vec{e}_1 + \cdots + \vec{e}_n$. Therefore, we can write:

$$BCC_n = 2\mathbb{Z}^n \cup (2\mathbb{Z}^n + \vec{e}) \tag{25.26}$$

Clearly $2\mathbb{Z}^n$ is proportional to the "cubic" lattice. Adding in the orbit of $\vec{e}$ produces one extra lattice vector inside each n-cube of side length 2, hence the name bcc.

Example 5: Now take $\mathbb{R}^2$ as a vector space but we do not use the Euclidean metric on $\mathbb{R}^2$ to induce the bilinear form on Λ. Rather we use the Minkowski signature metric

$$\mathbb{R}^{1,1} = \{(t, x) | \langle (t, x), (t, x) \rangle = -t^2 + x^2\} \tag{25.27}$$ eq:2DMink

Let $R > 0$, and consider the lattice $\Lambda(R)$ generated by

$$\begin{aligned} e_1 &= \frac{1}{\sqrt{2}}(\frac{1}{R}, \frac{1}{R}) \\ e_2 &= \frac{1}{\sqrt{2}}(-R, R) \end{aligned} \tag{25.28}$$ eq:nafi

Note that for any R we have simply:

$$G_{ij} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{25.29}$$

So, manifestly, as lattices these are all isomorphic, although as embedded lattices they depend on R.

♣ FIGURE OF e_1, e_2 IN THE PLANE ♣

Remarks: Solids comprised of a single element will form simple three-dimensional lattices in nature, at least in the limit that they are infinitely pure and large. Those on the LHS of the periodic table tend to be bcc e.g. the alkali metals of column one and Ba, Ra, while those towards the right tend to be fcc (e.g. Cu, Ni, Ag, Au, Pt, Ir, Al, Pb) or the column of noble gases (except He).

Exercise

If Λ is a lattice, let 2Λ be the lattice of elements divisible by 2, i.e., vectors $\vec{v}$ such that $\frac{1}{2}\vec{v} \in \Lambda$. Show that 2Λ is a subgroup of Λ. Suppose $\vec{v}$ is not divisible by 2. Is $2\Lambda + \vec{v}$ a subgroup?

Exercise *Automorphisms of* $\mathbb{Z}^n$

Check that the group of signed permutations is isomorphic to the semidirect product (25.13). Write explicitly $\alpha : S_n \to \mathrm{Aut}(\mathbb{Z}_2^n)$.

25.3 Some Invariants of Lattices

What can we say about the classification of lattices?

If we were allowed to take $S \in GL(n, \mathbb{R})$ then Sylvester's theorem guarantees that we can change basis to put the Gram matrix into diagonal form so that

$$\tilde{G}_{ij} = Diag\{-1^t, +1^s\} \tag{25.30}$$

This provides us with two important invariants of the lattice, the signature and the rank. The rank is

$$r = t + s \tag{25.31}$$

and we will define the signature to be

$$\sigma = s - t \tag{25.32}$$

When we work of $R = \mathbb{Z}$ there are going to be more invariants:

Example: Consider two lattices:

A. $\Lambda_A = \mathbf{e_1}\mathbb{Z} \oplus \mathbf{e_2}\mathbb{Z}$ with form:

$$G_A = \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \tag{25.33}$$

B. $\Lambda_B = \mathbf{e_1}\mathbb{Z} \oplus \mathbf{e_2}\mathbb{Z}$ with form:

$$G_B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{25.34}$$

We ask:

Can these be transformed into each other by a change of basis with $S \in GL(2, \mathbb{Z})$*?*

The answer is clearly "yes" over $\mathbb{R}$ because they both have Lorentzian signature.

The answer is clearly "no" over $\mathbb{Z}$ because the norm-square of any vector in Λ_B is even $(n_1 e_1 + n_2 e_2)^2 = 2n_1 n_2$, while this is not true of Λ_A.

The lattice Λ_B is an important lattice, it is denoted by $II^{1,1}$ or by $H(1)$. [66]

Definition. A lattice Λ is called an *even lattice* if, for all $\mathbf{x} \in \Lambda$

$$\langle \mathbf{x}, \mathbf{x} \rangle \in 2\mathbb{Z} \tag{25.35}$$

(Note: This does not preclude $\langle x, y \rangle$ being odd for $x \neq y$.) A lattice is called *odd* if it is not even.

[66] Some authors use the notation $U(1)$. We do not use this notation since it can cause confusion.

Note that under $G \to S^{tr}GS$,

$$(S^{tr}GS)_{ii} = \sum_k (S_{ki})^2 G_{kk} + 2\sum_{k<j} G_{kj}S_{ki}S_{ji} \tag{25.36}$$

Now, $S \in GL(n,\mathbb{Z})$, so the S_{ij} are integers, so if the diagonal elements of G are even in one basis then they are even in all bases.

In order to describe our next invariant of lattices we need to introduce the dual lattice.

Given a lattice Λ we can define the *dual lattice* [67]

$$\Lambda^* := \mathrm{Hom}_{\mathbb{Z}}(\Lambda, \mathbb{Z}) \tag{25.37}$$ eq:dualf

where Hom means a $\mathbb{Z}$-linear mapping.

As we have now seen several times, given the data of a bilinear form there is a $\mathbb{Z}$-linear map $\ell : \Lambda \to \Lambda^*$ defined by

$$\ell(v)(v') := \langle v, v' \rangle. \tag{25.38}$$

This has no kernel for a nondegenerate form and hence we can consider $\Lambda \subset \Lambda^*$ and so we may form:

$$D(\Lambda) := \Lambda^*/\Lambda \tag{25.39}$$

This abelian group is known as the *discriminant group*, or *glue group*.

Next we make Λ^* into a lattice by declaring ℓ to be an isometry onto its image:

$$\langle v, w \rangle_\Lambda = \langle \ell(v), \ell(w) \rangle_{\Lambda^*} \tag{25.40}$$ eq:isometry

We then extend to the rest of Λ^* to make it a lattice,

To make this more concrete suppose e_i is a basis for Λ and let $\hat{e}^i$ be the dual basis for Λ^* so that

$$\hat{e}^i(e_j) = \delta^i{}_j \tag{25.41}$$

It follows that

$$\ell(e_i) = \sum_j G_{ij}\hat{e}^j \tag{25.42}$$

Now, using (25.40) it follows that Λ^* has the Gram matrix

$$\langle \hat{e}^i, \hat{e}^j \rangle = G^{ij} \tag{25.43}$$

where $G^{ij}G_{jk} = \delta^i{}_k$.

Note that in general Λ^* is not an integral lattice since G^{ij} will be a rational matrix if G_{ij} is an integral matrix. Let us denote

$$g := \det G_{ij} \tag{25.44}$$

[67] The lattice Λ^* is closely related to the *reciprocal lattice* of solid state physics. However, there are some differences. Conceptually, the reciprocal lattice is an embedded lattice, embedded in momentum space. Moreover, there are some normalization differences by factors of 2π. More importantly, the reciprocal lattice depends on things like lattice spacings.

$$G^{ij} = \frac{1}{g}\widehat{G}^{ij} \tag{25.45}$$

where $\widehat{G}^{ij}$ is a matrix of integers.

Lemma The discriminant group is a *finite* abelian group of order g

Proof: To see that it is finite we note that

$$g\hat{e}^j = \ell(\widehat{G}^{ij}e_j) \tag{25.46}$$

and hence $[g\hat{e}^j] = 0$ in the discriminant group. Therefore, every element is torsion and hence the group is finite. By the classification of finite abelian groups we see that the order $|D(\Lambda)|$ divides g.

In fact $|D(\Lambda)| = g$ as the following argument shows:

If $\Lambda \subset \mathbb{R}^n$ is an embedded lattice of maximal rank, and both the dual pairing and the Gram-matrix are inherited from the standard Euclidean bilinear form on $\mathbb{R}^n$ and we may write:

$$\begin{aligned} \hat{e}^i &= \sum_j G^{ij}e_j \\ e_i &= \sum_j G_{ij}\hat{e}^j \end{aligned} \tag{25.47}$$

eq:dualbs

Now use the notion of fundamental domain defined in the next chapter. By comparing the volume of a unit cell of Λ^* to that of Λ we find:

$$|D(\Lambda)| = \frac{\sqrt{\det G_{ij}}}{\sqrt{\det G^{ij}}} = \det G_{ij} \tag{25.48}$$

concluding the proof ♠

♣Would be better to give an argument that does not use the fundamental domain. ♣

Moreover, $D(\Lambda)$ inherits a bilinear form valued in $\mathbb{Q}/\mathbb{Z}$. (Recall that an abelian group is a $\mathbb{Z}$-module and one can define bilinear forms modules over a ring.) Specifically, we define

$$b\left([v_1],[v_2]\right) = \langle v_1, v_2\rangle \mathrm{mod}\mathbb{Z} \tag{25.49}$$

The finite group $D(\Lambda)$ together with its bilinear form to $\mathbb{Q}/\mathbb{Z}$ is an invariant of the lattice.

Example 1: Consider $\Lambda = \nu\mathbb{Z} \subset \mathbb{R}$. We use the standard Euclidean metric on $\mathbb{R}$ so that ν^2 must be an integer n. Then $\Lambda^* = \frac{1}{\nu}\mathbb{Z}$. Note that $\Lambda \subset \Lambda^*$, indeed, $D(\Lambda) \cong \mathbb{Z}/n\mathbb{Z}$ so

$$[\Lambda^* : \Lambda] = n \tag{25.50}$$

There are only two choices of basis for Λ, namely $\mathbf{e_1} = \pm\nu$. The Gram matrix is $G_{11} = \nu^2 = n$. The bilinear form on the discriminant group is

$$b\left(\frac{r}{\nu} + \nu\mathbb{Z}, \frac{s}{\nu} + \nu\mathbb{Z}\right) = \frac{rs}{n}\mathrm{mod}\mathbb{Z} \tag{25.51}$$

Example 2: A_1^*. The Gram matrix is just the 1×1 matrix 2 so if $A_1 = \mathbb{Z}\alpha$ then $A_1^* = \mathbb{Z}\lambda$, with $\lambda = \frac{1}{2}\alpha$. The discriminant group is clearly $\mathbb{Z}_2$.

Example 3: A_2^*: Consider the Cartan matrix

$$C(A_2) = G_{ij} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{25.52}$$ eq:gmtx

and $\det G_{ij} = 3$.

We easily compute

$$G^{ij} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{25.53}$$ eq:gmtxop

and hence the dual basis is given by

$$\begin{aligned} \lambda^1 &= \frac{1}{3}(2\alpha_1 + \alpha_2) \\ \lambda^2 &= \frac{1}{3}(\alpha_1 + 2\alpha_2) \end{aligned} \tag{25.54}$$ eq:dualbas

The group $\Lambda^*/\Lambda \cong \mathbb{Z}_3$. Since $\alpha_1 = 2\lambda^1 - \lambda^2, \alpha_2 = -\lambda^1 + 2\lambda^2$, $2\lambda^1 = \lambda^2 \mathrm{mod}\Lambda$. So one set of representatives is given by $0\mathrm{mod}\Lambda, \lambda^1\mathrm{mod}\Lambda, 2\lambda^1\mathrm{mod}\Lambda$. Alternatively, we could take $\lambda^2\mathrm{mod}\Lambda$ as the generator.

If we take the embedding *** above then

$$\begin{aligned} \lambda^1 &= (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{6}}) \\ \lambda^2 &= (0, \sqrt{\frac{2}{3}}) \end{aligned} \tag{25.55}$$ eq:dualbasp

generate a triangular lattice.

♣ FIGURE ♣

As we shall see, the hexagonal lattice Λ is the root lattice of $SU(3)$, while Λ^* is the weight lattice.

Example 4: A_n^*. One could just invert the Cartan matrix and proceed as above. (See exercise below.) However, an alternative route is to view A_n as embedded in $\mathbb{Z}^{n+1} \otimes \mathbb{R}$ as above. Then relative to the basis α_i (25.16) above we will find a dual basis:

$$\lambda^i \cdot \alpha_j = \delta^i{}_j. \tag{25.56}$$

Writing out the equation in components one easily finds (and even more easily checks):

$$\lambda^1 = \left(-\frac{n}{n+1}, \underbrace{\frac{1}{n+1}, \ldots, \frac{1}{n+1}}_{n \text{ times}} \right) \tag{25.57}$$

Now that we have λ^1 it is also easy to solve for the λ^i, $i > 1$ in terms of λ^1 from:

$$\begin{aligned} \alpha_1 &= 2\lambda^1 - \lambda^2 \\ \alpha_2 &= -\lambda^1 + 2\lambda^2 - \lambda^3 \\ \vdots &\qquad \vdots \\ \alpha_n &= -\lambda^{n-1} + 2\lambda^n \end{aligned} \tag{25.58}$$

to get

$$\begin{aligned} \lambda^2 &= 2\lambda^1 - \alpha_1 \\ \lambda^3 &= 3\lambda^1 - 2\alpha_1 - \alpha_2 \\ \lambda^4 &= 4\lambda^1 - 3\alpha_1 - 2\alpha_2 - \alpha_1 \\ \vdots &\qquad \vdots \\ \lambda^{n+1} &= (n+1)\lambda^1 - n\alpha_1 - (n-1)\alpha_2 - \cdots - \alpha_n \end{aligned} \tag{25.59}$$

and explicit substitute of the vectors shows that $\lambda^{n+1} = 0$.

Thus

$$\lambda^i = \left(\underbrace{-\frac{i}{n+1}, \ldots, -\frac{i}{n+1}}_{j \text{ times}}, \underbrace{\frac{j}{n+1}, \ldots, \frac{j}{n+1}}_{i \text{ times}} \right) \tag{25.60}$$

where $i + j = n + 1$. ♣CHECK THIS!! ♣

Thus, the discriminant group is cyclic,

$$D(A_n) \cong \mathbb{Z}/(n+1)\mathbb{Z} \tag{25.61}$$

and is generated, for example, by $[\lambda^1]$. Therefore, to compute the quadratic form it suffices to compute

$$b([\lambda^1], [\lambda^1]) = -\frac{1}{n+1} \mathrm{mod}\mathbb{Z} \tag{25.62}$$

The inverse Cartan matrix is given in the exercise below.

Example 5: D_n^*: We claim the dual lattice of the "n-dimensional fcc lattice" is one half of the "n-dimensional bcc lattice":

$$D_n^* = \frac{1}{2} BCC_n \tag{25.63}$$

eq:fcc-dual

To see this note first that if $x \in BCC_n$ and $y \in D_n$ then $x \cdot y$ is even. This is obvious if all the x_i are even, and if they are all odd then $\sum_i x_i y_i = \sum_i (2n_i + 1) y_i = \sum_i y_i = 0 \mathrm{mod} 2$, by the definition of D_n. Therefore $\frac{1}{2} BCC_n \subset D_n^*$. Conversely, if $v \in D_n^*$ then $2v_i$ must be integer, since $2e_i \in D_n$, and moreover, looking at the products with $(1, -1, 0, \ldots, 0)$, $(0, 1, -1, 0, \ldots, 0)$ and so forth gives

$$\begin{aligned} v_1 - v_2 &= k_1 \\ v_2 - v_3 &= k_2 \\ &\vdots\vdots \\ v_{n-1} - v_n &= k_{n-1} \end{aligned} \tag{25.64}$$

Multiply these equations by 2. Then $2v_i$ are integers, and on the RHS we have even integers. Therefore the $2v_i$ are all even or all odd. Therefore $D_n^* \subset \frac{1}{2}BCC_n$, and this establishes (25.63).

♣So give the discriminant group! ♣

Remark: Combining Example 5 with the observation above on the periodic table we see that the periodic table has a (very approximate) self-duality! It works best for exchanging the first and last column (excluding H, He). A conceptual reason for this is the following. [68] The noble gases have a filled electron shell and to a good approximation act as hard spheres. So their crystal structure should be a minimal sphere packing in three-dimensional space. This means they should be hcp or fcc, and many-body effects break the degeneracy to fcc. On the other hand, in the first column we have a filled shell with a single electron. Therefore, to a good approximation the metals can be treated in the free one-electron picture. Then their Fermi surface is a sphere in *momentum space*. Now, this Fermi surface is a good approximation to the boundary of the Wigner-Seitz cell. Therefore the crystal structure is given by solving a sphere-packing problem in momentum space! Fcc in momentum space implies bcc in real space.

Exercise *Inverse of a generalized Cartan matrix*

Let a_α, $\alpha = 1, \ldots, r$ be a set of positive integers and consider the generalized Cartan matrix

$$G_{\alpha\beta} = a_\alpha \delta_{\alpha,\beta} - \delta_{\alpha+1,\beta} - \delta_{\alpha-1,\beta} \tag{25.65}$$

a.) Show that the inverse matrix is

$$G^{\alpha\beta} = \begin{cases} \frac{1}{n} q_\alpha p_\beta & 1 \le \alpha \le \beta \le r \\ \frac{1}{n} p_\alpha q_\beta & 1 \le \alpha \le \beta \le r \end{cases} \tag{25.66}$$

Where $n = \det G_{\alpha\beta}$ and the integers p_α, q_α and n are defined as follows.

Define $[x, y] := x - \frac{1}{y}$, then $[x, y, z] = [x, [y, z]]$, then $[x, y, z, w] = [x, [y, z, w]]$, etc. That is, these are continued fractions with signs. Now in terms of these we define p_α, q_α from

$$\begin{aligned} \frac{p_{j-1}}{p_j} &= [a_j, a_{j+1}, \ldots, a_r] \\ \frac{q_{j+1}}{q_j} &= [a_j, a_{j-1}, \ldots, a_1] \qquad\qquad 1 \le j \le r \end{aligned} \tag{25.67}$$

with boundary conditions $q_1 = 1, p_r = 1$.

b.) Show that $n = p_0$.

c.) Show that

$$[\underbrace{2, 2, \ldots, 2}_{r \text{ times}}] = \frac{r+1}{r} \tag{25.68}$$

♣So give the inverse Cartan matrix in matrix form ♣

---

[68] This argument was worked out with K. Rabe.

25.3.1 The characteristic vector

The next useful invariant of a lattice is based on a

Definition A *characteristic vector* on an integral lattice is a vector $w \in \Lambda$ such that

$$\langle v, v\rangle = \langle w, v\rangle \mathrm{mod} 2 \tag{25.69}$$

eq:chrctvctr

for every $v \in \Lambda$.

Lemma A characteristic vector always exists.

Proof: Consider the lattice $\Lambda/2\Lambda$. Denote elements in the quotient by $\bar{v}$. Note that the quadratic form $Q(v) = \langle v, v\rangle$ descends to a $\mathbb{Z}_2$-valued form $q(\bar{v}) = \langle \bar{v}, \bar{v}\rangle \mathrm{mod} 2$. Moreover, over the field $\kappa = \mathbb{Z}_2$ note that q is *linear*:

$$q(\bar{v}_1 + \bar{v}_2) = q(\bar{v}_1) + q(\bar{v}_2) + 2\langle \bar{v}_1, \bar{v}_2\rangle = q(\bar{v}_1) + q(\bar{v}_2) \tag{25.70}$$

But any linear function must be of the form $q(\bar{v}) = \langle \bar{v}, \bar{w}\rangle$. Now let $w \in \Lambda$ be any lift of $\bar{w}$. This will do. ♠

Note that characteristic vectors are far from unique. Indeed, if w is a characteristic vector and v is any other vector then $w + 2v$ is characteristic. Moreover, any characteristic vector is of this form if the form is nondegenerate when reduced mod two. Therefore the quantity

$$\mu(\Lambda) := \langle w, w\rangle \mathrm{mod} 8 \tag{25.71}$$

does not depend on the choice of w and is an invariant of the lattice Λ.

Remark There is a great deal of magic associated with the number 8 in lattice theory.

♣Say more? ♣

25.3.2 The Gauss-Milgram relation

♣ALSO DERIVE THIS FROM THE MODULAR TRANSFORMATION LAW FOR LATTICE THETA FUNCTIONS: TAKE LIMIT AS Im tau GOES TO INFINITY AND COMPARE ♣

The invariants we have just described are all related by a beautiful formula sometimes called the Gauss-Milgram sum formula:

Let Λ be an integral lattice. Choose a characteristic vector $w \in \Lambda$ and define the quadratic function

$$Q : \Lambda \otimes \mathbb{R} \to \mathbb{R} \tag{25.72}$$

by

$$Q(v) = \frac{1}{2}\langle v, v - w\rangle \tag{25.73}$$

Note that Q takes integral values on Λ and rational values on Λ^*. Moreover, if $x \in \Lambda^*$ note that

$$Q(x + v) = Q(x) + \langle x, v\rangle + \frac{1}{2}\langle v, v - w\rangle \tag{25.74}$$

eq:Qshift

and the second and third terms are in fact integral, so that if we may define

$$q : D(\Lambda) \to \mathbb{Q}/\mathbb{Z} \tag{25.75}$$

by

$$q(\bar{x}) := Q(x) \mathrm{mod} \mathbb{Z} \tag{25.76}$$

where x is any lift of $\bar{x}$ to Λ. Thanks to (25.74), $q(\bar{x})$ is well-defined.

This is an example of a quadratic function on a finite group. It satisfies

$$q(\bar{x}+\bar{y}) - q(\bar{x}) - q(\bar{y}) + q(\bar{0}) = b(\bar{x},\bar{y}) \tag{25.77}$$

Now, the Gauss-Milgram sum formula states that if $\mathcal{D} = D(\Lambda)$ then

$$\boxed{\sum_{\bar{x}\in\mathcal{D}} e^{2\pi i q(\bar{x})} = \sqrt{|\mathcal{D}|} e^{2\pi i \left(\frac{\sigma(\Lambda)-\mu(\Lambda)}{8}\right)}} \tag{25.78}$$

Proof:

Let us begin with the one-dimensional case $\Lambda = \nu\mathbb{Z}$ with $\nu^2 = n$ is a positive integer. Then, as we have seen $\mathcal{D} = \mathbb{Z}/n\mathbb{Z}$, and

$$b(\bar{x},\bar{y}) = \frac{\bar{x}\bar{y}}{n} \mathrm{mod}\ 1 \tag{25.79}$$

Moreover, we can take

♣Fix normalization? ♣

$$w = \begin{cases} 0 & n \quad \text{even} \\ 1 & n \quad \text{odd} \end{cases} \tag{25.80}$$

eq:charctr

So we have $Q(x) = \frac{x(x-w)}{2n}$, Let $q(x) = Q(x) \mathrm{mod} 1$.

Now we would like to evaluate:

$$S_n = \sum_{\mathcal{D}} e^{2\pi i q(x)} \tag{25.81}$$

Evaluation: Let $g(t) := \sum_{x=0}^{n-1} e^{2\pi i Q(x+t)}$. Note that $g(t+1) = g(t)$ so that

$$g(t) = \sum_{-\infty}^{+\infty} c_k e^{-2\pi i k t} \tag{25.82}$$

We want $g(0) = \sum c_k$. Write

$$c_k = \int_0^1 g(t) e^{2\pi i k t} dt = \int_0^n e^{2\pi i (Q(t)+kt)} dt \tag{25.83}$$

So now write

$$g(0) = \sum_{k\in\mathbb{Z}} c_k = \sum_{k=-\infty}^{\infty} \int_0^n e^{2\pi i Q(t)} e^{2\pi i k t} dt \tag{25.84}$$

But

$$Q(t+kn) = Q(t) + Q(kn) + tk \tag{25.85}$$

and $Q(kn)$ is an integer. Therefore,

$$\sum_{k=-\infty}^{+\infty} \int_0^n e^{2\pi i Q(t+kn)} dt = \int_{-\infty}^{+\infty} e^{2\pi i Q(t)} dt = \sqrt{\frac{\pi}{-i\pi/n}} e^{2\pi i \frac{w^2}{8n}} \tag{25.86}$$

so

$$\sum_{\mathcal{D}} e^{2\pi i q(x)} = \sqrt{n} = \sqrt{n} \exp[2\pi i \left(\frac{1}{8} - \frac{\langle w, w \rangle}{8} \right)] \tag{25.87}$$

Now, for the opposite signature we just take the complex conjugate.

Finally, to go to the general case note that we could have run a very similar argument by considering

$$g(t) = \sum_{\bar{x} \in \Lambda^*/\Lambda} e^{2\pi i Q(x+t)} \tag{25.88}$$

The Fourier analysis is very similar. Once we get to the Gaussian integral we can diagonalize it over $\mathbb{R}$ and then factorize the result into the one-dimensional case. ♠

♣Spell out the details some more here! ♣

Remarks

1. Note that it follows that for self-dual lattices $\mu(\Lambda) = \sigma(\Lambda) \mathrm{mod} 8$.

2. *Quadratic functions on finite abelian groups and quadratic refinements*

   ♣ Explain the general problem of finding a quadratic refinement of a bilinear form on a finite abelian group. ♣

3. *Gauss sums in general*

25.4 Self-dual lattices

Definition: An integral lattice is *self-dual*, or *unimodular* if $\Lambda = \Lambda^*$. Equivalently, Λ is unimodular if the determinant of the integral Gram matrix is $\det G_{ij} = \pm 1$.

Example 1: The Narain lattices generated by (25.28) above satisfy $\Lambda(R)^* = \Lambda(R)$ and are unimodular for all R.

Example 2: Of course, if Λ_1 and Λ_2 are unimodular then so is $\Lambda_1 \oplus \Lambda_2$. So $H(1) \oplus \cdots H(1)$ with d factors is an even unimodular lattice of signature (d, d). Similarly, $I^{t,s} \cong \mathbb{Z}^d$ with quadratic form:

$$Diag\{(-1)^t, (+1)^s\} \tag{25.89}$$

on $\mathbb{Z}^d$, $d = t + s$, is an odd unimodular lattice.

Example 3: *Positive definite even unimodular.* There is a class of very interesting positive definite even unimodular lattices which are of rank $r = 8k$ for k a positive integer. Introduce the vector

$$s = (\frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{Q}^{8k} \tag{25.90}$$

$$\begin{aligned}\Gamma_{8k} := &\Big\{(x_1,\cdots x_{8k}) \ \in \mathbb{Z}^{8k}| \qquad \sum x_i = 0(2)\Big\} \\ &\cup \Big\{(x_1,\cdots,x_8) \ \in \mathbb{Z}^{8k}+s| \qquad \sum x_i = 0(2)\Big\}\end{aligned} \tag{25.91}$$

Let us check this lattice is even unimodular:

a.) Integral: The only nonobvious part is whether the product of two vectors from $\mathbb{Z}^{8k}+s$ is integral. Write these as $x_i = n_i+\frac{1}{2}, y_i = m_i+\frac{1}{2}$ where $n_i, m_i \in \mathbb{Z}$ and $\sum n_i = 0(2)$ and $\sum m_i = 0(2)$. Then

$$\sum(n_i+\frac{1}{2})(m_i+\frac{1}{2}) = \sum n_i m_i + \frac{1}{2}(\sum n_i + m_i) + 2k \in \mathbb{Z} \tag{25.92}$$

b.) Even: Use $n_i^2 = n_i(2)$ for n_i integral.

c.) Self-dual: Suppose $(v_1,\ldots,v_{8k}) \in \Gamma_{8k}^*$. Then, $v_i \pm v_j \in \mathbb{Z}$ and therefore $2v_i \in \mathbb{Z}$ and moreover, the v_i are either all integral or all half-integral. Now,

$$s\cdot v = \frac{1}{2}\sum_i v_i \in \mathbb{Z} \tag{25.93}$$

implies $\sum v_i = 0(2)$, hence $v \in \Gamma_{8k}$, and hence $\Gamma_{8k}^* \subset \Gamma_{8k}$ implies it is unimodular.

The case $k=1$ defines what is known as the E_8-lattice, which is of particular interest in group theory and some areas of physics. Here is a particular lattice basis:

$$\begin{aligned}\alpha_1 &= \frac{1}{2}(e_1+e_8) - \frac{1}{2}(e_2+e_3+e_4+e_5+e_6+e_7) \\ \alpha_2 &= e_1+e_2 \\ \alpha_3 &= e_2-e_1 \\ \alpha_4 &= e_3-e_2 \\ \alpha_5 &= e_4-e_3 \\ \alpha_6 &= e_5-e_4 \\ \alpha_7 &= e_6-e_5 \\ \alpha_8 &= e_7-e_6\end{aligned} \tag{25.94}$$

The form $\alpha_i\cdot\alpha_j$ is the famous E_8 matrix:

$$\begin{bmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix} \tag{25.95}$$

One can check that this is in fact of determinant 1. This data is often encoded in a *Dynkin diagram* shown in Figure 20. A dot corresponds to a basis vector. Two dots are connected by a single line if $\alpha_i \cdot \alpha_j = -1$ (i.e. if the angle between them is $2\pi/3$). One gets

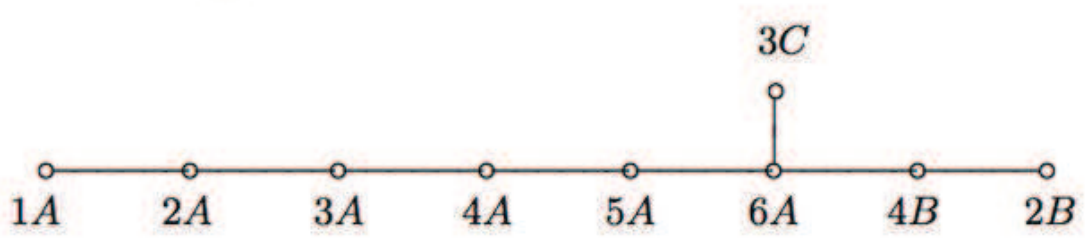


Figure 20: Dynkin diagram of the E_8 lattice. The numbers attached to the nodes have some interesting magical problems which will be discussed later.

fig:E8-Dynkin-

Remarks

1. The automorphism group of the E_8 lattice is an extremely intricate object. It is known as the Weyl group of E_8 and it is generated by the reflections in the hyperplanes orthogonal to the simple roots α_i listed above. There is an obvious subgroup isomorphic to

$$(\mathbb{Z}_2)^7 \ltimes S_8 \tag{25.96}$$

where the S_8 acts by permuting the coordinates and $(\mathbb{Z}_2)^7 \cong (\mathbb{Z}_2)^8/\mathbb{Z}_2$ is the group of sign-flips $x_i \to \epsilon_i x_i$ where an *even* number of signs are flipped. What is not obvious

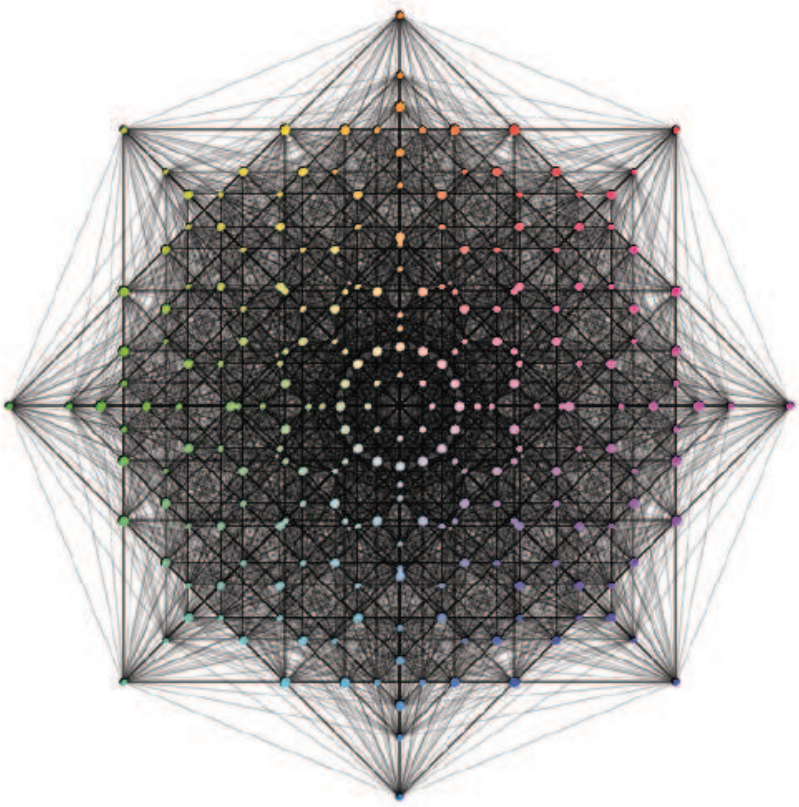


Figure 21: A projection of the 240 roots of the E8 root lattice in a two-dimensional plane. Copied from http://www.madore.org/ david/math/e8w.html. fig:ProjectedE

is that this group is only a subgroup and in fact the full Weyl group has order

$$\begin{aligned} |W(E_8)| &= 2^7 8! \times 135 \\ &= 8! \times (1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 4 \cdot 2 \cdot 3) \\ &= 2^{14} \times 3^5 \times 5^2 \times 7 \\ &= 696729600 \end{aligned} \tag{25.97}$$

2. ♣ SAY SOMETHING ABOUT VECTORS OF SQUARELENGTH TWO ♣

25.4.1 Some classification results

There are some interesting results on the classification of unimodular lattices. We now briefly review some of the most important ones.

The nature of the classification of lattices depends very strongly on the signature and rank of the form. For example, the classification of definite integral forms is a very difficult problem which remains unsolved in general.

By contrast, the classification is much simpler for indefinite signature: (i.e. $t > 0, s > 0$).

♣ EXPLAIN THE PROOF IN SERRE'S BOOK. THIS IS A BEAUTIFUL AND SIMPLE APPLICATION OF GENERAL IDEAS OF K-THEORY ♣

1. Odd unimodular lattices are unique. By change of basis we always get the lattice:

$$\Gamma \approx I^{t,s}. \tag{25.98}$$

2. Even unimodular lattices only exist for $(t-s) = 0 mod 8$ and are again unique for s, t both nonzero. They are denoted:

$$\Gamma \approx II^{t,s} \tag{25.99}$$

An explicit consturction of $II^{t,s}$ may be given by taking the lattice of d-tuples of points $(x_1, \ldots, x_d) \in \mathbb{R}^{t,s}$, with $d = t + s$, where the x_i are either all integral, or all half-integral, and in either case $\sum x_i = 0 \mathrm{mod} 2$.

Although the indefinite even unimodular lattices are unique, their *embedding* into $\mathbb{R}^{p,q}$ is highly nonunique. We have already seen this in example 3 above.

Note that $\Lambda(R) = \Lambda(1/R)$. The inequivalent embeddings of $II^{1,1}$ into $\mathbb{R}^{1,1}$ are parametrized by $R \geq 1$.

There are some partial results on positive definite even unimodular lattices.

1. In fact, they only exist for

$$\dim \Lambda = 0 \mathrm{mod} 8 \tag{25.100}$$

2. In any dimension there is a *finite* number of inequivalent lattices. In fact, we can count them! That is, we can count them using the Smith-Minkowski-Siegel "mass formula" which gives a formula for

$$\sum_{[\Lambda]} \frac{1}{|\mathrm{Aut}(\Lambda)|} = \frac{|B_{n/2}|}{n} \prod_{1 \leq j < n/2} \frac{|B_{2j}|}{4j} \tag{25.101}$$

The sum on the left is over inequivalent even unimodular positive definite lattice of dimension n. The B_{2j} are the Bernoulli numbers defined by

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} \pm \cdots \tag{25.102}$$

The growth of the Bernoulli number is given by Euler's result:

$$B_{2n} = (-1)^{n+1} \frac{2(2n)!}{(2\pi)^{2n}} \zeta(2n) \tag{25.103}$$

and hence the product on the RHS grows very fast. (Note that $\zeta(2n)$ is exponentially close to 1.

♣Say how RHS grows with the rank r. ♣

♣ Reference: A. Eskin, Z. Rudnik, and P. Sarnak, "A Proof of Siegel's Weight Formula," (There should be a simple topological field theory proof.) ♣

In higher dimensions there can be many inequivalent even integral unimodular lattices. The number $n(d)$ of such lattices is known to be:

$$\begin{aligned} n(8) &= 1 \\ n(16) &= 2 \\ n(24) &= 24 \\ n(32) &> 80 \times 10^6 \end{aligned} \tag{25.104}$$

eq:inequiv

Indeed, if we compute the RHS of the SMS formula for $n = 8$ then we get

$$\frac{B_4}{4} \times \frac{B_2}{4} \times \frac{B_4}{8} \times \frac{B_6}{12} = \left(\frac{1}{120}\right) \times \left(\frac{1}{24}\right) \times \left(\frac{1}{240}\right) \times \left(\frac{1}{504}\right) = \frac{1}{696729600} \tag{25.105}$$

This is exactly one over the order of the automorphism group of E_8, confirming $n(d) = 1$.

♣Surely there is a simpler proof of uniqueness... ♣

Of the 24 even unimodular lattices in dimension 24 one stands out, it is the *Leech lattice*, which is the unique lattice whose minimal length-square is 4.

There are many constructions of the Leech lattice, but one curious one is that we consider the light-like vector:

$$w = (70; 24, 23, \ldots, 3, 2, 1, 0) \tag{25.106}$$

in $II^{1,25} \subset \mathbb{R}^{1,25}$ and consider the lattice $w^\perp / w\mathbb{Z}$. This is a positive definite even integral lattice of rank 24. Note that the vectors of length-squared two are not orthogonal to w.

♣Need to explain why this construction gives a self-dual lattice. ♣

Remarks

1. Topology of 4-manifolds.
2. Abelian Chern-Simons theory.

Exercise
Compute the number of vectors of square-length = 2 in $\Lambda_R(E_8)$.

Exercise For any integer k construct an even unimodular lattice whose minimum length vector is 2^k.

Exercise *Narain lattices*
Let V be a vector space and $V^\vee$ be the dual space.

Using only the dual pairing one defines a natural signature (d,d) nondegenerate metric on $V \oplus V^\vee$:

$$\langle (x,\ell),(x',\ell')\rangle := \ell'(x) - \ell(x') \tag{25.107}$$

a.) Show that if $\Lambda \subset V$ is a lattice then

$$\Lambda_N := \{(p+w, p-w) | p \in \Lambda, w \in \Lambda^*\} \tag{25.108}$$

is an even unimodular lattice.

b.) Show that the space of inequivalent embedded lattices isomorphic to $II^{d,d}$ is $O(d,d;\mathbb{Z})\backslash O(d,d;\mathbb{R})$

♣This is out of place. We haven't done quotients yet. ♣

Exercise

Show that the lattice of $(x_1,\dots,x_d) \in \mathbb{R}^{t,s}$, with x_i all integral or all half-integral and $\sum x_i = 0(2)$ is an even self-dual lattice.

Hint: Use the same procedure as in the E_8 case above.

Exercise

Show that if $s > t$ then $II^{s,t}$ must be isomorphic as a lattice to a lattice of the form

$$II^{1,1} \oplus \cdots \oplus II^{1,1} \oplus E_8 \oplus \cdots \oplus E_8 \tag{25.109}$$

while if $t > s$ it is of the form

$$II^{1,1} \oplus \cdots \oplus II^{1,1} \oplus E_8(-1) \oplus \cdots \oplus E_8(-1) \tag{25.110}$$

of signature $((+1)^l, (-1)^{l+8m})$.

Exercise *Lattice Theta functions*

If Λ is a positive definite lattice one can associate to it the *Theta function*

$$\Theta_\Lambda := \sum_{v\in\Lambda} q^{\frac{1}{2}(v,v)} \tag{25.111}$$

This is a series in $q^{1/2}$ and it converges absolutely for $|q^{1/2}| < 1$. This function counts the number of lattice vectors of a given length.

For τ a complex number of positive imaginary part define $q := e^{2\pi i\tau}$. Using the Poisson summation formula show that the theta functions for Λ and Λ^* are related by:

$$\Theta_\Lambda(-1/\tau) = (-i\tau)^{\dim\Lambda/2}\frac{1}{\sqrt{|D(\Lambda)|}}\Theta_{\Lambda^*}(\tau) \tag{25.112}$$

♣♣ EXPLAIN RELATION OF FINITE HEISENBERG GROUPS AND THETA FUNCTIONS ♣♣

25.5 Embeddings of lattices: The Nikulin theorem

♣ Explain the terminology "glue group" ♣

25.6 References

Reference: For much more about lattices, see

J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices, and Groups.*

A beautiful and concise treatment of some of the material above can be found in:

J.-P. Serre, *A Course on Arithmetic*

26. Positive definite Quadratic forms

Criteria for A to be positive definite:

$A > 0$ iff the determinants of all minors is positive.

If A is a positive definite matrix with integer entries then it satisfies some remarkable properties:

1. Kronecker's theorem: $\| A \| \geq 2$ or $\| A \| = 2\cos(\pi/q)$, $q \geq 3$.

2. Perron-Frobenius theorem: A has a maximal positive eigenvalue and the eigenvector can be taken to have all positive entries. (Actually, the PF theorem is far more general.)

See V. Jones, et. al. *Coxeter graphs...* and Gantmacher, for a discussion.

Nice Application: GOOGLE search algorithm PageRank

27. Quivers and their representations

Nice application of linear algebra.

Preprint typeset in JHEP style - HYPER VERSION

Chapter 3: Transformations Groups, Orbits, And Spaces Of Orbits

Gregory W. Moore

Abstract: This chapter focuses on of group actions on spaces, group orbits, and spaces of orbits. Then we discuss mathematical symmetric objects of various kinds. May 3, 2019

-TOC- **Contents**

1. Introduction **2**

2. Definitions and the stabilizer-orbit theorem **2**
2.0.1 The stabilizer-orbit theorem 6
2.1 First examples 7
2.1.1 The Case Of $1+1$ Dimensions 11

3. Action of a topological group on a topological space **14**
3.1 Left and right group actions of G on itself 19

4. Spaces of orbits **20**
4.1 Simple examples 21
4.2 Fundamental domains 22
4.3 Algebras and double cosets 28
4.4 Orbifolds 28
4.5 Examples of quotients which are not manifolds 29
4.6 When is the quotient of a manifold by an equivalence relation another manifold? 33

5. Isometry groups **34**

6. Symmetries of regular objects **36**
6.1 Symmetries of polygons in the plane 39
6.2 Symmetry groups of some regular solids in $\mathbb{R}^3$ 42
6.3 The symmetry group of a baseball 43

7. The symmetries of the platonic solids **44**
7.1 The cube ("hexahedron") and octahedron 45
7.2 Tetrahedron 47
7.3 The icosahedron 48
7.4 No more regular polyhedra 50
7.5 Remarks on the platonic solids 50
7.5.1 Mathematics 51
7.5.2 History of Physics 51
7.5.3 Molecular physics 51
7.5.4 Condensed Matter Physics 52
7.5.5 Mathematical Physics 52
7.5.6 Biology 52
7.5.7 Human culture: Architecture, art, music and sports 53
7.6 Regular polytopes in higher dimensions 53

8. Classification of the Discrete subgroups of $SO(3)$ and $O(3)$ **53**
8.1 Finite subgroups of $O(3)$ 57
8.2 Finite subgroups of $SU(2)$ and $SL(2,\mathbb{C})$ 60

9. Symmetries of lattices **60**
9.1 The crystallographic restriction theorem 60
9.2 Three-dimensional lattices 61
9.3 Quasicrystals 64

10. Tesselations by Triangles **66**

11. Diversion: The Rubik's cube group **66**

12. Group actions on function spaces **67**

13. The simple singularities in two dimensions **69**

14. Symmetric functions **71**
14.1 Structure of the ring of symmetric functions 71
14.2 Free fermions on a circle and Schur functions 79
14.2.1 Schur functions, characters, and Schur-Weyl duality 85
14.3 Bosons and Fermions in 1+1 dimensions 86
14.3.1 Bosonization 86

---

1. Introduction

2. Definitions and the stabilizer-orbit theorem

♣LEFT G-ACTIONS WERE INTRODUCED AT THE BEGINNING OF CHAPTER 1 ♣

Let X be any set (possibly infinite). Recall the definition from chapter 1:

Definition 1: A *permutation* of X is a 1-1 and onto mapping $X \to X$. The set S_X of all permutations forms a group under composition.

Definition 2a: A *transformation group* on X is a subgroup of S_X.

This is an important notion so let's put it another way:

Definition 2b:

A *G-action on a set X* is a map $\phi : G \times X \to X$ compatible with the group multiplication law as follows:

A *left-action* satisfies:

$$\phi(g_1, \phi(g_2, x)) = \phi(g_1 g_2, x) \tag{2.1}$$

A *right-action* satisfies

$$\phi(g_1, \phi(g_2, x)) = \phi(g_2 g_1, x) \tag{2.2}$$

In addition in both cases we require that

$$\phi(1_G, x) = x \tag{2.3}$$ eq:identfix

for all $x \in X$.

A set X equipped with a (left or right) G-action is said to be a G-set.

Remarks:

1. If ϕ is a left-action then it is natural to write $g \cdot x$ for $\phi(g, x)$. In that case we have

$$g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x. \tag{2.4}$$

   Similarly, if ϕ is a right-action then it is better to use the notation $\phi(g, x) = x \cdot g$ so that

$$(x \cdot g_2) \cdot g_1 = x \cdot (g_2 g_1). \tag{2.5}$$

2. If ϕ is a left-action then $\tilde{\phi}(g, x) := \phi(g^{-1}, x)$ is a right-action, and vice versa. Thus there is no essential difference between a left- and right-action. However, in computations with nonabelian groups it is extremely important to be consistent and careful about which choice one makes.

3. A given set X can admit more than one action by the same group G. If one is working simultaneously with several different G actions on the same set then the notation $g \cdot x$ is ambiguous and one should write, for example, $\phi_g(x) = \phi(g, x)$ or speak of ϕ_g, etc. A good example of a set X with several natural G actions is the case of $X = G$ itself. Then there are the actions of left-multiplication, right-multiplication, and conjugation:

$$\begin{aligned} L(g, g') &= gg' \\ R(g, g') &= g'g \\ C(g, g') &= g^{-1} g' g \end{aligned} \tag{2.6}$$

   where on the RHS of these equations we use group multiplication. Note that our choice of C makes it a right-action of G on G.

There is some important terminology one should master when working with G-actions. First here are some terms used when describing a G-action on a set X:

Definitions:

1. A group action is *effective* or *faithful* if for any $g \neq 1$ there is some x such that $g \cdot x \neq x$. Equivalently, the only $g \in G$ such that ϕ_g is the identity transformation is $g = 1_G$. A group action is *ineffective* if there is some $g \in G$ with $g \neq 1$ so that $g \cdot x = x$ for all $x \in X$. The set of $g \in G$ that act ineffectively is a normal subgroup of G.

2. A group action is *transitive* if for any pair $x, y \in X$ there is some g with $y = g \cdot x$.

3. A group action is *free* if for any $g \neq 1$ then for *every* x, we have $g \cdot x \neq x$.

In summary:

1. *Effective*: $\forall g \neq 1, \exists x$ s.t. $g \cdot x \neq x$.

2. *Ineffective*: $\exists g \neq 1$,s.t. $\forall x\ g \cdot x = x$.

3. *Transitive*: $\forall x, y \in X, \exists g$ s.t. $y = g \cdot x$.

4. *Free*: $\forall g \neq 1, \forall x,\ g \cdot x \neq x$

In addition there are some further important definitions:

1. Given a point $x \in X$ the set of group elements:

$$\mathrm{Stab}_G(x) := \{g \in G : g \cdot x = x\} \tag{2.7}$$

is called the *isotropy group at* x. It is also called the *stabilizer group* of x. It is often denoted G^x. The reader should show that $G^x \subset G$ is in fact a subgroup. Note that a group action is free iff for every $x \in X$ the stabilizer group G^x is the trivial subgroup $\{1_G\}$.

2. A point $x \in X$ is a *fixed point* of the G-action if there exists some element $g \in G$ with $g \neq 1$ such that $g \cdot x = x$. So, a point $x \in X$ is a fixed point of G iff $\mathrm{Stab}_G(x)$ is not the trivial group. Some caution is needed here because if an author says "x is a fixed point of G" the author might mean that $\mathrm{Stab}_G(x) = G$. That would not be implied by our terminology.

3. Given a group element $g \in G$ the *fixed point set* of g is the set

$$\mathrm{Fix}_X(g) := \{x \in X : g \cdot x = x\} \tag{2.8}$$

The fixed point set of g is often denoted by X^g. Note that if the group action is free then for every $g \neq 1$ the set $\mathrm{Fix}_X(g)$ is the empty set.

4. The *orbit of G through a point x* is the set of points $y \in X$ which can be reached by the action of G:

$$O_G(x) = \{y : \exists g \quad \text{such that} \quad y = g \cdot x\} \tag{2.9}$$

Remarks:

1. If we have a G-action on X then we can define an equivalence relation on X by defining $x \sim y$ if there is a $g \in G$ such that $y = g \cdot x$. (Check this is an equivalence relation!) The orbits of G are then exactly the equivalence classes of under this equivalence relation.

2. The group action restricts to a transitive group action on any orbit.

3. If x, y are in the same orbit then the isotropy groups G^x and G^y are conjugate subgroups in G. Therefore, to a given orbit, we can assign a definite *conjugacy class* of subgroups.

Point 3 above motivates the

Definition If G acts on X a *stratum* is a set of G-orbits such that the conjugacy class of the stabilizer groups is the same. The set of strata is sometimes denoted $X \parallel G$.

Exercise

Recall that a group action of G on X can be viewed as a homomorphism $\phi : G \to S_X$. Show that the action is effective iff the homomorphism is injective.

Exercise

Suppose X is a G-set.

a.) Show that the subset H of elements which act ineffectively, i.e. the set of $h \in G$ such that $\phi(h, x) = x$ for all $x \in X$ is a normal subgroup of G.

b.) Show that G/H acts effectively on X.

Exercise

Let G act on a set X.

a.) Show that the stabilizer group at x, denoted G^x above, is in fact, a subgroup of G.

b.) Show that the G action is free iff the stabilizer group at every $x \in X$ is the trivial subgroup $\{1_G\}$.

c.) Suppose that $y = g \cdot x$. Show that G^y and G^x are conjugate subgroups in G. [1]

Exercise

a.) Show that whenever G acts on a set X one can canonically define a groupoid: The objects are the points $x \in X$. The morphisms are pairs (g, x), to be thought of as arrows $x \xrightarrow{g} g \cdot x$. Thus, $X_0 = X$ and $X_1 = G \times X$.

[1] *Answer*: If $y = g_0 \cdot x$ and $g \cdot x = x$ then $(g_0 g g_0^{-1}) \cdot y = y$ so $G^y = g_0 G^x g_0^{-1}$.

b.) What is the automorphism group of an object $x \in X$.
This groupoid is commonly denoted as $X// G$.

2.0.1 The stabilizer-orbit theorem

There is a beautiful relation between orbits and isotropy groups:

Theorem [Stabilizer-Orbit Theorem]: Each left-coset of G^x in G is in 1-1 correspondence with the points in the G-orbit of x:

$$\psi : Orb_G(x) \to G/G^x \tag{2.10}$$

for a $1-1$ map ψ.

Proof: Suppose y is in a G-orbit of x. Then $\exists g$ such that $y = g \cdot x$. Define $\psi(y) \equiv g \cdot G^x$. You need to check that ψ is actually well-defined.

$$y = g' \cdot x \qquad \to \qquad \exists h \in G^x \qquad g' = g \cdot h \qquad \to \qquad g'G^x = ghG^x = gG^x \tag{2.11}$$

Conversely, given a coset $g \cdot G^x$ we may define

$$\psi^{-1}(gG^x) \equiv g \cdot x \tag{2.12}$$

Again, we must check that this is well-defined. Since it inverts ψ, ψ is 1-1. ♠

Corollary: If G acts transitively on a space X then there is a $1-1$ correspondence between X and the set of cosets of H in G where H is the isotropy group of any point $x \in X$. That is, at least as sets: $X = G/H$. The isotropy groups for points in G/H are the conjugate subgroups of H in G.

Remark: Sets of the type G/H are called *homogeneous spaces*. This theorem is the beginning of an important connection between the *algebraic* notions of subgroups and cosets to the *geometric* notions of orbits and fixed points. Below we will show that if G, H are topological groups then, in some cases, G/H are beautifully symmetric topological spaces, and if G, H are Lie groups then, in some cases, G/H are beautifully symmetric manifolds.

Exercise *The Lemma that is not Burnside's*
Suppose a finite group G acts on a finite set X as a transformation group. A common notation for the set of points fixed by g is X^g. Show that the number of distinct orbits is the averaged number of fixed points:

$$|\{orbits\}| = \frac{1}{|G|} \sum_g |X^g| \tag{2.13}$$

For the answer see. [2]

Exercise *Jordan's theorem*
Suppose G is finite and acts transitively on a finite set X with more than one point. Show that there is an element $g \in G$ with no fixed points on X. [3]

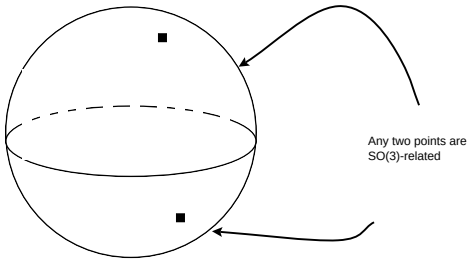


Figure 1: Transitive action of $SO(3,\mathbb{R})$ on the sphere. fig:orbii

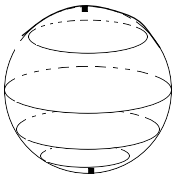


Figure 2: Orbits of $SO(2,\mathbb{R})$ on the two sphere. fig:orbiii

2.1 First examples

The concept of a G-action on a set is an extremely important concept, so let us consider a number of examples:

Examples

[2] *Answer*: Write

$$\sum_{g\in G} |X^g| = |\{(x,g)|g\cdot x = x\}| = \sum_{x\in X} |G^x| \tag{2.14}$$

Now use the stabilizer-orbit theorem to write $|G^x| = |G|/|\mathcal{O}_G(x)|$. Now in the sum

$$\sum_{x\in X} \frac{1}{|\mathcal{O}_G(x)|} \tag{2.15}$$

the contribution of each distinct orbit is exactly 1.

[3] Hint: Note that $X = G/H$ for some H and apply the Burnside lemma.

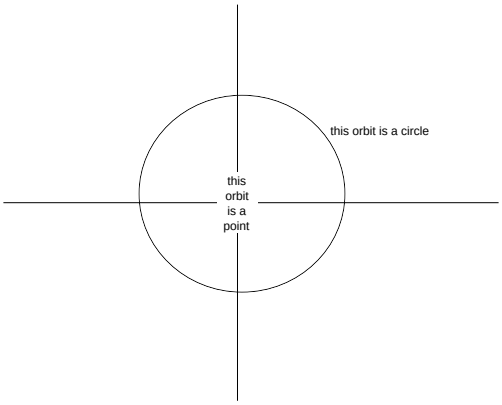


Figure 3: Notice not all orbits have the same dimensionality. There are two qualitatively different kinds of orbits of $SO(2,\mathbb{R})$. fig:sotwo

1. Let $X = \{1, \cdots n\}$, so $S_X = S_n$ as before. The action is effective and transitive, but not free. Indeed, the fixed point of any $j \in X$ is just the permutations that permute everything else, and hence $S_X^j \cong S_{n-1}$. Note that different j have different stabilizer subgroups isomorphic to S_{n-1}, but they are all conjugate.

2. *Group actions on the plane.* The group $G = GL(2,\mathbb{R})$ acts on the plane $X = \mathbb{R}^2$ by linear transformation. The action is effective: If $g \neq 1$ then it moves some vector a nonzero amount! There are only two orbits: Note that if $\vec{x} = 0$ then it remains zero under linear transformation, so $\{\vec{0}\}$ is an orbit. On the other hand, if $\vec{x}, \vec{y}$ are both nonzero then some linear transformation certainly will map $\vec{x}$ to $\vec{y}$. The action is therefore not transitive, and not free.

3. Similarly, $GL(n,\mathbb{R})$ acts on $\mathbb{R}^n$. If we act with a matrix on a column vector we get a left action. If we act on a row vector we get a right action. Either way, there are two orbits.

4. We can restrict the $GL(2,\mathbb{R})$ action on $\mathbb{R}^2$ to the action of the subgroup $G = SO(2,\mathbb{R})$. This completely changes the picture. The action is:

$$R(\phi) : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{2.16}$$

   The group action is effective. It is not free, and it is not transitive. There are now infinitely many orbits of $SO(2)$, and they are all distinguished by the invariant value of x^2+y^2 on the orbit. From the viewpoint of topology, there are two distinct "kinds" of orbits acting on $\mathbb{R}^2$. One has trivial isotropy group and one has isotropy group $SO(2)$. See Figure 3. These give two strata.

5. Orbits of $O(2)$. The two-dimensions orthogonal group $O(2,\mathbb{R})$ can be written as a semidirect product

$$O(2) = SO(2) \rtimes \mathbb{Z}_2 \tag{2.17}$$

where $\mathbb{Z}_2$ acts on $SO(2)$ by taking $R(\theta) \to R(-\theta)$. The group has two components which can be written as

$$O(2) = SO(2) \amalg P \cdot SO(2) \tag{2.18}$$

where P is not canonical and can be taken to be reflection in any line through the origin. The orbits of $SO(2)$ and $O(2)$ are the same.

6. Similarly, $SO(3,\mathbb{R})$ acts on $X = \mathbb{R}^3$. It is effective, not transitive, and not fixed-point-free. We can restrict the action to a sphere of any radius S^2_R. The action is then transitive on the sphere, The isotropy group of any point $x \in S^2_R$ is the subgroup of rotations about the axis through that point. That subgroup is isomorphic to $SO(2,\mathbb{R})$, but as x varies the particular subgroup varies. For example, with usual conventions, if x is on the x^3-axis then the subgroup is the subgroup of matrices of the form

$$\begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.19}$$

but if x is on the x^1-axis the subgroup is the subgroup of matrices of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} \tag{2.20}$$

and so on. There are two strata: Those with G^x congruent to $SO(2,\mathbb{R})$ and those with $G^x = SO(3,\mathbb{R})$.

7. By contrast consider a fixed $SO(2,\mathbb{R})$ subgroup of $SO(3,\mathbb{R})$, say, the subgroup defined by rotations around the z-axis. This subgroup also acts on the sphere - but <u>not</u> transitively. The G-orbits are shown in Figure 2.

8. If $G = \mathbb{Z}_2$ acts linearly on $\mathbb{R}^{n+1}$ (i.e. $V = \mathbb{R}^{n+1}$ is a representation of $\mathbb{Z}_2$) then we can choose coordinates so that the nontrivial element $\sigma \in G$ acts by

$$\sigma \cdot (x^1, \dots, x^{n+1}) = (x^1, \dots, x^p, -x^{p+1}, \cdots, -x^{p+q}) \tag{2.21}$$ eq:Z2-sphere

where $p + q = n + 1$. Note that this action preserves the equation of the sphere $\sum_i (x^i)^2 - 1 = 0$ and hence descends to a $\mathbb{Z}_2$-action on the sphere S^n. The case $p = 0, q = n + 1$ is the antipodal map, but there are many other natural actions of $\mathbb{Z}_2$ on S^n.

9. Let the group be $G = \mathbb{C}^*$. This acts on $X = \mathbb{C}^n$ by scaling all the coordinates. The set of orbits is not a nice manifold, but the set of orbits of the action on $\tilde{X} = \mathbb{C}^n - \{0\}$ is a good manifold (see below). It is called $\mathbb{CP}^{n-1}$. Consider a set of integers $(q_1, \dots, q_n) \in \mathbb{Z}^n$. Then for each such set of integers there is a $\mathbb{C}^*$-action on $\mathbb{CP}^{n-1}$ defined by

$$\mu \cdot [X^1 : \cdots : X^n] := [\mu^{q_1} X^1 : \cdots : \mu^{q_n} X^n] \tag{2.22}$$ eq:Cstar-CPn

for $\mu \in \mathbb{C}^*$. (Check it is well-defined!)

10. The group $G = SL(2,\mathbb{R})$ acts on the complex upper half plane:

$$\mathcal{H} = \{\tau | \mathrm{Im}\tau > 0\} \tag{2.23}$$

via

$$g \cdot \tau := \frac{a\tau + b}{c\tau + d} \tag{2.24}$$ eq:FracLin

where

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{2.25}$$

11. *Actions of* $\mathbb{Z}$. Let us consider $\mathbb{Z}$ to be the free group with one generator g_0. Then, given any invertible map $f : X \to X$ we can define a group action of $\mathbb{Z}$ on X by

$$g_0^n \cdot x = \begin{cases} \underbrace{f \circ \cdots \circ f}_{n \quad \text{times}}(x) & n > 0 \\ x & n = 0 \\ \underbrace{f^{-1} \circ \cdots \circ f^{-1}}_{|n| \quad \text{times}}(x) & n < 0 \end{cases} \tag{2.26}$$ eq:Z-action

Conversely, any $\mathbb{Z}$-action must be of this form since we can define $f(x) := g_0 \cdot x$.

12. Let G be any group and consider the group action defined by $\phi(g,x) = x$ for all $g \in G$. This is as ineffective as a group action can be: For every x, the istropy group is all of G, and for all $g \in G$, $\mathrm{Fix}(g) = X$. In particular, this situation will arise if X consists of a single point. This example is not quite as stupid as might at first appear, once one takes the categorical viewpoint, for $pt/\!/G$ is a very rich category indeed.

Exercise

Consider the action of $\mathbb{Z}_2$ on the sphere defined by (2.21).

a.) For which values of p, q is the action effective?

b.) For which values of p, q is the action transitive?

c.) Compute the fixed point set of the nontrivial element $\sigma \in \mathbb{Z}_2$.

d.) For which values of p, q is the action free?

Exercise

Consider the action of $G = \mathbb{C}^*$ on $\mathbb{CP}^{n-1}$ defined by (2.22).

a.) For which values of $(q_1, \ldots, q_n)$ is the action effective?

b.) For which values of $(q_1, \ldots, q_n)$ is the action transitive?

c.) What are the fixed points of the $\mathbb{C}^*$ action?

d.) What are the stabilizers at the fixed points of the $\mathbb{C}^*$ action?

Exercise

a.) Show that (2.24) above defines a left-action of $SL(2,\mathbb{R})$ on the complex upper half-plane. [4]

b.) Is the action effective?

c.) Is the action transitive?

d.) Which group elements have fixed points?

e.) What is the isotropy group of $\tau = \mathrm{i}$?

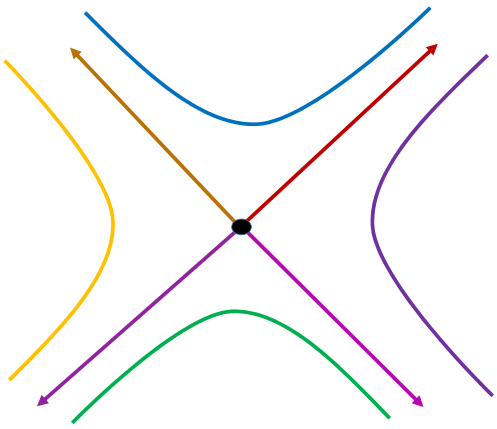


Figure 4: The distinct kinds of orbits of $SO(1,1,\mathbb{R})$ are shown in different colors. If we enlarge the group to include transformations that reverse the orientation of time and/or space then orbits of the larger group will be made out of these orbits by reflection in the space or time axis. fig:LorentzOrb

2.1.1 The Case Of $1+1$ Dimensions

Consider $1+1$-dimensional Minkowski space with coordinates $x = (x^0, x^1)$ and metric given by

$$\eta := \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.27}$$

[4] *Hint*: Show that $\mathrm{Im}(g\cdot\tau) = \frac{\mathrm{Im}\tau}{|c\tau+d|^2}$.

i.e. the quadratic form is $(x,x) = -(x^0)^2 + (x^1)^2$. The two-dimensional Lorentz group is defined by

$$O(1,1) = \{A | A^{tr}\eta A = \eta\} \tag{2.28}$$

This group acts on $\mathbb{M}^{1,1}$ preserving the Minkowski metric.

The connected component of the identity is the group of Lorentz boosts of rapidity θ:

$$x^0 \to \cosh\theta\ x^0 + \sinh\theta\ x^1 \tag{2.29}$$

$$x^1 \to \sinh\theta\ x^0 + \cosh\theta\ x^1 \tag{2.30}$$

that is:

$$SO_0(1,1;\mathbb{R}) \equiv \{B(\theta) = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} | -\infty < \theta < \infty\} \tag{2.31}$$

In the notation the S indicates we look at the determinant one subgroup and the subscript 0 means we look at the connected component of 1. This is a group since

$$B(\theta_1)B(\theta_2) = B(\theta_1 + \theta_2) \tag{2.32}$$

so $SO_0(1,1) \cong \mathbb{R}$ as groups. Indeed, note that

$$B(\theta) = \exp\left[\theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \tag{2.33}$$

It is often useful to define *light cone coordinates*: [5]

$$x^{\pm} := x^0 \pm x^1 \tag{2.34}$$

and the group action in these coordinates is simply:

$$x^{\pm} \to e^{\pm\theta}x^{\pm} \tag{2.35}$$

so it is obvious that $x^+x^- = -(x,x)$ is invariant.

It follows that the orbits of the Lorentz group are, in general, hyperbolas. They are separated by different values of the Lorentz invariant $x^+x^- = \lambda$, but this is not a complete invariant, since the sign (or vanishing) of x^+ and of x^- is also Lorentz invariant. For a real number r define

$$\text{sign}(r) := \begin{cases} +1 & r > 0 \\ 0 & r = 0 \\ -1 & r < 0 \end{cases} \tag{2.36}$$

Then $(\lambda, \text{sign}(x^+), \text{sign}(x^-))$ is a complete invariant of the orbits. That is, given this triple of data there is a unique orbit with these properties.

It is now easy to see what the different type of orbits are. They are shown in Figure 4: They are:

[5] Some authors will define these with a $1/2$ or $1/\sqrt{2}$. One should exercise care with this choice of convention.

♣Actually, the lightrays and hyperbolas have trivial stabilizer and hence are in the same strata. This is a problem with using strata. ♣

1. hyperbolas in the forward/backward lightcone and the left/right of the lightcone

2. 4 disjoint lightrays.

3. the origin: $x^+ = x^- = 0$.

It is now interesting to consider the orbits of the full Lorentz group $O(1,1)$ and its relation to the massless wave equations. This group has four components and is, group-theoretically

$$O(1,1) = SO_0(1,1) \rtimes (\mathbb{Z}_2 \times \mathbb{Z}_2) \tag{2.37}$$

We can write, noncanonically,

$$O(1,1) = SO_0(1,1) \amalg P \cdot SO_0(1,1) \amalg T \cdot SO_0(1,1) \amalg PT \cdot SO_0(1,1) \tag{2.38}$$

with

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.39}$$

The P and T operations map various orbits of $SO_0(1,1)$ into each other: P is a reflection in the time axis and T is a reflection in the space axis. Thus the orbits of the groups $SO(1,1)$, $SO_0(1,1) \amalg PT \cdot SO_0(1,1)$, and $O(1,1)$ all differ slightly from each other.

♣Should give more details here, or form an exercise. ♣

As an example of a physical manifestation of orbits let us consider the energy-momentum dispersion relation of a particle of mass m with energy-momentum $(E,p) \in \mathbb{R}^{1,1}$.

1. Massive particles: $m^2 > 0$ have (E,p) along an orbit in the upper quadrant:

$$\mathcal{O}^+(m) = \{(m\cosh\theta, m\sinh\theta) | \theta \in \mathbb{R}\} \tag{2.40}$$

2. Massless particles move at the speed of light. In 1+1 dimensions there is an interesting distinction: Left-moving particles with positive energy have support on [6] $p_+ = \frac{1}{2}(E+p) = 0$ and $p_- = \frac{1}{2}(E-p) \neq 0$. Right-moving particles with positive energy have support on $p^- = 0$ and $p^+ \neq 0$.

3. Tachyons have $E^2 - p^2 = m^2 < 0$ and have their support on the left or right quadrant. If we try to expand a solution to the wave-equation with $e^{\mathrm{i}(k_0 x^0 + k_1 x^1)}$ then $k_0^2 = \sqrt{k_1^2 + m^2}$ and so if the spatial momentum k_1 is sufficiently small then k_0 is pure imaginary and the wave grows exponentially, signaling and instability. This tells us our theory is out of control and some important new physical input is needed.

4. A massless "particle" of zero energy and momentum.

[6] Note the factors of two, so that $x^0 p_0 + x^1 p_1 = x^+ p_+ x^- p_-$. This is an example of the tricky factors of two one encounters when working with light-cone coordinates.

3. Action of a topological group on a topological space

If the group G is a topological group it is said to act continuously on a topological space X when $\phi : G \times X \to X$ is a continuous map. When working with topological groups acting on topological spaces, this is generally assumed. Note that $\phi_g : X \to X$ has a continuous inverse, namely $\phi_{g^{-1}}$ and therefore ϕ_g is a homeomorphism. Therefore, a topological group action on a topological space can be defined to be a homomorphism from the group G to the group of homeomorphisms on X. [7] One often wants to work with *proper* actions: This means that the ϕ is a proper continuous map.

Warning: There is some very important, but regrettably very confusing terminology associated with topological group actions on topological spaces. When G is a discrete group there is an important notion of a *properly discontinuous* action. The general definition is that the map $G \times X \to X \times X$ given by $(g, x) \mapsto (g \cdot x, x)$ is a proper continuous map. This does not mean the function taking $x \mapsto g \cdot x$ is discontinuous! On the contrary, as stated above, it is continuous. To make matters worse, one will find inequivalent definitions of the term "properly discontinuous" in textbooks and on the internet. The subtleties melt away when G is finite or when X is locally compact. We will follow the definitions used by W.P. Thurston, since he was one of the great masters of the subject. Specifically, Definition 3.5.1 of W.P. Thurston, *Three Dimensional Geometry and Topology*, vol 1, Princeton University Press 1997 includes:

Definition: Let G be a discrete group acting continuously on a topological space X. Then

1. The action *has discrete orbits* if every $x \in X$ has a neighborhood U such that the set of group elements $g \in G$ with $g \cdot x \in U$ is finite.
2. The action *is wandering* if every $x \in X$ has a neighborhood U such that the set of group elements $g \in G$ with $g \cdot U \cap U \neq \emptyset$ is finite.
3. If X is locally compact then the action is said to be *properly discontinuous* if for every compact set $K \subset X$ the set of g with $g \cdot K \cap K \neq \emptyset$ is finite.

If $\mathcal{M}$ is a measure space then a *discrete dynamical system* is a pair of $\mathcal{M}$ together with a measure-preserving group action of $\mathbb{Z}$ on $\mathcal{M}$. According to (2.26) this means we have a map $f : \mathcal{M} \to \mathcal{M}$ such that $\mu(f(\mathcal{A})) = \mu(\mathcal{A})$ for any measurable set $\mathcal{A}$. For example, in Hamiltonian dynamics one can take $\mathcal{M}$ to be phase space equipped with a symplectic form ω The natural measure is then the Liouville measure:

$$\mu(\mathcal{A}) := \int_{\mathcal{A}} \frac{\omega^n}{n!} \tag{3.1}$$

In particular, if f is symplectic, i.e. $f^*(\omega) = \omega$, or, in equations:

$$\omega_{\mu\nu}(f(x)) \frac{\partial f^\mu}{\partial x^\lambda} \frac{\partial f^\nu}{\partial x^\rho} = \omega_{\lambda\rho}(x) \tag{3.2}$$

[7] Continuous, with a suitable topology on the group of homeomorphisms of X.

then the corresponding $\mathbb{Z}$-action is a dynamical system.

One important result is the

Theorem [Poincaré recurrence theorem]. If $f : \mathcal{M} \to \mathcal{M}$ is a measure-preserving map and has bounded orbits then in any open U set there are points x such that for infinitely many n, $g^n x \in U$.

The proof is based on the idea that if this were not true then the volume of $\cup f^n(U)$ would be infinite, but that cannot be for a volume preserving map with bounded orbits. ♣Should give a better proof! ♣

There are various ways of expressing how "chaotic" a map is. A dynamical system is said to be *mixing* if for all pairs of (measurable) sets $\mathcal{A}, \mathcal{B} \subset \mathcal{M}$ we have

$$\lim_{n\to\infty} \mu(\mathcal{A}_n \cap \mathcal{B}) = \frac{\mu(\mathcal{A})\mu(\mathcal{B})}{\mu(\mathcal{M})} \tag{3.3}$$

where $\mathcal{A}_n = f^n(\mathcal{A})$. If $\mu(\mathcal{B}) \neq 0$ this means

$$\lim_{n\to\infty} \frac{\mu(\mathcal{A}_n \cap \mathcal{B})}{\mu(\mathcal{B})} = \frac{\mu(\mathcal{A})}{\mu(\mathcal{M})} \tag{3.4}$$

so that the "weight" of the set $\mathcal{A}$ is equally distributed over any $\mathcal{B}$. We will give an example of a dynamical system which is mixing below.

Let us examine some special cases of discrete dynamical systems:

1. First, take $X = \mathbb{R}^2$ and let

$$f(x_1, x_2) = (\lambda x_1, \lambda^{-1} x_2) \tag{3.5}$$ eq:lambda-act

where λ is a positive real number greater than 1. For $n > 0$, g_0^n stretches in the x_1 direction and flattens in the x_2 direction. For $n < 0$ the situation is reversed. This action clearly does *not* have discrete orbits, since the origin $(0, 0)$ is a fixed point for the entire group. This action can come up, for example in symplectic geometry: Note that this is a symplectic action with Poisson bracket $\{x_1, x_2\} = 1$, i.e. symplectic form $\omega = dx_1 \wedge dx_2$. It also comes up in some string theory models of cosmological singularities, where we view x_1, x_2 as light-cone coordinates in $1 + 1$-dimensional Minkowski space.

2. Next, consider $X = \mathbb{R}^2 - \{0, 0\}$. Now we can consider (3.5) to be a symplectic action with Poisson bracket for $\{x_1, x_2\} = 1$ just by restriction. However, having excised the point $(0, 0)$ we are now free to define the symplectic structure $\{x_1, x_2\} = x_1 x_2$, i.e.

$$\omega = \frac{dx_1}{x_1} \wedge \frac{dx_2}{x_2} \tag{3.6}$$

and since $x_1 x_2$ is preserved the action (3.5) is still symplectic. Now the action has discrete orbits and is in fact wandering: Consider a neighborhood of some point. We may assume the projection on the x_1 axis will be an open interval (a, b). Then

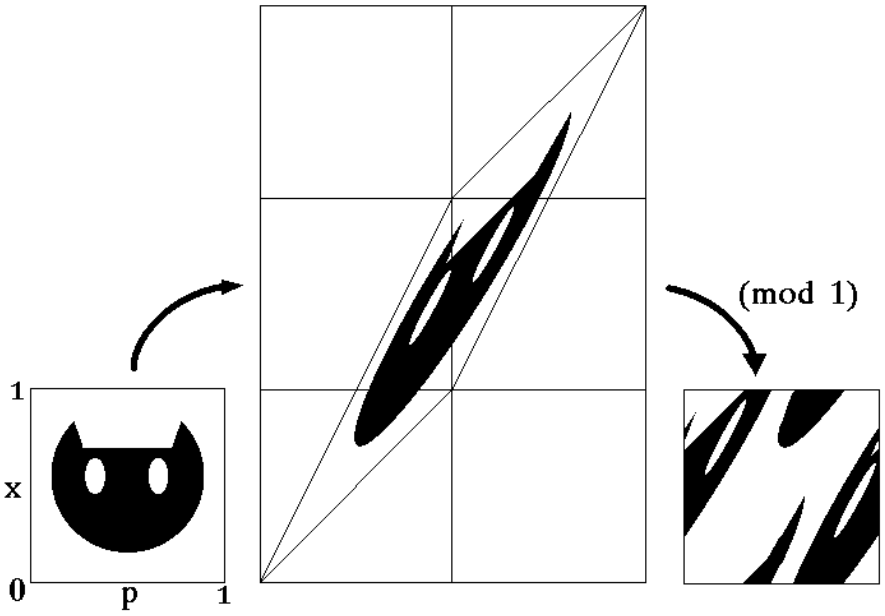


Figure 5: The famous Arnold cat map. The picture ultimately comes from the book V. Arnold and A. Avez, *Ergodic Problems in Classical Mechanics*. fig:ArnoldCatM

there are only a finite number of solutions to $a < \lambda^k b < b$ and only a finite number of solutions to $a < \lambda^k a < b$. On the other hand, the action is *not* properly discontinuous. To see this, consider the compact set K which is a closed line segment from $(1, 0)$ to $(0, 1)$. For any $n \neq 0$ there is a solution to

$$\begin{aligned} x + y &= 1 \\ \lambda^n x + \lambda^{-n} y &= 1 \end{aligned} \tag{3.7}$$

with $0 < x < 1$ and $0 < y < 1$. Therefore $g_0^n \cdot K \cap K \neq 0$ for all n.

3. *The Arnold cat map*: Consider the torus as $T^2 = [0, 1]^2/\sim$ with $x \sim x + 1$ and $y \sim y + 1$. Consider the transformation $f : T^2 \to T^2$ defined by

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3.8}$$

Note again that this is a symplectic transformation. This is known as *Arnold's cat map*, and is famous in dynamical systems theory and in discussions of chaos. Figure 5 shows one iteration of this map. It is shown in Arnold-Avez that the map is mixing.

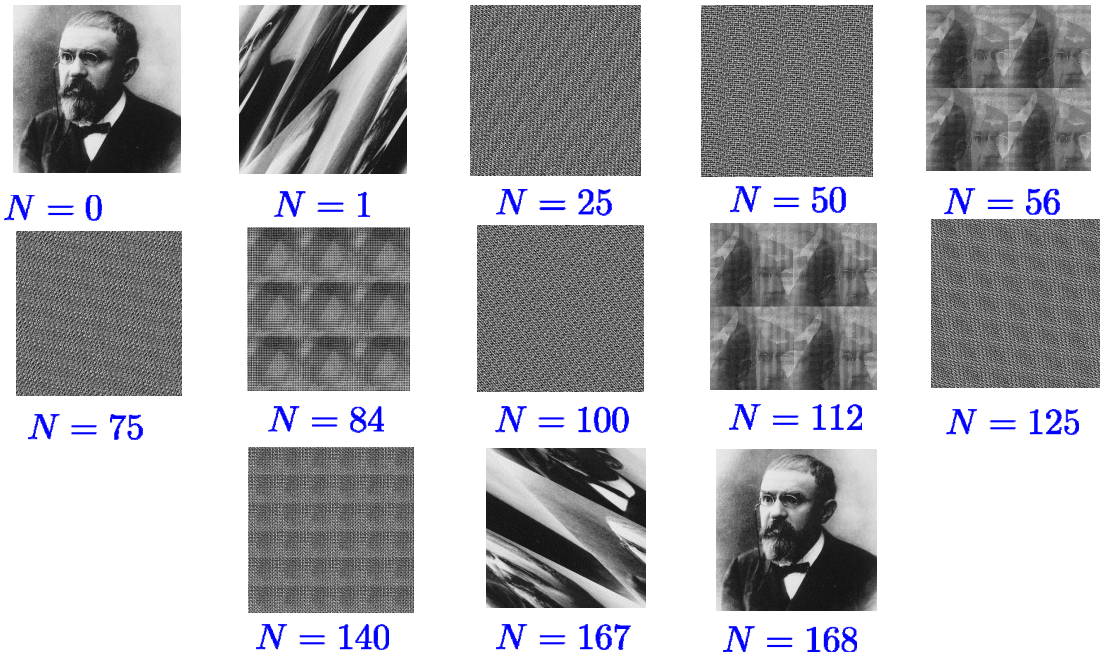


Figure 6: Illustrating Poincaré recurrence for the discrete cat map with $N \times N$ pixels where $N = 294$. Thanks to Andrew Moore for writing the code.

fig:PoincareRe

4. The cat map can also be used to give a dramatic illustration of Poincaré recurrence. A computer screen will have a finite number of pixels. Let us say it has $N \times N$ pixels. If these are black or white there will be $2^{N \times N}$ possible images. [8] Now, consider the discrete version of the cat map: We take

♣Should comment that this is clearly NOT mixing because periodicity means the limit doesn't really exist. ♣

$$x, y \in \frac{1}{N}\{0, 1, \ldots, N-1\} \tag{3.9}$$

and apply the above transformation. Since there are only $2^{N \times N}$ images the transformation must be periodic. See Figure 6 for $N = 294$. It is rather astonishing that the period is only 164. Due to some special number-theoretic aspects of this example (related to Fibonacci numbers) one can give an exact formula for a period (not necessarily the fundamental period), which in this case turns out to confirm $N = 164$. Using this formula it turns out that the period is $\leq 3N$. [9]

[8] Actually, there are 256 shades of grey, so actually there are 2^{8N^2} images on a typical computer screen.

[9] F.J. Dyson and H. Falk, "Period of a Discrete Cat Mapping," Amer. Math. Monthly, vol. 99 (1992), pp.603-614

5. In general, suppose that a $\mathbb{Z}$-action on a manifold M is generated by a differentiable map $g_0 \cdot x = f(x)$, where $f : M \to M$, and suppose that f has a fixed-point $f(x_0) = x_0$. Then, near the fixed point we can write, in local coordinates:

$$f(x_0 + \delta x) = f(x_0) + df(\delta x) = x_0 + df(\delta x) + \mathcal{O}((\delta x)^2) \tag{3.10}$$

where we identify an infinitesimal deviation with a tangent vector. In general $df : T_{x_0}X \to T_{x_0}X$ is just a real linear transformation and cannot be diagonalized. Suppose, however, that it can be diagonalized. Then there is a basis of $T_{x_0}X$ such that df has the matrix representation $A = \mathrm{Diag}\{\alpha_1, \ldots, \alpha_n\}$ for an n-dimensional manifold. If we consider orbits that begin close to x_0 then choose coordinates so that $x_0 = 0 \in \mathbb{R}^n$ and we have $x_{n+1} \cong (1 + A)x_n$. In the directions with $\alpha_j < 0$ the orbits contract to the fixed point. In the directions $\alpha_j > 0$ they expand away from the fixed point (and soon go beyond the linear approximation). If $\alpha_j = 0$ then we need to go to higher order to determine if the fixed point is isolated.

Remark: A very important dynamical system for quantum field theory is known as the "renormalization group." It has both discrete and continuous forms. One of the discrete forms is known as the block spin method. The dynamical system evolves on the infinite-dimensional space of all possible local couplings of the field theory. Fixed point loci are known as "scale invariant theories," and they are usually conformal field theories. One then defines local quantum field theory by a scaling of couplings near a fixed point. The key miracle is that, in this infinite dimensional space of couplings all but finitely many directions are attractive ("irrelevant operators") and only a finite number of directions are repulsive ("relevant operators") so, after a finite number of choices one has predictive power, at least for renormalizable quantum field theories.

Exercise

Consider an action of a discrete group on a topological space X. Show that properly discontinuous implies wandering implies discrete orbits.

Exercise

a.) Show that the Arnold cat map is a product of a shear by one unit in the x-direction followed by a shear by one unit in the y-direction.

b.) Find a square root of the Arnold cat map.

3.1 Left and right group actions of G on itself

Let G be any group. Then G acts on itself as a transformation group, in several ways.

To define the *left action of G on G* we associate to each $a \in G \to$ the *mapping* $L(a)$, $L(a) : G \to G$ defined by

$$L(a) : g \mapsto ag \tag{3.11}$$

where on the RHS we use the group multiplication of G. This mapping is 1-1 and onto, so $L(a) \in S_G$.

These transformations satisfy:

$$L(ab) = L(a)L(b) \tag{3.12}$$ eq:grplaw

and moreover $L(1_G)$ is the identity mapping.

Thus we have defined a left-action of G on itself. It is effective, free, and transitive.

Indeed, we can define a map $\mathcal{L} : a \mapsto L(a)$ which is a homomorphism $\mathcal{L} : G \to S_G$. Moreover,

$$L(a) = 1 \Longleftrightarrow a = 1 \tag{3.13}$$

and hence $\ker\mathcal{L} = \{1_G\}$, so the image of $\mathcal{L}$ is isomorphic to G. Thus, we have proved

Theorem ??.1 (Cayley's Theorem): Any group G is isomorphic to a subgroup of the full permutation group S_G. If $n = |G| < \infty$ then G is isomorphic to a subgroup of S_n.

Warning: For a fixed a, although $L(a)$ is a map $G \to G$, it is *not* a homomorphism! Note, for example that $L(a)$ takes 1_G to a. Do not confuse this with the fact that the mapping $\mathcal{L} : a \to L(a)$ is a homomorphism.

All of this can be repeated for right-actions: For $a \in G$ define the *right-translation operator* $R(a) : G \to G$ by $R(a) : g \mapsto g \cdot a$.

Then $\phi_a = R(a)$ defines a right-action of G on itself, and hence $\tilde{\phi}_a = R(a^{-1})$ defines a left-action of G on itself.

It should be fairly obvious that

$$R(a)L(b) = L(b)R(a) \tag{3.14}$$

for all $a, b \in G$. Thus, there is a left $G \times G$ action on G defined by:

$$\phi_{(g_1,g_2)} := L(g_1)R(g_2^{-1}) \tag{3.15}$$ eq:GGonG

Exercise

Since there is a left-action of $G \times G$ on $X = G$ there is a left-action of the diagonal subgroup $\Delta \subset G \times G$ where $\Delta = \{(g, g)|g \in G\}$ is is a subgroup isomorphic to G.

a.) Show that this action is given by $a \mapsto I(a)$, where $I(a)$ is the conjugation by a. (See, Chapter 1, Section ***)

b.) Show that the orbits of Δ are the conjugacy classes of G.
c.) What is the stabilizer subgroup of an element $g_0 \in G$?

---

4. Spaces of orbits

We now shift our focus from studying orbits to studying *the space of orbits*.

In general, if G has a right-action on a set X then the set of distinct G-orbits is denoted X/G. If G has a left-action then the set of distinct G orbits is denoted $G\backslash X$.

The study of such spaces of orbits is rather vast. Some of the ways spaces of orbits enter into physics are the following:

1. Spaces of orbits such as homogeneous spaces given beautifully symmetric manifolds (or orbifolds). Thus they form a rich source of geometrical constructions and can be used in constructing spacetimes or discussing moduli spaces of solutions to equations.

2. A natural source of orbits is *Hamiltonian dynamics*. The time evolution of a dynamical system naturally defines a system of $\mathbb{R}$ or $\mathbb{Z}$ -orbits on phase space.

3. If G is a global symmetry group in a field theory, then the vacua of the theory are a union of orbits of G, and hence a union of homogeneous spaces G/H. In general H will be different in different connected components. It is referred to as the *unbroken symmetry* in the theory of spontaneous symmetry breaking.

4. Often, it is convenient to introduce redundant variables in a physical problem to make some other property of the physics (such as locality) manifest. In this case G is a *gauge symmetry* and it acts on the set of redundant variables X while the space of orbits X/G are the physically inequivalent variables. The canonical example is obtained by identifying X with the set of all gauge potentials and G with the group of gauge transformations. For example, on $\mathbb{M}^{1,3}$ the gauge potentials A_μ are redundant variables. The group $\mathcal{G} = \mathrm{Map}(\mathbb{M}^{1,3}, \mathbb{R})$ acts by $A_\mu \to A_\mu + \partial_\mu \chi$. The orbit space is parametrized by the fieldstrength $F_{\mu\nu}$.

We cannot cover this topic in proper detail here, but just indicate some examples and definitions to give a taste of the subject.

Remarks

1. As an example, let us consider the case where $X = G$ is itself a group. Let $H \subset G$ be a subgroup and consider the right-action of H on the set X. The set of orbits is G/H. This is in accord with our notation for the set of left-cosets of a subgroup H in a larger group G. Note that the set of orbits still admits a left-action by G. If $K \subset G$ is another subgroup then the set of orbits of the left-action of K on G/H is known as a *double coset* and denoted $K\backslash G/H$.

2. Warning! The notation X/G is somewhat ambiguous, as a space X can admit more than one group action. For example, if $X = G$ and we use right-translation $R(g)$ then X/G is a single point. On the other hand, if $X = G$ and the G action is by conjugation:

$$\phi_g(h) := ghg^{-1} \tag{4.1}$$

then the space X/G is the set of conjugacy classes in G and always has more than one point, if G is nontrivial.

3. When G is a topological group acting on a topological space X we can make X/G into a topological space: We use the quotient topology under the equivalence relation of being G-related. Put differently, the topology on X/G is defined by requiring that $p : X \to X/G$ be a continuous map. That is, $\mathcal{U} \subset X/G$ is open if $p^{-1}(\mathcal{U})$ is open. Even when X and G are relatively "simple" the quotient can be quite subtle and complicated. For example, even if X and G are Hausdorff the quotient might not be Hausdorff. So, it is useful to have some criteria for when quotients are "well-behaved."

4. If G is a discrete group we defined the notion of a "properly discontinuous action" above. In this case we have a nice

   Theorem: Let G be a discrete group acting freely and properly discontinuously on a Hausdorff manifold X. Then the quotient X/G is a Hausdorff manifold.

   To prove this, just note that around each point $x \in X$ there is an open neighborhood U homeomorphic to $\mathbb{R}^n$ such that $g \cdot U \cap U = \emptyset$ for all $g \neq 1$. This U will serve as a local neighborhood. To check that it is Hausdorff show that if x_1, x_2 are on distinct orbits then we can then there are compact neighborhoods K_1 and K_2 of x_1, x_2 which are disjoint. For a more complete proof see Thurston, Proposition 3.5.7.

5. When G is a Lie group one useful criterion is the

   Theorem: [Quotient manifold theorem]: Let G be a Lie group acting smoothly, freely, and properly, on a smooth manifold X. Then X/G is a manifold of dimension $\dim X - \dim G$ and the projection map $p : X \to X/G$ is a submersion.

   For a proof see, J. Lee, *Introduction to smooth manifolds*, pp. 218-223.

♣SHOULD GIVE IDEA OF THE PROOF! ♣

4.1 Simple examples

Example 1 . $G = \mathbb{Z}$ acts properly discontinuously on $\mathbb{R}$ via $n \cdot x = x + n$. The orbits are in 1-1 correspondence with $[0, 1]/\sim$ where $\sim$ identifies $0 \sim 1$. Note that therefore the set of orbits is in one-one correspondence with $\mathbb{R}/\mathbb{Z}$, and can be identified with points on the circle S^1.

Example 2. This generalizes: Let Λ be a two-dimensional lattice. It acts properly discontinuously on $\mathbb{R}^2$ by translation. Then

$$\mathbb{R}^2/\Lambda \tag{4.2}$$

is geometrically realized as a doughnut, or torus. Of course, nothing is special about two dimensions in this example, if Λ is an n-dimensional lattice in $\mathbb{R}^n$ then it acts on $\mathbb{R}^n$ by translations and the set of orbits is an n-dimensional torus.

Example 3: If we take $X = S^2$ and $G = SO(2)$ acting by rotations around some axis then the space of orbits is the closed interval.

Example 4: If $X = \mathbb{R}^{n+1} - \{0\}$ and $G = \mathbb{R}^*$ acting by scalar multiplication then $X/G \cong \mathbb{RP}^n$. Similarly if $X = \mathbb{C}^{n+1} - \{0\}$ and $G = \mathbb{C}^*$ acting by scalar multiplication then $X/G \cong \mathbb{CP}^n$. In these cases the quotient manifold theorem is not so useful because the groups involved are not compact so it is not easy to check the map is proper. However, if we think of $\mathbb{CP}^n$ as the moduli space of lines through the origin then we can put the standard Hermitian structure on $\mathbb{C}^{n+1}$ and observe that for every line there exists a basis vector v of norm one: $v^* \cdot v = 1$. But

$$\{v \in \mathbb{C}^{n+1} | v^* \cdot v = 1\} \cong S^{2n+1} \tag{4.3}$$

This ON vector is not unique. We can define an action of $U(1)$ on S^{2n+1} by $v \to e^{\mathrm{i}\theta} v$. Then $S^{2n+1}/U(1)$ is a manifold by the quotient manifold theorem, and is another presentation of $\mathbb{CP}^n$. Now that we see that a given space of orbits can be realized as a quotient space if different ways we should note a third presentation: $GL(n+1, \mathbb{C})$ acts on $\mathbb{CP}^n$ by linear action on the homogeneous coordinates. The action is clearly transitive. Therefore, by the stabilizer-orbit theorem we need only find the stabilizer subgroup of a convenient point, such as $[1 : 0 : \cdots : 0]$. The stabilizer subgroup of this point is the subgroup B of $GL(n+1, \mathbb{C})$ consisting of matrices of the form:

$$\begin{pmatrix} g_{11} & g_{12} \cdots & g_{1,n+1} \\ 0_{n\times 1} & \tilde{g} & \end{pmatrix} \tag{4.4}$$

where $\tilde{g} \in GL(n, \mathbb{C})$ and $g_{11} \in \mathbb{C}^*$. Then $\mathbb{CP}^n$ can be identified with the complex homogeneous space $GL(n+1, \mathbb{C})/B$.

Example 5: If we take $X = \mathbb{R}^2$ with $G = SO_0(1,1)$ acting via Lorentz transformations then we saw above that the space of orbits X/G is a union of a 5 special points together with four copies of $(0, \infty)$. In this case X/G is not a Hausdorff space. See Section **** below.

♣Really belongs below in the section on how things can go wrong. ♣

4.2 Fundamental domains

One way to try to get a picture of the space of orbits of a group action on a space X is to form a *fundamental domain*. In order to define this notion recall that given a G action on X we can define an equivalence relation, saying $x \sim y$ ("x is G-related to y" if there is a $g \in G$ with $y = g \cdot x$. In these terms a fundamental domain is a subset $\mathcal{F} \subset X$ of elements providing a complete set of representatives of the distinct equivalence classes for this relation. Then:

$$X = \amalg_{g \in G} g \cdot \mathcal{F}, \tag{4.5}$$

so the points in $\mathcal{F}$ are in one-one correspondence with the points of X/G. Put differently:

Definition: A *fundamental domain* for a group action of G on X is a subset $\mathcal{F} \subset X$ such that:

1. No two distinct points in $\mathcal{F}$ are G-related.
2. Every point of X is G-related to some point in $\mathcal{F}$.

Example 1 . *Finite covers of the circle.* Consider:

$$X = U(1) \cong \{z : |z| = 1\} \cong S^1 \tag{4.6}$$

and let $G \cong \mathbb{Z}/N\mathbb{Z}$ be the subgroup of Nth roots of 1. This acts on X by multiplication. A fundamental domain is the set of $z = e^{i\theta}$ with θ varying over a half-open interval of length $2\pi/N$. For example we could take $0 \leq \theta < 2\pi/N$. Then X/G is another copy of S^1. The map $p : X \to X/G$ can be identified with the map $p(z) = z^N$.

Example 2 . If we consider $X/G = \mathbb{R}/\mathbb{Z} \cong S^1$ then a fundamental domain would be any set of points $\theta \leq x < \theta + 1$, for any θ. As we see from the previous two examples, fundamental domains are not unique.

Example 3 . If we consider $\mathbb{R}^2/\mathbb{Z}^2$ where $\mathbb{Z}^2$ acts by translations by independent vectors e_1, e_2 then one choice of fundamental domain would be the set of points $\vec{x}_0 + t_1 e_1 + t_2 e_2$ where we take the union

$$\{(t_1, t_2) : 0 < t_1, t_2 < 1\} \amalg \{(0, t_2) : 0 < t_2 < 1\} \amalg \{(t_1, 0) : 0 < t_1 < 1\} \amalg \{(0, 0)\} \tag{4.7}$$

Note that $\vec{x}_0$ is arbitrary, and any choice gives a fundmental domain.

Example 4 . In general *unit cell* $\bar{\mathcal{F}} \subset \mathbb{R}^n$ of an embedded lattice $\Lambda \subset \mathbb{R}^n$ is the closure of a fundamental domain for $\mathbb{R}^n/\Lambda$. One natural choice is given by *choosing* a basis e_i for the lattice together with a vector $\vec{x}_0 \in \mathbb{R}^n$ and defining

$$\bar{\mathcal{F}} = \vec{x}_0 + \{\sum_i t_i e_i | 0 \leq t_i \leq 1\}. \tag{4.8}$$

Since $\mathbb{R}^n$ has a natural Euclidean metric the lattice Λ inherits an inner product. In the basis e_i it is given by the symmetric form $G_{ij} := e_i \cdot e_j$. Note that the volume of the unit cell is

$$\mathrm{vol}(\bar{\mathcal{F}}) = \sqrt{\det G_{ij}} \tag{4.9}$$ eq:unitcell

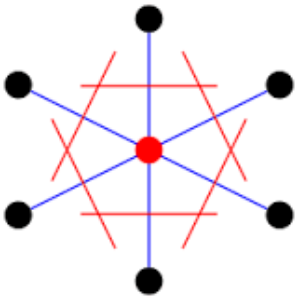


Figure 7: Constructing a Wigner-Seitz (or Voronoi) cell for the triangular lattice. The cells are regular hexagons. Figure from Wikipedia.

fig:WignerSeit

Example 5 . Given an embedded lattice $\Lambda \subset \mathbb{R}^n$ we can use the metric to produce a canonical (i.e. basis-independent) set of fundamental domains, known as *Voronoi cells* in mathematics and as *Wigner-Seitz cells* in physics. Choose any lattice point $v \in \Lambda$ and take $\bar{\mathcal{F}}$ to be the set of all points in $\mathbb{R}^n$ which are closer to v than to any other point. (If the points are equidistant to another lattice point we include them in the closure $\bar{\mathcal{F}}$. Thus, for the regular triangular lattice a Wigner-Seitz cell would be a regular hexagon centered on a lattice point. See Figure 7. In reciprocal space, the Wigner-Seitz cell for the reciprocal lattice is known in solid state physics as the *Brillouin zone*. Note that there is a clear algorithm for constructing $\bar{\mathcal{F}}$: Starting with v we look at all other points $v' \in \Lambda$. We consider the hyperplane perpendicular to the line between v and v' and take the intersection of all the half-planes containing v. It is also worth remarking that the concept of Voronoi cell does not require a lattice and applies to any collection of points, indeed, any collection of subsets of $\mathbb{R}^n$.

Example 6 . *The modular group.* The modular group is $PSL(2,\mathbb{Z}) := SL(2,\mathbb{Z})/\{\pm 1\}$, where $SL(2,\mathbb{Z})$ is the subgroup of $SL(2,\mathbb{R})$ of matrices all of whose matrix elements are integers. Recall that this group acts effectively on the complex upper half-plane $\mathcal{H}$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau := \frac{a\tau + b}{c\tau + d} \tag{4.10}$$

We will find a fundamental domain for this group action, and in the process prove that $SL(2,\mathbb{Z})$ is generated by the group elements S and T defined by:

$$S := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{4.11}$$

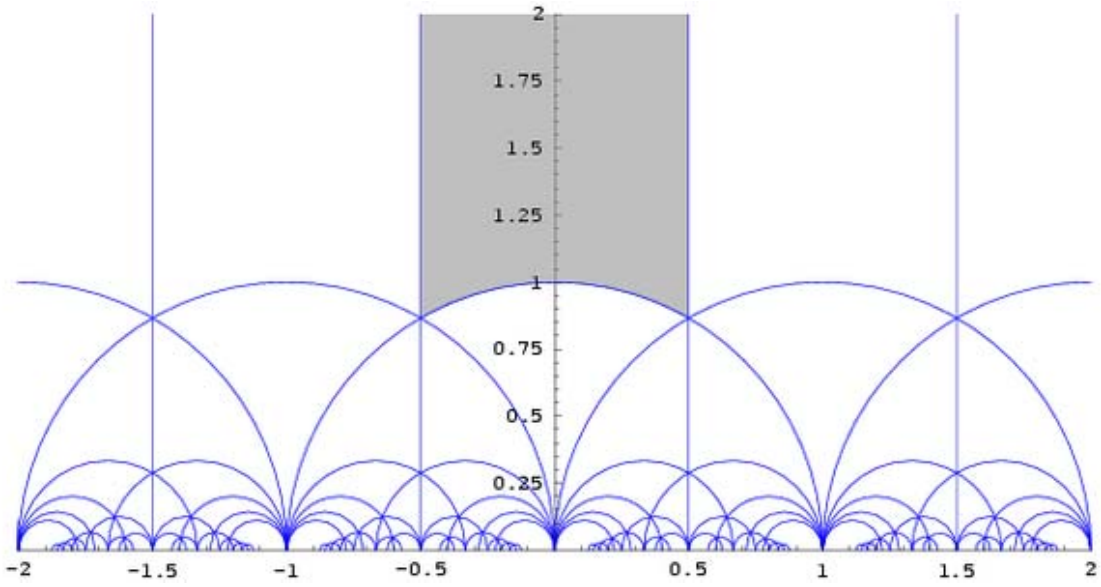


Figure 8: The keyhole region, a standard choice of fundamental domain for the action of $PSL(2,\mathbb{Z})$ on the complex upper half-plane. Figure from Wikipedia article on "Modular Group". fig:KeyholeReg

$$T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{4.12}$$

Denote their images in $PSL(2,\mathbb{Z})$ by $\bar{S}, \bar{T}$. [10]

Let:

$$\tilde{\mathcal{F}} := \{\tau \in \mathcal{H} || \tau| \geq 1 \quad \& \quad |\mathrm{Re}(\tau)| \leq \frac{1}{2}\} \tag{4.13}$$

This is almost, but not quite the canonical fundamental domain for the modular group. It is the famous *keyhole region* shown in Figure 8. Let $\bar{G}$ be the subgroup generated by $\bar{S}$ and $\bar{T}$. We claim that $\cup_{g\in\bar{G}} g\cdot\tilde{\mathcal{F}}$ is the entire half-plane. To prove this recall that, for any $g \in SL(2,\mathbb{Z})$,

$$\mathrm{Im}(g\cdot\tau) = \frac{\mathrm{Im}\tau}{|c\tau + d|^2} \tag{4.14}$$

Now, for any fixed $\tau \in \mathcal{H}$ the function $|c\tau + d|$ is bounded below on $SL(2,\mathbb{Z})$, and hence

[10] We are here following a very nice argument by J.-P. Serre, *A Course in Arithmetic*, Springer GTM 7, pp. 78-79.

on G. Indeed, decomposing $\tau = x + \mathrm{i}y$ into its real and imaginary parts

$$|c\tau + d|^2 = (cx+d)^2 + c^2y^2 \geq \begin{cases} y^2 & c \neq 0 \\ d^2 \geq 1 & c = 0 \end{cases} \tag{4.15}$$

Therefore, for any fixed τ there will exist a group element $g \in \bar{G}$ such that $\mathrm{Im}(g\cdot\tau)$ takes a maximal value as a function of g. Note that multiplying g on the left by a power of $\bar{T}$ or $\bar{T}^{-1}$ does not change this property, so there is not a unique g which maximizes $\mathrm{Im}(g\cdot\tau)$. We can fix the ambiguity by requiring $|\mathrm{Re}(g\cdot\tau)| \leq \frac{1}{2}$. Choose such a group element g. We claim that for this transformation, $\tau' = g\cdot\tau \in \tilde{\mathcal{F}}$. We need only check that $|\tau'| \geq 1$. If not, then $|\tau'| < 1$ but then $\mathrm{Im}(\bar{S}\cdot\tau') = \mathrm{Im}(\tau')/|\tau'|^2 > \mathrm{Im}(\tau')$, contradicting the definition of g. In conclusion, every element of the upper half-plane can be brought to $\tilde{\mathcal{F}}$ by a suitable element of $\bar{G}$.

Now we need two Lemmas:

Lemma 1: If $g \in SL(2,\mathbb{Z})$ and τ have the property that both $\tau \in \tilde{\mathcal{F}}$ and $g\cdot\tau \in \tilde{\mathcal{F}}$ then

1. $|\mathrm{Re}(\tau)| = \frac{1}{2}$ and $g\cdot\tau = \tau \pm 1$, or
2. $|\tau| = 1$

To prove this note that, WLOG, we may assume that $\mathrm{Im}(g\cdot\tau) \geq \mathrm{Im}\tau$. (If not replace $g \to g^{-1}$ and $\tau \to g^{-1}\tau$.) But this equation implies $1 \geq |c\tau + d|$ which in turn implies:

$$\begin{aligned} 1 &\geq |c\tau+d|^2 \\ &= (cx+d)^2 + c^2y^2 \\ &= c^2|\tau|^2 + 2cdx + d^2 \\ &\geq c^2|\tau|^2 - |cd| + d^2 \\ &= c^2(|\tau|^2 - \frac{1}{4}) + (|d| - \frac{1}{2}|c|)^2 \\ &\geq \frac{3}{4}c^2 + (|d| - \frac{1}{2}|c|)^2 \end{aligned} \tag{4.16}$$

eq:Inequals

From (4.16) we conclude:

1. $(c=0, d=\pm 1)$ or $(c=\pm 1, d=0,\pm 1)$.
2. The inequalities are saturated iff $2cdx = -|cd|$ and $|\tau| = 1$.

If $c=0$ and $d=\pm 1$ then $g\cdot\tau = \tau \pm 1$. In this case it is clear that $|\mathrm{Re}(\tau)| = \frac{1}{2}$. If $c=\pm 1$, and $d = 0, \pm 1$ then the inequality is saturated, and hence $|\tau| = 1$.

Lemma 2: If $\tau \in \tilde{\mathcal{F}}$ and $\bar{g}\cdot\tau = \tau$ with $\bar{g} \in PSL(2,\mathbb{Z})$ and $\bar{g} \neq 1$ then either

1. $\tau = \mathrm{i}$ and the stabilizer group is $\{1, \bar{S}\}$

2. $\tau = \omega = e^{2\pi \mathrm{i}/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i}$ and the stabilizer group is

$$\{1, \bar{S}\bar{T}, (\bar{S}\bar{T})^2\} \tag{4.17}$$

3. $\tau = -\omega^2 = e^{\pi \mathrm{i}/3} = \frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i}$ and the stabilizer group is

$$\{1, \bar{T}\bar{S}, (\bar{T}\bar{S})^2\} \tag{4.18}$$

In particular, for all other points $\tau \in \tilde{\mathcal{F}}$, the stabilizer group is the trivial group.

Lemma 2 follows quickly from Lemma 1: If $g\tau = \tau$ then we must be in the case $c = \pm 1$. If $d = 0$ then $a - 1/\tau = \tau$ for some integer a. But we must also have $|\tau| = 1$ and hence $a = \tau + \bar{\tau}$. Since a is an integer we quickly find that $a = 0$ with $\tau = \mathrm{i}$, or $a = \pm 1$ with $\tau = \omega$ or $-\omega^2$. If $d = \pm 1$ then from the saturation condition $2cdx = -|cd|$ we get $x = -\frac{1}{2}d/|d|$ and hence $\tau = \omega$ or $= -\omega^2$.

Now we can finally prove:

Theorem: $SL(2,\mathbb{Z})$ is generated by S and T.

Proof: Let τ_0 be in the <u>interior</u> of $\tilde{\mathcal{F}}$. Then choose any element $g \in SL(2,\mathbb{Z})$ with $\bar{g} \neq 1$. Then there is an element $g' \in \bar{G}$ so that $g'g \cdot \tau_0 \in \tilde{\mathcal{F}}$. Moreover, this element must be in the interior of $\tilde{\mathcal{F}}$ by Lemma 1 and and hence, by Lemma 2, $g'g = 1$ in $PSL(2,\mathbb{Z})$. Therefore $\bar{g} \in \bar{G}$, which means $\bar{G} = PSL(2,\mathbb{Z})$. Moreover, $S^2 = -1$, and hence S and T generate $SL(2,\mathbb{Z})$. ♠

1. The exact fundamental domain $\mathcal{F}$ must be chosen so that no two distinct points on the boundary are G-related. So, for example, we could choose the part of the boundary with $\mathrm{Re}(\tau) \geq 0$.

2. One can show that the relations on S and T are:

$$S^2 = -1 \qquad (ST)^3 = -1 \tag{4.19}$$

3. Given $g \in SL(2,\mathbb{Z})$ it is possible to write the word in S, T giving g by applying the Euclidean algorithm to (a, c) and interpreting the standard equations there in terms of matrices. See Chapter 1, Section 8.

4. Although the keyhole region is the standard fundamental domain there is no unique choice of fundamental domain. For example, one could equally well use any of the images shown in Figure 8 (and of course there are infinitely many such regions). Moreover, we could displace $\mathcal{F} \to \mathcal{F} + \epsilon$ and still produce a fundamental domain.

5. The action of $PSL(2,\mathbb{Z})$ is properly discontinuous on $\mathcal{H}$, but not quite free. If we consider finite-index subgroups that do not contain the stabilizer groups mentioned above then the action will be free and the quotient space will be a nice Riemann surface.

Exercise

a.) The *face-centered-cubic* lattice in $\mathbb{R}^3$ is the sublattice of $\mathbb{Z}^3$ of all points such that $\sum_i x_i = 0\text{mod}2$. Construct the Wigner-Seitz cell for the fcc lattice.

b.) The *body-centered-cubic* lattice in $\mathbb{R}^3$ is the sublattice of $\mathbb{Z}^3$ such that the coordinates x_i are all even or all odd. Construct the Wigner-Seitz cell for the bcc lattice.

Exercise

a.) The group $\Gamma_0(2)$ is the subgroup of $SL(2,\mathbb{Z})$ of matrices with $c = 0\text{mod}2$. Find a fundamental domain for $\bar{\Gamma}_0(2)$ acting on $\mathcal{H}$.

b.) The group $\Gamma(2)$ is the subgroup of $SL(2,\mathbb{Z})$ of matrices congruent to 1 modulo 2. Find a fundamental domain for $\bar{\Gamma}(2)$.

4.3 Algebras and double cosets

TO BE WRITTEN

4.4 Orbifolds

An interesting class of examples where the quotient space is "almost" a manifold are called "orbifolds" or "V-manifolds."

In a manifold, neighborhoods of points locally look like copies of $\mathbb{R}^n$.

In an orbifold neighborhoods of points locally look like copies of $\mathbb{R}^n/\Gamma$ where Γ is a finite group acting on $\mathbb{R}^n$. The finite group Γ can depend on the point in question. For example, for "most" points it might be trivial. But for some other points it might be nontrivial

Example 1: Let $\mathbb{Z}_2$ act on $\mathbb{R}^n$ by $\sigma \cdot \vec{x} = -\vec{x}$. Then in the quotient space $\mathbb{R}^n/\mathbb{Z}_2$ the neighborhood of every point is homeomorphic to $\mathbb{R}^n$ except for the origin. A neighborhood of $[\vec{0}]$ is a cone on $\mathbb{RP}^{n-1}$. Since $\mathbb{RP}^{n-1}$ is not homotopy equivalent to S^{n-1} for $n > 2$ this space cannot be homeomorphic to $\mathbb{R}^n$ for $n > 2$.

Example 2: Consider the n-dimensional torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$. Let $\Gamma = \mathbb{Z}_2$ act on T^n with the action induced from $\sigma \cdot \vec{x} = -\vec{x}$ on $\mathbb{R}^n$. There are now 2^n fixed points. At each fixed point the local neighborhood is homeomorphic to $\mathbb{R}^n/\mathbb{Z}_2$.

Example 3: Let $G \cong \mathbb{Z}/N\mathbb{Z}$ be the group of Nth roots of 1 acting on the complex plane by multiplication. The quotient $\mathbb{C}/\mathbb{Z}_N$ is an orbifold. The stabilizer group is trivial everywhere except the origin, where it is all of G. A neighborhood of 0 should be viewed as a cone with opening angle $2\pi/N$.

Example 4: Recall that we showed that we can identify $\mathcal{H} \cong SL(2,\mathbb{R})/SO(2)$. Now consider the double-coset

$$\Gamma\backslash SL(2,\mathbb{R})/SO(2) \tag{4.20}$$

where $\Gamma = SL(2,\mathbb{Z})$. We can identify this as the set of orbits of Γ on $\mathcal{H}$. From our analysis of the fundamental domain above we see that it is topologically a sphere with two orbifold singularities. The one at [i] has a neighborhood modeled on $\mathbb{C}/\mathbb{Z}_2$ and the one at $[\omega] = [-\bar{\omega}]$ has a neighborhood modeled on $\mathbb{C}/\mathbb{Z}_3$. In addition, there is one puncture (at $\tau = i\infty$).

Exercise *Weighted projective spaces*

Choose positive integers $p_1, \cdots p_{n+1}$. Then the *weighted projective space* $W\mathbb{P}[p_1, \ldots, p_{n+1}]$ is the space defined by $(\mathbb{C}^{n+1} - \{\vec{0}\})/\mathbb{C}^*$ with $\mathbb{C}^*$ action:

$$\lambda \cdot (z_1, \ldots, z_{n+1}) := (\lambda^{p_1} z_1, \ldots, \lambda^{p_{n+1}} z_{n+1}) \tag{4.21}$$

Show that the resulting quotient space is a well-defined orbifold. What are the orbifold singularities?

♣Do something with double cosets and Hecke algebras. ♣

4.5 Examples of quotients which are not manifolds

When we divide by a *noncompact group* like $\mathbb{C}^*$ or $GL(n,\mathbb{C})$ then, even when the action is free the quotient space M/G can be a "bad" (e.g. non-Hausdorff, or very singular and difficult-to-work-with) space.

Example 1 Consider the circle $S^1 \cong \mathbb{R}/\mathbb{Z}$. Consider translation $x \to x + \alpha$ where α is *irrational*. This generates an action of $\mathbb{Z}$ on S^1 (or $\mathbb{Z} \oplus \mathbb{Z}$ on $\mathbb{R}$) that is *not* properly discontinuous. The quotient space is *not* a manifold.

Example 2 . A similar example. Consider the two-dimensional torus, identified with $S^1 \times S^1 \cong \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$. Thus we can use coordinates (x, y) where x, y are defined modulo 1. Consider the group action by $\mathbb{R}$:

$$(x, y) \to (x + tv_1, y + tv_2) \tag{4.22}$$

where $t \in \mathbb{R}$, (v_1, v_2) is some vector. If the slope v_2/v_1 is a rational number the orbits are compact and the space of orbits is a nice space. (Exercise: What is this space?) But if the slope is irrational we again have a non-Hausdorff space since the orbits are all dense, so it is impossible to separate open sets around these orbits.

Example 3: Consider $X = \mathbb{C}^{n+1}$ and $G = \mathbb{C}^*$ acting as

$$\lambda \cdot \vec{z} = \lambda \cdot (z_1, \ldots, z_{n+1}) := (\lambda z_1, \ldots, \lambda z_{n+1}) \tag{4.23}$$

for $\lambda \in \mathbb{C}^*$. Then $X/\mathbb{C}^*$ is not Hausdorff. For consider the equivalence class $[\vec{0}]$. Let $\pi : X \to X/\mathbb{C}^*$ be the projection. An open neighborhood U of $[\vec{0}]$ is such that $\pi^{-1}(U)$ is

an open neighborhood of $\vec{0}$. Now consider any other point $[\vec{w}]$ where $\vec{w} \neq 0$. There will always be a $\lambda \in \mathbb{C}^*$ so that $\lambda\vec{w} \in \pi^{-1}(U)$. Therefore, $[\vec{w}] \in U$ for *all* points $[\vec{w}] \in X/\mathbb{C}^*$! In particular one cannot separate $[\vec{0}]$ and $[\vec{w}]$ by open sets. Clearly, there is one bad actor here, the point $\vec{0}$. Indeed if we eliminate it then the $\mathbb{C}^*$ action on $\mathbb{C}^* - \{0\}$ produces a good manifold

$$\mathbb{C}P^n = (\mathbb{C}^* - \{0\})/\mathbb{C}^* \tag{4.24}$$

Example 4: For a very similar example let us consider a $\mathbb{C}^*$ action on $\mathbb{C}^2$ with two coordinates ϕ_1, ϕ_2 and action:

$$\begin{aligned} \phi_1 &\to \lambda\phi_1 \\ \phi_2 &\to \lambda^{-1}\phi_2 \end{aligned} \tag{4.25}$$

eq:scalen

Generic orbits are labelled by the "gauge invariant" quantity $\phi_1\phi_2$. Note however that there are 3 special orbits corresponding to the value $\phi_1\phi_2 = 0$:

$$\begin{aligned} \mathcal{O}_1 &= \{(\phi_1, 0) | \phi_1 \neq 0\} \\ \mathcal{O}_2 &= \{(0, \phi_2) | \phi_2 \neq 0\} \\ \mathcal{O}_3 &= \{(0,0)\} \end{aligned} \tag{4.26}$$

Unlike projective space now the space

$$[\mathbb{C}^2 - \mathcal{O}_3]/\mathbb{C}^* \tag{4.27}$$

is NOT a Hausdorff space. Indeed the orbits $\mathcal{O}_1$ and $\mathcal{O}_2$ project to two distinct points $[\mathcal{O}_1]$ and $[\mathcal{O}_2]$ and we claim these cannot be separated from each other by open sets in the quotient topology. See the exercise below.

We could just consider the space

$$[\mathbb{C}^2 - (\mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{O}_3)]/\mathbb{C}^* \tag{4.28}$$

eq:easyquot

Now the quotient by $\mathbb{C}^*$ is Hausdorff and is a nice algebraic variety. It is just a copy of $\mathbb{C}^*$.

But we could compactify (4.28). One way to do this is to omit the axes $\phi_1 = 0$ and $\phi_2 = 0$, but include the origin, then the quotient space $\mathbb{C}^2 - (\mathcal{O}_1 \cup \mathcal{O}_2)$ is well-defined and in fact isomorphic to $\mathbb{C}$. But in fact, there are three ways of getting a good quotient by omitting any two of the three special orbits and then dividing by $\mathbb{C}^*$.

Remark: We did not use the complex numbers in an essential way in Example 4 so we can apply the very same ideas to the quotient of Minkowski space by the component of the identity of the Lorentz group. We can consider the real subspace $\mathbb{R}^2 \subset \mathbb{C}^2$ and interpret ϕ_1, ϕ_2 as light-cone coordinates. Then we consider the subgroup of $\mathbb{C}^*$ corresponding to $\mathbb{R}_+$. We studied the orbits in Section **** above. Using the same kind of reasoning as above we see that the quotient is not a Hausdorff space since the origin cannot be separated

from the lightlike orbits. Even if we remove the origin the quotient is not Hausdorff since we cannot separate timelike and spacelike orbits from lightlike orbits. If we remove the origin and the lightlike orbits *then* the quotient space is a Hausdorff space. It is naturally thought of as four copies of $\mathbb{R}_+$.

Example 5: Another example similar to the above is the following. Let us consider the set of conjugacy classes of all matrices in $M \in M_n(\mathbb{C})$ under the action of $S \in GL(n,\mathbb{C})$:

$$M \mapsto SMS^{-1} \tag{4.29}$$

Let us consider the conjugacy classes of matrices:

$$\mathcal{C}(n) := M_n(\mathbb{C})/GL(n,\mathbb{C}) \tag{4.30}$$

we could consider this as the set of linear transformations $T : V \to V$ up to change of basis, where V is an n-dimensional complex vector space.

The key fact here is that every matrix can be brought to Jordan canonical form: (See Linear Algebra User's Manual, chapter **** for more details, application, and a proof.)

Take the characteristic polynomial $p(x) = \det(x\mathbf{1} - M)$ and factor it into its distinct complex roots $p(x) = \prod_{i=1}^{s}(x-\lambda_i)^{r_i}$ where $r_i > 0$. Then, first of all we can bring M to block diagonal form $M = \oplus_{i=1}^{s} M_i$, where each block M_i is itself a block of Jordan matrices of type i, that is, there are positive integers n_α^i, $1 \le \alpha \le k_i$ with $\sum_\alpha n_\alpha^i = k_i$ such that

$$M_i = \oplus_{\alpha=1}^{k_i} J_{\lambda_i}^{(n_\alpha^i)} \tag{4.31}$$

where $J_\lambda^{(1)}$ is the 1×1 matrix with entry λ and, for $n > 1$, $J_\lambda^{(n)} = \lambda \mathbf{1}_{n\times n} + \sum_{j=1}^{n-1} e_{j,j+1}$ cannot be brought to a simpler form, and certainly cannot be diagonalized.

If a matrix M has nontrivial Jordan form then the quotient space $M_n(\mathbb{C})/GL(n,\mathbb{C})$ is non-Hausdorff in the neighborhood of $[M]$. To see why it suffices to consider the case $n = 2$: The characteristic polynomial $p(x) := \det(x\mathbf{1} - M)$ is now order two. When there are two distinct roots M is diagonalizable. When two roots coincide, say $p(x) = (x-\lambda)^2$, then there are two possibilities: Either $M = \lambda \mathbf{1}_{2\times 2}$ for some λ, or, M is conjugate to Jordan form:

$$M = S \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} S^{-1} \tag{4.32}$$

The existence of matrices with nontrivial Jordan blocks leads to non-Hausdorff behavior of the quotient. To see this it suffices to consider conjugation by matrices of the form:

$$g = \begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix} \in GL(2,\mathbb{C}) \tag{4.33}$$

then

$$g \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} g^{-1} = \begin{pmatrix} \lambda & z \\ 0 & \lambda \end{pmatrix} \tag{4.34}$$

with $z = t_1/t_2$ so

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \sim \begin{pmatrix} \lambda & z \\ 0 & \lambda \end{pmatrix} \tag{4.35}$$

for any $z \in \mathbb{C}^*$.

Now, suppose $\bar{V} \subset \mathcal{C}(2)$ is an open set in the quotient topology and moreover

$$[\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}] \in \bar{V} \tag{4.36}$$

$p^{-1}(\bar{V})$ is an open set in $M_2(\mathbb{C})$ containing the matrix $\lambda \mathbf{1}_{2\times 2}$. But any such open set must also contain

$$\begin{pmatrix} \lambda & z \\ 0 & \lambda \end{pmatrix} \tag{4.37}$$

for some sufficiently small z. Therefore,

$$[\begin{pmatrix} \lambda & z \\ 0 & \lambda \end{pmatrix}] = [\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}] \tag{4.38}$$

for $z \neq 0$, is a distinct point from

$$[\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}] \tag{4.39}$$

and yet any neighborhood of $[\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}]$ contains $[\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}]$. So, these distinct points cannot be separated by open sets.

If, by hand, we consider the subset of diagonalizable matrices in $M_2(\mathbb{C})$ then the quotient is better. If $m \in GL(2,\mathbb{C})$ is of the form:

$$m = g \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix} g^{-1} \qquad\qquad z_1 \neq z_2 \tag{4.40}$$

then $[m]$ is parametrized by the *unordered* pair $\{z_1, z_2\}$. This is just the configuration space $\mathbb{C}^2/\mathbb{Z}_2$ and is just an orbifold, with a $\mathbb{Z}_2$ orbifold singularity along the diagonal $\{z, z\}$.

Returning to the general case, if we consider $M_n^{\rm ss}(\mathbb{C}) \subset M_n(\mathbb{C})$ of diagonalizable matrices then the quotient $M_n^{\rm ss}(\mathbb{C})/GL(n,\mathbb{C})$ is much better behaved. It is an orbifold, with orbifold singularities corresponding to various symmetric groups along loci where eigenvalues coincide.

Remarks

1. Making nice quotients is part of the general subject of "geometric invariant theory." See, Fogarty, Kirwan, and Mumford, *Geometric Invariant Theory* for a sophisticated presentation of conditions in algebraic geometry for when quotients are "good."

2. The mathematics of Geometric Invariant theory finds several applications in supersymmetric field theory and string theory. First of all, it is important in forming various moduli spaces. (For example, we saw it was necessary in forming a good quotient space corresponding to the moduli space of lines.) Closely related to this, it is important in understanding the moduli spaces of vacua in SUSY field theory.

♣Explain more. Section in GMP2010 ch.5, section 6. ♣

3. The examples of quotients by ergodic actions such as the irrational rotation on a circle forms one of the primary examples in the study of *noncommutative geometry*. See, A. Connes, *Noncommutative Geometry* for much detailed discussion.

Exercise

Consider a small open neighborhood of $[\mathcal{O}_1]$:

$$\{[(1, z_2)] : |z_2| < \epsilon\} \tag{4.41}$$

and a small open neighborhood of $[\mathcal{O}_2]$:

$$\{[(z_1, 1)] : |z_1| < \epsilon'\} \tag{4.42}$$

a.) Show that when $n > 0$ in (4.25) these sets will intersect each other no matter how small we take ϵ, ϵ'

Therefore we cannot separate $[\mathcal{O}_1]$ from $[\mathcal{O}_2]$ and the quotient space is not Hausdorff.

4.6 When is the quotient of a manifold by an equivalence relation another manifold?

A natural question which arises from these examples is, more generally, when the quotient of a manifold by a general equivalence relation is another manifold.

An equivalence relation $\sim$ on any set X has a *graph* which is the subset $R \subset X \times X$ defined by

$$R = \{(x, y) | x \sim y\} \tag{4.43}$$

Conversely, one could define an equivalence relation from subsets of $X \times X$ satisfying certain properties.

We have already discussed how, if X is a topological space and $\sim$ is an equivalence relation then we can define a topological space with a continuous projection $p : X \to X/\sim$. A question which frequently arises is this: If M is a manifold, when is $M/\sim$ a manifold?

One criterion for answering this question is the following theorem: [11]

Theorem: If M is a smooth manifold and $\sim$ is an equivalence relation then the following are equivalent:

[11] For a proof see Theorem 8.3 of the notes "Differential Geometry," by R.L. Fernandes, in http://www.math.illinois.edu/~ruiloja/Math519/notes.pdf.

1. $M/\sim$ is a smooth manifold and $p : M \to M/\sim$ is a submersion.

2. The graph $R \subset M \times M$ is a proper smooth submanifold and the projection $p_1 : M \times M$ onto the first factor, when restricted to R, is a proper submersion.

♣SHOULD ILLUSTRATE HOW THIS THEOREM IS USEFUL. ♣

5. Isometry groups

♣AT THIS POINT PROBABLY SHOULD SPLIT OFF SEPARATE CHAPTERS ON ISOMETRY GROUPS AND REGULAR POLYTOPES AND ON SYMMETRIC FUNCTIONS. ♣

One natural way in which group actions appear in physics is via the set of transformations preserving some notion of distance. For example, we could consider transformations preserving lengths and angles. To fix ideas, let us consider $\mathbb{R}^3$ with Euclidean metric

$$\| \vec{x} \|^2 = (x^1)^2 + (x^2)^2 + (x^3)^2 \tag{5.1}$$

Definition: An *isometry* of the Euclidean metric is a map $T : \mathbb{R}^3 \to \mathbb{R}^3$ which is norm-preserving:

$$\| T(\vec{x}) - T(\vec{y}) \| = \| \vec{x} - \vec{y} \| \tag{5.2}$$

The group of all isometries is denoted $E(3)$. More generally, the group of isometries of Euclidean $\mathbb{R}^n$ is denoted $E(n)$.

Two examples:

1.) **Translations**

$$T_{\vec{a}} : \vec{x} \mapsto \vec{x} + \vec{a} \tag{5.3}$$

Now

$$T_{\vec{a}} \circ T_{\vec{b}} = T_{\vec{a}+\vec{b}} \tag{5.4}$$

so the $T_{\vec{a}}$'s form an abelian subgroup $\mathcal{T}$ of $E(3)$. We have $\mathcal{T} \cong \mathbb{R}^3$.

2.) **Rotations**

For $R \in O(3)$ we let $R \cdot \vec{x}$ be the vector with components $(R\vec{x})^i = R^i{}_j x^j$.

One can show [12] that the *full* isometry group of $\mathbb{R}^3$ is the semidirect product of $\mathcal{T}$ with $O(3)$:

$$E(3) \cong \mathcal{T} \ltimes O(3) \tag{5.5}$$

To see that this is a semidirect product let

$$(\vec{a}, R) \cdot \vec{x} = R\vec{x} + \vec{a}. \tag{5.6}$$

Then

$$(a_1, R_1)(a_2, R_2) = (a_1 + R_1 a_2, R_1 R_2) \tag{5.7}$$

[12] This is surprisingly hard. See P. Yale, *Geometry and Symmetry*, Holden-Day, San Francisco, California, 1968, for a proof.

clearly defines a twisted multiplication law. Put differently, every element of $E(3)$ is a product of a translation and a rotation and the subgroup of translations is a normal subgroup. Indeed, $O(3)$ acts as a group of automorphisms on $\mathcal{T}$ where $O(3) \to Aut(\mathcal{T})$ is defined by

$$\alpha_R : T_{\vec{a}} \to T_{R\vec{a}} \tag{5.8}$$

This follows from

$$RT_{\vec{a}}R^{-1} = T_{R\vec{a}} \tag{5.9}$$

We can define a homomorphism $\det : E(3) \to \mathbb{Z}_2$, by taking $\det R$. The kernel of this homomorphism are the *proper transformations* and the complement are the *improper transformations*.

It is useful to have simple pictures of the different group elements in $E(3)$. We begin with the following theorem:

Theorem:

a.) Every rotation in $SO(3)$ is a rotation about an axis.

b.) More generally, every rotation $R \in SO(n)$ in an odd number of dimensions fixes some line.

Proof: Note that for $R \in SO(n)$, $R - I = R(I - R^{tr})$ and $\det R = 1$ so

$$\det(R - I) = \det(I - R^{tr}) = \det(I - R) = (-1)^n \det(R - I) \tag{5.10}$$

so $\det(R - I) = 0$ for n odd. ♠

Note: In part (b) we are not asserting that R is a rotation about an axis.

Exercise *Geometrical interpretation of the elements and conjugacy classes of* $E(3)$

Denote by $C_{\vec{k}}(\theta)$ a rotation by θ about an axis through a vector $\vec{k}$.

a.) If $0 < \theta < 2\pi$ and $\vec{a}$ is perpendicular to $\vec{k}$ show that $(\vec{a}, C_{\vec{k}}(\theta))$ has an infinite number of fixed points, giving a straight line in $\mathbb{R}^3$ parallel to the axis through $\vec{k}$. Indeed $(\vec{a}, C_{\vec{k}}(\theta))$ is a rotation by θ around this axis.

b.) Show that every proper Euclidean transformation is a *screw displacement*. This is a rotation about some axis followed by a translation along that axis.

c.) Let $\sigma_{\vec{k}}$ be the reflection in a plane through the origin perpendicular to $\vec{k}$. Show that

$$\sigma_{\vec{k}}\vec{x} = \vec{x} - \frac{2(\vec{x} \cdot \vec{k})}{(\vec{k} \cdot \vec{k})}\vec{k} \tag{5.11}$$

d.) Show that any improper rotation in $O(3)$ can be written as $S_{\vec{k}}(\theta) = \sigma_{\vec{k}} C_{\vec{k}}(\theta)$.

e.) Show that $(\vec{a}, \sigma_{\vec{k}})$ is the reflection in a plane parallel to the plane through the origin and perpendicular to $\vec{k}$, and translation by a vector in that plane. This is called a *glide reflection*.

f.) Show that $(\vec{a}, S_{\vec{k}}(\theta))$ for $0 < \theta < 2\pi$ is a rotation-inversion about some point.

g.) The distinct conjugacy classes in $E(3)$ are:

1. Translations: $\{T_{\vec{a}} | \parallel \vec{a} \parallel = r\}$, $r > 0$. Infinite conjugacy class.
2. Screw displacements: $\{(\vec{a}, C_{\vec{k}}(\theta))$, $0 < \theta < \pi$. Parametrized by TS^2.
3. Glide reflections: $\{(\vec{a}, \sigma_{\vec{k}})\}$.
4. Rotation-inversions $(\vec{a}, S_{\vec{k}}(\theta))$ for $0 < \theta < 2\pi$. Parametrized by $\mathbb{R}^3 \times (SO(3) - \{1\})$.

A good reference for this material is
Miller, Symmetry Groups and Their Applications

6. Symmetries of regular objects

One very natural occurance of groups in physics, and elsewhere, is as symmetries of subsets of Euclidean space.

Suppose $\mathbb{R}^n$ has the Euclidean metric. Consider a subset $X \subset \mathbb{R}^n$. Then a *symmetry group of* X is a subgroup $G \subset E(n)$ such that G maps X to itself. The group of all isometries preserving X is the complete symmetry group of X.

Example : If X is the sphere then the complete symmetry group is $O(n)$.

In this section and the next we will examine the symmetry groups of various regular polygons and solids in $\mathbb{R}^2$ and $\mathbb{R}^3$. These are subgroups of $E(3)$. Now, classification of all subgroups would be very hard. For example, already in $SO(2)$ we could consider the group generated by several rotations by incommensurate irrational multiples of 2π.

It is good to restrict attention to *discrete groups*. These are groups such that if $B(r, \vec{x})$ is the sphere of radius r about $\vec{x}$ then for any $\vec{y}$, there are a finite number of points in the G-orbit of $\vec{y}$ in $B(r, \vec{x})$.

If X is a bounded region, then a symmetry group G of X must have some fixed point $\vec{y}_{cm}$ in $\mathbb{R}^3$. For a discrete group if $\vec{y} \in X$ the G-orbit of $\vec{y}$ must be finite (because X is bounded). Let us denote those points by $\{\vec{y}_1, \ldots, \vec{y}_n\}$. Then we can form the center of mass

$$\vec{y}_{cm} = \frac{1}{n} \sum_i \vec{y}_i \tag{6.1}$$

and it is easily seen to be a fixed point of G. By the classification of conjugacy classes above we see that G must be a subgroup of rotations or rotation-inversions about $\vec{y}_{cm}$, and is hence conjugate to a finite subgroup of $O(2)$ or $O(3)$.

subsec:sdisc

The finite subgroups of $SO(2)$ are just the cyclic groups of order n, denoted $C_n \subset SO(2)$. [13]

C_n is by definition the group of rotations about the origin $\vec{0}$ in $\mathbb{R}^2$ by $\theta = \frac{2\pi k}{n}$ $k = 0, 1, \cdots, n-1$ in the counterclockwise direction:

$$C_n := \{R(\frac{2\pi k}{n}) = \begin{pmatrix} \cos\frac{2\pi k}{n} & -\sin\frac{2\pi k}{n} \\ \sin\frac{2\pi k}{n} & \cos\frac{2\pi k}{n} \end{pmatrix}, k = 0, \ldots, n-1\} \tag{6.2}$$

[13] There are two common notations for subgroups of $E(3)$, the Schoenfliess notation and the international notation. We will be using the Schoenfliess notation.

when n is understood we denote the rotations by R_k.

Exercise
Show that $R_{k_1}R_{k_2} = R_{k_1+k_2 \mathrm{mod} n}$ and deduce the isomorphism: $\mathbb{Z}/n\mathbb{Z} \cong C_n$.

If we now consider $O(2)$ there is a new geometrical operation corresponding to parity. Suppose we take P to be reflection in the x-axis:

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{6.3}$$

Note that $P^2 = 1$ and

$$PR(\theta)P = R(-\theta) \tag{6.4}$$ eq:reflc

so the group generated by P and C_n is nonabelian for $n \geq 3$.

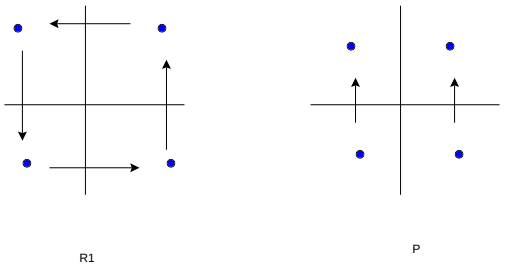


Figure 9: Operations of C_4 and P on 4 points. fig:prtbdiag

For $n = 4$ the action of R_1, P are shown in 9.

Moreover, a reflection in any other axis through the origin can be computed by rotating the axis to the x-axis, say by an angle ψ, then reflecting, and then rotating back. In equations this gives:

$$P_\psi = R(-\psi)PR(\psi) = PR(2\psi) = R(-2\psi)P \tag{6.5}$$

also satisfies

$$P_\psi^2 = 1 \qquad\qquad P_\psi R(\theta) P_\psi = R(-\theta) \tag{6.6}$$

Thus the groups generated by P_ψ and C_n, call them $D_{n,\psi}$ are all isomorphic as abstract groups, although they are different discrete subgroups of $O(2)$. Indeed, the $D_{n,\psi}$ are conjugate subgroups within $O(2)$. The common abstract group is known as the *dihedral group* and denoted as D_n.

Put more precisely D_n *is the group generated by* x, y *subject to the relations:*

$$\begin{aligned} x^2 &= 1 \\ y^n &= 1 \\ xyx^{-1} &= y^{-1} \end{aligned} \tag{6.7}$$ eq:rels

There is an isomorphism to the discrete subgroups of $O(2)$ given by $y \to R_1$ and $x \to P_\psi$.

Exercise *Generators and relations for the dihedral group D_n*

a.) Show that the relation $xyx^{-1} = y^{-1}$ can equivalently be written as

$$(xy)^2 = 1 \qquad \text{or} \qquad xy = y^{-1}x \qquad \text{or} \qquad xyx = y^{-1} \tag{6.8}$$

b.) Show that every element of D_n can be written as y^j or xy^j, and hence conclude that the order of D_n is $2n$.

c.) Show that there are n elements of order 2.

d.) The relations (6.7) show that D_n is nonabelian for $n \geq 3$. C_n is a normal subgroup. Show that

$$D_n/C_n \cong \mathbb{Z}_2 \tag{6.9}$$

e.) Indeed, define a homomorphism $\psi : D_n \to \mathbb{Z}_2$ whose kernel is C_n and deduce that

$$1 \to C_n \to D_n \to \mathbb{Z}_2 \to 1 \tag{6.10}$$

f.) Show that $\mathbb{Z}_2$ acts as a group of automorphisms on C_n and

$$D_n \cong C_n \ltimes \mathbb{Z}_2 \tag{6.11}$$

Exercise

Show that $D_2 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$

Exercise

Define a homomorphism $D_n \times \mathbb{Z}_2 \to D_{2n}$ as follows. Let

$$D_n = \langle x, z | x^2 = 1, z^n = 1, xzx = z^{-1} \rangle \tag{6.12}$$

$$D_{2n} = \langle x, y | x^2 = 1, y^{2n} = 1, xyx = y^{-1} \rangle \tag{6.13}$$

and let $\mathbb{Z}_2$ be generated by σ. Then define ϕ by its action on generators:

$$\begin{aligned} z &\to y^2 \\ x &\to x \\ \sigma &\to y^n \end{aligned} \tag{6.14}$$

a.) Show that ϕ is a homomorphism.

b.) Show that ϕ is an *isomorphism* for n odd, but is not an isomorphism for n even.

c.) Show that $\pi : D_{2n} \to D_n$ defined by $\pi : y \to z$ and $\pi : x \to x$ defines a homomorphism. Show that the kernel is isomorphic to $\mathbb{Z}_2$. Show that a section can be defined by

$$s : z \to y^{n+1} \qquad s : x \to x \tag{6.15}$$

and that this is a homomorphism for n odd, but not for n even.

d.) Show that D_{2n} is a central extension of D_n by $\mathbb{Z}_2$:

$$1 \to \mathbb{Z}_2 \to D_{2n} \to D_n \to 1 \tag{6.16}$$

And it is a split central extension for n odd.

6.1 Symmetries of polygons in the plane

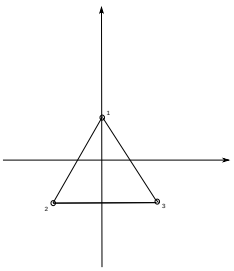


Figure 10: The symmetry group of the equilateral triangle is isomorphic to the symmetric group S_3. fig:triangle

Consider the symmetries of the equilateral triangle.

There are clearly 3 axes of reflection symmetry. If we label the vertices of the triangle $1, 2, 3$ they correspond to the three transpositions (12), (13), (23).

There are also 3-fold rotational symmetries about the center of mass of the triangle. These generate the subgroup C_3, and correspond to the cyclic transformations $1, (123), (132)$.

Thus, the group of symmetries is the full symmetric group S_3. On the other hand, it is also clearly the dihedral group D_3. Indeed, $S_3 \cong D_3$. The explicit isomorphism is:

$$\begin{aligned} 1 &\to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ (23) &\to \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ (123) &\to \begin{pmatrix} \cos(2\pi/3) & -\sin(2\pi/3) \\ \sin(2\pi/3) & \cos(2\pi/3) \end{pmatrix} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \end{aligned} \tag{6.17}$$

Note that this is a matrix representation of S_3.

Example .: The symmetries of the square.

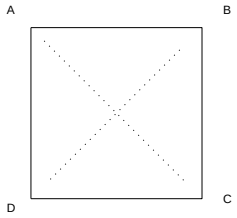


Figure 11: A square with labeled vertices. fig:symsq

Label the vertices A, B, C, D and consider the rotations and reflections in the plane preserving the square. Each such symmetry defines a permutation of the set $\{A, B, C, D\}$, and hence an element of S_4. Thus, the symmetries of the square form a subgroup of S_4.

Exercise

Label transformations as permutations of 1,2,3,4. Show there are 8 elements. Work out the multiplication table. Identify the symmetry group of the square as the dihedral group D_4.

Note that the full symmetry group is only a subgroup of S_4:

$$\{1, (1234)^j, (13), (24), (12)(34), (14)(23)\} \tag{6.18}$$

Exercise

D_4 can also be understood as a central extension of $\mathbb{Z}_2 \times \mathbb{Z}_2$ by $\mathbb{Z}_2$:

$$1 \to \mathbb{Z}_2 \to D_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \to 1 \tag{6.19}$$

Interpret this central extension geometrically.

Hint: $x \to \sigma_1, y \to \sigma_1\sigma_2$.

Figure 12: A regular n-gon. fig:ngoni

Finally, let us consider the regular n-gon in the plane, centered at the origin. It is clear that C_n is a subgroup, and there are reflection symmetries, so

The complete symmetry group of the regular n-gon in the plane is D_n.

Figure 13: Order two symmetry axis of the n-gon for n odd. fig:ngonii

Figure 14: Order two symmetry axis of the n-gon for n even. fig:ngoniii

Remarks *Conjugacy classes of D_n.* It is interesting to work out the conjugacy classes in D_n. First, let us think of it geometrically.

Now let us consider the elements of order 2, i.e. reflections in some axis.

1. If n is odd, each symmetry axis passes through a vertex and a face as in 13. There are n elements of order two, corresponding to PR_k, and they are all conjugate.

2. If n is even there are $n/2$ axes through antipodal vertices and $n/2$ axes through antipodal edge mid-points as in 14. Again we find n elements of order 2, but they separate into two sets of distinct conjugacy classes.

Now let us consider the question algebraically:

Note that $(xy^k)y^j(xy^k)^{-1} = y^{n-j}$. Geometrically, we are reflecting a rotation through an axis. Also note that $x(xy^j)x = xy^{n-j}$, while $y^k(xy^j)y^{-k} = xy^{j-2k}$. Now we must distinguish between n odd and n even.

For n odd we have

$$\begin{aligned} C(1) &= \{1\} \\ C(x) &= \{x, xy, \ldots, xy^{n-1}\} \\ C(y^j) &= \{y^j, y^{n-j}\} \qquad j = 1, \ldots, (n-1)/2 \end{aligned} \tag{6.20}$$ eq:dncci

This gives a total of $(n+3)/2$ distinct conjugacy classes.

For n even we have

$$\begin{aligned} C(1) &= \{1\} \\ C(x) &= \{x, xy^2, \ldots, xy^{n-2}\} \\ C(xy) &= \{xy, xy^3, \ldots, xy^{n-1}\} \\ C(y^j) &= \{y^j, y^{n-j}\} \qquad j = 1, \ldots, \frac{n}{2} - 1 \\ C(y^{n/2}) &= \{y^{n/2}\} \end{aligned} \tag{6.21}$$ eq:dncciev

This gives a total of $3 + n/2$ conjugacy classes. Note that there are two distinct conjugacy classes of order two elements. These are the two kinds of reflection axes.

Figure 15: A regular n-pyramid. fig:regpyr

6.2 Symmetry groups of some regular solids in $\mathbb{R}^3$

1. Consider the regular n-pyramid. Its base is a regular n-gon, and the distance to the vertex is not equal to the length of a side of the n-gon. A symmetry group is plainly C_n, rotations by multiples of $2\pi/n$ about the axis of the pyramid. To this we can add reflections in the symmetry planes to produce a subgroup of $O(3)$ known as C_{nv} in the Schoenfliess system to get the complete symmetry group of the n-pyramid. This group is isomorphic to D_n.

Figure 16: A regular n-prism. The isomorphic to D_n. The complete symmetry group is $D_n \times \mathbb{Z}_2$. I_3 is not a symmetry for n odd. fig:regprism

2. Consider the regular n-prism. Its base is a regular n-gon and the height is not equal to the length of a side of the n-gon.

The symmetries are generated by rotations about the axis by multiples of $2\pi/n$, reflections in the planes defined by the axes of symmetry of the n-gon times the z-direction, and reflection in the xy plane. For the latter we choose coordinates so that the midpoint of the axis of symmetry is at the origin, and the polygons are parallel to the xy plane.

Note that we have therefore the group $D_n \times \mathbb{Z}_2$ as the complete symmetry group of the prism. The D_n symmetries act in the 2×2 block leaving the z-coordinate fixed,

$$\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} \qquad g \in D_n \subset O(2) \tag{6.22}$$ eq:prsmsymm

while the $\mathbb{Z}_2$ can be taken to be

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{6.23}$$ eq:prsii

The group of *proper* symmetries is in fact D_n, but it is not the "same" D_n subgroup as in (6.22). Rather we take matrices of the form (6.22) for $g \in C_n$, but for g of the form xy^j we multiply by (6.23). Thus we have two nonconjugate copies of D_n in $O(3)$ acting as symmetries.

Note that in the case when n is even, $g = -1_{2\times 2} \in C_n$ and hence the reflection I_3 in all three coordinates is a symmetry.

For n odd, I_3 is *not* a symmetry. The complete group of proper symmetries is D_n. We can generate it by rotations about the axis, together with a product of reflections in the symmetry planes times a reflection in the xy plane. The complete symmetry group is then obtained by adding the reflection in the xy plane. Thus, the symmetry group is again $D_n \times \mathbb{Z}_2$. Notice that this group is isomorphic to D_{2n} (which is how it is presented in tables).

The complete symmetry group of the n-prism – the prismatic symmetry group - is denoted D_{nh} in the Schoenfliess system

6.3 The symmetry group of a baseball

Figure 17: Sketch of a baseball. fig:baseball

As a final example, let us consider the symmetry group of a baseball, drawn in 17. Note that we have symmetry elements acting as transformations in $O(3)$:

$$R = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{6.24}$$

and

$$P = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.25}$$

Check that this gives group isomorphic to D_4 since it can be presented using generators and relations:

$$\langle R, P : R^4 = 1, P^2 = 1, PRP^{-1} = R^{-1} \rangle \tag{6.26}$$

Note that this is a little tricky: R is an improper transformation but it is of order 4. This is the group D_{2d} in the Schoenfliess notation.

As we will see later, it is a direct sum of two irreducible representations of D_4.

7. The symmetries of the platonic solids

A *regular polyhedron* is a solid whose boundaries are regular polygons such that at each vertex, the nearest neighbor vertices form a regular polygon. There are 5 regular polyhedrons in three dimensions, much loved and studied since the time of the ancient Greek mathematicians. These are known as the platonic solids.

The 5 platonic solids are the tetrahedron, cube, octahedron, icosahedron, and dodecahedron.

The number of vertices V, edges E, and faces F is given by:

	Tetrahedron	Cube	Octahedron	Dodecahedron	Icosahedron
F	4	6	8	12	20
E	6	12	12	30	30
V	4	8	6	20	12

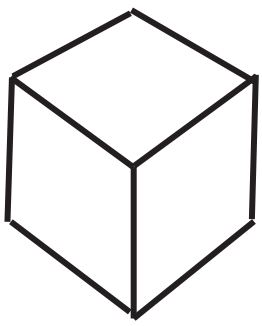


Figure 18: A cube viewed down an axis through antipodal vertices has 3-fold symmetry. fig:cubeone

Notice that $V - E + F = 2$, in each case, in accord with Euler's theorem. The boundary surfaces of these polyhedrons are all topologically spheres.

A convex solid has a dual given by the convex hull of the centers of its faces. Duality exchanges vertices and faces and preserves the number of edges. The tetrahedron is self-dual. The octahedron and cube are dual. The icosahedron and dodecahedron are dual. Dual solids have the same symmetry group.

When investigating the symmetries of the platonic solids it is helpful to keep in mind the theorem mentioned above that every rotation in $SO(3)$ is a rotation about some axis.

7.1 The cube ("hexahedron") and octahedron

In order to investigate the symmetries of the solids (and to construct them) it is easiest to begin with the cube 18.

If we draw the cube projected on the plane through an axis running through antipodal vertices then we get 18. The isotropy group of P is a discrete subgroup of $O(2)$ which is plainly seen to be D_3, since the figure is preserved by reflection and rotations of order 3.

Similarly, we can rotate by π through axes running through antipodal edge midpoints.

Finally we have 4-fold symmetry axes running through parallel faces.

Let us now account for the rotational symmetries of the cube:

There are

1. $V/2 = 4$ distinct axes of 3-fold symmetry yielding 8 distinct elements not equal to the identity.

2. $E/2 = 6$ axes through antipodal edge-midpoints of 2-fold symmetry yielding 6 distinct elements not equal to the identity.

3. $F/2 = 3$ axes through face centers of 4-fold symmetry yielding 9 distinct elements not equal to the identity.

Together with the identity we get a group of order

$$2 \times 4 + 1 \times 6 + 3 \times 3 + 1 = 24 \tag{7.1}$$

called the *octahedral group*, and denoted O. Since parity is a symmetry the full symmetry group of the cube is the group of order 48:

$$O_h = O \cup PO \tag{7.2}$$

eq:cubesym

Let us check this using orbits. The symmetry group of the cube acts transitively on the vertices. The isotropy group of a vertex is the group $D_3 \cong S_3$. There are 8 vertices so

$$|G|/|H| = 8 \tag{7.3}$$

so $|G| = 6 \times 8 = 48$, as expected.

Let us now understand the structure of the group O. We can label the vertices 1 to 8 so O_h is clearly a subgroup of S_8, but it is a rather small subgroup, consisting of only 48 out of $8! = 40320$ permutations. Notice for example that any two antipodal points

Figure 19: A cube with labeled vertices at $(\pm 1, \pm 1, \pm 1)$. fig:cubetwo

must remain antipodal, which would hardly be the case for a general permutation. Indeed, let us label pairs of antipodal points of the cube as A, B, C, D (or equivalently, the 3-fold symmetry axes as AA, BB, CC, DD). Any permutation of the axes can be implemented by some proper symmetry of the cube and hence we can identify the octahedral group as a subgroup of the symmetric group on 4 letters:

$$O \cong S_4 \tag{7.4}$$

As a check note that they both have order 24.

Since parity commutes with these rotations, the complete symmetry group of the octahedron, or cube, is

$$O_h \cong S_4 \times \mathbb{Z}_2 \tag{7.5}$$

and is a direct product.

Another point of view can be useful. Since the cube may be realized as the convex hull of the set of points

$$(\epsilon_1, \epsilon_2, \epsilon_3) \tag{7.6}$$

with $\epsilon_i = \pm 1$, as shown in 19, the group of orthogonal transformations preserving the cube is given by

$$(x_1, x_2, x_3) \to (\epsilon_1 x_{\sigma(1)}, \epsilon_2 x_{\sigma(2)}, \epsilon_3 x_{\sigma(3)}) \tag{7.7}$$

This exhibits the group as a semidirect product of S_3 with $(\mathbb{Z}_2)^3$.

Exercise

Identify each of the conjugacy classes of S_4 with symmetry operations on the cube.

Answer:

1. Rotations around 2-fold axes through edge midpoints; (ab)
2. Rotations by $\pi/2$ around a face: $(abcd)$
3. Rotation by π around a face: $(ac)(bd)$
4. Rotations on 3-fold axes: (abc).

Exercise *O as the symmetries of the octahedron*

a.) Show that the dual octahedron is the convex hull of the 6 vertices $(\epsilon_1, 0, 0), (0, \epsilon_2, 0), (0, 0, \epsilon_3)$.

b.) Show that there are

$V/2 = 3$ axes of 4-fold symmetry

$E/2 = 6$ axes of 2-fold symmetry

$F/2 = 4$ axes of 3-fold symmetry

c.) Label the 4 pairs of parallel faces of the octahedron A, B, C, D and identify the permutation associated with the different group elements.

7.2 Tetrahedron

We can obtain the tetrahedral symmetry group by embedding the tetrahedron in the cube. This can be done in two distinct ways:

Figure 20: Two embeddings of the tetrahedron in the cube. fig:embtet

One embedding has vertices (x_1, x_2, x_3) whose product is -1 and the other has vertices whose product is $+1$.

The complete symmetry group of the tetrahedron is thus the subgroup of O_h preserving these vertices. It follows immediately that

$$|T_d| = 24 \tag{7.8}$$

Let us consider the subgroup T generated by rotations. The group T of order 12 is known as the *tetrahedral group*. To account for the elements of T we have

1. $V = F = 4$ axes through vertices, or equivalently, face-centers, with 3-fold symmetry, yielding 8 distinct elements not equal to the identity.

2. $E/2 = 3$ axes through edge-midpoints of 2-fold symmetry yielding 3 distinct elements not equal to the identity.

Together with the identity these give 12 elements:

$$1 + 4 \times 2 + 3 \times 1 = 12 \tag{7.9}$$

Now, recall that $O \cong S_4$, so T is a subgroup of the symmetric group S_4 of order 12. By considering the map between the symmetries of the cube and elements of S_4 (see above exercise) we see that

1. Rotations through an axis of the tetrahedron passing through a vertex and an opposite face are rotations through a diagonal of the cube, and hence correspond to the conjugacy class of the shape (abc).

2. Rotations through edge-midpoints are rotations of 180 degrees through an axis of the cube passing through opposite faces. These correspond to the conjugacy class of type $(ac)(bd)$. Thus, we have the *even* permutations:

$$T \cong A_4 \tag{7.10}$$

Now, reflection in a plane that bisects the tetrahedron acts to permute just two of the vertices, so this gives the permutations of type (ab). Hence, the complete symmetry group of the tetrahedron is S_4. It is denoted T_d in the Schoenfliess system.

Note that the tetrahedron is the only one of the platonic solids which does not have inversion symmetry, so that T_d is not a direct product $T \times \mathbb{Z}_2$, a fact nicely encoded in the fact that $S_4 \neq A_4 \times \mathbb{Z}_2$.

Exercise

Show that although $O \cong S_4$ and $T_d \cong S_4$, these are not "the same" S_4. That is, show that T_d is not a subgroup of O, but only of O_h. Describe the embedding of T_d in O_h.

Thus, T_d and O are nonconjugate subgroups of $O(3)$, but both are isomorphic to S_4. They provide inequivalent 3-dimensional representations of S_4.

7.3 The icosahedron

The icosahedron is the regular solid with 20 triangular sides.

We can construct the icosahedron as follows.

Label alternate sides of the octahedron black and white and orient the edges so that black is on the left as we move along the arrow.

Divide the edges in the ratio given by the golden section. [14] Thus, the edge from $(0,1,0)$ to $(0,0,1)$ is $(0,a,b)$ with $a+b=1$ and we choose the point with

$$\frac{a}{b} = \frac{\sqrt{5}-1}{2} \tag{7.11}$$

The twelve points

$$(0,\pm a,\pm b), \qquad (\pm b,0,\pm a), \qquad (\pm a,\pm b,0) \tag{7.12}$$

form the vertices of the icosahedron.

[14] Recall: φ, the golden section, or golden ratio, is the ratio a/b defined by $\frac{a+b}{a} = \frac{a}{b}$. Thus $\varphi = \frac{1+\sqrt{5}}{2}$. Note that $\varphi^{-1} = \frac{\sqrt{5}-1}{2}$

Let us consider the rotational symmetries.

1. There are $V/2 = 6$ axes through antipodal vertices with 5-fold symmetry yielding 24 elements.

2. There are $E/2 = 15$ axes through antipodal edge-midpoints with 2-fold symmetry yielding 15 elements

3. There are $F/2 = 10$ axes through face centers with 3-fold symmetry yielding 20 elements.

Together with the identity these form the icosahedral group Y of

$$1 + 6 \times 4 + 15 \times 1 + 10 \times 2 = 60 \tag{7.13}$$

elements. It is a subgroup of $SO(3)$.

If we consider the complete symmetry group of the icosahedron then we need only include inversion $I_3 : \vec{x} \to -\vec{x}$ to get the complete symmetry group of the icosahedron:

$$Y_h = Y \cup PY \cong Y \times \mathbb{Z}_2 \tag{7.14}$$

of order 120.

Again we can check this with orbits: G is transitive on the vertices. The isotropy group of a vertex is isomorphic to a discrete subgroup of $O(2)$ isomorphic to D_5. There are 12 vertices so $|Y_h| = 120$, and we have accounted for all elements.

Figure 21: There are 5 triplets of dyad axes. Each triplet of axes is labelled AA,BB,CC,DD,EE. fig:tetrafive

Now, to describe the group structure of Y we focus on the notion of a dyad axis. This is an axis through antipodal edge midpoints. Using a model you can convince yourself that one can form an orthogonal coordinate system using 3 sets of dyad axes. This partitions up the $30/2 = 15$ dyad axes into 5 sets of 3 mutually orthogonal dyad axes. Let us call these 5 coordinate systems $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5$. Every symmetry in Y permutes the S_i and no symmetry in Y fixes all the $\mathcal{S}_i$, so Y is isomorphic to a subgroup of S_5. The only 60 element subgroup of S_5 is A_5, and hence we conclude

$$Y \cong A_5 \tag{7.15}$$

Remark Consider the $6 = 12/2$ *pentad axes* through antipodal vertices marked $\infty, 0, 1, 2, 3, 4$ in the figure. Interpreting these points as elements of the projective plane $\mathbb{PF}_5$ the sym-

metries of the icosahedron map to the action of $PSL(2, \mathbb{F}_5)$ on this plane by Mobius transformations.

Exercise
Is Y transitive on the vertices?

Exercise
Show that a $2\pi/5$ rotation about an axis passing through antipodal vertices maps to a cycle $(abcde)$ in A_5.

Exercise *The dodecahedron*
a.) Show that the Dodecahedron is the convex hull of the set of points

$$(0, \pm\varphi^{-1}, \pm\varphi), (\pm\varphi^{-1}, \pm\varphi, 0), (\pm\varphi, 0, \pm\varphi^{-1}), (\pm 1, \pm 1, \pm 1) \tag{7.16}$$

b.) Describe Y and Y_h as the group of symmmetries of the dodecahedron.

7.4 No more regular polyhedra

Suppose a regular polyhedron of p-gons have q faces meeting at vertices.

At each vertex of a regular p-gon the tangent vector makes a turn by angle $2\pi/p$. Therefore the internal angles of the p-gon is

$$(1 - \frac{2}{p})\pi \tag{7.17}$$

Now, in our regular polyhedron there will be q faces meeting at a vertex, so the sum of the angles around the vertex must be less than 2π. Therefore

$$\frac{1}{p} + \frac{1}{q} > \frac{1}{2} \tag{7.18}$$

There are only 5 solutions to this Diophantine inequality with $p, q > 2$, i.e. $p, q \geq 3$. Since $\frac{1}{6} + \frac{1}{3} = \frac{1}{2}$ we must have $p, q \leq 5$. It is now easy to enumerate the cases and find that out of the 9 possible pairs we must have:

$$(p, q) = (3, 3), (3, 4), (4, 3), (3, 5), (5, 3) \tag{7.19}$$

7.5 Remarks on the platonic solids

The platonic solids and their symmetries arise in many fields of human endeavor, as well as in Nature. We just give some indications here, with references to the literature.

7.5.1 Mathematics

1. For a description of the subgroups and conjugacy classes of the icosahedral group see Kramer and Haase, "Group Theory of Icosahedral Quasicrystals."

For a nice description of the icosahedral group and its representations see

2. J. Conway, *Monsters and Moonshine*, Math. Intell. **2**(1980) 165. Has a nice description of many different ways of looking at the icosahedral group.

More about the relation to finite fields and some amusing applications are in

3. B. Kostant, "Structure of the truncated icosahedron (such as fullerene or viral coatings) and a 60-element conjugacy class in $PSL(2,11)$," Proc. Natl. Acad. Sci. USA, Vol. 91, pp. 11714-11717, 1994.

4. The exceptional outer automorphism of S_6 is related to the icasahedron, as described in Joyner, sec. 8.6.

There is a rich literature on convex polytopes and their symmetry. See
1.P. du Val, *Homographies, Quaternions, and Rotations*
2. K. Lamotke, *Regular Solids and Isolated Singularities*
3. H.S.M. Coxeter, *Regular Polytopes*
4. H.S.M. Coxeter, *Regular Complex Polytopes*

7.5.2 History of Physics

In Plato's cosmology the four elements earth, air, fire and water were associated with four of the 5 platonic solids, and the fifth, the dodecahedron, corresponded to the heavens.

In Kepler's time there were only 6 known planets Mercury, Venus, Earth, Mars, Jupiter, and Saturn. Note that considering Earth to be a planet requires a heliocentric ordering of the solar system. By inscribing and circumscribing spheres on the 5 platonic solids Kepler wanted to derive the ratios of the radii of the orbits. The approximate success of this idea led him to support the Copernican theory. While this idea was obviously misguided, the broader principle - that there is an underlying beautiful mathematical basis of the laws of nature is still with us today. Moreover, Kepler's passionate pursuit of this idea led to his careful analysis of the orbits of planets, and hence the discovery that they are actually ellipsoidal, and not circular, and to the third law of planetary motion.

7.5.3 Molecular physics

Many molecules in nature have the symmetries we have discussed. For example, benzene C_6H_6 has dihedral symmetry, while CH_4, CCl_4 have tetrahedral symmetry. Uranium hexaflouride UF_6 has octahectral symmetry. These symmetries have implications for the spectrum of vibrations of these molecules.

See:
5. S. Sternberg, *Group Theory and Physics*

6. R.M. Hanson, *Molecular Origami*. This contains cutout models of many compounds in nature.

Recently there has been much excitement over the discovery of "buckyballs," or buckminsterfullerene. The molecule C_{60} is arranged in an icosahedron.

Some references:

8. Dresselhaus, et. al. *Science of Fullerenes and Carbon Nanotubes*

9. Kroto, Heath, et. al. Nature **318**(1985) 354

10. Fowler and Manolopoulos, *An Atlas of Fullerenes*

11. Battye and Sutcliffe, "Solitonic fullerene structures in light atomic nuclei," hep-th/0012215.

7.5.4 Condensed Matter Physics

As we will see below, crystals can have the symmetries T_d and O_h, but it turns out they cannot have icosahedral symmetry. Nevertheless, there is an interesting notion of icosahedral order.

A good reference is the book by David Nelson, "Defects and Geometry in Condensed Matter Physics." View online at

http://books.google.com

7.5.5 Mathematical Physics

1. There are some remarkable occurances of such structures in solitons in nonlinear sigma models in 3+1 dimensions. See, for example: Atiyah and Sutcliffe, arXiv:math-ph/0303071

2. The symmetries of the platonic solids also play and amusing role in the classification of unitary CFT's with $c < 1$: See Friedan, Qiu, and Shenker; Ginsparg; Dijkgraaf, Verlinde, Verlinde.

3. These symmetries also arise in algebraic geometry - classifying degenerations of $K3$ surfaces and simple singularities and have played an important role in string theory.

7.5.6 Biology

The regular polyhedra also occur in biology. For example Radiolaria species have shells in the form of regular polyhedra, e.g. *Circogonia icosahedra*, and *Circorrhegma dodecahedra*.

See the wonderful book:

Ernst Haeckel, *Kunstformen der Nature*: Look at the first radiolarian.

Available on-line at

http://caliban.mpiz-koeln.mpg.de/ stueber/haeckel/kunstformen/natur.html

http://www.flickr.com/photos/origomi/sets/72157601323433758/

A little surfing on the internet will teach you that the majority of viruses, including HIV, have protein coatings in the form of regular polyhedra, especially the icosahedron. [15]

See also:

https://simonsfoundation.org/multimedia/symmetric-structures/

[15] A good place to start is the Wikipedia article "Regular polyhedron."

7.5.7 Human culture: Architecture, art, music and sports

Regular solids have inspired architecture from the pyramids to the geodesic domes of Buckminster Fuller.

According to Weyl, in his book *Symmetry*, Princeton University Press, 1952, 4-fold symmetry in architecture is very common, while 5-fold symmetry is very rare, with the Pentagon in Washington D.C. providing a notable example. (As Weyl wryly notes, it thus provides an easy target for bombers, as indeed proved to be the case on 9/11/2001.) He goes on to say

"Leonardo da Vinci engaged in systematicaly determining the possible symmetries of a central building and how to attach chapels and niches without destroying the symmetry of the nucleus. In abstract modern terminology, his result is essentially [the classification of discrete subgroups of $O(2)$]."

The platonic solids also played a role in musical theory, according to Plato following on Pythagoras.

The platonic solids have appeared in artwork throughout the ages. Sometimes in subtle form. In the attached reprint of Albrecht Dürer's *Melancholia* you see a truncation of a platonic solid.

And let us not forget sports. Icosahedral dice were used in the Ptolemaic dynasty of ancient Egypt. If we truncate the icosahedron we get a soccar ball.

References:

1. Paul Calter, http://www.dartmouth.edu/ matc/math5.geometry/unit6/unit6.html#Plato

7.6 Regular polytopes in higher dimensions

An n-dimensional polytope is a generalization of the notion of polygon and polyhedron. It is a region of $\mathbb{R}^n$ bounded by $(n-1)$-dimensional polytopes.

A regular n-dimensional polytope can be defined recursively by requiring that its boundary is a union of regular $(n-1)$-dimensional polytopes, and its vertex-figure is a regular $(n-1)$-dimensional polytope. The vertex-figure is the $(n-1)$-dimensional polytope obtained by intersecting an $(n-1)$-dimensional hyperplane near an extremal vertex.

Ludwig Schläfli discovered that there are 6 regular 4-polytopes (or "polychorons") in 4-dimensions.

It turns out that in $n \geq 5$ dimensions there are only 3 regular polytopes:

1. The n-simplex, which is self-dual.
2. The n-dimensional cube.
3. The dual of the n-dimensional cube.

References:

1. Coxeter's book.
2. Wikipedia articles on Regular Polytopes.

8. Classification of the Discrete subgroups of $SO(3)$ and $O(3)$

Now we will prove that the symmetry groups above exhaust the discrete subgroups of

$SO(3)$. Of course, when classifying the discrete subgroups we understand two subgroups to be equivalent if they are conjugate.

Theorem The discrete subgroups of $SO(3)$ are

1. C_n, $n \geq 1$
2. D_n, $n \geq 1$
3. Symmetries of the tetrahedron, T
4. Symmetries of the cube = symmetries of the octahedron, O
5. Symmetries of the dodecahedron = symmetries of the icosahedron, Y

Note the similarity to the list of finite simple groups.

Figure 22: An element g of a finite subgroup $G \subset SO(3)$ fixes some point, and permutes other points in a G-orbit. fig:platonic

Proof: Any discrete subgroup of $SO(3)$ preserves the sphere, and any nontrivial element in the group is a rotation about an axis and therefore fixes two points on the sphere. Let X be the set of points on the sphere fixed by some nontrivial element of the finite subgroup $G \subset SO(3)$.

We begin by counting the elements of the set

$$\{(x, g) \in X \times G | g \neq 1 \qquad and \qquad gx = x\} \tag{8.1}$$ eq:countset

We count in two different ways. First, we project $(x, g) \to g$. The fiber of this map consists of two points because every element $g \neq 1$ fixes precisely two points. Denoting $|G| = n$, it is clear that this number is

$$2(n-1) \tag{8.2}$$

On the other hand, we can proceed by counting the number of group elements fixing each x. The points x will form orbits O_i, $i = 1, \ldots, r$ under G. Suppose $|O_i| = p_i$ and the isotropy group of a point in O_i is of order n_i. (Recall that the isotropy group of two points on the same orbit are conjugate subgroups, so this is well-defined.) Note that

$$p_i = n/n_i \tag{8.3}$$ eq:orbtord

by the stabilizer-orbit theorem. Thus (8.1) has

$$\sum_{i=1}^{r}(n_i-1)p_i = n\sum_{i=1}^{r}(1-\frac{1}{n_i}) \tag{8.4}$$ eq:otherfs

points, and therefore we obtain

$$2(1-\frac{1}{n}) = \sum_{i=1}^{r}(1-\frac{1}{n_i}) = r - \sum \frac{1}{n_i} \tag{8.5}$$ eq:relst

Now we analyze this Diophantine equation.

First, note that $r = 2$ or $r = 3$. To see this we first prove that it is impossible to have $n_i = 1$. Indeed, if $n_i = 1$ then for that x no $g \neq 1$ fixes x, but this is impossible by the definition of X. Therefore, all $n_i > 1$ and hence $\sum \frac{1}{n_i} < r/2$ so the RHS of (8.5) is $\geq r - \frac{1}{2}r = r/2$, but the LHS is less than 2. Thus, $r/2 < 2$ and hence $r = 1, 2, 3$. On the other hand, we cannot have $r = 1$: Note that we can assume $n > 1$, and hence the LHS is ≥ 1, but comparing to the RHS we see that this implies $r > 1$.

Now we examine the two cases $r = 2$ and $r = 3$:

$r = 2$: We rewrite the equation as

$$0 = n_1(n - n_2) + n_2(n - n_1) \tag{8.6}$$

Now $n_i \leq n$ so both terms are nonnegative. Since $n_1 > 1$ and $n_2 > 1$ it must be that $n_1 = n_2 = n$. Thus, X consists of two antipodal points. Each point is an orbit fixed by the entire group. These are the cyclic groups C_n acting by rotation around some axis.

Now consider $r = 3$. Then (8.5) simplifies to

$$\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} = 1 + \frac{2}{n} \tag{8.7}$$ eq:reslii

WLOG we assume $n_1 \leq n_2 \leq n_3$. Then the complete list of solutions is:

	n	(n_1, n_2, n_3)	p_1	p_2	p_3
D_k	$2k$	$(2, 2, k)$	k	k	2
T	12	$(2, 3, 3)$	6	4	4
O	24	$(2, 3, 4)$	12	8	6
I	60	$(2, 3, 5)$	30	20	12

This is easily proved as follows: We must have $n_1 = 2$, for if $n_1 > 2$ then the LHS of (8.7) is ≤ 1. Similarly, we cannot have $n_2 \geq 4$, since again the LHS (8.7) is ≤ 1. Therefore,

$n_2 = 2, 3$. If $n_2 = 2$ the equation becomes $n = 2n_3$. This is the first row. We have two orbits of length k and one orbit of length 2. Embedding the dihedral group D_k into $SO(3)$ by multiplying parity transformations of D_k with reflection in the x, y plane clearly gives such a solution.

If $n_2 = 3$ then

$$\frac{1}{n_3} = \frac{1}{6}(1 + \frac{12}{n}) \tag{8.8}$$

and hence $1/n_3 > 1/6$, so $n_3 < 6$. Now we simply plug in $n_3 = 3, 4, 5$.

Now we should show that each solution indeed corresponds to a discrete subgroup of $SO(3)$. But this follows from our discussion of the regular solids. ♠

It is interesting to work out the orbits on the sphere corresponding to the different symmetry groups of the regular solids.

Consider C_n embedded as rotations by $2\pi j/n$ around some axis. The set X indeed consists of two points, where the axis intersects the sphere, and the isotropy group at each point clearly has order n. This is the solution $r = 2$.

Now, for D_n we must embed it in $SO(3)$. We do this, as in the prismatic symmetry group by embedding

$$y^j \to \begin{pmatrix} R(2\pi j/n) & 0 \\ 0 & 1 \end{pmatrix} \tag{8.9}$$

$$x \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{8.10}$$

For the elements y^j the north and south poles are fixed points with isotropy group of order n. Note that x maps one to another, so they form an orbit of length two. The fixed points of xy^j necessarily lie in the $z = 0$ plane. We can find them by considering the circle to circumscribe the regular n-gon. Then consider the reflection axes of the n-gon. Each such vertex intersects the sphere in 2 points. The intersections of these axes with the sphere describe two regular n-gons. These are the other two orbits.

Next note that the last three entries nicely reproduce the number of faces, edges, and vertices of the platonic solids.

Consider, for example, the tetrahedral group T. If we embed a tetrahedron in the sphere, then the vertices form one orbit of length four. The inverted tetrahedron forms another orbit of length four. Indeed, the four 3-fold symmetry axes pierce the sphere in 2 sets of 4 points. These are the two orbits with $p_2 = p_3 = 4$. The 3 2-fold symmetry axes through edge midpoints pierce the sphere in 6 points. These form the orbit of $p_1 = 6$.

The other cases are similar. (See Lamotke's book for details.)

In conclusion we list without proof the generators and relations of the discrete subgroups of $SO(3)$: [16]

$$C_n = \langle \gamma | = \gamma^n = 1 \rangle \tag{8.11}$$

[16] See, e.g. Lamotke's book above.

$$D_n = \langle \beta, \gamma | \beta^2 = \gamma^n = (\beta\gamma)^2 = 1 \rangle \tag{8.12}$$

$$T = \langle \beta, \gamma | \beta^3 = \gamma^3 = (\beta\gamma)^2 = 1 \rangle \tag{8.13}$$

$$O = \langle \beta, \gamma | \beta^3 = \gamma^4 = (\beta\gamma)^2 = 1 \rangle \tag{8.14}$$

$$Y = \langle \beta, \gamma | \beta^3 = \gamma^5 = (\beta\gamma)^2 = 1 \rangle \tag{8.15}$$

Exercise
Verify the nontrivial relation on β, γ in each case above.

8.1 Finite subgroups of $O(3)$

Now we consider $O(3) = SO(3) \times \mathbb{Z}_2$.

We can reduce the classification of discrete subgroups of $O(3)$ to those of $SO(3)$ using the following construction.

Consider the determinant homomorphism $\det : O(3) \to \mathbb{Z}_2$. Suppose $G \subset O(3)$ is a discrete subgroup. If $\det : G \to \mathbb{Z}_2$ is trivial, then $G \subset SO(3)$, and we have already classified these. On the other hand, if $\det : G \to \mathbb{Z}_2$ is a nontrivial homomorphism (i.e. there are parity-reversing elements in G) then there is a kernel $K \subset G$ of order 2. It is a normal subgroup, and

$$G = K \amalg K g_0 \tag{8.16}$$

for some $g_0 \in G$. The elements of Kg_0 are rotation-inversions. We have an exact sequence

$$1 \to K \to G \to \mathbb{Z}_2 \tag{8.17}$$ eq:GtoK

If $I_3 : \vec{x} \to -\vec{x}$ is in G then we can simply take $g_0 = I_3$ and

$$G = K \amalg K I_3 \tag{8.18}$$

Note that I_3 is in the center of $O(3)$ and hence $G \cong K \times \mathbb{Z}_2$. The sequence (8.17) splits.

Now, it can happen that $I_3 \notin G$. For the most trivial example, consider a solid which is completely irregular except for a reflection plane of symmetry. (For example, the exterior of an idealized human being.) In this case $K = 1$ and $g_0 \neq I_3$. For more nontrivial examples consider the n-prism with n odd or the complete symmetry group of the tetrahedron, T_d.

If $I_3 \notin G$ then we can construct

$$G^+ := K \amalg K g_0 I_3. \tag{8.19}$$

Note that $K \cap K g_0 I_3 = \emptyset$ for otherwise there would be $k, k' \in K$ with $k = k' g_0 P$ which implies $k'' = g_0 I_3$ which implies $I_3 \in G$, since $g_0 \in G$, contrary to hypothesis. Moreover, we claim that G^+ is in fact a group. To verify this note that if $k_1, k_2 \in K$ then

$$k_1 g_0 I_3 k_2 g_0 I_3 = k_1 g_0 k_2 g_0 \tag{8.20}$$

but $k_1 g_0 k_2 g_0 \in G$, and yet it is a proper rotation, so it must be in K.

Moreover note that this calculation also shows that, as abstract groups, G^+ is isomorphic to G via the map

$$\begin{array}{lll} \psi: & g \to g & g \in K \\ & g \to I_3 g & g \notin K \end{array} \tag{8.21}$$

Warning: G^+ is a group of proper rotations, $G^+ \subset SO(3)$, but is distinct from G as a subgroup of $O(3)$. That is, there are inequivalent embeddings of G into $O(3)$.

Example: Consider the group denoted S_{2n} in the Schoenfliess classification. It is generated by

$$g_0 = \begin{pmatrix} R(\pi/n) & 0 \\ 0 & -1 \end{pmatrix} \tag{8.22}$$

Note that it is clearly isomorphic to C_{2n}. However, for n even $I_3 \notin S_{2n}$. Note that $K \cong C_n$ is generated by g_0^2. Moreover,

$$I_3 g_0 = \begin{pmatrix} R(\frac{n+1}{2n} 2\pi) & 0 \\ 0 & +1 \end{pmatrix} \tag{8.23}$$

generates $G^+ \cong C_{2n}$.

Now, using this construction we can construct the finite subgroups of $O(3)$. In addition to the finite subgroups of $SO(3)$ we have already listed, those with rotation-inversions are:

1. $C_n \amalg I_3 C_n$
2. $D_n \amalg I_3 D_n$
3. $T_h \cong T \amalg I_3 T$
4. 4. $O_h \cong O \amalg I_3 O$
5. $Y_h \cong Y \amalg I_3 Y$
6. $G^+ = C_{2n}, K = C_n$
7. $G^+ = D_n, K = C_n$
8. $G^+ = D_{2n}, K = D_n$
9. $G^+ = O, K = T$

To prove this we look for homomorphisms G^+ *onto* $\mathbb{Z}_2$ where G^+ runs over the discrete subgroups of $SO(3)$. Any element mapping to the nontrivial generator must have order two. C_n only has an order two element for n even, this gives $G^+ = C_{2n}$. For D_n we can use the reflection, or, if n is even a rotation by π. This gives the next two examples. Now note that $T \cong A_4$ and $Y \cong A_5$ are simple groups, so cannot have nontrivial homomorphic images. This leaves $G^+ = O$, which has only one possibility since S^4 has only one index two subgroup. If we think of O as the group which takes $(x_1, x_2, x_3) \to (\epsilon_1 x_{\sigma(1)}, \epsilon_2 x_{\sigma(2)}, \epsilon_3 x_{\sigma(3)})$ then the homomorphism is $g \to \epsilon_1\epsilon_2\epsilon_3$.

The following is a table of the finite subgroups of $O(3)$ which are not in $SO(3)$, together with examples of solids for which they are the complete symmetry group.

G^+	K	Schoenfliess	Solid
$C_n \amalg I_3 C_n$	C_n	S_{2n}, n odd	alternating $2n$-prism
$C_n \amalg I_3 C_n$	C_n	C_{nh}, n even	shaved n-prism
$D_n \amalg I_3 D_n$	D_n	D_{nh}, n even	n-prism
$D_n \amalg I_3 D_n$	D_n	D_{nd}, n odd	twisted n-prism
$T \amalg I_3 T$	T	T_h	decorated cube
$O \amalg I_3 O$	O	O_h	cube, octahedron
$Y \amalg I_3 Y$	Y	Y_h	dodecahedron, icosahedron
C_{2n}	C_n	S_{2n}, n even	alternating $2n$-prism
C_{2n}	C_n	C_{nh}, n odd	shaved n-prism
D_n	C_n	C_{nv}, $n \geq 2$	n-pyramid
D_{2n}	D_n	D_{nh}, n odd	n-prism
D_{2n}	D_n	D_{nd}, n even	twisted n-prism
O	T	T_d	tetrahedron

Note that several of the entries under G^+ are isomorphic to each other. However, the groups with different names in the Schoenfless classification are nonconjugate as subgroups of $O(3)$, despite being abstractly isomorphic as groups.

The twisted n-prism is obtained by gluing together two n-prisms rotated by π/n around their axis of symmetry. If we shave off the edges of an n-prism it receives and orientation and we get a shaved n-prism. If on a $2n$-prism we shave every other edge, we get an alternating n-prism. For pictures see pp. 96-97 of the book by Yale.

8.2 Finite subgroups of $SU(2)$ and $SL(2,\mathbb{C})$

In the next chapter we will study the rich relation between $SO(3), SU(2)$, and $SL(2,\mathbb{C})$. The finite subgroups are closely related (by a $2:1$ homomorphism) to the above. They have generators:

1. $\mathcal{C}_n$

$$\begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix} \qquad \omega = e^{\pi i/n} \tag{8.24}$$

2. $\mathcal{D}_{n+2}$: "binary dihedral"

$$\begin{pmatrix} \beta & 0 \\ 0 & \beta^{-1} \end{pmatrix} \qquad \beta = e^{\pi i/n}, \qquad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{8.25}$$

3. $\mathcal{T}$, "binary tetrahedral," is generated by $\mathcal{D}_4$ together with:

$$\frac{1}{\sqrt{2}} \begin{pmatrix} \epsilon^{-1} & \epsilon^{-1} \\ \epsilon^5 & \epsilon \end{pmatrix} \qquad \epsilon = e^{2\pi i/8} \tag{8.26}$$

4. $\mathcal{O}$, is generated by $\mathcal{T}$ together with:

$$\begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix} \qquad \epsilon = e^{2\pi i/8} \tag{8.27}$$

5. $\mathcal{I}$, is generated by:

$$-\begin{pmatrix} \eta^3 & 0 \\ 0 & \eta^2 \end{pmatrix}, \frac{1}{\eta^2 - \eta^3} \begin{pmatrix} \eta + \eta^4 & 1 \\ 1 & -\eta - \eta^4 \end{pmatrix} \qquad \eta = e^{2\pi i/5} \tag{8.28}$$

9. Symmetries of lattices

9.1 The crystallographic restriction theorem

The *crystallographic restriction theorem* is an important restriction on the possible subgroups of $SO(2)$ or $SO(3)$ which can be automorphisms of 2 and 3-dimensional lattices, respectively.

Let $g \in SO(2)$ be a symmetry of a two-dimensional lattice. This generates a discrete subgroup of $SO(2)$, which must be $\cong C_n$ corresponding to rotations by $k \cdot \frac{2\pi}{n}$.

In an *integral basis* for the lattice the action of $g \in SO(2)$ must be an *integral matrix*:

$$g : \begin{pmatrix} \vec{e}_1 \\ \vec{e}_2 \end{pmatrix} \to \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \vec{e}_1 \\ \vec{e}_2 \end{pmatrix} \tag{9.1}$$

with $a, b, c, d \in \mathbb{Z}$.

On the other hand, in a *Cartesian basis* the transformation g must be represented by a rotation:

$$R(g) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{9.2}$$ eq:rotgee

On the other hand, the trace is independent of basis, so (9.2) must have an integral trace. Therefore:

$$2\cos\theta \in \mathbb{Z} \tag{9.3}$$ eq:lattcond

This is only possible if $\cos\theta = 0, \pm\frac{1}{2}, \pm 1$.

$\cos\theta$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1
$-\pi < \theta \leq \pi$	π	$\pm\frac{2\pi}{3}$	$\pm\frac{\pi}{2}$	$\pm\frac{\pi}{3}$	0

In other words, $\theta = \pm\frac{2\pi}{n}$ with

$$n = 1, 2, 3, 4, \quad \text{or} \quad 6. \tag{9.4}$$ eq:twodee

A very similar argument applies to three-dimensional lattices. Once again we choose a lattice basis e_i. Then a point-group symmetry must take

$$g \cdot e_i = \sum_j C_{ji} e_j \tag{9.5}$$

where C_{ji} is a matrix of integers. On the other hand, g is conjugate to

$$\pm \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{9.6}$$

so

$$\mathrm{Tr}(g) = \pm(1 + 2\cos\phi) \in \mathbb{Z} \tag{9.7}$$

Once again we conclude that $2\cos\phi \in \mathbb{Z}$, and hence we again have the list (9.4).

9.2 Three-dimensional lattices

The symmetries of 3-dimensional lattices is an important subject in condensed matter physics. We summarize here the bare essentials:

Definition A subgroup of $E(3)$ mapping a three-dimensional lattice to itself and fixing a point of the lattice is called a *crystallographic point group*.

Because of the crystallographic restriction theorem there are only 32 possible crystallographic point groups. It is useful to separate those which contain I_3 from those which do not:

1. C_n, $n = 1, 2, 3, 4, 6$
2. D_n, $n = 2, 3, 4, 6$
3. T, O, *but not* Y!
4. C_{nh}, $n = 1, 3$
5. C_{nv}, $n = 2, 3, 4, 6$
6. D_{3h}
7. D_{2d}
8. T_d

9. S_n, $n = 2, 4, 6$
10. C_{nh}, $n = 2, 4, 6$
11. D_{nh}, $n = 2, 4, 6$
12. D_{nd} $n = 3$
13. T_h, O_h

Only the groups in 9-13, consisting of 11 groups, contain I_3.

Definition A *crystallographic space group* is a discrete subgroup $G \subset E(3)$ such that

$$L = G \cap \mathcal{T} \tag{9.8}$$

is a three-dimensional lattice group, i.e., isomorphic to $\mathbb{Z}^3$. Physically these classify lattice systems with a basis and are therefore symmetries of (idealized) crystals.

The question of how to classify crystallographic space groups is an intricate one. One's first guess is that we should classify crystallographic subgroups $G \subset E(3)$ up to conjugacy.

However, this leaves infinitely many parameters such as the lengths of basis vectors and the angles between them.

The standard way to classify space groups is via conjugacy within the group of all *affine transformations* of $\mathbb{R}^3$,

$$\vec{x} \to S\vec{x} + \vec{a} \tag{9.9}$$ eq:affinetmn

where $S \in GL(3, \mathbb{R})$.

The result of Fedorov, Schoenflies and Barrow from the 1890's is that:

1. There are 219 distinct isomorphism classes of space groups,

2. If we restrict the classification to conjugacy within the group of *proper* transformations, with $\det S > 0$ then there are 230 distinct classes of space groups. There are 11 conjugacy classes which are related by improper isometries, and hence crystallographers recognize 230 distinct space groups. In these 11 pairs it turns out that the conjugation taking one to the other changes the chirality of a screw displacement, but the chirality of the screw displacement does affect the physical properties of the crystal.

The lattice subgroup $L \subset G$ is a normal subgroup, and if an element $g \in G$ is $g = (\vec{a}, R)$ then conjugation by g acts as a group of automorphisms of L:

$$T_{\vec{a}} \in L \to g T_{\vec{a}} g^{-1} = T_{R\cdot\vec{a}} \tag{9.10}$$

Thus, $P = G/L$ acts a discrete subgroup of $O(3)$ on L.

Now, we have classified the discrete subgroups of $O(3)$, and we can use this information.

Definition The largest crystallographic point group at a point $\vec{x} \in L$ is called the *holohedry* of L at $\vec{x}$. Two lattices are in the same *crystal system* if their holohedries are conjugate subgroups of $E(3)$.

Lattices clearly have I_3 as a symmetry. Using this one can show that I_3 is always a member of the holohedry F. This shows that many of the crystallagraphic point groups cannot be holohedries. Using some detailed arguments (See, e.g. Miller's book or Sternberg, Appendix A) one can then show that of the 11 point groups containing I_3, only 7 can be holohedries.

There are only 7 crystal systems with holohedries S_2, C_{2h}, D_{3d}, O_h and D_{nh} for $n = 2, 4, 6$. These satisfy the subgroup relations

$$\begin{array}{ccccccc} S_2 & \subset & C_{2h} & \subset & D_{2h} & \subset & D_{4h} \subset O_h \\ & & \cap & & \cap & & \\ & & D_{3d} & \subset & D_{6h} & & \end{array} \tag{9.11}$$

Crystal System	Holohedry
triclinic	S_2
monoclinic	C_{2h}
orthorhombic	D_{2h}
tetragonal	D_{4h}
cubic	O_h
trigonal	D_{3d}
hexagonal	D_{6h}

If we now list the possible pairs (P, L) where P is a point group which is the maximal symmetry of L. The equivalence relation is that there exists an $S \in GL(3, \mathbb{R})$ with $\det S > 0$ such that $L' = SL$ and $P' = SPS^{-1}$. It turns out that the 7 crystal systems decompose into 14 "lattice types," known as the "Bravais lattices." [17]

For example, the generic lattice only has $P = S_2$ as its symmetry group. These are called the *triclinic* lattices.

We will not go through the detailed arguments here. See the references for that. Once we have the possible pairs (P, L) we can try to classify the possible crystallographic groups by studying the possible extensions:

$$1 \to L \to G \to P \to 1 \tag{9.12}$$ eq:extnesion

Not all of the extensions are split. The groups for which the sequence splits are called *symmorphic*.

Remarks:

1. The analogous question for 2-dimensional lattices leads to the 17 *wallpaper groups* which are discrete subgroups of $E(2)$ containing two independent translations. See the Wikipedia article http://en.wikipedia.org/wiki/Wallpaper-group for an account.

2. Hilbert's 18th problem asked if there are finitely many crystallographic space groups in higher dimensions. Bieberbach's theorem answered this in the affirmative.

References:

1. *Symmetry in Condensed Matter Physics*, L. Michel

2. Hahn, Theo, ed. (2002), International Tables for Crystallography, Volume A: Space Group Symmetry (5th ed.), Springer-Verlag

3. S. Sternberg, Group Theory and Physics

4. Miller, Symmetry Groups and Their Applications.

5. Wikipedia article on Space Groups

9.3 Quasicrystals

In particular, note that 5-fold symmetry of 2- and 3-dimensional lattices is forbidden. This is why the discovery of quasicrystals was such a shock! Certain materials have x-ray diffraction patterns with 5-fold symmetry.

For a reference on a group-theoretical approach to some quasicrystals see P. Kramer and R.W. Haase, "Group Theory of Icosahedral Quasicrystals," in *Introduction to the Mathematics of Quasicrystals* ed. by M. V. Jarić.

One way to produce almost periodic structures with symmetries unavailable for lattices is to *project* from a higher-dimensional lattice with forbidden symmetry.

[17] In 1835 Frankenheim published a paper listing 15 crystal types. In 1848, Bravais pointed out that two of them are isomorphic.

Example. Choose $\vec{v} = (v_1, 0; v_2, 0) \in \mathbb{R}^{2,2}$ and a rotation

$$\theta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix} \tag{9.13}$$ eq:thetr

This can be diagonalized to $\theta = Diag\{\xi, \xi^{11}, \xi^5, \xi^7\}$ with $\xi = e^{2\pi i/12}$. If

$$\begin{aligned} \langle \vec{v}, \vec{v} \rangle &= v_1^2 - v_2^2 = 2 \\ \langle \vec{v}, \theta \vec{v} \rangle &= \frac{\sqrt{3}}{2}(v_1^2 + v_2^2) = 2 \end{aligned} \tag{9.14}$$

then $\vec{v}, \theta\vec{v}, \theta^2\vec{v}, \theta^3\vec{v}$ generate a lattice in $\mathbb{R}^{2,2}$ with $\mathbb{Z}_{12}$ symmetry. Projecting to the positive definite 2-dimensional planes gives an aperiodic array with $\mathbb{Z}_{12}$ symmetry.

10. Tesselations by Triangles

11. Diversion: The Rubik's cube group

Reference: Joyner, *Adventures in Group Theory*

The Rubik's cube group G is generated by 90 degree rotations of the face slices. Note that rotations of the middle slice is equivalent to rotations of the two parallel face slices together with a rotation in space. Now notice that any rotation by a face slice leaves the central cube fixed. Thus, if the puzzle-solver is holding the cube we can describe the faces by U (top) D (bottom) F (facing the holder), B, (facing away from the holder), L, held by the left-hand and R held by the right-hand.

If one is looking down on a face the basic move is a 90 degree rotation in the *clockwise* direction. This defines six generators of the Rubik's cube group U, D, F, B, L, R.

These permute the $54 - 6 = 48$ remaining facets which are not centers of faces. Thus, the Rubik's cube group $\mathcal{R}$ is a subgroup of S_{48}. Now, the order of S_{48} is $48! \cong 1.24 \times 10^{61}$, and there are only about $\pi \times 10^7$ seconds in a year, so a little discretion is called for in solving the cube.

In fact, as with the symmetry groups of the platonic solids, we can cut down the group of potential transformations somewhat. Note that $\mathcal{R}$ separately permutes the 8 corner pieces (accounting for $8 \times 3 = 24$ facets) and the 12 edge pieces (accounting for $12 \times 2 = 24$ facets). In addition the facets have an orientation, the edge pieces have two possible orientations (since the painted faces always point outwards) and the corner pieces have three possible orientations. Therefore, the group $\mathcal{R}$ is a certain subgroup of

$$H = (C_3^{|V|} \ltimes S_V) \times (C_2^{|E|} \ltimes S_E) \tag{11.1}$$

Here V is the set of 8 vertex cubes and E is the set of 12 edge cubes. The factors C_3 and C_2 take care of the orientations of these vertex and edge cubes. $S_V \cong S_8$ is the group permuting the vertex cubes while S_E permutes the edge cubes. In the semidirect product $C_3^{|V|} \ltimes S_V$, S_V permutes the 8 factors of C_3.

The group H here is known as the *illegal Rubik's cube group* in Joyner's book. It is the group of permutations of the facets obtained by allowing one to take apart the cube and reassemble it, but one must keep the painted faces outward.

It is quite clear that $\mathcal{R}$ is a proper subgroup of H. Indeed, consider the basic move on a face. It only permutes the edge and corner pieces of that face which is being rotated. There is a relation between the permutation of the edge and face pieces: they are both cyclic permutations of the kind (1234). In particular, if $r \in S_V$ and $s \in S_E$ then

$$\epsilon(r) = \epsilon(s). \tag{11.2}$$

As for the orientation, choose a fixed orientation of the cube using some ordering of the central faces. For example, white side up, red side forward. Now in the solved position give each cube position a standard orientation, and each cube piece its intrinsic orientation. Comparing these two, for edge pieces we get an element of C_2 and for corner pieces and

element of C_3. Any configuration of the Rubik's cube defines an element $\vec{v} \in C_3^8$ and $\vec{w} \in C_2^{12}$. Note that the basic 90-degree rotation move preserves

$$\sum w_i = 0 \text{mod} 2 \qquad \sum v_i = 0 \text{mod} 3. \tag{11.3}$$

The nontrivial theorem in Joyner's book states that these are the only relations:

Theorem Denote an element of H by $(\vec{v}, r, \vec{w}, s)$ with $\vec{v} \in C_3^8, \vec{w} \in C_2^{12}, r \in S_8, s \in S_{12}$. Then the Rubik's cube group is the subgroup consisting of elements such that

1. $\epsilon(r) = \epsilon(s)$
2. $\sum v_i = 0 \text{mod} 3$
3. $\sum w_i = 0 \text{mod} 2$

From this one can deduce that the order of $\mathcal{R}$ is

$$\begin{aligned} \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} \times 3^8 \times 8! \times 2^{12} \times 12! &= 2^{27} \cdot 3^{14} \cdot 5^3 \cdot 7^2 \cdot 11 \\ &= 43252003274489856000 \\ &\cong 4.325 \times 10^{19} \end{aligned} \tag{11.4}$$

Exercise
Consider a 2×2 Rubik's cube. What is the Rubik's cube group?

---

12. Group actions on function spaces

Suppose X, Y are two sets. We can then consider the space of maps

$$\mathcal{F}_{X,Y} := \{f : X \to Y\} \tag{12.1}$$

Now, if X or Y admits a G-action for a group G then it follows that the space of maps $\mathcal{F}(X, Y)$ also admits a G-action.

Figure 23: If G acts on the target Y then we simply define $g \cdot f$ on functions $X \to Y$ by post-composing with the G action. fig:gactfunone

If $\varphi : G \times Y \to Y$ is a G action then we define a G-action $\Phi : G \times \mathcal{F} \to \mathcal{F}$ (abbreviating $\mathcal{F} = \mathcal{F}_{X,Y}$) by

$$\Phi(g, f)(x) := \varphi(g, f(x)) \tag{12.2}$$

Figure 24: If G acts on the target X then we simply define $g \cdot f$ on functions $X \to Y$ by pre-composing with the G action. fig:gactfuntwo

as in 23. One checks that if φ is a left(right)-action then Φ is also a left(right)-action.

Similarly, if X admits a G-action $\varphi : G \times X \to X$ then $\mathcal{F}$ admits a G action $\Phi : G \times \mathcal{F} \to \mathcal{F}$ defined by

$$\Phi(g, f)(x) := f(\varphi(g, x)) \tag{12.3}$$

as in 24. Now one checks that if φ is a left, resp. right-action on X the Φ is a right, resp. left-action on $\mathcal{F}$.

If we want to restrict attention to left-actions then we can say that

he function space $\mathcal{F}_{X,Y}$ also has a left-G-action defined by $f \mapsto g \cdot f$ where the new function $g \cdot f$ is defined by the values: $(g$

Exercise

a.) Show that

$$g_1 \cdot (g_2 \cdot f) = (g_1 g_2) \cdot f \tag{12.4}$$

and hence defines a left group action.

b.) Show that

$$(f \cdot g)(x) := f(g \cdot x) \tag{12.5}$$

defines a *right*-action of G on functions.

Example 1. $SO(2, \mathbb{R})$ acts on $\mathbb{R}^2$

$$R_\phi^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} \tag{12.6}$$

By our basic principle, $G = SO(2, \mathbb{R})$ also acts on the space of all functions

$$\mathcal{F} = \{f : \mathbb{R}^2 \to \mathbb{R}\} \tag{12.7}$$

$f \mapsto (R_\phi f)$ where the new function $(R_\phi f)$ is defined by the values $(R_\phi f)(x_1, x_2) \equiv f(x_1', x_2')$ and The association $f \to R_\phi f$ is an action of the group on $\mathcal{F}$.

The G-orbits in the space of functions are collections of functions which can be rotated into one another.

Exercise
a.) What is the image of the function $(x_1)^3$ under R_ϕ ?

Some orbits of $\mathcal{F}_{X,Y}$ are distinguished. These are the G-orbits consisting of one point, or, in plain English, the *invariant functions* In example 1 above these are simply the radially symmetric functions.

$$f(x^1, x^2) = f(r) \tag{12.8}$$

Remarks

1. The above principle is frequently used in quantum mechanics. The wavefunction $\psi : X \to \mathbb{C}$ where X is some configuration space of a dynamical system. If X admits a symmetry group then that group acts on the space of wavefunctions.

13. The simple singularities in two dimensions

There is a purely algebraic approach to geometry: Characterize a variety by the kinds of polynomial functions that make sense on it.

For $\mathbb{C}^2/\Gamma$, where Γ is a discrete subgroup of $SL(2,\mathbb{C})$ these varieties are very beautiful and important both in math and in string theory.

Example Consider $\mathbb{C}^2$ with the action of $\mathbb{Z}_k$:

$$(z_1, z_2) \to (\omega z_1, \omega^{-1} z_2) \tag{13.1}$$

where ω is a primitive k^{th} root of 1. This acts on the ring of polynomials $\mathbb{C}[z_1, z_2]$. *What is the ring of invariant polynomials?*

There are three obvious invariant polynomials:

$$X = z_1^k \qquad Y = z_2^k \qquad Z = z_1 z_2 \tag{13.2}$$

Any polynomial in X, Y, Z is therefore an invariant polynomial. Note that X, Y, Z are not independent, but satisfy

$$XY = Z^k \tag{13.3}$$

eq:akrel

A nontrivial fact is that the ring of invariant polynomials is *generated* by X, Y, Z, subject to the single relation (13.3), that is:

$$\mathbb{C}[z_1, z_2]^G = \mathbb{C}[X, Y, Z]/(XY - Z^k). \tag{13.4}$$

This describes the polynomial functions on the orbifold $\mathbb{C}^2/\mathbb{Z}_k$.

Example Now consider the binary dihedral group $\mathcal{D}_{n+2}$ of order $4n$. This acts on $\mathbb{C}^2$:

$$(z_1, z_2) \to (\beta z_1, \beta^{-1} z_2)$$
$$(z_1, z_2) \to (+z_2, -z_1) \tag{13.5}$$ eq:bindih

where $\beta = e^{i\pi/n}$. Again, let us ask for the ring of invariant polynomials. Some obvious invariants are:

$$Z = (z_1 z_2)^2 \qquad X = z_1^{2n} + z_2^{2n} \qquad Y = (z_1 z_2)(z_1^{2n} - z_2^{2n}) \tag{13.6}$$

Now note that

$$X^2 = z_1^{4n} + 2z_1^{2n} z_2^{2n} + z_2^{4n} \tag{13.7}$$
$$Y^2 = (z_1 z_2)^2 (z_1^{4n} - 2z_1^{2n} z_2^{2n} + z_2^{4n}) \tag{13.8}$$

so, again, these three have a single independent relation

$$ZX^2 - Y^2 = 4Z^{n+1} \tag{13.9}$$ eq:deerel

Again, it is true, but not obvious that all invariant polynomials are generated by X, Y, Z subject to the relation (13.9).

Remarks:

1. Similar, but more difficult arguments give the invariant polynomials $\mathbb{C}[z_1, z_2]^G$ for all the finite subgroups G of $SL(2, \mathbb{C})$. The ring is $\mathbb{C}[X, Y, Z]/I$ where I is the ideal generated by:

$\mathcal{C}_n$	$XY = Z^n$
$\mathcal{D}_{n+2}$	$ZX^2 + Y^2 = Z^{n+1}$
$\mathcal{T}$	$X^2 + Y^3 + Z^4 = 0$
$\mathcal{O}$	$X^2 + Y^3 + YZ^3 = 0$
$\mathcal{I}$	$X^2 + Y^3 + Z^5 = 0$

2. The groups $\mathcal{C}_n, \mathcal{D}_{n+2}, \mathcal{T}, \mathcal{O}, \mathcal{I}$ will turn out to be closely related to the Lie algebras $A_n, D_{n+2}, E_6, E_7, E_8$. The geometry of $\mathbb{C}^2/\Gamma$ has amazing relations to these Lie algebras. This is part of something called the McKay correspondence. We'll give a very rough idea of what happens here. The equation $XY - Z^n = 0$ is an equation for a

complex "surface" in $\mathbb{C}^3$. (That is the solutions space has 2 complex dimensions, or 4 real dimensions.) Note that it is a cone: If (X_0, Y_0, Z_0) is on the surface then so is $(\lambda^n X_0, \lambda^n Y_0, \lambda^2 Z_0)$. So the surface is singular. We can "resolve the singularity" by considering a deformation

$$XY - Z^n = \epsilon \tag{13.10}$$

This equation turns out to have n "independent" 2-dimensional spheres. (It might help to think about the real case, where a deformation of a cone $X^2 - Y^2 - Z^k = 0$ in $\mathbb{R}^3$ by putting ϵ on the RHS gives a smooth space, but the resolution has induced a noncontractible circle, hence nontrivial topology.) These spheres intersect in points within the surface. The pattern of intersection turns out to be the Dynkin diagram of the corresponding Lie group. That's just the beginning...

**** REFERENCES .... *****

Exercise

Describe the ring of invariant polynomials under

$$(z_1, z_2) \to (\omega z_1, \omega^k z_2) \tag{13.11}$$

where $\omega = e^{2\pi i/n}$.

Warning: This is hard.

---

14. Symmetric functions

14.1 Structure of the ring of symmetric functions

One important example of group actions on functions spaces is the theory of symmetric functions.

The permutation group S_n acts on $\mathbb{R}^n$:

$$\sigma \cdot (x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \tag{14.1}$$

Exercise

a.) Show that an orbit is uniquely labelled by a point in the subspace $x^1 \geq x^2 \geq \cdots \geq x^n$. That is, this subspace is a fundamental domain for the action of the symmetric group on $\mathbb{R}^n$.

b.) Compute the isotropy groups which arise.

By our basic principle S_n acts on functions on $\mathbb{R}^n$:

$$(\sigma^{-1}\psi)(x_1,\dots,x_n) = \psi(x_{\sigma(1)},\dots,x_{\sigma(n)}) \tag{14.2}$$ eq:funcs

The invariant functions are the *symmetric functions.*

One way these come up in physics is

1. As wavefunctions of identical bosonic particles. We will discuss this more below.

2. As characters of representations of Lie groups.

We now summarize a few of the main mathematical facts about symmetric functions. A standard reference is:

I.G. MacDonald, *Symmetric Functions and Hall Polynomials* Oxford 1995.

This book has the proofs of the theorems we quote.

Let us consider the ring of polynomials with integer coefficients [18]

$$R = \mathbb{Z}[x_1,\dots,x_n] \tag{14.3}$$

We would like to understand the structure of the ring of invariants

$$\Lambda_n := R^{S_n}. \tag{14.4}$$

What kind of symmetric polynomials can we make? For any polynomial f the sum

$$\sum_{\sigma\in S_n} \sigma\cdot f \tag{14.5}$$

is clearly a symmetric polynomial. Thus, the most obvious thing to do is to apply this construction to monomials. If $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{Z}_+^n$ is a sequence of nonnegative integers, then let

$$x^\alpha := x_1^{\alpha_1}\cdots x_n^{\alpha_n} \tag{14.6}$$

and consider the sum over distinct permutations of α. There will be a unique permutation of α so that it takes the form of a sequence of nonnegative integers $\lambda = (\lambda_1,\dots,\lambda_n)$ such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \tag{14.7}$$

We can label the resulting function m_λ.

Clearly, the m_λ form an integral basis for the symmetric polynomials. However, the symmetric polynomials form a (graded) ring, and to understand this ring it is better to use other functions.

In order to describe these functions let us recall the notion of a *partition* λ. This is a sequence of integers

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \tag{14.8}$$ eq:seg

such that $\lambda_i = 0$ for i larger than some integer. For a partition we can define $|\lambda| := \sum \lambda_i$. The partitions of n we met earlier in chapter 1 are the partitions with $|\lambda| = n$.

The functions m_λ are labeled by partitions.

[18] In fact, we can replace $\mathbb{Z}$ by any ring A.

Figure 25: Young diagrams fig:young

It is traditional to associate a *Young diagram* to partitions $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots)$. These are diagrams of boxes with λ_i boxes in the ith row. They are very useful in representation theory. Some examples are shown in 25. There are of course different conventions used by different authors. These conventions are related by the group D_4.

By reflecting in the diagonal we have a map of partitions $\lambda \to \lambda'$ the conjugate partition. For example the conjugate of (1^s) is $(s, 0, 0, \dots)$, or if $\lambda = (3, 1, 0, \dots)$ then $\lambda' = (2, 1, 1, 0, \dots)$. Formally, we may define

$$\lambda'_i = |\{\lambda_j | \lambda_j \geq i\}| \tag{14.9}$$

Figure 26: Ordering on Young diagrams fig:orderyoung

One can put various orderings on partitions. The only one we need is the partial ordering which says that $\lambda \geq \mu$ if

$$\lambda_1 + \cdots + \lambda_i \geq \mu_1 + \cdots + \mu_i \qquad \forall i \tag{14.10}$$

If in addition, for some i there is strict inequality we write $\lambda > \mu$. Note that this is a partial order. One cannot say that (31^3) is greater or less than (2^3).

Now to get a generating set for the ring of symmetric polynomials we consider the *Elementary symmetric functions* defined by

$$E(t) := \sum_{r=0}^{n} e_r t^r := \prod_{i=1}^{n} (1 + x_i t) \tag{14.11}$$ eq:esym

So, for example:

$$\begin{aligned} e_0 &= 1 \\ e_1 &= x_1 + \cdots + x_n \\ e_2 &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n \\ e_3 &= x_1x_2x_3 + \cdots \\ e_r &= \sum_{1\le i_1<i_2<\cdots<i_r\le n} x_{i_1}\cdots x_{i_r} \end{aligned} \tag{14.12}$$

eq:eet

The first fundamental theorem on symmetric functions is: **Theorem 1**. The ring of invariant polynomials, $\mathbb{Z}[x_1,\ldots,x_n]^{S_n}$, is itself a polynomial ring and

$$\mathbb{Z}[x_1,\ldots,x_n]^{S_n} = \mathbb{Z}[e_1,\ldots,e_n] \tag{14.13}$$

eq:rings

That is, *any* symmetric polynomial in n variables with integer coefficients can be uniquely written as a polynomial, with integer coefficients, in the functions $e_r(x)$.

Example:

$$\sum_{i=1}^{n} x_i^2 = e_1^2 - 2e_2 \tag{14.14}$$

Sketch of Proof: We have seen that the m_λ define an integral basis for the ring. For a partition $\lambda = (\lambda_1 \ge \lambda_2 \ge \cdots)$ define

$$e_\lambda := e_{\lambda_1}e_{\lambda_2}\cdots \tag{14.15}$$

Then the key identity to establish is

$$e_{\lambda'} = m_\lambda + \sum_{\mu<\lambda} a_{\lambda\mu}m_\mu \tag{14.16}$$

eq:keyrel

where $a_{\lambda\mu}$ are nonnegative integers.

To check this let us say $\lambda = (\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > 0 \ge 0\cdots)$ and $\lambda' = (\lambda'_1 \ge \lambda'_2 \ge \cdots \ge \lambda'_s > 0 \ge 0\cdots)$. Note that $r = \lambda'_1$ and $s = \lambda_1$. Similarly, the number of rows in λ' with at least two elements is λ_2, and so forth.

Now imagine evaluating the symmetric polynomial $e_{\lambda'}(x)$ in the regime were $x_1 \gg x_2 \gg \cdots \gg x_n$. Then the leading term in e_λ is

$$(x_1x_2\cdots x_{\lambda'_1}+\cdots)(x_1x_2\cdots x_{\lambda'_2}+\cdots)\cdots(x_1x_2\cdots x_{\lambda'_r}+\cdots) = x_1^{\lambda_1}x_2^{\lambda_2}\cdots x_n^{\lambda_n}+\cdots \tag{14.17}$$

To see this note that the leading power of x_1 is clears x_1^s. But $s = \lambda_1$. This accounts for the power of x_1. Next, the leading power of x_2 is the number of rows in λ' with at least two elements. But this is just λ_2. And so forth.

Moreover the other monomials one obtains from expanding out e_λ are subleading, which indeed means that the other monomials we obtain have leading term μ with $\lambda > \mu$. Thus we have (14.16).

Since this is an upper triangular transformation any symmetric function can be uniquely expressed as a polynomial in the e_λ. Thus they are algebraically independent. ♠

Another natural set of symmetric functions are the *complete symmetric functions*. These are defined by

$$H(t) := \prod_{i=1}^{n} \frac{1}{1 - x_i t} := \sum_{k=0}^{\infty} h_k t^k \tag{14.18}$$ eq:complete

That is, h_k is the sum, with coefficient 1, of all monomials of degree k in the x_i:

$$\begin{aligned} h_0 &= 1 \\ h_1 &= x_1 + \cdots + x_n \\ h_2 &= x_1^2 + \cdots + x_n^2 + x_1 x_2 + \cdots + x_{n-1} x_n \\ &\cdots\cdots \\ h_r &= \sum_{1 \le i_1 \le i_2 \le \cdots \le i_r \le n} x_{i_1} \cdots x_{i_r} \end{aligned} \tag{14.19}$$ eq:hfuncs

Note that

$$H(-t) = 1/E(t) \tag{14.20}$$ eq:htoee

and hence we can express the h_k in terms of the e_k by an upper triangular integral matrix with 1 on the diagonal:

$$\begin{aligned} h_0 &= e_0 \\ h_1 &= e_1 \\ h_2 &= -e_2 + e_1^2 \\ h_3 &= e_3 - 2e_1 e_2 + e_1^3 \\ &\cdots\cdots \end{aligned} \tag{14.21}$$

Therefore we can also say:

$$\mathbb{Z}[x_1, \ldots, x_n]^{S_n} = \mathbb{Z}[h_1, \ldots, h_n] \tag{14.22}$$ eq:ringsii

Conversely, using $E(t) = 1/H(-t)$ we can express the e_i as polynomials in the h_i with integral coefficients.

Another obvious way to make symmetric functions is to define the

Power sum functions:

$$s_r(x) := \sum_{i=1}^{n} x_i^r \tag{14.23}$$

These can be related to the other symmetric functions by considering the generating function

$$P(t) := \sum_{r=1}^{\infty} s_r(x) t^{r-1} = \sum_{i=1}^{n} \frac{x_i}{1 - x_i t} \tag{14.24}$$

Note that

$$P(t) = H'(t)/H(t) \tag{14.25}$$

or, equivalently

$$P(-t) = E'(t)/E(t) \tag{14.26}$$

This implies *Newton's identities*:

$$ne_n = \sum_{r=1}^{n} (-1)^{r-1} s_r e_{n-r} \tag{14.27}$$ eq:newtons

$$nh_n = \sum_{r=1}^{n} s_r h_{n-r} \tag{14.28}$$ eq:newtonsp

The first few give:

$$\begin{aligned}
e_1 &= s_1 \\
e_2 &= \frac{1}{2}(s_1^2 - s_2) \\
e_3 &= \frac{1}{3!}s_1^3 - \frac{1}{2!}s_1 s_2 + \frac{1}{3}s_3 \\
e_4 &= \frac{1}{4!}s_1^4 - \frac{1}{4}s_2 s_1^2 + \frac{1}{8}s_2^2 + \frac{1}{3}s_1 s_3 - \frac{1}{4}s_4 \\
e_5 &= \frac{1}{5!}(s_1^5 - 10 s_1^3 s_2 + 15 s_1 s_2^2 + 20 s_1^2 s_3 - 20 s_2 s_3 - 30 s_1 s_4 + 24 s_5)
\end{aligned} \tag{14.29}$$ eq:newiden

(See the exercise below for an efficient way to generate these polynomials.)
Thus we have proved

Theorem 2. *Any* symmetric polynomial in n variables (with rational coefficients) can be uniquely written as a polynomial in the functions $s_r(x)$. That is:

$$\mathbb{Q}[x_1, \dots, x_n]^{S_n} = \mathbb{Q}[s_1, \dots, s_n] \tag{14.30}$$

Remarks

1. Note that the relations between the h_i, e_i, s_i are *universal*. That is, they do not depend on the number of variables n of which h_i, e_i, s_i are functions. Thus we can, in a sense, take $n \to \infty$ and consider these as functions of infinitely many variables so long as $x_j = 0$ for j sufficiently large. The mathematically precise way to do this involves "injective limits."

2. There are many uses of these formulae. One common use is in discussions of class functions of matrices. Recall that by the Jordan canonical form theorem the conjugacy class of a matrix $A \sim SAS^{-1}$ has a representative given by Jordan canonical form. It is easy to evaluate traces and determinants in this form.

3. Symmetric functions appear very naturally in topology. The elementary symmetric functions e_i are related to Chern classes c_i of vector bundles, while the power functions are related to the Chern characters: $\mathrm{ch}_k \sim \frac{1}{k!}s_k$. Thus, the above identities relate Chern classes and characters of vector bundles.

Exercise
Show that

$$\exp\left(\sum_{r=1}^{\infty} s_r \frac{t^r}{r}\right) = \sum_{k=0}^{\infty} h_k t^k \tag{14.31}$$

$$\exp\left(-\sum_{r=1}^{\infty} s_r \frac{(-t)^r}{r}\right) = \sum_{k=0}^{\infty} e_k t^k \tag{14.32}$$

and thereby derive (14.29).

Exercise
Show that

$$\begin{aligned} s_1 &= e_1 \\ s_2 &= -2e_2 + e_1^2 \\ s_3 &= 3e_3 - 3e_2e_1 + e_1^3 \\ s_4 &= -4e_4 + 4e_3e_1 + 2e_2^2 - 4e_2e_1^2 + e_1^4 \\ s_5 &= 5(e_5 - e_4e_1 - e_2e_3 + e_3e_1^2 + e_2^2e_1 - e_2e_1^3) + e_1^5 \\ s_n &= (-1)^{n+1}ne_n + \cdots + e_1^n \end{aligned} \tag{14.33}$$

eq:chceeii

Exercise
a.) Write the e_i as polynomials in the h_i.
b.) Write the h_i as polynomials in the s_i.

Exercise
Show that any 2×2 matrix satisfies

$$\det A = \frac{1}{2}(\mathrm{Tr}A)^2 - \frac{1}{2}\mathrm{Tr}A^2 \tag{14.34}$$

Show that *any* 3×3 matrix A satisfies:

$$\det A = \frac{1}{3}\mathrm{Tr}A^3 - \frac{1}{2}(\mathrm{Tr}A)(\mathrm{Tr}A^2) + \frac{1}{6}(\mathrm{Tr}A)^3 \tag{14.35}$$

Show that any 4×4 matrix satisfies

$$\det A = \frac{1}{4!}(\mathrm{Tr}A)^4 - \frac{1}{4}(\mathrm{Tr}A^2)(\mathrm{Tr}A)^2 + \frac{1}{8}(\mathrm{Tr}A^2)^2 + \frac{1}{3}(\mathrm{Tr}A)(\mathrm{Tr}A^3) - \frac{1}{4}\mathrm{Tr}A^4 \tag{14.36}$$

And in general the expression for e_n as a polynomial in the s_i gives a relation between the determinant of an $n \times n$ matrix and its traces.

Exercise
Show that the roots $\theta_1, \ldots, \theta_n$ of a monic polynomial

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 \tag{14.37}$$

are related to the coefficients via the elementary symmetric functions

$$a_{n-j} = (-1)^j e_j(\theta_1, \ldots, \theta_n) \tag{14.38}$$

♣ EXERCISE ON EXPRESSING THE COEFFICIENTS IN THE CHARACTERISTIC POLYNOMIAL IN TERMS OF ELEMENTARY SYMMETRIC FUNCTIONS ♣

Exercise *The discriminant of a polynomial*
The discriminant of a polynomial $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots a_0$ with $a_i \in \mathbb{C}$ and $a_n \neq 0$ is defined as

$$a_n^{2n-2} \prod_{i<j} (\theta_i - \theta_j)^2 \tag{14.39}$$

where θ_i are the complex roots.

a.) Noting the relation of a_i to the elementary symmetric functions in the roots, express the discriminant as a polynomial in the coefficients a, b, c, d for the cubic polynomial $ax^3 + bx^2 + cx + d$:

$$b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd \tag{14.40}$$

For $x^3 + px + q$ this simplifies to

$$D = -4p^3 - 27q^2 \tag{14.41}$$

b.) Suppose a, b, c, d are real. Show that $D > 0$ implies there are three real roots and $D < 0$ implies there is one real and two complex roots. If $D = 0$ two roots coincide.

c.) By considering how the functions a_{n-j} scale if we scale the roots $\theta_j \to t\theta_j$ show that the monomial $a_0^{p_0}\cdots a_n^{p_n}$ can only appear in the discriminant if

$$n(n-1) = np_0 + (n-1)p_1 + \cdots + p_{n-1} \tag{14.42}$$

In fact, there is an explicit formula for the discriminant in terms of the resultant of $f(x)$ and $f'(x)$. In general the resultant of two polynomials $f(x) = a_n x^n + \cdots + a_0$ and $g(x) = b_m x^m + \cdots + b_0$ is defined as the determinant of the $(n+m)\times(n+m)$ Sylvester matrix

♣Clarify the structure of this matrix ♣

$$R(f,g) := \det \begin{pmatrix} a_n & a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 & 0 & 0 & \cdots & 0 \\ 0 & a_n & a_{n-1} & \cdots & a_2 & a_1 & a_0 & 0 & \cdots & 0 \\ 0 & 0 & a_n & \cdots & a_3 & a_2 & a_1 & a_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ b_m & b_{m-1} & b_{m-2} & \cdots & 1\cdot b_0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & b_m & b_{m-1} & b_{m-2} & \cdots & b_0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & b_m & b_{m-1} & b_{m-2} & \cdots & b_0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & b_m & b_{m-1} & b_{m-2} & \cdots & \cdots & b_0 \end{pmatrix} \tag{14.43}$$

The discriminant is proportional to $R(f,f')$.

14.2 Free fermions on a circle and Schur functions

Figure 27: Fermi sea and particle/hole excitations. fig:Fermisea

Consider a system of N free fermions on a circle $z = e^{i\theta}$, $\theta \sim \theta + 2\pi$, with Hamiltonian $H = -\sum_i \frac{d^2}{d\theta_i^2}$. The one-particle wavefunctions are

$$\psi_n(z) = z^n \tag{14.44}$$ eq:onpart

with energy n^2. Consider a state of N fermions occupying levels n_i. By Fermi statistics these levels must all be distinct and we can assume

$$n_N > n_{N-1} > \cdots > n_1 \tag{14.45}$$

The N-fermion wavefunction is the "Slater determinant"

$$\Psi_{\vec{n}}(z_1,\dots,z_N) = \det_{1\le i,j\le N} z_i^{n_j} \tag{14.46}$$ eq:slater

and has energy $\sum_i n_i^2$. States can be visualized as in 27.

For simplicity let us assume N is odd. Then there is a unique groundstate obtained by filling up the states $-n_F \le n \le n_F$ with

$$n_F = \frac{N-1}{2} \tag{14.47}$$ eq:fermil

(If N is even there are two groundstates.) This defines the Fermi sea. The ground state wavefunction is

$$\Psi_{gnd}(z_1,\dots,z_N) = \det_{1\le i,j\le N} z_i^{j-1-n_F} = (z_1\dots z_N)^{-n_F}\Delta_0(z) \tag{14.48}$$ eq:grnd

The ratio of the wavefunction to groundstate wavefunction

$$\frac{\Psi_{\vec{n}}}{\Psi_{gnd}} = \frac{\det_{1\le i,j\le N} z_i^{n_j}}{\det_{1\le i,j\le N} z_i^{j-1-n_F}} \tag{14.49}$$ eq:srat

is a totally symmetric function of the z_i. If $n_1 \ge -n_F$ then it is a symmetric polynomial known as a *Schur function*.

The Schur functions form a linear basis for $\mathbb{Z}[x_1,\dots,x_N]^{S_N}$ where basis elements are associated with partitions. To define them mathematically let us return to a monomial $x^\alpha = x_1^{\alpha_1}\cdots x_N^{\alpha_N}$ with the $\alpha_i \ge 0$ and now try to skew-symmetrize:

$$\sum_{\sigma\in S_N} \epsilon(\sigma)\sigma\cdot x^\alpha \tag{14.50}$$

Note that this function is totally antisymmetric, and moreover vanishes unless the α_i are all distinct. WLOG assume that

$$\alpha_N > \alpha_{N-1} > \cdots > \alpha_1 \ge 0. \tag{14.51}$$

The smallest such α is

$$N-1 > N-2 > \cdots > 1 > 0 \ge 0 \tag{14.52}$$

defining a vector $\delta := (N-1, N-2, \dots, 1, 0)$ i.e. $\delta_j = j-1$, $1 \le j \le N$. Using δ we define ν by

$$\alpha = \nu + \delta \tag{14.53}$$

Note that $\nu_j = \alpha_j - \delta_j \ge 0$ and moreover

$$\nu_j + j - 1 > \nu_{j-1} + j - 2 \qquad \Rightarrow \qquad \nu_j \ge \nu_{j-1} \tag{14.54}$$

Thus, ν is a partition. It is more convenient to use ν than α so we define:

$$\Delta_\nu(x) := \sum_{\sigma\in S_N} \epsilon(\sigma)\sigma\cdot x^{\nu+\delta} \tag{14.55}$$

Note that this can also be written as a determinant:

$$\begin{aligned}\Delta_\nu(x) &:= \det(x_i^{\nu_j+N-j}) \\ &= \det\begin{pmatrix} x_1^{\nu_1} & x_1^{\nu_2+1} & \cdots & x_1^{\nu_N+N-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_N^{\nu_1} & x_N^{\nu_2+1} & \cdots & x_N^{\nu_N+N-1} \end{pmatrix}\end{aligned} \tag{14.56}$$ eq:deltsun

Note that $\Delta_0(x)$ is the famous Vandermonde determinant:

$$\Delta_0(x) := \prod_{i>j}(x_i - x_j) \tag{14.57}$$

Exercise
Show that

$$\Delta_\nu(\sigma \cdot x) = \epsilon(\sigma)\Delta_\nu(x) \tag{14.58}$$

where $\epsilon(\sigma)$ is the sign homomorphism.

Now we define the *Schur function* to be the ratio:

$$\Phi_\nu(x) := \frac{\Delta_\nu(x)}{\Delta_0(x)} \tag{14.59}$$ eq:schurf

This is a totally symmetric function. In fact, it is a polynomial in the x_i. To see this consider it as a meromorphic function of a complex variable x_1. Note that it is in fact an entire function since the potential poles at $x_1 = x_i$ are cancelled by zeroes of the numerator. Next note that the growth at infinity is obviously x_1^m for an integer m.

Example. Put $N = 2$. Then $\nu_2 \geq \nu_1 \geq 0$,

$$\Delta_\nu(x) = (x_1x_2)^{\nu_1}(x_2^{\nu_2-\nu_1+1} - x_1^{\nu_2-\nu_1+1}) \tag{14.60}$$ eq:deltatwo

$$\Phi_\nu(x_1, x_2) = (x_1x_2)^{\nu_1}\frac{x_2^{\nu_2-\nu_1+1} - x_1^{\nu_2-\nu_1+1}}{x_2 - x_1} = (x_1x_2)^{\nu_2}h_{\nu_2-\nu_1}(x_1, x_2) \tag{14.61}$$ eq:phitwo

Returning to the free fermion example, we have

$$\frac{\Psi_{\vec{n}}}{\Psi_{\vec{0}}} = \frac{\Delta_\nu(z)}{\Delta_0(z)} \tag{14.62}$$ eq:fration

for $n_j + n_F = \nu_j + j - 1$, that is

$$\nu_j = n_j - (j-1) - n_F \tag{14.63}$$ eq:snet

Remarks

1. The second main theorem of symmetric polynomials is:

Theorem. The Schur functions $\Phi_\nu(x)$ form a linear integral basis for $\mathbb{Z}[x_1, \dots, x_N]^{S_N}$.

That is, any symmetric polynomial with integral coefficients can be written as a linear combination, with integral coefficients, of the $\Phi_\nu(x)$. Note that we are *not* forming polynomials in the Φ_ν.

2. The module A_N of skew-symmetric polynomials in $x_1, \dots, x_N$ is isomorphic to Λ_N via multiplication by Δ_0. Therefore the Δ_ν form a a $\mathbb{Z}$-basis for A_N. This close relation between completely symmetric and antisymmetric functions comes up in the theory of matrix models – integrals over space of $N \times N$ matrices. It also suggests a relation between bosons and fermions, at least in $1+1$ dimensions. That indeed proves to be the case – there is a nontrivial isomorphism known as *bosonization* which is an isomorphism of quantum field theories of bosons and fermions in 1+1 dimensions.

3. There is an elegant expression for the Schur functions Φ_ν as a determinant of a matrix whose entries involve the h_k. The expression is

$$\Phi_\nu = \det(h_{\nu_i - i + j}) \tag{14.64}$$ eq:phiach

where we take an $N \times N$ matrix for N variables and it is understood that $h_0 = 1$ and $h_i = 0$ for $i < 0$. Equivalently

$$\Phi_\nu = \det(e_{\nu'_i - i + j}) \tag{14.65}$$ eq:phiachp

For a proof see Macdonald, p.41.

Example Consider the partition $(\nu_1, \nu_2, 0, 0, \dots)$. Then applying 14.64 we get

$$\begin{aligned} \Phi_\nu &= \begin{pmatrix} h_{\nu_1} & h_{\nu_1+1} & h_{\nu_1+2} & h_{\nu_1+3} & h_{\nu_1+4} & \cdots \\ h_{\nu_2-1} & h_{\nu_2} & h_{\nu_2+1} & h_{\nu_2+2} & h_{\nu_2+3} & \cdots \\ 0 & 0 & 1 & h_1 & h_2 & \cdots \\ 0 & 0 & 0 & 1 & h_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \\ &= h_{\nu_1} h_{\nu_2} - h_{\nu_1+1} h_{\nu_2-1} \end{aligned} \tag{14.66}$$ eq:phitwoa

With a little algebra one can confirm that this is indeed (14.61). Note, however, that both sides of this equality make sense for $N > 2$.

4. Define the $\mathbb{Z} \times \mathbb{N}$ matrix Ξ whose $(p,q)^{th}$ entry is h_{q-p} where $h_0 = 1$ and $h_k = 0$ for $k < 0$. Here $-\infty < p < \infty$ labels the rows, while $q = 1, 2, \dots$ labels the columns. Thus the matrix looks like 11.

Figure 28: A matrix for defining Schur functions. fig:schrmtx

Define the integers $s_k = k - \nu_k$. These eventually become $s_k = k$. Then Φ_ν is the determinant of the matrix formed from Ξ by keeping only the rows labeled by s_k. This leads to an upper triangular matrix is the determinant of the finite matrix above.

5. The definition of Φ_ν only makes sense as a function of N such that $\nu_i = 0$ for $i > N$. However, the relations between Φ_ν and the h_i are universal.

6. We will return to Schur functions after we have defined a few terms in representation theory.

Exercise
Verify that (14.66) and (14.61) are equal for $N = 2$.

Exercise
a.) Let $\nu = (n, 0, 0, 0, \dots)$. Show that $\Phi_\nu = h_n$.
b.) Let $\nu = (1^n, 0, 0, \dots)$. Then $\Phi_\nu = e_n$.

Exercise
Using Cauchy's determinant identity:

$$\det\left(\frac{1}{1 - x_i y_j}\right) = \frac{\Delta_0(x_i)\Delta_0(y_j)}{\prod_{i,j}(1 - x_i y_j)} \tag{14.67}$$

Show that

$$\exp\left[\sum_k \frac{1}{k} s_k(x) s_k(y)\right] = \sum_\nu \Phi_\nu(x)\Phi_\nu(y) \tag{14.68}$$

Hint: Multiply both sides by $\Delta_0(x)\Delta_0(y)$. For the answer see Hammermesh, p. 195

Exercise *Laughlin's wavefunctions*

Another natural place where symmetric wavefunctions appear is in the quantum Hall effect, where the 1-body wavefunctions in the lowest Landau level are

$$\psi_n(z) = \frac{1}{\sqrt{\pi n!}} z^n e^{-\frac{1}{2}|z|^2} \tag{14.69}$$

Laughlin introduced a fascinating set of approximate eigenfunctions of 2d *interacting* electrons in a magnetic field:

$$\Psi = \prod_{i<j} (z_i - z_j)^{2n+1} e^{-\frac{1}{2}\sum |z_i|^2} \tag{14.70}$$

Express these in terms of Schur functions.

Many other interesting trial wavefunctions in the FQHE can be generated using theorems about symmetric functions.

****

FOLLOWING MATERIAL SHOULD BE MOVED TO CHAPTER ON REPRESENTATIONS. IT WOULD MAKE MORE SENSE TO TALK ABOUT REPS OF THE SYMMETRIC GROUP FIRST.

14.2.1 Schur functions, characters, and Schur-Weyl duality

Let V be the N-dimensional representation of $U(n)$. Consider the tensor spaces $V^{\otimes N}$. These are both representations of the symmetric group S_N as well as representations of $U(n)$. Then a beautiful result, *Schur-Weyl duality* states that

The relation between the Φ_ν and the power functions $s_k(x)$ is very nice, and important for representation theory. Let $(\ell) = (\ell_1, \ell_2, \ldots, \ell_N)$ denote a tuple of nonnegative integers. Then let

$$s_{(\ell)} := s_1^{\ell_1} s_2^{\ell_2} \ldots s_N^{\ell_N} \tag{14.71}$$

Then we have the important result of Frobenius:

$$\boxed{s_{(\ell)} = \sum_{\{\nu\}} \chi_\nu(\ell)\Phi_\nu} \tag{14.72}$$ eq:frobenius

We will prove this later as a consequence of Schur-Weyl duality. It will turn out that the coefficients $\chi_\nu(\ell)$ in this expansion are the characters of the representations of the symmetric group S_N in the conjugacy class labelled by $(1)^{\ell_1}(2)^{\ell_2}\cdots$. We can restrict to $\sum j\ell_j = N$ and $\sum \nu_j = N$ here.

3. Later we will view the $\Phi_\nu(x)$ as characters of representations of $U(N)$. Briefly, by Schur-Weyl duality we can label the representations of $U(N)$ by Young diagrams. Then for a given representation, its character is a function only of the conjugacy class of the matrix. Restricting to the maximal torus $Diag\{x_1, \ldots, x_N\}$ the character is the Schur polynomial in the x_i. (This is the Weyl character formula.)

Remark: If we think of Φ_ν as a character of a representation of $U(N)$. Then the Hamiltonian H is closely related to the quadratic Casimir of that representation. See Cordes et. al. eq. (4.3)

Exercise *Weyl dimension formula*
Show that

$$\Phi_\nu(1, 1, \ldots, 1) = \prod_{1\le i<j\le N} \frac{(\nu_i - \nu_j + j - i)}{(j-i)} \tag{14.73}$$ eq:weyldim

We will see later that this is a special case of the Weyl dimension formula.
Hint: Put $x_j = e^{jt}$ and let $t \to 0$.

14.3 Bosons and Fermions in 1+1 dimensions

14.3.1 Bosonization

Finally, we want to describe a truly remarkable phenomenon, that of bosonization in 1+1 dimensions.

Certain quantum field theories of bosons are equivalent to quantum field theories of fermions in 1+1 dimensions! Early versions of this idea go back to Jordan [19] The subject became important in the 1970's. Two important references are

S. Coleman, Phys. Rev. **D11**(1975)2088

S. Mandelstam, Phys. Rev. **D11** (1975)3026

The technique has broad generalizations, and plays an important role in string theory.

To get some rough idea of how this might be so, let us consider the *loop group* $S^1 \to U(1)$. For winding number zero, loop group elements can be written as:

$$z \to g(z) = \exp[i \sum_{n=-\infty}^{+\infty} j_n z^n] \tag{14.74}$$

with $j_n^* = j_{-n}$. This group acts on the one-body wavefunction by

$$z^n \to g(z) z^n \tag{14.75}$$

where $g(z)$ is in the loop group of $U(1)$.

Under such a change of one-body wavefunctions the Slater determinant changes by:

$$\begin{aligned} \Delta_0(z) &\to \det g(z_i) z_i^{j-1} \\ &= \det \begin{pmatrix} g(z_1) & g(z_1)z_1 & g(z_1)z_1^2 & \cdots \\ g(z_1) & g(z_2)z_2 & g(z_2)z_2^2 & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix} \\ &= \prod g(z_i) \Delta_0(z) \\ &= \exp[i \sum j_n s_n(z)] \Delta_0(z) \end{aligned} \tag{14.76}$$

eq:coherentone

Note that there are two Fermi levels, and if N is large they are "far apart" meaning that operators such as $\sum_i z_i^n$ for small n will not mix states near the two respective levels. Let us therefore imagine taking N large and focussing attention on one of the Fermi levels. Therefore we adjust our energy level so that the groundstate wavefunctions are $1, z^{-1}, z^{-2}, \ldots$ and we imagine that we have taken $N \to \infty$. Moreover, let us extend the $LU(1)$ action to $L\mathbb{C}^*$. Then we can separately consider the action of

$$g_-(z) = \exp[i \sum_{-\infty}^{0} \varphi_n z^n] \tag{14.77}$$

[19] P. Jordan, Z. Phys. **93**(1935)464.

and

$$g_+(z) = \exp[i\sum_1^\infty \varphi_n z^n] \tag{14.78}$$

on the groundstate Ψ_0. Note if we think of the groundstate wavefunction as a Slater determinant then acting with $g_-(z)$ takes one column to a linear combination of lower columns and hence does not change the wavefunction. On the other hand, by (14.76) acting with $g_+(z)$ has a nontrivial action and generates all possible symmetric functions of the z_i.

In this - rather heuristic - sense, the action of the loop group "generates the whole Hilbert space of fermionic states." Moreover, by mapping antisymmetric wavefunctions to symmetric wavefunctions we now view the Hilbert space as the space of polynomials in the s_n. [20]

Now we observe the following. The ring of symmetric functions (extending scalars to $\mathbb{C}$) is the polynomial ring $\mathbb{C}[s_1, s_2, \dots]$. We now make this space into a *Hilbert space*. We introduce the inner product on polynomials in s_i by

$$\langle f(s)|g(s)\rangle := \int \prod_{k=1}^\infty \frac{ds_k \wedge d\bar{s}_k}{2\pi i k} \overline{f(s)} g(s) e^{-\sum_{k=1}^\infty \frac{1}{k}|s_k|^2} \tag{14.79}$$

eq:innerpf

This is the coherent state representation of an infinite system of harmonic oscillators:

$$[a_k, a_j] = k\delta_{k+j,0} \qquad\qquad -\infty < j,k < \infty \tag{14.80}$$

with $a_{-j} = a_j^\dagger$. These are realized as follows: $a_k^\dagger$ is multiplication by s_k and a_k is $k\frac{\partial}{\partial s_k}$. The state

$$(a_1^\dagger)^{\ell_1}(a_2^\dagger)^{\ell_2}\cdots(a_N^\dagger)^{\ell_N}|0\rangle \tag{14.81}$$

corresponds to the symmetric function $s_{(\ell)}$.

Now one can form the quantum field:

$$\phi := i\sum_{k\neq 0} \frac{1}{k} a_k z^{-k} \tag{14.82}$$

Then the coherent states

$$|\phi(z)\rangle = \exp i\oint \phi(z) j(z)|0\rangle \tag{14.83}$$

give an action of the loopgroup on the bosonic Fock vacuum acting as

$$\exp[i\sum_{n=1}^\infty j_n a_n^\dagger]|0\rangle \tag{14.84}$$

eq:coherenttwo

thus producing a nontrivial isomorphism between a bosonic and fermionic Fock space.

[20] This statement is rather loose. See the Pressley-Segal book for a mathematically precise treatment. We are discussing an irreducible representation of a centrally extended group.

To make the isomorphism between bosonic and fermionic Fock spaces more explicit we introduce a second-quantized formalism. B^+_{-n} creates a fermionic state with wavefunction z^n, B_n annihilates it, so we introduce:

$$\begin{aligned}\Psi(\theta) &= \sum_{n\in\mathbb{Z}} B_n e^{in\theta} \\ \Psi^\dagger(\theta) &= \sum_{n\in\mathbb{Z}} B^+_{-n} e^{-in\theta}.\end{aligned} \tag{14.85}$$

eq:fermflds

The filled Fermi sea satisfies the constraints:

$$\begin{aligned}B_n|0\rangle &= 0, \qquad |n| > n_F \\ B^+_{-n}|0\rangle &= 0, \qquad |n| \le n_F.\end{aligned} \tag{14.86}$$

eq:fifer

When we have decoupled systems it is appropriate to define two independent sets of Fermi fields:

$$\begin{aligned}\Psi(\theta) &= e^{i(n_F+\frac{1}{2})\theta} b(\theta) + e^{-i(n_F+\frac{1}{2})\theta}\bar{b}(\theta) \\ \Psi^\dagger(\theta) &= e^{-i(n_F+\frac{1}{2})\theta} c(\theta) + e^{i(n_F+\frac{1}{2})\theta}\bar{c}(\theta).\end{aligned} \tag{14.87}$$

eq:bcsys

We introduce complex coordinates $z = e^{i\theta}$, and define the mode expansions:

$$\begin{aligned}b(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} b_n z^n \\ c(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} c_n z^n \\ \bar{b}(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} \bar{b}_n \bar{z}^n \\ \bar{c}(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} \bar{c}_n \bar{z}^n\end{aligned} \tag{14.88}$$

eq:bcsysi

$b_n, c_n, \bar{b}_n, \bar{c}_n$ are only unambiguously defined for $\mid n \mid << N$. That is, we focus on operators that do not mix excitations around the two Fermi levels. The peculiar half-integral moding is chosen to agree with standard conventions in CFT. In terms of the original nonrelativistic modes we have:

$$\begin{aligned}c_n &= B^+_{-n_F-\epsilon+n} \\ b_n &= B_{n_F+\epsilon+n} \\ \bar{c}_n &= B^+_{n_F+\epsilon-n} \\ \bar{b}_n &= B_{-n_F-\epsilon-n}\end{aligned} \tag{14.89}$$

eq:mapmodes

where $\epsilon = \frac{1}{2}$, so that

$$\{b_n, c_m\} = \delta_{n+m,0} \qquad \{\bar{b}_n, \bar{c}_m\} = \delta_{n+m,0} \tag{14.90}$$

eq:anticomms

and all other anticommutators equal zero.

We may now reinterpret the fields $b, c, ...$ Defining $z = e^{i\theta+\tau}$ we see that these may be extended to fields in two-dimensions, and that they are (anti-) chiral, that is, they satisfy the two-dimensional Euclidean Dirac equation. Upon continuation to Minkowski space we have $z \to e^{i(\theta+t)}$ and $\bar{z} \to e^{i(\theta-t)}$. We are thus discussing two relativistic massless Fermi fields in $1+1$ dimensions. The b, c fields generate a Fermionic Fock space built on the product of vacua $\mid 0\rangle_{bc} \otimes \mid \bar{0}\rangle_{\bar{b}\bar{c}}$ where $b_n \mid 0\rangle_{bc} = c_n \mid 0\rangle_{bc} = 0$ for $n > 0$. Explicitly, the b, c fields by themselves generate the Fock space:

$$\mathcal{H}_{bc} = \text{Span}\left\{\prod b_{n_i} \prod c_{m_i} \mid 0\rangle_{bc}\right\} \qquad . \tag{14.91}$$

eq:cftstspc

The space $\mathcal{H}_{bc}$ has a natural grading according to the eigenvalue of $\sum_{n\in\mathbb{Z}} : b_n c_{-n} = \oint bc$ (called "bc-number"):

$$\mathcal{H}_{bc} = \oplus_{p\in\mathbb{Z}} \mathcal{H}_{bc}^{(p)} \tag{14.92}$$

eq:grdspce

and the states obtained by moving a state from the Fermi sea to an excited level correspond to the subspace of zero bc number:

$$\mathcal{H}_{\text{chiral}} := \mathcal{H}_{bc}^{(0)} \tag{14.93}$$

eq:chrlspc

NEED TO GIVE HERE THE TRANSLATION BETWEEN a state corresponding to a Young Diagram and an explicit fermionic Fock state.

The bc system defines a conformal field theory. Indeed we may introduce

$$L_n = \sum_{m=-\infty}^{\infty} (n/2 + m) c_{-m} b_{m+n} \tag{14.94}$$

eq:frmvrop

and compute that the L_n satisfy the Virasoro algebra

$$[L_n, L_m] = (n-m)L_{n+m} + \frac{c}{12} n(n^2-1)\delta_{n+m,0} \tag{14.95}$$

eq:virasoroalg

with $c = 1$.

Now let us consider the second quantized operators corresponding to z^n. Before taking the limit these are given by:

$$\Upsilon_n = \sum z_i^n \qquad \longleftrightarrow \qquad \int d\theta \Psi^\dagger(\theta)\ e^{in\theta} \Psi(\theta). \tag{14.96}$$

eq:upsop

Now let us take the large N limit and focus on terms that do not mix the two Hilbert spaces of excitations around the two Fermi levels. Cross terms between barred and unbarred fields involve operators that mix the two Fermi levels. Since we are only interested in the case of the decoupled Fermi level excitations we may replace:

$$\begin{aligned}\Upsilon_n &\to \oint dz\ z^{-1-n} c\ b(z) + \oint d\bar{z}\ \bar{z}^{-1+n} \bar{c}\ \bar{b}(\bar{z}) \\ &= \sum_m c_{n-m} b_m + \bar{c}_{m-n} \bar{b}_{-m} \\ &= \alpha_n + \bar{\alpha}_{-n}\end{aligned} \tag{14.97}$$ eq:upsii

where we have introduced a field $bc = i\partial_z \phi(z)$ which has expansion

$$\begin{aligned}\partial_z \phi(z) &= i \sum_{m \in \mathbb{Z}} \alpha_m z^{m-1} \\ [\alpha_m, \alpha_n] = [\bar{\alpha}_m, \bar{\alpha}_n] &= m\delta_{m+n,0} \\ [\alpha_m, \bar{\alpha}_n] &= 0.\end{aligned} \tag{14.98}$$ eq:bosalg

In terms of α_n, we may introduce a representation of the Virasoro algebra:

$$L_n = \frac{1}{2} \sum \alpha_{n-m} \alpha_m \tag{14.99}$$ eq:virop

which satisfy (14.95), again with $c = 1$. Using the α we can define a vacuum $\alpha_n \mid 0\rangle = 0$ for $n \geq 0$ and a statespace

$$\mathcal{H}_\alpha = \text{Span} \left\{ \mid \vec{k}\rangle \equiv \prod (\alpha_{-j})^{k_j} \mid 0\rangle \right\}. \tag{14.100}$$ eq:achalph

Bosonization states that there is a natural isomorphism:

$$\mathcal{H}_\alpha \cong \mathcal{H}_{bc}^{(0)} \tag{14.101}$$ eq:bsnztin

We will not prove this but it can be made very plausible as follows. The Hilbert space may be graded by L_0 eigenvalue. The first few levels are:

$$\begin{array}{lccc} L_0 = 1 & \{b_{-\frac{1}{2}} c_{-\frac{1}{2}} \mid 0\rangle\} & & \{\alpha_{-1} \mid 0\rangle\} \\ L_0 = 2 & \{b_{-\frac{1}{2}} c_{-\frac{3}{2}} \mid 0\rangle, b_{-\frac{3}{2}} c_{-\frac{1}{2}} \mid 0\rangle\} & & \{\alpha_{-2} \mid 0\rangle, (\alpha_{-1})^2 \mid 0\rangle\} \end{array} \tag{14.102}$$ eq:firstlev

At level $L_0 = n$, the fermion states are labeled by Young diagrams Y with n boxes. At level $L_0 = n$, the Bose basis elements are labeled by partitions of n. We will label a partition of n by a vector $\vec{k} = (k_1, k_2, \dots)$ which has almost all entries zero, such that $\sum_j j k_j = n$. Bosonization states that the two bases are linearly related:

$$\mid Y\rangle = \sum_{\vec{k} \in \text{Partitions}(n)} \langle \vec{k} \mid Y\rangle \mid \vec{k}\rangle \tag{14.103}$$ eq:linrel

This last relation can be understood as the relation (14.72).

****

1. Need to explain this some more.

2. Look up papers of the Kyoto school, Jimbo et. al. for perhaps helpful ways of presenting this material.

****

For much more about this see the references below.

1. Pressley and Segal, *Loop Groups*, Chapter 10.

2. M. Stone, "Schur functions, chiral bosons, and the quantum-Hall-effect edge states," Phys. Rev. **B42** 1990)8399 (I have followed this treatment.)

3. S. Cordes, G. Moore, and S. Ramgoolam, "Lectures on 2D Yang-Mills Theory, Equivariant Cohomology and Topological Field Theories," Nucl. Phys. B (Proc. Suppl 41) (1995) 184, section 4. Also available at hep-th/9411210.

4. M. Douglas, "Conformal field theory techniques for large N group theory," hep-th/9303159.

Preprint typeset in JHEP style - HYPER VERSION

Chapter 4: Introduction to Representation Theory

Gregory W. Moore

Abstract: BASED MOSTLY ON GTLECT4 FROM 2009. BUT LOTS OF MATERIAL HAS BEEN IMPORTED AND REARRANGED FROM MATHMETHODS 511, 2014 AND GMP 2010 ON ASSOCIATED BUNDLES. SOMETHING ABOUT SU(2) REPS AND INDUCED REPS NOW RESTORED. BECAUSE OF IMPORTS THERE IS MUCH REDUNDANCY. THIS CHAPTER NEEDS A LOT OF WORK. April 27, 2018

-TOC- **Contents**

1. Symmetries of physical systems **3**

2. Basic Definitions **4**
2.1 Representation of a group 4
2.2 Matrix Representations 5
2.3 Examples 6
2.3.1 The fundamental representation of a matrix Lie group 6
2.3.2 The determinant representation 6
2.3.3 A representation of the symmetric groups S_n 6
2.3.4 $\mathbb{Z}$ and $\mathbb{Z}_2$ 7
2.3.5 The Heisenberg group 7

3. Unitary Representations **8**
3.1 Invariant Integration 9
3.2 Unitarizable Representations 10
3.3 Unitary representations and the Schrödinger equation 11

4. Projective Representations and Central Extensions **12**

5. Induced Group Actions On Function Spaces **13**

6. The regular representation **14**
6.1 Matrix elements as functions on G 16
6.2 R_G as a unitary rep: Invariant integration on the group 17
6.3 A More Conceptual Description 17

7. Reducible and Irreducible representations **18**
7.1 Definitions 18
7.2 Reducible vs. Completely reducible representations 21

8. Schur's Lemmas **22**

9. Orthogonality relations for matrix elements **23**

10. Decomposition of the Regular Representation **25**

11. Fourier Analysis as a branch of Representation Theory **31**
11.1 The irreducible representations of abelian groups 31
11.2 The character group 31
11.3 Fourier duality 33
11.4 The Poisson summation formula 36

11.5 Application: Bloch's Theorem in Solid State Physics 37
11.6 The Heisenberg group extension of $\hat{S} \times S$ for an abelian group S 39

12. Induced Representations 41
12.1 Induced Representations of Finite Groups: Frobenius Reciprocity 42

13. Representations Of $SU(2)$ 48
13.0.1 Matrix Reps Of $\mathfrak{su}(2)$ 51
13.1 Unitary structure 53
13.2 Lie algebra representations and differential operators 54
13.3 Coherent state formalism 54

14. Orbits of the Lorentz group and relativistic wave equations 55
14.1 Orbits, Representations, and Differential Equations 55
14.2 The massless case in $1+1$ dimensions 57
14.3 The case of d dimensions, $d > 2$ 58

15. Characters and the Decomposition of a Representation to its Irreps 61
15.1 Some basic definitions 61
15.2 Orthogonality relations on the characters 63

16. Some examples of representations of discrete groups and their character tables 65
16.1 S_3 65
16.2 Dihedral groups 68

17. The decomposition of tensor products 69

18. Algebras 71
18.1 Coalgebras, Bialgebras, and Frobenius algebras 72
18.2 When are two algebras equivalent? Introduction to Hochschild cohomology 73

19. The group ring and group algebra 73
19.1 Projection operators in the group algebra 75
19.2 The center of the group algebra and the subalgebra of class functions 76

20. Interlude: 2D Topological Field Theory and Frobenius algebras 77
20.1 Geometric Categories 77
20.2 Some general properties 80
20.3 Two dimensional closed topological field theory and Commutative Frobenius Algebras 82
20.4 Boundary conditions 85

21. Applications of the Projection Operators **89**
21.1 Decomposition of a representation into its isotypical parts 90
21.2 Block diagonalization of Hermitian operators 92
21.2.1 Projecting quantum wavefunctions 92
21.2.2 Finding normal modes in classical mechanics 93

22. Representations of the Symmetric Group **96**
22.1 Conjugacy classes in S_n 96
22.2 Young tableaux 96
22.2.1 Example 1: $G = S_3$ 98
22.2.2 Example 2: $G = S_4$ 99

23. Symmetric groups and tensors: Schur-Weyl duality and the irreps of $GL(d, k)$ **99**
23.1 Free fermions on a circle and Schur functions 103
23.1.1 Schur functions, characters, and Schur-Weyl duality 108
23.2 Bosons and Fermions in 1+1 dimensions 111
23.2.1 Bosonization 111

---

1. Symmetries of physical systems

♣Much of this is redundant with discussion of Wigner's theorem which should be done in general discussion of quantum mechanics in Chapter 2. ♣

One of the most important applications of group theory in physics is in quantum mechanics. The basic principle is that if G is a symmetry group of a physical system(e.g., rotational symmetry, translational symmetry, ...) then each element $g \in G$ corresponds to a *unitary operator* acting on the Hilbert space $\mathcal{H}$ of physical states, i.e., we have an association:

$$\forall g \in G \mapsto U(g) : \mathcal{H} \to \mathcal{H} \tag{1.1}$$

where $U(g)$ is a unitary operator. Moreover, these unitary operators should "act in the same way as the physical symmetry." Mathematically, this means that we have the operator equation:

$$U(g_1)U(g_2) = U(g_1 g_2). \tag{1.2}$$

eq:repi

Moreover, if we have a symmetry of the physical system we should have the same kind of time-development of two systems related by the symmetry, so

$$U(g)HU(g)^{-1} = H \tag{1.3}$$

where H is the Hamiltonian.

The equation (1.2) is the essential equation defining what is called a *representation of a group*, and the above principle is one of the main motivations in physics for studying representation theory.

2. Basic Definitions

subsec:ssFD

2.1 Representation of a group

Let V be a vector space over a field κ and recall that $GL(V)$ denotes the group of all nonsingular linear transformations $V \to V$.

Definition 2.1.1 A *representation* of G with *representation space* V is a homomorphism

$$\begin{array}{c} T : g \mapsto T(g) \\ G \to GL(V) \end{array} \tag{2.1}$$

eq:repii

The "dimension of the representation" is by definition the dimension $\dim V$ of the vector space V. This number can be finite or infinite. For convenience and mathematical rectitude we will often - but not always - focus on the case $\dim V < \infty$.

Terminology: We will often abbreviate "representation" to "rep." Moreover we sometimes refer to "the rep (T, V)" or to "the rep T," or to "the rep V," when the rest of the data is understood. Some authors call V the "carrier space."

Definition 2.1.2. A linear transformation $A : V \to V'$ between representations of G is called an *intertwiner* if for all $g \in G$ the diagram

$$\begin{array}{ccc} V & \stackrel{A}{\to} & V' \\ T(g)\downarrow & & \downarrow T'(g) \\ V & \stackrel{A}{\to} & V' \end{array} \tag{2.2}$$

commutes. Equivalently,

$$T'(g)A = AT(g) \tag{2.3}$$

eq:intertwin

for all $g \in G$.

Definition 2.1.3. Two representations are *equivalent* $(T, V) \cong (T', V')$ if there is an intertwiner which is an isomorphism. That is,

$$T'(g) = AT(g)A^{-1} \tag{2.4}$$

eq:wuivrep

for all $g \in G$.

Familiar notions of linear algebra generalize to representations:

1. The direct sum $\oplus$, tensor product $\otimes$ etc. of representations. Thus, the direct sum of (T_1, V_1) and (T_2, V_2) is the rep $(T_1 \oplus T_2, V_1 \oplus V_2)$ where the representation space is $V_1 \oplus V_2$ *and* the operators are:

$$(T_1 \oplus T_2)(g)\big(v_1, v_2\big) := \big(T_1(g)v_1, T_2(g)v_2\big) \tag{2.5}$$

2. Similarly, for the tensor product, the carrier space is $V_1 \otimes V_2$ and the group elements are represented by:

$$(T_1 \otimes T_2)(g)\big(v_1 \otimes v_2\big) := (T_1(g)v_1) \otimes (T_2(g)v_2) \tag{2.6}$$

3. Given a representation of V we get a dual representation on the dual space $V^\vee$ by demanding that under the pairing between V and $V^\vee$:

$$\langle T^\vee(g)\ell, T(g)v\rangle = \langle \ell, v\rangle, \tag{2.7}$$

where $\ell \in V^\vee, v \in V$. Thus, if V, W are representation spaces then so is $Hom(V, W)$.

4. If V is a complex vector space then the complex conjugate representation sends $g \to \bar{T}(g) \in GL(\bar{V})$. A *real representation* is one where $(\bar{T}, \bar{V})$ is equivalent to (T, V).

2.2 Matrix Representations

Definition 2.1.3 A *matrix representation* is a homomorphism:

$$T : G \to GL(n, \kappa) \tag{2.8}$$

Given a representation, *and* an ordered basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ for V we can associate to a representation a matrix representation. The point is, with a basis we can identify $GL(V) \cong GL(n, \kappa)$. Specifically, we get the matrix from the linear transformation by:

$$T(g)\vec{v}_k = \sum_{j=1}^{n} T(g)_{jk}\vec{v}_j \tag{2.9}$$

so the $T(g)_{jk}$ are the matrix elements of an element of $GL(n, \kappa)$. We will sometimes denote this matrix as $T(g)$ if it is understood we are using an ordered basis for V.

Now, if we change basis

$$\vec{v}_i = \sum_{j=1}^{n} S_{ji}\vec{v}'_j \tag{2.10}$$

then the matrices change to

$$T'(g) = ST(g)S^{-1} \tag{2.11}$$

and this motivates

Definition Two n-dimensional matrix representations T and T' are *equivalent*, denoted $T \cong T'$ if $\exists S \in GL(n, \kappa)$ with

$$T'(g) = ST(g)S^{-1} \tag{2.12}$$

Exercise *Complex conjugate and transpose-inverse representations*
Given a matrix representation of a group $g \to T(g)$ show that
a.) $g \to (T(g))^{tr,-1}$ is also a representation.

If we choose a basis v_i for V then this is the matrix representation for the dual representation in the dual basis $\hat{v}_i$.

b.) If T is a complex matrix representation wrt basis v_i then the complex conjugate representation with respect to $\bar{v}_i$ is: $g \to T(g)^*$.

c.) If T is a real representation, then there exists an $S \in GL(n, \mathbb{C})$ such that for all $g \in G$:

$$T^*(g) = ST(g)S^{-1} \tag{2.13}$$

Warning: The matrix elements $T(g)_{ij}$ of a real representation can of course fail to be real numbers!

2.3 Examples

2.3.1 The fundamental representation of a matrix Lie group

$GL(n,\kappa), SL(n,\kappa), O(n,\kappa), U(n)$ are all matrix representations of themselves! In the first three examples $V = \kappa^n$. In the fourth $V = \mathbb{C}^n$. These are called "the fundamental representation." Note that when $\kappa = \mathbb{C}^n$ and $n > 2$ the fundamental representation is not equivalent to its complex conjugate [1] so the fundamental representation is not the same as the "minimal-dimensional nontrivial representation."

2.3.2 The determinant representation

The general linear group $GL(n,\kappa)$ always has a family of one-dimensional real representations $\det^\mu$, $\mu \in \mathbb{R}$, given by

$$T(A) := |\det A|^\mu \tag{2.14}$$

eq:dtrmrep

This is a representation because:

$$T(AB) = |\det AB|^\mu = |\det A|^\mu |\det B|^\mu = T(A)T(B) \tag{2.15}$$

2.3.3 A representation of the symmetric groups S_n

The symmetric group S_n has an n-dimensional representation with $V = \mathbb{R}^n$ (or $\mathbb{C}^n$) defined by choosing the standard basis $\{e_i\}$ and defining

$$T(\sigma) \cdot e_i := e_{\sigma(i)} \tag{2.16}$$

Note that acting on coordinates $x_i e_i$ this takes $x_i \to x_{\sigma^{-1}(i)}$, i.e.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \to \begin{pmatrix} x_{\sigma^{-1}(1)} \\ \vdots \\ x_{\sigma^{-1}(n)} \end{pmatrix} \tag{2.17}$$

and thus the matrix representation has matrix elements which are just 0's and 1's with a single nonzero entry in each row and column.

[1] This is easily proven using characters, see below.

Explicitly for $n = 2$ we have, for examples:

$$(12) \to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.18}$$ eq:symmrep

while for $n = 3$ we have:

$$\begin{aligned} (12) &\to \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ (23) &\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ (123) &\to \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{aligned} \tag{2.19}$$ eq:symmrepii

2.3.4 $\mathbb{Z}$ and $\mathbb{Z}_2$

Sometimes relations come in families. If $\theta \in \mathbb{R}/\mathbb{Z}$ then we can define a representation T_θ of the group $\mathbb{Z}$ by $T_\theta(n) = e^{2\pi i n\theta}$.

At special values of parameters, special things can happen. For example note that when $\theta = 1/2$ the subgroup of even integers is represented trivially, and the representation "factors through" the sign representation of S_2. That is, $T_{1/2}$ is the composition of the projection $\mathbb{Z} \to \mathbb{Z}_2$ and the sign representation ε of $\mathbb{Z}_2$.

Similar, but more elaborate and interesting things happen for the representations of the braid group and its projection $B_n \to S_n$.

2.3.5 The Heisenberg group

An interesting group, the Heisenberg group, can be defined in terms of generators q, U, V where q is central and $UV = qVU$:

$$H = \langle q, U, V | qU = Uq, qV = Vq, UV = qVU \rangle \tag{2.20}$$ eq:heisn

If we we add the relations $q^N = U^N = V^N = 1$ then it is the finite Heisenberg group extension of $\mathbb{Z}_N \times \mathbb{Z}_N$ by $\mathbb{Z}_N$:

$$1 \to \mathbb{Z}_N \to \mathrm{Heis}_N \to \mathbb{Z}_N \times \mathbb{Z}_N \to 1 \tag{2.21}$$ eq:heisent

described in chapter 1, section 11.3.

Let ω be an Nth root of 1.

Let $U = Diag\{1, \omega, \omega^2, \ldots, \omega^{N-1}\}$ be the "clock operator."

Let V be the "shift operator:"

$$V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad V = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad V = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix} \tag{2.22}$$

It is easy to check that:

$$UV = qVU \tag{2.23}$$

with $q = \omega$.

If we let ω range over the Nth roots of unity we obtain N inequivalent $N \times N$ (irreducible) representations.

If we consider the group without imposing $q^N = U^N = V^N = 1$ there are infinitely many inequivalent representations.

These assertions are obvious since different values of the central element q cannot be conjugated into each other.

3. Unitary Representations

sec:sUR

Of particular importance in physics are the unitary representations. By Wigner's theorem described in section **** above we know that symmetry transformations will act as unitary or anti-unitary transformations. The basic reason for this is that these preserve norms and hence probability amplitudes.

Definition 3.1. Let V be an inner product space. A *unitary representation* is a representation (T, V) such that $\forall g \in G$, $T(g)$ is a unitary operator on V, i.e.,

$$\langle T(g)v, T(g)v \rangle = \langle v, v \rangle \qquad \forall g \in G, v \in V \tag{3.1}$$

Definition 3.2. A unitary matrix representation is a homomorphism

$$T : G \to U(n) \tag{3.2}$$

Exercise

a.) Show that if $T(g)$ is a rep on an inner product space then $T(g^{-1})^\dagger$ is a rep also.

b.) Suppose $T : G \to GL(V)$ is a unitary rep on an inner product space V. Let $\{\vec{v}_i\}$ be an ON basis for V. Show that the corresponding matrix rep $T(g)_{ij}$ is a unitary matrix rep.

c.) Show that for a unitary matrix rep the transpose-inverse and complex conjugate representations are equivalent.

Definition 3.2. If a rep (T, V) is equivalent to a unitary rep then such a rep is said to be *unitarizable*.

Example. A simple example of non-unitarizable reps are the $\det^\mu$ reps of $GL(n,k)$ described in section 2.3.4.

3.1 Invariant Integration

When proving facts about unitary representations a very important tool is the notion of *invariant integration*. We have already made use of it in our discussion of lattice gauge theory in Section *** of Chapter 1. Here is a reminder:

****

Important Remark: The notion of averaging over the group can be extended to a much larger class of groups than finite groups. It is given by *invariant integration* over the group which is the analog of the operation

$$f \to \frac{1}{|G|} \sum_g f(g) \tag{3.3}$$

on functions.

In general we replace

$$\frac{1}{|G|} \sum_g f(g) \to \int_G f(g) dg \tag{3.4}$$

where $\int_G f(g)dg$ should be regarded as a rule such that:

1. $\int_G f(g)dg$ is a complex number depending linearly on $f \in R_G$.
2. It satisfies the *left invariance* property:

$$\int_G f(hg)dg = \int_G f(g)dg \tag{3.5}$$

for all $h \in G$. This is the generalization of the rearrangement lemma. (We can define right invariance in a similar way. Compact groups admit integration measures which are simultaneously left- and right- invariant.)

Examples:

$$\begin{aligned} \int_{G=U(1)} f(g)dg &\equiv \int_0^{+2\pi} \frac{d\theta}{2\pi} f(\theta) \\ \int_{G=\mathbb{Z}} f(g)dg &\equiv \sum_{n\in\mathbb{Z}} f(n) \\ \int_{G=\mathbb{R}} f(g)dg &\equiv \int_{-\infty}^{+\infty} dx f(x) \end{aligned} \tag{3.6}$$

eq:intgroup

The notion of invariant integration also extends to all compact groups. For the important case of $G = SU(2)$ we can write it as follows. First, every element of $SU(2)$ can be written as:

♣Definition of β here is backwards from what we use when we describe reps using homogeneous polynomials below. ♣

$$g = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} \tag{3.7}$$

for 2 complex numbers α, β with

$$|\alpha|^2 + |\beta|^2 = 1. \tag{3.8}$$

In this way we identify the group as a manifold as S^3. That manifold has no globally well-defined coordinate chart. The best we can do is define coordinates that cover "most" of the group but will have singularities are some places. (It is always important to be careful about those singularities when using explicit coordinates!) One way to do this is to write

$$\begin{aligned} \alpha &= e^{i\frac{1}{2}(\phi+\psi)} \cos\theta/2 \\ \beta &= ie^{i\frac{1}{2}(\phi-\psi)} \sin\theta/2 \end{aligned} \tag{3.9}$$

With this definition we can write

$$\begin{aligned} g &= \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix} \begin{pmatrix} \cos\theta/2 & i\sin\theta/2 \\ i\sin\theta/2 & \cos\theta/2 \end{pmatrix} \begin{pmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2} \end{pmatrix} \\ &= e^{i\frac{1}{2}\phi\sigma^3} e^{i\frac{1}{2}\theta\sigma^1} e^{i\frac{1}{2}\psi\sigma^3} \end{aligned} \tag{3.10}$$

If we take $0 \leq \theta \leq \pi$, and identify $\phi \sim \phi + 2\pi$ and $\psi \sim \psi + 4\pi$ we cover the manifold once, away from the singular points at $\theta = 0, \pi$.

The normalized Haar measure for $SU(2)$ in these coordinates is [2]

$$[dg] = \frac{1}{16\pi^2} d\psi \wedge d\phi \wedge \sin\theta d\theta \tag{3.11}$$

For much more about this, see Chapter 5 below.

We give formulae for the Haar measures on the classical compact matrix groups in Chapter 6 below.

****

3.2 Unitarizable Representations

Proposition If T is a rep on an inner product space of a *compact group* G, then V is finite dimensional and T is unitarizable.

Proof for finite groups: If T is not already unitary with respect to the inner product $(\cdot,\cdot)_1$ then we can define a new inner product by:

$$\langle v, w\rangle_2 \equiv \frac{1}{\mid G \mid} \sum_{g\in G} \langle T(g)v, T(g)w\rangle_1 \tag{3.12}$$

This proof generalizes using a left- and right- invariant Haar measure, which always exists for a compact group.

[2] We have chosen an orientation so that, with a positive constant, this is $\mathrm{Tr}_{\mathbf{2}}(g^{-1}dg)^3)$.

Exercise

a.) Show that

$$\langle T(g)v, T(g)w\rangle_2 = \langle v, w\rangle_2 \tag{3.13}$$

and deduce that $T(g)$ is unitary with respect to $\langle\cdot,\cdot\rangle_2$.

b.) Show that if $T(g)$ is a finite-dimensional matrix rep of a finite group then it is equivalent to a unitary matrix rep.

3.3 Unitary representations and the Schrödinger equation

♣THIS SUBJECTION IS NOW OUT OF PLACE. ♣

Consider the Schrödinger equation for stationary states of energy E:

$$-\frac{\hbar^2}{2m}\Delta\psi(\vec{x}) + V(\vec{x})\psi(\vec{x}) = E\psi(\vec{x}) \tag{3.14}$$

eq:shro

and suppose V is rotationally invariant:

$$V(\vec{x}) = V(\mid \vec{x} \mid). \tag{3.15}$$

Now define

$$V_E = \{\psi(\vec{x}) : \psi solves (3.14)\ and\ is\ normalizable\} \tag{3.16}$$

Normalizable, or square integrable means:

$$\parallel \psi \parallel^2 = \int_{R^3} \mid \psi(\vec{x}) \mid^2 d^3x < \infty \tag{3.17}$$

so $V_E \subset L^2(\mathbb{R}^3)$.

Claim: V_E is a representation space of $O(3)$: $T : O(3) \to GL(V_E)$ given, as usual, by:

$$[T(g)\psi](\vec{x}) = \psi(g^{-1}\cdot\vec{x}) \tag{3.18}$$

Must check:

1. ψ solves (3.14) $\Rightarrow T(g)\psi$ solves (3.14).
2. $\parallel \psi \parallel^2 < \infty \Rightarrow \parallel T(g)\psi \parallel^2 < \infty$

Check of 1: If $x'_i = (g^{-1})^j_i x_j$ then

$$\Delta' = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i'^2} = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2} = \Delta \tag{3.19}$$

and therefore:

$$\Delta\big[T(g)\psi\big](x) = \Delta(\psi(x')) = \Delta'(\psi(x')) \tag{3.20}$$

Check of 2: $x' = g^{-1}x$ gives

$$d^3x' = \mid \det\frac{dx^{i'}}{dx^j} \mid d^3x = d^3x \tag{3.21}$$

so:

$$\| \psi \|^2 = \| T(g)\psi \|^2 \tag{3.22}$$

and hence this is an example of a unitary representation.

al states of a theory form unitary reps of the symmetry group. Since the group commutes with the Hamiltonian we can diag

Thus, for example, the wavefunctions in a spherically symmetric potential have quantum numbers, $|E, j, m, \cdots\rangle$.

Exercise *Helmholtz equation*
Consider the Helmholtz equation:

$$\Delta u(\vec{x}) + k^2 u(\vec{x}) = 0$$
$$\Delta = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2} \tag{3.23}$$ eq:hlmhltz

Let

$$V = \{u(\vec{x}) : u \ \text{ satisfies(3.23)}\} \ \texttt{eq:hlmhltziii} \tag{3.24}$$ eq:hlmhltziii

Show that V in (3.24) is a representation of the isometry group $E(3)$.

---

4. Projective Representations and Central Extensions

♣GIVEN THE EXTENSIVE MATERIAL ON CENTRAL EXTENSIONS IN CHAPTER ONE THIS NEEDS TO BE REWRITTEN ♣eq:projrep

A *projective representation* is a pair (T, V) where $T : G \to GL(V)$ "almost" is a homomorphism, but is allowed to deviate from a homomorphism by a phase:

$$T(g_1)T(g_2) = c(g_1, g_2)T(g_1 g_2) \tag{4.1}$$

where $c(g_1, g_2) \in \mathbb{C}^*$. For unitary representations $c(g_1, g_2) \in U(1)$.

From the associativity of a product of three transformations $T(g_1)T(g_2)T(g_3)$ we find that $c : G \times G \to \mathbb{C}^*$ is a 2-cocycle.

Recalling our discussion of central extensions we see that T should really be regarded as a representation of the central extension of G by $\mathbb{C}^*$ (or $U(1)$) constructed from the cocycle c. Put more formally

$$\hat{T}(z, g) := zT(g) \tag{4.2}$$ eq:cntpfjr

defines a true representation of the central extension $\hat{G}$ defined by c.

Moreover, if we allow ourselves the freedom to redefine $T(g) \to \tilde{T}(g) := f(g)T(g)$ where $f : G \to \mathbb{C}^*$ is a function then c changes by a coboundary.

Example Returning to our clock and shift operators, we can define a *projective* representation of $\mathbb{Z}_n \times \mathbb{Z}_n$ by

$$(\bar{s}, \bar{s}') \to U^{\bar{s}} V^{\bar{s}'} \tag{4.3}$$

but these only satisfy the group law up to a power of ω. In fact, what we have is a representation of $\mathrm{Heis}(\mathbb{Z}_n \times \mathbb{Z}_n)$.

5. Induced Group Actions On Function Spaces

:InducedAction

♣NOTE THERE IS OVERLAP OF THIS MATERIAL WITH EXAMPLES 5,6 OF FUNCTORS IN CATEGORY SECTION OF CHAPTER 1 ♣

Let X be a G-set and let Y be any set. There are natural left- and right- actions on the function space $\mathrm{Map}(X, Y)$. Given $\Psi \in \mathrm{Map}(X, Y)$ and $g \in G$ we need to produce a new function $\phi(g, \Psi) \in \mathrm{Map}(X, Y)$. The rules are as follows:

1. If G is a left-action on X then

$$\phi(g, \Psi)(x) := \Psi(g \cdot x) \qquad \text{right action on } \mathrm{Map}(X, Y) \tag{5.1}$$

2. If G is a left-action on X then

$$\phi(g, \Psi)(x) := \Psi(g^{-1} \cdot x) \qquad \text{left action on } \mathrm{Map}(X, Y) \tag{5.2}$$

3. If G is a right-action on X then

$$\phi(g, \Psi)(x) := \Psi(x \cdot g) \qquad \text{left action on } \mathrm{Map}(X, Y) \tag{5.3}$$

4. If G is a right-action on X then

$$\phi(g, \Psi)(x) := \Psi(x \cdot g^{-1}) \qquad \text{right action on } \mathrm{Map}(X, Y) \tag{5.4}$$

Example: Consider a spacetime $\mathcal{S}$. With suitable analytic restrictions the space of scalar fields on $\mathcal{S}$ is $\mathrm{Map}(\mathcal{S}, \kappa)$, where $\kappa = \mathbb{R}$ or $\mathbb{C}$ for real or complex scalar fields. If a group G acts on the spacetime, there is automatically an induced action on the space of scalar fields. To be even specific, suppose $X = \mathbb{M}^{1,d-1}$ is d-dimensional Minkowski space time, G is the Poincaré group, and $Y = \mathbb{R}$. Given one scalar field Ψ and a Poincaré transformation $g^{-1} \cdot x = \Lambda x + v$ we have $(g \cdot \Psi)(x) = \Psi(\Lambda x + v)$.

Similarly, suppose that X is any set, but now Y is a G-set. Then again there is a G-action on $\mathrm{Map}(X, Y)$:

$$(g \cdot \Psi)(x) := g \cdot \Psi(x) \qquad \text{or} \qquad \Psi(x) \cdot g \tag{5.5}$$

according to whether the G action on Y is a left- or a right-action, respectively. These are left- or right-actions, respectively.

We can now combine these two observations and get the general statement: We assume that both X is a G_1-set and Y is a G_2-set. We can assume, without loss of generality, that we have left-actions on both X and Y. Then there is a natural $G_1 \times G_2$-action on $\mathrm{Map}(X, Y)$ defined by:

$$\phi((g_1, g_2), \Psi)(x) := g_2 \cdot (\Psi(g_1^{-1} \cdot x)) \tag{5.6}$$

eq:GenAction

note that if one writes instead $g_2 \cdot (\Psi(g_1 \cdot x))$ on the RHS then we do not have a well-defined $G_1 \times G_2$-action (if G_1 and G_2 are both nonabelian). In most applications X and Y both have a G action for a single group and we write

$$\phi(g, \Psi)(x) := g \cdot (\Psi(g^{-1} \cdot x)) \tag{5.7}$$

This is a special case of the general action (5.6), with $G_1 = G_2 = G$ and specialized to the diagonal $\Delta \subset G \times G$.

Example: Again let $X = \mathbb{M}^{1,d-1}$ be a Minkowski space time. Take $G_1 = G_2$ and let $G = \Delta \subset G \times G$ be the diagonal subgroup, and take G to be the Poincaré group. Now let $Y = V$ be a finite-dimensional representation of the Poincaré group. Let us denote the action of $g \in G$ on V by $\rho(g)$. Then a field $\Psi \in \mathrm{Map}(X, Y)$ has an action of the Poincaré group defined by

$$g \cdot \Psi(x) := \rho(g)\Psi(g^{-1}x) \tag{5.8}$$

This is the standard way that fields with nonzero "spin" transform under the Poincaré group in field theory. As a very concrete related example, consider the transformation of electron wavefunctions in nonrelativistic quantum mechanics. The electron wavefunction is governed by a two-component function on $\mathbb{R}^3$:

$$\Psi(\vec{x}) = \begin{pmatrix} \psi_+(\vec{x}) \\ \psi_-(\vec{x}) \end{pmatrix} \tag{5.9}$$

Then, suppose $G = SU(2)$. Recall there is a surjective homomorphism $\pi : G \to SO(3)$ defined by $\pi(u) = R$ where

$$u\vec{x} \cdot \vec{\sigma} u^{-1} = (R\vec{x}) \cdot \vec{\sigma} \tag{5.10}$$

Then the (double-cover) of the rotation group acts to define the transformed electron wavefunction $u \cdot \Psi$ by

$$(u \cdot \Psi)(\vec{x}) := u \begin{pmatrix} \psi_+(R^{-1}\vec{x}) \\ \psi_-(R^{-1}\vec{x}) \end{pmatrix} \tag{5.11}$$

In particular, $u = -1$ acts trivially on $\vec{x}$ but nontrivially on the wavefunction.

6. The regular representation

Now that we have seen several examples of representations we would like to introduce a "universal example," of a representation of G, called the *regular representation.* We will see that from this representation we can learn about *all* of the representations of G, at least when G is compact.

Recall once again the principle of section 5: If X has G-action then the space $\mathcal{F}_{X,Y}$ of all functions $X \to Y$ also has a G-action. In particular, if Y is a vector space then $\mathcal{F}_{X,Y}$ is a vector space and we have a natural source of representations of G. In particular, we can make the simplest choice $Y = \mathbb{C}$, so the space of complex-valued functions on X, $\mathcal{F}_{X,\mathbb{C}}$ is a very natural source of representations of G.

♣THIS IS OLD. SHOULD MAKE IT A REPRESENTATION OF $G \times G$ FROM THE START. ♣

If G acts on nothing else, it certainly acts on itself as a group of transformations of $X = G$ with left or right action. Applying the above general principle we see that the space of complex-valued functions on G:

$$\mathcal{F}_{G,\mathbb{C}} := \{Maps \qquad \phi : G \to \mathbb{C}\} \tag{6.1}$$

forms a natural - "God given" - representation space of G. $\mathcal{F}_{G,\mathbb{C}}$ is also denoted by R_G. When G is infinite we usually want to put some conditions on the functions in question. If G admits a left-invariant integration measure then by R_G we understand $L^2(G)$ in this measure.

In fact, R_G is a representation space in two ways:

1. Left regular representation (LRR):

$$\phi \mapsto L(g) \cdot \phi \tag{6.2}$$

where $L(g) \cdot \phi$ has values defined by

$$(L(g) \cdot \phi)(h) \equiv \phi(g^{-1}h) \tag{6.3}$$

2. Right regular representation (RRR):

$$\phi \mapsto R(g) \cdot \phi \tag{6.4}$$

where $R(g) \cdot \phi$ has values defined by

$$(R(g) \cdot \phi)(h) \equiv \phi(hg) \tag{6.5}$$

Exercise

Check that

$$L(g_1)L(g_2) = L(g_1 g_2) \tag{6.6}$$

and that

$$R(g_1)R(g_2) = R(g_1 g_2) \tag{6.7}$$

Note too that for all $g, h \in G$

$$R(g)L(h) = L(h)R(g) \tag{6.8}$$

so, in fact, R_G is a representation of $G \times G = G_{\text{left}} \times G_{\text{right}}$.

Now let us assume $\mid G \mid < \infty$. Then R_G is a finite dimensional vector space. we can label a function f by its values $f(g)$.

Example. $G = \mathbb{Z}_2 = \{1, \sigma\}$. Then $R_{\mathbb{Z}_2} \cong \mathbb{C}^2$ as a vector space, since a function on $\mathbb{Z}_2$ is labelled by (a, b) where $a = f(1), b = f(\sigma)$.

Quite generally, we can introduce a basis of "delta functions" concentrated at g:

$$\delta_g(g') = \begin{cases} 1 & for g' = g \\ 0 & else \end{cases} \tag{6.9}$$

The functions $\delta_g \in R_G$ span R_G:

$$f = \sum_g f(g)\delta_g \tag{6.10}$$

Exercise

If G is a finite group, then show that

$$\dim R_G = |G| \tag{6.11}$$

Exercise

a.) Let $\delta_0, \delta_1, \delta_2$ be a basis of functions in the regular representation of $\mathbb{Z}_3$ which are 1 on $1, \omega, \omega^2$, respectively, and zero elsewhere. Show that ω is represented as

$$L(\omega) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{6.12}$$

b.) Show that

$$\begin{aligned} L(h) \cdot \delta_g &= \delta_{h \cdot g} \\ R(h) \cdot \delta_g &= \delta_{g \cdot h^{-1}} \end{aligned} \tag{6.13}$$

and conclude that for the left, or right, regular representation of a finite group the representation matrices in the δ-function basis have entries which are simply zeroes or ones in the natural basis.

6.1 Matrix elements as functions on G

Now we make an important conceptual step forward:

Let $g \to T_{ij}(g)$ be any matrix representation of G.

♣Actually, for current level of earlier chapters this is pretty obvious to the reader. ♣

> We may regard the matrix elements $T_{m\ell}$ as complex-valued *functions* on G. Therefore, we may regard $T_{m\ell}$ as *vectors* in the regular representation R_G.

6.2 R_G as a unitary rep: Invariant integration on the group

When G is a finite group we can turn the vector space R_G into an inner product space by the defining an inner product using averaging over the group:

$$\langle \phi_1, \phi_2 \rangle := \frac{1}{|G|} \sum_{g \in G} \phi_1^*(g) \phi_2(g) \tag{6.14}$$

eq:innprod

Theorem. With the inner product (6.14) the representation space R_G is *unitary* for both the RRR and the LRR.

Proof: Straightforward application of the rearrangement lemma.

6.3 A More Conceptual Description

♣MUCH OF THIS IS REDUNDANT WITH ABOVE BUT SOME THINGS ARE SAID MORE NICELY, esp. map of $End(V)$ into $L^2(G)$. ♣

As a very nice example of the idea of how group actions on a space induce group actions on the functions on that space we touch briefly on the Peter-Weyl theorem and the idea of induced representations.

We first describe the Peter-Weyl theorem:

Let G be a group. Then there is a left action of $G \times G$ on G: $(g_1, g_2) \mapsto L(g_1)R(g_2^{-1})$. Now let $Y = \mathbb{C}$. Then $\mathrm{Map}(G, \mathbb{C})$ is known as the *regular representation* of G because the induced left-action:

$$((g_1, g_2) \cdot \Psi)(h) := \Psi(g_1^{-1} h g_2) \tag{6.15}$$

converts the vector space $\mathrm{Map}(G \to \mathbb{C})$ of functions $\Psi : G \to \mathbb{C}$ into a representation space for $G \times G$.

♣Notation in this subsection is at variance with rest of chapter: Here ρ is the rep homomorphism and T is a general operator... ♣

Suppose, on the other hand that V is a linear representation of G. As mentioned above this means we have a group homomorphism $\rho : G \to GL(V)$. Then consider the vector space of linear transformations $\mathrm{End}(V)$ of V to itself. This is also a representation of $G \times G$ because if $T \in \mathrm{End}(V)$ then we can define a linear left-action of $G \times G$ on $\mathrm{End}(V)$ by:

$$(g_1, g_2) \cdot T := \rho(g_1) \circ T \circ \rho(g_2)^{-1} \tag{6.16}$$

Now, we have two representations of $G \times G$. How are they related? If V is finite-dimensional we have a map

$$\iota : \mathrm{End}(V) \to \mathrm{Map}(G, \mathbb{C}) \tag{6.17}$$

The map ι takes a linear transformation $T : V \to V$ to the complex-valued function $\Psi_T : G \to \mathbb{C}$ defined by

$$\Psi_T(g) := \mathrm{Tr}_V(T \rho(g^{-1})) \tag{6.18}$$

eq:PsiTee

If we choose a basis w_μ for V then the operators $\rho(g)$ are represented by matrices:

$$\rho(g) \cdot w_\nu = \sum_\mu D(g)_{\mu\nu} w_\mu \tag{6.19}$$

If we take $T = e_{\nu\mu}$ to be the matrix unit in this basis then Ψ_T is the function on G given by the matrix element $D(g^{-1})_{\mu\nu}$. So the Ψ_T's are linear combinations of matrix elements of

the representation matrices of G. The advantage of (6.18) is that it is completely canonical and basis-independent.

Note that $\iota : T \mapsto \Psi_T$ "commutes with the $G \times G$ action." What this means is that

$$(g_1, g_2) \cdot \Psi_T = \Psi_{(g_1,g_2)\cdot T} \tag{6.20}$$

eq:GxG-commute

(The reader should check this carefully.) Such a map is said to be *equivariant*. Put differently, denoting by $\rho_{\mathrm{End}(V)}$ the representation of $G \times G$ on $\mathrm{End}(V)$ and $\rho_{Reg.Rep.}$ the representation of $G \times G$ on $\mathrm{Map}(G, \mathbb{C})$ we get a commutative diagram:

$$\begin{array}{ccc} \mathrm{End}(V) & \xrightarrow{\iota} & \mathrm{Map}(G, \mathbb{C}) \\ \Big\downarrow{\scriptstyle \rho_{\mathrm{End}(V)}} & & \Big\downarrow{\scriptstyle \rho_{Reg.Rep.}} \\ \mathrm{End}(V) & \xrightarrow{\iota} & \mathrm{Map}(G, \mathbb{C}) \end{array} \tag{6.21}$$

In particular if we have a collection of finite-dimensional representations $\{V_\lambda\}$ of G then we have

$$\oplus_\lambda \mathrm{End}(V_\lambda) \hookrightarrow \mathrm{Map}(G, \mathbb{C}) \tag{6.22}$$

eq:PW-1

Thanks to the equivariance, the image of (6.22) is a $G \times G$-invariant subspace, i.e. a subrepresentation of $\mathrm{Map}(G, \mathbb{C})$. The very beautiful *Peter-Weyl theorem* states that, if G is a compact group, then, as representations of $G \times G$, (6.22) is an isomorphism if the sum is over the distinct isomorphism classes λ of irreducible representations V_λ of G and we restrict to the the subspace of $\mathrm{Map}(G, \mathbb{C})$ of L^2-normalizable functions with respect to a left-right-invariant measure on G. [3]

♣IRREPS NOT YET EXPLAINED ♣

7. Reducible and Irreducible representations

7.1 Definitions

Sometimes reps are "too big" and one wants to reduce them to their "essential parts." Reducing a representation to smaller representations is closely analogous to diagonalization of matrices.

For example – we will see that the natural representation of a group - its regular representation - is in fact highly reducible.

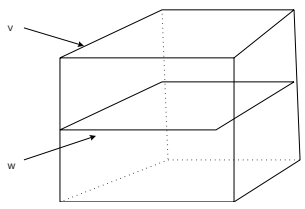


Figure 1: $T(g)$ preserves the subspace W.

fig:redrep

[3] In order for this to be completely correct we need to assume that G is a compact group. Then we introduce a left-right-invariant measure on G and replace the LHS by $L^2(G)$.

Definition. Let $W \subset V$ be a subspace of a group representation $T : G \to GL(V)$. Then W is *invariant* under T if $\forall g \in G, w \in W$

$$T(g)w \in W \tag{7.1}$$

This may be pictured as in 1. Example: Consider the three-dimensional representation of $SO(2)$ as rotations around the z-axis. Then the vector subspace of the xy plane is an invariant subspace.

Since T preserves the smaller vector space W, we can define a smaller group representation (T, W). This is called the *restriction of* T *to* W.

Remarks

•

If T is unitary on V then it is unitary on W.

• Both $\{\vec{0}\}$ and V are always invariant subspaces.

Definition. A representation T is called *reducible* if there is an invariant subspace $W \subset V$, under T, which is nontrivial, i.e., such that $W \neq 0, V$, If V is not reducible we say V is *irreducible*. That is, in an irreducible rep, the only invariant subspaces are $\{\vec{0}\}$ and V. We often shorten the term "irreducible representation" to "irrep."

Note:

• Given any nonzero vector $v \in V$, the linear span of $\{T(g)v\}_{g\in G}$ is an invariant subspace. In an irrep this will span all of V.

• If W is an subrepresentation of V then T descends to a representation on V/W.

Let us see how this works in terms of matrix representations. Suppose T is reducible. Then we can choose a basis

$\{v_1, \ldots v_k\}$ for W

and

$\{v_1, \ldots v_k, v_{k+1}, \ldots v_n\}$ for V

In this basis the matrix representation of T looks like:

$$T(g) = \begin{pmatrix} T_1(g) & T_{12}(g) \\ 0 & T_2(g) \end{pmatrix} \tag{7.2}$$ eq:redrep

where

$T_1(g) \in Mat_{k\times k}$

$T_{12}(g) \in Mat_{k\times n-k}$

$T_2(g) \in Mat_{(n-k)\times(n-k)}$

Writing out $T(g_1)T(g_2) = T(g_1g_2)$ we see that T_1 is the representation on W and T_2 is the representation on V/W. T_{12} transforms in a more complicated way.

Definition. A representation T is called *completely reducible* if it is isomorphic to

$$W_1 \oplus \cdots \oplus W_n \tag{7.3}$$

where the W_i are irreducible reps. Thus, there is a basis in which the matrices look like:

$$T(g) = \begin{pmatrix} T_1(g) & 0 & 0 & \cdots \\ 0 & T_2(g) & 0 & \cdots \\ 0 & 0 & T_3(g) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{7.4}$$

eq:comredrep

Examples

• $G = \mathbb{Z}_2$

$$\begin{aligned} 1 &\to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ (12) &\to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{7.5}$$

is a 2-dimensional *reducible* rep on $\mathbb{R}^2$ because $W = \{(x, x)\} \subset \mathbb{R}^2$ is a nontrivial invariant subspace. Indeed, σ^1 is diagonalizable, so this rep is equivalent to

$$\begin{aligned} 1 &\to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ (12) &\to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{7.6}$$

which is a direct sum of two 1×1 reps.

• Consider the representation of S_n on $\mathbb{R}^n$. Then the one-dimensional subspace $L = \{(x, \cdots, x)\}$ is a subrepresentation. Moreover we can take $L^\perp = \{x_1, \cdots x_n | \sum x_i = 0\}$. Then $L^\perp$ is an $(n-1)$-dimensional representation of S_n and the representation is block diagonal on $L \oplus L^\perp$.

• Invariant subspaces of the regular representation. Let T_{ij} be *any* matrix n-dimensional representation of G.

♣THIS IS OLD. MAYBE START WITH SPAN OF T_{ij} WHERE BOTH i, j VARY AS REP OF $G \times G$. WE CAN ALWAYS THEN RESTRICT TO $G \times 1$ AND $1 \times G$. ♣

Claim: The linear span of functions

$$\mathcal{R} := Span\{T_{ij}\}_{j=1,\ldots,n_\mu} \tag{7.7}$$

is an invariant subspace of R_G, considered as the RRR.

Proof:

$$\begin{aligned} (R(g) \cdot T_{ij})(h) &= T_{ij}(hg) \\ &= \sum_{s=1}^{n_\mu} T_{is}(h) T_{sj}(g) \end{aligned} \tag{7.8}$$

which is equivalent to the equation on functions:

$$\overbrace{R(g) \cdot T_{ij}}^{\text{function on G}} = \sum_{s=1}^{n_\mu} \underbrace{T_{sj}(g)}_{\text{matrix element for } R_G} \overbrace{T_{is}}^{\text{function on G}} \tag{7.9}$$

eq:mtxeltinvt

subsec:ssRedC

7.2 Reducible vs. Completely reducible representations

Irreps are the "atoms" out of which all reps are made. Thus we are naturally led to study the *irreducible* reps of G. In real life it can and does actually happen that a group G has representations which are reducible but not completely reducible. Reducible, but not completely reducible reps are sometimes called *indecomposable*.

Example . An example of an indecomposible rep which is not completely reducible is the rep

$$A \to \begin{pmatrix} 1 & \log|\det A| \\ 0 & 1 \end{pmatrix} \tag{7.10}$$

of $GL(n,\mathbb{R})$. As we will see, the Poincare group has indecomposable reps.

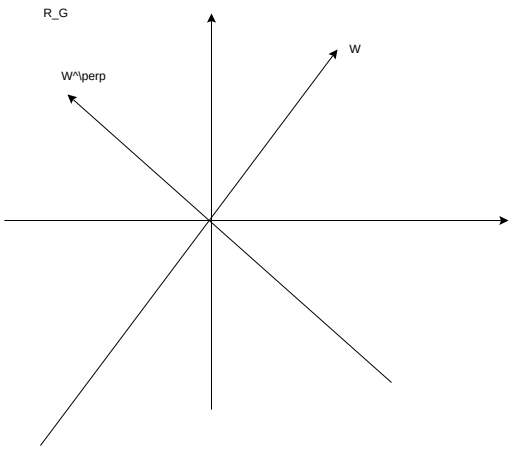


Figure 2: The orthogonal complement of an invariant subspace is an invariant subspace. fig:orthcomp

It is useful to have criteria for when this complication cannot occur:

Proposition 7.2.1 Suppose that the representation (T,V) is a unitary representation on an inner product space V and $W \subset V$ is an invariant subspace then $W^\perp$ is an invariant subspace:

Proof: Recall that

$$y \in W^\perp \Leftrightarrow \forall x \in W, \quad \langle y, x\rangle = 0 \tag{7.11}$$

Let $g \in G$, $y \in W^\perp$. Compute

$$\begin{aligned} \big(T(g)y, x\big) &= \big(y, T(g)^\dagger x\big) \\ &= \big(y, T(g^{-1})x\big) \end{aligned} \tag{7.12}$$

But, $T(g^{-1})x \in W$, since $W \subset V$ is an invariant subspace.

therefore: $\forall x \in W$, $\langle T(g)y, x\rangle = 0$.

therefore: $T(g)y \in W^\perp$
therefore: $W^\perp$ is an invariant subspace. ♠

Corollary: *Finite dimensional unitary representations are always completely reducible.*

In particular, by section 4:
1. Finite dimensional reps of finite groups are completely reducible.
2. More generally, finite dimensional reps of compact Lie groups are completely reducible.

It follows that for a finite group the regular representation R_G is completely reducible. In the next several sections we will show how to decompose R_G in terms of the irreducible representations of G.

If V is a completely decomposable representation (always true for compact groups) then we can write

$$V \cong \oplus_{\mu=1}^{r} \oplus_{i=1}^{a_\mu} T^{(\mu)} := \oplus_\mu a_\mu T^{(\mu)} \tag{7.13}$$

For a fixed μ let let

$$V^{(\mu)} := \oplus_{i=1}^{a_\mu} T^{(\mu)} \tag{7.14}$$

It contains the representation μ with the correct degeneracy in V. It is called the *isotypical component* belonging to μ. Note that it can be written as

$$V^{(\mu)} = \mathbb{C}^{a_\mu} \otimes T^{(\mu)} \tag{7.15}$$

eq:isotpy

where $\mathbb{C}^{a_\mu}$ is understood to be the trivial representation of G.

We abbreviate $\mathbb{C}^{a_\mu} \otimes T^{(\mu)}$ to $a_\mu T^\mu$ and with this understood we write the decomposition into isotypical components as:

$$V = \oplus a_\mu T^\mu \tag{7.16}$$

8. Schur's Lemmas

♣This section needs improvement: Schur's lemma says that the algebra of intertwiners of an irrep is a division algebra. So $\mathbb{R}, \mathbb{C}, \mathbb{H}$. What is here is correct, but not the best viewpoint. ♣

When thinking about irreducible representations it is important to understand how they are related to each other. A key technical tool is Schur's lemma. It is usually stated as two separate statements, each of which is practically a triviality.

Theorem 1: If $A : V \to V'$ is an intertwiner between two irreps then it is either an isomorphism or zero.

Proof: ker$A$ and Im A are both invariant subspaces. Of course, if V and V' are inequivalent then $A = 0$. ♠.

Theorem 2: If A is a complex matrix commuting with an irreducible matrix representation then A is proportional to the identity matrix.

Proof: Since we are working over the complex field A has a nonzero eigenvector $Av = \lambda v$. The eigenspace $C = \{w : Aw = \lambda w\}$ is therefore not the zero vector space. But it is also an invariant subspace. Therefore, it must be the entire carrier space. ♠.

Remarks:

• Consider the isotypical decomposition of a completely reducible representation. By Schur's lemma, the intertwiners of $V^{(\mu)}$ with itself are of the form $K \otimes 1$ where $K \in \mathrm{End}(\mathbb{C}^{a_\mu})$ is arbitrary.

• Schur's lemma is used in quantum mechanics very often. Note that a Hamiltonian invariant under some symmetry group $U(g)HU(g)^{-1} = H$ is an intertwiner. Therefore, if we decompose the Hilbert space of states into irreps of G then H must be scalar on each irrep. In particular this leads to important *selection rules* in physics. For example, if ΔH is an invariant operator which commutes with some symmetry G of a physical system then the matrix elements of ΔH between states in different irreducible representations of G must vanish.

9. Orthogonality relations for matrix elements

Let G be a finite group. We are interested in finding the reduction of R_G to its irreducible representations. Our next technical tool - very useful in a variety of contexts - are the orthogonality relations of matrix elements of irreps.

Label the distinct, i.e. inequivalent, *irreducible representations* by

$$T^{(\mu)} \qquad \mu = 1, 2, \ldots, r \tag{9.1}$$

and let $n_\mu = \dim T^{(\mu)}$. (We will prove later that $r < \infty$ and $n_\mu < \infty$.) In order not to overburden the notation we will denote the carrier space and the homomorphism by the same name $T^{(\mu)}$, relying on context to tell which is intended. Choose bases for $T^{(\mu)}$ and consider these to be matrix irreps.

Let $B \in Mat_{n_\mu \times n_\nu}$, and define a matrix $A \in Mat_{n_\mu \times n_\nu}$ by:

$$A \equiv \frac{1}{|G|} \sum_{g \in G} T^{(\mu)}(g) B T^{(\nu)}(g^{-1}) \tag{9.2}$$

♣SHOULD GIVE A MORE CONCEPTUAL DESCRIPTION OF WHAT IS GOING ON HERE USING $\sum_g T^\vee(g) \otimes T(g)$. ♣

We claim that A is an intertwiner, that is, $\forall h \in G$:

$$\boxed{T^{(\mu)}(h) A = A T^{(\nu)}(h)} \tag{9.3}$$

Proof:

$$\begin{aligned} T^{(\mu)}(h)A &= \frac{1}{|G|}\sum_{g\in G} T^{(\mu)}(h)T^{(\mu)}(g)BT^{(\nu)}(g^{-1}) \\ &= \frac{1}{|G|}\sum_{g\in G} T^{(\mu)}(hg)BT^{(\nu)}(g^{-1}) \\ &= \frac{1}{|G|}\sum_{g'\in G} T^{(\mu)}(g')BT^{(\nu)}((h^{-1}g')^{-1}) \\ &= \frac{1}{|G|}\sum_{g\in G} T^{(\mu)}(g)BT^{(\nu)}((g)^{-1})T^{(\nu)}(h) \\ &= AT^{(\nu)}(h) \end{aligned} \tag{9.4}$$

In the third line let $g' = hg$, and use the rearrangement lemma.

Therefore, by Schur's lemma:

$$A = \lambda\delta_{\mu,\nu}\mathbf{1}_{n_\nu\times n_\nu} \tag{9.5}$$ eq:orthii

where λ can depend on $\mu = \nu, B$.

Now, to extract useful information from this let $B = e_{\ell m}$ be a matrix unit, $1 \le \ell \le n_\mu$, $1 \le m \le n_\nu$. Then (9.2)(9.5) imply:

$$\sum_{g\in G} T^{(\mu)}_{i\ell}(g)T^{(\nu)}_{ms}(g^{-1}) = |G|\lambda\delta_{\mu\nu}\delta_{is} \tag{9.6}$$ eq:orthiii

where now λ depends on m, ℓ, μ, but does *not* depend on i, s.

We can determine λ by setting $\mu = \nu$, $i = s$ and summing on i:

Then (9.6) becomes:

$$\begin{aligned} \sum_{g\in G} T^{(\mu)}_{m\ell}(g^{-1}g) &= |G|\lambda n_\mu \\ |G|\delta_{\ell m} &= |G|\lambda n_\mu \\ \Rightarrow \lambda &= \frac{\delta_{\ell m}}{n_\mu} \end{aligned} \tag{9.7}$$

Putting it all together we get the main result of this section:

Theorem The orthogonality relations for the group matrix elements are:

$$\boxed{\frac{1}{|G|}\sum_{g\in G} T^{(\mu)}_{i\ell}(g)\hat{T}^{(\nu)}_{sm}(g) = \frac{1}{n_\mu}\delta_{\mu\nu}\delta_{is}\delta_{\ell m}} \tag{9.8}$$ eq:orthogmatrx

or, for a general compact group:

$$\boxed{\int_G T^{(\mu)}_{i\ell}(g)\hat{T}^{(\nu)}_{sm}(g)[dg] = \frac{1}{n_\mu}\delta_{\mu\nu}\delta_{is}\delta_{\ell m}} \tag{9.9}$$ eq:orthogmatrx

where the Haar measure $[dg]$ is normalized to unit volume. Recall that

$$\hat{T}^{(\nu)}_{sm}(g) = (T^{(\nu)}(g))^{-1,tr}_{sm} \tag{9.10}$$

is the matrix representation on the dual space.

Note: If the matrix reps $T^{(\mu)}_{i\ell}$ are *unitary* then $T(g)^{tr,-1} = T(g)^*$ so we can rewrite this in a nicer form: Define

$$\phi^{(\mu)}_{i\ell} \equiv n_\mu^{1/2} T^{(\mu)}_{i\ell} \tag{9.11}$$

then we have:

$$\boxed{\frac{1}{|G|}\sum_{g\in G} \phi^{(\mu)}_{i\ell}(g)(\phi^{(\nu)}_{sm}(g))^* = \delta_{\mu\nu}\delta_{is}\delta_{\ell m}} \tag{9.12}$$ eq:onrels

Or, for a general compact group

$$\boxed{\int_G (\phi^{(\nu)}_{sm}(g))^* \phi^{(\mu)}_{i\ell}(g)[dg] = \delta_{\mu\nu}\delta_{is}\delta_{\ell m}} \tag{9.13}$$ eq:onrels

Interpretation: The vectors $\phi^{(\mu)}_{ij}$ form an *orthonormal set* of vectors in R_G.

Below we will apply this to the decomposition of the regular representation into irreps.

10. Decomposition of the Regular Representation

sec:sDRR

As before we let $T^{(\mu)}_{m\ell}(g)$ be matrix elements of the distinct irreps of G.

Now recall our important conceptual step forward: We may regard these matrix elements as complex-valued *functions* on G. That is $T^{(\mu)}_{m\ell}$ are *vectors* in R_G.

Moreover, as we showed in the example surrounding (7.9) the linear span of functions

$$\mathcal{R}^{(\mu)}_i \equiv Span\{T^{(\mu)}_{ij}\}_{j=1,\ldots,n_\mu} \tag{10.1}$$

is an invariant subspace of R_G, considered as the RRR. Here we are holding μ, i fixed. Indeed, we have:

$$\overbrace{R(g)\cdot T^{(\mu)}_{ij}}^{\text{function on G}} = \sum_{s=1}^{n_\mu} \underbrace{T^{(\mu)}_{sj}(g)}_{\text{matrix element for } R_G} \overbrace{T^{(\mu)}_{is}}^{\text{function on G}} \tag{10.2}$$

We learn two things:

1. $\mathcal{R}^{(\mu)}_i$ is an invariant subspace, as claimed.
2. Moreover, the matrix elements are those of the rep $T^{(\mu)}$! Thus:

$$\mathcal{R}^{(\mu)}_i \cong T^{(\mu)} \tag{10.3}$$

independent of the value of i!

It follows that R_G is *highly reducible*. Indeed it contains invariant subspaces corresponding to each and every irrep of G. In fact, since the above holds for each $i = 1, \ldots, n_\mu$ the same irrep occurs n_μ times, where $n_\mu = \dim T^{(\mu)}$.

Let us define

$$
\begin{aligned}
W &:= \oplus_\mu \oplus_{i=1}^{n_\mu} \mathcal{R}_i^{(\mu)} \subset R_G \\
W &\cong \oplus_\mu \Big(\overbrace{T^\mu \oplus \cdots \oplus T^\mu}^{n_\mu \ times} \Big)
\end{aligned}
\tag{10.4}
$$

Note:

• $W = Span\{T_{ij}^{(\mu)}\}$.

• By the orthogonality relations the $T_{ij}^{(\mu)}$ are linearly independent. Thus we really do have a direct sum .

Exercise

Show that the set of functions $T_{ij}^{(\mu)}$ with μ, j held fixed forms a copy of the irrep $\overline{T^{(\mu)}}$ under the left regular rep. More on this below.

At this point the possibility remains that W is a proper subspace of R_G. We will now show that in fact $W = R_G$ exactly, by showing that $T_{ij}^{(\mu)}$ span R_G.

To do this recall the vector space R_G is an inner product space by the rule

$$
(\phi_1, \phi_2) \equiv \frac{1}{|G|} \sum_{g \in G} \phi_1^*(g) \phi_2(g) \tag{10.5}
$$

eq:innproda

Recall:

1. R_G is a *unitary* representation, wrt this product.

2. The normalized matrix elements ϕ_{ij}^μ are orthonormal wrt this inner product. This follows from (9.13).

By the proposition 7.2.1 we can use the inner product to decompose

$$
R_G \cong W \oplus W^\perp \tag{10.6}
$$

so $W^\perp$ is also a unitary representation. Moreover, again by 7.2.1 it can be reduced to a sum of irreducible representations, $T^{(\mu)}$.

Suppose

$$
Span\{f_i(g)\}_{i=1,\ldots,n_\mu} \subset W^\perp \tag{10.7}
$$

is a subspace transforming in a rep equivalent to T^μ. Then, by definition:

$$
\begin{aligned}
(R(h)f_i)(g) &= \sum_s T^\mu_{si}(h) f_s(g) \\
&\Rightarrow \\
f_i(gh) &= \sum_s T^\mu_{si}(h) f_s(g) \\
&\Rightarrow \\
f_i(g') &= \sum_s T^\mu_{si}(g^{-1}g') f_s(g) \\
&= \sum_{s,k} T^\mu_{sk}(g^{-1}) T^\mu_{ki}(g') f_s(g)
\end{aligned}
\tag{10.8}
$$

where we have set $g' = gh$.

Therefore, we have the equality of functions on G:

$$
\begin{aligned}
f_i(\cdot) &= \sum_k \alpha_k T^\mu_{ki}(\cdot) \\
\alpha_k &= \sum_s \Big(T^\mu_{sk}(g^{-1}) f_s(g) \Big) \qquad \text{any} \qquad g \in G
\end{aligned}
\tag{10.9}
$$

That is, f_i is a linear combination of the functions ϕ^μ_{ij}. But this contradicts the assumption that

$$
\{f_i\} \subset W^\perp \tag{10.10}
$$

so $W^\perp = 0$. ♠.

Thus, we have arrived at the decomposition of the RRR into its irreducible pieces:

$$
R_G \cong \oplus_\mu \Big(\overbrace{T^\mu \oplus \cdots \oplus T^\mu}^{n_\mu \ times} \Big) \tag{10.11}
$$

eq:rrrep

But we are not quite done. To arrive at the most beautiful statement we should study R_G as a representation under the the left regular rep. So we consider the linear subspaces where we hold μ, j fixed and consider the span:

$$
\mathcal{L}^{(\mu)}_j \equiv Span\{\phi^{(\mu)}_{ij}\}_{i=1,\ldots,n_\mu} \tag{10.12}
$$

This forms a copy of the dual rep:

$$
\begin{aligned}
(L(g)\cdot \phi^{(\mu)}_{ij})(h) &= \sum_s T^{(\mu)}_{is}(g^{-1}) \phi^{(\mu)}_{sj}(h) \\
&= \sum_s \hat{T}^{(\mu)}_{si}(g) \phi^{(\mu)}_{sj}(h)
\end{aligned}
\tag{10.13}
$$

so the span of the matrix elements of the rep μ

$$
Span\{\phi^{(\mu)}_{ij}\}_{i,j=1,\ldots,n_\mu} \tag{10.14}
$$

forms the rep $\hat{T}^{(\mu)} \otimes T^{(\mu)}$ of $G_L \times G_R$. Finally, we can identify

$$\hat{T}^{(\mu)} \otimes T^{(\mu)} \cong \mathrm{End}(T^{(\mu)}) \tag{10.15}$$

This is a representation of $G_L \times G_R$ of dimension n_μ^2 with action

$$T(g_L, g_R) \cdot \Phi = T^{(\mu)}(g_L) \Phi T^{(\mu)}(g_R^{-1}) \tag{10.16}$$

We have finally arrived at the beautiful :

Theorem 10.1 Let G be a finite group. The matrix elements ϕ_{ij}^μ of the distinct irreps of G decompose R_G into irreps of $G_{\mathrm{Left}} \times G_{\mathrm{Right}}$:

$$R_G \cong \oplus_\mu \hat{T}^{(\mu)} \otimes T^{(\mu)} \cong \oplus_\mu \mathrm{End}(T^{(\mu)}) \tag{10.17}$$

Moreover, if we use matrix elements of unitary irreps in an ON basis then the ϕ_{ij}^μ form an ON basis for R_G and

$$R_G \cong \oplus_\mu \overline{T^{(\mu)}} \otimes T^{(\mu)} \tag{10.18}$$

Corollary. Comparing dimensions we have the beautiful relation:

$$|G| = \sum_\mu n_\mu^2 \tag{10.19}$$

eq:grouporder

Remarks:

• This theorem generalizes beautifully to all compact groups and is known as the Peter-Weyl theorem.

• This leaves the separate question of actually *constructing* the representations of the finite group G. Until recently there has been no general algorithm for doing this.

Recently there has been a claim that it can be done. See

Vahid Dabbaghian-Abdoly, "An Algorithm for Constructing Representations of Finite Groups." Journal of Symbolic Computation Volume 39, Issue 6, June 2005, Pages 671-688

Exercise

Show that the number of distinct one-dimensional representations of a finite group G is the same as the index of the commutator subgroup $[G, G]$ in G. [4]

********** **********

MATERIAL FROM MATHMETHODS 511 2015:

For now, we just content ourselves with the statement of the theorem for G a finite group:

[4] Hint: A one-dimensional representation is trivial on $[G, G]$ and hence descends to a representation of the abelianization of G, namely the quotient group $G/[G, G]$.

Theorem: Let G be a finite group, and define an Hermitian inner product on $L^2(G) = \mathrm{Map}(G, \mathbb{C})$ by

$$(\Psi_1, \Psi_2) := \frac{1}{|G|} \sum_g \Psi_1^*(g)\Psi_2(g) \tag{10.20}$$

Then let $\{V_\lambda\}$ be a set of representatives of the distinct isomorphism classes of irreducible unitary representations for G. For each representation V_λ choose an ON basis $w_\mu^{(\lambda)}$, $\mu = 1, \ldots, n_\lambda := \dim_{\mathbb{C}} V_\lambda$. Then the matrix elements $D^\lambda_{\mu\nu}(g)$ defined by

$$\rho(g) w_\nu^{(\lambda)} = \sum_{\mu=1}^{n_\lambda} D^\lambda_{\mu\nu}(g) w_\mu^{(\lambda)} \tag{10.21}$$

form a complete orthogonal set of functions on $L^2(G)$ so that

$$(D^{\lambda_1}_{\mu_1\nu_1}, D^{\lambda_2}_{\mu_2\nu_2}) = \frac{1}{n_\lambda} \delta^{\lambda_1,\lambda_2} \delta_{\mu_1,\mu_2} \delta_{\nu_1,\nu_2} \tag{10.22}$$ eq:PW-2

Idea of proof: The proof is based on linear algebra and Schur's lemma. The normalization constant on the RHS of (10.22) is easily determined by setting $\lambda_1 = \lambda_2$ and $\nu_1 = \nu_2 = \nu$ and summing on ν, and using the hypothesis that these are matrix elements in a *unitary* representation. The relation to (6.22) is obtained by noting that the linear transformations $T = e_{\nu\mu}$, given by matrix units relative to the basis $w_\mu^{(\lambda)}$ form a basis for $\mathrm{End}(V_\lambda)$. ♠

Example 1: Let $G = \mathbb{Z}_2 = \{1, \sigma\}$ with $\sigma^2 = 1$. Then the general complex valued-function on G is specified by two complex numbers $(\psi_+, \psi_-) \in \mathbb{C}^2$:

$$\Psi(1) = \psi_+ \qquad \Psi(\sigma) = \psi_- \tag{10.23}$$

This identifies $\mathrm{Map}(G, \mathbb{C}) \cong \mathbb{C}^2$ as a vector space. There are just two irreducible representations $V_\pm \cong \mathbb{C}$ with $\rho_\pm(\sigma) = \pm 1$, because for any representation $\rho(\sigma)$ on a vector space V we can form orthogonal projection operators $P_\pm = \frac{1}{2}(1 \pm \rho(\sigma))$ onto direct sums of the irreps. They are obviously unitary with the standard Euclidean norm on $\mathbb{C}$. The matrix elements give two functions on the group $D^\pm$:

$$D^+(1) = 1 \qquad D^+(\sigma) = 1 \tag{10.24}$$

$$D^-(1) = 1 \qquad D^-(\sigma) = -1 \tag{10.25}$$

(Here and in the next examples when working with 1×1 matrices we drop the $\mu\nu$ subscript!) The reader can check they are orthonormal, and they are complete because any function Ψ can be expressed as:

$$\Psi = \frac{\psi_+ + \psi_-}{2} D^+ + \frac{\psi_+ - \psi_-}{2} D^- \tag{10.26}$$

Example 2: We can generalize the previous example slightly by taking $G = \mathbb{Z}/n\mathbb{Z} = \langle \omega | \omega^n = 1 \rangle$. Let us identify this group with the group of nth roots of unity and choose

a generator $\omega = \exp[2\pi \mathrm{i}/n]$. Since G is abelian all the representation matrices can be simultaneously diagonalized so all the irreps are one-dimensional. They are:

$V = \mathbb{C}$ and $\rho_m(\omega) = \omega^m$ where m is an integer. Note that $m \sim m + n$ so the set of irreps is again labeled by $\mathbb{Z}/n\mathbb{Z}$ and in fact, under tensor product the set of irreps itself forms a group isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

The matrix elements in the irrep (ρ_m, V) are

$$D^{(m)}(\omega^j) = \omega^{mj} = e^{2\pi \mathrm{i} \frac{mj}{n}} \tag{10.27}$$

Now we can check that indeed

$$\frac{1}{|G|} \sum_{g \in G} (D^{(m_1)}(g))^* D^{(m_2)}(g) = \delta_{m_1 - m_2 = 0 \mathrm{mod} n} \tag{10.28}$$

The decomposition of a function Ψ on the group G is known as the discrete Fourier transform.

Remark: The theorem applies to all compact Lie groups. For example, when $G = U(1) = \{z || z| = 1\}$ then the invariant measure on the group is just $-\mathrm{i}\frac{dz}{z} = \frac{d\theta}{2\pi}$ where $z = e^{\mathrm{i}\theta}$:

$$(\Psi_1, \Psi_2) = \int_0^{2\pi} \frac{d\theta}{2\pi} (\Psi_1(\theta))^* \Psi_2(\theta) \tag{10.29}$$

Now, again since G is abelian the irreducible representations are 1-dimensional and the unitary representations are (ρ_n, V_n) where $n \in \mathbb{Z}$, $V_n \cong \mathbb{C}$ and

$$\rho_n(z) := z^n \tag{10.30}$$

Now, the orthonormality of the matrix elements is the standard orthonormality of $e^{\mathrm{i}n\theta}$ and the Peter-Weyl theorem specializes to Fourier analysis: An L^2-function $\Psi(\theta)$ on the circle can be expanded in terms of the matrix elements of the irreps:

$$\Psi = \sum_{\text{Irreps } \rho_n} \hat{\Psi}_n D^{(n)} \tag{10.31}$$

When applied to $G = SU(2)$ the matrix elements are known as *Wigner functions* or *monopole harmonics*. They are the matrix elements

$$D^j_{m_L, m_R}(g) := \langle m_L | \rho^j(g) | m_R \rangle \tag{10.32}$$

in the standard ON basis of the unitary spin j representation diagonalizing the diagonal subgroup of $SU(2)$. So

$$j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots, \qquad m_L, m_R \in \{-j, -j+1, \ldots, j-1, j\} \tag{10.33}$$

Recall that $SU(2) \cong S^3$ as a manifold. Using the standard volume form, with unit volume we can define $L^2(SU(2))$. The entire theory of spherical harmonics and Legendre polynomials is easily derived from basic group theory.

♣MAKE GOOD ON THIS CLAIM BELOW. ♣

********** **********

11. Fourier Analysis as a branch of Representation Theory

11.1 The irreducible representations of abelian groups

Suppose G is abelian. Let T be an irrep of G. Fix $g_0 \in G$. Then

$$T(g_0)T(g) = T(g_0 g) = T(g g_0) = T(g)T(g_0) \tag{11.1}$$

Therefore, by Schur's lemma:

$$T(g_0) = \lambda(g_0)1 \tag{11.2}$$

Therefore, all representation matrices of irreducible representations are proportional to $1 \Rightarrow$

All finite-dimensional irreducible complex representations of an abelian group are 1-dimensional.

This is just simultaneous diagonalization of commuting operators. Note that the adjective "complex" is essential here. Consider $SO(2)$ acting on $\mathbb{R}^2$.

As we mentioned before noncompact groups can have indecomposable representations, e.g. $\mathbb{R}$ has the representation

$$x \to \begin{pmatrix} x & 1 \\ 0 & x \end{pmatrix} \tag{11.3}$$

Note that this is not unitary.

11.2 The character group

Notice that the space of 1 dimensional irreps of any group G is *itself* a group under $\otimes$ product:

$$T_1 \otimes T_2(g) = T_1(g)T_2(g) \tag{11.4}$$

The identity is the trivial representation $T_0(g) = 1$, and inverses exist

$$(T)^{-1}(g) = (T(g))^{-1} \tag{11.5}$$

By the same reasoning the space of 1-dimensional unitary irreps of a group is a group. Note that in this case $(T)^{-1}(g) = T^*(g)$ and $T(g)$ is always represented by a phase.

Definition The space of unitary irreps of a group G is called the *unitary dual* and is denoted by $\hat{G}$.

In the case when G is abelian, the unitary dual $\hat{G}$ is itself a group. It is sometimes called the *Pontryagin dual group.*

Examples

• Finally, let us consider $\mathbb{Z}_n$ thought of as multiplicative n^{th} roots of unity. For any integer s we can define a representation by taking

$$T^{(s)}(\omega) = \omega^s \tag{11.6}$$ eq:srep

where ω is any n^{th} root of unity. These are clearly unitary irreps. Note that

$$T^{(s+n)} = T^{(s)} \tag{11.7}$$

and

$$T^{(s_1)} \otimes T^{(s_2)} \cong T^{(s_1+s_2)} \tag{11.8}$$

so the dual group is another copy of $\mathbb{Z}_n$:

$$\widehat{\mathbb{Z}_n} \cong \mathbb{Z}_n \tag{11.9}$$

Note that the relation (10.19) is easily verified:

$$n = 1^2 + 1^2 + \cdots + 1^2 \tag{11.10}$$

• The theory we are discussing in this section extends to locally compact abelian groups such as $G = \mathbb{R}$. The unitary irreps are specified by a "momentum" $k \in \mathbb{R}$.

$$T^{(k)} : x \to e^{ikx} \tag{11.11}$$

This is a rep because $T^{(k)}(x+y) = T^{(k)}(x)T^{(k)}(y)$. It is unitary for k real.

Now notice that

$$T^{(k)} \otimes T^{(k')} = T^{(k+k')} \tag{11.12}$$

so we conclude:

$$\widehat{\mathbb{R}} \cong \mathbb{R} \tag{11.13}$$

• $G = \mathbb{Z}$. Unitary irreps of $\mathbb{Z}$ are labelled by a real number θ:

$$T^{(\theta)} : n \to e^{in\theta} \tag{11.14}$$

Note that

$$\begin{aligned} T^{(\theta)} &= T^{(\theta+2\pi)} \\ T^{(\theta)} \otimes T^{(\theta')} &= T^{(\theta+\theta')} \end{aligned} \tag{11.15}$$

eq:unirri

So the *space of representations* is parametrized by θ mod 2π, or, equivalently by $e^{i\theta} \in U(1)$. Moreover (11.15) shows that

$$\widehat{\mathbb{Z}} \cong U(1) \tag{11.16}$$

eq:dlgrp

• $G = \mathbb{R}/\mathbb{Z} \cong U(1)$. Now the unitary irreps are labelled by $n \in \mathbb{Z}$:

$$T^{(n)} : e^{i\theta} \to e^{2\pi i n\theta} \tag{11.17}$$

eq:unirrii

and we check:

$$T^{(n)} \otimes T^{(n')} = T^{(n+n')} \tag{11.18}$$

eq:unirriii

so

$$\widehat{U(1)} \cong \mathbb{Z} \tag{11.19}$$

eq:dlgrpp

• $G = \Lambda \subset \mathbb{R}^d$, a lattice. Given a vector $\vec{k} \in \mathbb{R}^d$ we can define a unitary irrep by:

$$T^{(\vec{k})} : \vec{n} \to e^{2\pi i \vec{k}\cdot\vec{n}} \tag{11.20}$$

where we use the standard inner product on $\mathbb{R}^d$.

As before $T^{(\vec{k})} \otimes T^{(\vec{k}')} = T^{(\vec{k}+\vec{k}')}$. Note however that

$$T^{(\vec{k}+\vec{b})} = T^{(\vec{k})} \tag{11.21}$$

if $\vec{b}$ is in the dual lattice:

$$\Lambda^* = \{\vec{b} \in \mathbb{R}^d | \forall \vec{n} \in \Lambda, \vec{b}\cdot\vec{n} \in \mathbb{Z}\} \tag{11.22}$$

Thus, the space of unitary irreps of Λ is a compact d-dimensional torus:

$$\widehat{\Lambda} \cong \mathbb{R}^d/\Lambda^* \tag{11.23}$$ eq:torirep

thus generalizing $\widehat{\mathbb{Z}} = U(1)$.

• . $G = \mathbb{R}^d/\Lambda$ is a d-dimensional torus.

Remarks Note that in all the above examples,

$$\widehat{\widehat{G}} = G \tag{11.24}$$

This is a general fact for abelian groups.

11.3 Fourier duality

Let us now return to Theorem 10.1 decomposing R_G in terms of the irreducible representations through the matrix elements of unitary irreps and apply it to finite abelian groups:

• $G = \mathbb{Z}_2 = \{1, \sigma\}$. A function $f : G \to \mathbb{C}$ is specified by its two values

$$f(1) = f_0 \qquad f(\sigma) = f_1 \tag{11.25}$$

The unitary irreps are just:

$$\begin{aligned} T^{(0)}(1) &= 1 & T^{(0)}(\sigma) &= 1 \\ T^{(1)}(1) &= 1 & T^{(1)}(\sigma) &= -1 \end{aligned} \tag{11.26}$$

Clearly:

$$f = \frac{f_0 + f_1}{2} T^{(0)} + \frac{f_0 - f_1}{2} T^{(1)} \tag{11.27}$$

Let us generalize this:

•

$G = \mathbb{Z}_n \cong \{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$ where $\omega = e^{2\pi i/n}$. We worked out the unitary irreps above.

The orthogonality relations on the irreducible matrix elements (11.6) flow from:

$$\begin{aligned}\frac{1}{n}\sum_{\ell} T^{(j)}(\omega^\ell)(T^{(k)}(\omega^\ell))^* &= \frac{1}{n}\sum_{\ell} e^{2\pi i(j\ell-k\ell)/n} \\ &= \delta_{j,k}\end{aligned} \tag{11.28}$$

where j, k are understood modulo n.

Now, according to the theorem, we can decompose any function on $G = \mathbb{Z}_n$ as:

$$\boxed{f = \sum_{j=0}^{n-1} \hat{f}_j T^{(j)}} \tag{11.29}$$

and we see that

$$\begin{aligned}\hat{f}_j &= \frac{1}{n}\sum_{s} f(\omega^s)(T^{(j)}(\omega^s))^* \\ &= \frac{1}{n}\sum_{s} f(\omega^s) e^{-2\pi i s j/n}\end{aligned} \tag{11.30}$$

is the *discrete Fourier transform.*

In fact, our discussion generalizes to all (locally compact[5] abelian groups including infinite groups like $\mathbb{R}, \mathbb{Z}, U(1)$. The main new ingredient we need is *integration over the group*, which is the analog of $\frac{1}{|G|}\sum_g$.

For these groups the integration of a function $f \in Fun(G)$ is the obvious one:

$$\begin{aligned}\int_{G=U(1)} f(g)dg &\equiv \int_0^{+2\pi} \frac{d\theta}{2\pi} f(\theta) \\ \int_{G=\mathbb{Z}} f(g)dg &\equiv \sum_{n\in\mathbb{Z}} f(n) \\ \int_{G=\mathbb{R}} f(g)dg &\equiv \int_{-\infty}^{+\infty} dx f(x)\end{aligned} \tag{11.31}$$

eq:intgroupp

With this in mind the above Theorem 10.1 becomes the following statements:

• Let us begin with $G = (\mathbb{R}, +)$ then $\hat{G} = (\mathbb{R}, +)$. Unitary irreps are labelled by $k \in \mathbb{R}$:

$$T^{(k)}(x) = \frac{1}{\sqrt{2\pi}} e^{2\pi i k x} \tag{11.32}$$

The orthogonality relations are:

$$\int_{\mathbb{R}} dx T^{(k)}(x)(T^{(\ell)}(x))^* = \delta(k-\ell) \tag{11.33}$$

[5] A Hausdorff space is called *locally compact* if every point has a compact neighborhood. All the abelian groups we will meet in this course are locally compact. **** GIVE AN EXAMPLE OF A non-locally-compact abelian group ***

The statement that any L^2 function on the group $\mathbb{R}$ can be expanded in the matrix elements of its unitary irreps is that statement that any L^2 function $f(x)$ can be expanded:

$$\boxed{f(x) = \int_{\widehat{\mathbb{R}}} dk \hat{f}(k) T^{(k)}(x)} \tag{11.34}$$

So we recognize the usual Fourier expansion.

Of course, since $\widehat{\widehat{\mathbb{R}}} \cong \mathbb{R}$ we have the inverse relation:

$$\hat{f}(k) = \int_{\mathbb{R}} dx f(x) \widehat{T}^{(x)}(k) \tag{11.35}$$

Of course, there remarks generalize to $\mathbb{R}^d$.

•

We showed that $\mathbb{Z}$ and $U(1)$ are dual groups. Thus, an arbitrary (L^2) function $f(\theta)$ on $G = U(1)$, can be decomposed into a sum of irreps of $U(1)$:

$$f(\theta) = \sum_{n \in \mathbb{Z}} \hat{f}(n) T^{(n)}(\theta) \tag{11.36}$$ eq:foursum

This is the usual Fourier series. and we recognize the Fourier decomposition. The orthogonality relations on matrix elements read:

$$\int_0^{2\pi} \frac{d\theta}{2\pi} (T^{(n)}(\theta))^* T^{(m)}(\theta) = \delta_{n,m} \tag{11.37}$$ eq:orthogrels

Conversely, an arbitrary L^2 function $\hat{f}(n)$ on the abelian group $G = \mathbb{Z}$ can be decomposed into a sum over the irreps $T^{(\theta)}$ of $\mathbb{Z}$, so we have the expansion:

$$\hat{f}(n) = \int_0^{2\pi} \frac{d\theta}{2\pi} f(\theta) T^{(\theta)}(n) \tag{11.38}$$ eq:invfour

The orthogonality relations read:

$$\begin{aligned} \sum_{n \in \mathbb{Z}} (T^{(\theta_1)}(n))^* T^{(\theta_2)}(n) &= \delta_{\text{periodic}}(\theta_1 - \theta_2) \\ &\equiv \sum_{m \in \mathbb{Z}} \delta(\theta_1 - \theta_2 - 2\pi m) \end{aligned} \tag{11.39}$$ eq:orthogrelsa

In summary, we have the

Theorem Let G be a locally compact group. Then there is a natural isomorphism $L^2(G) \cong L^2(\hat{G})$ which expands any function on G in terms of the unitary irreps of G:

$$f(g) := \int_{\hat{G}} d\chi \chi(g)^* \hat{f}(\chi) \tag{11.40}$$ eq:svnthm

Moreover, this isomorphism is an isometry:

$$\int_G dg f_1^* f_2 = \int_{\hat{G}} d\chi \hat{f}_1^* \hat{f}_2 \tag{11.41}$$

Remarks

• The functional analysis of $L^2(G)$ is often called *"nonabelian harmonic analysis."* For more about this point of view see,

A.A. Kirillov, *Elements of the Theory of Representations*, Springer-Verlag, 1976.

11.4 The Poisson summation formula

Let us return to the orthogonality relations for the irreps of Λ and $\mathbb{R}^d/\Lambda^*$.

These imply a beautiful formula known as the Poisson summation formula.

Suppose $f \in \mathcal{S}(\mathbb{R})$ (the Schwarz space of functions of rapid decrese for $\mid x \mid \to \infty$). Define the Fourier transform:

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(y) e^{-iky} dy \ . \tag{11.42}$$

Then:

Theorem [Poisson summation formula].

$$\sum_{n \in \mathbb{Z}} f(n) = \sum_{n \in \mathbb{Z}} \hat{f}(2\pi n) \tag{11.43}$$

Proof: The orthogonality relations tell us that

$$\boxed{\sum_{n \epsilon \mathbb{Z}} e^{2\pi i n t} = \sum_{k \epsilon \mathbb{Z}} \delta(k-t)} \tag{11.44}$$ eq:basicpsf

Now integrate both sides against $f(t)$. ♠

Remarks

• If it possible to give other proofs of this formula. We will content ourselves with its being a consequence of the orthogonality relations.

• The rapid decrease condition can be relaxed to

$$\mid f(x) \mid \ + \ \mid f'(x) \mid \ + \ \mid f''(x) \mid \ \leq \frac{const}{1+x^2} \tag{11.45}$$

Remark The above result has a beautiful generalization to arbitrary lattices in Euclidean space $\mathbb{R}^d$. The generalized Poisson summation formula states that:

$$\boxed{\sum_{\vec{v} \in \Lambda} e^{2\pi i \vec{v} \cdot \vec{x}} = \sum_{\vec{l} \in \Lambda^*} \delta^{(d)}(\vec{x} - \vec{l})} \tag{11.46}$$

or - in terms or functions

$$\sum_{\vec{n}\in Z^d} f(\vec{n}) = \sum_{\vec{m}\in Z^d} \int_{R^d} e^{-2\pi i \vec{m}\cdot\vec{t}} f(t)dt = \sum_{\vec{m}\in Z^d} \int_{R^d} e^{+2\pi i \vec{m}\cdot\vec{t}} f(t)dt \tag{11.47}$$ eq:poissonsumm

That is:

$$\sum_{\vec{v}\in\Lambda} f(\vec{v}) = \sum_{\vec{l}\in\Lambda^*} \hat{f}(\vec{l}) \ . \tag{11.48}$$

Remarks

• Since people have different conventions for the factors of 2π in Fourier transforms it is hard to remember the factors of 2π in the PSF. The equation (11.44) has no factors of 2π. One easy way to see this is to integrate both sides from $t = -1/2$ to $t = +1/2$.

• One application of this is the x-ray crystallography: The LHS is the sum of scattered waves. The RHS constitutes the bright peaks measured on a photographic plate.

• Another application is in the theory of elliptic functions and θ-functions described later.

Exercise

a.) Show that

$$\sum_{n\in Z} e^{-\pi a n^2 + 2\pi i b n} = \sqrt{\frac{1}{a}} \sum_{m\in Z} e^{-\frac{\pi(m-b)^2}{a}} \tag{11.49}$$ eq:formtwo

b.) If τ is in the upper half complex plane, and θ, ϕ, z are complex numbers define the *theta function*

$$\vartheta_{\theta,\phi} z = \sum_{n\in\mathbb{Z}} e^{i\pi\tau(n+\theta)^2 + 2\pi i (n+\theta)(z+\phi)} \tag{11.50}$$ eq:thefun

(usually, θ, ϕ are taken to be real numbers, but z is complex).

Using the Poisson summation formula show that

$$\vartheta[\begin{matrix}\theta\\ \phi\end{matrix}](\frac{-z}{\tau}|\frac{-1}{\tau}) = (-i\tau)^{1/2} e^{2\pi i\theta\phi} e^{i\pi z^2/\tau} \vartheta[\begin{matrix}-\phi\\ \theta\end{matrix}](z|\tau) \tag{11.51}$$ eq:esstmn

---

11.5 Application: Bloch's Theorem in Solid State Physics

Consider an electron moving in a crystalline material. Atoms or ions live on a lattice

$$\Lambda \subset \mathbb{R}^3 \tag{11.52}$$

and an electron (in the single-electron approximation) satisfies the Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V(\vec{x})\psi = E\psi. \tag{11.53}$$

Now, $V(\vec{x})$ will be complicated in general, but we do know that

$$V(\vec{x}+\vec{a}) = V(\vec{x}) \tag{11.54}$$

$\forall \vec{a} \in \Lambda$. Therefore the energy eigenspace $\mathcal{V}_E$ is some representation of the lattice translation group Λ.

As we have seen, the *one dimensional* reps of Λ are

$$\vec{a} \to e^{2\pi i \vec{k}\cdot\vec{a}} \tag{11.55}$$

and are labeled by $\vec{k} \in \hat{\Lambda} = \mathbb{R}^n/\Lambda^*$. Therefore, $\mathcal{V}_E$ can be written as a sum of reps labeled by $\vec{k}$:

$$\mathcal{V}_E = \oplus_{\vec{k}\in\hat{\Lambda}} \mathcal{V}_{E,\vec{k}} \tag{11.56}$$ eq:decompvx

where eigenfunctions in $\mathcal{V}_{E,\vec{k}}$ satisfy:

$$\psi(\vec{x}+\vec{a}) = e^{2\pi i \vec{k}\cdot\vec{a}}\psi(\vec{x}) \tag{11.57}$$

that is, $\psi(\vec{x}) = e^{2\pi i\vec{k}\cdot\vec{x}}u(\vec{x})$ where u is periodic. $\vec{k}$, the "crystal momentum," is only defined modulo Λ^*. It is properly an element of a torus.

Since we are working with infinite groups the sum on $\vec{k}$ in (11.56) in general must be interpreted as an integral.

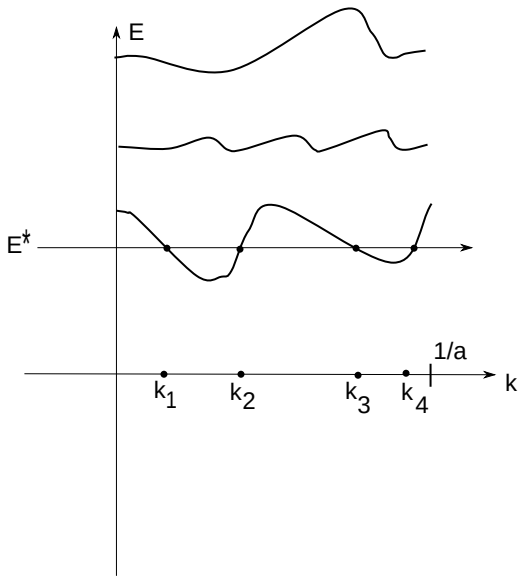


Figure 3: Four irreps occur at energy E_* in this one-dimensional band structure. For $d = 2, 3$ dimensional bandstructures there are infinity many k's forming a surface. The surface at the Fermi energy is the "Fermi surface." fig:bandstruct

Remarks

• If we substitute $\psi(\vec{x}) = e^{2\pi i \vec{k}\cdot\vec{x}} u(\vec{x})$ into the Schrödinger equation we obtain an elliptic eigenvalue problem on a compact space. It will therefore have a tower of discrete eigenvalues. This leads to the band structure illustrated schematically in 3.

• In solid state physics the dual lattice is referred to as the *reciprocal lattice*. Physicists generally choose a fundamental domain for the torus $\mathbb{R}^d/\Lambda^*$ and refer to it as a *Brillouin zone.*

• If the lattice Λ has a symmetry group then it will act on the torus. At fixed points there is enhanced symmetry and bands will cross.

11.6 The Heisenberg group extension of $\hat{S} \times S$ for an abelian group S

******

Change notation $S \to G$ below. Also need to explain that $G \times \hat{G} \to U(1)$ is a *perfect pairing.*

******

Let S be a locally compact abelian group with a measure. Let $\hat{S}$ be the dual group. There is a very natural Heisenberg extension of $S \times \hat{S}$ defined by the cocycle to choose the cocycle

$$c\big((s_1,\chi_1),(s_2,\chi_2)\big) = \frac{1}{\chi_1(s_2)} \tag{11.58}$$ eq:cocyclechoi

whose antisymmetrization is

$$s\big((s_1,\chi_1),(s_2,\chi_2)\big) = \frac{\chi_2(s_1)}{\chi_1(s_2)}. \tag{11.59}$$ eq:cocyclechoi

This gives

$$1 \to U(1) \to \mathrm{Heis}(S \times \hat{S}) \to S \times \hat{S} \to 1 \tag{11.60}$$ eq:heisext

There are two very natural representations of this group. First we consider $\mathcal{H} = L^2(S)$.

First of all $L^2(S)$ is a representation of S. After all for $s_0 \in S$ we can define the translation operator:

$$(T_{s_0}\psi)(s) := \psi(s + s_0). \tag{11.61}$$ eq:transop

On the other hand, $\mathcal{H}$ is also a representation of the Pontrjagin dual group of characters, denoted $\hat{S}$. If $\chi \in \hat{S}$ is a character on S then we define the multiplication operator M_χ on $\mathcal{H}$ via

$$(M_\chi\psi)(s) := \chi(s)\psi(s). \tag{11.62}$$ eq:multop

Note that $\mathcal{H}$ is *not* a representation of $S \times \hat{S}$. This simply follows from the easily-verified relation

$$T_{s_0} M_\chi = \chi(s_0) M_\chi T_{s_0}. \tag{11.63}$$ eq:noncomm

However, $\mathcal{H}$ *is* a representation of $\mathrm{Heis}(S \times \hat{S})$. It is defined so that

$$(z,(s,\chi)) \to zT_sM_\chi \tag{11.64}$$ eq:heishom

where $z \in U(1)$. Equation (11.64) defines a homomorphism into the group of invertible operators on $L^2(S)$.

Notice the complete symmetry in the construction between S and $\hat{S}$ (since the double-dual gives S again). Thus, we could also provide a representation from $\hat{\mathcal{H}} = L^2(\hat{S})$. The two representations are in fact equivalent under Fourier transform:

$$\hat{\psi}(\chi) := \int ds \chi(s)^* \psi(s) \tag{11.65}$$ eq:svnthmi

In fact, a theorem, the Stone-von Neumann theorem guarantees the essential uniqueness of the unitary representations of Heisenberg groups:

Theorem If $\tilde{A}$ is a central extension of a locally compact topological abelian group A by $U(1)$, then the unitary irreps of $\tilde{A}$ where $U(1)$ acts by scalar multiplication are in $1-1$ correspondence with the unitary irreps of the center of $\tilde{A}$, where $U(1)$ acts by scalar multiplication.

Example 1 Consider $L^2(\mathbb{Z}_n)$. Regarding $\mathbb{Z}_n$ as the integers modulo n we have

$$(T(\bar{s}_0)\cdot f)(\bar{s}) = f(\bar{s}+\bar{s}_0) \tag{11.66}$$

while the dual group $\chi_{\bar{t}}$ has character $\chi_{\bar{t}}(\bar{s}) = e^{2\pi i s t/n}$ acts by multiplication.

Example 2 $S = \mathbb{R}$ and $\hat{S} = \mathbb{R}$. Denote elements $q \in \mathbb{R}$ and $p \in \widehat{\mathbb{R}} \cong \mathbb{R}$.

$$(T(p)\cdot\psi)(q) = e^{2\pi i p q \theta}\psi(q) \tag{11.67}$$

$$(T(q_0)\cdot\psi)(q) = \psi(q+q_0) \tag{11.68}$$

Then

$$(T(q_0)(T(p)\cdot\psi))(q) = e^{2\pi i\theta p(q+q_0)}\psi(q+q_0) \tag{11.69}$$

$$(T(p)(T(q_0)\cdot\psi))(q) = e^{2\pi i\theta pq}\psi(q+q_0) \tag{11.70}$$

So $T(q_0)T(p) = e^{2\pi i p q_0}T(p)T(q_0)$. We recognize standard relations from quantum mechanics with θ playing the role of Planck's constant.

Example 3 $S = \mathbb{Z}$ and $\hat{S} = U(1)$ ... This generalizes to the duality between lattices and tori.

Remarks

• Most free field theories can be interpreted in this framework, with proper regard for the infinite-dimensional aspects of the problem. There is the interesting exception of self-dual field theories which are more subtle.

12. Induced Representations

♣MATERIAL HERE IMPORTED FROM GMP 2010 ASSUMES KNOWLEDGE OF BUNDLES. NEED TO EDIT THOSE PARTS AWAY. ♣

Induced representations are certain representations defined by taking sections of associated bundles. They form a unifying theme for a broad class of representations in physics:

1. Finite groups
2. Compact groups: The Borel-Weil-Bott Theorem
3. Lorentz groups: The Wigner construction.
4. Representations of Heisenberg groups and coherent states
5. Representations of Loop Groups

Suppose $H \subset G$ is a subgroup.

Suppose $\rho : H \to GL(V)$ is a representation of H with representation space V.

Note that

$$\begin{array}{ccc} H & \to & G \\ & & \downarrow \pi \\ & & G/H \end{array} \tag{12.1}$$

is a principal H bundle with total space $P = G$.

Therefore, according to what we have said above, using (ρ, V) we can form the associated vector bundle

$$G \times_H V \tag{12.2}$$

Now, the key point is that $\Gamma(G \times_H V)$ *is also a representation of* G. This representation is called the *induced representation* and sometimes denoted:

$$\Gamma(G \times_H V) = \mathrm{Ind}_H^G(V) \tag{12.3}$$

Using the grand tautology (**??**) we can equivalently define the induced representation as:

Definition Let (V, ρ) be a representation of H. As a vector space, $\mathrm{Ind}_H^G(V)$ is the set of all smooth functions $\psi : G \to V$ satisfying the equivariance condition:

$$\psi(gh) = \rho(h^{-1}) \cdot \psi(g) \tag{12.4}$$

for all $h \in H, g \in G$.

Note that, because G is a group we can multiply both on the left, and on the right. This allows us to define a representation of G in a way that does not spoil the equivariance condition:

Let us check that this is indeed a representation. If $\psi \in \mathrm{Ind}_H^G(\rho)$, and $g \in G$ we define $T(g) \cdot \psi$ as the new function whose values are given in terms of the old one by

$$(T(g) \cdot \psi)(g') := \psi(g^{-1} g') \tag{12.5}$$

eq:indrepi

The inverse is needed so that we actually have a representation:

$$\begin{aligned} (T(g_1)T(g_2) \cdot \psi)(g') &= (T(g_2) \cdot \psi)(g_1^{-1} g') \\ &= \psi(g_2^{-1} g_1^{-1} g') \\ &= (T(g_1 g_2) \cdot \psi)(g') \end{aligned} \tag{12.6}$$

However, the action (12.5) manifestly does not interfere with the equivariance property of ψ! In detail: if $h \in H$ then

$$(T(g)\cdot\psi)(g'h) = \psi(g^{-1}g'h) = \rho(h^{-1})\psi(g^{-1}g') = \rho(h^{-1})((T(g)\cdot\psi)(g')) \tag{12.7}$$ eq:explicit

So, $T(g)\cdot\psi$ is also equivariant.

Remarks: The associated vector bundle $(G\times_H V)\to G/H$ is an example of an *equivariant vector bundle*. In general if $E\to M$ is a vector bundle over a space M with a left G-action then E is an *equivariant vector bundle*, or *G-vector bundle* if there is an action of G on the total space of E, linear on the fibers, which covers the G action on M. Very often in physics (and mathematics) we are in a situation where there is a given G action on a space M and we must lift the action to make a vector bundle equivariant. This comes up, for example, in formulating Gauss laws.

12.1 Induced Representations of Finite Groups: Frobenius Reciprocity

The theory of induced representations is already interesting and nontrivial for G, H finite groups. In this case $G\to G/H$ is a finite cover (by H) of a discrete set of points. Nevertheless, the general geometrical ideas apply.

Let $\mathrm{Rep}(G)$ denote the category of finite-dimensional representations of G. Morphisms between $W_1, W_2 \in \mathrm{Rep}(G)$ are linear transformations commuting with G, i.e. G-intertwiners, and the vector space of all morphisms is denoted $\mathrm{Hom}_G(W_1, W_2)$. The induced representation construction defines a functor

$$\mathrm{Ind} : \mathrm{Rep}(H)\to\mathrm{Rep}(G). \tag{12.8}$$

(We denoted this by Ind_H^G before but H, G will be fixed in what follows so we simplify the notation.) On the other hand, there is an obvious functor going the other way, since any G-rep W is *a foriori* an H-rep, by restriction. Let us denote this "restriction functor"

$$R : \mathrm{Rep}(G)\to\mathrm{Rep}(H) \tag{12.9}$$

How are these two maps related? The answer is that they are "adjoints" of each other! This is the statement of Frobenius reciprocity:

$$\mathrm{Hom}_G(W, \mathrm{Ind}(V)) = \mathrm{Hom}_H(R(W), V) \tag{12.10}$$ eq:frobrecip

We can restate the result in another way which is illuminating because it helps to answer the question: How is $\mathrm{Ind}_H^G(V)$ decomposed in terms of irreducible representations of G? Let W_α denote the distinct irreps of G. Then Schur's lemma tells us that

$$\mathrm{Ind}_H^G(V) \cong \oplus_\alpha W_\alpha \otimes \mathrm{Hom}_G(W_\alpha, \mathrm{Ind}_H^G(V)) \tag{12.11}$$ eq:schurlemma

But now Frobenius reciprocity (12.10) allows us to rewrite this as

$$\mathrm{Ind}_H^G(V) \cong \oplus_\alpha W_\alpha \otimes \mathrm{Hom}_H(R(W_\alpha), V) \tag{12.12}$$ eq:frobrecipii

where the sum runs over the unitary irreps W_α of G, with multiplicity one.

The statement (12.12) can be a very useful simplification of (12.11) if H is "much smaller" than G. For example, G could be nonabelian, while H is abelian. But the representation theory for abelian groups is much easier! Similarly, G could be noncompact, while H is compact. etc.

Proof of Frobenius reciprocity:

In order to prove (12.12) we note that it is equivalent (see the exercise below) to the statement that the character of $\mathrm{Ind}_H^G(V)$ is given by

$$\chi(g) = \sum_{x \in G/H} \hat{\chi}(x^{-1} g x) \tag{12.13}$$ eq:charind

where x runs over a set of representatives and $\hat{\chi}$ is the character χ_V for H when the argument is in H and zero otherwise.

On the other hand, (12.13) can be understood in a very geometrical way. Think of the homogeneous vector bundle $G \times_H V$ as a collection of points $g_j H$, $j = 1, \ldots, n$ with a copy of V sitting over each point. Now, choose a representative $g_j \in G$ for each coset. Having chosen representatives g_j for the distinct cosets, we may write:

$$g \cdot g_j = g_{g \cdot j} h(g, j) \tag{12.14}$$ eq:permcos

where $j \mapsto g \cdot j$ is just a permutation of the integers $1, \ldots, n$, or more invariantly, a permutation of the points in G/H.

Now let us define a basis for the induced representation by introducing a basis v_a for the H-rep V and the equivariant functions determined by:

$$\psi_{i,a}(g_j) := v_a \delta_{i,j} \tag{12.15}$$

Geometrically, this is a section whose support is located at the point $g_i H$. The equivariant function is then given by

$$\psi_{i,a}(g_j h) := \rho(h^{-1}) v_a \delta_{i,j} \tag{12.16}$$

Now let us compute the action of $g \in G$ in this basis:

$$\begin{aligned} (g \cdot \psi_{i,a})(g_j) &= \psi_{i,a}(g^{-1} g_j) \\ &= \psi_{i,a}(g_{g^{-1} \cdot j} h(g^{-1}, j)) \\ &= \delta_{i, g^{-1} \cdot j} \rho(h(g^{-1}, j)^{-1}) \cdot v_a \end{aligned} \tag{12.17}$$

Fortunately, we are only interested in the trace of this G-action. The first key point is that *only the fixed points of the g-action on G/H contribute*. Note that the RHS above is supported at $j = g \cdot i$, but if we are taking the trace we must have $i = j$. But in

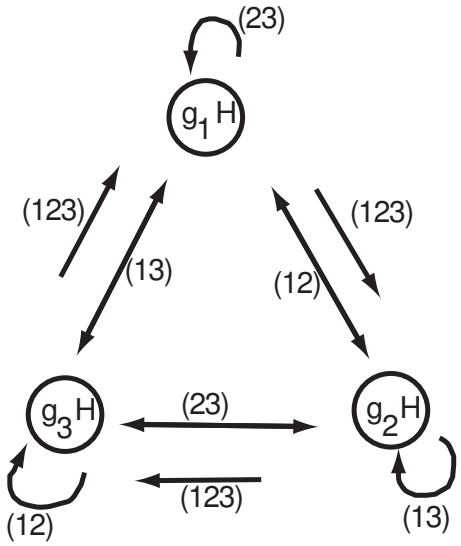


Figure 4: The left action of $G = S_3$ on G/H. In fact, this picture should be considered as a picture of a category, in this case, a groupoid. fig:indrepi

this case $gg_i = g_i h(g,i)$ and hence $g^{-1}g_i = g_i h(g,i)^{-1}$ so for fixed points we can simplify $h(g^{-1},i) = h(g,i)^{-1}$, and hence when we take the trace the contribution of a fixed point $gg_iH = g_iH$ is the trace in the H-rep of $h(g,i) = g_i^{-1}gg_i$, as was to be shown ♠

Remark: The Ind map does not extend to a ring homomorphism of representation rings.

Example A simple example from finite group theory nicely illustrates the general idea. Let $G = S_3$ be the permutation group. Let $H = \{1,(12)\} \cong \mathbb{Z}_2$ be a $\mathbb{Z}_2$ subgroup. G/H consists of 3 points. The left action of G on this space is illustrated in (12.5).

There are two irreducible representations of H, the trivial and the sign representation. These are both 1-dimensional. Call them $V(\epsilon)$, with $\epsilon = \pm$. Accordingly, we are looking at a line bundle over G/H and the vector space of sections of $G \times_H V(\epsilon)$ is 3-dimensional. A natural basis for the space of sections is given by the functions which are "δ-functions supported at each of the three points":

$$\begin{aligned} s^i(g_jH) &= \delta_{ij} \\ g_1H = (13)H &= \{(13),(123)\} \\ g_2H = (23)H &= \{(23),(132)\} \\ g_3H = (12)H &= \{1,(12)\} \end{aligned} \tag{12.18}$$ eq:basis

These sections correspond to equivariant functions on the total space. The space of all functions $F : G \to \mathbb{R}$ is a six-dimensional vector space. The equivariance condition:

$$\begin{aligned} F(12) &= \epsilon F(1) \\ F(123) &= \epsilon F(13) \\ F(132) &= \epsilon F(23) \end{aligned} \tag{12.19}$$ eq:equicond

cuts this six-dimensional space down to a three-dimensional space.

We can choose a basis of equivariant functions by *choosing* 3 representatives g_1, g_2, g_3 for the cosets in G/H and setting $F^i(g_j) = \delta_{ij}$. Using such a basis the representation of the group is easily expressed as a permutation representation.

In our example of $G = S_3$ it is prudent to choose $g_1 = (13), g_2 = (23), g_3 = (12)$ so that

$$\begin{aligned} (12)g_1 &= g_2(12) \\ (12)g_2 &= g_1(12) \\ (12)g_3 &= g_3(12) \\ (13)g_1 &= g_3(12) \\ (13)g_2 &= g_2(12) \\ (13)g_3 &= g_1(12) \\ (23)g_1 &= g_1(12) \\ (23)g_2 &= g_3(12) \\ (23)g_3 &= g_2(12) \end{aligned} \tag{12.20}$$ eq:multout

From this one easily gets the induced representation

$$\rho_{ind}(12) = \begin{pmatrix} 0 & \epsilon & 0 \\ \epsilon & 0 & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \tag{12.21}$$

$$\rho_{ind}(132) = \begin{pmatrix} 0 & 0 & \epsilon \\ \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \end{pmatrix} \tag{12.22}$$

and so forth.

Now let us look at Frobenius reciprocity. The irreducible representations of G are $W(\epsilon)$ defined by $(ij) \to \epsilon$, and W_2 defined by the symmetries of the equilateral triangle, embedded into $O(2)$:

$$\begin{aligned} \rho_2(12) &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ \rho_2(123) &= \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \end{aligned} \tag{12.23}$$ eq:rhotwo

As $H = \mathbb{Z}_2$ representations, we have $W(\epsilon) \cong V(\epsilon)$ and

$$W_2 \cong V(+1) \oplus V(-1) \tag{12.24}$$

Therefore

$$\begin{aligned} \mathrm{Hom}_H(W(\epsilon), V(\epsilon')) &= \delta_{\epsilon,\epsilon'}\mathbb{R} \\ \mathrm{Hom}_H(W_2, V(\epsilon)) &= \mathbb{R} \end{aligned} \tag{12.25}$$

By Frobenius reciprocity we have

$$\mathrm{Ind}_{S_2}^{S_3}(V(\epsilon)) = W(\epsilon) \oplus W_2 \tag{12.26}$$

eq:frobexple

Let us check this by computing characters from the geometric perspective.

The induced representation consists of functions on G/H valued in $V(\epsilon)$. The action of g acts as a permutation representation on the support of the functions.

Therefore, we can compute the character of g in the induced representation by looking at its action on fixed points:

$$\chi_{\rho_{ind}}(g) = \sum_{Fix(g)} \chi_{V(\epsilon)}(h(g,i)) \tag{12.27}$$

To find out how to decompose the representation $\mathrm{Ind}_{S_2}^{S_3}(V(\epsilon))$ in terms of $G = S_3$ irreps it suffices to compute the character for $g = (12)$ and $g = (123)$. Now, $g = (12)$ has exactly one fixed point, namely $g_3 H$ and $h(g, 3) = (12)$ for this element. Therefore,

$$\chi_{\rho_{ind}}(12) = \chi_V(12) = \epsilon \tag{12.28}$$

On the other hand, $g = (123)$ clearly has no fixed points, and therefore the character is zero. It follows immediately that we have the decomposition (12.26).

****

****

INCORPORATE:

Now let us turn to induced representations:

Let G be a group and H a subgroup. Suppose that $\rho : H \to \mathrm{End}(V)$ is a representation of the *subgroup* H. Then, as we have seen $\mathrm{Map}(G, V)$ is canonically a $G \times H$-space. To keep the notation under control we denote a general function in $\mathrm{Map}(G, V)$ by Ψ. Then the left-action of $G \times H$ defined by declaring that for $(g, h) \in G \times H$ and $\Psi \in \mathrm{Map}(G, V)$ the new function $\phi((g, h), \Psi) \in \mathrm{Map}(G, V)$ is the function $G \to V$ defined by:

$$\phi((g,h),\Psi)(g_0) := \rho(h) \cdot \Psi(g^{-1} g_0 h) \tag{12.29}$$

for all $g_0 \in G$. Now, we can consider the subspace of functions *fixed by the action of* $1 \times H$. That is, we consider the H-equivariant functions which satisfy

$$\boxed{\Psi(gh^{-1}) = \rho(h)\Psi(g)} \tag{12.30}$$

eq:EquivFun

for every $g \in G$ and $h \in H$. Put differently: There are two natural left-actions on $\mathrm{Map}(G,V)$ and we consider the subspace where they are equal. Note that the space of such functions is a linear subspace of $\mathrm{Map}(G,V)$. We will denote it by $\mathrm{Ind}_H^G(V)$. Moreover, it is still a representation of G since if Ψ is equivariant so is $(g,1)\cdot\Psi$.

The subspace $\mathrm{Ind}_H^G(V) \subset \mathrm{Map}(G,V)$ of H-equivariant functions, i.e. functions satisfying (12.30) is called the *induced representation of G, induced by the representation V of the subgroup H*. This is an important construction with a beautiful underlying geometrical interpretation. In physics it yields:

1. The irreducible unitary representations of space groups in condensed matter physics.

2. The irreducible unitary representations of the Poincaré group in QFT.

Example: Let us take $V = \mathbb{C}$ with the trivial representation of H, i.e. $\rho(h) = 1$. Then the induced representation is the vector space of functions on G which are invariant under right-multiplication by H. This is precisely the vector space of $\mathbb{C}$-valued functions on the homogeneous space G/H. For example, the invariant Wigner functions $D^j_{m_L,m_R}$ under right-action by the diagonal $U(1)$ subgroup of $SU(2)$ are $D^j_{m_L,0}(g)$. These descend to functions on $SU(2)/U(1) \cong S^2$ known (for j integral) as the *spherical harmonics*. The case of V a trivial representation generalizes in a beautiful way: When (ρ, V) is nontrivial the induced representation is interpreted not as a space of functions on G/H but rather as a vector space of sections of a homogeneous vector bundle over G/H determined by the data (ρ, V). See **** below.

Exercise
Prove (6.20).

Exercise

♣Physics 619, ch. 5, 2002 for more ♣

Let G be the symmetric group on $\{1,2,3\}$ and let $H = \{1,(12)\}$. Choose a representation of H with $V \cong \mathbb{C}$ and $\rho(\sigma) = +1$ or $\rho(\sigma) = -1$.

a.) Show that in either case, the induced representation $\mathrm{Ind}_H^G(V)$ is a three-dimensional vector space.

b.) Choose a basis for $\mathrm{Ind}_H^G(V)$ and compute the representation matrices of the elements of S_3 explicitly.

Exercise *Relation to the Peter-Weyl theorem*

Take H to be the trivial group and V the trivial representation and explain the relation of (12.12) to the Peter-Weyl theorem.

****

****

13. Representations Of $SU(2)$

♣TAKEN FROM GMP 2010 BUT ALL MATERIAL ASSUMING INDUCED REPRESENTATIONS, BUNDLE THEORY, AND BOREL-WEIL-BOTT IS SUPPRESSED ♣

Let us now consider the space of $\mathbb{C}^\infty$ sections $\Gamma(\mathcal{L}_k)$ as a representation of $SU(2)$. The right-action by $U(1)$ is given by

$$g \cdot e^{i\chi} := g e^{-i\chi\sigma^3} \tag{13.1}$$

where on the RHS we have ordinary matrix multiplication. According to our general principle, the sections of V_k are the same as equivariant functions $\psi : SU(2) \to \mathbb{C}$ such that

$$\psi(g e^{-i\chi\sigma^3}) = e^{-ik\chi}\psi(g) \tag{13.2}$$

eq:eqvfunct

In order to analyze (13.2) we must recall the Peter-Weyl theorem for $SU(2)$:

$$L^2(SU(2)) \cong \oplus_{j\in\frac{1}{2}\mathbb{Z}_+} \overline{\mathcal{D}^j} \otimes \mathcal{D}^j \tag{13.3}$$

as a representation of the $G_{\text{left}} \times G_{\text{right}}$. An orthonormal basis of functions is given by the matrix elements of $SU(2)$ in the unitary irreducible representations $\mathcal{D}^j$. Recall these are labeled by $j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots$ of dimension $n = 2j+1 = 1, 2, 3, 4, 5, \ldots$.

Using our parametrization of $SU(2)$ in terms of Euler angles from Chapter 5 we have:

$$\begin{aligned}
\langle j, m_L \mid T^{(j)}(g) \mid j, m_R\rangle &\equiv D^j_{m_L m_R}(\phi,\theta,\psi) \\
&= \langle j, m_L | T^{(j)}(e^{-i\frac{1}{2}\phi\sigma^3}) T^{(j)}(e^{-i\frac{1}{2}\theta\sigma^2}) T^{(j)}(e^{-i\frac{1}{2}\psi\sigma^3}) | j, m_R\rangle \\
&= e^{-im_L\phi} P^j_{m_L,m_R}(\cos\theta) e^{im_R\psi}
\end{aligned} \tag{13.4}$$

eq:sutmtrx

Here $P^j_{m_L,m_R}(x)$ is an associated Legendre polynomial and we have chosen the slightly unconventional representation of *antihermitian* generators:

$$J^k = -\frac{i}{2}\sigma^k \qquad\qquad [J^i, J^j] = \epsilon^{ijk} J^k \tag{13.5}$$

and diagonalized $J^3|j,m\rangle = im|j,m\rangle$.

Thus subspace of $L^2(SU(2))$ satisfying the equivariance condition (13.2) is that spanned by the D-functions with $k = -2m_R$.

Thus, an ON basis of functions satisfying the equivariance condition is $D^j_{m_L,-k/2}$, as j, m_L range over all possible values. Clearly, we must have $j \geq |k|/2$, and for fixed j the span of functions $m_L = -j, -j+1, \ldots, +j$ form a representation of $SU(2)$ and hence: From this description it is clear that the induced representation is infinite dimensional, and is given by a sum of representations of $SU(2)$,

$$\mathrm{Ind}_{U(1)}^{SU(2)}(V_k) \cong \oplus_{j' \geq \frac{1}{2}|k|} \overline{\mathcal{D}^{j'}} \tag{13.6}$$ eq:indrepsut

where the sum is over $j' = \frac{1}{2}|k|, \frac{1}{2}|k|+1, \frac{1}{2}|k|+2, \ldots$.

Remarks

• To explain the complex conjugate in (13.6) recall that under the left regular representation we have

$$\begin{aligned}(L(g_0)\cdot D^j_{m_L,m_R})(g) &:= D^j_{m_L,m_R})(g_0^{-1}g) \\ &= \langle j, m_L | T^{(j)}(g_0^{-1}) T^{(j)}(g) | j, m_R \rangle \\ &= \sum_{m'} \left(D^j_{m' m_L}(g_0) \right)^* D^j_{m' m_R}(g)\end{aligned} \tag{13.7}$$

• Let us check the Frobenius reciprocity statement:

$$\mathrm{Hom}_{SU(2)}(\mathcal{D}^{j'}, \mathrm{Ind}(V_k)) = \mathrm{Hom}_{U(1)}(R^*\mathcal{D}^{j'}, V_k) \tag{13.8}$$

The RHS is one-dimensional exactly for those values of j' given by $2j' - |k| \in 2\mathbb{Z}_+$.

• In the special case $k = 0$ we are talking about sections of the trivial line bundle. But these are just complex valued L^2 functions on S^2. Indeed, in this case $m_R = 0$ and the functions become independent of ψ and hence well-defined functions on S^2 rather than sections of nontrivial line bundles. In this case the functions are known as spherical harmonics and usually denoted:

$$D^\ell_{m,0} = Y_{\ell,m}(\theta, \phi) \tag{13.9}$$

• If we were studying the quantum mechanics of a charged particle of electric charge $+1$ moving on a sphere surrounding a magnetic monopole of magnetic charge k then the Hilbert space of the particle would be $\Gamma(\mathcal{L}_k) = \mathrm{Ind}_{U(1)}^{SU(2)}(V_k)$. Thus, we will sometimes refer to vectors in this representation space as "states" or "wavefunctions."

Now for $k \geq 0$ let

$$\mathcal{H}_k := \{\psi : \psi(\lambda u, \lambda v) = \lambda^k \psi(u,v)\} \subset \mathbb{C}[u,v] \tag{13.10}$$

be the vector space of homogeneous polynomials in two variables u, v of degree k.

Now, there is a natural basis of functions in $\mathcal{H}_k$, given by by the monomials $\psi_\ell(u,v) = u^\ell v^{k-\ell}$, $\ell = 0, 1, \ldots, \ell$. The matrix elements in this basis can be computed from

$$\begin{aligned}(g\cdot\psi_\ell)(u,v) &= \psi_\ell(g^{-1}\begin{pmatrix} u \\ v \end{pmatrix}) \\ &= \psi_\ell((\det g)^{-1}(g_{22}u - g_{12}v, -g_{21}u + g_{11}v)) \\ &= (\det g)^{-k}(g_{22}u - g_{12}v)^\ell(-g_{21}u + g_{11}v)^{k-\ell}\end{aligned} \tag{13.11}$$ eq:mtrxelem

By expanding out using the binomial theorem and collecting terms one obtains explicit matrix elements. We will do this momentarily.

What we have learned so far is that the holomorphically induced representation from B to $SL(2,\mathbb{C})$ for V_k is the $k+1$-dimensional representation which can canonically be viewed as the space of homogeneous polynomials of degree k. [*** Redundant??? ***]

Now, since $SU(2) \subset GL(2,\mathbb{C})$ we can restrict the functions in $\mathrm{HolInd}_B^{GL(2,\mathbb{C})}(V_k)$ to $SU(2)$. Since $SU(2) \cap B = T \cong U(1)$ they will be functions in the induced representation $\mathrm{Ind}_{U(1)}^{SU(2)}(V_k)$. The monomial basis functions become

$$\psi_\ell(g) = \alpha^\ell \beta^{k-\ell} \qquad \ell = 0, 1, \ldots, k \tag{13.12}$$

eq:monbasis

Now, let us try to identify this with one of the Wigner functions in $\mathrm{Ind}_{U(1)}^{SU(2)}(V_k)$.

In order to identify (13.12) with Wigner functions let us restrict the $GL(2,\mathbb{C})$ representation $\mathcal{H}_k$ for $k \geq 0$ to $SU(2)$. The representation remains irreducible. If we define j by $k = 2j$ then it is convenient to rename the basis slightly to

$$\tilde{f}_m(u,v) \equiv u^{j+m} v^{j-m} \tag{13.13}$$

eq:hombasp

for $m = -j, -j+1, -j+2, \cdots, j-1, j$. Note that m increases in steps of $+1$ and hence $j \pm m$ is always an integer even though j, m might be half-integer.

Now, for

$$g = \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix} \qquad |\alpha|^2 + |\beta|^2 = 1 \tag{13.14}$$

we continue the computation of the matrix elements (13.11) for this representation of $SU(2)$:

$$\begin{aligned} (g \cdot \tilde{f}_m)(u,v) &:= \tilde{f}_m(\bar{\alpha}u + \bar{\beta}v, -\beta u + \alpha v) \\ &= (\bar{\alpha}u + \bar{\beta}v)^{j+m}(-\beta u + \alpha v)^{j-m} \\ &:= \sum_{m'} \tilde{D}^j_{m'm}(g) \tilde{f}_{m'} \end{aligned} \tag{13.15}$$

eq:expbas

Note that for $g = \exp[-\frac{i}{2}\phi\sigma^3] = \exp[\phi J^3]$ (with our unconventional normalization of J^3) we have

$$g \cdot \tilde{f}_m = e^{im\phi} \tilde{f}_m \tag{13.16}$$

More generally, we can derive an explicit formulat for the matrix elements $\tilde{D}^j_{m'm}(g)$ as functions on $SU(2)$ by expanding out the two factors in (13.15) using the binomial theorem and collecting terms:

$$\tilde{D}^j_{m'm}(g) = \sum_{s+t=j+m'} \binom{j+m}{s} \binom{j-m}{t} \bar{\alpha}^s \alpha^{j-m-t} \bar{\beta}^{j+m-s} (-\beta)^t \tag{13.17}$$

eq:xplcttldee

Warning: This differs from standard Wigner functions by normalization factors. These are important when dealing with unitary representations. We return to unitarity below. For the moment we merely remark that $\tilde{D}^j_{m_L m}$ is proportional to $D^j_{m_L m}$.

Now, we found earlier that $\mathrm{Ind}_{U(1)}^{SU(2)}(V_k)$ is spanned by matrix elements with $m_R = -k/2$. In equation (13.17) this means we should put $m = -j$ which forces $s = 0$ and hence

$$\tilde{D}^j_{m,-j} = (-1)^{j+m}\binom{2j}{j+m}\alpha^{j-m}\beta^{j+m} \tag{13.18}$$

Up to a normalization constatn this is the explicit basis (13.12) we set out to identify.

Remarks

• If we consider a charged particle confined to a sphere and coupled to a Dirac monopole of charge k we will see that quantization of the system leads to a Hilbert space

$$\Gamma(\mathcal{L}_k) = \mathrm{Ind}_{U(1)}^{SU(2)}(V_k). \tag{13.19}$$

Since $SU(2)$ is a global symmetry of the quantum system the Hilbert space must be a representation of $SU(2)$. Above we have given the decomposition of this Hilbert space in terms of irreps.

• $SU(2)/U(1)$ is also a Kahler manifold and therefore we will be able to introduce the supersymmetric quantum mechanics with this target space with $N = 2$ supersymmetry. If the particle is coupled to a gauge field on a line bundle of first chern class k (i.e. if the particle is charged, with charge k) then the above representation has the physical interpretation as the space of BPS states.

****

****

13.0.1 Matrix Reps Of $\mathfrak{su}(2)$

As we have mentioned above, given a representation of a Lie group we can recover a representation of the corresponding Lie algebra. The representations corresponding to $\mathcal{H}^n$ with $n = 2j$ are the following:

Working to first order in ϵ we have:

$$T(e^{\epsilon X})\cdot f = T(1+\epsilon X)\cdot f = (T(1)+\epsilon T(X))\cdot f = f + \epsilon(T(X)\cdot f)$$

for any $X \in su(2)$. Taking $X = X_3 = \frac{i}{2}\sigma_3$ and evaluating the definition of the function on the LHS evaluated on (u, v) we get:

$$f(u - i\frac{1}{2}\epsilon u, v - i\frac{1}{2}\epsilon v) = f - \epsilon\frac{i}{2}\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)f$$

so we conclude:

$$T(X_3)\cdot f = -i\frac{1}{2}\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right)f$$

and, in particular,

$$iT(X^3)\cdot \tilde{f}_{j,m} = m\tilde{f}_{j,m}$$

Similarly we have

$$\begin{aligned} f + \epsilon T(X^1)\cdot f &= T(e^{\epsilon\frac{i}{2}\sigma^1})\cdot f \\ &= f - \epsilon\frac{i}{2}\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right)f \end{aligned} \tag{13.20}$$

since for $g = e^{\epsilon \frac{i}{2}\sigma^1}$ we have $\alpha = 1 + \mathcal{O}(\epsilon^2), \beta = -\frac{i}{2}\epsilon$. In just the same way we get:

$$\begin{aligned} f + \epsilon T(X^2) \cdot f &= T(e^{\epsilon \frac{i}{2}\sigma^2}) \cdot f \\ &= f - \epsilon \frac{1}{2}\left(v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}\right) f \end{aligned} \tag{13.21}$$

Let us define operators:

$$J^{\pm} \equiv iT(X^1) \pm T(X^2)$$

These act as

$$\begin{aligned} J^+ \cdot \tilde{f}_{j,m} &= u\frac{\partial}{\partial v}\tilde{f}_{j,m} \\ &= (j-m)\tilde{f}_{j,m+1} \\ J^- \cdot \tilde{f}_{j,m} &= v\frac{\partial}{\partial u}\tilde{f}_{j,m} \\ &= (j+m)\tilde{f}_{j,m-1} \end{aligned} \tag{13.22}$$

Or, as matrices:

$$J^+ = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 2 & 0 & \cdots \\ 0 & 0 & 0 & 3 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 2j \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{13.23}$$

eq:sutwomtrx

$$J^- = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ 2j & 0 & 0 & 0 & \cdots \\ 0 & 2j-1 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{13.24}$$

Exercise

Check that this is a representation of $su(2)$. It is different from what one finds in most textbooks on quantum mechanics because it is not a unitary matrix representation. See the next section.

****

****

13.1 Unitary structure

In Chapter 7 we observed that we can define an Hermitian metric on the holomorphic line bundle H^k such that the norm-squared of a section is

$$\| s \|^2 = \frac{1}{(1+|z|^2)^k}|s(z)|^2 \tag{13.25}$$

in the trivializations on $\mathcal{U}_\pm$. Therefore, we can define a unitary structure on the space of holomorphic sections by declaring:

$$\langle s, s\rangle := \int_{\mathbb{C}P^1} \| s \|^2 \, \mathrm{vol}\,(\mathbb{C}P^1) \tag{13.26}$$

where

$$\mathrm{vol}\,(\mathbb{C}P^1) = \frac{i}{2\pi}\frac{dzd\bar{z}}{(1+|z|^2)^2} \tag{13.27}$$

is the unit volume form. We claim that the action of $SU(2)$ on $H^0(\mathbb{C}P^1; H^k)$ we described above is unitary with respect to this metric.

To see this let us first note that we can identify the space $\mathcal{H}_k$ of homogeneous polynomials in two variables of degree k with the space $\mathcal{P}_k$ of polynomials in a single variable of degree $\leq k$ as follows:

If $\psi(u,v) \in \mathcal{H}_k$ then we map it to

$$s(z) := u^{-k}\psi(u,v) \tag{13.28}$$

Note that $s(z)$ will be a polynomial in the variable $z = v/u$ of degree $\leq k$.

If we require that the map $\mathcal{H}_k \to \mathcal{P}_k$ is an intertwiner of $SU(2)$ representations then we arrive at the formula for the $SU(2)$ action on $\mathcal{P}_k$: It takes a polynomial $s(z)$ to

$$(g\cdot s)(z) := (\bar{\beta}z + \bar{\alpha})^k s\left(\frac{\alpha z - \beta}{\bar{\beta}z + \bar{\alpha}}\right) \tag{13.29}$$

Now, the map

$$z \to z' = \frac{\alpha z - \beta}{\bar{\beta}z + \bar{\alpha}} \tag{13.30}$$

preserves the volume form $\mathrm{vol}\,(\mathbb{C}P^1)$, and a small computation using the identity

$$\frac{1}{(1+|z'|^2)} = \frac{|\bar{\beta}z + \bar{\alpha}|^2}{(1+|z|^2)} \tag{13.31}$$

shows that

$$\langle g\cdot s, g\cdot s\rangle = \langle s, s\rangle \tag{13.32}$$

In this way $H^0(\mathbb{C}P^1; H^k)$ becomes a Hilbert space on which $SU(2)$ acts unitarily.

The ON basis corresponding to $\tilde{f}_m$ is

$$\psi_{j,m} = \sqrt{\frac{(2j+1)!}{(j+m)!(j-m)!}}z^{j-m} \tag{13.33}$$

and can be taken to be $|j,m\rangle$ in our definition above of the Wigner functions $D^j_{m_L,m_R}$.

Exercise
Find the proportionality factor between $\tilde{D}^j_{m_L,m_R}$ and $D^j_{m_L,m_R}$.

****
****
UNITARIZE REP OF J^+ AND J^-
****
****

13.2 Lie algebra representations and differential operators

TO BE WRITTEN: Explain that

1. Generators are represented by differential operators.

2. The usual differential equations and recursion relations on special functions follow from this point of view.

13.3 Coherent state formalism

The above discussion of representations of $SU(2)$ is closely related to the subject of *coherent states*.

One reference for this material is

A. Perelomov, *Generalized Coherent States and Their Applications*, Springer Verlag 1986

Another reference is:

K. Fuji, Introduction to coherent states and quantum information theory, arXiv:quant-ph/0112090

One way to motivate this is to derive the representation of the Lie algebra in terms of differential operators.

If

$$J^+ = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{13.34}$$

then actiong on $\mathcal{P}_k$ we represent

$$J^+ = -\frac{\partial}{\partial z} \tag{13.35}$$

It is then natural to define coherent states:

$$|\zeta\rangle := \frac{1}{(1+|\zeta|^2)^j} \exp(\zeta J_+)|j,-j\rangle \tag{13.36}$$

One can show that

1.

$$|\zeta\rangle = \sum_m u_m(\zeta)|j,m\rangle \tag{13.37}$$

with

$$u_m(\zeta) = \sqrt{\frac{(2j)!}{(j+m)!(j-m)!}} \frac{\zeta^{j+m}}{(1+|\zeta|^2)^j} \tag{13.38}$$

******* IS SQRT CORRECT? ********

2. The inner product of two such polynomials is

$$\langle \xi | \eta \rangle = \frac{(1+\bar{\xi}\eta)^{2j}}{((1+|\xi|^2)^j (1+|\eta|^2)^j} \tag{13.39}$$

In particular $|\zeta\rangle$ is of unit norm.

3. If we regard polynomials in z of degree $\leq 2j$ as an irrep of $SU(2)$ as above then we may identify the state $|\zeta\rangle$ with the polynomial: ???

$$f(z) = \frac{(1+\zeta z)^{2j}}{(1+|\zeta|^2)^j} \tag{13.40}$$

Examples:

1. The wavefunction of the oxygen molecule!

2. Wavefunctions of electrons in the presence of a magnetic monopole will take values in this bundle.

3. Nice application of this coherent state formalism to the QHE in paper by Haldane.

Remarks:

1. The coherent states are good for taking the semiclassical limit of large j. Then the support of $|\zeta\rangle$ is sharply peaked at a point on $\mathbb{C}P^1$. **** SHOW *****

2.

3. Relate infinite dimensional induced rep to the Schwinger representation in terms of oscillators. *** What is this??? ***

14. Orbits of the Lorentz group and relativistic wave equations

14.1 Orbits, Representations, and Differential Equations

♣OVERLAP HERE WITH PETER-WEYL THEOREM IN CHAPTER 4 ON REPRESENTATIONS. MOVE THIS? ♣

Now let us turn to how these orbits are related to some important differential equations in field theory.

As we have seen, since $O(1,1)$ acts on $\mathbb{M}^{1,1}$, it follows that it acts on the space of fields $\mathrm{Map}(\mathbb{M}^{1,1}, \kappa)$, where $\kappa = \mathbb{R}$ or $\mathbb{C}$ for a real or complex-valued scalar field. For a *scalar* field recall the action of $A \in O(1,1)$ on the space of solutions is

$$(A \cdot \Psi)(x) := \tilde{\Psi}(x) := \Psi(A^{-1}x) \tag{14.1}$$

Note the A^{-1} in the argument of Ψ in the second equality. This is necessary to get a left-action of the group on the space of fields. If we use Ax then we get a right-action.

Now, quite generally, if V is a representation space for G, and $\mathcal{O} \in \mathrm{End}(V)$ is an invariant linear operator, i.e. an operator which commutes with the action of G,

$$\rho(g)\mathcal{O} = \mathcal{O}\rho(g) \tag{14.2}$$

♣For the case of *complex* scalar fields when A is in a component of $O(1,1)$ that reverses the orientation of time we have the choice of whether to include complex conjugation in the action. This would be adding a "charge conjugation" action. ♣

then any eigenspace, say $\{v \in V | \mathcal{O}v = \lambda v\}$ will be a sub-representation of G.

Consider the operator

$$\partial^\mu \partial_\mu = -\partial_0^2 + \partial_1^2 \tag{14.3}$$

acting on the space of scalar fields. This is an example of an invariant operator, as one confirms with a simple computation. It can be made manifest by writing

$$\partial^\mu \partial_\mu = -4\partial_+ \partial_- \tag{14.4}$$

where

$$\partial_\pm := \frac{\partial}{\partial x^\pm} = \frac{1}{2}(\partial_0 \mp \partial_1) \tag{14.5}$$

The *Klein-Gordon equation* for a complex or real scalar field $\Psi(x^0, x^1)$ is

$$\left(\partial^\mu \partial_\mu + m^2\right) \Psi = 0 \tag{14.6}$$

The space of fields satisfying the KG equation is a representation space of $O(1,1)$, by our general remark above.

Now we relate the orbits of $O(1,1)$ to the representations furnished by solutions of the KG equation:

For field configurations which are Fourier transformable we can write

$$\Psi(x^0, x^1) = \int dk_0 dk_1 \hat{\psi}(k) e^{\mathrm{i}k_0 x^0 + \mathrm{i}k_1 x^1} \tag{14.7}$$

If the field Ψ is on-shell then $\hat{\psi}(k)$ must have support on set

$$\{k : (k_0)^2 - (k_1)^2 = m^2\} \tag{14.8}$$

eq:MassShell

that is:

$$(k^2 + m^2)\hat{\psi}(k) = 0 \tag{14.9}$$

Thus, the support in Fourier space is an orbit. In physics the orbit with $k_0 > 0$ is called the *mass-shell*.

For example, suppose that $m^2 > 0$. Then on the orbit with $k^0 > 0$ we have $k = m(\cosh\theta, \sinh\theta)$ and we can write the general complex-valued solution as

$$\hat{\psi}(k) = \delta(k^2 + m^2)\bar{a}, \tag{14.10}$$

where the amplitude $\bar{a}$ should be regarded as a complex-valued function on the orbit. Then a complex-valued solution on the KG equation is generated by such a function on the orbit by:

$$\begin{aligned} \Psi(x) &= \int d^2k \delta(k^2 + m^2)\bar{a} e^{\mathrm{i}k\cdot x} \\ &= \int_{\mathbb{R}} \frac{dk_1}{2\sqrt{k_1^2 + m^2}} \bar{a}(k_1) e^{\mathrm{i}k\cdot x} \\ &= \frac{1}{2}\int_{-\infty}^{\infty} d\theta \bar{a}(\theta) e^{\mathrm{i}m(x^0 \cosh\theta + x^1 \sinh\theta)} \end{aligned} \tag{14.11}$$

eq:MomentumMod

where we chose a convention designation of the argument of the function $\bar{a}$ on the orbit.

Thus, we can identify a space of solutions to the KG equation with a space of functions on an orbit.

Now recall that the space of functions on an orbit is a special case of an induced representation. As we will soon see, induced representations are the right concept for generalization to other representations of the Lorentz group.

Remark 1: *The components of the Lorentz group:* If we use a particular orbit to generate representations we only represent the subgroup of the Lorentz group which preserves that orbit. Thus, for the orbit with $m > 0$ and $k_0 > 0$ we can include the parity-reversing component, but not the time-reversing component. Of course, there is a similar representation of $SO_0(1,1) \amalg P \cdot SO_0(1,1)$ given by the hyperbola with $k_0 < 0$. If we wish to represent the full $O(1,1)$ then we must include both orbits. Note that, if we wish to use *real* solutions of the KG equation we must include both hyperbolae and there is a reality condition relating the Fourier modes: $(\bar{a}(k))^* = a(-k)$

Remark 2: The representation can be made into a unitary representation. See below.

Exercise

Show that an invariant linear operator $\mathcal{O} \in \mathrm{End}(V)$ on a representation space V of G is a fixed point of the $G \times G$ action on $\mathrm{End}(V)$.

♣This, and the concept of an invariant operator really goes above in the general section on functions on groups. ♣

14.2 The massless case in $1+1$ dimensions

Having phrased things this way it is natural to ask what relativistic wave equations correspond to amplitudes supported on the light-like orbits.

In the massless case $m^2 = 0$ the general solution of the KG equation is easily written as

$$\Psi = \psi_L(x^+) + \psi_R(x^-) \tag{14.12}$$

Solutions of the form $f_R = 0$ are called *left-moving waves* because, as time evolves forward, the profile of the function f_L moves to the left on the x^1-axis. Similarly, solutions of the form $f_L = 0$ are called *right-moving waves*.

Remark: In the massless case there is a "more primitive" relativistic wave equation which is first order, and whose solutions are always solutions of the massless KG equation. Namely, we can consider the separate equations

$$\partial_+ \Psi = 0 \tag{14.13}$$

eq:anti-self-d

$$\partial_- \Psi = 0 \tag{14.14}$$

eq:self-dual

Note that these equations are themselves Lorentz invariant, even though the operators $\partial_\pm$ are not invariant differential operators. Solutions to (14.13) are right-moving scalar

fields and solutions to (14.14) are left-moving scalar fields. Such on-shell scalar fields are also known as *chiral scalar fields*. Equations (14.13) and (14.14) are notable in part because they play an important role in string theory and conformal field theory. It is also interesting to note that it is quite subtle to write an action principle that leads to such equations of motion. They are also called the anti-self-dual and self-dual equations of motion, respectively because, if we choose the orientation $dx \wedge dt$ then the Hodge star operation is $*dt = dx$ and $*dx = dt$ and hence $*(dx^{\pm}) = \pm dx^{\pm}$. Therefore if $\partial_{\pm}\Psi = 0$ its "fieldstrength" $F = d\Psi$ satisfies $*F = \mp F$.

The action of $P, T \in O(1,1)$ on a real scalar field is given by:

$$\begin{aligned}(P \cdot \Psi)(x,t) &:= \Psi(-x,t) \\ (T \cdot \Psi)(x,t) &:= \Psi^*(x,-t)\end{aligned} \tag{14.15}$$

The KG equation is separately P and T invariant, but the (anti)-self-dual equations are not. Nevertheless, the latter equations are PT invariant. This is a special case of the famous CPT theorem:

Roughly speaking, mathematical consistency implies that if a physical theory is invariant under group transformations in the neighborhood of the identity then it must be invariant under the transformations in the full connected component of the identity. But this does not mean the theory is invariant under disconnected components such as the components containing P and T. As a matter of fact, Nature chooses exactly that option in the standard model of the electroweak and strong interactions. However, if we also assume that there is a certain relation to "analytically continued" equations in Euclidean signature then, since PT is in the connected component of the identity of $O(2)$, such theories must in fact be PT invariant.

♣There is much more to discuss about the nature of the analytic continuation implied here. In what sense is $SO(2)$ and "analytic continuation" of $SO(1,1)$? It does not do simply to make the boost pure imaginary unless $x^{\pm} \to z, \bar{z}$. ♣

14.3 The case of d dimensions, $d > 2$

Now consider Minkowski space $\mathbb{M}^{1,d-1}$ with $d > 2$. The nature of the orbits is slightly different.

1. For $\lambda^2 > 0$ we can define

$$\mathcal{O}^+(\lambda) = \{x | (x^0)^2 - (\vec{x})^2 = \lambda^2 \qquad \& \qquad \text{sign}(x^0) = \text{sign}(\lambda)\} \tag{14.16}$$

   By the stabilizer-orbit theorem we can identify this with

$$SO_0(1, d-1)/SO(d-1) \tag{14.17}$$

   by considering the isotropy group at $(x^0 = \lambda, \vec{x} = 0)$. See Figure 5(a).

2. For $\mu^2 > 0$ we can define

$$\mathcal{O}^-(\lambda^2) = \{x | (x^0)^2 - (\vec{x})^2 = -\mu^2\} \tag{14.18}$$

   By the stabilizer-orbit theorem we can identify this with

$$SO_0(1, d-1)/SO_0(1, d-2) \tag{14.19}$$

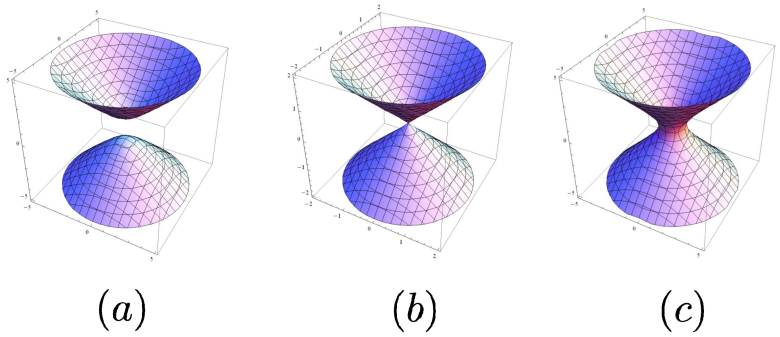


Figure 5: Illustrating orbits of the connected component of the identity in $O(1,3)$. In (a) the top and bottom hyperboloids are separate orbits, and if we include time-reversing transformations the orbits are unions of the two hyperboloids. In (b) there are three orbits shown with $x^0 > 0$ $x^0 < 0$ (the future and past, or forward and backward light cones), and the orbit consisting of the single point. In (c), once x^2 has been specified, there is just one orbit, for $d > 2$.

fig:LorentzOrb

by considering the isotropy group at $x = (x^0 = 0, x^1 = 0, \ldots, x^{d-2} = 0, x^{d-1} = \mu)$. The sign of μ does not distinguish different orbits for $d > 2$ because the sphere S^{d-2} is connected. See Figure 5(c).

3.

$$\mathcal{O}^{\pm} = \{x | x^2 = 0 \qquad \& \qquad \text{sign}(x^0) = \pm 1\} \tag{14.20}$$

Vectors in this orbit are of the form $(x^0, |x^0|\hat{n})$ where $\hat{n} \in S^{d-2} \subset \mathbb{R}^{d-1}$ and the sign of x^0 is invariant under the action of the identity component of $O(1,3)$. (Show this!). Note that, for $d = 2$ the sphere S^0 has two disconnected components, leading to left- and right-movers. But for $d > 2$ there is only one component. We can think of $\hat{n} \in S^{d-2}$ as parametrizing the directions of light-rays. That is, the point where the light ray hits the celestial sphere. In one spatial dimension, a light ray either moves left or right, and this is a Lorentz-invariant concept. In $d-1 > 1$ spatial dimensions, we can rotate any direction of light ray into any other. See Figure 5(b). One can show that these orbits too are homogeneous spaces: [6]

$$\mathcal{O}^{\pm} \cong SO_0(1, d-1)/\mathcal{I} \tag{14.21}$$

[6]The isotropy group of a light ray is $\mathcal{I} \cong ISO(d-2)$, where $ISO(d-2)$ is the Euclidean group on $\mathbb{R}^{d-2}$. The easiest way to show this is to use the Lie algebra of $so(1, d-1)$ and work with light-cone coordinates. Choosing a direction of the light ray along the x^{d-1} axis and introducing light-cone coordinates $x^{\pm} := x^0 \pm x^{d-1}$, and transverse coordinates x^i, $i = 1, \ldots, d-2$ if the lightray satisfies $x^- = 0$ then we have unbroken generators M^{+i} and M^{ij}.

4. The final orbit is of course $\{x = 0\}$.

As in $1+1$ dimensions we can identify a representation space of the Lorentz group associated with the space of solutions to the KG equation with functions on various orbits. The formulae are essentially the same. For example for the orbit $\mathcal{O}^+(m)$ a function $\bar{a} : \mathcal{O}^+(m) \to \mathbb{C}$ determines a solution:

$$\begin{aligned}\Psi(x) &= \int d^d k \delta(k^2 + m^2) \bar{a} e^{\mathrm{i} k\cdot x} \\ &= \int_{\mathbb{R}^{d-1}} \frac{d^{d-1}\vec{k}}{2\sqrt{\vec{k}^2 + m^2}} \bar{a}(\vec{k}) e^{\mathrm{i} k\cdot x}\end{aligned} \tag{14.22}$$

eq:MomentumMod

However, for scalar fields, there is no analog of the left- and right-chiral boson. There *are* analogs involving interesting first order equations such as the Dirac equation and the (anti-) self-dual equations for fields with spin.

Quite generally, we can define an inner product on the space of complex-valued solutions of the KG equation such that the action of the Lorentz group is unitary. Observe that, given any two complex-valued solutions Ψ_1, Ψ_2 the current

$$j_\mu := -\mathrm{i}\left(\Psi_1^* \partial_\mu \Psi_2 - (\partial_\mu \Psi_1)^* \Psi_2\right) \tag{14.23}$$

Note that, if Ψ_1 and Ψ_2 both satisfy the KG equation then

$$\partial^\mu j_\mu = 0 \tag{14.24}$$

is conserved. Therefore, if we choose a spatial slice with normal vector n^μ and induced volume form vol the inner product

$$(\Psi_1, \Psi_2) := \int_\Sigma n^\mu j_\mu \mathrm{vol} \tag{14.25}$$

is independent of the choice. So, fixing a Lorentz frame and taking Σ to be the slace at a fixed time we have

$$(\Psi_1, \Psi_2) := -\mathrm{i} \int_{\mathbb{R}^{d-1}} \left(\Psi_1^* \partial_0 \Psi_2 - (\partial_0 \Psi_1)^* \Psi_2\right) d^{d-1}\vec{x} \tag{14.26}$$

This is clearly not positive definite on the space of all solutions but does become positive definite when restricted to the space of complex solutions associated with a single orbit. Indeed, substituting the expansion in momentum space (14.22) we get

$$(\Psi_1, \Psi_2) = \int_{\mathbb{R}^{d-1}} \frac{d^{d-1}\vec{k}}{2\sqrt{\vec{k}^2 + m^2}} (\bar{a}_1(\vec{k}))^* (\bar{a}_2(\vec{k})) \tag{14.27}$$

Having phrased things this way, it is clear that there is an interesting generalization: We can choose other representations of $H = SO(d-1)$ and consider the induced representations of the Lorentz group. This indeed leads to the unitary representations, corresponding to particles with nontrivial spin.

♣Incorporate some of the following remarks: ♣

15. Characters and the Decomposition of a Representation to its Irreps

Given a group G there are some fundamental problems one wants to solve:

- Construct the irreducible representations.
- Given a representation, how can you tell if it is reducible?
- Given a reducible representation, what is its reduction to irreducible representations?

I am not aware of any systematic procedure for deriving the list of irreps of any group. However, if - somehow - the list is known then the answers to 2 and 3 are readily provided through the use of characters.

15.1 Some basic definitions

Definition 1 Let (T, V) be a representation of G. The functions on G defined by

$$\chi_{(T,V)}(g) \equiv \mathrm{Tr}_V T(g) \equiv \sum_i T_{ii}(g) \tag{15.1}$$

are called the *characters* of the representation T.

Note: $\chi_{(T,V)}$ is independent of basis. Moreover, if $(T, V) \cong (T', V')$ then

$$\chi_{(T,V)} = \chi_{(T',V')} \tag{15.2}$$

Exercise

Show that for a unitary representation

$$\chi(g^{-1}) = \chi(g)^* \tag{15.3}$$

A second motivation for looking at characters is the following: matrix elements are complicated. Sometimes one can extract all the information one is seeking by considering this simpler and more basic set of functions on G associated to a representation.

Recall that for any element of $g \in G$ the set

$$C(g) = \{h \in G : \exists k, g = khk^{-1}\} \tag{15.4}$$

is called the *conjugacy class* of g.

Definition A function on G that only depends on the conjugacy class is called a *class function*. In particular, characters are class functions. The space of class functions is a subspace of R_G.

Theorem. The χ_μ form a basis for the vector space of class functions in R_G.

Proof: This is a corollary of the previous section: Any function can be expanded in

$$F(g) = \sum \hat{F}^\mu_{ij} T^\mu_{ij}(g) \tag{15.5}$$

Therefore, it suffices to compute the average of $T^\mu_{ij}(g)$ over a conjugacy class:

$$(T^\mu_{ij})^{average}(g) \equiv \frac{1}{|G|} \sum_{h \in G} T^{(\mu)}_{ij}(hgh^{-1}) \tag{15.6}$$

Expanding this out and doing the sum on h using the ON relations shows that

$$(T^\mu_{ij})^{average}(g) = n_\mu \chi_\mu(g) \tag{15.7}$$

On the other hand, if F is a class function then

$$F^{\text{average}} = F \tag{15.8}$$

so

$$F(g) = \sum_{ij} \hat{F}^\mu_{ij} n_\mu \chi_\mu(g) \tag{15.9}$$

So that χ_μ form a basis of class functions. ♠

Now, there is another obvious basis of the space of class functions: Denote the distinct conjugacy classes by $C_i, i = 1, \ldots r$. For each C_i can define a class function δ_{C_i} in R_G to be the characteristic function of C_i. That is it is 1 on the class C_i and zero on all other group elements. Comparing these two ways of finding the dimension we get:

Theorem The number of conjugacy classes of G is the same as the number of irreducible representations of G.

Since the number of representations and conjugacy classes are the same we can define a *character table*:

$$\begin{array}{cccccc} & m_1 C_1 & m_2 C_2 & \cdots & \cdots & m_r C_r \\ \chi_1 & \chi_1(C_1) & \cdots & \cdots & \cdots & \chi_1(C_r) \\ \chi_2 & \chi_2(C_1) & \cdots & \cdots & \cdots & \chi_2(C_r) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \chi_r & \chi_r(C_1) & \cdots & \cdots & \cdots & \chi_r(C_r) \end{array} \tag{15.10}$$

Here m_i denotes the order of C_i.

As we will see, character tables are a very useful summary of important information about the representation theory of a group.

Exercise

In a previous exercise we computed the average number of fixed points of a finite group G acting on a finite set X.

Rederive the same result by interpreting the number of elements of X fixed by g as a character $\chi(g)$ of the representation given by the functions on X.

15.2 Orthogonality relations on the characters

Let χ_μ be the character of $T^{(\mu)}$. The orthogonality relations for matrix elements implies:

$$\frac{1}{|G|}\sum_{g\in G}\chi_\nu(g^{-1})\chi_\mu(g) \equiv (\chi_\nu,\chi_\mu) = \delta_{\mu\nu} \tag{15.11}$$

To prove this, simply put $i=\ell$ and $s=m$ in (9.9) and sum on i and s.

Now, we argued before - abstractly - that *any* finite dimensional representation (T,V) of G is completely reducible:

$$T \cong a_1 T^{(1)} \oplus a_2 T^{(2)} \oplus \cdots \oplus a_r T^{(r)} = \oplus_\mu a_\mu T^{(\mu)} \tag{15.12}$$ eq:reducible

The a_i are nonnegative integers, they measure the number of times a rep appears and are referred to as the "multiplicities."

From the orthogonality relation we get:

$$a_\mu = (\chi_{(T,V)},\chi_\mu) \tag{15.13}$$ eq:multipl

Thus we arrive at a key fact:

Theorem

Two representations are equivalent iff they have the same character.

If you have two reps and you want to see whether they are inequivalent then an efficient test is to compute the characters.

The orthogonality relations on characters can be stated more beautifully if we use the fact that characters are class functions:

Theorem Orthogonality Relations for characters Denote the distinct conjugacy classes $C_i \in \mathcal{C}$, and the (distinct) characters of the distinct irreducible unitary representations χ_μ. Then

$$\frac{1}{|G|}\sum_{C_i\in\mathcal{C}} m_i \chi_\mu(C_i)\chi_\nu(C_i)^* = \delta_{\mu\nu} \tag{15.14}$$ eq:orthogrel

where $m_i = |C_i|$ is the order of the conjugacy class C_i. We also have

$$\sum_\mu \chi_\mu(C_i)^*\chi_\mu(C_j) = \frac{|G|}{m_i}\delta_{ij} \tag{15.15}$$ eq:orthogreli

Proof: Let us first prove (15.14). From unitarity we have $\chi_\mu(g^{-1}) = (\chi_\mu(g))^*$. Since χ_μ are class functions the sum over the elements in the group can be written as a sum over the conjugacy classes $C_i \in \mathcal{C}$.

The equation (15.14) can be interpreted as the statement that the $r \times r$ matrix

$$S_{\mu i} := \sqrt{\frac{m_i}{|G|}}\chi_\mu(C_i) \qquad \mu = 1, \ldots, r \quad i = 1, \ldots, r \tag{15.16}$$

satisfies

$$\sum_{i=1}^{r} S_{\mu i} S^*_{\nu i} = \delta_{\mu\nu} \tag{15.17}$$

Therefore, $S_{\mu i}$ is a unitary matrix. The left-inverse is the same as the right-inverse, and hence we obtain (15.15). ♠

Exercise

Show that right-multiplication of the character table by a diagonal matrix produces a unitary matrix.

Exercise

Using the orthogonality relations on matrix elements, derive the more general relation on characters:

$$\frac{1}{|G|}\sum_{g\in G} \chi_\mu(g)\chi_\nu(g^{-1}h) = \frac{\delta_{\mu\nu}}{n_\mu}\chi^{(\nu)}(h) \tag{15.18}$$

eq:convolchar

We will interpret this more conceptually later.

Exercise

A more direct proof of (15.15) goes as follows. Consider the operator $L(g_1) \otimes R(g_2)$ acting on the regular representation. We will compute

$$\mathrm{Tr}_{R_G}[L(g_1) \otimes R(g_2)] \tag{15.19}$$

in two bases.

First consider the basis ϕ^μ_{ij} of matrix elements. We have:

$$L(g_1) \otimes R(g_2) \cdot \phi^\mu_{ij} = \sum_{i',j'} T^\mu_{j'j}(g_2) T^\mu_{i'i}(g_1^{-1}) \phi^\mu_{i'j'} \tag{15.20}$$

so the trace in this basis is just:

$$\mathrm{Tr}_{R_G}[L(g_1)\otimes R(g_2)] = \sum_\mu \chi_\mu(g_1)^*\chi_\mu(g_2) \tag{15.21}$$

On the other hand, we can use the delta-function basis: δ_g. Note that

$$L(g_1)\otimes R(g_2)\cdot\delta_g = \delta_{g_1 g g_2^{-1}} \tag{15.22}$$

So in the delta function basis we get a contribution of $+1$ to the trace iff $g = g_1 g g_2^{-1}$, that is iff $g_2 = g^{-1}g_1 g$, that is iff g_1 and g_2 are conjugate, otherwise we get zero.

Now show that

$$\#\{g : g_2 = g^{-1}g_1 g\} = |C(g_1)| = m_1 \tag{15.23}$$

eq:weight

by considering $Z(g_1)$, the centralizer subgroup of g_1 in G and identifying the cosets $G/Z(g_1)$ with the elements of the conjugacy class $C(g_1)$.

So we get the second orthogonality relation. ♠

---

16. Some examples of representations of discrete groups and their character tables

16.1 S_3

The group S_3 has three obvious irreducible representations:

$\mathbf{1}_+$. $1\to 1$ $(ij)\to 1$ $(123)\to 1$, etc.

$\mathbf{1}_-$. $1\to 1$ $(ij)\to -1$ $(123)\to +1$, etc. This is the sign representation, or the ϵ homomorphism to $\mathbb{Z}_2$ mentioned in lecture 1.

2. Symmetries of the triangle.

$$\begin{aligned} 1 &\to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ (23) &\to \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ (123) &\to \begin{pmatrix} \cos(2\pi/3) & \sin(2\pi/3) \\ -\sin(2\pi/3) & \cos(2\pi/3) \end{pmatrix} = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \end{aligned} \tag{16.1}$$

These reps are irreducible. Note that by applying group elements in the third representation to any vector we obtain a set of vectors which spans $\mathbb{R}^2$.

Indeed we can check:

$$6 = 1^2 + 1^2 + 2^2 \tag{16.2}$$

to conclude that there are no other irreps.

It is now straightforward to write out the character table for S_3.

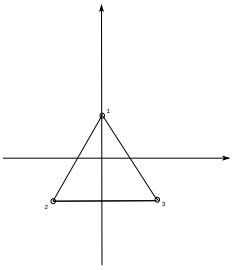


Figure 6: The symmetries of the equilateral triangle give the 2-dimensional irrep of S_3. fig:triangle

$$\begin{array}{rccc} & [1] & 3[(12)] & 2[(123)] \\ \chi_{\mathbf{1}_+}: & 1 & 1 & 1 \\ \chi_{\mathbf{1}_-}: & 1 & -1 & 1 \\ \chi_2: & 2 & 0 & -1 \end{array} \tag{16.3}$$

Here is an example of the use of characters to decompose a representation:

Now let us see how we can use the orthogonality relations on characters to find the decomposition of a reducible representation.

Example 1 Consider the 3×3 rep generated by the $\natural$ action of the permutation group S_3 on $\mathbb{R}^3$. We'll compute the characters by choosing one representative from each conjugacy class:

$$1 \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (12) \to \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (132) \to \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{16.4}$$

From these representatives the character of $V = \mathbb{R}^3$ is easily calculated:

$$\chi_V(1) = 3 \qquad \chi_V([(12)]) = 1 \qquad \chi_V([(132)] = 0 \tag{16.5}$$

Using the orthogonality relations we compute

$$a_{\mathbf{1}_+} = (\chi_{\mathbf{1}_+}, \chi) = \frac{1}{6}3 + \frac{3}{6}1 + \frac{2}{6}0 = 1 \tag{16.6}$$

$$a_{\mathbf{1}_-} = (\chi_{\mathbf{1}_-}, \chi) = \frac{1}{6}3 + \frac{3}{6}(-1)\cdot 1 + \frac{2}{6}0 = 0 \tag{16.7}$$

$$a_{\mathbf{2}} = (\chi_{\mathbf{2}}, \chi) = \frac{1}{6}3\cdot 2 + \frac{3}{6}0\cdot 1 + \frac{2}{6}(-1)\cdot 0 = 1 \tag{16.8}$$

Therefore:

$$\chi_V = \chi_{\mathbf{1}_+} + \chi_{\mathbf{2}} \tag{16.9}$$

showing the decomposition of $\mathbb{R}^3$ into irreps.

Example 2 Consider S_3 acting by permuting the various factors in the tensor space $V \otimes V \otimes V$ for any vector space V. Now, if $\dim V = d$ then we have

$$\begin{aligned} \chi([1]) &= d^3 \\ \chi([(ab)]) &= d^2 \\ \chi[(abc)] &= d \end{aligned} \tag{16.10}$$

eq:vcubed

So we can compute

$$a_{\mathbf{1}_+} = (\chi_{\mathbf{1}_+}, \chi) = \frac{1}{6}d^3 + \frac{3}{6}d^2 + \frac{2}{6}d = \frac{1}{6}d(d+1)(d+2) \tag{16.11}$$

$$a_{\mathbf{1}_-} = (\chi_{\mathbf{1}_-}, \chi) = \frac{1}{6}d^3 + \frac{3}{6}(-1)\cdot d^2 + \frac{2}{6}d = \frac{1}{6}d(d-1)(d-2) \tag{16.12}$$

$$a_{\mathbf{2}} = (\chi_{\mathbf{2}}, \chi) = \frac{1}{6}2d^3 + \frac{3}{6}0\cdot d^2 + \frac{2}{6}(-1)\cdot d = \frac{1}{3}d(d^2-1) \tag{16.13}$$

Thus, *as a representation of* S_3, we have

$$V^{\otimes 3} \cong \frac{d(d+1)(d+2)}{6}\mathbf{1}_+ \oplus \frac{d(d-1)(d-2)}{6}\mathbf{1}_- \oplus \frac{d(d+1)(d-1)}{3}\mathbf{2} \tag{16.14}$$

Note that the first two dimensions are those of S^3V and $\Lambda^3 V$, respectively, and that the dimensions add up correctly.

We are going to develop this idea in much more detail when we discuss Schur-Weyl duality in section ****

Remarks

• Notice that this does not tell us *how* to block-diagonalize the matrices. We will see how to do that later.

Exercise

Repeat example 2 to decompose $V^{\otimes 2}$ as a representation of S_2.

Exercise

Write out the unitary matrix $S_{\mu i}$ for $G = S_3$.

Exercise

a.) Suppose we tried to define a representation of S_3 by taking $(12) \to 1$ and $(23) \to -1$. What goes wrong?

b.) Show that for any n there are only two one-dimensional representations of S_n

---

16.2 Dihedral groups

The irreps of the dihedral group D_n are easily written down.

Recall that D_n is the group generated by x, y subject to the relations:

$$\begin{aligned} x^2 &= 1 \\ y^n &= 1 \\ xyx^{-1} &= y^{-1} \end{aligned} \tag{16.15}$$

eq:rels

We can define two-dimensional complex representations $T^{(j)}$ by:

$$\begin{aligned} T^{(j)} : y &\to \begin{pmatrix} \omega^j & \\ & \omega^{-j} \end{pmatrix} \\ x &\to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{16.16}$$

eq:twodee

where $\omega = e^{2\pi i/n}$. Clearly, these satisfy the defining relations of D_n.

Note that $T^{(j)} = T^{(j+n)}$, and moreover, $T^{(j)} \cong T^{(n-j)}$ are equivalent representations. Note that $(T^{(j)})^* = T^{(-j)}$ and hence these are all real representations. Indeed, by a similarity transformation we can make y map to a rotation matrix.

Suppose n is odd. Then the representations $T^{(j)}$ are irreducible and inequivalent for $j = 1, \ldots, (n-1)/2$. For $j = 0$ the representation is reducible to $\mathbf{1}_+$ given by $y \to 1, x \to 1$ and $\mathbf{1}_-$ given by $y \to 1, x \to -1$. Altogether we have

$$2n = \frac{n-1}{2} \cdot 2^2 + 1^2 + 1^2 \tag{16.17}$$

so these are the only irreps.

Suppose n is even. Then the representations $T^{(j)}$ are irreducible and inequivalent for $j = 1, \ldots, \frac{n}{2} - 1$. For $j = 0$ the representation is reducible to $\mathbf{1}_{++}$ given by $y \to 1, x \to 1$ and $\mathbf{1}_{+-}$ given by $y \to 1, x \to -1$. For $j = n/2$ the representation is reducible to $\mathbf{1}_{-+}$ given by $y \to -1, x \to 1$ and $\mathbf{1}_{--}$ given by $y \to -1, x \to -1$. Altogether we have

$$2n = (\frac{n}{2} - 1) \cdot 2^2 + 1^2 + 1^2 + 1^2 + 1^2 \tag{16.18}$$

so we have all the representations.

The character tables are:

For $n = 2k + 1$ odd:

	$[1]$	$2k[x]$	$2[y]$	$\cdots$	$2[y^k]$
$\mathbf{1}_+$	1	1	1	$\cdots$	1
$\mathbf{1}_-$	1	-1	1	$\cdots$	1
$\mathbf{2^{(j)}}$	2	0	$\omega^j + \omega^{-j}$	$\cdots$	$\omega^{jk} + \omega^{-jk}$

where $j = 1, \ldots, k$.
For $n = 2k$ even we have
MISSING TABLE
where $j = 1, \ldots, k-1$
**** SHOULD apply these tables to solve some problem ***

17. The decomposition of tensor products

A very common problem one faces is the following:

Suppose we have two representations V_1, V_2. Then the tensor product $V_1 \otimes V_2$ is also a representation.

Question: *If we know how to decompose V_1, V_2 into irreps, how does $V_1 \otimes V_2$ fall apart into its irreducible pieces?*

An example of this general question arises in the quantum mechanical addition of angular momenta. For this reason it is sometimes called the "Clebsch-Gordon problem."

(**** Following should be moved to place where we talk about complete reducibility ****)

If V is a completely decomposable representation (always true for compact groups) then we can write

$$V \cong \oplus_{\mu=1}^{r} \oplus_{i=1}^{a_\mu} T^{(\mu)} := \oplus_\mu a_\mu T^{(\mu)} \tag{17.1}$$

For a fixed μ let let

$$V^{(\mu)} := \oplus_{i=1}^{a_\mu} T^{(\mu)} \tag{17.2}$$

It contains the representation μ with the correct degeneracy in V. It is called the *isotypical component* belonging to μ. Note that it can be written as

$$V^{(\mu)} = \mathbb{C}^{a_\mu} \otimes T^{(\mu)} \tag{17.3}$$

eq:isotpyp

where $\mathbb{C}^{a_\mu}$ is understood to be the trivial representation of G. Thus by Schur's lemma, the intertwiners of $V^{(\mu)}$ with itself are of the form $K \otimes 1$ where $K \in \mathrm{End}(\mathbb{C}^{a_\mu})$ is arbitrary.

We abbreviate $\mathbb{C}^{a_\mu} \otimes T^{(\mu)}$ to $a_\mu T^\mu$ and with this understood we write the decomposition into isotypical components as:

$$V = \oplus a_\mu T^\mu \tag{17.4}$$

Now suppose:

$$V_1 = \oplus a_\mu T^\mu \qquad V_2 = \oplus b_\nu T^\nu \tag{17.5}$$

then

$$V_1 \otimes V_2 = \oplus_{\mu,\nu} a_\mu b_\nu T^\mu \otimes T^\nu \tag{17.6}$$

so in finding the isotypical decomposition of $V_1 \otimes V_2$ it suffices to find the decomposition of products of irreps.

Definition *Fusion Coefficients* The isotypical decomposition of the product of irreps is of the form:

$$T^{(\mu)} \otimes T^{(\nu)} = \oplus_\lambda \Big(\overbrace{T^{(\lambda)} \oplus \cdots \oplus T^{(\lambda)}}^{N^\lambda_{\mu\nu}\ times} \Big) = \oplus_\lambda N^\lambda_{\mu\nu} T^{(\lambda)} \tag{17.7}$$ eq:decomp

The multiplicities $N^\lambda_{\mu\nu}$ are known as *fusion coefficients.*

Exercise
Show that

$$\chi_{T_1 \otimes T_2}(g) = \chi_{T_1}(g)\chi_{T_2}(g) \tag{17.8}$$

Now let us put this observation to use and solve our problem. Taking the trace of (17.7) we get:

$$\chi_\mu(g)\chi_\nu(g) = \sum_\lambda N^\lambda_{\mu\nu}\chi_\lambda(g) \tag{17.9}$$ eq:chardec

Now, taking the inner product we get:

$$N^\lambda_{\mu\nu} = (\chi_\lambda, \chi_\mu\chi_\nu) \tag{17.10}$$ eq:fusion

We can write this out as:

$$N^\lambda_{\mu\nu} = \frac{1}{|G|}\sum_{g\in G} \chi_\mu(g)\chi_\nu(g)\chi_\lambda(g^{-1}) \tag{17.11}$$ eq:fusioni

This can be written in a different way, which we will explain conceptually in the next section. For simplicity choose unitary irreps, and recall that the orthogonality relations on characters is equivalent to the statement that

$$S_{\mu i} := \sqrt{\frac{m_i}{|G|}}\chi_\mu(C_i) \qquad \mu = 1, \ldots, r \quad i = 1, \ldots, r \tag{17.12}$$

is a unitary matrix. Notice that if we take $\mu = 0$ to correspond to the trivial 1-dimensional representation then $S_{0i} = \sqrt{\frac{m_i}{|G|}}$. Thus we can write

$$N^\lambda_{\mu\nu} = \sum_i \frac{S_{\mu i} S_{\nu i} S^*_{\lambda i}}{S_{0i}} \tag{17.13}$$ eq:fusionii

This is a prototype of a celebrated result in conformal field theory known as the "Verlinde formula."

Equations (17.10)(17.11)(17.13) give a very handy way to get the numbers $N^\lambda_{\mu\nu}$. Note that, by their very definition the coefficients $N^\lambda_{\mu\nu}$ are nonnegative integers, although this is hardly obvious from, say, (17.13).

18. Algebras

♣First part with definitions was copied into chapter 2. No need to repeat here. ♣

We first need the general abstract notion: **Definition** An *algebra* over a field k is a vector space V over k with a notion of multiplication of two vectors

$$V \times V \to V \tag{18.1}$$

denoted:

$$v_1, v_2 \in V \to v_1 \cdot v_2 \in V \tag{18.2}$$

which has a ring structure compatible with the scalar multiplication by the field. Concretely, this means we have axioms:

i.) $(v_1 + v_2) \cdot v_3 = v_1 \cdot v_3 + v_2 \cdot v_3$

ii.) $v_1 \cdot (v_2 + v_3) = v_1 \cdot v_2 + v_1 \cdot v_3$

iii.) $\alpha(v_1 \cdot v_2) = (\alpha v_1) \cdot v_2 = v_1 \cdot (\alpha v_2), \qquad \forall \alpha \in k.$

The algebra is *unital*, i.e., it has a unit, if $\exists 1_V \in V$ (*not* to be confused with the multiplicative unit $1 \in k$ of the ground field) such that:

iv.) $1_V \cdot v = v \cdot 1_V = v$

If, in addition, the product satisfies:

$$(v_1 \cdot v_2) \cdot v_3 = v_1 \cdot (v_2 \cdot v_3) \tag{18.3}$$

for all $v_1, v_2, v_3 \in V$ then V is called an *associative algebra*. In general, a *nonassociative algebra* means a not-necessarily associative algebra. In any algebra we can introduce the *associator*

$$[v_1, v_2, v_3] := (v_1 \cdot v_2) \cdot v_3 - v_1 \cdot (v_2 \cdot v_3) \tag{18.4}$$

eq:associator

Note that it is trilinear.

In general, if $\{v_i\}$ is a basis for the algebra then the *structure constants* are defined by

$$v_i \cdot v_j = \sum_k c_{ij}^k v_k \tag{18.5}$$

A *representation* of an algebra V is a vector space W and a map $T : V \to \mathrm{End}(W)$ so that

$$\begin{aligned} T(\alpha v + \beta w) &= \alpha T(v) + \beta T(w) \\ T(v_1 \cdot v_2) &= T(v_1) \circ T(v_2) \end{aligned} \tag{18.6}$$

eq:repalgebra

Example. $M_n(k)$ is a vector space over k of dimension n^2. It is also an associative algebra because matrix multiplication defines an algebraic structure. An important fact is that the only irreducible representation of the algebra $M_n(k)$ is the natural one acting on $W = k^{\oplus n}$. This leads to a concept called "Morita equivalence" of algebras.

Exercise

a.) Write out a basis and structure constants for the algebra $M_n(k)$.

b.) Show that the set of $n \times n$ matrices over an associative algebra, A, denoted $M_n(A)$, is itself an associative algebra.

Exercise

a.) If A is an algebra, then it is a module over itself, via the left-regular representation (LRR). $a \to L(a)$ where

$$L(a) \cdot b := ab \tag{18.7}$$

Show that if we choose a basis a_i then the structure constants

$$a_i a_j = c_{ij}{}^k a_k \tag{18.8}$$

define the matrix elements of the LRR:

$$(L(a_i))_j{}^k = c_{ij}^k \tag{18.9}$$

An algebra is said to be *semisimple* if these operators are diagonalizable.

b.) If A is an algebra, then it is a bimodule over $A \otimes A^o$ where A^o is the opposite algebra.

18.1 Coalgebras, Bialgebras, and Frobenius algebras

Definition A *Frobenius algebra* A is an associative algebra over a field k with a trace $\theta : A \to k$ such that the quadratic form $A \otimes A \to k$ given by $a \otimes b \to \theta(ab)$ defines a nondegenerate bilinear form on A.

Matrix algebras over a field are examples of Frobenius algebras where the nondegenerate bilinear form is the trace map. Frobenius algebras have nice properties with their duals.

The dual space of an algebra V^* has a structure known as a *coalgebra*. The dual of the multiplication is a comultiplication

$$\Delta : V^* \to V^* \otimes V^* \tag{18.10}$$

and the dual of associativity says $(1 \otimes \Delta)\Delta = (\Delta \otimes 1)\Delta$.

The dual of the unit is the counit $\epsilon : V^* \to k$. The dual of the axioms for the unit are

$$(\epsilon \otimes 1)\Delta = (1 \otimes \epsilon)\Delta = 1. \tag{18.11}$$

DIAGRAMS.

A *bialgebra* is an algebra which is simultaneously a coalgebra so that the structures are all compatible, that is, the comultiplication Δ and the counit ϵ are algebra homomorphisms.

Finally, a *Hopf* algebra is a bialgebra with an extra bit of structure $S : H \to H$. It is a generalization of the algebra of functions on a group and is a starting point for the theory of quantum groups.

Exercise

a.) In the literature one sometimes sees a Frobenius algebra defined as an algebra A with a nondegenerate bilinear form

$$\sigma : A \otimes A \to k \tag{18.12}$$ eq:frob

such that $\sigma(xy, z) = \sigma(y, zx)$. Show that this is equivalent to our definition.

b.) Show that if (A, θ) is a Frobenius algebra then the dual algebra A^* is a left A-module which is isomorphic to A as a left A-module.

18.2 When are two algebras equivalent? Introduction to Hochschild cohomology

Deformation - which are trivial? Nontrivial deformations measured by Hochschild coho. View as a flatness condition.

19. The group ring and group algebra

The group algebra is a useful tool for addressing the fundamental questions at the beginning of section ****

1. Useful for producing explicit reduction of matrix reps by the method of projection operators.

2. Useful for producing the representations of the symmetric group S_n.

3. Important and beautiful mathematical structure.

The regular representation is an algebra. Indeed, the space of complex-valued functions $Fun(X)$ on *any* space X is an algebra under pointwise multiplication:

$$(f_1 \cdot f_2)(x) := f_1(x) f_2(x) \tag{19.1}$$

Note that this is a commutative algebra. In particular, we can apply this remark to $X = G$.

Now, the *group ring*, usually denoted $\mathbb{Z}[G]$ is, as an abelian group the free abelian group generatred by G. That is, it is the space of formal linear combinations

$$\mathbb{Z}[G] := \{x = \sum_{g \in G} x_g g : x_g \in \mathbb{Z}\} \tag{19.2}$$

The ring structure is defined by taking

$$\delta_{g_1} *_2 \delta_{g_2} = \delta_{g_1 \cdot g_2} \tag{19.3}$$

and extending by linearity. Thus, we have

$$\begin{aligned} \left(\sum_{g\in G} x(g)g\right) *_2 \left(\sum_{g\in G} y(g)g\right) &:= \sum_{g,h\in G} x(g)y(h) \overbrace{g\cdot h}^{\text{group multiplication}} \\ &= \sum_{k\in G}\left[\sum_{g,h\in G: gh=k} x(g)y(h)\right]k \end{aligned} \tag{19.4}$$

eq:groupalgmul

Replacing $\mathbb{Z} \to \mathbb{C}$ (more properly, taking the tensor product with $\mathbb{C}$) gives the group algebra $\mathbb{C}[G]$.

As a vector space $\mathbb{C}[G]$ is naturally dual to the regular representation: $R_G \cong \mathbb{C}[G]^*$. The natural pairing is

$$\langle f, x\rangle := \sum_{g\in G} x_g f(g) \tag{19.5}$$

Of course, there is a natural basis for $\mathbb{C}[G]$, namely the elements g themselves, and the dual basis is the basis of delta functions δ_g. Thus the isomorphism takes $f : G \to \mathbb{C}$ to $\sum_g f(g)g$.

Now we see that there are *two* algebra structures on R_G and $\mathbb{C}[G]$.

First of all, pointwise multiplication defines an algebra structure on $\mathbb{C}[G]$ since

$$\begin{aligned} f_1 *_1 f_2(g) &\equiv f_1(g)f_2(g) \Rightarrow \\ \delta_{g_1} *_1 \delta_{g_2} &= \delta_{g_1} \qquad \text{if} g_1 = g_2 \\ &= 0 \qquad\quad \text{otherwise} \end{aligned} \tag{19.6}$$

eq:pointwise

which implies

$$g_1 *_1 g_2 = \begin{cases} g_1 & g_1 = g_2 \\ 0 & g_1 \neq g_2 \end{cases} \tag{19.7}$$

eq:pointwiseii

However, the multiplication (19.7) is *not* what is usually understood when one speaks of the group algebra, rather what is meant is the multiplication (19.4).

Under the isomorphism $R_G \cong \mathbb{C}[G]$ the product $*_2$ must correspond to a *second* algebra structure on R_G, and indeed there is one - called the *convolution product*.

$$(f_1 *_2 f_2)(g) := \sum_{h\in G} f_1(h)f_2(h^{-1}g) \tag{19.8}$$

eq:covolution

Note that $*_2$ is therefore, in general, a *noncommutative* algebra product, whereas $*_1$ is a *commutative* algebra structure.

Example $\mathbb{Z}_2 = \{1, \sigma\}$. The group algebra is

$$R_{\mathbb{Z}_2} = \{a \cdot 1 + b \cdot \sigma : a, b \in \mathbb{C}\} \tag{19.9}$$

Now the algebra structure is:

$$(a \cdot 1 + b \cdot \sigma) \cdot (a' \cdot 1 + b' \cdot \sigma) = (aa' + bb')1 + (ba' + ab')\sigma \tag{19.10}$$

These are the "double-numbers," a lesser known cousin of the complex numbers.

Exercise
Which elements of $R_{\mathbb{Z}_2}$ are invertible?

19.1 Projection operators in the group algebra

The group algebra is of course a representation of $G_{\text{left}} \times G_{\text{right}}$ acting as a left action:

$$T(g_1, g_2) \cdot x = g_1 x g_2^{-1} \tag{19.11}$$

for $x \in \mathbb{C}[G]$.

We know from the Peter-Weyl theorem that this representation is highly reducible. We can now construct explicit projection operators which project onto the irreducible components. We assume that the distinct irreps $T^{(\mu)}_{ij}$ are somehow known.

Let us define:

$$P^{\mu}_{ij} \in \mathbb{C}[G] \qquad 1 \le i, j \le n_\mu \tag{19.12}$$ eq:idemps

by

$$P^{\mu}_{ij} := \frac{n_\mu}{|G|} \sum_{g \in G} \hat{T}^{(\mu)}_{ij}(g) g \tag{19.13}$$ eq:projops

Recall that the dual representation is the transpose inverse of the original one so $\hat{T}^{(\mu)}_{ij}(g) = T^{(\mu)}_{ji}(g^{-1})$

These elements satisfy the simple product law:

$$P^{\mu}_{ij} P^{\mu'}_{i'j'} = \delta_{\mu\mu'} \delta_{ji'} P^{\mu}_{ij'} \tag{19.14}$$ eq:prodlaw

They are therefore *orthogonal projection operators* in the group algebra acting on itself.

Moreover, again using similar manipulations one can show that:

$$g \cdot P^{\mu}_{ij} = \sum_s T^{(\mu)}_{si}(g) P^{\mu}_{sj} \tag{19.15}$$ eq:geerep

$$P^{\mu}_{ij} \cdot g^{-1} = \sum_s \hat{T}^{(\mu)}_{sj}(g) P^{\mu}_{is} \tag{19.16}$$ eq:geerepii

and thus the projection operators P^{μ}_{ij} explicitly decompose $\mathbb{C}[G]$ into the irreps $T^{(\mu)} \otimes \hat{T}^{(\mu)}$ of $G_{\text{left}} \times G_{\text{right}}$. This will be useful below in showing how to decompose representations into irreps.

Exercise

Prove (19.14)(19.15) and (19.16) using the orthogonality conditions The manipulations are very similar to those used to prove theorem 10.1.

19.2 The center of the group algebra and the subalgebra of class functions

As we have noted, the group algebra $\mathbb{C}[G]$ is a noncommutative algebra. Correspondingly, R_G is a noncommutative algebra under the convolution product.

A little thought shows that the *center* of the group algebra $Z[\mathbb{C}[G]]$ is spanned by the elements $c_i = \sum_{g \in C_i} g$ where we sum over the conjugacy class C_i of g. This simply follows from

$$g c_i g^{-1} = \sum_{h \in C_i} g h g^{-1} = c_i \tag{19.17}$$

Correspondingly, the center of R_G under the convolution product are the class functions.

Indeed, we can now interpret the exercise with equation (15.18) above as saying that the characters satisfy:

$$\chi_\mu *_2 \chi_\nu = \frac{\delta_{\mu\nu}}{n_\mu} \chi_\mu \tag{19.18}$$

The product is commutative, and in fact, diagonal in this basis. On the other hand, the convolution product is nondiagonal in the conjugacy class basis:

$$\delta_{C_i} *_2 \delta_{C_j} = \sum_k M^k_{ij} \delta_{C_k} \tag{19.19}$$

It is interesting to compare with the pointwise product, where the reverse is true

$$\chi_\mu *_1 \chi_\nu = \sum_\lambda N^\lambda_{\mu\nu} \chi_\lambda \tag{19.20}$$

while

$$\delta_{C_i} *_1 \delta_{C_j} = \delta_{ij} C_i \tag{19.21}$$

So the unitary matrix $S_{\mu i}$ is a kind of Fourier transform which exchanges the product for which the characters of representations or characteristic functions of conjugacy classes is diagonalized:

$$\chi_\mu = \sum_i \sqrt{\frac{|G|}{m_i}} S_{i\mu} \delta_{C_i} \tag{19.22}$$

Exercise

Show that the structure constants in the basis δ_{C_i} in the convolution product can be written as:

$$M_{ij}^{k} = \frac{S_{0i}S_{0j}}{S_{0k}} \sum_{\mu} \frac{S_{i\mu}^{*} S_{j\mu}^{*} S_{k\mu}}{n_{\mu}} \tag{19.23}$$

eq:convolusc

Exercise

An algebra is always canonically a representation of itself in the left-regular representation. The representation matrices are given by the structure constants, thus

$$L(\chi_{\mu})_{\nu}{}^{\lambda} = N_{\mu\nu}^{\lambda} \tag{19.24}$$

shows that the fusion coefficients are the structure constants of the pointwise multiplication algebra of class functions.

Note that since the algebra of class functions is commutative the matrices $L(\chi_\mu)$ can be simultaneously diagonalized.

Show that the unitary matrix $S_{i\mu}$ is in fact the matrix which diagonalizes the fusion coefficients.

---

20. Interlude: 2D Topological Field Theory and Frobenius algebras

The group algebra illustrates the notions of bialgebra and Frobenius algebra very nicely. We would like to explain briefly the relation of these algebraic structures to two-dimensional *topological field theory*.

The axioms of topological field theory give a caricature of what one wants from a field theory: Spaces of states and transition amplitudes. By stripping away the many complications of "real physics" one is left with a very simple structure. Nevertheless, the resulting structure is elegant, it is related to beautiful algebraic structures which, at least in two dimensions, which have surprisingly useful consequences. This is one case where one can truly "solve the theory."

Of course, we are really interested in more complicated theories. But the basic framework here can be adapted to any field theory. What changes is the geometric category under consideration.

20.1 Geometric Categories

What are the most primitive things we want from a physical theory in d spacetime dimensions?

In a physical theory one often decomposes spacetime into space and time as in 7. If space is a $(d-1)$-dimensional manifold time Y then, in quantum mechanics, we associate to it a vector space of states $\mathcal{H}(Y_d)$.

Figure 7: A spacetime $X_d = Y \times \mathbb{R}$. Y is $(d-1)$-dimensional space. fig:tfti

In a generic physical theory the vector space has a lot of extra structure: It is a Hilbert space, there are natural operators acting on this Hilbert space such as the Hamiltonian. The spectrum of the Hamiltonian and other physical observables depends on a great deal of data. Certainly they depend on the metric on spacetime since a nonzero energy defines a length scale

$$L = \frac{\hbar c}{E} \tag{20.1}$$

In topological field theory one ignores most of this structure, and focuses on the dependence of $\mathcal{H}(Y)$ on the topology of Y. For simplicity, we will initially assume Y is compact.

So: We have an association:

$(d-1)$-manifolds Y to vector spaces: $Y \to \mathcal{H}(Y)$, such that homeomorphic manifolds map to isomorphic vector spaces.

Now, we also want to incorporate some form of locality, at the most primitive level. Thus, if we take disjoint unions

$$\mathcal{H}(Y_1 \amalg Y_2) = \mathcal{H}(Y_1) \otimes \mathcal{H}(Y_2) \tag{20.2}$$ eq:disunion

Note that (20.2) implies that we should assign to $\mathcal{H}(\emptyset)$ the field of definition of our vector space. For simplicity we will take $\mathcal{H}(\emptyset) = \mathbb{C}$, although one could use other ground fields.

Figure 8: Time development in X_d gives a linear map. fig:tftitime

In addition, in physics we want to speak of transition amplitudes. If there is a spacetime X_d interpolating between two time-slices, then mathematically, we say there is a cobordism between Y and Y'. That is, a *cobordism* from Y to Y' is a d-manifold with boundary and a partition of its boundary

$$\partial X_d = (\partial X_d)_{\rm in} \cup (\partial X_d)_{\rm out} \tag{20.3}$$

so that there is a homeomorphism $(\partial X_d)_{\rm in} \cong Y$ and $(\partial X_d)_{\rm out} \cong Y'$.

Here we will be considering *oriented* manifolds. Then an oriented cobordism from Y_1 to Y_2 must have

$$\partial M = Y_2 - Y_1 \tag{20.4}$$

where the minus sign indicates the opposite orientation on Y_1. (Recall that an orientation on M determines one on ∂M. We will adopt the convention "outward normal first" - ONF - "One Never Forgets.").

If X_d is an oriented cobordism from Y to Y' then the Feynman path integral assigns a linear transformation

$$F(X_d) : \mathcal{H}(Y) \to \mathcal{H}(Y'). \tag{20.5}$$

Again, in the general case, the amplitudes depend on much more than just the topology of X_d, but in topological field theory they are supposed only to depend on the topology. More precisely, if $X_d \cong X_d'$ are homeomorphic by a homeomorphism $= 1$ on the boundary of the cobordism, then

$$F(X_d) = F(X_d') \tag{20.6}$$

One key aspect of the path integral we want to capture - again a consequence of locality - is the idea of summing over a complete set of intermediate states. In the path integral formalism we can formulate the sum over all paths of field configurations from t_0 to t_2 by composing the amplitude for all paths from t_0 to t_1 and then from t_1 to t_2, where $t_0 < t_1 < t_2$, and then summing over all intermediate field configurations at t_1. We refer to this property as the "gluing property." The gluing property is particularly obvious in the path integral formulation of field theories.

Figure 9: Gluing cobordisms

fig:tftii

In topological field theory this is formalized as:

If M is a cobordism from Y_1 to Y_2 with

$$\partial M = Y_2 - Y_1 \tag{20.7}$$

and M' is another oriented cobordism from Y_2 to Y_3

$$\partial M' = Y_3 - Y_2 \tag{20.8}$$

then we can compose $M' \circ M$ as in 9 to get a cobordism from Y_1 to Y_3.

Naturally enough we want the associated linear maps to compose:

$$F(M \circ M') = F(M') \circ F(M) : \mathcal{H}(Y_1) \to \mathcal{H}(Y_3) \tag{20.9}$$

What we are describing, in mathematical terms is a functor between categories.

Definition A *functor* between categories C_1, C_2 is an association between categories which preserves structure. Thus there is a map on objects: $F : Obj(C_1) \to Obj(C_2)$. Moreover, we have a

a.) *Covariant functor* if $F : Arr(C_1) \to Arr(C_2)$ such that $F(a \to b)$ is an arrow from $F(a)$ to $F(b)$ for all $a, b \in Obj(C_1)$ and

$$F(\phi_2 \circ \phi_1) = F(\phi_2) \circ F(\phi_1) \tag{20.10}$$

b.) *Contravariant functor* if $F : Arr(C_1) \to Arr(C_2)$ such that $F(a \to b)$ is an arrow from $F(b)$ to $F(a)$ for all $a, b \in Obj(C_1)$ and

$$F(\phi_2 \circ \phi_1) = F(\phi_1) \circ F(\phi_2) \tag{20.11}$$

The above axioms can be concisely summarized by defining a cobordism category whose objects are homeomorphism classes of oriented $(d-1)$-manifolds and whose morphisms are oriented cobordisms, two morphisms being identified if they differ by a homeomorphism which is an identity on the boundary. Let us call this $\mathbf{Cob}(d)$.

Definition A d-dimensional topological field theory is a functor from the category $\mathbf{Cob}(d)$ to the category **VECT** of vector spaces and linear transformations.

Actually, these are both *tensor categories* and we want a functor of tensor categories.

At this point we begin to see how we can incorporate other field theories. We can change the geometric category to include manifolds with Riemannian metric, spin structure, etc.

20.2 Some general properties

Let us deduce some simple general facts following from the above simple remarks.

First note that if M is closed then it can be regarded as a cobordism from $\emptyset$ to $\emptyset$. Therefore $F(M)$ must be a linear map from $\mathbb{C}$ to $\mathbb{C}$. But any such linear map is canonically associated to a complex number. We define the *partition function* of M, $Z(M)$ to be this complex number.

There is one cobordism which is distinguished, namely $Y \times [0, 1]$. This corresponds to a linear map $P : \mathcal{H}(Y) \to \mathcal{H}(Y)$. Indeed, physically it is just the operator

$$\exp[-TH] \tag{20.12}$$

where H is the Hamiltonian, and T is the Euclidean time interval (in some presumed metric).

Evidently, by the axioms of topological field theory, $P^2 = P$ and therefore we can decompose

$$\mathcal{H}(Y) = P\mathcal{H}(Y) \oplus (1 - P)\mathcal{H}(Y) \tag{20.13}$$ eq:decompp

All possible transitions are zero on the second summand since, topologically, we can always insert such a cylinder. It follows that it is natural to assume that

$$F(Y \times [0,1]) = Id_{\mathcal{H}(Y)} \tag{20.14}$$ eq:identmap

One can think of this as the statement that the Hamiltonian is zero.

Figure 10: Deforming the cylinder. fig:dfscyl

Now, the cobordism (20.14) is closely related to the cobordism $\emptyset \to Y \cup -Y$ thus defining a map

$$\delta : \mathbb{C} \to \mathcal{H}(Y) \otimes \mathcal{H}(-Y) \tag{20.15}$$

and also to a cobordism $Y \cup -Y \to \emptyset$ thus defining a quadratic form:

$$Q : \mathcal{H}(Y) \otimes \mathcal{H}(-Y) \to \mathbb{C} \tag{20.16}$$

Figure 11: Composing $\delta \otimes 1$ and $1 \otimes Q$ fig:scobord

Let us now compose these cobordisms we get the identity map as in 11. It then follows from some linear algebra that Q is a *nondegenerate* pairing, so we have an isomorphism to the linear dual space:

$$\mathcal{H}(-Y) \cong \mathcal{H}(Y)^*, \tag{20.17}$$

under which Q is just the dual pairing. Moreover $\delta(1)$ has the explicit formula:

$$\delta : 1 \to \sum \phi_\mu \otimes \phi^\mu \tag{20.18}$$

where we can let ϕ_μ be any basis for $\mathcal{H}(Y)$, and then define the dual basis by $Q(\phi_\mu, \phi^\nu) = \delta_\mu{}^\nu$.

Exercise

Prove this formula. Write $\delta(1) = \sum \alpha^{\mu\nu}\phi_\mu \otimes \phi_\nu$ and compute the process in 9.

Figure 12: Composing Q with δ gives the dimension: $\dim\mathcal{H}(Y) = Z(Y \times S^1)$. fig:dimcirc

Now consider the diagram in 12. Geometrically this is a morphism from $\emptyset$ to $\emptyset$, so the functor maps this to a linear map from $\mathbb{C}$ to $\mathbb{C}$. Such a linear map is completely determined by a complex number, and this complex number is the composition $Q\delta$. From our formula for $\delta(1)$ we see that the value $Z(Y \times S^1)$ is just the dimension $\dim\mathcal{H}(Y)$. Thus, in topological field theories, the vector spaces are necessarily finite dimensional.

Note that if we change the category to the category of manifolds with Riemannian structure and we take the product Riemannian structure on $Y \times S^1$ then

$$Z(Y \times S^1) = \mathrm{Tr} e^{-\beta H} \tag{20.19}$$ eq:traccrc

where β is the radius of the circle and H is the Hamiltonian. If in addition we enrich the category to include spin structures then there will be several kinds of traces.

Exercise

Show that a 1-dimensional TFT is completely specified by choosing a single vector space V and a vector in that space.

Exercise

Show that any cobordism of $\emptyset$ to Y defines a *state* in the space $\mathcal{H}(Y)$. This is a primitive version of the notion of the "Hartle-Hawking" state in quantum gravity. It is also related to the state/operator correspondence in conformal field theory.

20.3 Two dimensional closed topological field theory and Commutative Frobenius Algebras

Some beautiful extra structure shows up when we consider the case $d = 2$, due to the relatively simple nature of the topology of 1-manifolds and 2-manifolds. To begin, we restrict attention to closed $(d-1)$-manifolds, that is, we consider a theory of closed strings.

In this case, the spatial $(d-1)$ manifolds are necessarily of the form $S^1 \cup S^1 \cup \cdots \cup S^1$, i.e. disjoint unions of n S^1's.

If we assign $\mathcal{H}(S^1) = \mathcal{C}$ then

$$\mathcal{H}(S^1 \amalg S^1 \amalg \cdots \amalg S^1) = \mathcal{C}^{\otimes n} \tag{20.20}$$

Thus, we have a single basic vector space.

Figure 13: Typical Riemann surface amplitude fig:twodeei

Transition amplitudes can be pictured as in 13. We get

$$F(\Sigma) : \mathcal{C}^{\otimes n} \to \mathcal{C}^{\otimes m} \tag{20.21}$$

where Σ is a Riemann surface. There are n ingoing circles and m outgoing circles.

Now, topology dictates that the vector space $\mathcal{C}$ in fact must carry some interesting extra structure.

Figure 14: The sphere with 3 holes fig:product

In 14 we see that the sphere with three holes defines a product

$$m : \mathcal{C} \otimes \mathcal{C} \to \mathcal{C} \tag{20.22}$$

With the other orientation we get a co-product:

$$\Delta : \mathcal{C} \to \mathcal{C} \otimes \mathcal{C} \tag{20.23}$$

In 15 we see that there is a trace:

$$\theta : \mathcal{C} \to \mathbb{C} \tag{20.24}$$

In 16 we see that there is a map $\mathbb{C} \to \mathcal{C}$. This is completely determined by its value on $1 \in \mathbb{C}$. From the diagram in 17 we see that the image of 1 must be in fact a unit for the muiltiplication.

Figure 15: the trace map fig:trace

Figure 16: The disk defines the unit. fig:unit

Figure 17: Showing that $1_{\mathcal{C}}$ really is the unit. fig:uniti

Moreover, from 18 we see the multiplication is associative.

Finally, we can make a diffeomorphism of the disk with 2 holes, holding the outer circle fixed and rotating the inner two circles. This shows that the product must be *commutative*.

Any oriented surface can be decomposed into the basic building blocks we have used above. However, the same surface can be decomposed in many different ways. When we have different decompositions we get algebraic relations on the basic data $m, \Delta, \theta_{\mathcal{C}}$. At this point you might well ask: "Can we get more elaborate relations on the algebraic data by cutting up complicated surfaces in different ways?" The beautiful answer is: "No, the above relations are the only independent relations." The algebraic structure we have discovered is just that of a *Frobenius algebra* of section *** !

The Sewing Theorem. To give a 2d topological field theory is equivalent to giving a commutative associative finite dimensional Frobenius algebra.

The proof (which is not difficult) depends on Morse theory and will not be given here. For a detailed discussion see, for example,

Figure 18: Associativity

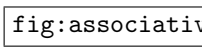
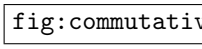


Figure 19: Commutativity. In the supercase we will have graded commutativity: $\phi_1\phi_2 = (-1)^{\deg\phi_1 \deg\phi_2}\phi_2\phi_1$.



G. Segal and G. Moore, "D-branes and K-theory in 2D topological field theory" hep-th/0609042; To appear in *Mirror Symmetry II*, Clay Mathematics Institute Monograph, by P. Aspinwall et. al.

20.4 Boundary conditions

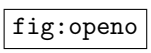
Now let us enrich our theory by allowing the time-slices Y to be manifolds with boundary. There will be a set of boundary conditions $\mathcal{B}_0$, and we will attach an element of $\mathcal{B}_0$ to each boundary component of ∂Y.

A cobordism X from Y_0 to Y_1 will thus have two kinds of boundaries:

$$\partial X = Y_0 \cup Y_1 \cup \partial_{\rm cstr} X \tag{20.25}$$

where $\partial_{\rm cstr}X$ is the time-evolution of the spatial boundaries. We will call this the "constrained boundary."

Figure 20: Morphism space for open strings: $\mathcal{O}_{ab}$.



In $d = 2$, in this enlarged geometric category the initial and final state-spaces are associated with circles, as before, and now also with intervals. The boundary of each interval carries a label $a, b, c, \ldots$ from the set $\mathcal{B}_0$.

Definition: We denote the space $\mathcal{O}_{ab}$ for the space associated to the interval $[0,1]$ with label b at 0 and a at 1.

In the theory of D-branes, the intervals are open strings ending on submanifolds of spacetime. That is why we call the time-evolution of these boundaries the "constrained boundaries" – because the ends are constrained to live in the D-brane worldvolume.

Figure 21: Basic cobordism of open strings. fig:openi

Figure 22: A disk defining an element $1_a \in \mathcal{O}_{aa}$ fig:dsdki

Figure 23: The trace element: $\theta_a : \mathcal{O}_{aa} \to \mathbb{C}$. fig:tftiii

As in the closed case, the cobordism $[0,1] \times [0,1]$ defines $P_{ab} : \mathcal{O}_{ab} \to \mathcal{O}_{ab}$, and we can assume WLOG that it is $P_{ab} = 1$.

Now consider the cobordism in 21. This clearly gives us a bilinear map

$$\mathcal{O}_{ab} \times \mathcal{O}_{bc} \to \mathcal{O}_{ac} \tag{20.26}$$

As in the closed case we see that these maps satisfy an associativity law. Moreover, as in the closed case, 1_a is an identity for the multiplication.

Comparing with the definition of a category we see that we should interpret $\mathcal{B}_0$ as the space of objects in a category $\mathcal{B}$, whose morphism spaces $Hom(b,a) = \mathcal{O}_{ab}$. Note that the morphism spaces are vector-spaces, so we have a $\mathbb{C}$-linear category. In fact, this category

has a very special property as we learn from considering the S-shaped cobordism (the open string analog of 11). We learn that $\theta_a : \mathcal{O}_{aa} \to \mathbb{C}$ defines a nondegenerate inner product:

$$Q_a(\psi_1, \psi_2) = \theta_a(\psi_1 \psi_2) \tag{20.27}$$

Thus, the $\mathcal{O}_{aa}$ are Frobenius algebras.

Moreover, using the S-shaped cobordism analogous to 11 we learn that $\mathcal{O}_{ab}$ is dual to $\mathcal{O}_{ba}$. In fact we have

$$\begin{aligned} \mathcal{O}_{ab} \otimes \mathcal{O}_{ba} \to \mathcal{O}_{aa} &\stackrel{\theta_a}{\to} \mathbb{C} \\ \mathcal{O}_{ba} \otimes \mathcal{O}_{ab} \to \mathcal{O}_{bb} &\stackrel{\theta_b}{\to} \mathbb{C} \end{aligned} \tag{20.28}$$

eq:dualpair

are perfect pairings with

$$\theta_a(\psi_1 \psi_2) = \theta_b(\psi_2 \psi_1) \tag{20.29}$$

eq:cyclic

for $\psi_1 \in \mathcal{O}_{ab}, \psi_2 \in \mathcal{O}_{ba}$.

Definition A *Frobenius category* is a $\mathbb{C}$-linear category in which there is a perfect pairing of $Hom(a,b)$ with $Hom(b,a)$ for all $a, b \in Ob(C)$ by a pairing which factorizes through the composition in either order.

Remark: It is important to note that the argument for commutativity fails in the open case: The algebras $\mathcal{O}_{aa}$ are in general *noncommutative*.

Figure 24: The open-closed transition maps

fig:openclosed

So, to give an open and closed TFT involves giving a Frobenius cateogry. But the open and closed strings must be related to each other. The essential new information is a pair of linear maps

$$\begin{aligned} \iota_a &: \mathcal{C} \to \mathcal{O}_{aa} \\ \iota^a &: \mathcal{O}_{aa} \to \mathcal{C} \end{aligned} \tag{20.30}$$

eq:linmaps

corresponding to the open-closed string transitions of 24.

By drawing pictures we can readily discover the following necessary algebraic conditions:

1. ι_a is an algebra homomorphism

Figure 25: Factorization of the open string loop on closed string exchange. Also known as the "Cardy condition." fig:cardycond

$$\iota_a(\phi_1\phi_2) = \iota_a(\phi_1)\iota_a(\phi_2) \tag{20.31}$$ eq:centerh

2. The identity is preserved

$$\iota_a(1_{\mathcal{C}}) = 1_a \tag{20.32}$$ eq:ident

3. Moreover, ι_a is central in the sense that

$$\iota_a(\phi)\psi = \psi\iota_b(\phi) \tag{20.33}$$ eq:center

for all $\phi \in \mathcal{C}$ and $\psi \in \mathcal{O}_{ab}$

4. ι_a and ι^a are adjoints:

$$\theta_{\mathcal{C}}(\iota^a(\psi)\phi) = \theta_a(\psi\iota_a(\phi)) \tag{20.34}$$ eq:adjoint

for all $\psi \in \mathcal{O}_{aa}$.

5. The "Cardy conditions."[7] Define $\pi_b{}^a : \mathcal{O}_{aa} \to \mathcal{O}_{bb}$ as follows. Since $\mathcal{O}_{ab}$ and $\mathcal{O}_{ba}$ are in duality (using θ_a or θ_b), if we let ψ_μ be a basis for $\mathcal{O}_{ba}$ then there is a dual basis ψ^μ for $\mathcal{O}_{ab}$. Then we define

$$\pi_b{}^a(\psi) = \sum_\mu \psi_\mu\psi\psi^\mu, \tag{20.35}$$ eq:dblttw

and we have the "Cardy condition":

$$\pi_b{}^a = \iota_b \circ \iota^a. \tag{20.36}$$ eq:cardycon

This is illustrated in 25.

Exercise

[7] These are actually generalization of the conditions stated by Cardy. One recovers his conditions by taking the trace. Of course, the factorization of the double twist diagram in the closed string channel is an observation going back to the earliest days of string theory.

Draw pictures associated to the other algebraic conditions given above.

Theorem Open-Closed Sewing Theorem. The above conditions are the complete set of sewing constraints on the algebraic data.

This is proved in the reference cited above.

Example: Let G be a finite group. In section 8 we saw that the class functions on G form a commutative Frobenius algebra. Now we take our category of boundary conditions $\mathcal{B}$ to be the category of finite dimensional complex representations V of G, with $\mathcal{O}_{VV} = End(V)$ and $\theta_V : \mathcal{O}_{VV} \to \mathbb{C}$ given by

$$\theta_V(\psi) = \frac{1}{|G|}\mathrm{Tr}(\psi) \tag{20.37}$$

Exercise

Write out the full set of open-closed string data for the example of a finite group and check the sewing conditions.

---

21. Applications of the Projection Operators

Let us return to the projection operators in $\mathbb{C}[G]$. In this section we will assume - as we may - that the irreps T^μ are unitary so

$$\hat{T}^{(\mu)}_{ij}(g) = (T^{(\mu)}_{ij}(g))^* \tag{21.1}$$

Thus we have

$$P^\mu_{ij} = \frac{n_\mu}{|G|}\sum_{g\in G}(T^{(\mu)}_{ij}(g))^* g \tag{21.2}$$ eq:projopsa

$$P^\mu_{ij}P^{\mu'}_{i'j'} = \delta_{\mu\mu'}\delta_{ji'}P^\mu_{ij'} \tag{21.3}$$ eq:prodlawa

$$g\cdot P^\mu_{ij} = \sum_s T^{(\mu)}_{si}(g)P^\mu_{sj} \tag{21.4}$$ eq:geerepa

$$P^\mu_{ij}\cdot g^{-1} = \sum_s (T^{(\mu)}_{sj}(g))^* P^\mu_{is} \tag{21.5}$$ eq:geerepiia

21.1 Decomposition of a representation into its isotypical parts

While characters tell us the irreps into which a representation can be decomposed, they do not tell us *how* to decompose the representation explicitly. To do this, we need the method of projection operators. We assume that the matrix elements $T^{(\mu)}_{ij}$ are known.

For simplicity we consider finite groups $\mid G \mid < \infty$, but the same techniques extend to compact groups.

Now, let $T(g)$ be *any* representation on V. Extending by linearity we obtain a representation of the algebra R_G. Define operators on V by applying T to (19.12) to get [8]

$$\mathcal{P}^{\mu}_{ij} = T(P^{\mu}_{ij}) = \frac{n_\mu}{|G|}\sum_{g\epsilon G}(T^{(\mu)}_{ij}(g))^* T(g) \tag{21.6}$$

Now we imediately get the properties (19.14), (19.15), (19.16) for the operators $\mathcal{P}^{\mu}_{ij}$.

Therefore, if $\mathcal{P}^{lk}_{\mu}\vec{v} = \vec{w}_l \neq 0$ then

$$\{\vec{w}_j\} = \left\{\mathcal{P}^{jk}_{\mu}\vec{v}\right\}_{j=1,\cdots n_\mu} \tag{21.7}$$

span an invariant subspace of V transforming in the rep $T^{(\mu)}$ according to the representation T^{μ}. Using these operators we can in principle decompose a representation into its irreps.

Exercise *Calculating projection operators*

Now we return to the example of the 3-dimensional reducible representation of S_3 acting on $\mathbb{R}^3$ in section *** above.

Referring to the three irreducible representations of S_3, called $1_+, 1_-, 2$ in section **** show that:

$$\begin{aligned}
P_{1_+} &= \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad P^2_{1_+} = P_{1_+} \\
P_{1_-} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
P^{1,1}_{\mathbf{2}} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix} \qquad (P^{1,1}_{\mathbf{2}})^2 = P^{1,1}_{\mathbf{2}} \\
P^{2,1}_{\mathbf{2}} &= \begin{pmatrix} 0 & 1/\sqrt{3} & -1/\sqrt{3} \\ 0 & -1/2\sqrt{3} & 1/2\sqrt{3} \\ 0 & -1/2\sqrt{3} & 1/2\sqrt{3} \end{pmatrix}
\end{aligned} \tag{21.8}$$

[8] We choose $T^{(\mu)}_{ij}$ = unitary matrices for simplicity.

Therefore,

$$P_{\mathbf{1}+}\begin{pmatrix}x\\y\\z\end{pmatrix} = (\frac{x+y+z}{3})\begin{pmatrix}1\\1\\1\end{pmatrix} \tag{21.9}$$

so:

$$V_1 = Span\left\{\vec{V}_1 = (1,1,1)\right\} = \{(x,x,x) : x \in \mathbb{R}\} \tag{21.10}$$

Similarly $P_{\mathbf{2}}$ projects onto the orthogonal subspace

$$P^{1,1}_{\mathbf{2}}\begin{pmatrix}x\\y\\z\end{pmatrix} = \frac{1}{2}(y-z)\begin{pmatrix}0\\1\\-1\end{pmatrix} \qquad P^{2,1}_{\mathbf{2}}\begin{pmatrix}x\\y\\z\end{pmatrix} = \frac{\sqrt{3}}{2}(y-z)\begin{pmatrix}2\\-1\\-1\end{pmatrix} \tag{21.11}$$

So:

$$V_2 = Span\left\{\vec{V}_2 = (0,1,-1)\vec{V}_3 = (2,-1,-1)\right\} \tag{21.12}$$

are the two invariant subspaces. Check

$$S = \begin{pmatrix}1 & 1 & 1\\0 & 1 & -1\\2 & -1 & -1\end{pmatrix} \tag{21.13}$$

Conjugates the rep into block form. Thus, the decomposition into irreps is:

$$\mathbb{R}^3 \cong V_{\mathbf{1}+} \oplus V_{\mathbf{2}} \tag{21.14}$$ eq:expldec

---

An important shortcut: In general the characters of irreps are much easier to calculate than the matrix elements. Therefore the following route to decomposition is an important shortcut:

If

$$V \cong \oplus_\mu a_\mu T^{(\mu)} \tag{21.15}$$

then we can *easily* project onto the isotypical component as follows: Note that for unitary $T(g)$, $(\mathcal{P}^{ij}_\mu)^* = \mathcal{P}^{ij}_\mu$, so

$$\mathcal{P}^\mu := \sum_{i=1}^{n_\mu} \mathcal{P}^\mu_{ii} \tag{21.16}$$ eq:charporj

form a system of *orthogonal, Hermitian, projection operators*:

$$\begin{aligned}(\mathcal{P}_\mu)^2 &= \mathcal{P}_\mu\\(\mathcal{P}_\mu)^\dagger &= \mathcal{P}_\mu\\\mathcal{P}_\mu\mathcal{P}_\nu &= 0 \qquad \mu \neq \nu\end{aligned} \tag{21.17}$$ eq:projopsaa

They project onto the invariant subspaces $V^{(\mu)}$.

The point is: from the definition (21.16) we get:

$$\mathcal{P}_\mu = \frac{n_\mu}{|G|} \sum_{g \in G} \overline{\chi_\mu(g)} T(g) \tag{21.18}$$

eq:calcpmu

so, we need only know the characters to compute the $\mathcal{P}_\mu$.

Example: In the above example, one easily computes

$$\mathcal{P}_{\mathbf{2}} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \tag{21.19}$$

21.2 Block diagonalization of Hermitian operators

Suppose H is an $N \times N$ Hermitian operator acting on the inner product space $\mathbb{C}^N$ with standard innder product. Suppose we have a group representation T of G with $\mathbb{C}^N$ as carrier space and that moreover, G is a symmetry of H in the sense that

$$\forall g \in G \qquad [T(g), H] = 0 \tag{21.20}$$

Then

$$[\mathcal{P}^\mu_{ij}, H] = 0 \tag{21.21}$$

Then, the projection operators P_λ onto the different eigenspaces of H commute with the different projection operators $\mathcal{P}^\mu_{ij}$ and in particular with $\mathcal{P}^\mu$.

Therefore, in this situation, by reducing $\mathbb{C}^N$ to the irreps of G, we have partially block-diagonalized H.

Of course, this is very useful in quantum mechanics where one wants to diagonalize Hermitian operators acting on wavefunctions.

21.2.1 Projecting quantum wavefunctions

Suppose G acts on X as a tranformation group, and $\{\psi_a\}$ is a collection of functions on X transforming according to some representation T of G. (For example, energy eigenfunctions in a Schrödinger problem.)
Then:

$$\psi^{ij}_{a,\mu}(x) \equiv \frac{n_\mu}{\mid G \mid} \sum_{h \epsilon G} (T^\mu_{ij}(h))^* \psi_a(h^{-1} \cdot x) \tag{21.22}$$

for fixed μ, j, a, letting $i = 1, \ldots, n_\mu$ be a collection of functions transforming according to the irrep T^μ: This is a special case of the general statement we made above. Here, again, is the explicit calculation:

$$\begin{aligned}\psi_a^{ij}(g^{-1}\cdot x) &= \frac{n_\mu}{\mid G \mid}\sum_{h\epsilon G}(T_{ij}^\mu(h))^*\psi_a(h^{-1}\cdot g^{-1}\cdot x)\\ &= \frac{n_\mu}{\mid G \mid}\sum_{h\epsilon G}(T_{ij}^\mu(g^{-1}h))^*\psi_a(h^{-1}\cdot x)\\ &= \sum_{s=1}^{n_\mu}(T_{is}^\mu(g^{-1}))^*\psi_a^{sj}(x)\\ &= \sum_{s=1}^{n_\mu}T_{si}^\mu(g)\psi_a^{sj}(x) \qquad \text{if } T^\mu \text{ is unitary.}\end{aligned} \tag{21.23}$$

Example 2 As a somewhat trivial special case of the above consider $\mathbb{Z}$ acting on $\mathbb{R}$ $x \to -x$ therefore acts on $Fun(\mathbb{R})$. Decompose into irreps:

$\psi =$ function on $\mathbb{R}$

$$\begin{aligned}\psi^1(x) &= \frac{1}{2}(\psi(x)+\psi(-x))\\ \psi^2(x) &= \frac{1}{2}(\psi(x)-\psi(-x))\end{aligned} \tag{21.24}$$

transform acccording to the two irreps of $\mathbb{Z}_2$.

21.2.2 Finding normal modes in classical mechanics

Consider a classical mechanical system with degrees of freedom

$$\vec{q} = (q^1, \ldots, q^n) \tag{21.25}$$

Suppose we have a generalized harmonic oscillator so that the kinetic and potential energies are:

$$T = \frac{1}{2}m_{ij}\dot{q}^i\dot{q}^j \qquad\qquad V = \frac{1}{2}U_{ij}q^iq^j \tag{21.26}$$

where U_{ij} is independent of q^i. We can obtain solutions of the classical equations of motion by taking

$$q^i(t) = \mathrm{Re}(\gamma^i e^{i\omega t}) \tag{21.27}$$

where

$$(-\omega^2 m_{ij} + U_{ij})\gamma^j = 0 \tag{21.28}$$

If a group G acts as $(T(g)\cdot q)^i = T(g)^{ij}q^j$ and

$$[T(g), m] = [T(g), U] = 0 \tag{21.29}$$

then we can partially diagonalize $H = -\omega^2 m + U$ by choosing $\vec{\gamma}$ to be in irreps of G.

Consider n beads on a ring connected by springs of the same force constant and hence the same frequency ω as in 26.

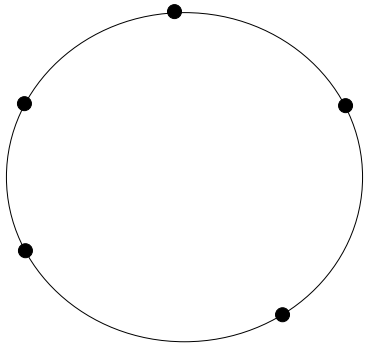


Figure 26: A system of beads and springs. fig:normalmode

The Lagrangian is

$$L = \frac{1}{2}m(\dot{q}^i)^2 - \frac{1}{2}k\sum_i (q^i - q^{i+1})^2 \tag{21.30}$$

where we understand the superscript on q^i to modulo n: $q^{i+n} = q^i$.

We will consider the system where the angle coordinates q^i have been lifted to the universal cover, i.e. $q^i \in \mathbb{R}$ so that we can consider $\vec{q}$ to be in a vector space. [9]

The normal mode problem is

$$(-m\omega^2 + kA)\gamma = 0 \tag{21.31}$$

where

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & \cdots & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & -1 \\ -1 & 0 & \cdots & 0 & 0 & -1 & 2 \end{pmatrix} \tag{21.32}$$

Thus, we reduce the problem of finding normal modes to the problem of diagonalizing this matrix.

The problem has an obvious $\mathbb{Z}_n$ symmetry where a generator $T(\omega)$ takes

$$T(\omega) : q^1 e_1 + \cdots q^n e_n \to q^1 e_2 + \cdots + q^n e_{n-1} + q^1 e_n \tag{21.33}$$

Here e_i is the standard basis for $\mathbb{C}^n$ so $T(\omega^\ell)e_i = e_{i+\ell}$ and again the subscript is understood modulo n.

We know the irreps of $\mathbb{Z}_n$: $T^\mu(\omega^j) = \omega^{j\mu}$, $\mu = 0, \ldots, n-1$. The projection operators are thus

$$\mathcal{P}^\mu = \frac{1}{n}\sum_{\ell=0}^{n-1} \omega^{-\mu\ell} T(\omega^\ell) \tag{21.34}$$

[9] This alters the problem physically. We should add a constraint which says, roughly speaking that $\sum(q^{i+1} - q^i) = 2\pi$. We will ignore this constraint in what follows.

In terms of matrix units e_{ij} we have

$$\mathcal{P}^\mu = \frac{1}{n}\sum_{i,\ell} \omega^{-\mu\ell} e_{i+\ell,i} \tag{21.35}$$

On the other hand,

$$A = 2 - \sum_i (e_{i,i+1} + e_{i+1,i}) \tag{21.36}$$

Now we carry out the multiplication:

$$A\mathcal{P}^\mu = 2\mathcal{P}^\mu - \sum_{i,\ell} \omega^{-\mu\ell}(e_{i+\ell,i+1} + e_{i+\ell,i-1}) \tag{21.37}$$

After some relabeling of the subscripts we easily find

$$\begin{aligned} A\mathcal{P}^\mu &= (2 - \omega^{-\mu} - \omega^{\mu})\mathcal{P}^\mu \\ &= (2\sin\frac{\pi\mu}{n})^2\mathcal{P}^\mu \end{aligned} \tag{21.38}$$

Similarly,

$$\mathcal{P}^\mu \vec{q} = \frac{1}{n}\sum_{\ell,s} \omega^{-\mu\ell} q^s e_{s+\ell} = \frac{1}{n}\sum_s \omega^{\mu s} q^s \left(\sum_j \omega^{-\mu j} e_j\right) \tag{21.39}$$

So the normal mode frequencies are

$$\omega_\mu^2 = 4\frac{k}{m}(2\sin\frac{\pi\mu}{n})^2 \tag{21.40}$$

and the normal modes are:

$$\vec{q}_{(\mu)} = \mathrm{Re}(\alpha_\mu e^{i\omega_\mu t}\sum_j \omega^{-\mu j} e_j) \tag{21.41}$$

Note that there is a *degeneracy* in the mode frequencies

$$\omega_\mu = \omega_{n-\mu}. \tag{21.42}$$

eq:funnydegen

It is a general principle that when you have unexplained degeneracies you should search for a further symmetry in the problem.

Indeed, in this case, we have overlooked a further symmetry:

$$q^i \to q^{n-i} \tag{21.43}$$

This leads to D_n symmetry of the problem which explains the degeneracy (21.42).

Exercise

Using the character tables above for D_n construct the normal modes which are in representations of D_n.

---

22. Representations of the Symmetric Group

22.1 Conjugacy classes in S_n

Recall from chapter one that the conjugacy classes in S_n are labeled by partitions of n: A cycle decomposition has the form

$$(1)^{\nu_1}(2)^{\nu_2}\cdots(n)^{\nu_n} \tag{22.1}$$

where ν_j is the number of cycles of length j. Clearly since we must account for all n letters being permuted:

$$n = \nu_1 + 2\nu_2 + \cdots + n\nu_n = \sum j\nu_j \tag{22.2}$$

To every cycle decomposition we can associate a partition as follows:

$$\begin{aligned}
\lambda_1 &:= \nu_1 + \nu_2 + \cdots + \nu_n \\
\lambda_2 &:= \nu_2 + \cdots + \nu_n \\
\lambda_3 &:= \nu_3 + \cdots + \nu_n \\
&\cdots\cdots \\
\lambda_n &:= \nu_n
\end{aligned} \tag{22.3}$$

eq:lambdaTOnu

Note

1. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$
2. $\sum \lambda_i = n$
3. Given a partition of n we can recover the cycle decomposition $\prod (j)^{\nu_j}$ since the transformation (22.3) is invertible over the integers: $\nu_j = \lambda_j - \lambda_{j+1}$.

As we saw when discussing symmetric functions, to a partition of n we can uniquely associate a picture, called a *Young diagram*, consisting of rows of λ_i boxes, for a total of n boxes.

classes, and hence the irreps of S_n, are in 1-1 correspondence with the partitions of n and hence in 1-1 correspondence with Y

22.2 Young tableaux

Young diagrams can be used to define projection operators onto the irreducible representations in the group algebra $\mathbb{C}[S_n]$ using the following procedure. We only give the barest minimal description here. See the references at the end of this section for a full account.

Define a *Young tableau* to be a Young diagram in which each of the integers $1, 2, \ldots, n$ has been placed into a box. These integers can be placed in any order. A *standard tableau* is one in which the integers are increasing in each row, left to right, and in each column, top to bottom. For any given partition there are $n!$ different Young tableau.

Let T denote a Young tableau. Then we define two subgroups of S_n, $R(T)$ and $C(T)$. $R(T)$ consists of permutations in S_n which only permute numbers within each row of T.

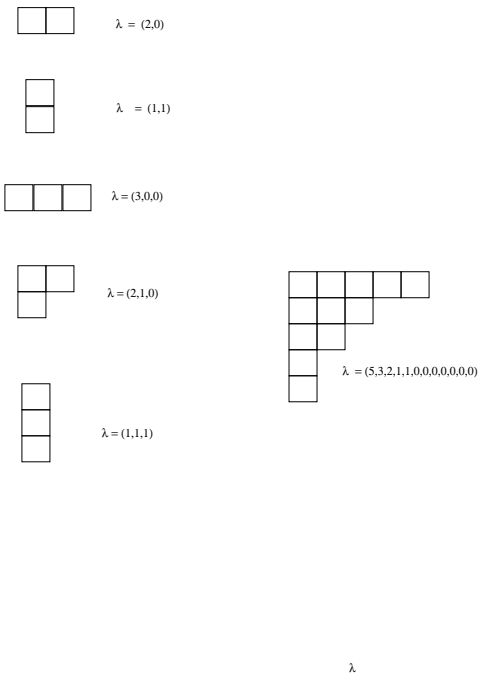


Figure 27: Examples of Young diagrams. fig:youngone

Similarly, $C(T)$ consists of permutations which only permute numbers within each column of T.

For a Young tableau T define the following elements in the group ring $\mathbb{Z}[S_n]$:

$$P := \sum_{p \in R(T)} p \tag{22.4}$$

$$Q := \sum_{q \in C(T)} \epsilon(q) q \tag{22.5}$$

$$P(T) := PQ = \sum_{p \in R(T), q \in C(T)} \epsilon(q) pq \tag{22.6}$$

Then we have the following rather nontrivial statements:

1. $P(T)^2 = c_T P(T)$ where c_T is an integer.

2. $P(T)P(T') = 0$ if T and T' correspond to different partitions.

3. If T, T' are two different standard tableaux corresponding to the same partition then $P(T)P(T') = 0$.

4. $P(T)$ projects onto an irreducible representation of S_n. That is $\mathbb{C}[S_n] \cdot P(T)$ transforms as an irreducible representation $R(T)$ of S_n in the left-regular representation. All of the irreducible representations are equivalent to one of these. Representations $R(T)$ for different partitions are inequivalent, while two representations for the same partition are equivalent.

5. The integer c_T in $P(T)^2 = c_T P(T)$ is given by

$$c_T = \frac{n!}{\dim R(T)} \tag{22.7}$$

$\dim R(T)$ can also be characterized as the number of standard tableaux corresponding to the underlying partition.

6. Another formula for $\dim R(T)$ is the *hook length formula*. For a box in a Young diagram define its hook length to be the number of squares to the right in its row, and underneath in its column, counting that box only once. Then

$$\dim R(T) = \frac{n!}{\prod \text{hooklengths}} \tag{22.8}$$

Figure 28: Four standard Young tableaux for S_3. fig:exmplesthr

22.2.1 Example 1: $G = S_3$

There are 3 partitions, 3 Young diagrams, and 4 standard tableaux shown in 28.

Clearly $P(T_1) = \sum_{p\in S_3} p$ and $P(T_1)^2 = 6P(T_1)$. Moreover $\mathbb{Z}[S_3]\cdot P(T_1)$ is one-dimensional. This is the trivial representation.

Similarly, $P(T_4) = \sum \epsilon(q)q$ spans the one-dimensional sign representation.

There are two standard tableaux corresponding to $\lambda = (2,1)$ with

$$P(T_2) = (1+(12))(1-(13)) = 1+(12)-(13)-(132) \tag{22.9}$$ eq:projteetwo

$$P(T_3) = (1+(13))(1-(12)) = 1-(12)+(13)-(123) \tag{22.10}$$

One checks that $P(T_2)^2 = 3P(T_2)$ and $P(T_3)^2 = 3P(T_3)$ and $P(T_2)P(T_3) = 0$. These are consistent with the above statements about c_T and the hooklength formula.

Now consider $\mathbb{Z}[S_3]\cdot P(T_2)$. We compute

$$\begin{aligned}
(12)\cdot P(T_2) &= P(T_2) := v_1 \\
(13)\cdot P(T_2) &= -1+(13)-(23)+(123) := v_2 \\
(23)\cdot P(T_2) &= -(12)+(23)-(123)+(132) = -v_1 - v_2 \\
(123)\cdot P(T_2) &= -1+(13)-(23)+(123) = v_2 \\
(132)\cdot P(T_2) &= -(12)+(23)-(123)+(132) = -v_1 - v_2
\end{aligned} \tag{22.11}$$

Thus the space is two-dimensional.

From this it is easy to compute: $(12)\cdot v_2 = -v_1 - v_2$, $(13)v_2 = v_1$ and $(23)v_2 = (123)P(T_2) = v_2$ and so in this basis we have a representation generated by

$$\rho(12) = \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \qquad \rho(13) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho(23) = \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \tag{22.12}$$

One easily checks that the character is the same as that we computed above for the **2**.

Of course, similar statements hold for $P(T_3)$, which the reader should check as an exercise.

Figure 29: 10 standard Young tableaux for S_4. fig:exmplesthr

22.2.2 Example 2: $G = S_4$

References

1. For further discussion of the above material see the books by Miller, Hammermesh, Curtis+Reiner, Fulton+Harris *Representation Theory.*

2. There is a second, elegant method for constructing the irreps of the symmetric group using induced representations. See the book by Sternberg for an account.

23. Symmetric groups and tensors: Schur-Weyl duality and the irreps of $GL(d, k)$

Consider a vector space V of dimension d.

In previous sections we have seen that

$$V^{\otimes n} \equiv V \otimes V \otimes \cdots \otimes V \tag{23.1}$$

is a representation of S_n by permuting the factors:

$$\sigma \cdot \ (w_1 \otimes w_2 \otimes \cdots \otimes w_n) = w_{\sigma(1)} \otimes w_{\sigma(2)} \otimes \cdots \otimes w_{\sigma(n)} \tag{23.2}$$

*** ?? should we use σ^{-1} here ?? ***

Now recall that we also characterized vectors in $V^{\otimes n}$ as tensors. If V is a representation of a matrix subgroup G of $GL(d, \mathbb{R})$ or $GL(d, \mathbb{C})$ then $V^{\otimes n}$ is also a representation of G via

$$\rho(g) \cdot (w_1 \otimes w_2 \otimes \cdots \otimes w_n) = g \cdot w_1 \otimes g \cdot w_2 \otimes \cdots \otimes g \cdot w_n \tag{23.3}$$

For example, suppose that $V = k^d$ is also the fundamental representation of $GL(d, \mathbb{R})$ or $GL(d, \mathbb{C})$ for $k = \mathbb{R}$ or $k = \mathbb{C}$. Given a basis $\{v_i\}$ for V a typical element can be expanded in the basis as:

$$v = \sum_{i_1, i_2, \ldots, i_n} t^{i_1, i_2, \ldots, i_n} v_{i_1} \otimes \cdots \otimes v_{i_n} \tag{23.4}$$

We will often assume the summation convention where repeated indices are automatically summed. Under the action of $GL(d, k)$:

$$g \cdot v_i = g_{ji} v_j \tag{23.5}$$

we therefore have:

$$(g \cdot t)^{i_1 \cdots i_n} = t^{j_1 \cdots j_n} g_{j_1 i_1} \cdots g_{j_n i_n} \tag{23.6}$$

eq:tensor

that is, elements in $V^{\otimes n}$ transform as tensors and $V^{\otimes n}$ is *also* a representation of $GL(d, \mathbb{R})$: On the other hand

$$(\sigma \cdot t)^{i_1 \ldots i_n} = t^{i_{\sigma^{-1}(1)} \cdots i_{\sigma^{-1}(n)}} \tag{23.7}$$

Now, note that the actions of $GL(d, k)$ and S_n *commute*, so that $V^{\otimes n}$ is a representation of $GL(d, k) \times S_n$.

Considered as a representation of S_n, we have complete reducibility:

$$V^{\otimes n} \cong \oplus_\lambda \mathcal{V}_\lambda \otimes R_\lambda \tag{23.8}$$

where R_λ is the irrep of S_n corresponding to the partition λ. By Schur's lemma we can characterize the multiplicity space $\mathcal{V}_\lambda$ as

$$\mathcal{V}_\lambda = \mathrm{Hom}_{S_n}(R_\lambda, V^{\otimes n}). \tag{23.9}$$

This is the vector space of S_n-equivariant maps from R_λ to $V^{\otimes n}$.

Now, because $GL(d, k)$ and S_n commute we know that $\mathcal{V}_\lambda$ will also be a representation of $GL(d, k)$.

The beautiful result of Schur and Weyl – known as the Schur-Weyl duality theorem – is

Theorem The representations $\mathcal{V}_\lambda$ are irreducible representations of $GL(d, k)$ and, up to a power of the determinant representation, all irreducible representations may be obtained in this way. Moreover, these statements hold for $U(d) \subset GL(d, \mathbb{C})$.

Proof For complete proofs see Segal, Miller, Fulton-Harris. We will just sketch the idea from Segal's lectures.

We can see fairly easily why all the representations of $U(d)$ must occur if we accept the Peter-Weyl theorem. Note that the representations which we obtain from $V^{\otimes d}$ have matrix elements which are algebraic expressions in the matrix elements a_{ij}. Indeed, they lie in $\mathbb{C}[a_{ij}]$ (the a_{ij} occur individually of course from V itself). But this forms a dense subalgebra of $L^2(U(d))$.

Now let us prove the $\mathcal{V}_\lambda$ is irreducible. Recall that the center of a matrix algebra $End(W)$ is just $\mathbb{C}$. Now we note that

$$\mathrm{End}_{S_n}(V^{\otimes n}) = ((\mathrm{End}(V))^{\otimes n})^{S_n} \tag{23.10}$$

Since $Gl(d)$ forms a dense open set in $End(V)$, the operators $T(g)$ acting on $V^{\otimes n}$ generate an algebra which is all of $\mathrm{End}_{S_n}(V^{\otimes n})$. However, this means that the operators that commute with both G and S_n, i.e.

$$\mathrm{End}_{G\times S_n}(V^{\otimes n}) \tag{23.11}$$

must lie in the center of $\mathrm{End}_{S_n}(V^{\otimes n})$. On the other hand,

$$\mathrm{End}_{S_n}(V^{\otimes n}) = \oplus_\lambda \mathrm{End}(\mathcal{V}_\lambda) \tag{23.12}$$

and

$$\mathrm{End}_{G\times S_n}(V^{\otimes n}) = \oplus_\lambda \mathrm{End}_G(\mathcal{V}_\lambda \tag{23.13}$$

Therefore $\mathrm{End}_G(\mathcal{V}_\lambda$ must be contained in the center of $\mathrm{End}(\mathcal{V}_\lambda)$, but the latter is just $\mathbb{C}$. Therefore $\mathrm{End}_G(\mathcal{V}_\lambda$ is just $\mathbb{C}$. But, by Schur's lemma, this must mean that $\mathcal{V}_\lambda$ is irreducible. ♠

We can construct $\mathcal{V}_\lambda$ explicitly by taking a Young diagram and its corresponding symmetrizer $P(T)$ (for some tableau T) and taking the image of $P(T)$ acting on $V^{\otimes n}$. This projects onto tensors of a definite symmetry type. These tensors will transform in irreducible representations of $GL(d,k)$. As representations of $GL(d,k)$ we have

$$P(T)V^{\otimes n} \cong P(T')V^{\otimes n} \tag{23.14}$$

if T and T' correspond to the same partition, i.e. the same Young diagram.

With some work it can be shown that - as representations of $GL(d,k)$ we have an orthogonal decomposition

$$V^{\otimes n} = \oplus_T P(T)V^{\otimes n} \tag{23.15}$$

where T runs over the *standard* tableaux of S_n. Of course, tableaux with columns of length $\geq d$ will project to the zero vector space and can be omitted.

Example

• If we take the partition $\lambda = (n, 0, 0, \dots)$ then $P(\lambda)$ projects to totally symmetric tensors and we get $S^n(V)$. Weyl's theorem tells us that if $V = k^d$ is the fundamental representation of $GL(d,k)$ then $S^n(V)$ is an irreducible representation. We computed before that its dimension is

$$\dim S^n(V) = \binom{n+d-1}{n} \tag{23.16}$$

Note in particular that we have infinitely many irreducible representations.

• If we take the partition $\lambda = (1, 1, \ldots, 1)$ then $P(\lambda)V^{\otimes n} = \Lambda^n V$ is the subspace of totally antisymmetric tensors of dimension

$$\dim\Lambda^n(V) = \binom{d}{n} \tag{23.17}$$

Note that this subspace vanishes unless $\dim V = d \geq n$.

• Let V be of dimension d and consider $V^{\otimes 3}$ as a representation of S_3. Using characters we showed above that

$$V^{\otimes 3} \cong \frac{d(d+1)(d+2)}{6}\mathbf{1}_+ \oplus \frac{d(d-1)(d-2)}{6}\mathbf{1}_- \oplus \frac{d(d+1)(d-1)}{3}\mathbf{2} \tag{23.18}$$

Clearly, the first and second summands correspond to the totally symmetric and totally antisymmetric tensors, respectively.

Applying the projector (22.9) to a tensor $t^{ijk}v_i \otimes v_j \otimes v_k$ produces a tensor with mixed symmetry:

$$P(T_2)v_i \otimes v_j \otimes v_k = v_i \otimes v_j \otimes v_k + v_j \otimes v_i \otimes v_k - v_k \otimes v_j \otimes v_i - v_k \otimes v_i \otimes v_j \tag{23.19}$$

and therefore the components are of the form:

$$t^{ijk} + t^{jik} - t^{kji} - t^{jki} \tag{23.20}$$

for arbitrary t^{ijk}. This tensor space is the space of tensors $\tilde{t}^{ijk}$ which satisfy the identities:

$$\tilde{t}^{ijk} + \tilde{t}^{jki} + \tilde{t}^{kij} = 0 \tag{23.21}$$

eq:cyclsym

together with

$$\tilde{t}^{ijk} = -\tilde{t}^{kji} \tag{23.22}$$

eq:iksymm

The first set of equations (23.21) cuts out a space of dimension $\frac{2}{3}d(d^2-1)$. The next set of equations (23.22) cuts it down by half so we get $\frac{1}{3}d(d^2-1)$ as the dimension of this space of tensors.

Similarly,

$$P(T_3)v_i \otimes v_j \otimes v_k = v_i \otimes v_j \otimes v_k - v_j \otimes v_i \otimes v_k + v_k \otimes v_j \otimes v_i - v_j \otimes v_k \otimes v_i \tag{23.23}$$

and therefore the components are of the form:

$$t^{ijk} - t^{jik} + t^{kji} - t^{kij} \tag{23.24}$$

for arbitrary t^{ijk}. Thus, this tensor space is the space of tensors $\tilde{t}^{ijk}$ which satisfy the identities:

$$\tilde{t}^{ijk} + \tilde{t}^{jki} + \tilde{t}^{kij} = 0 \tag{23.25}$$

$$\tilde{t}^{ijk} = -\tilde{t}^{jik} \tag{23.26}$$

Remarks The algebra of operators $\mathcal{A}$ generated by linear combinations of σ acting on $V^{\otimes n}$ commutes with the algebra of operators $\mathcal{B}$ generated by $g \in GL(d, k)$ acting on $V^{\otimes n}$. In fact these algebras are full commutants of each other: Any operator commuting with $\mathcal{A}$ is in $\mathcal{B}$ and vice versa.

23.1 Free fermions on a circle and Schur functions

Figure 30: Fermi sea and particle/hole excitations. fig:Fermisea

Consider a system of N free fermions on a circle $z = e^{i\theta}$, $\theta \sim \theta + 2\pi$, with Hamiltonian $H = -\sum_i \frac{d^2}{d\theta_i^2}$. The one-particle wavefunctions are

$$\psi_n(z) = z^n \tag{23.27}$$ eq:onpart

with energy n^2. Consider a state of N fermions occupying levels n_i. By Fermi statistics these levels must all be distinct and we can assume

$$n_N > n_{N-1} > \cdots > n_1 \tag{23.28}$$

The N-fermion wavefunction is the "Slater determinant"

$$\Psi_{\vec{n}}(z_1, \ldots, z_N) = \det_{1 \le i,j \le N} z_i^{n_j} \tag{23.29}$$ eq:slater

and has energy $\sum_i n_i^2$. States can be visualized as in 30.

For simplicity let us assume N is odd. Then there is a unique groundstate obtained by filling up the states $-n_F \le n \le n_F$ with

$$n_F = \frac{N-1}{2} \tag{23.30}$$ eq:fermil

(If N is even there are two groundstates.) This defines the Fermi sea. The ground state wavefunction is

$$\Psi_{gnd}(z_1, \ldots, z_N) = \det_{1 \le i,j \le N} z_i^{j-1-n_F} = (z_1 \ldots z_N)^{-n_F} \Delta_0(z) \tag{23.31}$$ eq:grnd

The ratio of the wavefunction to groundstate wavefunction

$$\frac{\Psi_{\vec{n}}}{\Psi_{gnd}} = \frac{\det_{1 \le i,j \le N} z_i^{n_j}}{\det_{1 \le i,j \le N} z_i^{j-1-n_F}} \tag{23.32}$$ eq:srat

is a totally symmetric function of the z_i. If $n_1 \ge -n_F$ then it is a symmetric polynomial known as a *Schur function*.

The Schur functions form a linear basis for $\mathbb{Z}[x_1, \ldots, x_N]^{S_N}$ where basis elements are associated with partitions. To define them mathematically let us return to a monomial $x^\alpha = x_1^{\alpha_1} \cdots x_N^{\alpha_N}$ with the $\alpha_i \geq 0$ and now try to skew-symmetrize:

$$\sum_{\sigma \in S_N} \epsilon(\sigma) \sigma \cdot x^\alpha \tag{23.33}$$

Note that this function is totally antisymmetric, and moreover vanishes unless the α_i are all distinct. WLOG assume that

$$\alpha_N > \alpha_{N-1} > \cdots > \alpha_1 \geq 0. \tag{23.34}$$

The "smallest" such α is

$$N-1 > N-2 > \cdots > 1 > 0 \geq 0 \tag{23.35}$$

defining a vector $\delta := (N-1, N-2, \ldots, 1, 0)$ i.e. $\delta_j = j-1$, $1 \leq j \leq N$. Using δ we define λ by

$$\alpha = \lambda + \delta \tag{23.36}$$

Note that $\lambda_j = \alpha_j - \delta_j \geq 0$ and moreover

$$\lambda_j + j - 1 > \lambda_{j-1} + j - 2 \qquad \Rightarrow \qquad \lambda_j \geq \lambda_{j-1} \tag{23.37}$$

Thus, λ is a partition. It is more convenient to use λ than α so we define:

$$\Delta_\lambda(x) := \sum_{\sigma \in S_N} \epsilon(\sigma) \sigma \cdot x^{\lambda + \delta} \tag{23.38}$$

Note that this can also be written as a determinant:

$$\begin{aligned} \Delta_\lambda(x) &:= \det(x_i^{\lambda_j + j - 1}) \\ &= \det \begin{pmatrix} x_1^{\lambda_1} & x_1^{\lambda_2+1} & \cdots & x_1^{\lambda_N+N-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_N^{\lambda_1} & x_N^{\lambda_2+1} & \cdots & x_N^{\lambda_N+N-1} \end{pmatrix} \end{aligned} \tag{23.39}$$

eq:deltsun

Note that $\Delta_0(x)$ is the famous Vandermonde determinant:

$$\Delta_0(x) := \prod_{i>j} (x_i - x_j) \tag{23.40}$$

Exercise
Show that

$$\Delta_\lambda(\sigma \cdot x) = \epsilon(\sigma) \Delta_\lambda(x) \tag{23.41}$$

where $\epsilon(\sigma)$ is the sign homomorphism.

Now we define the *Schur function* to be the ratio:

$$\Phi_\lambda(x) := \frac{\Delta_\lambda(x)}{\Delta_0(x)} \tag{23.42}$$ eq:schurf

This is a totally symmetric function. In fact, it is a polynomial in the x_i. To see this consider it as a meromorphic function of a complex variable x_1. Note that it is in fact an entire function since the potential poles at $x_1 = x_i$ are cancelled by zeroes of the numerator. Next note that the growth at infinity is obviously x_1^m for an integer m.

Example. Put $N = 2$. Then $\lambda_2 \geq \lambda_1 \geq 0$,

$$\Delta_\lambda(x) = (x_1 x_2)^{\lambda_1}(x_2^{\lambda_2-\lambda_1+1} - x_1^{\lambda_2-\lambda_1+1}) \tag{23.43}$$ eq:deltatwo

$$\Phi_\lambda(x_1, x_2) = (x_1 x_2)^{\lambda_1} \frac{x_2^{\lambda_2-\lambda_1+1} - x_1^{\lambda_2-\lambda_1+1}}{x_2 - x_1} = (x_1 x_2)^{\lambda_2} h_{\lambda_2-\lambda_1}(x_1, x_2) \tag{23.44}$$ eq:phitwo

Returning to the free fermion example, we have

$$\frac{\Psi_{\vec{n}}}{\Psi_{\vec{0}}} = \frac{\Delta_\lambda(z)}{\Delta_0(z)} \tag{23.45}$$ eq:fration

for

$$\begin{aligned} \alpha_j &= n_j + n_F \\ &= \lambda_j + j - 1 \end{aligned}, \tag{23.46}$$

that is

$$\lambda_j = (n_j + n_F) - (j - 1) \tag{23.47}$$ eq:snet

Remarks

• The second main theorem of symmetric polynomials is:

Theorem. The Schur functions $\Phi_\lambda(x)$ form a linear integral basis for $\mathbb{Z}[x_1, \ldots, x_N]^{S_N}$.

That is, any symmetric polynomial with integral coefficients can be written as a linear combination, with integral coefficients, of the $\Phi_\lambda(x)$. Note that we are *not* forming polynomials in the Φ_λ.

• The module A_N of skew-symmetric polynomials in $x_1, \ldots, x_N$ is isomorphic to Λ_N via multiplication by Δ_0. Therefore the Δ_λ form a a $\mathbb{Z}$-basis for A_N. This close relation between completely symmetric and antisymmetric functions comes up in the theory of matrix models – integrals over space of $N \times N$ matrices. It also suggests a relation between bosons and fermions, at least in $1+1$ dimensions. That indeed proves to be the case – there

is a nontrivial isomorphism known as *bosonization* which is an isomorphism of quantum field theories of bosons and fermions in 1+1 dimensions.

•

There is an elegant expression for the Schur functions Φ_λ as a determinant of a matrix whose entries involve the h_k. The expression is

$$\Phi_\lambda = \det(h_{\lambda_i - i+j}) \tag{23.48}$$ eq:phiach

where we take an $N \times N$ matrix for N variables and it is understood that $h_0 = 1$ and $h_i = 0$ for $i < 0$. Equivalently

$$\Phi_\lambda = \det(e_{\lambda'_i - i+j}) \tag{23.49}$$ eq:phiacha

For a proof see Macdonald, p.41.

Example Consider the partition $(\lambda_1, \lambda_2, 0, 0, \dots)$. Then applying (23.48) we get

$$\begin{aligned}\Phi_\lambda &= \begin{pmatrix} h_{\lambda_1} & h_{\lambda_1+1} & h_{\lambda_1+2} & h_{\lambda_1+3} & h_{\lambda_1+4} & \cdots \\ h_{\lambda_2-1} & h_{\lambda_2} & h_{\lambda_2+1} & h_{\lambda_2+2} & h_{\lambda_2+3} & \cdots \\ 0 & 0 & 1 & h_1 & h_2 & \cdots \\ 0 & 0 & 0 & 1 & h_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \\ &= h_{\lambda_1} h_{\lambda_2} - h_{\lambda_1+1} h_{\lambda_2-1}\end{aligned} \tag{23.50}$$ eq:phitwoa

With a little algebra one can confirm that this is indeed (23.44). Note, however, that both sides of this equality make sense for $N > 2$.

Figure 31: A matrix for defining Schur functions. fig:symsq

• Define the $\mathbb{Z} \times \mathbb{N}$ matrix Ξ whose $(p,q)^{th}$ entry is h_{q-p} where $h_0 = 1$ and $h_k = 0$ for $k < 0$. Here $-\infty < p < \infty$ labels the rows, while $q = 1, 2, \dots$ labels the columns. Thus the matrix looks like 31.

Define the integers $s_k = k - \lambda_k$. These eventually become $s_k = k$. Then Φ_λ is the determinant of the matrix formed from Ξ by keeping only the rows labeled by s_k. This leads to an upper triangular matrix is the determinant of the finite matrix above.

• The definition of Φ_λ only makes sense as a function of N such that $\lambda_i = 0$ for $i > N$. However, the relations between Φ_λ and the h_i are universal.

Exercise
Verify that (23.50) and (23.44) are equal for $N = 2$.

Exercise
a.) Let $\lambda = (n, 0, 0, 0, \dots)$. Show that $\Phi_\lambda = h_n$.
b.) Let $\lambda = (1^n, 0, 0, \dots)$. Then $\Phi_\lambda = e_n$.

Exercise
Using Cauchy's determinant identity:

$$\det\left(\frac{1}{1 - x_i y_j}\right) = \frac{\Delta_0(x_i)\Delta_0(y_j)}{\prod_{i,j}(1 - x_i y_j)} \tag{23.51}$$

Show that

$$\exp\left[\sum_k \frac{1}{k} s_k(x) s_k(y)\right] = \sum_\lambda \Phi_\lambda(x)\Phi_\lambda(y) \tag{23.52}$$

Hint: Multiply both sides by $\Delta_0(x)\Delta_0(y)$. For the answer see Hammermesh, p. 195

Exercise *Laughlin's wavefunctions*
Another natural place where symmetric wavefunctions appear is in the quantum Hall effect, where the 1-body wavefunctions in the lowest Landau level are

$$\psi_n(z) = \frac{1}{\sqrt{\pi n!}} z^n e^{-\frac{1}{2}|z|^2} \tag{23.53}$$

Laughlin introduced a fascinating set of approximate eigenfunctions of 2d *interacting* electrons in a magnetic field:

$$\Psi = \prod_{i<j} (z_i - z_j)^{2n+1} e^{-\frac{1}{2}\sum |z_i|^2} \tag{23.54}$$

Express these in terms of Schur functions.
Many other interesting trial wavefunctions in the FQHE can be generated using theorems about symmetric functions.

23.1.1 Schur functions, characters, and Schur-Weyl duality

Now it is interesting to combine Schur functions with Schur-Weyl duality.

We have seen that irreducible representations of $GL(d,\mathbb{C})$ (and of $U(d)$) can be labeled by Young diagrams with $\leq d$ rows. We called these S_λ above.

It turns out that the $\Phi_\lambda(x)$ are characters of the representations S_λ. That is if $g \in GL(d,\mathbb{C})$ can be diagonalized to $Diag\{x_1,\ldots,x_d\}$ then

$$\mathrm{Tr}_{\mathcal{V}_\lambda}\rho(g) = \Phi_\lambda(x_1,\ldots x_d) \tag{23.55}$$

This is known as the Weyl character formula and will be derived later.

The relation between the Φ_λ and the power functions $s_k(x)$ is a very nice application of the Schur-Weyl duality theorem.

Suppose that a group element $\sigma \in S_N$ has cycle decomposition $(1)^{\ell_1}(2)^{\ell_2}\cdots$ where $\sum j\ell_j = N$. As we have discussed this determines a conjugacy class $C[\ell]$ as well as a partition. Now, suppose $g \in GL(d,\mathbb{C})$ is diagonalizable to $\mathrm{Diag}\{x_1,\ldots,x_d\}$ and let us evaluate

$$\mathrm{Tr}_{V^{\otimes N}}(\sigma g) \tag{23.56}$$

eq:keytrace

where $V = \mathbb{C}^d$ is the fundamental representation of $GL(d,\mathbb{C})$.

We will evaluate (23.56) in two ways. On the one hand, it is clear that this should just be

$$(\mathrm{Tr}(g))^{\ell_1}(\mathrm{Tr}(g^2))^{\ell_2}(\mathrm{Tr}(g^3))^{\ell_3}\cdots \tag{23.57}$$

eq:prodns

On the other hand, since

$$V^{\otimes N} \cong \oplus_\lambda S_\lambda \otimes R_\lambda \tag{23.58}$$

the trace must also be

$$\sum_\lambda \chi_\lambda(\sigma)\Phi_\lambda(x) \tag{23.59}$$

Thus, recalling the symmetric functions which are the power sum functions $s_j = \sum_i x_i^j$ we have the following identity due to Frobenius:

Let $(\ell) = (\ell_1,\ell_2,\ldots,\ell_N)$ denote a tuple of nonnegative integers. Then let

$$s_{(\ell)} := s_1^{\ell_1}s_2^{\ell_2}\ldots s_N^{\ell_N} \tag{23.60}$$

Then:

$$\boxed{s_{(\ell)} = \sum_{\{\lambda\}} \chi_\lambda(C[\ell]))\Phi_\lambda} \tag{23.61}$$

eq:frobenius

Returning to the free fermion interpretation we see that the operator

$$\mathcal{O}(\ell) = \prod_j (\mathrm{Tr}(g^j))^{\ell_j} \tag{23.62}$$

acts on the groundstate wavefunction $\Psi_{gnd}(x)$ to produce the quantum states:

$$\mathcal{O}(\ell)\Psi_{gnd}(x) = \sum_{\lambda} \chi_\lambda(C[\ell]))\Psi_{\vec{n}} \tag{23.63}$$ eq:transamp

where the n_i are the fermion occupation numbers corresponding to the partition associated to λ as in (23.47). Thus the characters of the symmetric group can be regarded as "transition amplitudes" in the free fermion problem!

Indeed, we can now write the character table of the symmetric group S_N:

$$\chi_\lambda(C[\ell]) = \left[\Delta_0(x)s_{(\ell)}(x)\right]_{\alpha_1,\ldots,\alpha_d} \tag{23.64}$$ eq:charactsymm

where the subscript means we extract the coefficient of $x_1^{\alpha_1}\cdots x_d^{\alpha_d}$ and we recall that

$$\alpha_j = \lambda_j + j - 1 \tag{23.65}$$

This is the famous formula of Frobenius.

Remark: If we think of Φ_λ as a character of a representation of $U(N)$. Then the Hamiltonian H is closely related to the quadratic Casimir of that representation. That in turn is related to the solution of two-dimensional Yang-Mills theory. See Cordes et. al. eq. (4.3)

Exercise *Dimensions of irreps of S_N*

a.) Using the Frobenius formula show that the character of 1, and hence the dimension of R_λ is

$$\dim R_\lambda = \frac{N!}{\alpha_1!\cdots\alpha_N!}\prod_{i>j}(\alpha_i - \alpha_j) \tag{23.66}$$

b.) Using this formula, derive the hook-length formula.

Warning: This is hard.

Exercise *Weyl dimension formula*

Show that

$$\Phi_\lambda(1,1,\ldots,1) = \prod_{1\le i<j\le N}\frac{(\lambda_i - \lambda_j + j - i)}{(j-i)} \tag{23.67}$$ eq:weyldim

We will see later that this is a special case of the Weyl dimension formula.

Hint: Put $x_j = e^{jt}$ and let $t \to 0$.

References:

There are a lot of beautiful connections here with quantum field theory. In particular there are close connections to two-dimensional Yang-Mills theory and to the phenomenon of bosonization in two dimensions. See

For much more about this see the references below.

1. Pressley and Segal, *Loop Groups*, Chapter 10.

2. M. Stone, "Schur functions, chiral bosons, and the quantum-Hall-effect edge states," Phys. Rev. **B42** 1990)8399 (I have followed this treatment.)

3. S. Cordes, G. Moore, and S. Ramgoolam, "Lectures on 2D Yang-Mills Theory, Equivariant Cohomology and Topological Field Theories," Nucl. Phys. B (Proc. Suppl 41) (1995) 184, section 4. Also available at hep-th/9411210.

4. M. Douglas, "Conformal field theory techniques for large N group theory," hep-th/9303159.

5. Papers of Jimbo, Miwa, ... (Kyoto school) on Toda theory.

****

FOLLOWING MATERIAL SHOULD BE MOVED TO CHAPTER ON REPRESENTATIONS. IT WOULD MAKE MORE SENSE TO TALK ABOUT REPS OF THE SYMMETRIC GROUP FIRST.

23.2 Bosons and Fermions in 1+1 dimensions

23.2.1 Bosonization

Finally, we want to describe a truly remarkable phenomenon, that of bosonization in 1+1 dimensions.

Certain quantum field theories of bosons are equivalent to quantum field theories of fermions in 1+1 dimensions! Early versions of this idea go back to Jordan [10] The subject became important in the 1970's. Two important references are

S. Coleman, Phys. Rev. **D11**(1975)2088

S. Mandelstam, Phys. Rev. **D11** (1975)3026

The technique has broad generalizations, and plays an important role in string theory.

To get some rough idea of how this might be so, let us consider the *loop group* $S^1 \to U(1)$. For winding number zero, loop group elements can be written as:

$$z \to g(z) = \exp[i \sum_{n=-\infty}^{+\infty} j_n z^n] \tag{23.68}$$

with $j_n^* = j_{-n}$. This group acts on the one-body wavefunction by

$$z^n \to g(z) z^n \tag{23.69}$$

where $g(z)$ is in the loop group of $U(1)$.

Under such a change of one-body wavefunctions the Slater determinant changes by:

$$\begin{aligned} \Delta_0(z) &\to \det g(z_i) z_i^{j-1} \\ &= \det \begin{pmatrix} g(z_1) & g(z_1)z_1 & g(z_1)z_1^2 & \cdots \\ g(z_1) & g(z_2)z_2 & g(z_2)z_2^2 & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix} \\ &= \prod g(z_i) \Delta_0(z) \\ &= \exp[i \sum j_n s_n(z)] \Delta_0(z) \end{aligned} \tag{23.70}$$

eq:coherentone

Note that there are two Fermi levels, and if N is large they are "far apart" meaning that operators such as $\sum_i z_i^n$ for small n will not mix states near the two respective levels. Let us therefore imagine taking N large and focussing attention on one of the Fermi levels. Therefore we adjust our energy level so that the groundstate wavefunctions are

[10] P. Jordan, Z. Phys. **93**(1935)464.

$1, z^{-1}, z^{-2}, \dots$ and we imagine that we have taken $N \to \infty$. Moreover, let us extend the $LU(1)$ action to $L\mathbb{C}^*$. Then we can separately consider the action of

$$g_-(z) = \exp[i \sum_{-\infty}^{0} \varphi_n z^n] \tag{23.71}$$

and

$$g_+(z) = \exp[i \sum_{1}^{\infty} \varphi_n z^n] \tag{23.72}$$

on the groundstate Ψ_0. Note if we think of the groundstate wavefunction as a Slater determinant then acting with $g_-(z)$ takes one column to a linear combination of lower columns and hence does not change the wavefunction. On the other hand, by (23.70) acting with $g_+(z)$ has a nontrivial action and generates all possible symmetric functions of the z_i.

In this - rather heuristic - sense, the action of the loop group "generates the whole Hilbert space of fermionic states." Moreover, by mapping antisymmetric wavefunctions to symmetric wavefunctions we now view the Hilbert space as the space of polynomials in the s_n. [11]

Now we observe the following. The ring of symmetric functions (extending scalars to $\mathbb{C}$) is the polynomial ring $\mathbb{C}[s_1, s_2, \dots]$. We now make this space into a *Hilbert space*. We introduce the inner product on polynomials in s_i by

$$\langle f(s)|g(s)\rangle := \int \prod_{k=1}^{\infty} \frac{ds_k \wedge d\bar{s}_k}{2\pi i k} \overline{f(s)} g(s) e^{-\sum_{k=1}^{\infty} \frac{1}{k}|s_k|^2} \tag{23.73}$$

eq:innerpf

This is the coherent state representation of an infinite system of harmonic oscillators:

$$[a_k, a_j] = k\delta_{k+j,0} \qquad\qquad -\infty < j, k < \infty \tag{23.74}$$

with $a_{-j} = a_j^\dagger$. These are realized as follows: $a_k^\dagger$ is multiplication by s_k and a_k is $k\frac{\partial}{\partial s_k}$. The state

$$(a_1^\dagger)^{\ell_1}(a_2^\dagger)^{\ell_2} \cdots (a_N^\dagger)^{\ell_N}|0\rangle \tag{23.75}$$

corresponds to the symmetric function $s_{(\ell)}$.

Now one can form the quantum field:

$$\phi := i \sum_{k \neq 0} \frac{1}{k} a_k z^{-k} \tag{23.76}$$

Then the coherent states

$$|\phi(z)\rangle = \exp i \oint \phi(z) j(z) |0\rangle \tag{23.77}$$

give an action of the loopgroup on the bosonic Fock vacuum acting as

[11] This statement is rather loose. See the Pressley-Segal book for a mathematically precise treatment. We are discussing an irreducible representation of a centrally extended group.

$$\exp[i\sum_{n=1}^{\infty} j_n a_n^\dagger]|0\rangle \tag{23.78}$$ eq:coherenttwo

thus producing a nontrivial isomorphism between a bosonic and fermionic Fock space.

To make the isomorphism between bosonic and fermionic Fock spaces more explicit we introduce a second-quantized formalism. B_{-n}^+ creates a fermionic state with wavefunction z^n, B_n annihilates it, so we introduce:

$$\begin{aligned}\Psi(\theta) &= \sum_{n\in\mathbb{Z}} B_n e^{in\theta} \\ \Psi^\dagger(\theta) &= \sum_{n\in\mathbb{Z}} B_{-n}^+ e^{-in\theta}.\end{aligned} \tag{23.79}$$ eq:fermflds

The filled Fermi sea satisfies the constraints:

$$\begin{aligned}B_n|0\rangle &= 0, \qquad |n| > n_F \\ B_{-n}^+|0\rangle &= 0, \qquad |n| \le n_F.\end{aligned} \tag{23.80}$$ eq:fifer

When we have decoupled systems it is appropriate to define two independent sets of Fermi fields:

$$\begin{aligned}\Psi(\theta) &= e^{i(n_F+\frac{1}{2})\theta} b(\theta) + e^{-i(n_F+\frac{1}{2})\theta}\bar{b}(\theta) \\ \Psi^\dagger(\theta) &= e^{-i(n_F+\frac{1}{2})\theta} c(\theta) + e^{i(n_F+\frac{1}{2})\theta}\bar{c}(\theta).\end{aligned} \tag{23.81}$$ eq:bcsys

We introduce complex coordinates $z = e^{i\theta}$, and define the mode expansions:

$$\begin{aligned}b(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} b_n z^n \\ c(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} c_n z^n \\ \bar{b}(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} \bar{b}_n \bar{z}^n \\ \bar{c}(z) &= \sum_{n\in\mathbb{Z}+\frac{1}{2}} \bar{c}_n \bar{z}^n\end{aligned} \tag{23.82}$$ eq:bcsysi

$b_n, c_n, \bar{b}_n, \bar{c}_n$ are only unambiguously defined for $\mid n \mid << N$. That is, we focus on operators that do not mix excitations around the two Fermi levels. The peculiar half-integral moding is chosen to agree with standard conventions in CFT. In terms of the original nonrelativistic modes we have:

$$\begin{aligned}c_n &= B_{-n_F-\epsilon+n}^+ \\ b_n &= B_{n_F+\epsilon+n} \\ \bar{c}_n &= B_{n_F+\epsilon-n}^+ \\ \bar{b}_n &= B_{-n_F-\epsilon-n}\end{aligned} \tag{23.83}$$ eq:mapmodes

where $\epsilon = \frac{1}{2}$, so that

$$\{b_n, c_m\} = \delta_{n+m,0} \qquad \{\bar{b}_n, \bar{c}_m\} = \delta_{n+m,0} \tag{23.84}$$

eq:anticomms

and all other anticommutators equal zero.

We may now reinterpret the fields $b, c, ...$ Defining $z = e^{i\theta+\tau}$ we see that these may be extended to fields in two-dimensions, and that they are (anti-) chiral, that is, they satisfy the two-dimensional Euclidean Dirac equation. Upon continuation to Minkowski space we have $z \to e^{i(\theta+t)}$ and $\bar{z} \to e^{i(\theta-t)}$. We are thus discussing two relativistic massless Fermi fields in $1+1$ dimensions. The b, c fields generate a Fermionic Fock space built on the product of vacua $\mid 0\rangle_{bc} \otimes \mid \bar{0}\rangle_{\bar{b}\bar{c}}$ where $b_n \mid 0\rangle_{bc} = c_n \mid 0\rangle_{bc} = 0$ for $n > 0$. Explicitly, the b, c fields by themselves generate the Fock space:

$$\mathcal{H}_{bc} = \text{Span}\left\{\prod b_{n_i} \prod c_{m_i} \mid 0\rangle_{bc}\right\} \qquad . \tag{23.85}$$

eq:cftstspc

The space $\mathcal{H}_{bc}$ has a natural grading according to the eigenvalue of $\sum_{n\in\mathbb{Z}} : b_n c_{-n} = \oint bc$ (called "bc-number"):

$$\mathcal{H}_{bc} = \oplus_{p\in\mathbb{Z}} \mathcal{H}_{bc}^{(p)} \tag{23.86}$$

eq:grdspce

and the states obtained by moving a state from the Fermi sea to an excited level correspond to the subspace of zero bc number:

$$\mathcal{H}_{\text{chiral}} := \mathcal{H}_{bc}^{(0)} \tag{23.87}$$

eq:chrlspc

NEED TO GIVE HERE THE TRANSLATION BETWEEN a state corresponding to a Young Diagram and an explicit fermionic Fock state.

The bc system defines a conformal field theory. Indeed we may introduce

$$L_n = \sum_{m=-\infty}^{\infty} (n/2 + m) c_{-m} b_{m+n} \tag{23.88}$$

eq:frmvrop

and compute that the L_n satisfy the Virasoro algebra

$$[L_n, L_m] = (n-m)L_{n+m} + \frac{c}{12} n(n^2-1)\delta_{n+m,0} \tag{23.89}$$

eq:virasoroalg

with $c = 1$.

Now let us consider the second quantized operators corresponding to z^n. Before taking the limit these are given by:

$$\Upsilon_n = \sum z_i^n \qquad \longleftrightarrow \qquad \int d\theta \Psi^\dagger(\theta)\ e^{in\theta} \Psi(\theta). \tag{23.90}$$

eq:upsop

Now let us take the large N limit and focus on terms that do not mix the two Hilbert spaces of excitations around the two Fermi levels. Cross terms between barred and unbarred fields involve operators that mix the two Fermi levels. Since we are only interested in the case of the decoupled Fermi level excitations we may replace:

$$\begin{aligned}\Upsilon_n &\to \oint dz\ z^{-1-n} c\ b(z) + \oint d\bar{z}\ \bar{z}^{-1+n} \bar{c}\ \bar{b}(\bar{z}) \\ &= \sum_m c_{n-m} b_m + \bar{c}_{m-n} \bar{b}_{-m} \\ &= \alpha_n + \bar{\alpha}_{-n}\end{aligned} \tag{23.91}$$ eq:upsii

where we have introduced a field $bc = i\partial_z \phi(z)$ which has expansion

$$\begin{aligned}\partial_z \phi(z) &= i \sum_{m\in\mathbb{Z}} \alpha_m z^{m-1} \\ [\alpha_m, \alpha_n] = [\bar{\alpha}_m, \bar{\alpha}_n] &= m\delta_{m+n,0} \\ [\alpha_m, \bar{\alpha}_n] &= 0.\end{aligned} \tag{23.92}$$ eq:bosalg

In terms of α_n, we may introduce a representation of the Virasoro algebra:

$$L_n = \frac{1}{2} \sum \alpha_{n-m} \alpha_m \tag{23.93}$$ eq:virop

which satisfy (23.89), again with $c = 1$. Using the α we can define a vacuum $\alpha_n \mid 0\rangle = 0$ for $n \geq 0$ and a statespace

$$\mathcal{H}_\alpha = \text{Span} \left\{ \mid \vec{k}\rangle \equiv \prod (\alpha_{-j})^{k_j} \mid 0\rangle \right\}. \tag{23.94}$$ eq:achalph

Bosonization states that there is a natural isomorphism:

$$\mathcal{H}_\alpha \cong \mathcal{H}_{bc}^{(0)} \tag{23.95}$$ eq:bsnztin

We will not prove this but it can be made very plausible as follows. The Hilbert space may be graded by L_0 eigenvalue. The first few levels are:

$$\begin{array}{lccc} L_0 = 1 & & \{b_{-\frac{1}{2}} c_{-\frac{1}{2}} \mid 0\rangle\} & \{\alpha_{-1} \mid 0\rangle\} \\ L_0 = 2 & \{b_{-\frac{1}{2}} c_{-\frac{3}{2}} \mid 0\rangle, b_{-\frac{3}{2}} c_{-\frac{1}{2}} \mid 0\rangle\} & & \{\alpha_{-2} \mid 0\rangle, (\alpha_{-1})^2 \mid 0\rangle\} \end{array} \tag{23.96}$$ eq:firstlev

At level $L_0 = n$, the fermion states are labeled by Young diagrams Y with n boxes. At level $L_0 = n$, the Bose basis elements are labeled by partitions of n. We will label a partition of n by a vector $\vec{k} = (k_1, k_2, \dots)$ which has almost all entries zero, such that $\sum_j j k_j = n$. Bosonization states that the two bases are linearly related:

$$\mid Y\rangle = \sum_{\vec{k}\in \text{Partitions}(n)} \langle \vec{k} \mid Y\rangle \mid \vec{k}\rangle \tag{23.97}$$ eq:linrel

This last relation can be understood as the relation (23.61).

****

1. Need to explain this some more.

2. Look up papers of the Kyoto school, Jimbo et. al. for perhaps helpful ways of presenting this material.

****

For much more about this see the references below.

1. Pressley and Segal, *Loop Groups*, Chapter 10.

2. M. Stone, "Schur functions, chiral bosons, and the quantum-Hall-effect edge states," Phys. Rev. **B42** 1990)8399 (I have followed this treatment.)

3. S. Cordes, G. Moore, and S. Ramgoolam, "Lectures on 2D Yang-Mills Theory, Equivariant Cohomology and Topological Field Theories," Nucl. Phys. B (Proc. Suppl 41) (1995) 184, section 4. Also available at hep-th/9411210.

4. M. Douglas, "Conformal field theory techniques for large N group theory," hep-th/9303159.

Next time – include

Verma modules – how the more standard treatment of SU(2) reps fits in.

basis

$$f_m(z) = \frac{z^{j+m}}{\sqrt{(j+m)!(j-m)!}} \leftrightarrow \quad \mid j,m > \tag{23.98}$$

for $-j \leq m \leq j$

is orthonormal wrt:

$$(f,g) = \frac{(2j+1)!}{\pi} \int d^2z \;\; \frac{1}{(1+\mid z \mid^2)^{2j+1}} \;\; f^*_{(z)} g(z) \tag{23.99}$$

Preprint typeset in JHEP style - HYPER VERSION

Lecture 8: Lie Algebras from Lie Groups

Gregory W. Moore

Abstract: Not updated since November 2009. March 27, 2018

-TOC- **Contents**

1. Introduction **2**

2. Geometrical approach to the Lie algebra associated to a Lie group **2**
2.1 Lie's approach 2
2.2 Left-invariant vector fields and the Lie algebra 4
2.2.1 Review of some definitions from differential geometry 4
2.2.2 The geometrical definition of a Lie algebra 5

3. The exponential map **8**

4. Baker-Campbell-Hausdorff formula **11**
4.1 Statement and derivation 11
4.2 Two Important Special Cases 17
4.2.1 The Heisenberg algebra 17
4.2.2 All orders in B, first order in A 18
4.3 Region of convergence 19

5. Abstract Lie Algebras **19**
5.1 Basic Definitions 19
5.2 Examples: Lie algebras of dimensions $1, 2, 3$ 23
5.3 Structure constants 25
5.4 Representations of Lie algebras and Ado's Theorem 26

6. Lie's theorem **28**

7. Lie Algebras for the Classical Groups **34**
7.1 A useful identity 35
7.2 $GL(n,k)$ and $SL(n,k)$ 35
7.3 $O(n,k)$ 38
7.4 More general orthogonal groups 38
7.4.1 Lie algebra of $SO^*(2n)$ 39
7.5 $U(n)$ 39
7.5.1 $U(p,q)$ 42
7.5.2 Lie algebra of $SU^*(2n)$ 42
7.6 $Sp(2n)$ 42

8. Central extensions of Lie algebras and Lie algebra cohomology **46**
8.1 Example: The Heisenberg Lie algebra and the Lie group associated to a symplectic vector space 47
8.2 Lie algebra cohomology 48

9. Left-invariant differential forms **50**

10. The Maurer-Cartan equation **51**

11. Examples **53**
11.1 $SU(2)$ 53
11.1.1 The Heisenberg group 55

12. Metrics on Lie groups **56**
12.1 Simple Lie groups and the index of a representation 58
12.2 Metrics on abelian groups 60
12.3 Geodesics 60
12.4 Compact and noncompact real forms: The signature of the CK metric 60
12.5 Example: $SL(2,R)$ 61

13. Some Remarks on Infinite Dimensional Lie Algebras **62**
13.1 Generalities 62
13.2 Groups of operators on Hilbert space 62
13.3 Gauge Groups 62
13.3.1 Loop algebras 62
13.4 Diffeomorphism Groups 63
13.4.1 Gravity in 1+1 dimensions 64
13.4.2 The exponential map for $Diff(S^1)$ 66

---

1. Introduction

Quite generally, when one is confronted with a nonlinear object or phenomenon it is often useful to reduce the problem to a linear problem, at the cost of restricting the domain of applicability.

Lie groups are not linear – they are curved manifolds. If one chooses coordinates then the group multiplication law is, in general, given by a complicated power series in the coordinates.

Nevertheless, Lie's theorem reduces many questions about Lie groups to questions about *Lie algebras*. Questions about curved manifolds turn out to be equivalent to questions about linear algebra. This is a profound simplification, and it leads to a very rich theory.

2. Geometrical approach to the Lie algebra associated to a Lie group

2.1 Lie's approach

A good way to approach the subject is the way Sophus Lie did himself. A Lie group is a group with continuous (or smooth) parameters. We convert the associativity of the group law into a differential equation, and study the integrability of that differential equation.

Suppose $x^1, \ldots, x^n$ are coordinates in a neighborhood $\mathcal{U}$ of 1_G, where $n = \dim G$. We take $x = 0$ to correspond to 1_G. Thus, we have a smooth parametrization of group elements $g(x)$. The product of two group elements near the identity will be another group element near the identity. Thus, there is a neighborhood $\mathcal{U}' \subset \mathcal{U}$ so that if $g(x), g(y) \in \mathcal{U}'$ then we can write the group law as:

$$g(x)g(y) = g(\phi(x,y)) \tag{2.1}$$ eq:grplaw

for some smooth functions $\phi^i(x,y)$, $i = 1, \ldots n$.

The group laws can be expressed in terms of ϕ:

1. Associativity:

$$\phi(x, \phi(y,z)) = \phi(\phi(x,y), z) \tag{2.2}$$ eq:associa

2. Identity: $\phi(0,x) = \phi(x,0) = x$

3. Inverse: $\phi(x, x_0) = \phi(x_0, x) = 0$ is solvable for a unique x in terms of x_0.

Now, let us differentiate the associativity condition (2.2) with respect to the z^k:

$$\frac{\partial \phi^i(x,\phi(y,z))}{\partial \phi^j(y,z)} \frac{\partial \phi^j(y,z)}{\partial z^k} = \frac{\partial \phi^i}{\partial z^k}(\phi(x,y), z) \tag{2.3}$$ eq:slieone

where repeated indices are summed.

Now define an $n \times n$ matrix function of one variable:

$$u^i{}_k(x) := \left. \frac{\partial \phi^i(x,y)}{\partial y^k} \right|_{y=0} \tag{2.4}$$ eq:defnsu

Then, setting $z = 0$ in (2.3) gives

$$\frac{\partial \phi^i(x,y)}{\partial y^j} u^j{}_k(y) = u^i{}_k(\phi(x,y)) \tag{2.5}$$ eq:lieseq

Note that (2.5) is equivalent to the equality of first order differential operators:

$$u^j{}_k(y) \frac{\partial}{\partial y^j} = u^i{}_k(\phi(x,y)) \frac{\partial}{\partial \phi^i} \tag{2.6}$$ eq:livfone

Here we are holding x fixed and viewing $y^i \to \phi^i(x,y)$ as a nonlinear change of coordinates. Nevertheless the equation is true for any x and hence we can write (2.6) as

$$u^j{}_k(y) \frac{\partial}{\partial y^j} = u^i{}_k(\phi) \frac{\partial}{\partial \phi^i} \tag{2.7}$$ eq:livfonep

Now, (2.7) is a remarkable equation because the LHS depends only on y and the RHS depends only on ϕ. We will apply this observation in one moment.

A second observation is that if we denote

$$I_k(y) := u^j{}_k(y) \frac{\partial}{\partial y^j} \tag{2.8}$$ eq:invtvf

then we can compute

$$
\begin{aligned}
[I_k(y), I_m(y)] &= \left(u^j{}_k \partial_j u^l{}_m - u^j{}_m \partial_j u^l{}_k\right) \frac{\partial}{\partial y^l} \\
&= \left(u^j{}_k \partial_j u^l{}_m - u^j{}_m \partial_j u^l{}_k\right) (u^{tr,-1})_l{}^p I_p(y) \\
&:= f^p_{km}(y) I_p(y)
\end{aligned}
\tag{2.9}
$$

eq:commis

Now, by (2.7) $I_k(y) = I_k(\phi)$ so it follows that $f^p_{mk}(y) = f^p_{mk}(\phi)$, and since $\phi = \phi(x, y)$ with arbitrary x, it follows that $f^p_{mk}(y)$ is *constant*! Let us consider the finite-dimensional vector space spanned by the first order differential operators) $I_k(y)$ in (2.8). That is, we consider the linear combinations of the $I_k(y)$ with constant (real) coefficients. Note that this vector space is *closed under commutator*. We denote this vector space $L(G)$ or sometimes $\mathfrak{g}$. Note that

$$
\mathrm{dim} G = \mathrm{dim} L(G) \tag{2.10}
$$

On the LHS we have the dimension of a manifold, and on the RHS the dimension of a vector space. Moreover, it follows from general properties of differential operators that if X_1, X_2, X_3 are any three vector fields in $L(G)$ then

$$
[[X_1, X_2], X_3] + [[X_3, X_1], X_2] + [[X_2, X_3], X_1] = 0 \tag{2.11}
$$

eq:jacobi

This crucial identity is known as the *Jacobi identity*.

These are the crucial properties which are abstracted into the general definition:

Definition : An abstract Lie algebra $\underline{\mathbf{g}}$ over a field k is a vector space over k together with a product $(v_1, v_2) \to [v_1, v_2] \in \mathfrak{g}$, such that for all $v_1, v_2, v_3, \in \mathfrak{g}$ $\alpha, \beta \in k$.

1.) $[v_1, v_2] = -[v_2, v_1]$
2.) $[\alpha v_1 + \beta v_2, v_3] = \alpha[v_1, v_3] + \beta[v_2, v_3]$
3.) $[[v_1, v_2], v_3] + [[v_3, v_1], v_2] + [[v_2, v_3], v_1] = 0.$

2.2 Left-invariant vector fields and the Lie algebra

We will now rephrase Lie's argument in the language of modern differential geometry.

2.2.1 Review of some definitions from differential geometry

Tangent vectors are directional derivatives along paths. If we imagine $M \subset \mathbb{R}^N$ then we literally take a tangent plane. In general if $p \in M$ let $C^1(p)$ be the functions defined in some neighborhood of $p \in M$, which are differentiable at p. A directional derivative along a curve $\gamma(t)$ such that $\gamma(0) = p$ is a linear functional on this space defined by

$$
\dot{\gamma}(f) := \frac{d}{dt}|_{t=0} f(\gamma(t)) \tag{2.12}
$$

Then *the space of tangent vectors at* p, T_pM, is the linear span of these linear functionals $C^1(p) \to \mathbb{R}$. A *vector field* is a continuous system of tangent vectors.

Now suppose that $\phi : M_1 \to M_2$ is a map between manifolds. Let us study how geometric objects behave with respect to such maps.

First of all, functions are contravariant. That is we define the *pullback* on functions

$$\phi^* : Fun(M_2) \to Fun(M_1) \tag{2.13}$$

by $\phi^*(f) = f \circ \phi$. Here $Fun(M)$ is the space of all (say, continuous) functions from M to (say) the real numbers.

By contrast, vectorfields *push forward*:

$$\phi_* : T_p M_1 \to T_{\phi(p)} M_2 \tag{2.14}$$

the definition is that $\phi_*(\dot{\gamma})$ is the directional derivative along the curve $\phi(\gamma(t))$.

Two facts we will need below are:

• The commutator of two first-order differential operators is a first order differential operator. This defines the commutator of two vector fields:

$$[\xi_1, \xi_2] \tag{2.15}$$

Thus, $Vect(M)$ is a Lie algebra, for any manifold M. It is infinite-dimensional. It is not difficult to show that

$$\phi_*[\xi_1, \xi_2] = [\phi_*(\xi_1), \phi_*(\xi_2)] \tag{2.16}$$

•

Vector fields act on functions to produce new functions. From the above definitions it follows that:

$$\phi_*(V)(f) = V(\phi^*(f)) \tag{2.17}$$

Exercise

In general, functions do *not* push forward. They only pull back. However, if ϕ is an invertible map between manifolds show that it makes sense to define $\phi_*(f) = f \circ \phi^{-1}$. This is used when pushing forward the product of a function and a vector field, fV in the next section.

2.2.2 The geometrical definition of a Lie algebra

Definition. A *left translation*, or *right translation* by an element $g \in G$ is the diffeomorphism:

$$\begin{aligned} L_g &: G \to G \qquad & h &\mapsto g \cdot h \\ R_g &: G \to G \qquad & h &\mapsto h \cdot g^{-1} \end{aligned} \tag{2.18}$$

eq:leftrighttr

Exercise
Show that L_g, R_g define injections of groups $G \hookrightarrow Diff(G)$.

Now we introduce the important idea of left and right invariance: A left-invariant function satisfies:

$$L_g^*(f) = f \tag{2.19}$$ eq:lftinvfun

for all $g \in G$. Of course, this just means that f is constant, because if we evaluate (2.19) at $g = 1$ then

$$L_g^*(f)|_{h=1} = f|_{h=1} \Rightarrow f(g \cdot 1) = f(1) \tag{2.20}$$ eq:lftinvfunp

However, the notion of left- or right- invariant tensors still leaves room for very interesting examples.

Definition. A vector field $\xi \in Vect(G)$ is left (or right) invariant if

$$\begin{aligned}(L_{g_0})_*(\xi_g) &= \xi_{g_0 \cdot g} \\ (R_{g_0})_*(\xi_g) &= \xi_{g g_0^{-1}}\end{aligned} \tag{2.21}$$ eq:livr

is satisfied for all $g, g_0 \in G$, respectively.

The picture is: For a general vector field ξ_g is the directional derivative at g to some curve $\gamma_g(t)$ going through g, while $\xi_{g_0 g}$ is the directional derivative of some -*a priori* unrelated - curve $\tilde{\gamma}_{g_0 g}(t)$ going through $g_0 g$. Then the condition of left-invariance:

$$(L_g)_*(\xi_g) = \xi_{g_0 g} \tag{2.22}$$

means $\tilde{\gamma}_{g_0 g}(t)$ has the same directional derivative at $t = 0$ as the curve $g_0 \cdot \gamma_g(t)$.

We now recognize our first order operators $I_k(y)$ defined in (2.8) as vector fields on the group G. The equation (2.7) is the statement that these are left-invariant vector fields.

Now, let us return to the general situation. Note that if ξ_1, ξ_2 are two LIVF's then

$$(L_g)_*[\xi_1, \xi_2] = [(L_g)_*(\xi_1), (L_g)_*(\xi_2)] = [\xi_1, \xi_2] \tag{2.23}$$

This leads to the

Geometric definition of the Lie algebra: We define the Lie algebra $L(G)$ of a Lie group G to be the Lie algebra of left invariant vector fields on G.

Examples

• $G = U(1)$. The general vector field on $U(1)$ is

$$\xi = f(\theta) \frac{d}{d\theta} \tag{2.24}$$

Left invariant (and right-invariant) vector fields satisfy $f(\theta+\theta_0)=f(\theta)$ for all θ_0. That is, $f(\theta)$ must be constant.

• $G=GL(n,k)$, $k=\mathbb{R},\mathbb{C}$. We can choose as global coordinates on the manifold the matrix elements g_{ij} of $g\in G$. Introduce an $n\times n$ matrix of vector fields $\frac{\partial}{\partial g}$ whose components are:

$$(\frac{\partial}{\partial g})_{ij}:=\frac{\partial}{\partial g_{ij}} \tag{2.25}$$

$(\frac{\partial}{\partial g_{ij}})_{g_0}$ is the directional derivative along the curve g_0+te_{ij} where e_{ij} is a matrix unit. Note that if g_0 is invertible then this curve is indeed in $GL(n,k)$ for sufficiently small $|t|$. One computes

$$(L_{g_0})_*(\frac{\partial}{\partial g})=g_0^{tr}\cdot\frac{\partial}{\partial g} \tag{2.26}$$

where matrix multiplication on the RHS is understood. To do this, let $f_{kl}(g)=g_{kl}$ be the function which picks out a matrix element. Then

$$(L_{g_0})_*(\frac{\partial}{\partial g_{ij}})(f_{kl})=\frac{\partial}{\partial g_{ij}}f_{kl}\circ L_{g_0}=\frac{\partial}{\partial g_{ij}}\sum_s(g_0)_{ks}f_{sl}=(g_0)_{ki}\delta_{j,l} \tag{2.27}$$

Therefore,

$$(L_{g_0})_*(\frac{\partial}{\partial g_{ij}}|_g)=\sum_k(g_0)_{ki}\frac{\partial}{\partial g_{kj}}|_{g_0g} \tag{2.28}$$

Therefore

$$\xi_{ij}:=(g^{tr}\frac{\partial}{\partial g})_{ij}=\sum_k g_{ki}(\frac{\partial}{\partial g})_{kj} \tag{2.29}$$

is a matrix of left-invariant vector fields on $GL(n)$. Proof:

$$\begin{aligned}(L_{g_0})_*(g^{tr}\frac{\partial}{\partial g})&=(g_0^{-1}g)^{tr}g_0^{tr}\cdot\frac{\partial}{\partial g}\\&=g^{tr}\cdot\frac{\partial}{\partial g}\end{aligned} \tag{2.30}$$

The n^2 vector fields ξ_{ij} are linearly independent (simply consider their values at $g_0=1$) and form a basis for all LIVF's on $GL(n)$, by an argument given below. Finally, note that by direct computation we find

$$[\xi_{ij},\xi_{kl}]=\delta_{jk}\xi_{il}-\delta_{li}\xi_{kj} \tag{2.31}$$ eq:glsc

and the structure constants are indeed constant.

Returning to the general case, let us now suppose ξ is a LIVF. Then $\xi_g=(L_g)_*(\xi_1)$, so a LIVF is completely determined by its value at $g=1$, and hence we can identify $L(G)$ with the tangent space at $g=1$: $L(G)\cong T_1(G)$, at least, as a vector space. Conversely, given a tangent vector $X\in T_1(G)$. Then to $X\in T_1(G)$ we can associate a curve $\gamma_X(t)$ through $g=1$ whose tangent vector at $g=1$ is X. (As we will see below for a matrix group we can take $\gamma_X(t)$ to be the curve $\exp[tX]$ where we literally exponentiate the matrix.) If

G is a Lie group, then the *right* action of G on itself defines a global system of *left-invariant* vector fields on G as follows: If $X \in T_1(G)$ is the directional derivative along a curve $\gamma_X(t)$ passing through $g = 1$ then then the curves $\gamma_{X,g}(t) := g\gamma_X(t)$ through g have directional derivatives defining a vector field $\xi(X) \in T_gG$. This vector field is left-invariant, and in this way we define the map from T_1G to the left-invariant vector fields on G.

The left-invariant vector fields $\xi(X)$ are called the *fundamental vector fields* on G. Every LIVF of G is of the form $\xi(X)$ for some unique X (indeed $X = \xi_1$) so we can define $[X, Y] \in T_1G$ by

$$[\xi(X), \xi(Y)] = \xi([X, Y]) \tag{2.32}$$ eq:LieTone

This defines T_1G as a Lie algebra, and then ξ defines a homomorphism of Lie algebras $T_1(G) \to Vect(G)$.

To get a better feel for these vector fields consider again the example of $G = GL(n, \mathbb{F})$. The Lie algebra is $\mathfrak{g} = Mat_n(\mathbb{F})$. Let e_{ij} be the matrix unit with a 1 in the ith row and j^{th} column, and zero in all other matrix elements. Then one easily computes the vector field by considering the curve $g\exp[te_{ij}]$:

$$\xi(e_{ij}) = \sum_{k=1}^{n} g_{ki} \frac{\partial}{\partial g_{kj}} \tag{2.33}$$ eq:vertcli

where we are regarding the matrix elements g_{ij} as coordinates on the group. Thus the fundamental vector fields on G are just the vector fields ξ_{ij} we examined above.

Then, combining (2.31) and (2.32) we have

$$\xi\left([e_{ij}, e_{i'j'}]\right) = [\xi(e_{ij}), \xi(e_{i'j'})] = \delta_{ji'}\xi(e_{ij'}) - \delta_{ij'}\xi(e_{i'j}) \tag{2.34}$$

and we conclude that $[e_{ij}, e_{i'j'}]$ coincides with matrix commutator, as it should.

If $G \subset GL(n, \mathbb{F})$ is a subgroup of a matrix group then the matrix elements g_{ij} are not all independent, so some vectors $\frac{\partial}{\partial g_{kj}}$ will be expressed in terms of others. Put differently, a path such at $1 + te_{ij}$ will not be a path within the matrix subgroup in general. These linear dependences are most easily computed using the dual cotangent space T_1^*G and the Maurer-Cartan form, as discussed below.

Exercise
Write out the right-invariant vector fields on $GL(n, \mathbb{F})$.

---

3. The exponential map

We have discussed in chapter 2 how to exponentiate matrices. We would now like to generalize this notion to define a map

$$\exp : L(G) \to G \tag{3.1}$$

Since not all Lie groups are matrix groups it will be a little more abstract.

Consider a homomorphism $f : \mathbb{R} \to GL(n, \mathbb{R})$. Such a map must satisfy

$$\begin{aligned} \frac{d}{dt} f(t) &= \lim_{h \to 0} h^{-1}[f(t+h) - f(t)] \\ &= \lim_{h \to 0} h^{-1}[f(h) - 1] f(t) \\ &= A f(t) \end{aligned} \tag{3.2}$$

so $f(t)$ is the unique solution to this differential equation with $f(0) = 1$. We write

$$f(t) = e^{tA} \tag{3.3}$$

Recall that for $A \in M_n(k) \quad k = \mathbb{R}$ or $\mathbb{C}$, we defined

$$\exp A := \sum_{j \geq 0} \frac{A^j}{j!} \in M_n(k) \tag{3.4}$$

defines the *exponential of a matrix*. In general any function defined by a power series can be evaluated for matrix arguments, as above.

In a sufficiently small neighborhood $\mathcal{U}$ of 0 the map $\exp : M_n(\mathbb{R}) \to GL(n, \mathbb{R})$ is invertible with

$$\log g = -\sum_{k=1}^{\infty} (1-g)^k / k \tag{3.5}$$

as its unique inverse.

As we have said, not all Lie groups are matrix groups. So we define exp more generally (and more intrinisically) as follows:

Theorem: If G is a Lie group there is a 1-1 correspondence between the tangent space $T_1 G$ and group homomorphisms $f : \mathbb{R} \to G$ (aka "1-parameter subgroups").

Proof: A homomorphism clearly determines a tangent vector $X \in T_1 G$. Conversely, given a tangent vector $X \in T_1 G$ we consider the left-invariant vector field $\xi(X)$. We study the tangent curves to $\xi(X)$: these are curves $\gamma_X(t)$ such that

$$\dot{\gamma}_X(t) = \xi(X)|_{\gamma_X(t)}. \tag{3.6}$$

eq:tangentcurv

We would like to show that the tangent curve through $g = 1$ defines a group homomorphism of $\mathbb{R}$ into G.

We can take the tangent curve to satisfy $\gamma_X(0) = 1_G$. Equation (3.6) is just an ODE. By the theory of ODE's we know that we can find $\gamma_X(t)$ on a sufficiently small neighborhood $t \in (-\epsilon, \epsilon)$. Now consider a t_0 in this interval. Similarly, we can study this differential equation near $\gamma_X(t_0)$. Now $\gamma_X(t+t_0)$ and $\gamma_X(t_0)\gamma_X(t)$ are both solutions, since $\xi(X)$ is left-invariant. By the uniqueness of solutions to ODE's

$$\gamma_X(t + t_0) = \gamma_X(t_0)\gamma_X(t) \tag{3.7}$$

for $t, t_0, t+t_0$ in the interval, so $\gamma_X(t)$ is a local homomorphism. To extend the definition of $\gamma_X(t)$ to large values of t we take

$$\gamma_X(t) := (\gamma_X(t/n))^n \tag{3.8}$$

eq:largetee

for sufficiently large n. (Exercise: Show that it doesn't matter what n you pick.) ♠

Evaluation of $\gamma_X(t)$ at $t = 1$ defines the map

$$\exp : T_1 G \to G \tag{3.9}$$

whose derivative at 1_G is the identity. In this way we define exp, even if G is not a matrix group.

Remarks:

• $\exp : \mathbb{R} \to U(1)$ given by $x \to e^{2\pi i x}$ shows that exp is in general not 1-1.

•

A Lie group can be given the structure of a Riemannian manifold in a canonical way. The geodesics through 1_G are precisely the 1-parameter subgroups. For a compact Lie group, we get a complete Riemannian manifold, and a general theorem of differential geometry shows that any two points on a complete Riemannian manifold can be joined by a geodesic. Therefore, for a compact Lie group, exp is onto.

• For a noncompact Lie group exp need not be onto. We will give counterexamples when we survey the Lie algebras of matrix groups below.

• As we will see, finite dimensional Lie algebras can be exponentiated to form Lie groups. However, there exist examples of Lie algebras which cannot be "exponentiated" to form Lie groups.

Exercise

Show that the choice of n does not matter in (3.8).

Exercise *Exponentials and inverses*

a.) Show that for the abstract exp map we have:

$$e^A e^{-A} = 1 \tag{3.10}$$

for any $A \in \mathfrak{g}$.

b.) Define $Ad(G)$ on $\mathfrak{g}$ ..... *** NEED SOME DETAILS *** Show that

$$\exp[BAB^{-1}] = B\exp(A)B^{-1} \tag{3.11}$$

---

4. Baker-Campbell-Hausdorff formula

sec:sBCH

The exponential map is neither one-one nor onto in general. However, for a finite dimensional Lie group, in a sufficiently small neighborhood of the identity $1_G \in \mathcal{U}_2 \subset G$ and $0 \in \mathcal{U}_1 \subset T_1G$, $\exp : \mathcal{U}_1 \to \mathcal{U}_2$ is 1-1 and onto.

It is therefore natural to ask what the full group law looks like in terms of the linear space $\mathfrak{g} := T_1G$. Thus we would like to express

$$C(A, B) = \log e^A e^B \tag{4.1}$$

in terms of A, B.

If we imagine A, B are sufficiently small (we will say they are "order one") then we can expand in Taylor series

$$C(A, B) = A + B + \frac{1}{2}b(A, B) + \cdots \tag{4.2}$$

where $b(A, B) \in \mathfrak{g}$ are the terms of order 2 and $\cdots$ are the terms of order ≥ 3. The factor of $\frac{1}{2}$ is for later convenience. [1]

Since $C(A, 0) = A$ and $C(0, B) = B$, $b(A, B)$ has no terms of order A^2 or B^2 and hence is a bilinear map:

$$b : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{4.3}$$

Moreover, since $\exp(-A) = (\exp(A))^{-1}$ we have $C(-B, -A) = -C(A, B)$ and therefore $b(A, B) = -b(B, A)$, that is, b is skew-symmetric. Thus, working infinitesimally, we derive the Lie product on T_1G.

Now, the amazing thing about Lie groups is that

the entire Taylor series $C(A, B)$ can be expressed solely in terms of b.

The formula that does this is known as the BCH formula, and we will now derive it for the case of matrix groups in the next section.

4.1 Statement and derivation

The BCH formula allows us to understand the Lie group operation in terms of the matrices in the exponential. It also provides a convenient way to understand the relationship between Lie groups and Lie algebras. [2]

The problem we want to solve in this section is:

Let A, B be $n \times n$ matrices. Find a matrix C such that $e^C = e^A e^B$ where C is expressed as a (possibly infinite) series in A, B

[1] To be more precise, we can introduce parameters t_1, t_2 which are real and small and consider the multiplication of $\exp[t_1 A]\exp[t_2 B]$. The "degree" is the total degree of t_1 and t_2 in the Taylor expansion.

[2] The proof in this section follows Miller, *Symmetry Groups and their applications.*

The series we will obtain is convergent for "small" matrices. We can make this precise by, for example defining $\| A \| = \max|a_{ij}|$ and demanding that $\| A \|$, $\| B \|$ be sufficiently small. We comment on the radius of convergence below.

Lemma 1: Let A be a constant matrix in t. Then

$$\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A \tag{4.4}$$

Proof: Series expansion. ♠

To state the next lemma we introduce some notation:

Definition 4.1: For $A \in M_n(k)$ we denote by $\mathrm{Ad}(A)$ the linear transformation $M_n(k) \to M_n(k)$ defined by

$$\mathrm{Ad}(A) : B \mapsto [A, B] \tag{4.5}$$

We also denote:

$$(\mathrm{Ad}(A))^m B = \overbrace{[A, [A, \cdots [A, B] \cdots]}^{m \text{ times}} \tag{4.6}$$

where there are m commutators on the RHS.

Lemma 2:

$$\begin{aligned} e^A B e^{-A} &= e^{\mathrm{Ad}A}(B) = \Sigma \frac{(\mathrm{Ad}A)^j}{j!} B \\ &= B + [A, B] + \frac{1}{2}[A, [A, B]] + \cdots \end{aligned} \tag{4.7}$$

Proof: Let $B(t) = e^{tA} B e^{-tA}$.

Note $B(0) = B$

$\dot{B}(t) = Ae^{tA} B e^{-tA} - e^{tA} B e^{-tA} A = (\mathrm{Ad}A)(B(t))$

Now, by induction

$$(\frac{d}{dt})^j B(t) = (\mathrm{Ad}A)^j B(t) \tag{4.8}$$

$B(t)$ is analytic in t, so we can write the Taylor series around zero. ♠

Now we come to the much more nontrivial: **Lemma 3**: Let

$$f(z) = \frac{e^z - 1}{z} = 1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots \tag{4.9}$$

Then

$$-(\frac{d}{dt}e^{A(t)})e^{-A(t)} = e^{A(t)}\frac{d}{dt}e^{-A(t)} = -f(\mathrm{Ad}(A(t))) \cdot \dot{A}(t) \tag{4.10}$$

where $A(t)$ is any differentiable matrix function of t.

Proof:
This is nontrivial because $\dot{A}(t)$ does not commute with $A(t)$ in general!

$$\begin{aligned} B(s,t) &\equiv e^{sA(t)}\frac{d}{dt}e^{-sA(t)} \\ \frac{\partial B}{\partial s} &= A(t)e^{sA(t)}\frac{d}{dt}e^{-sA(t)} - e^{sA(t)}\frac{d}{dt}\left[e^{-sA(t)}A(t)\right] \\ &= \mathrm{Ad}(A(t))B(s,t) - \dot{A}(t) \\ \frac{\partial^j B}{\partial s_j} &= (\mathrm{Ad}(A(t)))^j B(s,t) - (\mathrm{Ad}A(t))^{j-1}\dot{A}(t) \end{aligned} \tag{4.11}$$

$B(0,t) = 0$ therefore again by Taylor:

$$\frac{1}{j!}\frac{\partial^j}{\partial s^j}B(s,t)\mid_{s=0} = -\mathrm{Ad}(A(t))^{j-1}\dot{A}(t) \qquad j \geq 1 \tag{4.12}$$

So

$$e^{sA(t)}\frac{d}{dt}(e^{-sA(t)}) = -\sum_{j=1}^{\infty}\frac{s^j(\mathrm{Ad}(A(t)))^{j-1}}{j!}\dot{A}(t) \tag{4.13}$$

Now set s=1. ♠
Note: you can rewrite this lemma as the statement:

$$\frac{d}{dt}e^{-A(t)} = -\int_0^1 e^{-(1-s)A(t)}\dot{A}(t)e^{-sA(t)}ds \tag{4.14}$$

eq:intfrm

because:

$$\begin{aligned} \frac{d}{dt}e^{-A(t)} &= -\int_0^1 e^{-(1-s)A(t)}\dot{A}(t)e^{-sA(t)}ds \\ &= -e^{-A(t)}\int_0^1 e^{sAd(A(t))}ds\dot{A}(t) \\ &= -e^{-A(t)}\left(\frac{e^{Ad(A(t))}-1}{Ad(A(t))}\right)\dot{A}(t) \end{aligned} \tag{4.15}$$

eq:intfrmi

Note that (4.14) is an intuitively appealing formula. For a finite product we have:

$$\begin{aligned} \frac{d}{dt}\Big[M_1(t)M_2(t)M_3(t)\cdots M_n(t)\Big] &= (\frac{d}{dt}M_1(t))M_2(t)M_3(t)\cdots \\ &+ (M_1(t))(\frac{d}{dt}M_2(t))M_3(t)\cdots \\ &+ (M_1(t))(M_2(t))(\frac{d}{dt}M_3(t))\cdots \\ &+ \cdots + M_1(t)M_2(t)\cdots(\frac{d}{dt}M_n(t)) \end{aligned} \tag{4.16}$$

Now regard

$$e^{A(t)} = \prod_{i=1}^{N} [e^{A(t)\Delta s}] \tag{4.17}$$

where $\Delta s = 1/N$. Now take $N \to \infty$.

Now we are finally ready to state the main theorem:

Theorem: (Baker-Campbell-Hausdorff formula)
Let:

$$g(z) = \frac{\log z}{z-1} = \sum_{j=0}^{\infty} \frac{(1-z)^j}{j+1} \tag{4.18}$$

be a power series about 1. Then when A, B are $n \times n$ matrices with $\| A \|$, $\| B \|$ sufficiently small, the matrix C given by the expansion:

$$\boxed{C = B + \int_0^1 g(e^{t\mathrm{Ad}A} e^{\mathrm{Ad}B})(A) dt} \tag{4.19}$$

satisfies $C = \log(e^A e^B)$.

Explicitly the first few terms are: [3]

$$\boxed{C = A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \frac{1}{12}[B,[B,A]] + \frac{1}{24}[A,[B,[A,B]]] + \cdots} \tag{4.20}$$

where the next terms are order ϵ^5 if we scale A, B by ϵ.[4]

Proof:

Introduce:

$$e^{C(t)} = e^{tA} e^{B} \tag{4.21}$$

eq:defcee

$C(0) = B$, and $C(1)$ is the matrix we want. We derive a differential equation for $C(t)$. By Lemma 4 we have:

$$e^{C(t)} \frac{d}{dt} e^{-C(t)} = -f(\mathrm{Ad}C(t)) \dot{C}(t) \tag{4.22}$$

with

$$f(z) = \frac{e^z - 1}{z} \tag{4.23}$$

On the other hand, plugging in the definition (4.21),

$$e^{C(t)} \frac{d}{dt} e^{-C(t)} = e^{tA} \frac{d}{dt} e^{-tA} = -A \tag{4.24}$$

[3] It is useful to note that $[A,[B,[A,B]]] = -[B,[A,[B,A]]] = B^2A^2 - A^2B^2$

[4] One can find an algorithm for generating the higher order terms in Varadarajan's book on group theory.

by using Lemma 2. Therefore:

$$f(\mathrm{Ad}C(t))\dot{C}(t) = A \tag{4.25}$$

Now, f is a power series about 1 so it immediately follows that

$$\dot{C}(t) = f(Ad(C(t)))^{-1}A \tag{4.26}$$ eq:invtser

Let us make this more explicit.
Introduce the power series about 1

$$g(w) = \frac{\log w}{w-1} = \sum_{j=0}^{\infty} \frac{(1-w)^j}{j+1} \tag{4.27}$$

which satisfies the equation:

$$f(z)g(e^z) = \frac{e^z - 1}{z} \cdot \frac{z}{e^z - 1} = 1 \tag{4.28}$$

regarded as an identity of power series in z. Now we can substitute for z any operator $\mathcal{O}$, and use

$$g(e^{\mathcal{O}}) = f(\mathcal{O})^{-1}, \tag{4.29}$$ eq:invrl

and therefore we can solve for $\dot{C}$:

$$\begin{aligned} \dot{C}(t) &= f(\mathrm{Ad}(C(t)))^{-1} \cdot A \\ &= g(\exp(\mathrm{Ad}(C(t))) \cdot A \end{aligned} \tag{4.30}$$ eq:first

where we applied (4.29) with $\mathcal{O} = (Ad(C(t))$. This hardly seems useful, since we still don't know $C(t)$, but now since we have power series we can say

$$e^{\mathcal{O}} = e^{Ad(C(t))} = e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)} \tag{4.31}$$ eq:formeff

Proof of claim:
Note that for all H:

$$\begin{aligned} e^{\mathrm{Ad}C(t)}H &= e^{C(t)}He^{-C(t)} && \text{by lemma 2} \\ &= e^{tA}e^{B}He^{-B}e^{-tA} && \text{def.ofC} \\ &= e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)}H && \text{by lemma 2} \\ \Rightarrow e^{\mathrm{Ad}(C(t))} &= e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)} \end{aligned} \tag{4.32}$$

Therefore:

$$\dot{C}(t) = g(e^{\mathrm{Ad}(tA)}e^{\mathrm{Ad}(B)}) \cdot A \tag{4.33}$$ eq:dffq

Now we integrate equation (4.33)

$$C(t) = C(0) + \int_0^t g(e^{\mathrm{Ad}(sA)} e^{\mathrm{Ad}(B)}) A ds \tag{4.34}$$

but $C(0) = B$, so

$$C = C(1) = \log(e^A e^B) = B + \int_0^1 g(e^{\mathrm{Ad}(s\ A)} e^{\mathrm{Ad}(B)}) A\ ds \tag{4.35}$$

which is what we wanted to show. ♠.

Remarks For suitable operators A, B on Hilbert space the BCH formula continues to hold.

Exercise
Use the BCH theorem to show that if

$$g_1 = e^{t_1 A_1}, \quad g_2 = e^{t_2 A_2} \tag{4.36}$$

the *group commutator*, $g_1 g_2 g_1^{-1} g_2^{-1}$ corresponds to the Lie algebra commutator:

$$g_1 \cdot g_2 \cdot g_1^{-1} \cdot g_2^{-1} = 1 + t_1 t_2 [A_1, A_2] + \mathcal{O}(t_1^2, t_2^2) \tag{4.37}$$

Thus we say a Lie algebra is "abelian" if $[A_1, A_2] = 0$ for all $A_1, A_2 \in \mathfrak{g}$. Otherwise it is "nonabelian."

Moreover, if $A_i = \frac{d}{dt}|_0 g_i(t)$ then $[A_1, A_2]$ is the Lie algebra element associated to the curve

$$g_{12}(t) = g_1(\sqrt{t}) \cdot g_2(\sqrt{t}) \cdot g_1^{-1}(\sqrt{t}) \cdot g_2^{-1}(\sqrt{t}) \tag{4.38}$$

Exercise
Work out the BCH series to order 5 in A, B.

Exercise
Show that we can also write:

$$C = \log(e^A e^B) = A + \int_0^1 g(e^{-\mathrm{Ad}(s\ B)} e^{-\mathrm{Ad}(A)}) B\ ds \tag{4.39}$$

4.2 Two Important Special Cases

4.2.1 The Heisenberg algebra

Suppose $[A, B] = z \cdot 1$ where z is a *scalar*. Then all higher commutators vanish and

$$e^A e^B = e^{A+B+\frac{1}{2}z\cdot 1} \tag{4.40}$$

hence:

$$e^A e^B = e^z \cdot e^B \cdot e^A \tag{4.41}$$

One very common application of this formua is in *quantum mechanics*. For example, if we consider the quantum mechanics of a single particle on a line then A, B can be scalar multiples of $\hat{q}, \hat{p}$, respectively where

$$[\hat{p}, \hat{q}] = -i\hbar 1 \tag{4.42}$$

The above formula says:

$$e^{\alpha\hat{p}} e^{\beta\hat{q}} = e^{-i\hbar\alpha\beta} e^{\beta\hat{q}} e^{\alpha\hat{p}} \tag{4.43}$$ eq:heisgroup

More generally,

$$e^{\alpha_1\hat{p}+\beta_1\hat{q}} e^{\alpha_2\hat{p}+\beta_2\hat{q}} = e^{-\frac{i}{2}(\alpha_1\beta_2-\alpha_2\beta_1)} e^{(\alpha_1+\alpha_2)\hat{p}+(\beta_1+\beta_2)\hat{q}} \tag{4.44}$$ eq:heisgroupi

---

Exercise *The magnetic translation group*

Consider the problem of an electron confined to a two-dimensional plane x_1, x_2 Usually in quantum mechanics translations in x^1, x^2 by a^1, a^2 are represented by

$$U_1 = e^{ia^1 p_1} \qquad\qquad U_2 = e^{ia^2 p_2} \tag{4.45}$$ eq:transl

with $U_1 U_2 = U_2 U_1$, because $[p_1, p_2] = 0$.

Now suppose that there is a constant magnetic field B perpendicular to the x_1, x_2 space, as in the quantum Hall effect.

The Hamiltonian for a free electron in the presence of the magnetic field is:

$$H = \frac{1}{2m}\left((p_1 - eA_1)^2 + (p_2 - eA_2)^2\right) \tag{4.46}$$ eq:hamil

In a uniform magnetic field we can choose a gauge so that this can be put in the form:

$$H = \frac{1}{2m}\left((p_1 - \frac{eBx_2}{2})^2 + (p_2 + \frac{eBx_1}{2})^2\right) \tag{4.47}$$ eq:hamila

Now p_1, p_2 *do not commute with* H. This is hardly surprising since H is no longer translation invariant. Moreover the gauge invariant momenta $\tilde{p}_i := p_i - eA_i$ also do not commute with the Hamiltonian.

a.) Show that $[\tilde{p}_i, H] \neq 0$.

b.) Show that if we define the *magnetic translation operators*:

$$\pi_1 := p_1 + \frac{eBx_2}{2} \qquad \pi_2 := p_2 - \frac{eBx_1}{2} \tag{4.48}$$

eq:pies

then these satisfy $[\pi_i, H] = 0$. In fact, $[\pi_i, \tilde{p}_j] = 0$.

c.) The price we pay is that:

$$[\pi_1, \pi_2] = i\hbar eB \tag{4.49}$$

eq:picomm

The "magnetic translation group" is the obtained from the operators

$$\begin{aligned} U_1 &= \exp[ia^1\pi_1] \\ U_2 &= \exp[ia^2\pi_2] \end{aligned} \tag{4.50}$$

d.) Show that U_1, U_2 satisfy the relations:

$$U_1U_2 = \exp[-i\hbar eBa^1a^2]U_2U_1 \tag{4.51}$$

eq:heisgrp

****

Comment on the "noncommutative plane"

****

4.2.2 All orders in B, first order in A

Consider the following problem: Find $e^Ae^B = e^C$, when $A = \epsilon$ is infinitesimal, to all orders in B but to first order in ϵ.

$$\begin{aligned} BCH \Rightarrow C &= B + \int_0^1 g(e^{t\mathrm{Ad}\epsilon}e^{\mathrm{Ad}B})(\epsilon)dt \\ &= B + \int_0^1 g\big(e^{\mathrm{Ad}B}\big)(\epsilon)dt + \mathcal{O}(\epsilon^2) \\ &= B + \frac{\mathrm{Ad}B}{e^{\mathrm{Ad}B} - 1}(\epsilon) \\ = B + \epsilon &- \frac{1}{2}[B, \epsilon] + \frac{1}{12}[B, [B, \epsilon]] \\ &- \frac{1}{720}[B, [B, [B, [B, \epsilon]]]] + \cdots \end{aligned} \tag{4.52}$$

Note:

$$\begin{aligned} \frac{x}{e^x - 1} &= 1 - \frac{1}{2}x + \frac{1}{12}x^2 - \frac{x^4}{720} + \frac{x^6}{30240} - \frac{x^8}{1209600} + \frac{x^{10}}{47900160} \\ &- \frac{691}{1307674368000}x^{12} + \cdots \\ &\equiv \sum_{n=0}^{\infty} \frac{B_n x^n}{n!} \end{aligned} \tag{4.53}$$

eq:bernnum

is an important expansion in classical function theory -the numbers B_n are known as the Bernoulli numbers

There are many applications of this formula. One in particle physics is to spontaneous symmetry breaking where the formula above gives the chiral transformation law of the pion field. Here $B = \pi(x)$ is the pion field and ϵ is the chiral transformation parameter.

Exercise
Show that $e^C = e^A e^\epsilon$ is given to first order in ϵ by

$$C = A - \frac{\mathrm{Ad}(A)}{e^{-\mathrm{Ad}(A)} - 1}\epsilon \tag{4.54}$$

4.3 Region of convergence

The BCH formula has a finite radius of convergence. This is clear from the finite radius of convergence of the series expansion for $\log z$ around $z = 1$. Therefore it suffices to check that $e^{t\mathrm{Ad}A}e^{\mathrm{Ad}B}$ has characteristic values sufficiently close to 1.

In general, if A is diagonalizable with eigenvalues λ_i, then the eigenvalues of $\mathrm{Ad}A$ are $\lambda_i - \lambda_j$.

Exercise
a.) Let B be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that $Ad(B)$ acting on the space of all $n \times n$ complex matrices has the n^2 eigenvalues $\lambda_i - \lambda_j$ $1 \leq i, j \leq n$.

b.) Show that the Baker-Campbell-Hausdorff series can fail to converge if the eigenvalues of A differ by $2\pi i\mathbb{Z}$.

---

5. Abstract Lie Algebras

ubsec:ssGenDef

5.1 Basic Definitions

Our discussion of vector fields on Lie groups led to the general notion of a Lie algebra. We repeat the definition:

Definition 5.1.1: An abstract Lie algebra $\underline{\mathfrak{g}}$ over k is a vector space over k together with a product $(v_1, v_2) \to [v_1, v_2] \in \mathfrak{g}$, such that for all $v_1, v_2, v_3, \in \mathfrak{g}$ $\alpha, \beta \in k$.

1.) $[v_1, v_2] = -[v_2, v_1]$
2.) $[\alpha v_1 + \beta v_2, v_3] = \alpha[v_1, v_3] + \beta[v_2, v_3]$
3.) $[[v_1, v_2], v_3] + [[v_3, v_1], v_2] + [[v_2, v_3], v_1] = 0.$

Remarks

• You might be tempted to write

$$[v_1, v_2] \stackrel{?}{=} v_1 \cdot v_2 - v_2 \cdot v_1 \tag{5.1}$$ eq:tempted

but in the abstract definition (and in some examples) there is no sense in which we have defined a second product $v_1 \cdot v_2$ on $\underline{\mathbf{g}}$. Of course, this is true in some examples as in the Lie algebra $M_n(\mathbb{R})$ of all $n \times n$ matrices, but is certainly not true in general. For example, as we will see, the Lie algebra of $U(n)$ is the real vector space of $n \times n$ antihermitian matrcies. However, the product of antihermitian matrices is not antihermitian. $M_n(\mathbb{R})$ has more structure. Another example is the Lie algebra of vector fields. The product of vector fields is a second order differential operator, and is certainly not a vector field.

•

Any set of matrices that is closed under commutator satisfies the Jacobi identity:

$$[[X_1, X_2], X_3] + [[X_3, X_1], X_2] + [[X_2, X_3], X_1] = 0 \tag{5.2}$$

Simply because the product $X_1 X_2$ *is* defined for matrices. More generally, *any* associative algebra has a Lie algebra product by $[a_1, a_2] = a_1 a_2 - a_2 a_1$. Note that associativity is crucial. Moreover, if $\mathcal{A}$ is an associative algebra then the matrix commutator defines a Lie algebra structure on $M_n(\mathcal{A})$.

Exercise
Verify the statement of Remark 2.

Exercise
Considering Lie algebras from the viewpoint of algebras in general show that
a.) Lie algebras are nonassociative. Indeed, show that the associator is

$$[x_1, x_2, x_3] = [x_2, [x_3, x_1]] \tag{5.3}$$

b.) If a Lie algebra has a unit then it is trivial (over a field of characteristic $\neq 2$).

Just as for groups we can define a notion of homomorphism or isomorphism of Lie algebras, etc.

Definition 1 A *homomorphism* of two Lie algebras $\mu : \underline{\mathbf{g}}_1 \to \underline{\mathbf{g}}_2$ is
1.) a linear transformation (remember the $\underline{\mathbf{g}}_i$ are vector spaces)
2.) which preserves the Lie bracket:

$$\mu([X, Y]) = [\mu(X), \mu(Y)] \tag{5.4}$$

A one-one onto homomorphism is an isomorphism.

Some of the operations of linear algebra generalize nicely to Lie algebras. For example, a sub-Lie algebra $\mathfrak{h} \subset \mathfrak{g}$ is a vector subspace such that

$$[X, Y] \in \mathfrak{h} \tag{5.5}$$

for all $X, Y \in \mathfrak{h}$. Furthermore, a *direct sum* of two Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$, written $\mathfrak{g}_1 \oplus \mathfrak{g}_2$, is the direct sum as a vector space, but also has the property that

$$[X_1, X_2] = 0 \tag{5.6}$$ eq:dirctsum

for all $X_1 \in \mathfrak{g}_1$ and $X_2 \in \mathfrak{g}_2$.

We can also make a *semidirect sum* of Lie algebras, analogous to the semidirect product of Lie groups. In the exercise below we introduce the notion of a derivation of an algebra. If there is a homomorphism of Lie algebras from $\mathfrak{g}$ to the Lie algebra of derivations of $\mathfrak{h}$, say $Y \in \mathfrak{g} \to D_Y \in \mathrm{Der}(\mathfrak{h})$, (note that D_Y need not be an inner derivation), then we can define a Lie bracket on $\mathfrak{h} \oplus \mathfrak{g}$ by:

$$[(X_1, Y_1), (X_2, Y_2)] = ([X_1, X_2] + D_{Y_1}(X_2) - D_{Y_2}(X_1), [Y_1, Y_2]) \tag{5.7}$$ eq:semidrctsum

In the exercises you check that this indeed defines a Lie algebra structure.

In general, if $\mathfrak{h} \subset \mathfrak{g}$ is a sub-Lie algebra it is not true that the vector space quotient $\mathfrak{g}/\mathfrak{h}$ is another Lie algebra. However, if $\mathfrak{h}$ is an *ideal*, then $\mathfrak{g}/\mathfrak{h}$ can be given the structure of a Lie algebra.

Definition 2 An *ideal*, or, *invariant subalgebra* in $\mathfrak{g}$ is a sub Lie algebra $\mathfrak{h}$ such that for all $X \in \mathfrak{h}$, and all $Y \in \mathfrak{g}$,

$$[X, Y] \in \mathfrak{h} \tag{5.8}$$

Note that there is no distinction between left- and right-ideals in a Lie algebra.

Some of the operations of linear algebra *do not* generalize to Lie algebras. For example, the tensor product of Lie algebras $\mathfrak{g}_1 \otimes \mathfrak{g}_2$ makes sense as a vector space, but does not have any natural Lie bracket.

Finally, suppose $\mu : \mathfrak{g}_1 \to \mathfrak{g}_2$ is a Lie-algebra homomorphism. Then it is easy to show that $\ker\mu \subset \mathfrak{g}_1$ is not only a Lie subalgebra but an ideal and, just as for groups, if μ is surjective then

$$\mathfrak{g}_1/\ker\mu \cong \mathfrak{g}_2 \tag{5.9}$$

Exercise *Ideals*
Show that if $\mathfrak{h} \subset \mathfrak{g}$ is an ideal then $\mathfrak{g}/\mathfrak{h}$ is a well-defined Lie algebra.
Give a counterexample when $\mathfrak{h}$ is not an ideal.

Exercise *Tensor products*
Is the tensor product of two Lie algebras a Lie algebra?

Exercise *Derivations*
A *derivation* of an algebra $\mathcal{A}$ over a field k is a k-linear map $D : \mathcal{A} \to \mathcal{A}$ which satisfies the "Leibniz rule"

$$D(ab) = D(a)b + aD(b) \tag{5.10}$$

eq:leibniz

a.) Show that if D_1, D_2 are two derivations of $\mathcal{A}$ then

$$[D_1, D_2] = D_1 D_2 - D_2 D_1 \tag{5.11}$$

is a derivation (although $D_1 D_2$ is not a derivation). Conclude that the set of derivations $Der(\mathcal{A})$ itself forms a Lie algebra.

b.) If $a \in \mathcal{A}$ and $\mathcal{A}$ is associative, then the *inner derivation* D_a defined by

$$D_a(b) := ab - ba \tag{5.12}$$

is indeed a derivation, and that

$$[D_{a_1}, D_{a_2}] = D_{[a_1, a_2]} \tag{5.13}$$

c.) If $\mathcal{A}$ is a Lie algebra then show that the inner derivation

$$D_y(x) := [y, x] \tag{5.14}$$

is indeed a derivation:

$$D_y([x_1, x_2]) = [D_y(x_1), x_2] + [x_1, D_y(x_2)] \tag{5.15}$$

and moreover

$$[D_{y_1}, D_{y_2}] = D_{[y_1, y_2]} \tag{5.16}$$

d.) Show that if D is any derivation then

$$[D, D_a] = D_{D(a)} \tag{5.17}$$

and conclude that the inner derivations form an ideal in the Lie algebra of derivations.

Exercise *Semidirect sums*

a.) Show that (5.7) is a valid Lie bracket: It is antisymmetric and satisfies the Jacobi relation.

b.) Going back to chapter one, let us consider the semidirect product $H \ltimes G$ of Lie groups where $\alpha : G \to Aut(H)$ is a continuous homomorphism. Show that if we write $\alpha_{1+\epsilon Y+\cdots} = 1 + \epsilon D_Y + \cdots$ then the Lie algebra of $H \ltimes G$ is indeed the semidirect sum

$$\mathfrak{h} \tilde{\oplus} \mathfrak{g} \tag{5.18}$$

5.2 Examples: Lie algebras of dimensions $1, 2, 3$

1.) One dimensional real Lie algebras $= \mathbb{R}$. Let $\vec{e}_1$ be the a basis vector. Then $[\vec{e}_1, \vec{e}_1] = 0$. The corresponding Lie group is $(\mathbb{R}^*, \times) \cong (\mathbb{R}, +)$.

2.) Two dimensional real Lie algebras are $\mathfrak{g} = \mathbb{R}^2$, as vector spaces. Let $\vec{e}_1, \vec{e}_2$ be basis vectors. We can define:

a.) The abelian Lie algebra:

$$[\vec{e}_1, \vec{e}_2] = 0. \tag{5.19}$$

The corresponding Lie group is $\mathbb{R}^* \times \mathbb{R}^*$.

b.) The nonabelian Lie algebra:

$$[\vec{e}_1, \vec{e}_2] = \vec{e}_1 \tag{5.20}$$

Proposition. Any 2-dimensional Lie algebra is isomorphic to (a) or (b), and these are not isomorphic.

Proof: (We follow Talman and Wigner):

Suppose $\mathfrak{g} =$ 2-dimensional nonabelian Lie algebra. Let $\vec{a}, \vec{b} \in \mathfrak{g}$ be linearly independent vectors. Then $[\vec{a}, \vec{b}] = \vec{e}_1 \neq 0$. <u>*Any*</u> two vectors in $\mathfrak{g}$ can be written

$$\vec{w}_1 = \alpha_1 \vec{a} + \beta_1 \vec{b} \tag{5.21}$$

$$\vec{w}_2 = \alpha_2 \vec{a} + \beta_2 \vec{b} \tag{5.22}$$

$$[\vec{w}_1, \vec{w}_2] = (\alpha_1 \beta_2 - \beta_1 \alpha_2) \vec{e}_1. \tag{5.23}$$

Choose a <u>*new*</u> basis $\vec{e}_1, \vec{k}$ where $\vec{k}$ is linearly independent of $\vec{e}_1$.

$$[\vec{e}_1, \vec{k}] = c\vec{e}_1 \qquad c \neq 0 \tag{5.24}$$

put $\vec{k} = c\vec{e}_2$ then $[\vec{e}_1, \vec{e}_2] = \vec{e}_1$. ♠

Exercise

Consider the Lie algebra spanned by: $\{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\}$.

Show

$$\vec{e}_1 \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{5.25}$$

$$\vec{e}_2 \mapsto \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.26}$$

is a Lie algebra isomorphism with the Lie algebra in example 2b above.
Describe the corresponding Lie group.

Example 3 The three-dimensional Lie algebras. By similar arguments one can list the 6 inequivalent 3D real Lie Algebras.

a.) $[\vec{e}_1, \vec{e}_2] = [\vec{e}_1, \vec{e}_3] = [\vec{e}_2, \vec{e}_3] = 0$ b.) $[\vec{e}_1, \vec{e}_2] = [\vec{e}_1, \vec{e}_3] = 0 \quad [\vec{e}_2, \vec{e}_3] = \vec{e}_1$ c.) $[\vec{e}_1, \vec{e}_2] = 0 \quad [\vec{e}_1, \vec{e}_3] = \vec{e}_1 \quad [\vec{e}_2, \vec{e}_3] = \lambda\vec{e}_2$ d.) $[\vec{e}_1, \vec{e}_2] = 0 \quad [\vec{e}_1, \vec{e}_3] = \vec{e}_1 \quad [\vec{e}_2, \vec{e}_3] = \vec{e}_1 + \vec{e}_2$ e.) $[\vec{e}_1, \vec{e}_2] = 0 \quad [\vec{e}_1, \vec{e}_3] = \lambda\vec{e}_1 - \vec{e}_2 \quad [\vec{e}_2, \vec{e}_3] = \vec{e}_1 + \lambda\vec{e}_2$

f.) $[\vec{e}_1, \vec{e}_2] = \vec{e}_3 \quad [\vec{e}_2, \vec{e}_3] = \vec{e}_1 \quad [\vec{e}_3, \vec{e}_1] = \vec{e}_2$

Remarks

• The examples c,e exhibit an interesting phenomenon. Note that there is a continuous parameter λ so there are infinitely many inequivalent 3d Lie algebras, parametrized by λ. Parameter spaces labelling inequivalent mathematical objects are often called moduli spaces.

• (b) is called the *Heisenberg algebra*

• (f) is the Lie algebra of $sl(2)$, or of $su(2)$ depending on whether we work over $\mathbb{C}$ or over $\mathbb{R}$.

For a proof of the above classification, together with matrix representations, see the book of Wigner and Talman.

Exercise

Show that the matrices e_{13}, e_{23} and $-(e_{11} + \lambda e_{22})$ give a matrix representation of the Lie algebra of type (c) above.

5.3 Structure constants

Suppose we choose a basis $\{\vec{e}_i\}$ for $\underline{\mathbf{g}}$ as a vector space. Then there exist constants

$$[\vec{e}_i, \vec{e}_j] = \sum_k f_{ij}^k \vec{e}_k \tag{5.27}$$

eq:structurecc

Definition f_{ij}^k are called the *structure constants of* $\underline{\mathbf{g}}$ *wrt the basis* $\{\vec{e}_i\}$.

Remark: It sometimes happens that Lie algebras are defined in terms of a special basis together with relations (5.27). One must always be sure to check that such a presentation in fact satisfies the Jacobi identity. There have been cases where someone proposes a new Lie algebra, but the Jacobi identity fails (and hence no such Lie algebra exists).

Exercise *Properties of structure constants*

Show that

(a) $f_{jk}^i = -f_{kj}^i$

(b)

$$\sum_\ell \left[f_{ij}^\ell f_{\ell k}^m + f_{jk}^\ell f_{\ell i}^m + f_{ki}^\ell f_{\ell j}^m \right] = 0 \qquad \forall i,j,k,m \tag{5.28}$$

In (a) there is in general no symmetry property relating i to j,k. (b) is equivalent to the "Jacobi identity."

Exercise

A symmetric bilinear form $h : \mathfrak{g} \times \mathfrak{g} \to k$ is said to be an *invariant form* if

$$h(Ad(X)Y, Z) + h(Y, Ad(X)Z) = 0 \tag{5.29}$$

for all X, Y, Z.

Choosing a basis e_i for $\mathfrak{g}$ let $h_{ij} = h(e_i, e_j)$ and define

$$f_{ijk} := h_{i\ell} f_{jk}^\ell \tag{5.30}$$

Show that if h is an invariant form then f_{ijk} is totally antisymmetric on all three indices.

5.4 Representations of Lie algebras and Ado's Theorem

In order to discuss the main theorem of this chapter, Lie's theorem, we need a few definitions and an auxiliary result, Ado's theorem.

The space End(V) of endomorphisms of a vector space V is a Lie algebra. Thus we can define:

Definition: A *representation* of a Lie algebra $\mathfrak{g}$ with representation space V is a Lie algebra homomorphism

$$T : \mathfrak{g} \to End(V) \tag{5.31}$$

where $End(V)$ is the space of all linear transformations on V. We also say V *is a* $\mathfrak{g}$*-module.* A matrix representation is a homomorphism $T : \mathfrak{g} \to M_n(k)$ for some n.

Put differently, we have a vector space V together with a multiplication

$$\begin{aligned} &\mathfrak{g} \times V \to V \\ &(X, v) \to X \cdot v \end{aligned} \tag{5.32}$$

which is bilinear in X, v and such that

$$[X, Y] \cdot v = X \cdot (Y \cdot v) - Y \cdot (X \cdot v) \tag{5.33}$$

Upon choosing a basis for V a $\mathfrak{g}$-module provides a matrix representation of $\mathfrak{g}$.

Two representations

$$T_i : \mathfrak{g} \to End(V_i) \tag{5.34}$$

are said to be *equivalent* if there is an isomorphism $S : V_1 \to V_2$ such that $S^{-1}T_2(X)S = T_1(X)$ for all $X \in \mathfrak{g}$. If the representation defines an isomorphism of $\mathfrak{g}$ with its image, i.e. if $\ker T = \{0\}$, then the representation is said to be *faithful.*

Remarks

• Given any matrix representation of a Lie group G we automatically get a matrix representation of the Lie algebra $L(G)$: If $X = \frac{d}{dt}g(t)|_0 \in \mathfrak{g}$ then

$$T(X) := \frac{d}{dt}T(g(t))|_0 \tag{5.35}$$

eq:diffgrp

Exercise

Show that (5.35) defines a Lie algebra representation $T([X, Y]) = [T(X), T(Y)]$ by using the BCH formula.

• Any Lie algebra has a canonical representation: $V = \mathfrak{g}$,

$$X \in \mathfrak{g} \to T(X) = Ad(X) \in End(\mathfrak{g}) \tag{5.36}$$

where

$$Ad(X) : \mathfrak{g} \to \mathfrak{g} \tag{5.37}$$

is the linear operator:

$$Ad(X)(Y) := [X, Y] \tag{5.38}$$ eq:adrep

This is called the *adjoint representation*.

Exercise *The adjoint representation*

a.) Show that (5.38) *is* a representation. To do this you must show:

$$[Ad(X), Ad(Y)] = Ad([X, Y]) \tag{5.39}$$ eq:adrepii

b.) Show that the matrix elements of the adjoint representation wrt a basis X_i for $\mathfrak{g}$ are the given by the *structure constants* relative to that basis:

$$(Ad(X_i))_j^{\ k} = f_{ij}^k \tag{5.40}$$ eq:strccnt

c.) Show that the kernel of the homomorphism $X \to Ad(X)$ is the *center* of the Lie algebra:

$$Z(\mathfrak{g}) := \{X | Ad(X)Y = 0 \qquad \forall Y \in \mathfrak{g}\} \tag{5.41}$$

Thus the adjoint representation is in general not a faithful representation of $\mathfrak{g}$.

d.) Suppose $\mathfrak{g}$ is abelian. Describe the adjoint representation.

e.) Define the symmetric bilinear form

$$h(X, Y) = \mathrm{Tr}_{\mathfrak{g}} Ad(X)Ad(Y) \tag{5.42}$$

Show that $h(X, Y)$ is an invariant bilinear form. This is known as the Cartan-Killing metric, and is very important in determining the structure of semisimple Lie algebras.

Evidently, although every Lie algebra has a canonical representation - the adjoint representation - this need not be a faithful representation.

Theorem[Ado's Theorem]. Every finite dimensional Lie algebra has a faithful matrix representation.

Proof: The proof is somewhat involved. See Jacobsen, *Lie Algebras*, p. 202. The main idea is this: The kernel of the adjoint representation is the center $Z(\mathfrak{g})$. Therefore, if one can find some representation V of $\mathfrak{g}$ which is faithful on $Z(\mathfrak{g})$ then we can take a direct sum with the adjoint representation: $V \oplus \mathfrak{g}$ to produce a faithful representation of $\mathfrak{g}$.

One begins by constructing a faithful representation of $Z(\mathfrak{g})$. If it is c-dimensional then in a $c+1$-dimensional vector space we choose a nilpotent operator N such that $N^c \neq 0$. Then $Z(\mathfrak{g})$ is represented by the commutative Lie algebra generated by $N, N^2, \ldots, N^c$. Indeed, we can choose any basis for $Z(\mathfrak{g})$ and map the basis elements to the powers of N. Such a representation is faithful.

Thus we have produced a faithful representation, $V \oplus \mathfrak{g}$, of $Z(\mathfrak{g}) \oplus (\mathfrak{g}/Z(\mathfrak{g}))$. Next, one needs to upgrade this into a representation of $\mathfrak{g}$ itself. For the details, see Jacobsen. ♠

********** should there be a discussion of irreps and schur's lemma here? **********

6. Lie's theorem

We have now seen how Lie algebras emerge from Lie groups. What about going the other way, i.e. from Lie algebras to Lie groups?

Consider the case of matrix Lie algebras. If *commutators* are closed under some property then exponentiating matrices will (locally) form a group, by the BCH formula. For example, if A is anti-hermitian, $A = -A^\dagger$, then $U = \exp(A)$ is unitary:

$$UU^\dagger = (e^A)(e^{A^\dagger}) = e^A e^{-A} = 1 \tag{6.1}$$

Conversely, consider the collection of unitary matrices of the form

$$\tilde{U}(N) = \{e^A : A^\dagger = -A\} \tag{6.2}$$

Claim: $\tilde{U}(N)$ is locally closed under multiplication, i.e., it is closed as long as the BCH formula converges. The main thing we must check is that if we write

$$e^A e^B = e^C \tag{6.3}$$

with C given by the BCH formula then C is antihermitian. Suppose A, B antihermitian $A^\dagger = -A, B^\dagger = -B$. Then

$$[A, B]^\dagger = -[A, B] \tag{6.4}$$

is antihermitian. Now, for the same reason, all the terms in the BCH series are antihermitian since the expansion coefficients in the Taylor series expansion are *real*. Thus, just using BCH we would be tempted to conclude that $\tilde{U}(N)$ forms a group. However, we should worry about three problems: The BCH series does not always converge, the exponential map can fail to be surjective, and the exponential map is generally not 1-1. In fact, it turns out that the exponential map for $U(N)$ is actually surjective, and hence $\tilde{U}(N)$ above is a group after all, but we need extra information to reach this conclusion, and our discussion would also fail for certain noncompact groups.

Thus, for the general case we are going to have to work harder.

The central theorem of the subject, Lie's theorem, is the following: [5]

Theorem:

a.) Every finite dimensional Lie algebra $\mathfrak{g}$ arises from a unique (up to isomorphism) connected and simply connected Lie group G.

b.) Under this correspondence, Lie group homomorphisms $f : G_1 \to G_2$ are in $1-1$ correspondence with Lie algebra homomorphisms $\mu : T_1G_1 \to T_1G_2$.

[5] We follow closely G. Segal's elegant exposition in *Lectures on Lie Groups and Lie Algebras*, by R. Carter, G. Segal, and I. MacDonald.

Remarks

• Locally isomorphic groups have the same Lie algebra, so we need to use the simply connected cover to get a 1-1 correspondence.

• **Warning**: It follows from the theorem that if $\mathfrak{h}$ is a sub Lie algebra of $\mathfrak{g} = T_1 G$ then there is a Lie group H such that $\mathfrak{h} \cong T_1 H$, and moreover there is a group homomorphism $H \to G$. However, H need not be homomorphic to any topological subgroup of G. For example, take $G = SU(n)$, $n > 2$ and consider a $U(1) \times U(1)$ subgroup K of the maximal torus, say

$$Diag\{\lambda, \mu, (\lambda\mu)^{-1}, 1, \ldots, 1\} \qquad \lambda, \mu \in U(1) \tag{6.5}$$

eq:embedding

The group K has as Lie algebra the abelian Lie algebra $\mathbb{R} \oplus \mathbb{R}$. The corresponding Lie algebra homomorphism is

$$\nu : (x, y) \to Diag\{2\pi i x, 2\pi i y, -2\pi i(x + y), 0, \ldots, 0\} \tag{6.6}$$

Now, for $\mathfrak{h}$ take the abelian Lie algebra $\mathbb{R}$. Its corresponding connected and simply connected Lie group is $(\mathbb{R}, +)$. Choose the homomorphism μ to be obtained by embedding $\mathbb{R}$ into $\mathbb{R} \oplus \mathbb{R}$ via $1 \to (1, \alpha) \in \mathbb{R} \oplus \mathbb{R}$. Considering $\mathbb{R} \oplus \mathbb{R}$ to be a subalgebra of $L(SU(n))$ using ν we have

$$\mu : x \to Diag\{2\pi i x, 2\pi i \alpha x, -2\pi i(1 + \alpha)x, 0, \ldots, 0\} \tag{6.7}$$

The corresponding Lie group homomorphism $(\mathbb{R}, +) \to SU(n)$ guaranteed by the theorem takes

$$x \to Diag\{e^{2\pi i x}, e^{2\pi i \alpha x}, e^{-2\pi i(1+\alpha)x}, 1, \ldots, 1\} \tag{6.8}$$

If α is rational, the image is a topological subgroup of $SU(n)$. Now consider the case where α is irrational. The image is a copy of the Lie group $\mathbb{R}$ embedded in $SU(n)$, but this image is not a topological subgroup of $SU(n)$ since it is not a submanifold. Indeed, it densely fills the $U(1) \times U(1)$ subgroup K.

• Lie's theorem has an elegant and concise statement in the mathematical language of categories. There is a category $\mathcal{C}_1$ whose objects are finite dimensional connected and simply connected Lie groups, and whose morphisms are group homomorphisms. There is a category $\mathcal{C}_2$ whose objects are finite dimensional Lie algebras, and whose morphisms are Lie algebra homomorphisms. In categorical language, Lie's theorem is simply the statement that the functor $G \to L(G)$ is an equivalence of the categories $\mathcal{C}_1$ and $\mathcal{C}_2$. Indeed, the point of the theorem is to construct the functor $L(G) \to G$ such that the composition of the two functors is the identity. [6]

Proof: Let us now sketch the proof of Lie's theorem.

The hard part is showing that you can always exponentiate a finite dimensional Lie algebra to form a Lie group.

[6] In general to prove an equivalence of categories one only needs a natural transformation of FG and GF to the identity.

We use Ado's theorem to assume that $\mathfrak{g} \subset M_n(\mathbb{R})$ is a sub Lie-algebra of $M_n(\mathbb{R})$ for some n. [7]

Our Lie group will be built by thinking about the (infinite dimensional!) space of smooth paths

$$\gamma : [0,1] \to \mathfrak{g} \tag{6.9}$$

with $\gamma(0) = \gamma(1) = 0$. To such a path we can associate a group element $g_\gamma \in GL(n,\mathbb{R})$ by the "holonomy"

$$g_\gamma = P\exp \int_0^1 \gamma(t')dt' \tag{6.10}$$

This can be defined using the path ordered product or as the solution $g_\gamma(1)$ to the first order ODE:

$$\frac{d}{dt} g_\gamma(t) = \gamma(t) g_\gamma(t) \tag{6.11}$$ eq:dffqa

For bounded paths the path ordered exponential always converges, so $g_\gamma(1)$ is well-defined.

Figure 1: A closed loop in the Lie algebra maps to an open path in $GL(n,\mathbb{R})$. fig:looplie

We would like the set of elements g_γ to form a subgroup. To do this we can add the technical condition that all derivatives $\frac{d^n}{dt^n}\gamma$, $n \geq 1$, vanish at $t = 0, 1$.

Define concatenation of paths:

$$\gamma_1 * \gamma_2 = \begin{cases} 2\gamma_2(2t), & 0 \leq t \leq \frac{1}{2} \\ 2\gamma_1(2t-1), & \frac{1}{2} \leq t \leq 1 \end{cases} \tag{6.12}$$ eq:compone

The solution of (4.33) for this concatenation is

$$g_{\gamma_1 * \gamma_2}(t) = \begin{cases} g_{\gamma_2}(2t), & 0 \leq t \leq \frac{1}{2} \\ g_{\gamma_1}(2t-1) g_{\gamma_2}(1), & \frac{1}{2} \leq t \leq 1 \end{cases} \tag{6.13}$$ eq:componea

and hence:

$$g_{\gamma_1 * \gamma_2} = g_{\gamma_1} g_{\gamma_2} \tag{6.14}$$ eq:comptwo

[7] This can be dispensed with, but the proof is harder. See below.

Thus, group elements of the form g_γ are closed under multiplication. Define $G' \subset GL(n,\mathbb{R})$ to be the set of matrices g_γ corresponding to paths of the above type. From (6.14) we see that G' is closed under group multiplication. Moreover, running γ backwards gives the inverse element, so G' is indeed a subgroup of $GL(n,\mathbb{R})$. Unfortunately, G' is not quite the group we are after, because it might not be a topological subgroup of $GL(n,\mathbb{R})$, as the above example with irrational α shows. Therefore, we cannot conclude it is a Lie group.

To take care of this we define $\mathcal{P}$ to be the space of all closed paths $\gamma : [0,1] \to \mathfrak{g}$ with all derivatives vanishing at $t = 0, 1$:

$$\mathcal{P} = \{\gamma : [0,1] \to \mathfrak{g} | \left(\frac{d}{dt}\right)^n \gamma|_{t=0,1} = 0, n \geq 0\} \tag{6.15}$$ eq:curlype

This is an infinite-dimensional manifold. To compute the tangent space consider a path of elements in $\mathcal{P}$ going through some $\gamma_*(t)$. That is, a path of paths $\gamma(s,t)$ with $\gamma(0,t) = \gamma_*(t)$. Differentiating a path of such paths gives

$$T_{\gamma_*(t)}\mathcal{P} \cong \mathcal{P} \tag{6.16}$$ eq:curlypea

Moreover $\mathcal{P}$ has a multiplication $\gamma_1 * \gamma_2$. But it is not quite an infinite-dimensional Lie group - there is no identity or inverse. The natural identity would be the zero path $\gamma_0(t) = 0$, but $\gamma_0 * \gamma$ is only homotopic to γ. ($\mathcal{P}$ is in fact an infinite-dimensional *homotopy*-Lie group.)

Now, equation (6.14) shows that $\mathcal{P}$ has a map to G' preserving multiplication. The group we are after is obtained from $\mathcal{P}$ by imposing an equivalence relation:

$$\gamma_1 \sim \gamma_2 \qquad \Leftrightarrow \qquad g_{\gamma_1} = g_{\gamma_2} \tag{6.17}$$ eq:equivrel

We claim that $G := \mathcal{P}/\sim$ is a finite dimensional manifold, locally modeled on $\mathfrak{g}$, and moreover it is a Lie group. We will compute its tangent space and Lie algebra in a moment. Note that the quotient relation allows us to show that G really is a group. Moreover, it is clearly connected and simply-connected, because we can shrink paths.

Thus, G will be a nice smooth Lie group, although its image $G' \subset GL(n,\mathbb{R})$ might be very bad.

Now we would like to compute the tangent space to G, that is, the Lie algebra of G. We work in the neighborhood of the zero path $\gamma_0(t) = 0$ in $\mathcal{P}$.

Recall that there is a neighborhood $\mathcal{U}$ of the identity in $GL(n,\mathbb{R})$ so that log is well-defined. By definition of the topology of $\mathcal{P}$, there is an open neighborhood $\tilde{\mathcal{U}}$ of the zero path in $\mathcal{P}$ obtained by considering those paths so that $g_\gamma(t)$ lies in $\mathcal{U}$ for all t and hence $\log g_\gamma(t)$ is well-defined. Define $\eta_\gamma(t)$ for such a path by

$$g_\gamma(t) = e^{\eta_\gamma(t)}. \tag{6.18}$$ eq:defeta

At this point we only know that $\eta_\gamma : [0,1] \to M_n(\mathbb{R})$. Note that although $\gamma(t)$ are closed paths in $\mathfrak{g}$, $\eta_\gamma(t)$ in general will be an open path. We can take $\eta_\gamma(0) = 0$, but in general $g_\gamma \neq 1$ and in this case $\eta_\gamma(1)$ cannot vanish.

Now we show that $\eta_\gamma(t)$ is in fact valued in $\mathfrak{g}$, not just $M_n(\mathbb{R})$. To do this, note that

$$\gamma(t) = (\frac{d}{dt} g_\gamma(t)) g_\gamma(t)^{-1} = \frac{e^{\mathrm{Ad}(\eta_\gamma(t))} - 1}{\mathrm{Ad}(\eta_\gamma(t))} (\dot{\eta}_\gamma(t)) \tag{6.19}$$ eq:diffleta

Now suppose $\gamma(t)$ is scaled by ϵ. Then there a solution to (6.19) given by a power series in ϵ: $\eta_{\epsilon\gamma(t)} = \epsilon\eta_1(t) + \epsilon^2\eta_2(t) + \cdots$ where we solve

$$\begin{aligned} \dot{\eta}_1 &= \gamma(t) \\ \dot{\eta}_2 + \frac{1}{2}[\eta_1, \dot{\eta}_1] &= 0 \\ \dot{\eta}_3 + \frac{1}{2}[\eta_1, \dot{\eta}_2] + \frac{1}{2}[\eta_2, \dot{\eta}_1] + \frac{1}{6}[\eta_1, [\eta_1, \dot{\eta}_1]] &= 0 \\ \cdots &= 0 \end{aligned} \tag{6.20}$$

By induction we see that $\eta_\gamma(t) \in \mathfrak{g}$ for all t.

Now we remark that by (6.18) and the fact that exp is $1 - 1$ in our neighborhood, the equivalence relation (6.17) is the same as $\eta_{\gamma_1}(1) = \eta_{\gamma_2}(1)$. The local neighborhood $\mathcal{K} = \pi(\tilde{\mathcal{U}})$ of the identity in G is precisely

$$\mathcal{K} \cong \{\eta : [0,1] \to \mathfrak{g} | \eta(0) = 0\} / \sim \tag{6.21}$$ eq:newlocnb

where

$$\eta_1 \sim \eta_2 \qquad \Leftrightarrow \qquad \eta_1(1) = \eta_2(1) \tag{6.22}$$

But this means that an equivalence class is uniquely parametrized by the element of $\mathfrak{g}$ given by the endpoint of the path. Moreover, the space $\mathcal{P}$ is trivially a Lie algebra: The Lie bracket is just

$$[\eta_1, \eta_2](t) := [\eta_1(t), \eta_2(t)] \tag{6.23}$$ eq:liebrckt

and this Lie bracket descends to the equivalence relation. Thus, the space (6.21) is isomorphic to $\mathfrak{g}$ as a Lie algebra.

This establishes part (a) of the theorem.

To show part (b), first note that a homomorphism of Lie groups easily determines a homomorphism of Lie algebras by considering one-parameter subgroups through 1. Now for the converse, suppose we are given $\mu : \mathfrak{g}_1 \to \mathfrak{g}_2$. Using part (a) we can identify $\mathfrak{g}_1 \cong T_1 G_1$ and $\mathfrak{g}_2 \cong T_1 G_2$ where G_1, G_2 are connected and simply connected.

We would like to define $f(g)$ by writing $g = e^X$ and taking

$$f(e^X) := e^{\mu(X)} \tag{6.24}$$ eq:triy

Certainly, by BCH, if (6.24) makes sense then given that μ is a homomorphism of Lie algebras, f is a homomorphism of Lie groups. The problem is that X need not exist, and even if it does, it need not be unique. So, if $e^X = e^{X'}$ then it might not be true that $\mu(X) = \mu(X')$ and hence the attempt to define f by (6.24) in fact fails.

Instead we define $f(g)$ as follows. To begin, we choose a path $g(t)$, now in G_1, not $\mathfrak{g}_1$, with $g(0) = 1$ and $g(1) = g$. We get a corresponding path in $\mathfrak{g}_1$:

$$\gamma(t) := \dot{g}(t)g^{-1}(t). \tag{6.25}$$

Let $\tilde{\gamma}(t) := \mu(\gamma(t))$ be the corresponding path in $\mathfrak{g}_2$. We can solve

$$\frac{d}{dt}\tilde{g}(t) = \tilde{\gamma}(t)\tilde{g}(t) \tag{6.26}$$

with boundary condition $\tilde{g}(0) = 1_{G_2}$, and attempt to define

$$f(g) := \tilde{g}(1). \tag{6.27}$$ eq:trydefii

Note that if we have two group elements g_1, g_2 then, as we have seen $g_1 \cdot g_2$ corresponds to $\gamma_1 * \gamma_2$, hence to $\tilde{\gamma}_1 * \tilde{\gamma}_2$ hence to $\tilde{g}_1\tilde{g}_2$, so we do get a homomorphism.

The problem is, we need to show that $f(g)$ in (6.27) is well-defined since we made a choice of path $g(t)$ to define it. It is exactly at this point we make use of the fact that G_1 is simply connected. Suppose that $g_0(t), g_1(t)$ are two paths with $g_0(1) = g_1(1)$. Since G_1 is simply connected we can write a homotopy $g(s,t)$ between these two paths.

That is, there is a smooth map $g(s,t)$ mapping $I \times I \to G$ such that $g(0,t) = g_0(t)$, $g(1,t) = g_1(t)$, $g(s,0) = 1$ and $g(s,1) = g$:

Then we define:

$$\begin{aligned} \gamma(s,t) &:= \frac{\partial}{\partial t} g(s,t)g(s,t)^{-1} \\ \eta(s,t) &:= \frac{\partial}{\partial s} g(s,t)g(s,t)^{-1} \end{aligned} \tag{6.28}$$ eq:twodervs

Note that these satisfy the Maurer-Cartan equation

$$\frac{\partial}{\partial s}\gamma - \frac{\partial}{\partial t}\eta = [\eta, \gamma] \tag{6.29}$$ eq:mceq

An important point is that the converse also holds. Namely, (6.29) is the integrability condition for the existence of a solution $g(s,t)$ to (6.28) for a given γ, η.

Therefore, we apply the Lie algebra homomorphism μ to (6.29) to get

$$\frac{\partial}{\partial s}\tilde{\gamma} - \frac{\partial}{\partial t}\tilde{\eta} = [\tilde{\eta}, \tilde{\gamma}] \tag{6.30}$$ eq:mceqa

and, using the converse result, this is the integrability condition for the existence of a solution $\tilde{g}(s,t)$ to

$$\begin{aligned} \frac{\partial}{\partial t}\tilde{g}(s,t) &:= \tilde{\gamma}(s,t)\tilde{g}(s,t) \\ \frac{\partial}{\partial s}\tilde{g}(s,t) &:= \tilde{\eta}(s,t)\tilde{g}(s,t) \end{aligned} \tag{6.31}$$

eq:twodervsa

Now, $g(s,1) = g$ is independent of s. Therefore $\eta(s,1) = 0$. Therefore $\tilde{\eta}(s,1) = 0$. Therefore $\frac{\partial}{\partial s}\tilde{g}(s,1) = 0$, so (6.27) is well-defined. ♠

Remarks

• As we mentioned, it is possible to avoid Ado's theorem, and proceed directly from $\mathfrak{g}$ to the definition of the simply connected Lie group. To do this, we return to $\mathcal{P}$, but now we introduce a more subtle equivalence relation. Now we say that $\gamma_0 \sim \gamma_1$ if there exists a homotopy $\gamma(s,t)$ of paths in $\mathcal{P}$ such that there is a partner $\eta(s,t)$ satisfying the Maurer-Cartan equation:

$$\frac{\partial}{\partial s}\gamma - \frac{\partial}{\partial t}\eta = [\eta, \gamma] \tag{6.32}$$

for some $\eta(s,t) \in \mathfrak{g}$. Then we claim that $\mathcal{P}/\sim := G$ is the desired Lie group. The hard part is to show that the Lie algebra of G is indeed $\mathfrak{g}$. To do this we solve

$$f(Ad\eta_\gamma(t))\dot{\eta}_\gamma(t) = \gamma(t) \tag{6.33}$$

and show that $\gamma_1 \sim \gamma_2$ iff $\eta_{\gamma_1}(1) = \eta_{\gamma_2}(1)$. Then

$$L(G) = \{\eta : [0,1] \to \mathfrak{g}\}/\sim \tag{6.34}$$

is isomorphic to $\mathfrak{g}$.

• We are really using here the theory of flat connections. The Maurer-Cartan equation is the statement that the gauge field $A = \gamma dt + \eta ds$ is a flat gauge field: $F = dA + A^2 = 0$.

• One of the advantages of the above presentation is that it can be generalized to infinite-dimensional Lie groups. See

J. Milnor, "Remarks On Infinite Dimensional Lie Groups," *In *Les Houches 1983, Proceedings, Relativity, Groups and Topology, Ii*, 1007-1057*

7. Lie Algebras for the Classical Groups

Now we can review the Lie algebras of some matrix groups, including the classical groups. As we have seen, if G is a matrix group, that is, a Lie subgroup of $GL(n,k)$ then the Lie algebra associated to the Lie group G is:

$$L(G) \equiv \{A = \frac{d}{dt}\Big|_0 g(t) : g(t) = any \qquad curve \quad g(t) \quad with \quad g(0) = 1\} \tag{7.1}$$

One way to think about Lie algebras related to the classical groups is as "infinitesimal group elements." Since Lie groups are manifolds we can consider a neighborhood of the

identity element 1_G. In this neighborhood the manifold looks like a linear space, namely, the tangent space T_1G. We now want to get a better understanding of how the structure of the group translates into an algebraic structure on T_1G in the classical matrix groups.

For example, if G is a matrix group then matrices $g \in G$ close to the identity can be written as

$$g \cong 1 + \epsilon A \tag{7.2}$$

where ϵ is small. Thus infinitesimal group elements defines a matrix A in the Lie algebra. Thus, for example, if A is anti-hermitian then $1 + \epsilon A$ is unitary to order ϵ^2:

$$(1 + \epsilon A)(1 + \epsilon A)^\dagger = 1 + \mathcal{O}(\epsilon^2) \tag{7.3}$$

We will now list the relevant properties for the classical matrix groups. Our discussion will parallel the survey we did of the matrix groups:

1. Definition and dimension.
2. Properties of the exponential map.
3. A natural choice of a *Cartan subalgebra* $\mathfrak{t} \subset \mathfrak{g}$. By definition, a *Cartan subalgebra* is a maximal abelian subalgebra of $\mathfrak{g}$. It is of supreme importance in the theory of roots and weights and representations. A Cartan subalgebra is the Lie algebra of a corresponding Cartan torus.

7.1 A useful identity

The following identity is extremely useful: Given *any* $n \times n$ matrix $A \in M_n(k)$, the determinant and the trace may be related as follows:

$$\boxed{\det(\exp(A)) = \exp(\mathrm{Tr}A)} \tag{7.4}$$ eq:trcelog

This is sometimes written

$$\mathrm{logdet}C = \mathrm{Trlog}C \tag{7.5}$$

but the latter identity only makes sense when the log is well-defined.

We have already proven (7.4) above as an application of Jordan canonical form.

7.2 $GL(n,k)$ and $SL(n,k)$

The Lie algebra of $GL(n,k)$ is just the Lie algebra of all $n \times n$ matrices:

$$\boxed{gl(n,k) \equiv L(GL(n,k)) = M_n(k)} \tag{7.6}$$

Let us check this by producing a one-parameter family. Let $A \in M_n(k)$ and consider $g(t) := \exp(tA)$. These matrices are all invertible, and

$$\frac{d}{dt}|_{t=0} g(t) = A \tag{7.7}$$

Thus we get the full Lie algebra of $n \times n$ matrices.

Again by $\det e^A = \exp(\mathrm{Tr}(A))$ we get

$$\boxed{sl(n,k) \equiv L(SL(n,k)) = \{A \in M_n(k) \mid \mathrm{tr}\, A = 0\}} \tag{7.8}$$

We have a direct sum of Lie algebras:

$$gl(n,k) = k \oplus sl(n,k) \tag{7.9}$$

A natural choice of Cartan subalgebra is the subalgebra of diagonal matrices.

Remarks

• The exponential map is *not* onto for $SL(n,\mathbb{R})$. Let us consider $SL(2,\mathbb{R})$. If $g = e^A$ then g lies on some one-parameter subgroup $g(t) = e^{tA}$. Any such can be conjugated into a maximal abelian subgroup. From our classification of the conjugacy classes of $SL(2,\mathbb{R})$ it follows from $hgh^{-1} = e^{hAh^{-1}}$ that A can be conjugated to one of three possible forms:

$$A = x \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7.10}$$ eq:hyperbol

$$A = x \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{7.11}$$ eq:elliptic

$$A = x \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{7.12}$$ eq:parabol

It is easy to check that the trace of e^A for A conjugate to (7.10)(7.11)(7.12) is ≥ -2, and the trace is $= -2$ only for $g = -1 = \exp[i\pi\sigma^2]$. In particular *The parabolic conjugacy classes with $Tr(g) = -2$ and the hyperbolic conjugacy classes with $Tr(g) < -2$ are not in the image of the exponential map.*

Exercise *Structure constants for $GL(n,k)$.*

The Lie algebra has a natural basis given by the matrix units: e_{kj} = The matrix with zero entries except = 1 at $(kj)^{th}$ matrix element.

Calculate the structure constants

$$[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta_{il} e_{kj} \tag{7.13}$$

Exercise *Structure constants for $sl(2,\mathbb{C})$ and $sl(2,\mathbb{R})$*

Show that we may choose a basis

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad E^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad E^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{7.14}$$ eq:spc

with structure constants

$$\begin{aligned} [E^+, E^-] &= H \\ [H, E^\pm] &= \pm 2E^\pm \end{aligned} \tag{7.15}$$ eq:sltwo

We will see a lot more of this later.

Exercise *The exponential map is onto for* $GL(n, \mathbb{C})$
Show that

$$\exp : gl(n, \mathbb{C}) \to GL(n, \mathbb{C}) \tag{7.16}$$ eq:expmapon

is *surjective*.

Answer: It suffices to check this statement for Jordan form. Therefore, try to write $\lambda + N$, where $N = e_{12} + e_{23} + \cdots$ is a $d \times d$ Jordan block in the image of the exponential map. Recall $N^d = 0$ but $N^{d-1} \neq 0$. Note that

$$\exp[x1 + \alpha_1 N + \alpha_2 N^2 + \cdots + \alpha_{d-1} N^{d-1}] = \tag{7.17}$$

$$= e^x \left(1 + \alpha_1 N + (\alpha_2 + \frac{1}{2}\alpha_1^2) N^2 + \cdots + (\alpha_{d-1} + \cdots + \frac{\alpha_1^{d-1}}{(d-1)!} N^{d-1} \right) \tag{7.18}$$

and note that we can set $e^x = \lambda$ and then solve the upper-triangular system of equations for the α_i.

Exercise
a.) Show that matrices of the form

$$\begin{pmatrix} -x & 0 \\ 0 & -1/x \end{pmatrix} \tag{7.19}$$

for $x > 0$ are in $SL(2, \mathbb{R})$, are in the image of the exponential map for $SL(2, \mathbb{C})$, but not in the image of the exponential map for $SL(2, \mathbb{R})$.

b.) Show that

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \tag{7.20}$$

is in $SL(2, \mathbb{C})$, is in the image of the exponential map for $GL(2, \mathbb{C})$, but not for $SL(2, \mathbb{C})$.

7.3 $O(n,k)$

Suppose $g(t) = e^{tA}$ is a family of orthogonal matrices. Then

$$g(t)g(t)^{tr} = 1 \tag{7.21}$$ eq:orthone

since they are orthogonal for each t. Differentiate (7.21) wrt t to get:

$$\frac{d}{dt}|_{t=0} e^{tA}e^{tA^{tr}} = A + A^{tr} = 0 \tag{7.22}$$ eq:orthlie

and we can conclude from (7.22) that the Lie algebra of $O(n,k)$ is the Lie algebra of $n \times n$ skew matrices:

$$\boxed{o(n,k) \equiv L(O(n,k)) = \{A \in M_n(k) \mid A^{tr} = -A\}} \tag{7.23}$$

Note that $o(n,k) = so(n,k)$ because the groups $O(n,k)$ and $SO(n,k)$ have the same neighborhood of the identity. In general locally isomorphic Lie groups have the same Lie algebra.

Note that the exponential map is *not* one to one. For example:

$$\exp[2\pi \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}] = 1 \tag{7.24}$$

Let us give a basis for the Lie algebra. Introduce the matrices:

$$T_{ij} := e_{ij} - e_{ji} \tag{7.25}$$

so that $T_{ij} = -T_{ji}$. Then T_{ij} for $1 \leq i < j \leq n$ form a basis for $o(n,k)$. We compute the structure constants:

$$[T_{ij}, T_{kl}] = \delta_{jk}T_{il} - \delta_{li}T_{kj} - \delta_{ik}T_{jl} + \delta_{lj}T_{ki} \tag{7.26}$$ eq:soscs

N.B. T_{ij} are elements of a Lie algebra - they are NOT matrix elements!

For $o(2r,k)$, or $o(2r+1,k)$, a natural basis for a Cartan subalgebra is

$$T_{1,2}, T_{3,4}, \ldots, T_{2r-1,2r}. \tag{7.27}$$

7.4 More general orthogonal groups

Recall that for any symmetric bilinear form D we can define an orthogonal group $O(D)$. If we choose a basis and represent D as a symmetric matrix D_{ij} then $O(D)$ becomes the set of matrices with

$$gDg^{tr} = D \tag{7.28}$$

Following the procedure above we find the Lie algebra is the set of matrices A such that

$$AD + DA^{tr} = 0. \tag{7.29}$$

In particular, specializing to $D = \eta = ((+1)^p, (-1)^q)$ we can rewrite this condition as:

$$(\eta A)^{tr} = -(\eta A) \tag{7.30}$$

Thus, the matrices are skew-symmetric *after* multiplication by η. Thus we could choose a basis $\hat{T}_{ij} \equiv \eta T_{ij}$, for $o(p,q)$ where T_{ij} are a basis for $o(p+q)$. Thus a basis for the Lie algebra is:

$$\hat{T}_{ij} = \eta_{ii} e_{ij} - \eta_{jj} e_{ji}, \qquad 1 \leq i < j \leq n \tag{7.31}$$

We can read off immediately the structure constants for $o(p,q)$:

$$[\hat{T}_{ij}, \hat{T}_{kl}] = \eta_{jk}\hat{T}_{il} - \eta_{li}\hat{T}_{kj} - \eta_{ik}\hat{T}_{jl} + \eta_{lj}\hat{T}_{ki} \tag{7.32}$$

eq:soscsp

Again, we caution that the subcripts do not label matrix elements but elements in a basis for the Lie algebra. In particular it is always true that:

$$\hat{T}_{ij} = -\hat{T}_{ji}. \tag{7.33}$$

However, note that if $\eta_{ii} = \eta_{jj}$ then $\hat{T}_{ij}$ is an antisymmetric matrix, but if $\eta_{ii} = -\eta_{jj}$ then $\hat{T}_{ij}$ is a *symmetric* matrix.

Exercise
Compute the structure constants for the Lorentz group

7.4.1 Lie algebra of $SO^*(2n)$

There is one other "real form" of $SO(2n, \mathbb{C})$ given by matrices of the type:

$$\begin{pmatrix} A_+ + iA_- & B_+ + iB_- \\ -B_+ + iB_- & A_+ - iA_- \end{pmatrix} \tag{7.34}$$

where $A_\pm, B_\pm$ are real $n \times n$ matrices with

$$(A_\pm)^{tr} = -A_\pm \qquad\qquad (B_\pm)^{tr} = \pm B_\pm \tag{7.35}$$

For more discussion about this consult Gilmore, p.343.

7.5 $U(n)$

Proceeding to unitary matrices $g(t) = e^{tA}$ is unitary if:

$$e^{tA^\dagger} e^{tA} = 1 \tag{7.36}$$

Differentiating at $t = 0$ we conclude that:

$$A^\dagger = -A \tag{7.37}$$

So:

$$\boxed{\begin{array}{c} u(n) = L(U(n)) = \{A \in M_n(\mathbb{C}) \mid A^\dagger = -A\} \\ su(n) = L(SU(n)) = \{A \in L(U(n)) : \operatorname{tr} A = 0\} \\ u(n) = \mathbb{R} \oplus su(n) \end{array}} \tag{7.38}$$

A Cartan subalgebra is generated by matrices ie_{aa}, $a = 1, \ldots n$ for $u(n)$, but for $su(n)$ we need to have traceless matrices. One choice of basis is $i(e_{aa} - e_{a+1,a+1})$, $1 \leq a \leq n-1$.

Remarks

• Physicists usually separate out a factor of $i = \sqrt{-1}$ and write

$$U = e^{iH} \tag{7.39}$$

where H is <u>hermitian</u>.

•

Note that the condition $A^\dagger = -A$ is only preserved by multiplication of A by real numbers so we must consider these as Lie algebras over $\mathbb{R}$.

Exercise

Let us consider $SU(2)$ in particular. Show that the Lie algebra can be thought of as the real span of three anti-hermitian matrices (iJ_k), $k = 1, 2, 3$, with J_k Hermitian, such that

$$[J_i, J_j] = i\epsilon_{ijk} J_k \tag{7.40}$$

where ϵ_{ijk} is the antisymmetric tensor on three indices. Take $J_i = +\frac{1}{2}\sigma_i$.

Note that the structure constants $-\epsilon_{ijk}$ of this Lie algebra are *real* if we use the real basis $\frac{i}{2}\sigma^i$.

Exercise *Relating $o(4)$ and $su(2)$*

We saw in the previous lecture that at the level of groups:

$$SO(4) = \frac{SU(2) \times SU(2)}{\mathbb{Z}_2} \tag{7.41}$$

eq:sutwotwo

This must be reflected in the structure of the Lie algebras.

The Lie algebra version of (7.41) is that

$$o(4) \cong su(2) \oplus su(2) \tag{7.42}$$

Do this by defining

$$\tilde{A}_{\mu\nu} := \frac{1}{2}\epsilon_{\mu\nu\lambda\rho} A_{\lambda\rho} \tag{7.43}$$

for any antisymmetric tensor $A_{\mu\nu} = -A_{\nu\mu}$. In particular, for the generators of $so(4)$ we can define:

$$\tilde{T}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}T_{\lambda\rho} \tag{7.44}$$

Now form the "self-dual" and "anti-self-dual" combinations of generators:

$$A^{\pm}_{\mu\nu} = \frac{1}{2}[T_{\mu\nu} \pm \tilde{T}_{\mu\nu}] \tag{7.45}$$

a.) Show that

$$\tilde{A}^{\pm}_{\mu\nu} = \pm A^{\pm}_{\mu\nu} \tag{7.46}$$

b.) Show that $[A^+, A^-] = 0$ and that

$$\begin{aligned} J_1 \to J_1^+ &:= A^+_{23} = A^+_{14} = \frac{1}{2}(T_{23} + T_{14}) \\ J_2 \to J_2^+ &:= A^+_{31} = A^+_{24} = \frac{1}{2}(T_{31} + T_{24}) \\ J_3 \to J_3^+ &:= A^+_{12} = -A^+_{34} = \frac{1}{2}(T_{12} + T_{34}) \end{aligned} \tag{7.47}$$

defines an isomorphism of the Lie subalgebra of the $A^+_{\mu\nu}$ and $su(2)$. Show that the corresponding statement for the $A^-_{\mu\nu}$ is

$$\begin{aligned} J_1 \to J_1^- &:= \frac{1}{2}(T_{23} - T_{14}) \\ J_2 \to J_2^- &:= \frac{1}{2}(T_{31} - T_{24}) \\ J_3 \to J_3^- &:= \frac{1}{2}(T_{12} - T_{34}) \end{aligned} \tag{7.48}$$

Exercise *The 't Hooft symbols*

When working in Euclidean 4-dimensional physics the 't Hooft symbol is often of great use. This is defined by

$$\alpha^{\pm,i}_{\mu\nu} := \frac{1}{2}(\pm\delta_{i\mu}\delta_{\nu 4} \mp \delta_{i\nu}\delta_{\mu 4} + \epsilon_{i\mu\nu}) \tag{7.49}$$ eq:thooft

where $1 \le \mu, \nu \le 4$, $1 \le i \le 3$ and it is understood that $\epsilon_{i\mu\nu}$ is zero unless $1 \le \mu, \nu \le 3$.

a.) Show that in the previous exercise we can take:

$$J^{\pm,i} = \frac{1}{2}\alpha^{\pm,i}_{\mu\nu}T_{\mu\nu} \tag{7.50}$$

b.) Show that if we regard $\alpha^{\pm,i}$ as 4×4 matrices with matrix elements $\alpha^{\pm,i}_{\mu\nu}$ then

$$[\alpha^{\pm,i}, \alpha^{\pm,j}] = -\epsilon^{ijk}\alpha^{\pm,k} \tag{7.51}$$ eq:tfhti

$$[\alpha^{\pm,i}, \alpha^{\mp,j}] = 0 \tag{7.52}$$ eq:tfhtii

$$\{\alpha^{\pm,i}, \alpha^{\pm,j}\} = -\frac{1}{2}\delta^{ij} \tag{7.53}$$ eq:tfhtiia

See Belitsky, VanDoren, VanNieuwenhuizen, hep-th/0004186, for many more identities of this nature. A closely related tensor are the so-called "'t Hooft symbols" $\eta^a_{\mu\nu}$. Definitions differ, but roughly $\eta = 2\alpha$.

Exercise *Spinor indices and 't Hooft symbols*

7.5.1 $U(p,q)$

The discussion here parallels that for $O(p,q)$. Introduce η as before. Now a one-parameter subgroup through 1 in $U(p,q)$ satisfies

$$e^{tA^\dagger}\eta e^{tA} = \eta \tag{7.54}$$

for $A \in u(p,q)$. Thus,

$$(\eta A)^\dagger = -(\eta A) \tag{7.55}$$

and ηA is antihermitian.

Etc.

7.5.2 Lie algebra of $SU^*(2n)$

7.6 $Sp(2n)$

Following the same procedure as before we take a family of symplectic matrices $g(t) = e^{tA}$. They must satisfy:

$$e^{tA^{tr}} J e^{tA} = J \tag{7.56}$$

Differentiating at $t = 0$ gives:

$$A^{tr}J + JA = 0 \tag{7.57}$$ eq:sympcond

Conversely, (7.57) and $J^2 = -1$ imply:

$$-Je^{A^{tr}}J = e^{-JA^{tr}J} = e^{-A} \quad \Rightarrow \quad e^{A^{tr}}J = Je^{-A} \quad \Rightarrow \quad e^A \in Sp(2n,k) \tag{7.58}$$

by pulling J to left and using (7.57) successively. so:

$$\boxed{sp(2n,k) \equiv L(Sp(2n,k)) = \{A \in M_{2n}(k) \mid (JA)^{tr} = +(JA)\}} \tag{7.59}$$

Now we can easily count the dimension of $Sp(2n,\mathbb{R})$. We need only take a real symmetric $2n \times 2n$ matrix and identify it with JA. Thus the dimension is:

$$\dim_{\mathbb{R}} Sp(2n,\mathbb{R}) = \frac{1}{2}2n(2n+1) \tag{7.60}$$ eq:dimen

in accord with a direct analysis at the group level.

Recall that the Cartan torus of $USp(2n)$ is given by diagonal matrices of the form

$$\begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} \tag{7.61}$$

where Λ is a diagonal matrix of phases. The corresponding Cartan subalgebra is spanned by

$$i(e_{aa} - e_{a+n,a+n}) \qquad a = 1, \ldots, n \tag{7.62}$$

Exercise

Define a basis $m_{\alpha\beta}$ for the Lie algebra from

$$Jm_{\alpha\beta} = -(e_{\alpha\beta} + e_{\beta\alpha}) \tag{7.63}$$

i.e.

$$m_{\alpha\beta} = J(e_{\alpha\beta} + e_{\beta\alpha}) \tag{7.64}$$

Show that

$$[m_{\alpha\beta}, m_{\gamma\delta}] = J_{\beta\gamma}m_{\alpha\delta} + J_{\alpha\gamma}m_{\beta\delta} + J_{\beta\delta}m_{\alpha\gamma} + J_{\alpha\delta}m_{\beta\gamma} \tag{7.65}$$

One way to remember this is that the result must be *symmetric* on (α, β) as well as (γ, δ) and *antisymmetric* on $(\alpha, \beta) \leftrightarrow (\gamma, \delta)$.

Exercise

It is also worthwhile to examine the condition (7.57) in terms of the block decomposition. Represent the matrix A in block form:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{7.66}$$

Show that (7.57) implies:

$$A_{11}^{tr} = -A_{22} \tag{7.67}$$

$$A_{12}^{tr} = A_{12} \qquad A_{21}^{tr} = A_{21} \tag{7.68}$$

Recall that if $g \in Sp(2n)$ then $g^{tr} \in Sp(2n)$. Prove this using the Lie algebra.

Exercise $su(1,1) \cong sl(2,\mathbb{R}) \cong sp(2,\mathbb{R})$

a.) Show that a basis for the *real* Lie algebra $su(1,1)$ can be taken to be

$$E^{\pm} = \frac{1}{2}(\sigma^1 \mp i\sigma^3) \tag{7.69}$$

$$H = \sigma^2 \tag{7.70}$$

and these satisfy the above standard commutation relations of $sl(2,\mathbb{R})$.

b.) Note that $sl(2,\mathbb{R}) \cong sp(2,\mathbb{R})$ trivially from the above characterization of block diagonal form.

Exercise $su(2) \cong usp(2,\mathbb{R})$

Exercise *Solving the general quadratic Hamiltonian in quantum mechanics*

Show that symplectic transformations allow us to solve the general quadratic Hamiltonian in quantum mechanics for a finite number of degrees of freedom:

Consider the general quadratic Hamiltonian for N degrees of freedom $r = 1, \ldots, N$.

$$H = M_{rs}P_rP_s + A_{rs}P_rX_s + A^*_{rs}X_sP_r + G_{rs}X_rX_s \tag{7.71}$$ eq:ghen

where, after quantization $[P_i, Q_j] = -i\delta_{ij}\hbar$, $[P_i, P_j] = [X_r, X_s] = 0$. We restrict to Hermitian Hamiltonians so that M_{rs} and G_{rs} are real symmetric. A_{rs} is an arbitrary complex matrix.

a.) By adding a c-number we can assume that A_{rs} is real.

b.) Write the Hamiltonian as

$$H = \begin{pmatrix} X_r & P_r \end{pmatrix} J \begin{pmatrix} -A_{rs} & -M_{rs} \\ G_{rs} & A^{tr}_{rs} \end{pmatrix} \begin{pmatrix} X_s \\ P_s \end{pmatrix} \tag{7.72}$$ eq:symphamil

and observe that it is of the form

$$H = \begin{pmatrix} X_r & P_r \end{pmatrix} Jh \begin{pmatrix} X_s \\ P_s \end{pmatrix} \tag{7.73}$$ eq:sntne

where $h \in sp(2N,\mathbb{R})$.

Conclude that by a complex symplectic transformation

$$\begin{pmatrix} X_s \\ P_s \end{pmatrix} \to g \begin{pmatrix} X_s \\ P_s \end{pmatrix} \tag{7.74}$$

we can bring $h \to g^{-1}hg$ into the Lie algebra of the Cartan torus. That is, we can bring it to the form

$$\begin{pmatrix} i\omega & 0 \\ 0 & -i\omega \end{pmatrix} \tag{7.75}$$

where ω is diagonal and real. In this basis the Hamiltonian takes the form

$$H = \sum_{i=1}^{N} \omega_i \{A_i^\dagger, A_i\} \tag{7.76}$$

c.) Finding the relevant transformation explicitly can be hard. The Hamiltonian can be simplified and the problem can be reduced to the diagonalization of an $N \times N$ Hermitian matrix as follows:

By a symplectic transformation $X \to TX, P \to T^{tr,-1}P$ we can bring M_{rs} to the form δ_{rs}. Then, by conjugating with $\exp[i\frac{1}{2}S_{nm}X_nX_m]$ we can make A_{rs} antisymmetric. Thus we can bring H to the form

$$H = P_rP_r + A_{rs}(P_rX_s + X_sP_r) + G_{rs}X_rX_s = (P_r + A_{rs}X_s)^2 + (G + A^tA)_{rs}X_rX_s \tag{7.77}$$

eq:ghens

with A_{rs} real antisymmetric. (This may be interpreted as charged particles in a constant magnetic field in a quadratic potential.)

The Heisenberg equations of motion

$$-i\hbar \frac{\partial}{\partial t}\mathcal{O}(t) = [H, \mathcal{O}] \tag{7.78}$$

become:

$$-i\partial_t \begin{pmatrix} X_r \\ P_r \end{pmatrix} = K \begin{pmatrix} X_s \\ P_s \end{pmatrix} \tag{7.79}$$

eq:heisenb

where

$$K = -2i \begin{pmatrix} A & 1 \\ -G & A \end{pmatrix} \tag{7.80}$$

eq:kay

1. K is in the symplectic Lie algebra $sp(2N)$ and hence by a symplectic transformation can be brought to the Cartan subalgebra:

$$U^{-1}KU = \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \tag{7.81}$$

eq:cartfrm

where ω is $N \times N$ diagonal. This is the transformation to creation/annihilation operators.

To find U note that eigenvectors of K are of the form

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{7.82}$$

with

$$L(E)v_1 = \frac{E^2}{4}v_1 \qquad v_2 = -(A - \frac{i}{2}E)v_1 \tag{7.83}$$

where $L(E) = G + A^2 - iEA$. This is an Hermitian matrix and can be diagonalized with eigenvalues $f_i(E)$. We then solve the quadratic equations $f_i(E) = E^2/4$.

---

8. Central extensions of Lie algebras and Lie algebra cohomology

The ideas here parallel those for central extensions of groups:

A central extension of a Lie algebra $\mathfrak{g}$ by an abelian Lie algebra $\mathbf{z}$ is a Lie algebra $\widetilde{\mathfrak{g}}$ such that we have an exact sequence of Lie algebras:

$$0 \to \mathbf{z} \to \widetilde{\mathfrak{g}} \to \mathfrak{g} \to 0 \tag{8.1}$$

with $\mathbf{z}$ mapping into the center of $\widetilde{\mathfrak{g}}$. As a *vector space* (but not necessarily as a Lie algebra) $\widetilde{\mathfrak{g}} = \mathbf{z} \oplus \mathfrak{g}$ so we can denote elements by (z, X) and the Lie bracket has the form

$$[(z_1, X_1), (z_2, X_2)] = (c(X_1, X_2), [X_1, X_2]) \tag{8.2}$$

where $c : \Lambda^2\mathfrak{g} \to \mathbf{z}$ is known as a *two-cocycle* on the Lie algebra. That is $c(X, Y)$ is bilinear, it satisfies

$$c(X, Y) = -c(Y, X) \tag{8.3}$$ eq:twococyci

and the Jacobi relation requires

$$c([X_1, X_2], X_3) + c([X_3, X_1], X_2) + c([X_2, X_3], X_1) = 0. \tag{8.4}$$ eq:twococycia

Two different cocycles can define isomorphic Lie algebras. If there is a *linear* function $f : \mathfrak{g} \to \mathbf{z}$ such that

$$c(X, Y) = b_f(X, Y) = f([X, Y]) \tag{8.5}$$ eq:cobodn

then the cocycle is said to be trivial, and the central extension is isomorphic to $\mathbf{z} \oplus \mathfrak{g}$ as a Lie algebra. Indeed,

$$X \to (f(X), X) \tag{8.6}$$

is an explicit Lie algebra homomorphism splitting the sequence. (Check!)

More generally, if two cocycles differ by a cocycle of the form b_f then they define isomorphic Lie algebras. Thus, again, classifying isomorphism classes of central extensions is a cohomology problem, in this case, Lie algebra cohomology.

Remark: Suppose that $\widetilde{G}$ is a Lie group central extension of a Lie group G. Then a 2-cochain $c(X,Y)$ defines a left-invariant 2-form on G, and a 2-cocycle defines a closed left-invariant two-form ω on G. A central extension of the Lie group, defined by the group cocycle $f(g_1,g_2)$, also defines a corresponding 2-cocycle on the Lie algebra by

$$c(X_1,X_2) = \frac{1}{2\pi i}\frac{d}{dt_1}|_0\frac{d}{dt_2}|_0 \log\left[\frac{f(e^{t_1X_1},e^{t_2X_2})}{f(e^{t_2X_2},e^{t_1X_1})}\right] \tag{8.7}$$ eq:twoconscy

8.1 Example: The Heisenberg Lie algebra and the Lie group associated to a symplectic vector space

In general, given any real vector space V and a skew bilinear form $\omega: V\times V\to\mathbb{R}$ we can view V as an abelian Lie algebra and use ω to define a central extension

$$0\to\mathbb{R}\to \mathrm{heis}(V,\omega)\to V\to 0 \tag{8.8}$$ eq:heislielg

We simply use the cocycle $c(v_1,v_2)=\omega(v_1,v_2)$.

When ω is a symplectic form, *i.e.* ω is nondegenerate, then the corresponding Lie algebra is called a *Heisenberg Lie algebra*.

The corresponding Lie group $\mathrm{Heis}(V,\omega)$ is $U(1)\times V$ as a set with multiplication

$$(\lambda_1,v_1)(\lambda_2,v_2) = (\lambda_1\lambda_2 e^{i\omega(v_1,v_2)}, v_1+v_2) \tag{8.9}$$

Let us return to the corollary (4.44) of the BCH formula:

$$e^{\alpha_1\hat{p}+\beta_1\hat{q}}e^{\alpha_2\hat{p}+\beta_2\hat{q}} = e^{-\frac{i}{2}(\alpha_1\beta_2-\alpha_2\beta_1)}e^{(\alpha_1+\alpha_2)\hat{p}+(\beta_1+\beta_2)\hat{q}} \tag{8.10}$$ eq:heisgroupi

We can take $V=\mathbb{R}^2=\{(\alpha,\beta)\}$ and we recognize

$$\omega((\alpha_1,\beta_1),(\alpha_2,\beta_2)) = -\frac{1}{2}(\alpha_1\beta_2-\alpha_2\beta_1) \tag{8.11}$$

as the symplectic form in this case.

(Of course, we could take $V=\mathbb{C}^2$ and consider it as a 4-dimensional real vector space. If we take α,β pure imaginary then the operators are unitary in the usual representation on $L^2(\mathbb{R})$.)

Exercise

Check that $e^{i\omega(v_1,v_2)}$ is a 2-cocycle on the abelian group V.

Exercise

Referring to the exercise in section **** on the magnetic translation group

Show that the magnetic translation group generated by the U_i for all $a^i \in \mathbb{R}$ is a Heisenberg group, extending the abelian group of translations in the plane. Show that the skew bilinear form is

$$\omega(\vec{a}, \vec{b}) = -eB(\vec{a} \times \vec{b}) \tag{8.12}$$ eq:skewfrm

---

8.2 Lie algebra cohomology

The notion of a two-cocycle above fits into a more general theory of Lie algebra cohomology. So to put it into proper context let us briefly consider the topic more generally.

We consider the simplest case of central extensions by a one-dimensional abelian Lie algebra k, $\mathbf{z} = k$.

We define a complex

$$d : \Lambda^k \mathfrak{g}^* \to \Lambda^{k+1} \mathfrak{g}^* \tag{8.13}$$ eq:lacoh

by taking as differential:

$$d\omega(X_1, \dots, X_{k+1}) := \sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], \dots \hat{X}_i, \dots, \hat{X}_j \dots) \tag{8.14}$$ eq:extderi

The resulting differential (8.13) may also be expressed by identifying $\Lambda^* \mathfrak{g}^* = \Lambda^*[\theta^a]$ where θ^a are of degree 1 and a runs over a basis for the Lie algebra. We then *define* the differential to be:

$$d\theta^a := -\frac{1}{2} f_{bc}{}^a \theta^b \theta^c \tag{8.15}$$ eq:lacohii

---

Exercise

Check that this is a differential, that is, that $d^2 = 0$.

The cohomology of the complex $(\Lambda^* \mathfrak{g}^*, d)$ is known as *Lie algebra cohomology* and denoted $H^*(\mathfrak{g})$. Note that it can be formulated purely algebraically. The differential defined by (8.14) or, equivalently, (8.15) is sometimes called the *Chevalley-Eilenberg differential.*

It is useful to translate this into the physicist's language:

Suppose we have a Lie algebra $\mathfrak{g}$ with basis T_a, a is an index running over the generators. Let us introduce the Clifford algebra:

$$\begin{aligned} \{c^a, c^{a'}\} &= 0 \\ \{b_a, b_{a'}\} &= 0 \\ \{c^a, b_{a'}\} &= \delta^a{}_{a'} \end{aligned} \tag{8.16}$$ eq:ghosti

where $c^a, b_{a'}$ are referred to as ghosts and antighosts, respectively.

We can quantize the Clifford algebra by choosing a Clifford vacuum

$$b_{a'}|0\rangle = 0 \tag{8.17}$$ eq:ghostii

and the resulting Hilbert space is spanned by $|0\rangle, c^a|0\rangle, ...$

The Hilbert space is graded by the "ghost number operator" $N = \sum_a c^a b_a$, and we have an isomorphism of the vector space of states of ghost number k with $\Lambda^k \mathfrak{g}^*$:

$$\omega \leftrightarrow \frac{1}{k!}\omega_{a_1 \cdots a_k} c^{a_1} \cdots c^{a_k}|0\rangle \tag{8.18}$$ eq:ghostiii

Under the isomorphism (8.18) the Chevalley-Eilenberg differential becomes what is known as the *BRST operator*:

$$Q := -\frac{1}{2} f_{a_2 a_3}^{\ \ a_1} c^{a_2} c^{a_3} b_{a_1} \tag{8.19}$$ eq:ghostiv

The BRST cohomology is the cohomology of Q, and is graded by ghost number.

BRST cohomology enters physics in the quantization of theories with local gauge symmetry. For the moment let us note a very natural generalization. Suppose we have a representation ρ of the Lie algebra $\mathfrak{g}$. We can then consider the complex

$$\Lambda^* \mathfrak{g}^* \otimes V \tag{8.20}$$

and introduce a differential

$$Q = c^a t_a - \frac{1}{2} f_{a_2 a_3}^{\ \ a_1} c^{a_2} c^{a_3} b_{a_1} \tag{8.21}$$ eq:ghostv

where $t_a = \rho(T_a)$ are the representation matrices of the rep. V.

Geometrically, the cohomology $H^*_Q(\Lambda^* \mathfrak{g}^* \otimes V)$ can be identified, for G compact and connected, with the cohomology $H^*_{DR}(G; \mathcal{V})$ of a homogeneous vector bundle over the group G.

Exercise

Check that $Q^2 = 0$

Exercise

Compute the degree 2 Lie algebra cohomology of $su(2)$ and show that it is trivial.

---

9. Left-invariant differential forms

We introduced the Lie algebra of a Lie group by thinking about the left-invariant vector fields on the group. It is quite useful to think about the dual picture of left-invariant differential forms:

Definition: A differential form $\omega \in \Omega^*(G)$ is *left invariant* if

$$(L_g)^*(\omega) = \omega \tag{9.1}$$

for all $g \in G$.

Suppose first that $G = GL(n, \mathbb{F})$. Then we can construct an explicit basis of left-invariant 1-forms as follows. The basis of one-forms dual to $\xi(e_{ij})$ is

$$\Theta_{i'j'} = (g^{-1})_{i'k} dg_{kj'} \tag{9.2}$$ eq:dualbasis

One can check explicitly that $\Theta_{i'j'}$ is a dual basis to $\xi(e_{ij})$ by computing the contraction:

$$\iota(\xi(e_{ij}))(\Theta_{i'j'}) = \langle \Theta_{i'j'}, \xi(e_{ij}) \rangle = \delta_{ii'}\delta_{jj'} \tag{9.3}$$ eq:contractior

It is convenient to assemble the 1-forms Θ_{ij} into a matrix of one-forms

$$\Theta := \Theta_{ij} e_{ij} = g^{-1} dg. \tag{9.4}$$ eq:maurercarta

Let us check the left-invariance directly in matrix notation:

$$(L_{g_0})^*(g^{-1}dg) = (g_0 g)^{-1} d(g_0 g) = g^{-1} g_0^{-1} g_0 dg = g^{-1} dg \tag{9.5}$$

The matrix of 1-forms (9.4) is known as the *Maurer-Cartan form.* Note that, if $X \in gl(n, \mathbb{F})$ then we have

$$\boxed{\iota(\xi(X))(\Theta) = X} \tag{9.6}$$ eq:contr

Suppose now that $G \subset GL(n, \mathbb{F})$ is a Lie subgroup. Under restriction to G the matrix elements g_{ij} are not all independent, but satisfy polynomial relations. Nevertheless, we can pull back the expression $\Theta = g^{-1}dg$ to G. Formally, we should introduce the inclusion map $j : G \hookrightarrow GL(n, \mathbb{F})$ and speak of $\Theta_G = j^*(\Theta)$. But to avoid cluttered notation we drop this.

Now, for $g \in G$, the matrix of 1-forms will be valued in the lie algebra $\mathfrak{g}$ of G considered as a subalgebra of $Mat_n(\mathbb{F})$. Indeed, if

$$\gamma_X(t) = g_0 e^{tX} \tag{9.7}$$

is a curve through g_0 whose directional derivative is $\xi(X)$ then

$$\gamma_X^*(g^{-1}dg) = e^{-tX} g_0^{-1} d(g_0 e^{tX}) = X dt. \tag{9.8}$$

This shows that

$$\langle \xi(X), \Theta \rangle = X \in \mathfrak{g} \tag{9.9}$$

In this way $\Theta = g^{-1}dg$ defines the Maurer-Cartan forms on any Lie subgroup of a matrix group.

Example $G = SO(2)$ If

$$g = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{9.10}$$

so

$$g^{-1}dg = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} d\theta \tag{9.11}$$

so

$$\langle \frac{\partial}{\partial\theta}, g^{-1}dg \rangle = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{9.12}$$

10. The Maurer-Cartan equation

A very important property of the left-invariant 1-forms is the *Maurer-Cartan equation*. Differentiating, using $d^2 = 0$ and the Leibniz rule gives:

$$d\Theta = -g^{-1}dg \wedge g^{-1}dg = -\Theta \wedge \Theta \tag{10.1}$$ eq:derivmc

We thus have the *Maurer-Cartan equation*:

$$\boxed{d\Theta + \Theta \wedge \Theta = 0} \tag{10.2}$$ eq:mcequations

Since in general the matrix product of two Lie algebra elements is not in the Lie algebra, it might not be immediately obvious to the reader that $\Theta \wedge \Theta$ is valued in $\mathfrak{g}$. To make this obvious we proceed as follows:

It is often useful to choose a basis T_a for $\mathfrak{g}$. Then we can define *structure constants*:

$$[T_a, T_b] = f_{ab}{}^c T_c \tag{10.3}$$

Now, since $g^{-1}dg$ is valued in the Lie algebra $\mathfrak{g}$ we can expand:

$$g^{-1}dg = e^a T_a \tag{10.4}$$

and this defines a system of 1-forms e^a on G which span $T^*_g(G)$ at any point g. Then we compute

$$\begin{aligned} 0 &= de^a T_a + e^a T_a \wedge e^b T_b \\ &= de^a T_a + \frac{1}{2} e^a \wedge e^b [T_a, T_b] \end{aligned} \tag{10.5}$$

In terms of the structure constants we have:

$$\boxed{de^a + \frac{1}{2} f_{bc}{}^a e^b \wedge e^c = 0} \tag{10.6}$$ eq:structure

If follows from the above that the dual basis of vectorfields e_a is precisely the system of fundamental vector fields described above

$$e_a = \xi(T_a) \tag{10.7}$$ eq:funda

Remarks:

• Although we have taken a rather concrete approach. The equations (10.6) and (10.7) hold for all Lie groups. (Not all finite dimensional Lie groups are matrix groups. Heisenberg groups give a counterexample. However, by the Peter-Weyl theorem, all compact Lie groups are subgroups of some unitary group, and hence are matrix groups.)

• Later on, we will interpret this as a zero-fieldstrength condition for a pure gauge field.

• Note that we can write:

$$\Theta \wedge \Theta = \frac{1}{2}[\Theta, \Theta] \tag{10.8}$$

where it is useful to introduce the notation

$$[\alpha, \beta] := \alpha \wedge \beta - (-1)^{pq}\beta \wedge \alpha \tag{10.9}$$ eq:gradcomm

when α is a matrix-valued p-form and β is a matrix-valued q-form. Thus, the ordinary commutator would read $\Theta\Theta - \Theta\Theta = 0$, but since we are discussing matrices whose elements are in a *graded* algebra the natural notion of commutator is the "graded commutative" (10.9).

• For future use in our discussion of connections on principal bundles, note that (10.2) implies

$$\mathcal{L}_{\xi(X)}\Theta = (d\iota(X) + \iota(X)d)\Theta = -\iota(X)\Theta^2 = -Ad(X)(\Theta) \tag{10.10}$$ eq:liederv

where in the last equality we consider Θ to be a Lie-algebra valued one-form, valued in the adjoint representation.

Exercise

One way of defining the exterior derivative of a differential k-form field is via the formula:

$$\begin{aligned} d\omega(X_1, \dots, X_{k+1}) &:= \sum_{i=1}^{k+1}(-1)^{i+1}X_i\big(\omega(\dots, \hat{X}_i, \dots)\big) \\ &+ \sum_{i<j}(-1)^{i+j}\omega([X_i, X_j], \dots \hat{X}_i, \dots, \hat{X}_j \dots) \end{aligned} \tag{10.11}$$ eq:extderiv

where $\hat{X}$ means that entry is deleted.

a.) Using this formula, show that the Maurer-Cartan equation is equivalent to the fact that the left-invariant vector fields form a closed Lie algebra.

b.) In particular, show that (10.6) is compatible with

$$[e_a, e_b] = +f_{ab}^c e_c \tag{10.12}$$

c.) Since left invariant forms are determined by their value at $g = 1$ we can identify $\Omega_L^k(G) \cong \Lambda^k \mathfrak{g}^*$, as vector spaces. This gives us a new perspective on Lie algebra cohomology defined above.

(Hint: If ω is left-invariant and X_i are left-invariant then $\omega(X_1, \ldots, X_k)$ is left-invariant, and hence constant.)

Exercise *Right-invariance*

a.) Show that

$$\tilde{\Theta} = dgg^{-1} \tag{10.13}$$

is a matrix of right-invariant 1-forms.

b.) Show that

$$d\tilde{\Theta} - \tilde{\Theta} \wedge \tilde{\Theta} = 0 \tag{10.14}$$

c.) Show that if we expand

$$\Theta = \omega_L^a T_a \tag{10.15}$$

$$\tilde{\Theta} = \omega_R^a T_a \tag{10.16}$$

then

$$\omega_L^a = (Ad(g))^a{}_b \omega_R^b \tag{10.17}$$

so the bases of left- and right-invariant forms are related by the adjoint representation.

---

11. Examples

11.1 $SU(2)$

Introduce angular coordinates:

$$\begin{aligned} g &= e^{\frac{i}{2}\sigma^3\phi} e^{\frac{i}{2}\sigma^2\theta} e^{\frac{i}{2}\sigma^3\psi} \\ &= e^{i\frac{1}{2}\phi\sigma_3} \begin{pmatrix} \cos\theta/2 & \sin\theta/2 \\ -\sin\theta/2 & \cos\theta/2 \end{pmatrix} e^{i\frac{1}{2}\psi\sigma_3} \qquad (11.1) \quad \texttt{eq:angl} \\ 0 \le \theta \le \pi, \quad \phi &\sim \phi + 2\pi \quad \psi \sim \psi + 4\pi \end{aligned}$$

Then

$$\Theta = g^{-1}dg = \frac{i}{2}\sigma^a e^a \tag{11.2}$$

$$
\begin{aligned}
e^1 &= \cos\psi \sin\theta d\phi - \sin\psi d\theta \\
e^2 &= \sin\psi \sin\theta d\phi + \cos\psi d\theta \\
e^3 &= d\psi + \cos\theta d\phi
\end{aligned}
\tag{11.3}
$$

eq:leftforms

which is conveniently written as

$$
g^{-1}dg = \frac{i}{2}\sigma^3(d\psi + \cos\theta d\phi) + \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} e^{-i\psi}(d\theta + i\sin\theta d\phi) - \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} e^{i\psi}(d\theta - i\sin\theta d\phi)
\tag{11.4}
$$

With the basis $T^a = \frac{i}{2}\sigma^a$ the structure constants are $f_{ab}^{\ c} = -\epsilon_{abc}$, and one easily verifies the Maurer-Cartan equations directly:

$$
de^3 = e^1 \wedge e^2 \qquad + cyclic
\tag{11.5}
$$

eq:sutmc

The dual basis of vector fields is

$$
\begin{aligned}
e_3 &= \frac{\partial}{\partial\psi} \\
e_2 &= \cos\psi\frac{\partial}{\partial\theta} + \frac{\sin\psi}{\sin\theta}\frac{\partial}{\partial\phi} - \sin\psi\cot\theta\frac{\partial}{\partial\psi} \\
e_1 &= -\sin\psi\frac{\partial}{\partial\theta} + \frac{\cos\psi}{\sin\theta}\frac{\partial}{\partial\phi} - \cos\psi\cot\theta\frac{\partial}{\partial\psi}
\end{aligned}
\tag{11.6}
$$

eq:leftvfs

and one checks with an explicit computation that

$$
[e_a, e_b] = -\epsilon_{abc}e_c
\tag{11.7}
$$

eq:explicit

Remark: Note that $d\psi$ is *not* a globally well-defined form on $SU(2)$, nor is $\cos\theta d\phi$. However, the e^a *are* globally well-defined. Similarly, the e_a are globally well-defined vector fields.

Exercise

It is often useful to introduce coordinates on $\mathbb{C}^2$:

$$
\begin{aligned}
z^1 &= re^{i(\psi+\phi)/2}\cos\theta/2 \\
z^2 &= re^{i(\psi-\phi)/2}\sin\theta/2
\end{aligned}
\tag{11.8}
$$

which satisfy $|z^1|^2 + |z^2|^2 = r^2$. Setting $r = 1$ gives a coordinate system the covers $SU(2)$ (but degenerates at the poles).

Write out the forms in terms of these variables.

Exercise

Show that under $g \to g^{-1}$ we have

$$e^a_L(g) = -e^a_R(g^{-1}) \tag{11.9}$$

and that in coordinates this is $\phi \leftrightarrow -\psi, \theta \to -\theta$

Deduce that

$$\begin{aligned} e^1_R &= -\cos\phi \sin\theta d\psi + \sin\phi d\theta \\ e^2_R &= \sin\phi \sin\theta d\psi + \cos\phi d\theta \\ e^3_R &= d\phi + \cos\theta d\psi \end{aligned} \tag{11.10}$$

eq:rightforms

with dual basis of vector fields

$$\begin{aligned} e^R_3 &= \frac{\partial}{\partial\phi} \\ e^R_2 &= \cos\phi\frac{\partial}{\partial\theta} + \frac{\sin\phi}{\sin\theta}\frac{\partial}{\partial\psi} - \sin\phi\cot\theta\frac{\partial}{\partial\phi} \\ e^R_1 &= -\sin\phi\frac{\partial}{\partial\theta} - \frac{\cos\phi}{\sin\theta}\frac{\partial}{\partial\psi} + \cos\phi\cot\theta\frac{\partial}{\partial\phi} \end{aligned} \tag{11.11}$$

eq:leftvfsa

11.1.1 The Heisenberg group

An interesting example which is not a compact Lie group are the Heisenberg groups.

The three-dimensional Heisenberg group is the group of matrices

$$g = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \tag{11.12}$$

with $x, y, z \in \mathbb{R}$. The multiplication law is

$$(x, y, z) \cdot (x', y'z') = (x + x', y + y', z + z' + xy') \tag{11.13}$$

(note it is not abelian.)

An easy computation shows that

$$g^{-1}dg = \begin{pmatrix} 0 & dx & dz - xdy \\ 0 & 0 & dy \\ 0 & 0 & 0 \end{pmatrix} \tag{11.14}$$

12. Metrics on Lie groups

A very natural class of metrics on Lie groups are those which are left or right invariant.

As with forms and vector fields (and indeed all tensors), invariant metrics are uniquely determined by their value at $g = 1$, where they are bilinear forms on the Lie algebra. A natural way to make such a metric is to choose a representation ρ of the Lie group and define:

$$\begin{aligned} ds^2 &= \mathrm{Tr}_\rho(g^{-1}dg) \otimes (g^{-1}dg) \\ &= \mathrm{Tr}_\rho(dgg^{-1}) \otimes (dgg^{-1}) \end{aligned} \tag{12.1}$$

eq:natmetric

Using equation (4.4)(??) above we see that this corresponds to the metric

$$(X_1, X_2) = \mathrm{Tr}_\rho(\rho(X_1)\rho(X_2)) \tag{12.2}$$

eq:natmetricii

on the Lie algebra.

Warning: While it is natural, the left-right invariant metric (12.1) might be zero. For example, consider the group of upper triangular matrices with 1 on the diagonal. Similarly, consider the Heisenberg group above. Also, for infinite-dimensional groups, the cyclicity of the trace used in (12.1) might be invalid.

Metrics which are simultaneously left- and right- invariant correspond to Ad-invariant metrics on the Lie algebra. Every Lie algebra has a canonical representation, namely, the Adjoint representation. This defines the Ad-invariant Cartan-Killing metric:

$$g_{CK}(X_1, X_2) := \mathrm{Tr}_{\mathfrak{g}}(Ad(X_1)Ad(X_2)) \tag{12.3}$$

eq:cartkill

Example 1: $O(n)$ has a basis for the Lie algebra $T_{ij} = e_{ij} - e_{ji}$, $1 \le i, j \le n$. Then

$$g_{CK}(T_{ij}, T_{kl}) = -2(n-2)\delta_{i,k}\delta_{j,l} \tag{12.4}$$

eq:onck

Example 2: For $SU(2)$, if we regard $SU(2) = S^3$ as the unit sphere in Euclidean $\mathbb{R}^4$ then the metric is expressed in terms of the left-invariant forms is $1/2$ of the metric in the 2-dimensional representation:

$$\begin{aligned} d\Omega_3^2 &= \frac{1}{4}\sum_{a=1}^{3} e^a \otimes e^a \\ &= \frac{1}{4}\Big[(d\psi + \cos\theta d\phi)^2 + (d\theta)^2 + \sin^2\theta (d\phi)^2\Big] \\ &= \frac{1}{4}\Big[(d\psi)^2 + (d\theta)^2 + (d\phi)^2 + 2\cos\theta d\psi d\phi\Big] \end{aligned} \tag{12.5}$$

eq:sumtrc

Exercise

Using the round metric of $SU(2)$ compute the inner products of the left and right invariant vector fields and show that

$$\begin{aligned} g(e_a^L, e_b^L) &= g(e_a^R, e_b^R) = \delta_{ab} \\ g(e_a^L, e_b^R) &= R_{ab}(\psi, \theta, \phi) \end{aligned} \tag{12.6}$$

eq:lrips

where R_{ab} is the $SO(3)$ rotation matrix associated with the Euler angles.

Exercise

Show that if we write the metric on a Lie algebra as

$$(T_a, T_b) := I_{ab} \tag{12.7}$$

then Ad-invariance is equivalent to:

$$f_{ac}{}^d I_{db} = I_{ad} f_{cb}{}^d \tag{12.8}$$

Exercise

Show that if we choose a basis T_a for $\mathfrak{g}$ then the metric components of g_{CK} are given by

$$(g_{CK})_{ab} = f_{ac}{}^d f_{bd}{}^c \tag{12.9}$$

eq:carkillii

Exercise

Compute the CK form for $U(n)$, $SU(n)$, $USp(2n)$.

Exercise

Show that the volume of $SU(N)$ in the metric

$$ds^2 = -Tr_N[(g^{-1}dg) \otimes (g^{-1}dg)] \tag{12.10}$$

eq:metric

is

$$\mathrm{vol}\,(SU(N)) = \frac{\sqrt{N}}{2\pi}(2\pi)^{\frac{1}{2}N(N+1)}\frac{1}{1!2!3!\cdots(N-1)!} \tag{12.11}$$

Hint:

Consider $SU(N)/SU(N-1) = S^{2N-1}$ and relate the local coordinates of S^{2N-1} and $SU(N)$ via

$$g = 1 + iy^N Diag\{\epsilon, \ldots, \epsilon, 1\} + \sum_i (z^i e_{iN} - \bar{z}^i e_{Ni}) + \cdots \tag{12.12}$$ eq:localcoord

at $g = 1$. (Here $\epsilon = -1/(N-1)$.). Here we are thinking of S^{2N-1} as the solutions to $\sum_{i=1}^N |z^i|^2 = 1$ and $z^N = x^N + iy^N$. In these coordinates, the group metric (12.10) becomes

$$ds^2 = 2\sum_{i=1}^{N-1}[(dx^i)^2 + (dy^i)^2] + \frac{N}{N-1}(dy)^2 \tag{12.13}$$ eq:metrspher

Volumes of G and G/K for compact symmetric spaces (in the Killing metric) were computed in A. Kojun and Y. Ichiro, "Volumes of compact symmetric spaces," Tokyo J. Math. **20** 1997)p.87.

12.1 Simple Lie groups and the index of a representation

If we define a simple Lie algebra as one which has no nontrivial invariant subalgebras then we have the following key result:

Exercise *Uniqueness of the Cartan-Killing form*

a.) Show that the *Ad*-invariant bilinear forms on a finite-dimensional simple Lie algebra are unique up to a constant.

b.) Show that if T_a is an ON basis in the CK form, then the most general invariant metric on a simple Lie group is

$$ds^2 = \Omega^2 \sum_{a=1}^{\dim G} e^a \otimes e^a \tag{12.14}$$

where Ω^2 is a constant.

Given the uniqueness of the CK form up to scale, for any representation ρ of $\mathfrak{g}$ set

$$g_\rho(X_1, X_2) = \mathrm{Tr}_\rho(\rho(X_1)\rho(X_2)) = \ell(\rho) g_{CK}(X_1, X_2) \tag{12.15}$$ eq:liemetric

where $\ell(\rho)$ is a constant called the index of the representation. This is often useful to have handy when working with instanton effects in gauge theory.

Example 1: For $SU(2)$ we can easily compute the index of the irreducible representations by simply computing the norm-squared of J_3. For the adjoint representation (spin =1) we have

$$g_{CK}(J_3, J_3) = 1^2 + 0^2 + (-1)^2 = 2 \tag{12.16}$$ eq:cksut

and for spin j we have

$$g_{\mathbf{j}}(J_3, J_3) = \sum_{m=-j}^{+j} m^2 = \sum_{\ell=0}^{2j} (-j+\ell)^2 = \frac{1}{3} j(j+1)(2j+1) \tag{12.17}$$ eq:cksutii

and therefore,

$$f(\mathbf{j}) = \frac{1}{6} j(j+1)(2j+1) \tag{12.18}$$

Example 2 For $SU(N)$, $\mathfrak{g} = su(N)$ is the algebra of $N \times N$ traceless antihermitian matrices.

$$\mathrm{Tr}_{adj} = 2N \mathrm{Tr}_N \tag{12.19}$$ eq:indxemb

Remarks

• Tables of in the indexes of representations are given in books by Patera.

Exercise

Show that

$$\frac{\ell(\rho_1)}{\ell(\rho_2)} = \frac{\dim \rho_1 \cdot C_2(\rho_1)}{\dim \rho_2 \cdot C_2(\rho_2)} \tag{12.20}$$

where $C_2(\rho)$ is the value of the Casimir in the representation ρ.

Exercise

Suppose that the decomposition of a tensor product of representations is given by

$$\rho_1 \otimes \rho_2 = \sum_\lambda N^\lambda \rho_\lambda \tag{12.21}$$

Show that

$$\ell(\rho_1) D(\rho_2) + \ell(\rho_2) D(\rho_1) = \sum_\lambda N^\lambda \ell(\rho_\lambda) \tag{12.22}$$ eq:relateindx

where $D(\rho)$ is the dimension of ρ.

Exercise

Show that for a semisimple Lie algebra the degree two Lie algebra cohomology vanishes: There are no nontrivial central extensions.

12.2 Metrics on abelian groups

In contrast to the simple groups, where the invariant metrics are unique up to an overall constant, the metrics on abelian groups are far from unique. Indeed, the torus $U(1)^n$ has a moduli space of invariant metrics $GL(n,\mathbb{R})/O(n)$.

12.3 Geodesics

Geodesics through the origin are of the form $\gamma(t) = \exp[tX]$ for $X \in \mathfrak{g}$.

********** EXPLAIN WHY **********

12.4 Compact and noncompact real forms: The signature of the CK metric

For a good discussion on this topic see Gilmore's book.

By the KAN decomposition (Gram-Schmidt decomposition) we see that every connected Lie group is (topologically) a product of its maximal compact subgroup and a Euclidean space.

A general result identifies the compact subalgebra with the negative definite generators and the noncompact generators with the positive definite subspace.

Key example: Compact generator

$$X = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus 0 \oplus 0 \oplus \cdots \tag{12.23}$$

This is compact in the sense that its 1-parameter subgroup forms a compact orbit:

$$\exp[tX] = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \oplus 1 \oplus 1 \oplus \cdots \tag{12.24}$$

Note that:

$$\mathrm{Tr}(X)^2 = -2 \tag{12.25}$$

while a noncompact generator

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus 0 \oplus 0 \oplus \cdots \tag{12.26}$$

has

$$\exp[tX] = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} \oplus 1 \oplus 1 \oplus \cdots \tag{12.27}$$

and

$$\mathrm{Tr}(X)^2 = +2 \tag{12.28}$$

12.5 Example: $SL(2,R)$

The Lie algebra $sl(2,\mathbb{R})$ is the Lie algebra of traces 2×2 real matrices.

One natural basis of the Lie algebra has commutation relations:

$$[K_0, K_\pm] = \pm K_\pm, \qquad [K_+, K_-] = -2K_0 \ . \tag{12.29}$$ eq:sucom

In the 2-dimensional representation we could take, for example:

$$K_0 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad K_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad K_- = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \tag{12.30}$$

Maurer-Cartan forms:

$$\begin{aligned} U &:= e^{C_+K_+}e^{C_0K_0}e^{C_-K_-} \\ dUU^{-1} &= e^{-C_0}dC_-K_- + (dC_0 - 2e^{-C_0}C_+dC_-)K_0 + (dC_+ + e^{-C_0}C_+^2dC_- - C_+dC_0)K_+ \\ U^{-1}dU &= dC_+e^{-C_0}K_+ + (dC_0 - 2C_-e^{-C_0}dC_+)K_0 + (dC_- - C_-dC_0 + C_-^2e^{-C_0}dC_+)K_- \end{aligned} \tag{12.31}$$ eq:mcforms

The Killing metric is:

$$(aK_0 + bK_+ + cK_-, aK_0 + bK_+ + cK_-) = 2(a^2 - 4bc) \tag{12.32}$$

Note that this is a Lorentzian signature metric.

Casimir:

$$4K_0^2 - 2(K_+K_- + K_-K_+) \tag{12.33}$$

The compact generator is $K_+ + K_-$ and the KAN decomposition is:

$$SO(2)\begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \tag{12.34}$$

Exercise

a.) Compute the Maurer-Cartan forms in the global coordinates of the KAN decomposition.

b.) Suppose we write the Gauss decomposition:

$$g = \begin{pmatrix} 1 & 0 \\ \chi & 0 \end{pmatrix}\begin{pmatrix} e^{\phi} & 0 \\ 0 & e^{-\phi} \end{pmatrix}\begin{pmatrix} 1 & \psi \\ 0 & 1 \end{pmatrix} \tag{12.35}$$

(warning: These are not globally valid coordinates on $SL(2,\mathbb{R})$.) Show that

$$\mathrm{Tr}(g^{-1}dg)^3 = 2e^{2\phi}d\phi\wedge d\psi\wedge d\chi \tag{12.36}$$

References:

1. A comprehensive reference on the geometry (but not the topology) of Lie groups is Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, 1978.

2. A less systematic, but more readable account for physicists is R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications*, John Wiley, 1974.

13. Some Remarks on Infinite Dimensional Lie Algebras

13.1 Generalities

The study of infinite dimensional Lie algebras and their representations is an active subject in modern mathematical physics. Infinite dimensional Lie algebras arise in conformal field theory, string theory, gauge theory of elementary particles, and gravitational theory classical and quantum mechanical. They have also play a role in some field theoretical applications of condensed matter physics such as the quantum Hall effect, and the Kondo problem.

****

Banach Lie groups. Comments from Milnor.

****

13.2 Groups of operators on Hilbert space

A very important infinite-dimensional group is the group $U(\mathcal{H})$ of unitary operators on Hilbert space. The topology of this group depends strongly on how we define the group precisely. The inductive limit $\lim_{n\to\infty} U(n)$ has lots of topology (it is a classifying space for K^1).

On the other hand, there is a very deep and remarkable theorem, known as Kuiper's theorem $U(\mathcal{H})$ which states that in the norm topology $U(\mathcal{H})$ is *contractible*!

13.3 Gauge Groups

For another set of examples, consider a Lie group G, and a manifold M. The space of all smooth maps $M \to G$ forms an infinite-dimensional group with Lie algebra the space of all maps $M \to \mathfrak{g}$. Except for the case where M is one-dimensional, very little is known about these groups. The one-dimensional case is of central importance in string theory and low-dimensional field theory.

13.3.1 Loop algebras

Suppose $\mathfrak{g}$ is a finite dimensional simple Lie algebra. We can associate to it an infinite dimensional Lie algebra whose elements are maps

$$f : S^1 \to \mathfrak{g} \tag{13.1}$$

We will be vague about the precise class of maps - it should be some completion of the space of Laurent polynomials in $z = e^{i\theta}$. The set of all such maps is itself a Lie algebra for

we can define

$$[f_1, f_2](z) := [f_1(z), f_2(z)] \tag{13.2}$$

This infinite dimensional Lie algebra is known as the loop algebra $L\mathfrak{g}$.

It is often useful to choose a basis T^a for $\mathfrak{g}$, for this induces an obvious basis for $L\mathfrak{g}$:

$$T^a_n := T^a z^n \tag{13.3}$$

with commutation relations

$$[T^a_n, T^b_m] = f^{ab}_c T^c_{n+m} \tag{13.4}$$

Loop algebras admit a very interesting central extension: [8]

$$0 \to \mathbb{R} \to \widetilde{L\mathfrak{g}} \to L\mathfrak{g} \to 0 \tag{13.5}$$

Elements of $\widetilde{L\mathfrak{g}}$ are pairs (f, ξ) where f is a map, and $\xi \in \mathbb{C}$. The bracket of the central extension is

$$[(f_1, \xi_1), (f_2, \xi_2)] := ([f_1, f_2], \omega(f_1, f_2)) \tag{13.6}$$

where

$$\omega(f_1, f_2) := \frac{1}{2\pi i} \oint (f_1'(z), f_2(z))_{\mathfrak{g}} dz \tag{13.7}$$

where in the integral we use an ad-invariant bilinear form $(\cdot, \cdot)_{\mathfrak{g}}$, such as the Cartan-Killing form on $\mathfrak{g}$. (Recall that for G simple all such forms are the same up to scale, consistent with the uniqueness of the central extension.)

Again, it is useful to write this out in terms of a basis. Using the bilinear form on $\mathfrak{g}$ we define

$$(T^a, T^b)_{\mathfrak{g}} = g^{ab} \tag{13.8}$$

and we define $K := (0, 1) \in \widetilde{L\mathfrak{g}}$. Then

$$\begin{aligned} [T^a_n, T^b_m] &= f^{ab}_c T^c_{n+m} + n g^{ab} \delta_{n+m,0} K \\ [K, T^a_n] &= 0 \end{aligned} \tag{13.9}$$

eq:kmalg

This is the way the algebras are often written in the physics literature.

13.4 Diffeomorphism Groups

A nice example is given by the space of vector fields on a manifold and the related group of diffeomorphisms of a manifold.

Already, the simplest special case is highly nontrivial. Consider the group diffeomorphisms of the circle.

A natural basis of vector fields on the circle is given by Fourier decomposition:

$$\ell_n = i e^{in\theta} \frac{d}{d\theta} \tag{13.10}$$

[8] It is an easy result that, up to isomorphism, this is the unique nontrivial central extension when $\mathfrak{g}$ is simple.

They satisfy the Lie algebra:

$$[\ell_n, \ell_m] = (n-m)\ell_{n+m} \tag{13.11}$$

One can show that there is a one-dimensional cohomology space and a representative cocycle is

$$\omega(\ell_n, \ell_m) = \delta_{n+m,0} n^3 \tag{13.12}$$ eq:repcocycl

Note this is cohomologous to

$$\omega(\ell_n, \ell_m) = \delta_{n+m,0}(n^3 - n) \tag{13.13}$$ eq:repcocycla

The unique central extension of the Lie algebra of vector fields on S^1 is known as the Virasoro algebra.

The Virasoro algebra can be defined in terms of a basis L_n, $n \in \mathbb{Z}$ and a central element c, with structure constants:

$$\begin{aligned} [c, c] &= 0 \\ [c, L_n] &= 0 \\ [L_n, L_m] &= (n-m)L_{n+m} + \frac{c}{12}(n^3 - n)\delta_{n+m,0} \end{aligned} \tag{13.14}$$ eq:virasoro

The normalization of the central element c is convenient in physical applications. The choice of cocycle is made so that the Lie algebra generated by $L_{\pm 1}, L_0$ forms a copy of $sl(2)$ and is not extended.

Exercise

a.) Check the Jacobi relations of the Virasoro algebra.

b.) Show that $\omega(\ell_n, \ell_m) = n\delta_{n+m,0}$ is a coboundary.

c.) Show that the Lie algebra cohomology of the conformal group is one-dimensional and hence show that the above is the most general central extension.

13.4.1 Gravity in 1+1 dimensions

Consider quantum gravity in $1+1$ dimensions. If space is compact and connected it must be S^1, we are therefore interested in invariant states under $G = Diff^+(S^1)$. A natural basis for the corresponding (complexified) Lie algebra is given by $\ell_n = e^{in\theta}\frac{d}{d\theta}$, $n \in \mathbb{Z}$. Actually, because this is an infinite-dimensional group we should allow for anomalies and work with a central extension. The corresponding Lie algebra is the *Virasoro algebra* given above. Let us see how the central extension comes about physically.

Let us consider the BRST operator. We introduce (diffeomorphism) ghosts and antighosts c_n, b_n, $n \in \mathbb{Z}$ with $\{c_n, b_m\} = \delta_{n+m,0}$.

Because we are working with an infinite-dimensional Clifford algebra there are many inequivalent quantizations. The one relevant to radial quantization in the complex plane of string theory is [9]

$$\begin{aligned} c_n|0\rangle &= 0 \qquad n > 1 \\ b_n|0\rangle &= 0 \qquad n > -2 \end{aligned} \tag{13.15}$$

eq:ghostvac

The idea here is that the "Dirac sea" is filled by $c_2 \wedge c_3 \wedge c_4 \wedge \cdots$. For this reason, elements of the ghost Hilbert space are referred to as "semi-infinite differential forms" in the math literature.

Following the formula (8.19) above we have in this case:

$$Q_{\rm ghost} = -\frac{1}{2}\sum_{m,n\in\mathbb{Z}} (m-n) : c_{-m}c_{-n}b_{m+n} : \tag{13.16}$$

eq:virq

The normal ordering is necessary so that Q has a well-defined action on the ghost Hilbert space.

Now one finds a big surprise: $Q^2 \neq 0$ due to the need to regularize the infinite sums! Using ζ-function regularization one computes

$$Q^2_{\rm ghost} = \sum_{m,n} \frac{(m-13m^3)}{12}\delta_{m+n,0} c_m c_n \tag{13.17}$$

eq:virqiii

If we consider the Lie algebra cohomology in a nontrivial Virasoro module with central charge $\rho(c)$ then the BRST operator is:

$$Q = \sum c_n \rho(L_n) - \frac{1}{2}\sum_{m,n\in\mathbb{Z}} (m-n) : c_{-m}c_{-n}b_{m+n} : -a c_0 \tag{13.18}$$

eq:virqiv

where we have allowed for a normal-ordering shift $\rho(L_0) \to \rho(L_0) - a$. Then

$$Q^2 = \frac{1}{2}\sum_{m,n\in\mathbb{Z}} \left[[\rho(L_n), \rho(L_m)] - (n-m)\rho(L_{n+m}) + \frac{((1+12a)m - 13m^3)}{6}\delta_{m+n,0} \right] c_m c_n \tag{13.19}$$

eq:virqiiia

so we get $Q^2 = 0$ for $a = 1$ and $\rho(c) = 26$.

In the worldsheet approach to string theory we work with $1+1$ dimensional quantum gravity coupled to matter and this is one way of understanding the restriction to matter with $c = 26$. Only in this case is there a well-defined spectrum of string states from BRST cohomology.

[9] In general, a primary field of dimension Δ has an expansion $\Phi = \sum_n \Phi_n z^{-n-\Delta}$ and hence $\Phi_n|0\rangle = 0$ for $n > -\Delta$, so that $\lim_{z\to 0}\Phi(z)|0\rangle$ is well-defined. The ghost field $c(z)$ has $\Delta = -1$ and the antighost $b(z)$ has dimension $\Delta = +2$.

13.4.2 The exponential map for $Diff(S^1)$

• For $Diff(S^1)$ the Exp map is neither 1-1 nor onto, even in an arbitrarily small neighborhood of the origin!

• For $Vir_{\mathbb{C}}$ there is no corresponding Lie group.

For a discussion of (1) above see Pressley+Segal, Loop Groups, Sections 3.3.1 and the article by Milnor. Briefly, to show exp is not $1-1$ consider rotation $R_{2\pi/n}$. Consider the centralizer $Z(R_{2\pi/n})$. This consists of diffeomorhisms which are periodic in the sense that $\phi(\theta + 2\pi/n) = \phi(\theta) + 2\pi/n \mathrm{mod} 2\pi$. Now for any ϕ in the centralizer, $\phi SO(2)\phi^{-1}$ is a one parameter subgroup of $Diff(S^1)$ consisting of elements $\{\phi R_\theta \phi^{-1} | \theta \in \mathbb{R}\}$. However, $R_{2\pi/n}$ lies on all these one-parameter subgroups! Thus, the exponential map is not 1-1. Note that we can make n arbitrarily large, so this failure of injectivity holds arbitrarily close to the identity.

Regarding (2) consider the following diffeomorphism

$$f(\theta) = \theta + \pi/n + \epsilon \sin^2(n\theta) \tag{13.20}$$

where $0 < \epsilon < 1/n$. One can prove that this diffeomorphism has precisely one periodic orbit of length $2n$: $0 \to \pi/n \to 2\pi/n \to \cdots \to 2\pi(n-1)/n \to 0$. Indeed, if $0 < \theta_0 < \pi/n$ then defining $\theta_{j+1} = f(\theta_j)$ one can check with a little calculus that

$$\theta_0 < \theta_1 - \pi/n < \theta_2 - 2\pi/n < \cdots < \pi/n \tag{13.21}$$

so no θ_j is ever equal to θ_0 modulo 2π. Thus, our diffeomorphism has precisely one periodic orbit, and it is of length $2n$. It is remarkable that it has only one orbit with an even number of elements in the orbit. Why?

Consider a diffeomorphism F of *any* manifold. It is useful to compare the periodic orbits of F with those of $F \circ F$. If F has an odd periodic orbit then, $F \circ F$ has the same odd periodic orbit, essentially because the order of the orbit and 2 are relatively prime. (Draw the case of $n = 3, 5$.) On the other hand, if F has an *even* periodic orbit of order $n = 2m$, then the periodic orbits of $F \circ F$ split up into two distinct periodic orbits of length m. (Check the cases $n = 4, 6$.)

Now, we apply this remark to note that $f(\theta)$ above cannot be of the form $F \circ F$ for any diffeomorphism F, since the number of periodic orbits of even period must be even, but $f(\theta)$ has a single periodic orbit of even period. But this in turn means that $f(\theta)$ cannot be on any one-parameter subgroup $F_t(\theta)$ since if we had $F_{t_*}(\theta) = f(\theta)$ then we would have had $f = F \circ F$ with $F = F_{\frac{1}{2}t_*}$.

For a discussion of point (2) above see section 3.3.2 of Pressley-Segal.

♣Need to explain why. Helpful remarks in Graeme's email from November 19, 2009 on this. ♣

Preprint typeset in JHEP style - HYPER VERSION

Chapter 10:: Clifford algebras

Abstract: NOTE: THESE NOTES, FROM 2009, MOSTLY TREAT CLIFFORD ALGEBRAS AS UNGRADED ALGEBRAS OVER $\mathbb{R}$ OR $\mathbb{C}$. A CONCEPTUALLY SUPERIOR VIEWPOINT IS TO TREAT THEM AS $\mathbb{Z}_2$-GRADED ALGEBRAS. SEE REFERENCES IN THE INTRODUCTION WHERE THIS SUPERIOR VIEWPOINT IS PRESENTED.
April 3, 2018

-TOC- **Contents**

1. Introduction **1**

2. Clifford algebras **1**

3. The Clifford algebras over $\mathbb{R}$ **3**
3.1 The real Clifford algebras in low dimension 4
3.1.1 $C\ell(1_-)$ 4
3.1.2 $C\ell(1_+)$ 5
3.1.3 Two dimensions 6
3.2 Tensor products of Clifford algebras and periodicity 7
3.2.1 Special isomorphisms 9
3.2.2 The periodicity theorem 10

4. The Clifford algebras over $\mathbb{C}$ **14**

5. Representations of the Clifford algebras **15**
5.1 Representations and Periodicity: Relating Γ-matrices in consecutive even and odd dimensions 18

6. Comments on a connection to topology **20**

7. Free fermion Fock space **21**
7.1 Left regular representation of the Clifford algebra 21
7.2 The Exterior Algebra as a Clifford Module 22
7.3 Representations from maximal isotropic subspaces 22
7.4 Splitting using a complex structure 24
7.5 Explicit matrices and intertwiners in the Fock representations 26

8. Bogoliubov Transformations and the Choice of Clifford vacuum **27**

9. Comments on Infinite-Dimensional Clifford Algebras **28**

10. Properties of Γ-matrices under conjugation and transpose: Intertwiners **30**
10.1 Definitions of the intertwiners 31
10.2 The charge conjugation matrix for Lorentzian signature 31
10.3 General Intertwiners for $d = r + s$ even 32
10.3.1 Unitarity properties 32
10.3.2 General properties of the unitary intertwiners 33
10.3.3 Intertwiners for $d = r + s$ odd 35
10.4 Constructing Explicit Intertwiners from the Free Fermion Rep 35
10.5 Majorana and Symplectic-Majorana Constraints 36

10.5.1 Reality, or Majorana Conditions 36
10.5.2 Quaternionic, or Symplectic-Majorana Conditions 37
10.5.3 Chirality Conditions 38

11. $\mathbb{Z}_2$ graded algebras and modules 39
11.1 $\mathbb{Z}_2$ grading on the algebra 39
11.2 The even subalgebra $C\ell^0(r_+, r_-)$ 40
11.3 $\mathbb{Z}_2$ graded tensor product of Clifford algebras 41
11.4 $\mathbb{Z}_2$ graded modules 42
11.5 K-theory over a point 46
11.6 Graded tensor product of modules and the ring structure 46
11.6.1 The Grothendieck group 46

12. Clifford algebras and the division algebras 46

13. Some sources 47

---

1. Introduction

****

Note added April 3, 2018: For the reader willing to invest some time first learning some $\mathbb{Z}_2$-graded- , or super- linear algebra a much better treatment of this material can be found in:

http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf (Chapter 13)

http://www.physics.rutgers.edu/∼gmoore/PiTP-LecturesA.pdf (Section 2.3)

****

Clifford algebras and spinor representations of orthogonal groups naturally arise in theories with fermions. They also play an important role in supersymmetry and supergravity. They are used quite frequently in the various connections of physics to geometry. It also turns out that they are central to modern topology.

While the details that follow can get rather intricate, it is worth the investment it takes to learn them.

It is important to have a clear understanding about the meaning of a complex structure and a quaternionic structure on a real vector space, what realification means etc. See chapter two for these linear algebra basics.

2. Clifford algebras

Definition. Let V be a vector space over a field F and let Q be a quadratic form on V valued in F. The *Clifford algebra* $C\ell(Q)$ is the algebra over F generated by V and defined by the relations:

$$v_1 v_2 + v_2 v_1 = 2Q(v_1, v_2) \cdot 1 \tag{2.1}$$

where 1 is the unit, considered to be the multiplicative unit in the ground field F.

Remarks

• One can give a much more formal definition by taking the quotient of the tensor algebra $T(V)$ by the 2-sided ideal generated by $v_1 \otimes v_2 + v_2 \otimes v_1 - 2Q(v_1, v_2)1$. Unlike the tensor algebra the Clifford algebra is *not* $\mathbb{Z}$-graded, since two vectors can multiply to a scalar. Nevertheless it is $\mathbb{Z}_2$-graded, and this $\mathbb{Z}_2$-grading is important. We can define an algebra automorphism λ on $C\ell(Q)$ by taking $\lambda(v) = -v$ for $v \in V$ and extending this to be an algebra automorphism. The even and odd parts of the $\mathbb{Z}_2$ grading are the $\lambda = \pm 1$ subspaces. We will use this grading later in this chapter and in the next.

• Since the relation is quadratic the embedding $V \to C\ell(Q)$ has no kernel. To lighten the notation we will usually not distinguish between a vector $v \in V$ and its image in $C\ell(Q)$, trusting to context to resolve the ambiguity.

• If $Q = 0$ the resulting algebra is called a *Grassmann algebra* and is isomorphic to the exterior algebra on V. (See more on this below). Usually, when people speak of a "Clifford algebra" it is understood that Q is nondegenerate.

To write a vector space basis of $C\ell(Q)$ let $\{e^\mu | \mu = 1, \dots d\}$ be a basis of V. Then the e^μ generate the algebra. The quadratic form in this basis will be $Q^{\mu\nu}$. Since it is nondegenerate we can define an inverse $Q_{\mu\nu}$ so that $Q_{\mu\nu}Q^{\nu\lambda} = \delta_\mu{}^\lambda$. We will also use the notation $e_\mu = Q_{\mu\nu}e^\nu$.

For simplicity, let us choose a basis in which Q is diagonal [1]

When we multiply $e_{\mu_1} \cdots e_{\mu_p}$ we can use the relations to move vectors with the same index next to each other and then "annihilate them" (i.e., use the relation) to get scalars. Define

$$e^{\mu_1 \cdots \mu_p} := e^{\mu_1} \cdots e^{\mu_p} \tag{2.2}$$

eq:multiindx

when the indices are all different. Note that this expression is completely antisymmetric on $\mu_1, \dots, \mu_p$.

Clearly the $e^{\mu_1 \cdots \mu_p}$ for $\mu_1 < \mu_2 < \cdots < \mu_p$ are all linearly independent and moreover form a vector space basis over F for $C\ell(Q)$. In particular we find

$$\dim_F C\ell(Q) = 2^d = \sum_{p=0}^{d} \binom{d}{p} \tag{2.3}$$

[1] We can also speak of Clifford modules over a ring. In this case we might not be able to diagonalize Q. This might happen, for example, if we considered Clifford rings over $\mathbb{Z}$.

Note that using this basis we can define an isomorphism of $C\ell(Q)$ with $\Lambda^*(V)$ as *vector spaces*:

$$\frac{1}{p!}\omega_{\mu_1\ldots\mu_p}e^{\mu_1\cdots\mu_p} \to \frac{1}{p!}\omega_{\mu_1\ldots\mu_p}e^{\mu_1}\wedge\cdots\wedge e^{\mu_p} \tag{2.4}$$ eq:isomf

where $\omega_{\mu_1\ldots\mu_p}$ is a totally antisymmetric tensor.

Even though $\Lambda^*(V)$ is an algebra, (2.4) does *not* define an algebra isomorphism since $(e^\mu)^2 \neq 0$ in the Clifford algebra (when Q is diagonal) while $e^\mu \wedge e^\mu = 0$ in $\Lambda^*(V)$. Indeed, $\Lambda^*(V)$ is isomorphic to the Grassmann algebra, as an algebra.

Remarks

• An important anti-automorphism, the *transpose* is defined as follows: $1^t = 1$ and $v^t = v$ for $v \in V$. Now we extend this to be an anti-automorphism so that $(\phi_1\phi_2)^t = \phi_2^t\phi_1^t$. In particular:

$$(e^{\mu_1}e^{\mu_2}\cdots e^{\mu_k})^t = e^{\mu_k}e^{\mu_{k-1}}\cdots e^{\mu_2}e^{\mu_1} \tag{2.5}$$

A little computation shows that

$$(e^{\mu_1\cdots\mu_k})^t = e^{\mu_k\cdots\mu_1} = (-1)^{\frac{1}{2}k(k-1)}e^{\mu_1\cdots\mu_k} \tag{2.6}$$

• The functions $f(k) = (-1)^{\frac{1}{2}k(k-1)}$ and $g(k) = (-1)^{\frac{1}{2}k(k+1)}$ appear frequently in the following. Note that $f(k)$ and $g(k)$ only depend on $k\mathrm{mod}4$, $f(k) = g(-k)$ and

$$(-1)^{\frac{1}{2}k(k-1)} = \begin{cases} +1 & k = 0, 1\mathrm{mod}4 \\ -1 & k = 2, 3\mathrm{mod}4 \end{cases} \tag{2.7}$$ eq:qnfen

$$(-1)^{\frac{1}{2}k(k+1)} = \begin{cases} +1 & k = 0, 3\mathrm{mod}4 \\ -1 & k = 1, 2\mathrm{mod}4 \end{cases} \tag{2.8}$$ eq:qnfena

3. The Clifford algebras over $\mathbb{R}$

In quantum mechanics we work over the complex numbers. Nevertheless, in theories of fermions it is often important to take into account reality constraints, and hence the properties of the Clifford algebras over $\mathbb{R}$ is quite relevant to physics, particularly when we discuss supersymmetric field theories and string theories in various spacetime dimensions. Moreover, in the physics-geometry interaction the beautiful structure of the Clifford algebras over $\mathbb{R}$ plays an important role. It is thus well worth the extra effort to understand the structures here.

By Sylvester's theorem, when working over $\mathbb{R}$ it suffices to consider

$$Q(e^\mu, e^\nu) = \eta^{\mu\nu} \tag{3.1}$$

where $\eta^{\mu\nu}$ is diagonal with ± 1 entries. Since there are different conventions in the literature, we will denote the Clifford algebra with

$$\begin{aligned} (e^1)^2 = \cdots = (e^{r_+})^2 = +1 \\ (e^{r_++1})^2 = \cdots = (e^{r_++s_-})^2 = -1 \end{aligned} \tag{3.2}$$

by $C\ell(r_+, s_-)$. The Clifford algebra is thus specified by a pair of nonnegative integers labelling the number of $+1$ and -1 eigenvalues. If the form is definite we simply denote $C\ell(r_+)$ or $C\ell(s_-)$. [2]

Exercise

Let p_μ be a vector on the pseudo-sphere

$$p_\mu p_\nu \eta^{\mu\nu} = R^2. \tag{3.3}$$

Show that $(p_\mu e^\mu)^2 = R^2 \cdot 1$.

Exercise *The Clifford volume element*

a.) There is a basic anti-automorphism, the transpose automorphism that acts as $(\phi_1\phi_2)^t = \phi_2^t\phi_1^t$. Show that:

$$e^{\mu_1\cdots\mu_k} = (-1)^{\frac{1}{2}k(k-1)} e^{\mu_k\cdots\mu_1} \tag{3.4}$$

b.) Show that the volume element in $C\ell(r_+, s_-)$

$$\omega = e^1 e^2 \cdots e^d \tag{3.5}$$

$d = r_+ + s_-$, satisfies

$$\omega^2 = (-1)^{\frac{1}{2}(s_- - r_+)(s_- - r_+ + 1)} = \begin{cases} +1 & for (s_- - r_+) = 0, 3 \mathrm{mod} 4 \\ -1 & for (s_- - r_+) = 1, 2 \mathrm{mod} 4 \end{cases} \tag{3.6}$$

[Answer: The easiest way to compute is to write

$$\omega \cdot \omega = (-1)^{\frac{1}{2}d(d-1)} \omega\omega^t = (-1)^{\frac{1}{2}d(d-1)+s} = (-1)^{\frac{1}{2}(s-r)(s-r+1)}] \tag{3.7}$$

c.) Show that under a change of basis $e^\mu \to f^\mu = \sum g^{\nu\mu} e^\nu$ where $g \in O(\eta)$ we have $\omega' = \det g \omega$, so that ω indeed transforms as the volume element.

d.) $\omega e^\mu = (-1)^{d+1} e^\mu \omega$. Thus ω is central for d odd and is not central for d even.

Note:

1. $d_T = s_- - r_+$ generalizes the number of dimensions *transverse* to the light cone in Lorentzian geometry.

2. $\omega^2 = 1$ and ω is central only for $d_T = s - r = 3 \mathrm{mod} 4$.

---

[2] I find it impossible to remember whether the first or second entry should be the number of positive or negative signs. Therefore, I will explicitly put a subscript $\pm$ to indicate the signature. Lawson-Michelsohn take $C\ell(r, s) = C\ell(r_-, s_+)$.

3.1 The real Clifford algebras in low dimension

In 0 dimensions $C\ell(0) = \mathbb{R}$.

3.1.1 $C\ell(1_-)$

To see this note that the general element in $C\ell(1_-)$ is $a+be^1$. But we have a faithful matrix representation $e^1 \to \sqrt{-1}$. (Note that in fact we have *two* inequivalent representations, depending on the sign of the squareroot). Thus $a + be^1 \to a + ib$ defines an isomorphism to $\mathbb{C}$, regarded as an algebra over $\mathbb{R}$. Thus

$$C\ell(1_-) \cong \mathbb{C} \tag{3.8}$$

There is only one real irreducible representation up to equivalence: $V \cong \mathbb{R}^2$ and

$$\rho(a + be^1) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{3.9}$$

Note that

$$\rho'(a + be^1) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{3.10}$$

defines an equivalent representation.

Over $\mathbb{C}$ these matrices can be diagonalized, leading to *two* inequivalent complex representations of $\mathbb{C}\ell(1)$. Applying the rule for the complexification of a real vector space with complex structure we have $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \oplus \overline{\mathbb{C}}$ so the inequivalent representations are:

$$\rho_+(a + be^1) = a + bi \tag{3.11}$$ eq:inqa

and

$$\rho_-(a + be^1) = a - bi \tag{3.12}$$ eq:inqb

3.1.2 $C\ell(1_+)$

The multiplication law of elements $a + be^1$ is simply

$$(a + be^1) \cdot (a' + b'e^1) = (aa' + bb') + (ab' + a'b)e^1 \tag{3.13}$$

In this dimension we can introduce central projection operators

$$P_\pm = \frac{1}{2}(1 \pm e^1) \tag{3.14}$$

Then $P_+C\ell(1_+)$ and $P_-C\ell(1_+)$ are subalgebras and we can write a direct sum of algebras:

$$C\ell(1_+) = P_+C\ell(1_+) \oplus P_-C\ell(1_+) \tag{3.15}$$

In this case each of the subalgebras is one-dimensional:

$$a + be^1 = (a + b)(\frac{1 + e^1}{2}) + (a - b)(\frac{1 - e^1}{2}) \tag{3.16}$$

We have

$$C\ell(1_+) = \mathbb{R} \oplus \mathbb{R} \tag{3.17}$$

This algebra is also known as the "double numbers."

There are two inequivalent real representations:

$$\rho_+(a + be^1) = a + b \tag{3.18}$$

$$\rho_-(a + be^1) = a - b \tag{3.19}$$

These representations are not faithful. We have a faithful matrix rep:

$$a + be^1 \to \begin{pmatrix} a & b \\ b & a \end{pmatrix} \tag{3.20}$$ eq:doublnumone

Of course the representation (3.20) is in fact reducible. it is equivalent to matrices of the form

$$\begin{pmatrix} a+b & 0 \\ 0 & a-b \end{pmatrix} \tag{3.21}$$ eq:doublenumi

However, one needs both diagonal entries to get a faithful representation. Later we will talk about $\mathbb{Z}_2$ graded representations. The minimal $\mathbb{Z}_2$ graded representation is 2-dimensional and given by (3.20).

Finally, over $\mathbb{C}$, i.e. for $\mathbb{C}\ell_1$ we could also have taken $(e^1)^2 = +1$ and

$$\rho_+(a + be^1) = a + b \tag{3.22}$$ eq:inqc

and

$$\rho_-(a + be^1) = a - b \tag{3.23}$$ eq:inqd

with $a, b \in \mathbb{C}$ reflecting the complexification of the two representations of $C\ell(1_+)$. Thus

$$\mathbb{C}\ell(1) \cong \mathbb{C} \oplus \mathbb{C} \tag{3.24}$$ eq:cclxone

3.1.3 Two dimensions

Let us introduce the very useful notation

$$\mathbb{K}(n) := Mat_{n\times n}(\mathbb{K}) \tag{3.25}$$ eq:kayenn

where $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$. This is an algebra over $\mathbb{R}$ of real dimension $n^2, 2n^2, 4n^2$, respectively.

In two dimensions we have

$$\begin{aligned} C\ell(2_+) &= \mathbb{R}(2) \\ C\ell(1_+, 1_-) &= \mathbb{R}(2) \\ C\ell(2_-) &= \mathbb{H} \end{aligned} \tag{3.26}$$ eq:tdclff

To see this consider first $C\ell(2_+)$: Give a faithful matrix rep:

$$e^1 \to \sigma_1 \qquad e^2 \to \sigma_3 \tag{3.27}$$

then

$$e^1 e^2 = -i\sigma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{3.28}$$

Now we can write an arbitrary 2×2 real matrix as a linear combination of $1, \sigma_1, \sigma_3, -i\sigma_2$:

$$\rho(a + be^1 + ce^2 + de^1 e^2) = \begin{pmatrix} a+c & b-d \\ b+d & a-c \end{pmatrix} \tag{3.29}$$

Next for $C\ell(2-)$. Map to the imaginary unit quaternions:

$$\begin{aligned} e^1 &\to \mathbf{i} \\ e^2 &\to \mathbf{j} \\ e^1 e^2 &\to \mathbf{k} \end{aligned} \tag{3.30}$$

Finally $C\ell(1,1)$. Again we can provide a faithful matrix rep:

$$e^1 \to \sigma_1 \qquad e^2 \to i\sigma_2 \tag{3.31}$$

thus

$$\rho(a + be^1 + ce^2 + de^1 e^2) = \begin{pmatrix} a+d & b+c \\ b-c & a-d \end{pmatrix} \tag{3.32}$$

and the algebra is that of $\mathbb{R}(2)$. ♠

Remarks

• When we complexify there is no distinction between the signatures. Any of the above three algebras can be used to show that

$$\mathbb{C}\ell(2) \cong \mathbb{C}(2) \tag{3.33}$$ eq:ccxclff

• The representation matrices are always denoted as Γ matrices in the physics literature. Thus, for example, what we are saying above is that in $1+1$ dimensions we could choose an irreducible real representation

$$\Gamma^0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \Gamma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.34}$$

Exercise

We have now obtained two algebra structures on the vector space $\mathbb{R}^4$: $\mathbb{R}(2)$ and $\mathbb{H}$. Are they isomorphic? (Hint: Is $\mathbb{R}(2)$ a division algebra?)

3.2 Tensor products of Clifford algebras and periodicity

We will now examine how the Clifford algebras in different dimensions are related to each other. This will enable us to express the Clifford algebra in terms of matrix algebras for all (r_+, s_-). These relations are also very useful in physics in dimensional reduction.

There are two kinds of tensor products one could define, the graded and ungraded tensor product. For now we will focus on the *ungraded* tensor product. [3] The tensor product is then the standard tensor product of vector spaces. We define a Clifford multiplication on the tensor product by the rule:

$$(\phi_1 \otimes \psi_1) \cdot (\phi_2 \otimes \psi_2) := \phi_1 \cdot \phi_2 \otimes \psi_1 \cdot \psi_2 \tag{3.35}$$

This is the standard tensor product on two algebras. In section *** below we will discuss the *graded* tensor product which differs by some important sign conventions.

Lemma:

• $C\ell(r_+, s_-) \otimes C\ell(2_+) \cong C\ell((s+2)_+, r_-)$

• $C\ell(r_+, s_-) \otimes C\ell(1_+, 1_-) \cong C\ell((r+1)_+, (s+1)_-)$

• $C\ell(r_+, s_-) \otimes C\ell(2_-) \cong C\ell(s_+, (r+2)_-)$

Proofs:

• Let e^μ be generators of $C\ell(r_+, s_-)$, f^α, $\alpha = 1, 2$ be generators of $C\ell(2+)$. Note that the obvious set of generators $e^\mu \otimes 1$ and $1 \otimes f_\alpha$, do not satisfy the relations of the Clifford algebra, because they do not anticommute. On the other hand if we take

$$\tilde{e}_\mu := e_\mu \otimes f_{12} \qquad \tilde{e}_{d+\alpha} := 1 \otimes f_\alpha \tag{3.36}$$

where $f_{12} = f_1 f_2$, then $\tilde{e}_M$, $M = 1 \ldots, d+2$ satisfy the Clifford algebra relations and also generate the tensor product. Now note that $(f_{12})^2 = -1$ and hence:

$$(e_\mu \otimes f_{12})^2 = -(e_\mu)^2 \tag{3.37}$$

(no sum on μ). The same proof works for item 3 above.

• Once again we can take generators as above, now we need only notice that in signature $(1_+, 1_-)$ we have $(f_{12})^2 = +1$ and hence:

$$(e_\mu \otimes f_{12})^2 = +(e_\mu)^2 \tag{3.38}$$

(no sum on μ). ♠

Remarks These isomorphisms, and the consequences below are very useful because they relate Clifford algebras and spinors in different dimensions. Notice in particular, item 2, which relates the Clifford algebra in a spacetime to that on the transverse space to the lightcone.

---

[3] In section *** below we discuss the graded tensor product.

Exercise
Show that $C\ell((s+1)_+, r_-) \cong C\ell((r+1)_+, s_-)$.

Exercise

Show $C\ell(r_+, s_-) \cong \mathbb{R}(2^r)$ when $r = s$. This is always a matrix algebra over the reals. Further understanding of why this is so comes from the model for Clifford algebras in terms of contraction and wedge product of differential forms (and free fermions) described below.

In general, two algebras related by $A \cong B \otimes \mathrm{Mat}_n(\mathbb{R})$ for some n are said to be *Morita equivalent*.

3.2.1 Special isomorphisms

For $K = \mathbb{R}, \mathbb{C}, \mathbb{H}$ let $K(n)$ denote the algebra over $\mathbb{R}$ of all $n \times n$ matrices with entries in K. We have the following special isomorphisms of matrix algebras over $\mathbb{R}, \mathbb{C}, \mathbb{H}$:

$$\mathbb{R}(n) \otimes_{\mathbb{R}} \mathbb{R}(m) \cong \mathbb{R}(nm) \tag{3.39}$$ eq:specisoma

$$\mathbb{R}(n) \otimes_{\mathbb{R}} K \cong K(n) \tag{3.40}$$ eq:specisomb

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \oplus \mathbb{C} \tag{3.41}$$ eq:specisomc

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathbb{C}(2) \tag{3.42}$$ eq:specisomd

$$\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathbb{R}(4) \tag{3.43}$$ eq:specisome

To prove (3.41) note that the tensor product is generated (*over* $\mathbb{R}$) by $1 \otimes 1$, $1 \otimes i$, $i \otimes 1$, $i \otimes i$. Now we have an explicit isomorphism

$$\begin{aligned}
(1,0) &\to \frac{1}{2}(1 \otimes 1 + i \otimes i) \\
(i,0) &\to \frac{1}{2}(i \otimes 1 - 1 \otimes i) \\
(0,1) &\to \frac{1}{2}(1 \otimes 1 - i \otimes i) \\
(0,i) &\to \frac{1}{2}(i \otimes 1 + 1 \otimes i)
\end{aligned} \tag{3.44}$$

This should be compared with the isomorphism

$$\mathbb{C} \otimes_{\mathbb{C}} \mathbb{C} \cong \mathbb{C} \tag{3.45}$$

The isomorphism (3.42) follows from the familiar representation of quaternions in terms of complex 2×2 matrices that gives us the quaternions as $x_\mu \tau_\mu$. If we now take x_μ to be complex then we get all 2×2 complex matrices.

For isomorphism (3.43) we identify $\mathbb{H} \cong \mathbb{R}^4$ and note that to an $q_1 \otimes q_2 \in \mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}$ we can associate a linear map $\mathbb{R}^4 \to \mathbb{R}^4$ given by

$$x \to q_1 x \bar{q}_2 \tag{3.46}$$

Extending by linearity this defines an algebra homomorphism $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \to \mathrm{End}(\mathbb{R}^4) = \mathbb{R}(4)$. We claim the map is an isomorphism. To see this let us try to compute the kernel. This would be an element $\sum_{\mu\nu} a_{\mu\nu} \tau_\mu \otimes \bar{\tau}_\nu \in \mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}$ so that for all $x \in \mathbb{H}$

$$\sum_{\mu\nu} a_{\mu\nu} \tau_\mu x \bar{\tau}_\nu = 0 \tag{3.47}$$ eq:allsnx

By conjugation, if $a_{\mu\nu}$ satisfies (3.47) so does its transpose, so we can separate the equations into $a_{\mu\nu}$ symmetric and antisymmetric. Now note that the equation is $SO(4) \times SO(4)$ covariant, if (3.47) is satisfied then for any four unit quaternions q_1, q_2, p_1, p_2 we have

$$\sum_{\mu\nu} a_{\mu\nu} q_1 \tau_\mu \bar{q}_2 x p_1 \bar{\tau}_\nu \bar{p}_2 = 0 \tag{3.48}$$ eq:allsnxi

and hence if $a_{\mu\nu}$ satisfies (3.47) so does $R_1 a R_2$ where $R_1, R_2 \in SO(4)$. Thus for the symmetric case we can diagonalize a so that (3.47) becomes

$$\sum_{\mu} \lambda_\mu \tau_\mu x \bar{\tau}_\mu = 0 \qquad\qquad \forall x \in \mathbb{H} \tag{3.49}$$

Substituting $x = 1, \mathbf{i}, \mathbf{j}, \mathbf{k}$ gives four linear equations which easily imply $\lambda_\mu = 0$. Similarly if $a_{\mu\nu}$ is anti-symmetric it can be skew-diagonalized and then it is easy to show that the skew eigenvalues vanish. Thus the kernel of the map $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \to \mathrm{End}(\mathbb{R}^4) = \mathbb{R}(4)$ is zero and since both domain and range have dimension 16 the map is an isomorphism.

Exercise
Show that

$$C\ell(3_+) = C\ell(1_-) \otimes C\ell(2_+) \cong \mathbb{C}(2) \tag{3.50}$$ eq:clthrp

$$C\ell(3-) = C\ell(1_+) \otimes C\ell(2_-) \cong \mathbb{H} \oplus \mathbb{H} \tag{3.51}$$ eq:clthrm

---

3.2.2 The periodicity theorem

Now we combine the above isomorphisms to produce some useful relations between the Clifford algebras:

On the one hand we can say

$$C\ell(r_+,s_-)\otimes C\ell(2_+)\otimes C\ell(2_-)\cong C\ell((s+2)_+,r_-)\otimes C\ell(2_-)\cong C\ell(r_+,(s+4)_-) \tag{3.52}$$ eq:pera

On the other hand we can say

$$C\ell(r_+,s_-)\otimes C\ell(2_-)\otimes C\ell(2_+)\cong C\ell(s_+,(r+2)_-)\otimes C\ell(2_+)\cong C\ell((r+4)_+,s_-) \tag{3.53}$$ eq:perb

Finally we note that

$$C\ell(2_+)\otimes C\ell(2_-)=\mathbb{R}(2)\otimes\mathbb{H}\cong\mathbb{H}(2) \tag{3.54}$$ eq:perc

Summarizing:

$$C\ell((r+4)_+,s_-)\cong C\ell(r_+,(s+4)_-)\cong C\ell(r_+,s_-)\otimes\mathbb{H}(2) \tag{3.55}$$ eq:perd

Thus, if we wish to understand the structure of $C\ell(r_+,s_-)$ we can use tensor product with $C\ell(1_+,1_-)$ to reduce the algebra to the (Morita equivalent) form $C\ell(n_+)$ or $C\ell(n_-)$, depending on whether $r-s\geq 0$ or $r-s\leq 0$, respectively. Then we can use (3.55) to bring $0\leq n_\pm\leq 3$. On the other hand, we have explicitly determined the algebras in this range using the above computations. In this way we can list the full set of algebras.

The mod-four periodicity under $\otimes_{\mathbb{R}}\mathbb{H}(2)$ can be iterated to produce a basic mod-eight periodicity which is a little easier to think about. Note that since $\mathbb{H}(2)\otimes\mathbb{H}(2)\cong\mathbb{H}\otimes\mathbb{H}\otimes\mathbb{R}(4)\cong\mathbb{R}(16)$ the identities (3.55) immediately imply:

$$\boxed{\begin{array}{l} C\ell(r_+,(s+8)_-)\cong C\ell(r_+,s_-)\otimes\mathbb{R}(16)\\ C\ell((r+8)_+,s_-)\cong C\ell(r_+,s_-)\otimes\mathbb{R}(16)\end{array}} \tag{3.56}$$

Thus, combining the computation of the first four algebras with (3.55) we produce the following table:

$r=$	1	2	3	4	5	6	7	8
$C\ell(r_+)$	$\mathbb{R}\oplus\mathbb{R}$	$\mathbb{R}(2)$	$\mathbb{C}(2)$	$\mathbb{H}(2)$	$\mathbb{H}(2)\oplus\mathbb{H}(2)$	$\mathbb{H}(4)$	$\mathbb{C}(8)$	$\mathbb{R}(16)$
$C\ell(r_-)$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{H}\oplus\mathbb{H}$	$\mathbb{H}(2)$	$\mathbb{C}(4)$	$\mathbb{R}(8)$	$\mathbb{R}(8)\oplus\mathbb{R}(8)$	$\mathbb{R}(16)$

Thus, implementing the above procedure we have the following result:

1. If $r \geq s$ and $r - s = \alpha + 8k$, $0 \leq \alpha \leq 7, k \geq 0$, then

$$C\ell(r_+, s_-) \cong C\ell(\alpha_+) \otimes \mathbb{R}(2^{\frac{1}{2}(d-\alpha)}) \tag{3.57}$$ eq:perioa

2. If $s \geq r$, and $s - r = \beta + 8k$, $0 \leq \beta \leq 7, k \geq 0$, then

$$C\ell(r_+, s_-) \cong C\ell(\beta_-) \otimes \mathbb{R}(2^{\frac{1}{2}(d-\beta)}) \tag{3.58}$$ eq:periob

(Note that $d - \alpha$ and $d - \beta$ are both even.)

Now, we can unify the two results (3.57) and (3.58) by defining the *type* of the Clifford algebra to be $\mathbb{R}, \mathbb{R} \oplus \mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{H} \oplus \mathbb{H}$. That is, these are the basic Morita equivalence classes which appear.

Define the "transverse dimension" to be

$$d_T := s - r \tag{3.59}$$ eq:transversed

for $C\ell(r_+, s_-)$. Observe from the above table that the Morita equivalence class *only depends on* $d_T \mathrm{mod} 8$! For example $C\ell(1_+)$ with $d_T = -1$ is of the same type as $C\ell(7_-)$ with $d_T = +7$.

Figure 1: The Clifford stop sign. fig:clffstp

We can now list the real Clifford algebras $C\ell(r_+, s_-)$ in a unified way. The type of the algebra follows the basic pattern for $d_T = 0, 1, 2, \ldots$ of

$$\mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{H} \oplus \mathbb{H}, \mathbb{H}, \mathbb{C}, \mathbb{R}, \mathbb{R} \oplus \mathbb{R}, \mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{H} \oplus \mathbb{H}, \mathbb{H}, \ldots \tag{3.60}$$

and can be visualized as in 1. The first three entries are easy to remember since you are adding squareroots of -1. If you remember the general pattern then you can easily reconstruct the type. Next you can construct the algebra since the total dimension over $\mathbb{R}$ is just 2^d where $d = r + s$ is the dimension.

To summarize. Define $\ell \equiv [d/2]$. Then:

$(s-r)\mathrm{mod}8$	$C\ell(r_+, s_-)$
0	$\mathbb{R}(2^\ell)$
1	$\mathbb{C}(2^\ell)$
2	$\mathbb{H}(2^{\ell-1})$
3	$\mathbb{H}(2^{\ell-1}) \oplus \mathbb{H}(2^{\ell-1})$
4	$\mathbb{H}(2^{\ell-1})$
5	$\mathbb{C}(2^\ell)$
6	$\mathbb{R}(2^\ell)$
7	$\mathbb{R}(2^\ell) \oplus \mathbb{R}(2^\ell)$
8	$\mathbb{R}(2^\ell)$

(3.61) eq:cliffalgs

Remarks

- A useful table of the Lorentzian signature Clifford algebras is

$d = s+1$	$C\ell(s_+, 1_-)$	$C\ell(1_+, s_-)$
0+1	$\mathbb{C}$	$\mathbb{R} \oplus \mathbb{R}$
1 +1	$\mathbb{R}(2)$	$\mathbb{R}(2)$
2 +1	$\mathbb{R}(2) \oplus \mathbb{R}(2)$	$\mathbb{C}(2)$
3 +1	$\mathbb{R}(4)$	$\mathbb{H}(2)$
4 +1	$\mathbb{C}(4)$	$\mathbb{H}(2) \oplus \mathbb{H}(2)$
5 +1	$\mathbb{H}(4)$	$\mathbb{H}(4)$
6 +1	$\mathbb{H}(4) \oplus \mathbb{H}(4)$	$\mathbb{C}(8)$
7 +1	$\mathbb{H}(8)$	$\mathbb{R}(16)$
8 +1	$\mathbb{C}(16)$	$\mathbb{R}(16) \oplus \mathbb{R}(16)$
9 +1	$\mathbb{R}(32)$	$\mathbb{R}(32)$
10 +1	$\mathbb{R}(32) \oplus \mathbb{R}(32)$	$\mathbb{C}(32)$
11 +1	$\mathbb{R}(64)$	$\mathbb{H}(32)$

•

Note in particular:

$$\begin{aligned} C\ell(1_+,3_-) &= \mathbb{H}(2) \qquad C\ell(3_+,1_-) = \mathbb{R}(4) \\ C\ell(0,4_-) &= \mathbb{H}(2) \qquad C\ell(4_+,0) = \mathbb{H}(2) \end{aligned} \tag{3.62}$$

Exercise

It appears that the convention of whether we assign time to directions of positive or negative norm leads to different Clifford algebras.

Why is this convention irrelevant when working with fermions in physics?

Exercise *Clifford algebras appearing in supergravity theories*

Verify the following isomorphisms:

$$\begin{aligned} C\ell(1_+,3_-) &= \mathbb{H}(2) \\ C\ell(0,4) &= \mathbb{H}(2) \\ C\ell(1,4) &= \mathbb{H}(2)\oplus\mathbb{H}(2) \\ C\ell(1_-,5_+) &= \mathbb{H}(2)\otimes\mathbb{R}(2)\cong\mathbb{H}(4) \\ C\ell(1_+,5_-) &= \mathbb{H}(2) \\ C\ell(0,6) &= \mathbb{R}(8) \\ C\ell(1_+,9_-) &= \mathbb{R}(32) \\ C\ell(9_+,1_-) &= \mathbb{R}(32) \\ C\ell(1_+,10_-) &= \mathbb{C}(16) \\ C\ell(10_+,1_-) &= \mathbb{R}(32)\oplus\mathbb{R}(32) \\ C\ell(11_+) &= \mathbb{C}(32) \\ C\ell(11_-) &= \mathbb{H}(16)\oplus\mathbb{H}(16) \end{aligned} \tag{3.63}$$

eq:clff

Exercise

Prove the isomorphisms:

$$C\ell(r_+,s_-)\cong C\ell((r+4)_+,(s-4)_-) \tag{3.64}$$

$$C\ell((r+1)_+,s_-)\cong C\ell((s+1)_+,r_-) \tag{3.65}$$

---

4. The Clifford algebras over $\mathbb{C}$

When we change the ground field from $\mathbb{R}$ to $\mathbb{C}$ the structure of the Clifford algebras simplifies considerably.

Now there is no need to take account of the signature. We can take $e_i e_j + e_j e_i = 2\delta_{ij}$. Or we could instead take $e_i e_j + e_j e_i = -2\delta_{ij}$. Both conventions lead to equivalent results. Since we are working over the complex numbers we could also send $e_i \to \sqrt{-1} e_i$.

We have already shown that in low dimensions:

$$\begin{aligned} \mathbb{C}\ell(0) &= \mathbb{C} \\ \mathbb{C}\ell(1) &\cong \mathbb{C} \oplus \mathbb{C} \\ \mathbb{C}\ell(2) &\cong \mathbb{C}(2) \end{aligned} \tag{4.1}$$

Next, the periodicity is simplified considerably. We now have:

$$\mathbb{C}\ell(d+2) \cong \mathbb{C}\ell(d) \otimes_{\mathbb{C}} \mathbb{C}\ell(2) \cong \mathbb{C}\ell(d) \otimes_{\mathbb{C}} \mathbb{C}(2) \tag{4.2}$$

from which we obtain the basic result:

$d = 0 \mathrm{mod} 2$:

$$\mathbb{C}\ell(d) = \mathbb{C}(2^{[d/2]}) = \mathbb{C}(2^{d/2}) \tag{4.3}$$ eq:cplxev

$d = 1 \mathrm{mod} 2$:

$$\mathbb{C}\ell(d) = \mathbb{C}(2^{[d/2]}) \oplus \mathbb{C}(2^{[d/2]}) = \mathbb{C}(2^{(d-1)/2}) \oplus \mathbb{C}(2^{(d-1)/2}) \tag{4.4}$$ eq:cplxodd

Recall that we defined the *Clifford volume form* to be $\omega := e_1 e_2 \cdots e_d$, and showed moreover that

$$\omega^2 = (-1)^{\frac{1}{2}d(d-1)} = \begin{cases} +1 & d = 0, 1 \mathrm{mod} 4 \\ -1 & d = 2, 3 \mathrm{mod} 4 \end{cases} \tag{4.5}$$

(If we had taken $e_i e_j + e_j e_i = -2\delta_{ij}$ then we would have gotten $(-1)^{\frac{1}{2}d(d+1)}$.) When working over $\mathbb{C}$ we can always find a scalar ξ so that $\omega_c := \xi\omega$ satisfies $(\omega_c)^2 = 1$. Explicitly

$$\omega_c = \begin{cases} \omega & d = 0, 1 \mathrm{mod}\ 4 \\ i\omega & d = 2, 3 \mathrm{mod}\ 4 \end{cases} \tag{4.6}$$

Then we can form *projection operators*

$$P_{\pm} = \frac{1}{2}(1 \pm \omega_c) \tag{4.7}$$

When d is odd these projection operators are central - they commute with everything in the algebra - and this gives a decomposition of the algebra into a direct sum of subalgebras as in (4.4). That is elements of the form $P_+ \phi$ form a sub*algebra*, not just a vector sub*space*, *because* P_+ is central. This subalgebra is isomorphic to $\mathbb{C}\ell(2^{[d/2]})$ and similarly for the subalgebra of elements of the form $P_- \phi$.

5. Representations of the Clifford algebras

Let us now consider representations of the Clifford algebras. This assigns $\phi \in C\ell \to \rho(\phi)$, where $\rho(\phi)$ is a linear transformation such that $\rho(\phi_1)\rho(\phi_2) = \rho(\phi_1\phi_2)$. The matrix representations of the generators $\rho(e^\mu) = \Gamma^\mu$ are called Γ-matrices in the physics literature.

We have seen how to write all the Clifford algebras in terms of matrix algebras over the division algebras. That is, they are all of the form $\mathbb{K}(n)$ or $\mathbb{K}(n) \oplus \mathbb{K}(n)$.

Now we only need a standard result of algebra which says that $\mathbb{K}(n)$ is a simple algebra (it has no nontrivial two-sided ideals) and hence has a unique representation (up to isomorphism) when $\mathbb{K}$ is a division algebra. To justify this recall

****

FIX FOLLOWING: DID WE NOT COVER THIS IN AN EARLIER CHAPTER WITH ALGEBRAS?

****

If V is an irreducible representation then we have a homomorphism $A \to \mathrm{End}(V)$. Since V is irreducible the image cannot commute with any nontrivial projection operators and hence must be the full algebra $\mathrm{End}(V)$. On the other hand, the kernel would be a two-sided ideal in A. Now, if $A = M_n(D)$, where D is a division algebra then it is simple. To prove this: Consider any ideal $I \subset M_n(D)$. If $X = x_{ij}e_{ij}$ is in I with some $x_{kl} \neq 0$ then $e_{kk}Xe_{ll} \in I$, but this means $e_{kl} \in I$ but now the ideal generated by e_{kl} is all of $M_n(D)$.

The nature of the representation depends very much on the fields we are working over.

Let us consider first the complex Clifford algebras, $\mathbb{C}\ell(d)$. These are of the form $\mathbb{C}(n)$ or $\mathbb{C}(n) \oplus \mathbb{C}(n)$. The unique irrep of $\mathbb{C}(n)$ is just the defining representation space $V = \mathbb{C}^n$. Thus, $\mathbb{C}\ell(d)$ has a unique irrep for d even, $V = \mathbb{C}^n$ with

$$n = 2^{d/2} \qquad d \qquad \text{even} \tag{5.1}$$

While for d odd $\mathbb{C}\ell(d)$ has two inequivalent irreps $V_\pm$. As vector spaces $V_\pm \cong \mathbb{C}^n$ with

$$n = 2^{(d-1)/2} \qquad d \qquad \text{odd} \tag{5.2}$$

where $\rho(\phi_1 \oplus \phi_2) = \phi_1$ on V_+ and $\rho(\phi_1 \oplus \phi_2) = \phi_2$ on V_-. Another way to characterize these is to consider ω_c, with $(\omega_c)^2 = 1$. Then

$$\rho_+(\omega_c) = +1 \qquad on \qquad V_+ \tag{5.3}$$

$$\rho_-(\omega_c) = -1 \qquad on \qquad V_- \tag{5.4}$$

Now let us consider the Clifford algebras $C\ell(r_+, s_-)$ over $\mathbb{R}$. These are all of the form $\mathbb{K}(n)$ or $\mathbb{K}(n) \oplus \mathbb{K}(n)$. Now, one can form an $\mathbb{K}$-linear representation of $\mathbb{K}(n)$ by having the matrices act on $\mathbb{K}^n$. Here $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$. Note that for the quaternions we must take care that if the matrix multiplication acts from the left on a column vector then the scalar multiplication by $\mathbb{H}$ acts on the right. Again $\mathbb{K}(n)$ is a simple algebra and the unique irrep up to isomorphism is $\mathbb{K}^n$.

Of course, we can consider $\mathbb{K}^n$ to be a real vector space of real dimension $n, 2n, 4n$ for $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$.

Conversely, a real vector space V is said to have a complex structure if there is a linear transformation $I \in End_{\mathbb{R}}(V)$ which satisfies $I^2 = -1$. With a choice of I, V can be made into a complex vector space of dimension $\frac{1}{2}\dim_{\mathbb{R}}(V)$.

Similarly, a real vector space V is said to have a quaternionic structure if there are linear transformations I, J, K with

$$I^2 = J^2 = K^2 = IJK = -1 \tag{5.5}$$

A choice of I, J, K makes V have a multiplication by quaternions, and defines an isomorphism to $\mathbb{H}^d$ with $d = \frac{1}{4}\dim_{\mathbb{R}}(V)$. For example on $\mathbb{H}^n$ the linear operators I, J, K would be *right*-multiplication by $\mathbf{i}, \mathbf{j}, \mathbf{k}$, respectively.

We say that the real representation V of $C\ell(r_+, s_-)$ has a complex structure I if $[\rho(\phi), I] = 0$. Similarly, we say that it has a quaternionic structure if

$$[\rho(\phi), I] = [\rho(\phi), J] = [\rho(\phi), K] = 0 \tag{5.6}$$

We can now trivially read off the basic properties of Clifford algebra representations from the above tables. Recall that $\ell := [d/2]$.

• If $d_T = s - r \neq 3\text{mod}4$ then $C\ell(r_+, s_-)$ is a simple algebra and there is a unique irrep. It is $\mathbb{K}(N)$ acting on $\mathbb{K}^N$ where $N = 2^\ell$ or $N = 2^{\ell-1}$.

• If $d_T = s - r = 3\text{mod}4$ then $C\ell(r_+, s_-)$ is a sum of simple algebras and there are two inequivalent irreps $\rho_\pm$. Note that this is the case where the volume element satisfies $\omega^2 = 1$ *and* ω is central. So we can characterize the representations as $\rho_\pm(\omega) = \pm 1$. Since ω is central these rep's cannot be equivalent.

• For $d_T = 6, 7, 8\text{mod}8$, $\dim_{\mathbb{R}} V = 2^\ell$. For $d_T = 6, 7, 8\text{mod}8$ we can represent Γ^μ by $2^\ell \times 2^\ell$ *real* matrices. In physics this is called a *Majorana representation* of the Clifford algebra.

• Multiplying Γ matrices by a factor of i changes the signature. Therefore for $d_T = 0, 1, 2\text{mod}8$ we can represent Γ^μ by $2^\ell \times 2^\ell$ *pure imaginary* matrices. In physics this is called a *Pseudo-Majorana representation.*

• For $d_T = 1, 5\text{mod}8$ (i.e. $d_T = 1\text{mod}4$) $\dim_{\mathbb{C}} V = 2^\ell$ and hence $\dim_{\mathbb{R}} V = 2^{\ell+1}$, *and* V carries a complex structure commuting with the Clifford action. But if we use $2^\ell \times 2^\ell$ matrices they must be complex.

• For $d_T = 2, 3, 4\text{mod}8$, $\dim_{\mathbb{H}} V = 2^{\ell-1}$, so $\dim_{\mathbb{C}} V = 2^\ell$, and hence $\dim_{\mathbb{R}} V = 2^{\ell+1}$, *and* V carries a quaternionic structure commuting with the Clifford action. Thus we can write a reresentation by $2^{\ell-1} \times 2^{\ell-1}$ quaternionic matrices. If we choose to write the quaternions as 2×2 complex matrices A such that $A^* = \sigma_2 A \sigma_2$ then we can represent the Γ^μ by $2^\ell \times 2^\ell$ complex matrices such that

$$\Gamma^{\mu,*} = J\Gamma^\mu J^{-1} \tag{5.7}$$

with $J = i\sigma^2 \oplus \cdots \oplus i\sigma^2$. Note that $JJ^* = J^2 = -1$ and $J^{tr} = -J$.

****

confusions:

2. Traubenberg has a notion of Pseudo-Symplectic Majorana. Does this make sense?

****

Thus the pattern of representations is the following: ν= number of irreps of the algebra, d= real dimension of the irrep,

s	$C\ell(s_-)$	$\nu(s)$	$d(s)$	Structure	K	$\dim_{\mathbb{C}}$(Irreps of $\mathbb{C}\ell(s)$)
1	$\mathbb{C}$	1	2	$\mathbb{C}$	$\mathbb{Z}$	1
2	$\mathbb{H}$	1	4	$\mathbb{H}$	$\mathbb{Z}$	2
3	$\mathbb{H}\oplus\mathbb{H}$	2	4	$\mathbb{H}$	$\mathbb{Z}\oplus\mathbb{Z}$	2
4	$\mathbb{H}(2)$	1	8	$\mathbb{H}$	$\mathbb{Z}$	4
5	$\mathbb{C}(4)$	1	8	$\mathbb{C}$	$\mathbb{Z}$	4
6	$\mathbb{R}(8)$	1	8	$\mathbb{R}$	$\mathbb{Z}$	8
7	$\mathbb{R}(8)\oplus\mathbb{R}(8)$	2	8	$\mathbb{R}$	$\mathbb{Z}\oplus\mathbb{Z}$	8
8	$\mathbb{R}(16)$	1	16	$\mathbb{R}$	$\mathbb{Z}$	16
9	$\mathbb{C}(16)$	1	32	$\mathbb{C}$	$\mathbb{Z}$	16
10	$\mathbb{H}(16)$	1	64	$\mathbb{H}$	$\mathbb{Z}$	32
11	$\mathbb{H}(16)\oplus\mathbb{H}(16)$	2	64	$\mathbb{H}$	$\mathbb{Z}\oplus\mathbb{Z}$	32
12	$\mathbb{H}(32)$	1	128	$\mathbb{H}$	$\mathbb{Z}$	64

$d=s+1$	$C\ell(s_+,1_-)$	$\nu(s)$	$d(s)$	Structure	K
0+1	$\mathbb{C}$	1	2	$\mathbb{C}$	$\mathbb{Z}$
1 +1	$\mathbb{R}(2)$	1	2	$\mathbb{R}$	$\mathbb{Z}$
2 +1	$\mathbb{R}(2)\oplus\mathbb{R}(2)$	2	2	$\mathbb{R}$	$\mathbb{Z}\oplus\mathbb{Z}$
3 +1	$\mathbb{R}(4)$	1	4	$\mathbb{R}$	$\mathbb{Z}$
4 +1	$\mathbb{C}(4)$	1	8	$\mathbb{C}$	$\mathbb{Z}$
5 +1	$\mathbb{H}(4)$	1	16	$\mathbb{H}$	$\mathbb{Z}$
6 +1	$\mathbb{H}(4)\oplus\mathbb{H}(4)$	2	16	$\mathbb{H}$	$\mathbb{Z}\oplus\mathbb{Z}$
7 +1	$\mathbb{H}(8)$	1	32	$\mathbb{H}$	$\mathbb{Z}$
8 +1	$\mathbb{C}(16)$	1	32	$\mathbb{C}$	$\mathbb{Z}$
9 +1	$\mathbb{R}(32)$	1	32	$\mathbb{R}$	$\mathbb{Z}$
10 +1	$\mathbb{R}(32)\oplus\mathbb{R}(32)$	2	32	$\mathbb{R}$	$\mathbb{Z}\oplus\mathbb{Z}$
11 +1	$\mathbb{R}(64)$	1	64	$\mathbb{R}$	$\mathbb{Z}$

Note that for $C\ell(s_-)$, $d(s+8k) = 16^k d(s)$.

5.1 Representations and Periodicity: Relating Γ-matrices in consecutive even and odd dimensions

The mod-two periodicity of the Clifford algebras over $\mathbb{C}$ is reflected in the representation theory as follows:

If γ^i is an irrep of $\mathbb{C}\ell(2n-1)$, and hence $\gamma^1 \cdots \gamma^{2n-1}$ is a scalar, then we get irreps of $\mathbb{C}\ell(2n)$ and $\mathbb{C}\ell(2n+1)$ by defining new representation matrices:

$$\begin{aligned} \Gamma^i &= \gamma^i \otimes \sigma^1 = \begin{pmatrix} 0 & \gamma^i \\ \gamma^i & 0 \end{pmatrix} \qquad i = 1, \ldots, 2n-1 \\ \Gamma^{2n} &= 1_{2^{n-1}} \otimes \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ \Gamma^{2n+1} &= 1_{2^{n-1}} \otimes \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{5.8}$$

eq:evenodd

Iterating this procedure gives an explicit matrix representation of $\mathbb{C}\ell(d)$ in terms of $2^{[d/2]} \times 2^{[d/2]}$ complex matrices. Note that if we start with $\mathbb{C}\ell(1)$ with $\gamma^1 = 1$ then the matrices we generate will be both Hermitian and unitary, satisfying $\{\Gamma^\mu, \Gamma^\nu\} = 2\delta^{\mu\nu}$.

Of course, this by no means a unique way of relating representations in dimensions $d, d+1, d+2$. In our discussion of the relation to oscillators below we will see another one.

Similarly, working over $\mathbb{R}$, if γ^μ is an irrep of $C\ell(r_+, s_-)$ then

$$\begin{aligned} &\gamma^\mu \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &1 \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &1 \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{5.9}$$

eq:oneoneclf

gives an irrep of $C\ell((r+1)_+, (s+1)_-)$.

****

See Van Proeyen, Tools for susy ; Or Trubenberger for explicit formulae of this type.

****

Remarks

• A representation of the form (5.8) with off-diagonal Γ matrices and diagonal volume element is known as a chiral basis of complex gamma matrices. When we study the spin representations the generators of spin are block diagonal and $\Gamma^1, \ldots, \Gamma^{2n}$ exchange chirality.

• If $\gamma^1 \cdots \gamma^{2n-1} = \xi_{2n-1}$ then

$$\Gamma^1 \cdots \Gamma^{2n+1} = i\xi_{2n-1} \tag{5.10}$$

• There are two distinct irreps of $\mathbb{C}\ell(2n+1)$. We can get the other one by switching the sign of an odd number of Γ matrices above.

6. Comments on a connection to topology

****

THIS SECTION IS OUT OF PLACE - You should go directly to the description of representations in terms of free fermions.

This section is actually more allied with the Z2-graded structure section and should go before or after.

****

Consider a representation of $\mathbb{C}\ell(d)$ by anti-Hermitian gamma matrices Γ^μ such that $\{\Gamma^\mu, \Gamma^\nu\} = -2\delta^{\mu\nu}$, where $\mu = 1, \ldots, d$.

Suppose x_0, x_μ, $\mu = 1, \ldots, d$ are coordinates on the unit sphere S^d embedded in $\mathbb{R}^{d+1}$.

Consider the function

$$T(x) := x_0 1 + x_\mu \Gamma^\mu \tag{6.1}$$

eq:tachfld

Note that

$$T(x)T(x)^\dagger = 1 \tag{6.2}$$

and therefore $T(x)$ is a unitary matrix for every $x \in S^{d-1}$. We can view $T(x)$ as describing a map $T : S^d \to U(2^{[d/2]})$. Now, sometimes this map is topologically trivial and sometimes it is topologically nontrivial. By nontrivial we mean that it represents a nontrivial element of the homotopy group $\pi_d(U(2^{[d/2]}))$.

For example: If $d = 1$ then one of the two irreducible representations is $\Gamma = i$. If $x_0^2 + x_1^2 = 1$ then

$$T(x) = x_0 + i x_1 \tag{6.3}$$

is a map $S^1 \to U(1)$ of winding number 1. If $d = 3$ then we may choose $\Gamma^i = \sqrt{-1}\sigma^i$ and

$$T(x) = x_0 + x_i \tau_i \tag{6.4}$$

is our representation of $SU(2)$. Thus the map $T : S^3 \to SU(2)$ is the identity map. It has winding number one and generates $\pi_3(SU(2)) = \mathbb{Z}$.

Here is one easy criterion for triviality: Suppose we can introduce another anti-Hermitian $2^{[d/2]} \times 2^{[d/2]}$ gamma matrix, Γ so that $(\Gamma)^2 = -1$ and $\{\Gamma, \Gamma^\mu\} = 0$. Now consider the unit sphere

$$S^{d+1} = \{(x_0, x_\mu, y) | x_0^2 + \sum_{\mu=1}^{d} x_\mu x_\mu + y^2 = 1\} \subset \mathbb{R}^{d+2} \tag{6.5}$$

Then we can define

$$\tilde{T}(x, y) = x_0 + x_\mu \Gamma^\mu + y\Gamma \tag{6.6}$$

Figure 2: The map on the equator extends to the northern hemisphere, and is therefore homotopically trivial. fig:clffstpa

When restricted to $S^{d+1} \subset \mathbb{R}^{d+2}$, $\tilde{T}$ is also unitary and maps $S^{d+1} \to U(2^{[d/2]})$. Moreover $\tilde{T}(x,0) = T(x)$ while $\tilde{T}(0,1) = \Gamma$. Thus $\tilde{T}(x,y)$ provides an explicit homotopy of $T(x)$ to the constant map.

Thus, we can see that *if the irreducible representation of* $\mathbb{C}\ell(d)$ *is the restriction of an irreducible representation of* $\mathbb{C}\ell(d+1)$ *to* $\mathbb{C}\ell(d)$ *then* $T(x)$ *is topologically trivial.*

Amazingly, it turns out that the converse is also true. If the irrep cannot be extended, then $T(x)$ is homotopically nontrivial. In fact, it represents a generator of $\pi_{d-1}(U(2^{[d/2]}))$.

Now, let us look at the Clifford representation theory. When $d = 2p+1$ is odd there are two irreps of complex dimension 2^p. Restricting to the action of $\mathbb{C}\ell_{2p}$ on either irrep gives the 2^p-dimensional representation. On the other hand, if $d = 2p$ is even, then $\mathbb{C}\ell_{2p-1}$ has irreps of complex dimension 2^{p-1}. Thus, the action of $\mathbb{C}\ell_{2p-1}$ on the complex 2^p-dimensional irrep of $\mathbb{C}\ell_{2p}$ does not give an irrep. Or, put differently, the 2^{p-1}-dimensional irrep of $\mathbb{C}\ell_{2p-1}$ does *not* extend to an irrep of $\mathbb{C}\ell_{2p}$.

These facts are compatible with the statement in topology that

$$\pi_{2p-1}(U(N)) = \mathbb{Z} \qquad N \geq p \tag{6.7}$$ eq:piodd

$$\pi_{2p}(U(N)) = 0 \qquad N > p \tag{6.8}$$ eq:pieven

Note that these equations say that for N sufficiently large, the homotopy groups do not depend on N. These are called the *stable homotopy groups* of the unitary groups and can be denoted $\pi_k(\mathbf{U})$. The mod two periodicity of $\pi_k(\mathbf{U})$ as a function of k is known as Bott periodicity.

Similarly, there are stable homotopy groups $\pi_k(\mathbf{O})$ and $\pi_k(\mathbf{Sp})$ which are mod-8 periodic in k.

****

Explain about vector fields on spheres

****

References:

1. Atiyah, Bott, and Shapiro,

2. Lawson and Michelson, *Spin Geometry*

7. Free fermion Fock space

7.1 Left regular representation of the Clifford algebra

The Clifford algebra acts on itself, say, from the left. On the other hand, it is a vector space. Thus, as with any algebra, it provides a representation of itself, called the left-regular representation.

Note that this representation is 2^d dimensional, and hence rather larger than the $\sim 2^{[d/2]}$ dimensional irreducible representations. Hence it is highly reducible. In order to find irreps we should "take a squareroot" of this representation.

We will now describe some ways in which one can take such a "squareroot."

7.2 The Exterior Algebra as a Clifford Module

We have noted that

$$C\ell(r_+, s_-) \cong \Lambda^* \mathbb{R}^d \tag{7.1}$$ eq:vspi

as a vector spaces. Also, while the exterior algebra $\Lambda^*\mathbb{R}^d$ is an algebra we stressed that (7.1) is *not* an algebra isomorphism.

Nevertheless, since (7.1) is a vector space isomorphism this means that $\Lambda^*(\mathbb{R}^d)$ must be a Clifford *module*, that is, a representation space of the Clifford algebra. We now describe this representation.

If $v \in \mathbb{R}^{r,s}$ then we can define the contraction operator by

$$\iota(v)(v_{i_1} \wedge \cdots \wedge v_{i_k}) := \sum_{s=1}^{k} (-1)^{s-1} Q(v, v_{i_s}) v_{i_1} \wedge \cdots \wedge \widehat{v_{i_s}} \wedge \cdots \wedge v_{i_k} \tag{7.2}$$ eq:contract

where the hat superscript means we omit that factor. Similarly, we can define the wedge operator by

$$w(v)(v_{i_1} \wedge \cdots \wedge v_{i_k}) := v \wedge v_{i_1} \wedge \cdots \wedge v_{i_k} \tag{7.3}$$ eq:wedgeoper

These operators are easily shown to satisfy the algebra:

$$\begin{aligned} \{\iota(v_1), \iota(v_2)\} &= 0 \\ \{w(v_1), w(v_2)\} &= 0 \\ \{\iota(v_1), w(v_2)\} &= Q(v_1, v_2) \end{aligned} \tag{7.4}$$ eq:rccrs

Using these relations we see that we can represent Clifford multiplication by v by the operator:

$$\rho(v) = \iota(v) + w(v) \tag{7.5}$$

Since the $v \in V$ generate the Clifford algebra we can then extend this to a representation of the entire Clifford algebra by taking $\rho(\phi_1 \cdot \phi_2) = \rho(\phi_1)\rho(\phi_2)$.

7.3 Representations from maximal isotropic subspaces

We still need to take a "squareroot" of our representation. One example where this can be done is the following.

Let W be a real vector space and consider $V = W \oplus W^*$. Note that V admits a natural nondegenerate quadratic form of signature $(+1^n, -1^n)$ where we take W, W^* to be isotropic and use the pairing $W \times W^* \to \mathbb{R}$. That is, if we choose a basis w_i for W and a dual basis $\hat{w}^i$ for W^* then with respect to this basis

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{7.6}$$

The resulting Clifford algebra is $C\ell(n_+, n_-) \cong \mathbb{R}(2^n)$ where $n = \dim W$.

One way to construct the irrep of dimension 2^n is by taking the representation space to be $\Lambda^* W^*$.

Now $\rho(\hat{w})$ for $\hat{w} \in W^*$ is defined by wedge product, $\hat{w}\wedge$ and $\rho(w)$ for $w \in W$ is defined by $\rho(w) = \iota(w)$ where $\iota(w)$ is the contraction operator:

$$\iota(w)(\hat{w}^{i_1} \wedge \cdots \wedge \hat{w}^{i_n}) = \sum_{j=1}^{n} (-1)^{j-1} \langle w, \hat{w}^{i_j} \rangle \hat{w}^{i_1} \wedge \cdots \wedge \hat{w}^{i_{j-1}} \wedge \hat{w}^{i_{j+1}} \wedge \cdots \wedge \hat{w}^{i_n} \tag{7.7}$$ eq:contracta

A simple computation shows that

$$\begin{aligned} \{\rho(w), \rho(w')\} &= 0 \\ \{\rho(\hat{w}), \rho(\hat{w}')\} &= 0 \\ \{\rho(w), \rho(\hat{w}')\} &= \langle w, \hat{w}' \rangle \end{aligned} \tag{7.8}$$ eq:sensf

and thus the Clifford relations are satisfied.

Example: For example, if M is a manifold we can consider $TM \oplus T^*M$ which has a natural quadratic form of signature (n, n) since TM and T^*M are dual spaces. Note that $W = TM$ a maximal isotropic subspace, and a natural choice of complementary isotropic subspace is $U = T^*M$. Then the Clifford algebra acts on the DeRham complex $\Lambda^* T^* M$. $\rho(v^i) = dx^i \wedge$ acts by wedge product, and $\rho(w_i) = \iota(\frac{\partial}{\partial x^i})$ acts by contraction. These represent the Clifford algebra:

$$\begin{aligned} \{v^i, v^j\} &= 0 \\ \{v^i, w_j\} &= \delta^i{}_j \\ \{w_i, w_j\} &= 0 \end{aligned} \tag{7.9}$$ eq:tensors

The above construction can be generalized as follows:

Suppose V is $2n$-dimensional with a nondegenerate metric of signature (n, n). Thus $C\ell(n_+, n_-) \cong \mathbb{R}(2^n)$ and we wish to construct the 2^n-dimensional irrep. Suppose we have

a decomposition of V into two maximal isotropic subspaces $V = W \oplus U$ where W, U are maximal isotropic. That is, with respect to this decomposition we have

$$Q = \begin{pmatrix} 0 & q \\ q^\dagger & 0 \end{pmatrix} \tag{7.10}$$

where $q : U \to W$ is an isomorphism.

Note that there is a family of such decompositions, parametrized by some kind of Grassmannian.

Then, we claim, the exterior algebra $\Lambda^*(V/W)$ is naturally a 2^n dimensional representation of the Clifford algebra on V.

$u \in U$ acts on $\Lambda^*(V/W)$ by wedge product: Note that V/W acts via wedge product. Since U is a subspace of V it descends to a subspace of V/W and hence it acts by wedge product. On the other hand, $w \in W$ acts by contraction

$$\iota(w)([v_{i_1}] \wedge \cdots \wedge [v_{i_n}]) = \sum_{j=1}^{n} (-1)^{j-1} Q(w, v_{i_j}) [v_{i_1}] \wedge \cdots \wedge [v_{i_{j-1}}] \wedge [v_{i_{j+1}}] \wedge \cdots \wedge [v_{i_n}] \tag{7.11}$$

eq:contractaa

Note that the expression $Q(w, v_{i_j})$ is unambiguous because W is isotropic.

There is an alternative description of the same representation since one can show that $V/W \cong W^*$. To see this note that given v, $\ell_v : w \mapsto (v, w)$ is an element of W^* and $\ell_v = \ell_{v+w}$ for $w \in W$ (since W is isotropic). Thus we could also have represented the Clifford algebra on $\Lambda^* W^*$. Elements of W act by contraction and elements of U act by wedge product (where one needs to use the isomorphism $V/W \cong W^*$.)

Remarks

• The above construction appears to depend on a *choice* of decomposition of a vector space into a sum of isotropic subspaces. In finite dimensions all these representations are equivalent. But this is not so in infinite dimensions.

•

The choice of W is a choice of "Dirac sea." Different choices are related by "Bogoliubov transformation."

Exercise

Describe the analogous representation on $\Lambda^* TM$.

7.4 Splitting using a complex structure

Another way we can split the space V into half-dimensional spaces is by using a complex structure. Our goal in this section is to construct a representation of $\mathbb{C}\ell(d)$ for $d = 2n$ using the ideas of the previous two sections.

For definiteness, let us take $V = \mathbb{R}^{2n}$ with Euclidean signature:

$$Q(e^i, e^j) = +\delta^{ij} \tag{7.12}$$

and choose a complex structure compatible with the metric:

Given a choice of complex structure compatible with the metric there exists a basis e^i, $i = 1, \ldots, 2n$ of V such that

$$\begin{aligned} Ie^{2j-1} &= -e^{2j} \\ Ie^{2j} &= e^{2j-1} \qquad\qquad j = 1, \ldots, n \end{aligned} \tag{7.13}$$

That is, in this basis:

$$I = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{7.14}$$

(with n summands).

By extending scalars to $\mathbb{C}$ we can diagonalize I. Put differently, $V \otimes_{\mathbb{R}} \mathbb{C} \cong W \oplus \bar{W}$:

$$\begin{aligned} Iv^j &= iv^j \\ Iw_j &= -iw_j \qquad\qquad j = 1, \ldots, n \end{aligned} \tag{7.15}$$

where $v^j = \frac{1}{2}(e^{2j-1} + ie^{2j}), w_j = \frac{1}{2}(e^{2j-1} - ie^{2j})$.

Moreover, if we extend the metric $\mathbb{C}$-linearly then it becomes:

$$\begin{aligned} Q(v^j, v^k) &= 0 \\ Q(w_j, w_k) &= 0 \\ Q(v^j, w_k) &= \delta^j{}_k \qquad\qquad j, k = 1, \ldots, n \end{aligned} \tag{7.16}$$

eq:nnmetr

so that on the complex vector space we have a metric of signature (n, n).

Thus, the Clifford algebra elements corresponding to v^j, w_k satisfy the canonical commutation relations of fermionic oscillators. That is, if we define:

$$\begin{aligned} \bar{a}_j = v^j &= \frac{1}{2}(e^{2j-1} + ie^{2j}) \\ a_j = w_j &= \frac{1}{2}(e^{2j-1} - ie^{2j}) \end{aligned} \tag{7.17}$$

Then the above relations are the standard fermion CCR's:

$$\begin{aligned} \{a_j, a_k\} &= 0 \\ \{\bar{a}_j, \bar{a}_k\} &= 0 \\ \{a_j, \bar{a}_k\} &= \delta_{jk} \end{aligned} \tag{7.18}$$

We will now use the above insight about the importance of isotropic subspaces to construct representations of the Clifford algebras over $\mathbb{C}$. For example we could take the representation space to be $\Lambda^* \bar{W}$ with $\bar{a}_j$ acting by wedge product and a_j acting by contraction.

Our construction can be related to the standard construction by identifying the "vacuum state" $|0\rangle$ with $1 \in \Lambda^* \bar{W}$. Then

$$a_j|0\rangle = 0 \tag{7.19}$$ eq:vacchoice

The vacuum state is also known as a *Dirac sea* or a *Clifford vacuum*. We can build a representation of the Clifford algebra by acting on the Clifford vacuum with creation operators $\bar{a}_j$. In this way we obtain a natural basis for the representation of the Clifford algebra by creating fermionic states:

$$\bar{a}_{j_1} \cdots \bar{a}_{j_s}|0\rangle \tag{7.20}$$

Thus, the representation is isomorphic to $\Lambda^* \bar{W} \cong \mathbb{C}^{2^n}$, and hence is an irrep of $\mathbb{C}\ell_{2n}$.

We can put a Hermitian structure on the representation space and then $\bar{a}_j = (a_j)^\dagger$.

Remarks

Repeat for other signatures –

7.5 Explicit matrices and intertwiners in the Fock representations

Let us use the oscillator approach to form an explicit basis for the irreducible representations of $\mathbb{C}\ell(d)$.

Then

$$\begin{aligned} a_j^\dagger &= \frac{1}{2}(e_{2j-1} + ie_{2j}) \\ a_j &= \frac{1}{2}(e_{2j-1} - ie_{2j}) \qquad j = 1, \ldots, n \end{aligned} \tag{7.21}$$

so $\{a_j, a_k^\dagger\} = \delta_{jk}$.

Consider first the case $n = 1$. we use the representation

$$|0\rangle := |-\frac{1}{2}\rangle \qquad a^\dagger|0\rangle := |+\frac{1}{2}\rangle \tag{7.22}$$

This labelling will be useful later for representations of the spin group.

Now taking

$$x_1|+\rangle + x_2|-\rangle \to \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{7.23}$$

we have the representation

$$\rho(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 0 & -i \\ +i & 0 \end{pmatrix} \tag{7.24}$$

Now, with n oscillator pairs we choose a basis for a 2^n dimensional vector space

$$(a_n^\dagger)^{s_n + \frac{1}{2}} (a_{n-1}^\dagger)^{s_{n-1} + \frac{1}{2}} \cdots (a_1^\dagger)^{s_1 + \frac{1}{2}} |0\rangle \tag{7.25}$$

where $s_i = \pm\frac{1}{2}$. We identify these states with the basis for the tensor product of representations

$$|s_n, s_{n-1}, \dots, s_1\rangle = |s_n\rangle \otimes |s_{n-1}\rangle \otimes \cdots \otimes |s_1\rangle \tag{7.26}$$

(The vector $(s_n, \dots, s_1)$ is what is called a *spinor weight*. In the theory of representations of semi-simple Lie algebras the space is graded by the action of the Cartan subalgebra and the grading is called the weight. The vectors $(s_n, \dots, s_1)$ are the weights of the spinor representations of $so(2n; \mathbb{C})$. See Chapter *** below.)

Let $\Gamma^j_{(n-1)}$ be the $2^{n-1} \times 2^{n-1}$ representation matrices of e_j for a collection of $(n-1)$ oscillators. Then when we add the nth oscillator pair we get

$$\begin{aligned}
\rho_n(e_j) = \Gamma^j_{(n)} &= \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \otimes \Gamma^j_{(n-1)} \qquad\qquad j = 1, \dots, 2n-2 \\
\rho_n(e_{2n-1}) = \Gamma^{2n-1}_{(n)} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes 1_{2^{n-1}} \\
\rho_n(e_{2n}) = \Gamma^{2n}_{(n)} &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \otimes 1_{2^{n-1}}
\end{aligned} \tag{7.27}$$

We take the complex volume form to be

$$\Gamma_\omega = (-i)^n \Gamma^1 \cdots \Gamma^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7.28}$$ eq:cplxvol

where there are n factors.

For $d = 2n+1$ we still take n pairs of oscillators and set $\Gamma^{2n+1} = \Gamma_\omega$.

8. Bogoliubov Transformations and the Choice of Clifford vacuum

It is important to note that our decomposition into creation and annihilation operators depends on a *choice* of complex structure.

On $\mathbb{R}^{2n}$ we can produce other complex structures from $I' = RIR^{-1}$ with $R \in GL(2n, \mathbb{R})$. The complex structure will be compatible with the Euclidean metric if $R \in O(2n)$.

If we use I' rather than I then the new oscillators $b_j, \bar{b}_j$ will be related to the old ones by a *Bogoliubov transformation*:

$$\begin{aligned}
\bar{b}_i &= A_{ij}\bar{a}_j + B_{ij}a_j \\
b_i &= C_{ij}\bar{a}_j + D_{ij}a_j
\end{aligned} \tag{8.1}$$ eq:bogoliub

For a general complex linear combination (8.1) the CCR's are preserved iff the matrix

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{8.2}$$

satisfies

$$g\begin{pmatrix}0&1\\1&0\end{pmatrix}g^{tr}=\begin{pmatrix}0&1\\1&0\end{pmatrix} \tag{8.3}$$

That is, iff

$$\begin{aligned}AD^{tr}+BC^{tr}&=1\\BA^{tr}+AB^{tr}&=0\\CD^{tr}+DC^{tr}&=0\end{aligned} \tag{8.4}$$

That is, $g\in O(n,n;\mathbb{C})$.

****

SO: WHAT IS THE POINT? HOW IS g RELATED TO R ?

****

The new Clifford vacuum is of course the same if g amounts to a complex transformation of the a_j to the b_j and the $\bar{a}_j$ to the $\bar{b}_j$. Thus, $B=C=0$ and $AD^{tr}=1$. Thus, the space of Dirac vacua is $O(n,n;\mathbb{C})/GL(n,\mathbb{C})$, a homogeneous space of complex dimension $n(n-1)$.

Remark: Note $O(n,n;\mathbb{C})\cong O(2n;\mathbb{C})$. Indeed it is useful to make this isomorphism explicit. Let

$$T=\frac{1}{\sqrt{2}}\begin{pmatrix}1&i\\1&-i\end{pmatrix} \tag{8.5}$$

then

$$\begin{pmatrix}0&1\\1&0\end{pmatrix}=TT^{tr} \tag{8.6}$$

so that $\hat{g}=T^{-1}gT$ is in $O(2n;\mathbb{C})$, i.e. $\hat{g}\hat{g}^{tr}=1$.

If we impose a conjugation operation $a_j\to\bar{a}_j$ which is preserved by the Bogoliubov transformation then $A=D^*$ and $B=C^*$ and we get AB^{tr} is antisymmetric and $AA^\dagger+BB^\dagger=1$. Note that in this case g is a unitary transformation $g\in U(2n)$. Now, the matrix T defined above is also unitary so in this case $\hat{g}\in U(2n)\cap O(2n;\mathbb{C})\cong O(2n;\mathbb{R})$. Moreover, the group preserving the Clifford vacuum is the subgroup with $B=0$ is indeed isomorphic to $U(n)$. Thus, the space of Clifford vacua obtained by unitary transformations is $O(2n;\mathbb{R})/U(n)$, as space of real dimension $n(n-1)$. This is the space of complex structures on $\mathbb{R}^{2n}$ compatible with the Euclidean metric.

We will discuss this further after we have introduced the Spin group.

****

To add:

1. More Discussion of polarization. see Pressley and Segal, *Loop Groups*, Oxford, p. 233. Segal's Stanford notes.

2. J. Strathdee, Int. J. Mod. Phys. **A2** 273 (1987)

9. Comments on Infinite-Dimensional Clifford Algebras

In quantum field theories of fermions we encounter infinite-dimensional Clifford algebras. The construction of the irreducible representations are a little different from the finite-dimensional case explained above.

Here we follow some notes of G. Segal. ("Stanford Lectures")

The typical situation is that we have a vector space of one-particle wavefunctions of fermions on a spatial slice. Call this E. It could be, for example, the L^2-spinors on a spatial slice. Canonical quantization calls for a representation of the infinite-dimensional Clifford algebra $C\ell(E \oplus E^*)$. It turns out to be physically wrong to consider $\Lambda^* E$ or $\Lambda^* E^*$. Rather, the physically correct construction is more subtle.

Typically, there is a self-adjoint operator D on E with spectrum of eigenvalues λ_k which goes to $\pm\infty$ for $k \to \pm\infty$. (In the physical applications, D would be the spatial Dirac operator.) Now one wants to consider a "Dirac sea." This is a formal element of $\Lambda^* E$ given by taking the wedge-product with all the negative energy levels. Let e_k be an ON basis of eigenvectors of D. Let us assume that $\lambda = 0$ is not in the spectrum, and let us label the eigenvectors so that e_k has $\lambda_k > 0$ for $k > 0$ and $\lambda_k < 0$ for $k < 0$.

Now we try to define the "Fock vacuum."

$$\Omega = e_0 \wedge e_{-1} \wedge e_{-2} \wedge \cdots \tag{9.1}$$

This is a "semi-infinite form" The Fock space will then be spanned by elements $e_{\vec{k}}$ where $\vec{k} = \{k_0, k_{-1}, k_{-2}, \cdots\}$ is a strictly decreasing series of integers $k_0 > k_{-1} > k_{-2} > \cdots$ which only differs from $\{0, -1, -2, \cdots\}$ by a finite number of elements. The resulting space is even a Hilbert space and is acted on by $C\ell(E \oplus E^*)$ in the way we have described above.

However, there is a problem problem with this definition. The problem is that the resulting vector space is only well-defined as a *projective space*. The reason is that we could choose a different eigenbasis $\tilde{e}_k = u_k e_k$, with $|u_k| = 1$. In general $\prod u_k$ is not well defined.

Now, this becomes a real problem when we consider *families* of operators D. For example, we could consider families of metrics on our space(time), or we could consider the Dirac operator coupled to a gauge field and consider the family parametrized by the gauge field. We will find in such situations that there is no unambiguous way to choose a well-defined Fock vacuum throughout the entire family.

One mathematical approach to making the construction well-defined is the following. We first introduce the notion of a polarization on a vector space. This is a *family* of decompositions $E = E^+ \oplus E^-$ where $E^\pm$ are – very roughly speaking – half of E and different decompositions in the family are "close" to one another. We certainly want to consider two decompositions $E = E^+ \oplus E^-$ and $E = \tilde{E}^+ \oplus \tilde{E}^-$ to be in the same polarization if they differ by finite-dimensional spaces.

Definition: A *coarse polarization*, denoted $\mathcal{J}$, of E is a class of operators $P : E \to E$ such that

1. $P^2 = 1$ modulo compact operators.

2. For any two elements $P, P' \in \mathcal{J}$, $P' - P$ is compact.

3. $\mathcal{J}$ does not contain ± 1.

****

COMPACT OR HILBERT-SCHMIDT ?

****

Given a polarization, one defines the restricted Grassmannian $Gr(E)$ to be the set of spaces E^- which arise in the decompositions allowed by the polarization.

In our example of spinors on spacetime, an operator such as D above defines a polarization by considering the projection onto positive and negative eigenvalues of D to be in $\mathcal{J}$. If Y is the boundary of some spacetime $Y = \partial X$ then the boundary values of solutions of the Dirac equation on X will define an element of the corresponding restricted Grassmannian. As we vary, say, the metric on X we will obtain a family of vector spaces in $Gr(E)$.

Now, the polarization defined by D let $E = E^+ \oplus E^-$ be an allowed decomposition. Then we can make the above Dirac sea precise by considering the Fock space

$$\mathcal{F}_{E^-}(E) := \Lambda^*((E^-)^*) \otimes \Lambda^*(E/E^-) \tag{9.2}$$

where now $C\ell(E \oplus E^*)$ acts as follows: $e \in E$ can be decomposed as $e = e_- \oplus e_+$ and it acts by

$$\rho(e) = 1 \otimes (e_+ \wedge) + \iota(e_-) \otimes 1 \tag{9.3}$$

while $\hat{e} \in E^*$ has a decomposition $\hat{e}_- \oplus \hat{e}_+$ and it acts as

$$\rho(\hat{e}) = (\hat{e}_- \wedge) \otimes 1 + 1 \otimes \iota(\hat{e}_+) \tag{9.4}$$

Now the vector space $\mathcal{F}_{E^-}(E)$ has a canonically defined vacuum, namely, 1, just as in finite dimensions. However, the price we pay is that for different elements E_1^- and E_2^- in the Grassmannian the isomorphism

$$\mathcal{F}_{E_1^-}(E) \to \mathcal{F}_{E_2^-}(E) \tag{9.5}$$

is only defined up to a scalar. The line bundle of Fock vacua will be nontrivial.

****

Explain the last two statements.

****

In finite dimensions there is a unique irrep of $C\ell(E \oplus E^*)$. Indeed, we constructed an isomorphism between different representations constructed using different complex structures, or using different maximal isotropic subspaces using a Bogoliubov transformations. But in infinite dimensions different polarizations can lead to inequivalent representations of $C\ell(E \oplus E^*)$.

10. Properties of Γ-matrices under conjugation and transpose: Intertwiners

In physical applications of Clifford algebras it is often important to have a good understanding of the Hermiticity, unitarity, (anti-)symmetry and complex conjugation properties of the gamma matrices.

The reason we care about this apparently dull question is that these properties are very important for:

1. Imposing reality conditions on fermions - needed to get proper numbers of degrees of freedom in various theories, especially supersymmetric theories.

2. Forming action principles for theories of fermions.

3. Constructing supersymmetry algebras.

4. Group theory manipulations such as decomposition of tensor products of spinor representations. This is crucial for understanding what super-Poincare and superconformal algebras one can construct.

10.1 Definitions of the intertwiners

A useful ref. for this section is Kugo-Townsend pp. 360-375

Given a representation provided by gamma matrices Γ^μ we always have another representations:

$$\Gamma^\mu \to -\Gamma^\mu \tag{10.1}$$

Moreover, we can take the transpose

$$\Gamma^\mu \to (\Gamma^\mu)^{tr} \tag{10.2}$$

For representations by matrices over the complex numbers or over the quaternions we can take the conjugation:

$$\Gamma^\mu \to (\Gamma^\mu)^* \tag{10.3}$$

Finally, we can combine these operations. For example

$$\pm(\Gamma^\mu)^\dagger \tag{10.4}$$

is another representation.

These representations might, or might not be equivalent, depending on dimension and signature.

When the the representations are equivalent (and this is guaranteed when there is a unique irrep, and hence when $d = r + s$ is even) Schur's lemma guarantees that we can define intertwiners:

$$\begin{aligned} -\Gamma^\mu &= \Gamma\Gamma^\mu\Gamma^{-1} \\ \xi(\Gamma^\mu)^\dagger &= A_\xi\Gamma^\mu A_\xi^{-1} \\ \xi(\Gamma^\mu)^* &= B_\xi\Gamma^\mu B_\xi^{-1} \\ \xi(\Gamma^\mu)^{tr} &= C_\xi\Gamma^\mu C_\xi^{-1} \end{aligned} \tag{10.5}$$

where $\xi = \pm 1$. Here $\Gamma = \rho(\omega) = \Gamma^1 \cdots \Gamma^d$ is the volume element. These equations define A, B, C only up to a nonzero scalar multiple. If we know two of the intertwiners we can easily obtain a third, e.g. we could take $A_+ = C_\xi^* B_\xi$ etc.

The Γ^μ can be chosen to be unitary. If this is done then...
****

Be more systematic about the relations between the interwiners.
****

10.2 The charge conjugation matrix for Lorentzian signature

As a simple example of why we wish to know about symmetry properties of the transpose let us consider the special case of $C\ell(s_+, 1_-)$ appropriate to Lorentzian signature spacetime.

As we will see we can (????? NOT FOR ALL ODD DIMENSIONS! ???) define the charge conjugation matrix C with

$$-(\Gamma^\mu)^{tr} = C\Gamma^\mu C^{-1} \tag{10.6}$$ eq:chargeconj

It follows from explicit constructions above that we can, and will, choose a basis such that Γ^0 is anti-hermitian and Γ^i are Hermitian, where 0 denotes the negative signature direction. For *real* gamma matrices we have:

$$(\Gamma^0)^{tr} = -\Gamma^0 \qquad\qquad (\Gamma^i)^{tr} = +\Gamma^i \tag{10.7}$$ eq:cconja

where 0 denotes the direction with $\eta^{00} = -1$ and $i = 1, \ldots, s$. Now, using the properties of the Clifford algebra we can take $C = \Gamma^0$ Note that $C^{-1} = -C$, $C^{tr} = -C$.

Moreover, we have

$$(C\Gamma^\mu)^{tr} = +(C\Gamma^\mu) \tag{10.8}$$ eq:cconjb

That is, the symmetric product of the real representation of the Clifford algebra contains the vector. This is important: It means that when we work with real fields we can make a real action:

$$\int \mathrm{vol}\Psi^{tr} C\Gamma^\mu \partial_\mu \Psi + \cdots \tag{10.9}$$

Note that it is important that Ψ is anti-commuting. If $C\Gamma^\mu$ were anti-symmetric with anti-commuting Ψ, or if $C\Gamma^\mu$ were symmetric, with commuting Ψ then the action density would be a total derivative.

The symmetry of $C\Gamma^\mu$ also allows the definition of supersymmetry algebras:

$$\{Q_\alpha, Q_\beta\} = (C\Gamma^\mu)_{\alpha\beta} P_\mu + \cdots \tag{10.10}$$

Exercise
Show that in this situation

$$C\Gamma^{\mu_1 \cdots \mu_k} = \begin{cases} \text{Symmetric} & fork = 1, 2 \mathrm{mod}\ 4 \\ \text{Antisymmetric} & fork = 3, 4 \mathrm{mod}\ 4 \end{cases} \tag{10.11}$$

10.3 General Intertwiners for $d = r + s$ even

10.3.1 Unitarity properties

When representing $C\ell(r_+, s_-)$ by complex matrices we can always choose Γ^μ to be unitary. To prove this, note that we constructed such representations for $C\ell(1_\pm)$, $C\ell(2_\pm)$ and $C\ell(1_+, 1_-)$ and $C\ell(3\pm)$. Then it follows from the tensor product construction.

The Clifford relations imply $(\Gamma^\mu)^{-1} = \eta^{\mu\mu}\Gamma^\mu$, and hence if we choose our matrices to be unitary then

$$(\Gamma^\mu)^\dagger = \eta^{\mu\mu}\Gamma^\mu = \begin{cases} +\Gamma^\mu & \mu = 1, \dots, r \\ -\Gamma^\mu & \mu = r+1, \dots, r+s \end{cases} \tag{10.12}$$ eq:herm

with no sum on μ. Thus, as opposed to representations of $\mathbb{C}\ell(d)$, we are *not* free to choose the Hermiticity properties.

Exercise

Define

$$U_+ = \Gamma^1 \dots \Gamma^r \tag{10.13}$$

$$U_- = \Gamma^{r+1} \dots \Gamma^{r+s} \tag{10.14}$$

Show that

$$(\Gamma^\mu)^\dagger = \begin{cases} U_+ \Gamma^\mu U_+^{-1} & r = 1 \mathrm{mod} 2 \\ U_- \Gamma^\mu U_-^{-1} & s = 0 \mathrm{mod} 2 \end{cases} \tag{10.15}$$

$$-(\Gamma^\mu)^\dagger = \begin{cases} U_+ \Gamma^\mu U_+^{-1} & r = 0 \mathrm{mod} 2 \\ U_- \Gamma^\mu U_-^{-1} & s = 1 \mathrm{mod} 2 \end{cases} \tag{10.16}$$

The $U_\pm$ are used to construct intertwiners for complex conjugation and transpose below.

10.3.2 General properties of the unitary intertwiners

We will write some explicit intertwiners using our oscillator representation betlow. In this section we derive some general properties of the intertwiners.

First, by taking the Hermitian conjugate of the defining relations and applying Schur's lemma we see that $A^\dagger A, B^\dagger B, C^\dagger C$ must be proportional to the unit matrix. Moreover that scalar must be positive and therefore, WLOG we can always take A, B, C to be unitary.

Second, by iterating the equations for B and C we can see that

Now, again by Schur,

$$C_\xi^{-1} C_\xi^{tr} = \epsilon 1 \tag{10.17}$$ eq:ceexirel

and

$$B_\xi^* B_\xi = \epsilon' 1 \tag{10.18}$$ eq:beexirel

for some scalars ϵ and ϵ'. Note, moreover that consistency of (10.17) means $\epsilon = \pm 1$ and, if we also use unitarity of B_ξ then $\epsilon' = \pm 1$. Moreover, if we make a definite choice of A_+ and use this to relate C_ξ and B_ξ then ϵ and ϵ' are not independent.

It turns out that ϵ cannot be chosen arbitrarily, but is determined by the dimension $d\mathrm{mod}8$ and ξ.

Using the definition of C_ξ we get:

$$(C_\xi \Gamma^{\mu_1 \cdots \mu_k})^{tr} = \xi^k (-1)^{\frac{1}{2}k(k-1)} C_\xi \Gamma^{\mu_1 \cdots \mu_k} C_\xi^{-1} C_\xi^{tr} \tag{10.19}$$ eq:symmetry

It follows that all the matrices

$$C_\xi \Gamma^{\mu_1 \cdots \mu_k} \tag{10.20}$$

are either symmetric or antisymmetric:

$$(C_\xi \Gamma^{\mu_1 \cdots \mu_k})^{tr} = \epsilon \xi^k (-1)^{\frac{1}{2}k(k-1)} C_\xi \Gamma^{\mu_1 \cdots \mu_k} \tag{10.21}$$ eq:sympar

Now, we are working with Clifford matrices over $\mathbb{C}$ and so the Clifford algebra is just $\mathbb{C}(2^{d/2})$. That algebra contains:

- $2^{d-1} + 2^{\frac{d}{2}-1}$ *symmetric matrices*
- $2^{d-1} - 2^{\frac{d}{2}-1}$ *antisymmetric matrices*

On the other hand, we can enumerate the number of symmetric or antisymmetric matrices combining (10.21) with the sums:

$$\begin{aligned}
\sum_{k=0(4)} \binom{d}{k} &= 2^{d-2} + 2^{\frac{1}{2}d-1} \cos(\frac{\pi d}{4}) \\
\sum_{k=1(4)} \binom{d}{k} &= 2^{d-2} + 2^{\frac{1}{2}d-1} \sin(\frac{\pi d}{4}) \\
\sum_{k=2(4)} \binom{d}{k} &= 2^{d-2} - 2^{\frac{1}{2}d-1} \cos(\frac{\pi d}{4}) \\
\sum_{k=3(4)} \binom{d}{k} &= 2^{d-2} - 2^{\frac{1}{2}d-1} \sin(\frac{\pi d}{4})
\end{aligned} \tag{10.22}$$

(This can be proved by applying the binomial expansion to $(1+\zeta)^d$ for the four distinct fourth roots of 1. It holds for d even or odd.)

The result is:

$\xi = +1, \epsilon = +1$: $d = 0, 2\mathrm{mod}8$

$\xi = +1, \epsilon = -1$: $d = 4, 6\mathrm{mod}8$

$\xi = -1, \epsilon = +1$: $d = 0, 6\mathrm{mod}8$

$\xi = -1, \epsilon = -1$: $d = 2, 4\mathrm{mod}8$

For examples, if we want $\xi = +1, \epsilon = +1$ then we can get the number of symmetric matrices by summing on $k = 0, 1\text{mod}4$, this will correctly give the number of symmetric matrices if

$$\cos(\frac{\pi d}{4}) + \sin(\frac{\pi d}{4}) = 1 \tag{10.23}$$

which is the case for $d = 0, 2\text{mod}8$ but not for $d = 4, 6\text{mod}8$.

****

GIVE CONDITIONS ON ϵ'

****

10.3.3 Intertwiners for $d = r + s$ odd

In this case there are two distinct irreducible representations of the same dimension. Now Γ^μ and $-\Gamma^\mu$ are *not* equivalent.

Similarly, the other intertwiners only exist in certain dimensions modulo 8:

1. Since $\Gamma^1 \cdots \Gamma^d$ is represented as a scalar $\xi(\Gamma^\mu)^{tr}$ can only be equivalent if $\xi^d(-1)^{\frac{1}{2}d(d-1)} = 1$ That is:

$\xi = +1$ and $d = 1\text{mod}4$

$\xi = -1$ and $d = 3\text{mod}4$

****

CONTINUE

****

10.4 Constructing Explicit Intertwiners from the Free Fermion Rep

Let us return to the Free fermion representation constructed in equations **** and **** above.

In our explicit basis Γ^i are real and symmetric for i odd, and imaginary ($= i\times$ real) and antisymmetric for i even. Our explicit intertwiners are

$$\begin{aligned} A\Gamma^i A^{-1} &= (\Gamma^i)^\dagger \\ B_\pm \Gamma^i B_\pm^{-1} &= \pm(\Gamma^i)^* \\ C_\pm \Gamma^i C_\pm^{-1} &= \pm(\Gamma^i)^{tr} \end{aligned} \tag{10.24}$$

We can take $A = 1$. Note that in this basis we can take $B_\pm = C_\pm$.

Let $U := \Gamma^2\Gamma^4 \cdots \Gamma^{2n}$. Then we have

$$C_+ = B_+ = \begin{cases} U & neven \\ \Gamma_\omega U & nodd \end{cases} \tag{10.25}$$ eq:plusinter

$$C_- = B_- = \begin{cases} \Gamma_\omega U & neven \\ U & nodd \end{cases} \tag{10.26}$$ eq:plusintera

To see this note that

$$\begin{aligned} U\Gamma^{2j}U^{-1} &= (-1)^{n-1}\Gamma^{2j} = (-1)^n(\Gamma^{2j})^* \\ U\Gamma^{2j-1}U^{-1} &= (-1)^n\Gamma^{2j-1} = (-1)^n(\Gamma^{2j-1})^* \end{aligned} \tag{10.27}$$

It is now straightforward to compute

$$UU^* = U^*U = \begin{cases} +1 & n = 0,3 \text{ mod}4 \\ -1 & n = 1,2 \text{ mod}4 \end{cases} \tag{10.28}$$

$$(\Gamma_\omega U)(\Gamma_\omega U)^* = (\Gamma_\omega U)^*(\Gamma_\omega U) = \begin{cases} +1 & n = 0,1 \text{ mod}4 \\ -1 & n = 2,3 \text{ mod}4 \end{cases} \tag{10.29}$$

Recall that for $d = 2n+1$ we take $\Gamma^{2n+1} = \Gamma_\omega$. Then

$$U\Gamma^{2n+1}U^{-1} = (-1)^n\Gamma^{2n+1} = (-1)^n(\Gamma^{2n+1})^* \tag{10.30}$$

and also

$$(\Gamma_\omega U)\Gamma^{2n+1}(\Gamma_\omega U)^{-1} = (-1)^n\Gamma^{2n+1} = (-1)^n(\Gamma^{2n+1})^* \tag{10.31}$$

Now, if n is even we *must* use B_+, while if n is odd we *must* use B_-.

That is:

• For the Dirac representation the intertwiner B_+ exists for $d = 0,1,2\text{mod}4$, but *not* for $d = 3\text{mod}4$.

• For the Dirac representation the intertwiner B_- exists for $d = 0,2,3\text{mod}4$, but *not* for $d = 1\text{mod}4$.

• The sign of $B_\xi B_\xi^*$, computed above, depends on $d\text{mod}8$. In this way we compute:

$d$mod8	$B_+^*B_+$	$B_-^*B_-$
0	+1	+1
1	+1	*
2	+1	−1
3	*	−1
4	−1	−1
5	−1	*
6	−1	+1
7	*	+1

The * is in the entry where the matrix does not exist.

10.5 Majorana and Symplectic-Majorana Constraints

The above representations are complex. However, we know that the irreducible representations are sometimes real. How do we see that using intertwiners in the Dirac representation?

10.5.1 Reality, or Majorana Conditions

We attempt to define a real vector subspace invariant under the Clifford action defined above. The real subspace will be defined to be the vectors satisfying the constraint

$$\psi^* = B\psi \tag{10.32}$$

eq:majorana

Notice that this equation is only consistent if $B^*B = 1$.

Mathematically, we are trying to introduce a *real structure*. That is, a $\mathbb{C}$-*anti*-linear operator $\mathcal{C}$ such that $\mathcal{C}^2 = 1$. We would define $\mathcal{C} : \psi \mapsto B\psi^*$ and the real subspace is the $+1$ eigenspace of $\mathcal{C}$.

The subspace $\mathcal{C} = 1$ defined by (10.32) is a *real* vector space. Next, to get a real representation of the Clifford algebra we need to know that it is preserved by the Clifford action. Thus we need to check that if ψ satisfies (10.32) then

$$(\Gamma^i\psi)^* \stackrel{?}{=} B\Gamma^i\psi \tag{10.33}$$

Equivalently, we want the Clifford action to commute with $\mathcal{C}$. This will only be the case if we choose B_+.

A glance at the table for $C\ell(r_+)$ confirms that we only expect such representations for $d = 0, 1, 2\text{mod}8$.

This can be confirmed using the explicit free fermion representation. Consistency of (10.32), or equivalently, $\mathcal{C}^2 = 1$ can be checked by computing the following:

$$B_+^* B_+ = \begin{cases} +1 & d = 0, 1, 2 \text{ mod}8 \\ -1 & d = 4, 5, 6 \text{ mod}8 \end{cases} \tag{10.34}$$

eq:majcond

$$B_-^* B_- = \begin{cases} +1 & d = 6, 7, 8 \text{ mod}8 \\ -1 & d = 2, 3, 4 \text{ mod}8 \end{cases} \tag{10.35}$$

eq:majcondi

As shown in table *** above. (THIS IS REDUNDANT!)

10.5.2 Quaternionic, or Symplectic-Majorana Conditions

In the case of $d_T = 3, 4, 5\text{mod}8$ we have a representation by quaternionic matrices. We can always represent a quaternion as a 2×2 complex matrix A such that $A^* = \sigma^2 A \sigma^2$.

In this way we convert $2^N \times 2^N$ quaternionic matrices to $2^{N+1} \times 2^{N+1}$ complex matrices such that

$$\Gamma^{\mu,*} = J\Gamma^\mu J^{-1} \tag{10.36}$$

with $J = i\sigma^2 \oplus \cdots \oplus i\sigma^2$. Thus, the complex conjugate representation matrices are equivalent to conjugation with J. This is the case of $\xi = +1$ and $B_\xi^* B_\xi = -1$.

Now, vectors in the irreducible representation are N-component vectors of quaternions where $N = 2^{\ell-1}$. Therefore, when we write in terms of complex vectors we have

$$\begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix} \to \begin{pmatrix} z_1 & w_1 \\ -\bar{w}_1 & \bar{z}_1 \\ \vdots & \vdots \\ z_N & w_N \\ -\bar{w}_N & \bar{z}_N \end{pmatrix} \tag{10.37}$$

Thus, we get a *pair* of complex spinors of dimension 2^{N+1} from the two columns. Moreover, denoting these two columns by ψ_1 and ψ_2 we see they are related by

$$\psi_1^* = J\psi_2 \tag{10.38}$$

We can introduce a spinor-index notation for the $SU(2)$ subgroup of the quaternions with a, b running over $1, 2$ and then we can write our condition as

$$(\psi_{i,a})^* = \epsilon^{ab} J_{ij} \psi_{jb} \tag{10.39}$$ eq:sympmajorar

where $\epsilon^{12} = +1$. Note that $J = J^*$, and $JJ^* = J^2 = -1$, as is required by consistency with (10.39).

Conversely, when there is an intertwiner $B_\xi B_\xi^* = -1$ we can introduce a "symplectic Majorana" spinor by taking a *pair* of spinors ψ_a, $a = 1, 2$, and imposing the reality condition:

$$(\psi^*)^a := (\psi_a)^* = \epsilon^{ab} B_\xi \psi_b \tag{10.40}$$ eq:sympmajii

Consistency of this equation requires $B_\xi B_\xi^* = -1$. Again, the subspace defined by (10.40) will only be preserved by Γ^μ, that is

$$(\Gamma^\mu \psi_a)^* = \epsilon^{ab} B_\xi \Gamma^\mu \psi_b, \tag{10.41}$$

if we choose $\xi = +1$. As we see from (10.39) above, this is the same as saying that the representation is quaternionic.

Remark: More generally, in the physics literature if Ω^{ab}, $a, b = 1, \ldots, u$ is an antisymmetric symplectic matrix, so $\Omega^2 = -1$ we can impose a condition:

$$(\psi^*)^a := (\psi_a)^* = \Omega^{ab} B \psi_b \tag{10.42}$$

and these are also referred to as symplectic Majorana spinors. This is useful in 6d supersymmetry for reasons connected with the tensor product of spin reps we come to next.

10.5.3 Chirality Conditions

When we come to representations of the *Spin group* we will see that we want to represent just the *even* Clifford algebra. Then we might wish to impose both a Majorana *and* a Weyl condition. This is only possible if $d = 2n$ is even. In this case we have

$$P_\pm = \frac{1}{2}(1 \pm \Gamma_\omega) \tag{10.43}$$

and we compute

$$B P_\pm B^{-1} = \begin{cases} P_\pm & d = 0 \bmod 4 \\ P_\mp & d = 2 \bmod 4 \end{cases} \tag{10.44}$$ eq:bonjc

Combining with (10.34)(10.35) we see that we can impose Majorana and Weyl for $d = 0 \bmod 8$, that is, the chiral representation is self-conjugate.

If $d = 2, 6 \bmod 8$ then complex conjugation switches the chirality.

As we have explained above, when working with field theories of fermions one also needs to know the transpose properties because it is sometimes important to know the (anti)symmetry of $C_{\pm}\Gamma^{\mu_1\cdots\mu_k}$. These are easily obtained from

$$U^{tr} = (-1)^{\frac{1}{2}n(n+1)}U \tag{10.45}$$

$$(\Gamma_\omega U)^{tr} = (-1)^{\frac{1}{2}n(n-1)}\Gamma_\omega U \tag{10.46}$$

Note that the signs depend on $n$mod4 and hence on $d$mod8.

Exercise
Repeat this exercise for Lorentzian signature.

---

11. $\mathbb{Z}_2$ graded algebras and modules

11.1 $\mathbb{Z}_2$ grading on the algebra

As we mentioned at the beginning of this chapter, there is a natural $\mathbb{Z}_2$ grading defined on generators by $\lambda : e^\mu \to -e^\mu$, and then extending λ so that it is an algebra homomorphism. This makes the Clifford algebra a $\mathbb{Z}_2$-graded algebra (i.e. a superalgebra):

$$C\ell(Q) \cong C\ell^0(Q) \oplus C\ell^1(Q) \tag{11.1}$$

where $C\ell^0(Q)$ is the subalgebra generated by products of even numbers of e^μ's

Let us stress that for d odd the decomposition

$$\mathbb{C}\ell(d) \cong \mathbb{C}(2^{[d/2]}) \oplus \mathbb{C}(2^{[d/2]}) \tag{11.2}$$ eq:dirsumalg

is *not* compatible with the $\mathbb{Z}_2$-grading, that is, with the decomposition

$$\mathbb{C}\ell(d) \cong (\mathbb{C}\ell(d))^0 \oplus (\mathbb{C}\ell(d))^1 \tag{11.3}$$ eq:dirsumgrad

The volume form ω_c is odd, and therefore $P_\pm$ do not commute with the grading.

Using the form (11.2) of the Clifford algebra we can consider to be the subalgebra of $\mathbb{C}(2N)$ where $N = 2^{[d/2]}$ of matrices of the form

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \tag{11.4}$$

where $a, b \in \mathbb{C}(N)$. In this block decomposition we have

$$\omega_c = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{11.5}$$

Since ω_c is odd, we see this block decomposition is *not* (11.3). In the block decomposition (11.2) the grading operator is given by *conjugation* with

$$\begin{pmatrix} 0 & 1_N \\ 1_N & 0 \end{pmatrix} \tag{11.6}$$

and hence the even subalgebra consists of matrices with $a = b$ and the odd subspace consists of matrices of the form $a = -b$.

Sometimes we prefer to use the block decomposition (11.3). In this decomposition the Clifford algebra consists of matrices of the form

$$\begin{pmatrix} A & B \\ B & A \end{pmatrix} \tag{11.7}$$

eq:gradcd

with $A = \frac{1}{2}(a+b) \in \mathbb{C}(N)$ and $B = \frac{1}{2}(a-b) \in \mathbb{C}(N)$. Note that the set of matrices (11.7) is indeed a subalgebra of $\mathbb{C}(2N)$:

$$\begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} A' & B' \\ B' & A' \end{pmatrix} = \begin{pmatrix} AA' + BB' & AB' + BA' \\ BA' + AB' & AA' + BB' \end{pmatrix} \tag{11.8}$$

The even subalgebra corresponds to the matrices of the type

$$\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \tag{11.9}$$

and is indeed isomorphic to $\mathbb{C}(N)$.

In the block decomposition (11.3) the volume form ω_c is odd and we can take it to be

$$\omega_c = \begin{pmatrix} 0 & 1_N \\ 1_N & 0 \end{pmatrix} \tag{11.10}$$

where $N = 2^{[d/2]}$, While the grading operator is *conjugation* with $\Pi = (-1)^{\deg}$ where

$$\Pi = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{11.11}$$

We now explore some consequences of this $\mathbb{Z}_2$-grading.

11.2 The even subalgebra $C\ell^0(r_+, r_-)$

The first important remark to make is that the even subalgebra is isomorphic to another Clifford algebra. To see this, choose any basis element e^{μ_0} and consider the algebra generated by

$$\tilde{e}^{\nu} = e^{\mu_0 \nu} \qquad \nu \neq \mu_0 \tag{11.12}$$

Note that

$$\tilde{e}^{\nu}\tilde{e}^{\rho} + \tilde{e}^{\rho}\tilde{e}^{\nu} = -2\eta^{\mu_0\mu_0}\eta^{\nu\rho} \qquad \nu, \rho \neq \mu_0 \tag{11.13}$$

and therefore [4]

$$\begin{aligned} C\ell^0(r_+, s_-) &\cong C\ell(r_+, (s-1)_-) \qquad s \geq 1 \\ C\ell^0(r_+, s_-) &\cong C\ell(s_+, (r-1)_-) \qquad r \geq 1 \end{aligned} \tag{11.14}$$

eq:twoevens

Over the complex numbers we have $\mathbb{C}\ell^0(d) \cong \mathbb{C}\ell(d-1)$.

Example: The even subalgebra of $C\ell(2,0)$ is the algebra of matrices:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{11.15}$$

and is isomorphic to $\mathbb{C}$.

Remarks

• This observation is useful when discussing $\mathbb{Z}_2$-graded representations and when we discuss representations of the Spin group.

•

Exercise

Show that when both $r \geq 1$ and $s \geq 1$ then the two equations in (11.14) are compatible.

Exercise

Show that

$$(C\ell(r_+, s_-))^0 \cong (C\ell(s_+, r_-))^0 \tag{11.16}$$

11.3 $\mathbb{Z}_2$ graded tensor product of Clifford algebras

Since the Clifford algebra is $\mathbb{Z}_2$ graded we can also define a graded tensor product. Of course

$$(C\ell(Q_1)\hat{\otimes}C\ell(Q_2))^0 \cong (C\ell(Q_1))^0 \otimes (C\ell(Q_2))^0 \oplus (C\ell(Q_1))^1 \otimes (C\ell(Q_2))^1 \tag{11.17}$$

as vector spaces (with a similar formula for the odd part).

[4] Note that this implies that we must have $C\ell((r+1)_+, s_-) \cong C\ell((s+1)_+, r_-)$ for all $r, s \geq 0$. One can indeed prove this is so using the periodicity isomorphisms and the observation that $C\ell(2_+) \cong C\ell(1_+, 1_-) \cong \mathbb{R}(2)$. Nevertheless, at first site this might seem to be very unlikely since the transverse dimensions are $r-s-1$ and $s-r-1$ and in general are *not* equal modulo 8. Note that the sum of the transverse dimensions is $-2 = 6\text{mod}8$. Thus, we have the pairs $(0,6)$, $(1,5)$, $(2,4)$, and $(3,3)$. One can check from the table that these all do in fact have the same Morita type! Of course, the dimensions are the same, so they must in fact be isomorphic.

The important new point is the sign rule in the Clifford multiplication on the graded tensor product. We define the graded Clifford multiplication to be:

$$(\phi\hat{\otimes}\xi)\cdot(\phi'\hat{\otimes}\xi') := (-1)^{\deg\xi\cdot\deg\phi'}\phi\phi'\hat{\otimes}\xi\xi' \tag{11.18}$$

eq:gradedcliff

Quite generally, if $V = V_1 \oplus V_2$ and $Q = Q_1 \oplus Q_2$ then

$$C\ell(Q) = C\ell(Q_1)\hat{\otimes}C\ell(Q_2) \tag{11.19}$$

eq:grdedtenspr

It follows that

$$C\ell(r_+, s_-) = C\ell(1_+)^{\hat{\otimes}r}\hat{\otimes}C\ell(1_-)^{\hat{\otimes}s} \tag{11.20}$$

Note that (11.19) is *not* true for the ungraded tensor product.

The graded tensor product is not as useful for constructing representations because when one takes a tensor product of matrices one normally uses the ordinary tensor product. This is why we used the ordinary tensor product to construct the structure of the algebras and discuss their representations above.

11.4 $\mathbb{Z}_2$ graded modules

A $\mathbb{Z}_2$-graded module M for the Clifford algebra is a representation which itself a $\mathbb{Z}_2$-graded vector space

$$M = M^0 \oplus M^1 \tag{11.21}$$

such that the Clifford action respects the $\mathbb{Z}_2$ grading. Thus,

$$C\ell(Q)^A \cdot M^B \subset M^{A+B\mathrm{mod}2} \tag{11.22}$$

A $\mathbb{Z}_2$-graded module said to be of graded-dimension $(\mathrm{dim}M^0|\mathrm{dim}M^1)$.

One useful way to think about a grading is the following: A $\mathbb{Z}_2$ grading means we have an operator Π on M with $\Pi^2 = 1$ such that Π anti-commutes with all the Clifford generators γ_i. So, Π can be considered to be another gamma-matrix!

Let us now examine the $\mathbb{Z}_2$-graded modules of the Clifford algebras. We begin with the complex Clifford algebras for simplicity.

It is very useful to speak of the free abelian group generated by Clifford modules. The direct sum of modules is a module so we can let

$$nM := M^{\oplus n} \tag{11.23}$$

for $n > 0$. Then we introduce formally an additive inverse so that we can write

$$nM + n'M = (n + n')M \tag{11.24}$$

for *all* integers $n, n' \in \mathbb{Z}$.

Let $\mathcal{M}_k^c$ denote the free abelian group of graded $\mathbb{C}\ell_k$ modules and $\mathcal{N}_k^c$ denote the free abelian group of ungraded $\mathbb{C}\ell_k$ modules

Before giving the general construction let us look at some low-dimensional examples. It will also be interesting to consider the following problem: We can consider $\mathbb{C}\ell_k^c$ to be

a subalgebra of $\mathbb{C}\ell^c_{k+1}$ by identifying the first k generators of the latter algebra to be the generators $e_1, \ldots, e_k$ of the former. Therefore, every $\mathbb{C}\ell^c_{k+1}$ module can be considered to be a $\mathbb{C}\ell^c_k$ module, and hence we have a group homomorphism $\iota^* : \mathcal{M}^c_{k+1} \to \mathcal{M}^c_k$. The problem – of great utility in topology - is to determine the abelian group

$$\mathcal{M}^c_k / \iota^*(\mathcal{M}^c_{k+1}) \tag{11.25}$$

(For some idea of why this might be so, see section *** above and section *** below.)

We begin with zero dimensions. $\mathbb{C}\ell(0) = \mathbb{C}$ has one (irreducible) ungraded module, $\mathbb{C}$. We will refer to it as N_0 below. However, $\mathbb{C}\ell(0)$ has *two* irreducible graded modules $M_{0,+}$ of dimension $(1|0)$ and $M_{0,-}$ of dimension $(0|1)$. Thus,

$$\begin{aligned} \mathcal{N}_0 &\cong \mathbb{Z} \\ \mathcal{M}_0 &\cong \mathbb{Z} \oplus \mathbb{Z} \end{aligned} \tag{11.26}$$

Now consider $\mathbb{C}\ell(1) \cong \mathbb{C} \oplus \mathbb{C}$. Choosing the generator so $e_1 e_1 = -1$ the two subalgebras are obtained by the projection operators $P_\pm = \frac{1}{2}(1 \pm ie)$. Thus note that neither subalgebra has homogeneous grading.

Now, $\mathbb{C}\ell(1)$ has two inequivalent ungraded modules, denoted $N_{1,\pm}$ depending on whether we represent e by $\mp i$. That is $N_{1,\pm} = P_\pm C\ell^c_1$. Each module $N_{1,\pm}$ is a one-dimensional complex vector space. On the other hand, $\mathbb{C}\ell(1)$ has only one irreducible graded module. We will denote it M_1. An even vector must be mapped by e_1 to an odd vector, and vice versa. Thus M_1 has dimension $(1|1)$ and is $\mathbb{C} \oplus \mathbb{C}$ as a vectorspace with e_1 represented by

$$\rho(e_1) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{11.27}$$

Now the even subalgebra $(\mathbb{C}\ell(1))^0$ is just $\mathbb{C}$, and is generated by the scalars.

It is useful to note that we could define M_1 by the formula

$$M_1 = \mathbb{C}\ell(1) \otimes_{(\mathbb{C}\ell(1))^0} N_0 \tag{11.28}$$

A vector space basis is $v_1 = 1 \otimes 1$ generating the even subspace $(M_1)^0$ and $v_2 = e \otimes 1$, generating the odd subspace $(M_1)^1$, and it is clear that in this basis

$$e \to \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix} \tag{11.29}$$

Note that, as is required, e is an odd operator - it is off-diagonal. Thus,

$$\begin{aligned} \mathcal{N}_1 &\cong \mathbb{Z} \oplus \mathbb{Z} \\ \mathcal{M}_1 &\cong \mathbb{Z} \end{aligned} \tag{11.30}$$

We can also determine the homomorphism $\iota^* : \mathcal{M}_1 \to \mathcal{M}_0$. Since the generator M_1 has dimension $(1|1)$ the map $\mathbb{Z} \to \mathbb{Z} \oplus \mathbb{Z}$ is just $n \to (n, n)$.

Now let us turn to $\mathbb{C}\ell(2) \cong \mathbb{C}(2)$. Clearly, there is only one irreducible ungraded module $N_2 \cong \mathbb{C}^2$. However, there are two inequivalent $\mathbb{Z}_2$ graded modules: $M_2^\pm$. To see

this first note that the even subalgebra is $(\mathbb{C}\ell(2))^0 \cong \mathbb{C} \oplus \mathbb{C}$ the isomorphism being given by $P_\pm = \frac{1}{2}(1 \mp ie_1e_2)$. But there are two inequivalent ungraded modules for $(\mathbb{C}\ell(2))^0$ and these induce inequivalent graded modules via

$$M_2^\pm = \mathbb{C}\ell(2) \otimes_{(\mathbb{C}\ell(2))^0} N_{1,\pm} \tag{11.31}$$

As a vector space $M_2^\pm \cong \mathbb{C}^2$. We can choose a vector space basis $v_1 = 1 \otimes 1$, generating the even part $(M_2^\pm)^0$ and $v_2 = e_1 \otimes 1$, generating the odd part $(M_2^\pm)^1$. Note $e_2 \otimes 1$ is not a linearly independent vector because

$$\begin{aligned} e_2 \otimes 1 &= -e_1(e_1e_2) \otimes 1 \\ &= -e_1 \otimes e_1 \cdot 1 \qquad \text{definition of tensor product} \\ &= e_1 \otimes \mp i \\ &= \mp i e_1 \otimes 1 \end{aligned} \tag{11.32}$$

A small computation shows that in this basis

$$e_1 \to \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad e_2 \to \begin{pmatrix} 0 & \mp i \\ \mp i & 0 \end{pmatrix} \tag{11.33}$$

eq:cliffacttwo

If we normalize the volume form so it squares to 1, $\omega = ie_1e_2$, then, although both the graded and ungraded modules are $\mathbb{C}^2$ as vector spaces, the two inequivalent graded modules are distinguished by the sign of ω acting on the ungraded component $(M_2^\pm)^0$.

Thus

$$\begin{aligned} \mathcal{N}_2 &\cong \mathbb{Z} \\ \mathcal{M}_2 &\cong \mathbb{Z} \oplus \mathbb{Z} \end{aligned} \tag{11.34}$$

Moreover, we can determine $\iota^* : \mathcal{M}_2 \to \mathcal{M}_1$ easily since (11.33) gives the action of e_1, on $M_{2,\pm}$ which is identical to that of $\mathbb{C}\ell_1^c$ on M_1. Thus the map is $(1,0) \to 1$ and $(0,1) \to 1$.

Finally, let us look at $\mathbb{C}\ell(3) \cong \mathbb{C}(2) \oplus \mathbb{C}(2)$, with projectors $P = \frac{1}{2}(1 \mp e_1e_2e_3)$. The even subalgebra $(\mathbb{C}\ell(3))^0 \cong \mathbb{C}(2)$ is generated by, say, $\hat{e}_1 = e_1e_2$ and $\hat{e}_2 = e_2e_3$. Clearly $\mathbb{C}\ell(3)$ has two ungraded modules, isomorphic to $\mathbb{C}^2$ as vector spaces. We can call these $N_{3,\pm}$. Explicitly, they are represented by $e_j \to \pm i\sigma^j$ where σ^j are Pauli matrices. The two different modules differ by the *sign* of the volume element. On the other hand, there is only one graded module, which we can again take to be

$$M_3 = \mathbb{C}\ell(3) \otimes_{(\mathbb{C}\ell(3))^0} N_2 \tag{11.35}$$

M_3 is $\mathbb{C}^4$ as a vector space. to be explicit we can choose a basis v_1, v_2 for N_2 as above, then we have a representation of $\hat{e}_1 = e_1e_2, \hat{e}_2 = e_2e_3$, as above. Then we choose

$$w_1 = 1 \otimes v_1, w_2 = 1 \otimes v_2, w_3 = e_1 \otimes v_1, w_4 = e_1 \otimes v_2 \tag{11.36}$$

Note that

$$e_2 \otimes v_1 = -e_1e_1e_2 \otimes v_1 = -e_1 \otimes (e_1e_2)v_1 = -e_1 \otimes v_2 \tag{11.37}$$

and in a similar fashion we get

$$\begin{aligned} e_2 \otimes v_2 &= e_1 \otimes v_1 \\ e_3 \otimes v_1 &= -ie_1 \otimes v_1 \\ e_3 \otimes v_2 &= -ie_1 \otimes v_2 \end{aligned} \tag{11.38}$$

so we compute

$$\begin{aligned} e_1 : w_1 &\to w_3 \\ w_2 &\to w_4 \\ w_3 &\to -w_1 \\ w_4 &\to -w_2 \end{aligned} \tag{11.39}$$

etc.

Exercise

Show that restriction of M_3 as a graded $\mathbb{C}\ell(2)$ module is

$$\iota^*(M_3) = M_{2,+} \oplus M_{2,-} \tag{11.40}$$

Now we turn to the general story:

$\mathbb{C}\ell(2n) \cong \mathbb{C}(2^n)$ has

1. Unique ungraded irrep N_{2n}, $\cong \mathbb{C}^{2^n}$ as a vector space.

2. Two inequivalent graded irreps $M_{2n,\pm}$, $\cong \mathbb{C}^{2^{n+1}}$ as vector spaces. [??? Aug. 28, 2011: Dimension seems wrong. ???].

while $\mathbb{C}\ell(2n+1) = \mathbb{C}(2^n) \oplus \mathbb{C}(2^n)$ has

1. Two inequivalent ungraded irreps $N_{2n+1,\pm}$, $\cong \mathbb{C}^{2^n}$ as vector spaces.

2. A unique graded irrep M_{2n+1}, $\cong \mathbb{C}^{2^{n+1}}$ as a vector space.

That is:

$$\begin{aligned} \mathcal{N}_{2k} &\cong \mathbb{Z} \\ \mathcal{N}_{2k+1} &\cong \mathbb{Z} \oplus \mathbb{Z} \end{aligned} \tag{11.41}$$

$$\begin{aligned} \mathcal{M}_{2k} &\cong \mathbb{Z} \oplus \mathbb{Z} \\ \mathcal{M}_{2k+1} &\cong \mathbb{Z} \end{aligned} \tag{11.42}$$

Moreover, if $\iota : \mathbb{C}\ell(n) \to \mathbb{C}\ell(n+1)$ then we examine $\iota^*(M)$ or $\iota^*(N)$, the restriction of the module to the Clifford subalgebra, and we have

$$\iota^*(M_{2n+1}) = M_{2n,+} \oplus M_{2n,-} \tag{11.43}$$

$$\iota^*(M_{2n,\pm}) = M_{2n-1} \tag{11.44}$$

If M is a graded module for $\mathbb{C}\ell(k+1)$ then M^0 is a graded module for $(\mathbb{C}\ell(k+1))^0 \cong \mathbb{C}\ell(k)$. The inverse relation is that if N is an ungraded module for $\mathbb{C}\ell(k)$ then we can produce a graded module for $\mathbb{C}\ell(k+1)$ by

$$M = \mathbb{C}\ell(k+1) \otimes_{(\mathbb{C}\ell(k+1))^0} N \tag{11.45}$$

Remarks

• To make contact with our discussion of representations in section ****, note the following: For d odd, ω^c is central and there are two inequivalent representations of $\mathbb{C}\ell(d)$ depending on the sign of ω^c. For d even, there is only one inequivalent representation. These are the ungraded representations. For the graded representations we can look at the even subalgebra acting on the even space M^0. Now for d even, the even subalgebra has a volume form ω_c^0 which is central in the even subalgebra and hence acting on the even part of the module, M^0, has ω_c is represented by a scalar ± 1. This defines the two kinds of inequivalent graded modules. For d odd the even subalgebra has no such central term and there is only one inequivalent graded irreducible graded module.

• There is an analogous, but more intricate, discussion for the modules of the real Clifford algebra.

11.5 K-theory over a point

****

Explain about brane/antibrane and how invertible tachyon fields produce annihilation.
****

One can thus consider $\mathbb{Z}_2$-graded $\mathbb{C}\ell_n^c$-modules with an odd anti-hermitian operator.
Need to mod out by invertible operators, and homotopy relation....
Thus we consider the space

$$K^{-j}(pt) = \mathcal{M}_j / \iota^*(\mathcal{M}_{j+1}) \tag{11.46}$$

So $M_{2n,+} \sim -M_{2n,-}$ in K-theory
From the above

$$K^{-j}(pt) = \begin{cases} \mathbb{Z} & j = 0 \text{mod} 2 \\ 0 & j = 1 \text{mod} 2 \end{cases} \tag{11.47}$$

11.6 Graded tensor product of modules and the ring structure

and $M_{2n,+} \cong (M_{2,+})^{\otimes n}$ as tensor products of $\mathbb{Z}_2$-graded modules. So $M_{2,+}$ is the Bott element $\pm u^{-1}$.

Moreover, as we have seen $C\ell^0((r+1)_+) \cong C\ell(r_+)$. Thus we can list the free abelian groups generated by irreducible $\mathbb{Z}_2$ graded modules as:

11.6.1 The Grothendieck group

Now describe Z_2-graded modules modulo those which extend to Z_2 graded modules one dimension higher.

Tensor product of $\mathbb{Z}_2$-graded modules. Ring structure of the Grothendieck group.

12. Clifford algebras and the division algebras

There is an intimate relation between Clifford algebras and the division algebras. See Kugo-Townsend. Also see

A. Sudbury, "Division algebras, (pseudo)orthogonal groups and spinors," J. Phys. A Math. Gen. **17** (1984) 939.

Interesting constructions of the Clifford algebras for $Spin(7)$ and $Spin(8)$ using the mutliplication table of the octonions can be found in:

M. Gunaydin and F. Gursey, J. Math. Phys. **14** (1973)1651

R. Dundarer, F. Gursey, and C.H. Tze, J. Math. Phys. **25** (1984)1496

13. Some sources

Some references:

1. E. Cartan, *The theory of Spinors*

2. Chevalley,

2'. P. Deligne, "Notes on spinors," in *Quantum Fields and Strings: A Course for Mathematicians*

3. One of the best treatments is in Atiyah, Bott, and Shapiro, "Clifford Modules"

4. A textbook version of the ABS paper can be found in Lawson and Michelson, *Spin Geometry*, ch.1

5. Freund, Introduction to Supersymmetry

6. M. Sohnius, "Introducing Supersymmetry" Phys. Rept.

7. T. Kugo and P. Townsend, "Supersymmetry and the division algebras," Nuc. Phys. **B221** (1983)357.

8. M. Rausch de Traubenberg, "Clifford Algebras in Physics," arXiv:hep-th/0506011.