Preface

Time to gather stones...

In his well-known popular lectures R. Feynman [1] reflects on the way physical theories are built up and distinguishes two such ways or, rather, two stages in the process of their construction: *(i)* the "Babylonian" stage and *(ii)* the "Greek" stage.

It is not difficult to guess that the term "Babylonian" refers to ancient Babylon and the corresponding physical theory is just geometry. A Babylonian geometer (the words "mathematician" or "physicist" were not yet coined) knew many facts about circles, triangles, and other figures, and his understanding was not purely empirical because he could also *relate* different such facts with each other. He could e.g. heuristically derive a fact A on the basis of some other facts B and C or maybe the fact C on the basis of B and some other fact D, etc. He was actually quite efficient in his main job: the calculation of the areas of land estates. In other words, his theory described the observed experimental facts well and had direct practical applications.

Our Babylonian colleague was lacking, however, a consistent structured system in which a set of basic simple facts are chosen as *axioms* and all others are rigorously derived as *theorems*. Constructing such a system meant the advanced "Greek" stage of understanding geometry. At this stage a physical theory becomes a branch of mathematics where the primary criterion for Truth is not Experiment, but rather Rigor of derivations and internal logical Consistency. Feynman writes that a modern physicist is a

Babylonian rather than a Greek in this respect: he does not care too much about Rigor, and his God and ultimate Judge is Experiment.

Strictly speaking, this is not quite correct (a general statement never is). Some branches of classical and also of quantum physics have now quite reached the Greek stage. As a result, many problems concerning classical and quantum dynamics of Hamiltonian systems are studied now in mathematics departments rather than in physics departments. If you want, rendering a theory understandable to mathematicians may be considered a final goal of research in physics.

Regarding quantum chromodynamics and quantum field theory in general, we are living now in interesting times when we go over from the Babylonian to the Greek stage. A basic concept in quantum field theory is the path integral. Mathematicians still do not understand what it is: they cannot provide a rigorous definition for it. My personal impression, however, is that such a mathematical definition will be given soon.

A transitional period when a physical theory is already essentially constructed and understood at a semi-heuristic level but has not become yet a branch of mathematics, is a proper time to write textbooks. And many such textbooks were written recently. Is there really a need to write another one? It is difficult for me to judge whether this particular book is worth reading, but if you ask me why I have written it, the answer is the following:

There are very good recent monographs [2; 3] on quantum field theory. They are modern and comprehensive but are rather voluminous, and this might be considered a disadvantage. The large size of these books is due to two major choices which their authors make:

- The basics of quantum field theory are described from scratch. No preliminary acquaintance with the subject is required.
- *Many* different quantum field theories are considered there: quantum electrodynamics, quantum chromodynamics, and electroweak theory. The third volume of Weinberg's book is devoted to supersymmetric theories.

Besides general comprehensive monographs, there are also books (cf. [4]–[7]) written in a different style and specially devoted to Yang–Mills theory and to QCD. All these books are "recommended reading" and many pages in my own book were actually written under their influence. That especially concerns Faddeev's and Slavnov's book which was one of my own

primary sources when learning QCD.

All these books concentrate, however, on the *perturbative* aspects of QCD, whereas my primary goal is to convey to the reader what is currently known about *nonperturbative* QCD dynamics. In particular, I discuss in some details the instanton solution, the nonperturbative (*viz.* lattice) definition of the path integral including also the "Ginsparg–Wilson revolution" — the modern way to define fermionic path integrals. I further discuss the θ vacuum picture, theoretical and phenomenological aspects of quark–hadron duality, the QCD phase transitions brought about by heat or by squeeze, and, last but not least, confinement. To make the book reasonably self-contained, the perturbative dynamics of QCD is also considered, however. We outline the derivation of the Feynman rules, dwell upon the ultraviolet regularization and renormalization issues, and also review the structure of infrared and collinear singularities in QCD.

Only quantum chromodynamics is discussed * and we assume the reader to be already familiar with QED, the book [8] being our main reference point. That means the acquaintance with the standard operator approach to quantization, derivation of the Feynman rules using Wick's theorem, calculation of Feynman integrals, etc. We describe, however, path integral quantization in field theory and in quantum mechanics — a material not covered by the book [8], which can, of course, be found in almost *any* modern textbook on field theory.

We put emphasis on the principal foundational aspects of QCD and provide only a limited discussion of QCD phenomenology. Actually, one of our assumptions is that the reader has a rough idea about hadron phenomenology, knows about flavor $SU(3)$ symmetry (still, the most basic facts about unitary groups are given in the Appendix) and its implications for the spectrum of mesons and baryons, has heard about confinement, etc.

All these limitations together with a concise manner of presentation make the book relatively thin (which was one of my goals). Being influenced by the style of the Landau course, I have put in a number of **Problems** accompanied by **Solutions**. Each such **Problem** examines a particular (usually, technical and, more often than not, relatively simple) theoretical issue.

*To be more exact, we discuss sometimes QED and also some hypothetical (not describing the real World) field theories, but we do it only to clarify and to illustrate what happens in QCD. Purely theoretical issues having no direct implications for QCD are not discussed.

The book does not contain many references. I have decided *not* to give references to classical original papers whose results are now well known and are discussed also in other books on QCD. Vanity-ridden, I sometimes do not refer to my own papers. I tried to quote, however, the *names* of the people who obtained the results I write about. Also, I give references to some less known papers, or to papers which are known well but are more recent and did not get into the textbooks yet. In other words, the references are quite chaotic and always (with one exception) have a nature of personal recommendation to *read* this or that particular book or a paper.

This book has grown up from the lecture course on the foundations of QCD which I have given for graduate students in high energy theory in ITEP in Moscow during the Spring of 1998, and which I later put in writing. Of course, when actually writing the book, I have added much more material than I was able to squeeze into the real time lectures and I also added some "off-shell" lectures which were never actually given. Thus, the book which you hold now in your hands differs from the original draft in an essential way. Still, I have decided to preserve the form of the virtual lecture course.

It is a pleasure to thank C. Bender, M. Chernodub, J. Hoppe, M. Lavelle, P. Mansfield, M. Peskin, K. Selivanov, M. Shifman, A. Slavnov, A. Vainshtein, and L. Yaffe for discussions and useful remarks concerning the book. Special thanks are due to M. Staudacher, who took pains to read carefully the whole manuscript, for many valuable remarks and comments and to my wife E. Savvina for her very essential help with editing the manuscript. I appreciate also the artistic contribution of Olga Feigelman who has drawn the headpieces for the parts.

Introduction: Some History

Quantum Chromodynamics is the fundamental theory of the strong interactions. In everyday life, the strong interactions have the important task of binding protons and neutrons inside the nucleus. Back in 1932, immediately after the neutron had been discovered, it was realized that the forces keeping together the nucleons in nuclei have nothing to do with the electromagnetic forces, which are much weaker. The official birthdate of QCD is 1974 — it took therefore more than 40 years to understand what is the fundamental theory describing the dynamics of the nuclear forces and to explain the immense bulk of the accelerator data accumulated by that time.

The gradual progress in our understanding of the dynamics of nucleons and other strongly interacting particles, called *hadrons*,[†] cannot be visually represented by a reasonably straight line. It may be better described by a kind of fractal which, starting from Ignorance and passing through innumerable Fallacies, ends up eventually at the predestined final point of Truth.

Maybe we would still be far from the truth were it not for the constant pressure exerted by experimental data. QCD was created while constantly profiting from genuinely physical feedback. Unfortunately, this was probably the last example of such an efficient cooperation between theorists and experimentalists. Those fundamental laws of Nature which are still not known refer to a scale of energies currently not accessible to experiment. As for the most intriguing question of what is going on at the Planck scale $\sim 10^{19}$GeV — it presumably will *never* be answered experimentally. We

[†]This term coined by L. Okun is derived from the Greek *hadros*, which means "bulky" as opposed to *leptos* meaning "small".

will have to guess the fundamental Theory of Everything without much experimental feedback. The situation is difficult but, of course, not hopeless. Everybody knows an example when such a purely theoretical quest was successful: the (classical) theory of gravitation (or general relativity). It was created by Einstein in 1916 by *fiat*, confirmed by experiment soon thereafter, and, since then, there has been little essential progress in further understanding gravity.

Going back to the "fractal" history of QCD, this fractal smoothens out in our perception as time passes by and the details which are no longer relevant are forgotten. It is still wiggly enough, but its length seems to be finite and the main line of development can (and will) be described in the few pages that follow.

The first theory of strong interactions was constructed by Yukawa in 1935. By analogy with QED, where the interaction between charged particles is mediated by electromagnetic fields whose quanta are photons, Yukawa assumed that the interaction of the nucleons is mediated by some new field whose quanta (the *mesons*) represent a new kind of particle. This new theory differed from QED in three points:

- The electromagnetic field is vectorial in nature. This leads to the repulsion of same-sign charges, and only charges of opposite signs attract. However, the nucleus involves only one kind of particle (in two isotopic forms) and the interaction is attractive. That can only be realized if the quanta of the mediating field have even spin. The simplest possibility is that the spin is zero.
- Photons are massless and the Coulomb interaction is long-range. But the nuclear forces have the short range of order 1 fm. That means that the "meson" should be massive, with mass μ being of order $(1\text{ fm})^{-1} \sim 200$ MeV.
- We are dealing with *strong* interactions, and this means that the meson–nucleon coupling constant (the analog of electric charge) is large.

The theory of Yukawa was brilliantly confirmed in 1947 when the meson with the required properties was discovered in cosmic rays.[‡] That was, of course, the π meson. Three charged states of the π meson (π^+, π^0, and π^-)

[‡]Before that the muon was discovered and the attempts to associate it with the Yukawa mesons lead to considerable confusion. This is one of the "wiggles" we disregard here.

were found which meant that it is an isovector rather than an isoscalar (the apparatus of the isotopic symmetry was already well developped by that time). Also, it turned out to be pseudoscalar rather than scalar. Bearing all this in mind, the Lagrangian of the meson–nucleon theory takes the form

$$\mathcal{L}_{\text{meson}} = i\bar{N}\partial\!\!\!/ N - m\bar{N}N + \frac{1}{2}(\partial_\mu\phi)^2 - \frac{\mu^2}{2}\phi^2 - ig\bar{N}\gamma^5\tau_i N\phi_i \ , \qquad (0.1)$$

where $\tau_{1,2,3}$ are the Pauli matrices.

Soon after that, it was understood that, even though the theory (0.1) certainly correctly describes some features of nuclear interactions, it cannot be the final truth. First of all, after the discovery of the π mesons many more mesons and baryons were discovered, first in cosmic rays and then in accelerator experiments. The simple Lagrangian (0.1) had no place for them.

Secondly, the theory (0.1) had severe problems of a purely theoretical nature. The main problem was that the coupling constant g was rather big ($g^2/4\pi \sim 14$ which is ~ 1000 times larger than $e^2/4\pi \equiv \alpha$). As a result, practically no calculations were possible. Indeed, at that time (we are in the fifties now) the only known way to work with quantum field theory was to use the perturbative expansions. This works marvellously well for QED, where the expansion parameter is small, but fails completely in the case of the meson theory. Furthermore, at the beginning of the fifties, the so called zero charge phenomenon was discovered by Landau, Abrikosov, and Khalatnikov. It was found that the effective charge grows with the characteristic momentum. This means that the "continuum limit", where the ultraviolet cutoff is sent to infinity, cannot be reached: if the physical charge is kept fixed, we run into the *Landau pole* and the bare coupling constant becomes infinite at some finite Λ_{ultr}. On the other hand, if we keep g^2_{bare} finite, the physical charge goes to zero.

Thus, it became clear that a theory like (0.1), as well as QED, is sick and has no meaning beyond perturbation theory. In QED the latter works and, if we are not afraid of the inevitable trouble which we would meet if we tried to explore the unattainably high energies $E \sim M_{\text{pl}}$, we can obtain nontrivial theoretical predictions and compare them with experiment. In the meson theory, we cannot do that and, since the theory just makes no sense when the coupling is big, nonperturbative methods do not work either. This means that the Lagrangian (0.1) should be put into the waste basket.

Actually, the prevailing attitude of physicists for almost 20 years, from the mid-fifties to the mid-seventies, was that not only *that* Lagrangian, but also all other Lagrangians, the entire Lagrangian method of field theory should be discarded. It was assumed that the correct theory should be founded on some other principles. The idea of the *bootstrap* (that there are no fundamental particles and no fundamental fields whatsoever, and everything depends on itself in some self-consistent way) was popular. The theory of the S-matrix was being intensely developped. We now understand that the purely kinematical requirements of unitarity, causality, space–time and internal symmetries are not restrictive enough to determine the form of the S-matrix in a four-dimensional theory, but that *was* the hope. One of the positive developments in this "medieval" period was Regge theory. It is perceived now as good phenomenological effective theory describing hadron scattering at high energies and small tranverse momenta. But, in old days, this was considered *the* theory, the most refined and ingenious product of kinematical methods.

Thus, physicists, leaving aside the ambition to reveal a fundamental field theory, occupied themselves with a more pragmatic task — to handle the rapidly growing zoo of new "elementary" particles and resonances, if not in a refined theoretical fashion, then at least in some phenomenological way. It was basically the same job as had been done by Mendeleev for the zoo of chemical elements a century earlier. The "Mendeleev Table" of hadrons had been constructed by the early sixties by the efforts of many people, of which M. Gell-Mann probably deserves most credit. It was found that all known hadrons can be grouped into some octets and decuplets representing multiplets of $SU(3)$. This $SU(3)$ symmetry [now called *flavor* $SU(3)$] was not exact, but slightly broken in a controllable way. The model resulted in a lot of non-trivial predictions. The most famous one was the prediction of the existence of the Ω^--hyperon, including an estimate of its mass; this prediction was soon confirmed by experiment. This discovery played the same role as the discovery of missing chemical elements which filled out the empty cells in Mendeleev's Table. One could recall as well the classic story of the discovery of the planet Neptune.

We observe in experiment the octets and the decuplets, but where are the triplets and anti-triplets, the particles belonging to the fundamental representation of $SU(3)$ and its conjugate? These mysterious unobserved

particles were called *quarks*.[§] The constituent quark model for hadrons was created. From a purely algebraic result concerning tensor products of several fundamental representations: $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} + \mathbf{1}$ and $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} + \mathbf{8} + \mathbf{8} + \mathbf{1}$, it was inferred that mesons (grouped in octets) represent bound states of a quark and an antiquark, while the baryons (grouped in octets and decuplets) consist of three quarks. Assuming that the quarks are fermions and ascribing them the proper quantum numbers: the fractional electric charges $e_u = 2/3$, $e_d = e_s = -1/3$, the strangeness -1 and the isospin 0 for the strange quark and the strangeness 0 and the isospin 1/2 for the u and d quarks, and assuming that the hadrons are composed of quarks in roughly the same manner as the nuclei are composed of nucleons, the quantum numbers of all observed hadrons were correctly reproduced.

That was, of course, an important success, but the dynamics of the forces binding together the quarks in hadrons remained completely unknown. Even if one did not try to answer this question in precise terms, there were two points of trouble. First, it was absolutely unclear why the quarks are always bound and not present in Nature in free unbound form. (A *lot* of experimental efforts were devoted to the search of fractionally charged particles in accelerators, in cosmic rays, in meteorites and in the ice of Greenland: to no avail.) The second difficulty was more technical. Take e.g. the Δ^{++}-isobar which, according to the constituent quark model, is made of three u quarks in S wave with the same spin orientation. But the Pauli principle forbids it (by the same token, it forbids the existence of all other baryons including nucleons).

The remedy for the second trouble was found rather soon. It was assumed that the quark of each flavor (u, d, and s) is found in Nature in three *color*[¶] forms. Thus, on top of $SU(3)$ flavor, we have $SU(3)$ color, which is exact. Pauli's principle is no longer violated because we can write the

[§] At first, there were three of them: "up", "down", and "strange" quarks (or just u, d, and s quarks). Now we know three more: "charmed", "bottom" (or "beautiful"), and "top" quarks. The experimental values for the quark masses are

$$m_u \approx 4 \text{ MeV}\,, \quad m_d \approx 7 \text{ MeV}\,, \quad m_s \approx 150 \text{ MeV}\,,$$
$$m_c \approx 1.35 \text{ GeV}\,, \quad m_b \approx 4.8 \text{ GeV}\,, \quad m_t \approx 170 \text{ GeV}\,. \tag{0.2}$$

[¶] This is not the first time the word "color" appeared in this book. It is actually present right in its title in a disguised form. The Greek word for "color" is *chroma* and *chromo*dynamics means the dynamics of color.

wave function for the baryons in the required antisymmetric form $\epsilon^{ijk}q^i q^j q^k$ ($i, j, k = 1, 2, 3$ are the color indices).

It was postulated that by some miraculous reason only the color–singlet states are allowed to exist in Nature in unbound form. The quarks are color–triplets and are therefore doomed to stay confined within the hadrons. The word *confinement* was coined, but no satisfactory explanation for this phenomenon was suggested. In the middle of the sixties (and even much later) many theorists found this picture so crazy that they did not believe in the existence of quarks. Their position was pretty much the same as the position of Copernicus who said that the astronomical data could be explained more easily if adopting a mathematical model that the Earth and other planets go around the Sun. He never said it *is* the case, indeed (and that is why Copernicus in contrast to Galilei was never bothered by the Church).

This viewpoint actually prevailed until 1969 when experiments on deep inelastic electron scattering were performed at SLAC. It was shown that the total cross section of the deep inelastic scattering of virtual photons on nucleons (recall that the deep inelastic scattering is the inclusive process $e + N \to e +$ anything or, which is the same, $\gamma^* N \to$ anything, where γ^* is the virtual photon emitted by an electron) behaved in the same way as the cross section of the scattering of γ^* on a charged pointlike particle. That meant that there *are* some pointlike constituents within the proton. They were at first called *partons*, but it soon became clear that everything works wonderfully well if one assumes that these partons have the quantum numbers of quarks. Thus, quarks surpassed the status of an amusing mathematical model and displayed themselves as dynamical entities which interact and, hence, are really there (*interago ergo sum*).

But what about field theory? Was it indeed true that theorists, being frustrated by the Landau pole problem, abandoned all field theory studies? Not, of course. The latter were no longer streamline, but the nature of science is such that one can always find some people standing aside the crowd, doing unpopular things. Sometimes, this happens to be the right thing to do.

The field theory on whose basis QCD was eventually built, Yang–Mills theory, had actually been invented a long time ago, in 1954. Yang and Mills generalized the principle of local gauge invariance to the non-Abelian gauge group $SU(2)$ and wrote the celebrated Lagrangian $\mathcal{L}_{YM} = -(1/4)F^a_{\mu\nu}F^{\mu\nu\,a}$.

The spectrum of this theory involves massless self-interacting vector particles belonging to the triplet of $SU(2)$. Yang and Mills added the mass term $-m^2(A^a_\mu)^2/2$ to the Lagrangian and tried to associate the resulting massive vector particles with the ρ mesons. This attempt did not work. As a result, Yang–Mills theory was not considered to be a realistic physical theory during the following 10–15 years.

But the model as such was interesting and many theorists continued to play with it. It turned out that the quantization of such a theory is far from being easy. Proceeding in a naïve way, one gets nonsensical results, including violation of unitarity. At the beginning of the sixties, Feynman tried to resolve this problem (he was interested in Yang–Mills theory as a model for quantum gravity: the nonlinear interactions of gauge bosons in the former carry some features of the nonlinear interactions of gravitons in the latter). Feynman did not succeed by himself, but he had some important insights, which helped Faddeev and Popov to develop their *ghost* method in 1967. Including the ghosts made perturbation theory self-consistent and it became possible to calculate scattering amplitudes and other quantities at, in principle, any order in the coupling constant.

While some people tried to construct a consistent theoretical scheme for the quantization of Yang–Mills theory, some other researchers contemplated on its possible phenomenological applications. Nobody was too much in a hurry.

When the idea to describe ρ mesons as Yang–Mills quanta was abandonned, people did not further think about the possibility of applying this theory to the physics of strong interactions, but rather turned to the weak interactions. Fermi theory, involving four-fermion vertices with dimensionful coupling, was sick: nonrenormalizable and nonunitary. The idea was to consider it as the effective low-energy Lagrangian of some other theory involving only dimensionless couplings of fermions with massive charged vector bosons. Charged fundamental vector particles are most naturally described as Yang–Mills particles (ρ mesons are not well described in this way, but neither are they fundamental). Of course, in pure Yang–Mills theory the quanta are massless and one should have guessed how to give them mass in some benign way in order not to spoil renormalizability and unitarity of the massless theory (which, as a matter of fact, were not yet proven at that time!)

An almost correct model of the weak interactions was constructed along these lines by Glashow back in 1961 and the fully correct one, which unites

weak and electromagnetic interactions, by Weinberg and Salam in 1967. The $W^{\pm}$ and Z^0 bosons were endowed with masses by the non-Abelian Higgs mechanism. The renormalizability and unitarity of this model were proven later by 't Hooft and Veltman.

The idea that Yang–Mills theory describes not only the physics of weak interactions, but also the strong interaction dynamics was put forward in 1973 in three independent papers by Pati and Salam; Fritzsch, Gell-Mann, and Leutwyler; and Weinberg. They suggested that the quarks interact with each other by the exchange of *gluons*, massless vector particles belonging to the octet representation of $SU(3)$. The gluons also interact between themselves, as is dictated by the Yang–Mills Lagrangian with the $SU(3)$ color gauge group. A real breakthrough occurred when the charge renormalization in this theory was calculated by Gross, Wilczek, and Politzer, and the phenomenon of asymptotic freedom was discovered: the effective charge does not grow with energy as is the case for Landau-pole-plagued theories like QED, but rather decreases.

History is rather dramatic at this point. The first person who correctly calculated the renormalization of the coupling constant in pure Yang–Mills theory was Khriplovich as early as 1969. He calculated the 1-loop gauge boson polarization operator $\Pi_{\mu\nu}$ in Coulomb gauge, which does not involve ghosts. In Coulomb gauge, $\Pi_{\mu\nu}$ is the *only* source of the charge renormalization; the contribution of the vertices is irrelevant. Khriplovich did not understand, however, the meaning of his result, and did not think of applying Yang–Mills theories to physics of strong interaction. There was an even earlier paper in 1965 by Terent'ev and Vanyashin, who found indications that charge may decrease with energy, but they did the calculation in a nonrenormalizable theory (without ghosts, which were not yet known at that time) and ascribed the strange behavior of the charge to this. Finally, 't Hooft also calculated it about a year before Gross, Wilczek, and Politzer, but did not, at first, understand the overwhelming significance of his result and did not publish it.[‖]

At any rate, by the end of 1974 the dust settled, and the picture became clear. The asymptotic freedom of the theory:

(1) Made it self-consistent: there *is* a continuum limit where the bare charge falls off as the ultraviolet cutoff increases in such a way that

[‖]The interested reader can read about all that in more details in a historical review article by M. Shifman written for the volume [11].

the effective charge defined at a given energy scale is kept fixed.

(2) Explained SLAC deep inelastic data. When the energy is large, the interaction between quarks is relatively weak, we can treat them as free and that is what was seen in experiment. Moreover, a quantitative theory of "scaling violations", i.e. the theory of deviations from the free quark–parton model due to interaction between quarks was developped by Politzer in 1974. These deviations turned out to be rather weak. Now experiments are precise enough to see them, and they perfectly well agree with the theory.

(3) Explained why strong interactions are strong. If the coupling falls off at small distances, it grows at large distances, becoming roughly of order 1 at the distances of order 1 fm, which are relevant to the nuclear physics.

(4) Did not explain the phenomenon of confinement (to derive it from the first principles of QCD is still a major unresolved task), but made it probable. If the interaction is strong at large distances, who knows how it behaves there? It is quite conceivable that the spectrum of asymptotic physical states does not resemble the set of the fundamental fields and, in particular, does not include colored states. Some modern ideas on the possible mechanisms of confinement will be discussed at the end of this book.

Since that time, many other experimental tests of QCD were carried out, and there is not a slightest doubt anymore that QCD is correct. Probably, the most spectacular confirmation comes from jet physics. Consider the process of e^+e^- annihilation with creation of a bunch of hadrons in the final state. If the energy of colliding electron and positron is large, the process goes, according to the theory, in two stages. At the first stage, the fundamental particles, the quarks and gluons are produced. This fundamental process can be $e^+e^- \to q\bar{q}$, $e^+e^- \to q\bar{q}g$, $e^+e^- \to q\bar{q}q\bar{q}$, etc. The cross section of these processes can be calculated in the framework of QCD perturbation theory.

But since quarks and gluons do not exist in free form, we observe only hadrons in the final state. Thus, at the second stage each energetic quark or gluon creates a *jet* of hadrons going roughly in the same direction as the initial colored particles (see Fig. 0.1).

Such jet processes are indeed seen. The most common is the process when only two jets are created. The underlying fundamental process is

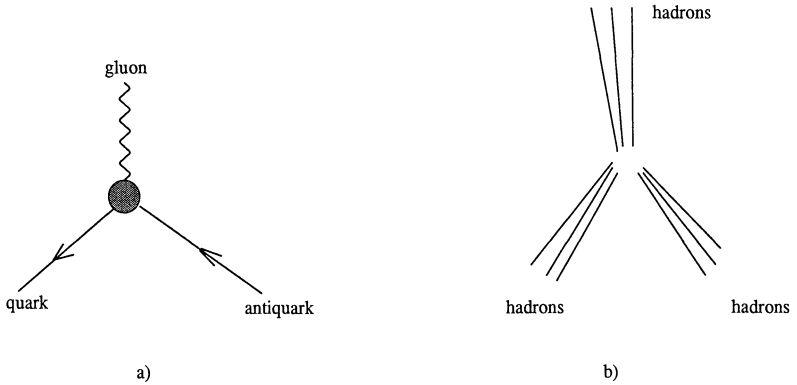


Fig. 0.1 Production of three jets in e^+e^- annihilation. *a)* quark and gluon stage, *b)* hadron stage.

$e^+e^- \to q\bar{q}$. The three-jet process (corresponding to the QCD process $e^+e^- \to q\bar{q}g$) is less probable because the cross section involves an extra power of the strong coupling constant α_s (which is small at high energies). An event where four jets are created (corresponding to the QCD processes $e^+e^- \to q\bar{q}gg$ and $e^+e^- \to q\bar{q}q\bar{q}$) is still less probable, etc. Experimentalists can measure jet cross sections, *viz.* the probabilities of producing, say, 3 jets with given total energies (E_{jet} is the sum of the energies of all particles forming the jet) and forming given angles. There is quantitative agreement between such data and the theoretical result for the differential cross section of the process $e^+e^- \to \bar{q}qg$.

There are many other tests of QCD: a series of tests for heavy quarkonium systems, for the decays of W and Z bosons, for the masses, widths, static and dynamical characteristics of low lying hadron resonances (they were calculated in QCD numerically by doing the functional integrals on the lattice and semianalytically using the ingenious *ITEP sum rules* technique) as well as many other things. The voices of the last "heretics" who have still expressed doubts whether QCD is the true theory of strong interactions, faded away by the end of the seventies.

Contents

Preface vii

Notation and Conventions xv

Introduction: Some History xix

PART 1: FOUNDATIONS

Lecture 1 Yang–Mills Field 3
1.1 Path Ordered Exponentials. Invariant Actions. 3
1.2 Classical Solutions. 11

Lecture 2 Instantons 15
2.1 Topological Charge. 15
2.2 Explicit Solutions. 19

Lecture 3 Path Integral in Quantum Mechanics 25
3.1 Conventional Approach. 25
3.2 Euclidean Path Integral. 30
3.3 Holomorphic Representation. 32
3.4 Grassmann Dynamic Variables. 34

Lecture 4 Quantization of Gauge Theories 39
4.1 Dirac Quantization Procedure. 39
4.2 Path Integral on the Lattice. 49

Lecture 5 θ–Vacuum **55**
5.1 Quantum Pendulum. 55
5.2 Large Gauge Transformations in Non-Abelian Theory. 62

PART 2: PERTURBATION THEORY

Lecture 6 Diagram Technique in Simple and Complicated Theories **69**
6.1 Feynman Rules from Path Integral 70
6.2 Fixing the Gauge 76

Lecture 7 When the Gauge is Fixed ... **85**
7.1 Gribov Copies. 85
7.2 Ward Identities. 88
7.3 Ghosts and Unitarity. 92
7.4 BRST Quantization. 96

Lecture 8 Regularization and Renormalization **103**
8.1 Different Regularization Schemes. 103
8.2 Renormalized Theory as an Effective Theory. Slavnov–Taylor Identities. 109

Lecture 9 Running Coupling Constant **119**
9.1 One-loop Calculations. 119
9.2 Renormalization Group. Asymptotic Freedom and Infrared Slavery. 127
9.3 Observables. Ambiguities. Anomalous Dimensions. 133

Lecture 10 Weathering Infrared Storms **143**
10.1 Bloch–Nordsieck Cancellation. 144
10.2 Non-Abelian Complications. Coherent States. 151

Lecture 11 Collinear Singularities: Theory and Phenomenology **157**
11.1 Double Logarithmic Asymptotics. 157
11.2 Jet Cross Sections. 159
11.3 DIS and KLN. 165

PART 3: NONPERTURBATIVE QCD

Lecture 12 Symmetries: Anomalous and Not **177**
12.1 Conformal Symmetry and its Breaking. 179
12.2 Anomalous Chiral Symmetry. 184
12.3 Nonsinglet Chiral Symmetry and its Spontaneous Breaking. . . 193
12.4 Effective Chiral Langrangian. 196

Lecture 13 Quarks on Euclidean Lattice **199**
13.1 Nielsen–Ninomiya's No-go Theorem. 199
13.2 Ways to Go. The Ginsparg–Wilson Way. 203

Lecture 14 Aspects of Chiral Symmetry **211**
14.1 QCD Inequalities. Vafa–Witten Theorem. 211
14.2 Euclidean Dirac Spectral Density. 214
14.3 Infrared Face of Anomaly. 219
14.4 Chiral Symmetry Breaking and Confinement. 223

Lecture 15 Mesoscopic QCD **227**
15.1 Partition Function: $N_f = 1$. 227
15.2 Partition Function: $N_f \geq 2$. 230
15.3 Spectral Sum Rules. 232
15.4 Instanton Gas and Instanton Liquid. 234

Lecture 16 Fairy QCD **241**
16.1 Finite θ. 241
16.2 Large N_c. 247

Lecture 17 ITEP Sum Rules: the Duality Festival **255**
17.1 The Method. 255
17.2 Nucleon Mass and Residue. 257
17.3 Pion Formfactor and Nucleon Magnetic Moments. 268

Lecture 18 Hot and Dense QCD **275**
18.1 Lukewarm Pion Gas. Restoration of Chiral Symmetry. 277
18.2 Quark–Gluon Plasma. 282
18.3 Finite Baryon Density. Color Superconductivity. 287

Lecture 19 Confinement **291**
19.1 Weak Confinement and Strong Confinement. Wilson Criterium. 292
19.2 Schwinger Model. 295
19.3 Polyakov Model. 298
19.4 Dual Superconductivity. 302

Appendix Unitary groups **307**

Bibliography **315**

Index **319**

Lecture 1

Yang–Mills Field

1.1 Path Ordered Exponentials. Invariant Actions.

We start from the well known Lagrangian of electrodynamics involving a charged scalar field $\phi(x)$:

$$\mathcal{L} = -\frac{1}{4e^2} F_{\mu\nu} F^{\mu\nu} + \mathcal{D}_\mu \phi \mathcal{D}^\mu \phi^* \ , \tag{1.1}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and $\mathcal{D}_\mu = \partial_\mu - iA_\mu$ is the covariant derivative (a more conventional definition for A_μ differs from ours by the factor e). The Lagrangian (1.1) is invariant under *gauge transformations*

$$\begin{aligned} A_\mu &\rightarrow A_\mu + \partial_\mu \chi(x) \ , \\ \phi &\rightarrow e^{i\chi(x)} \phi \ . \end{aligned} \tag{1.2}$$

These transformations form a one-parameter Abelian group which is $U(1)$. Yang–Mills theory is a generalization of QED to the case of non-Abelian gauge groups. We will now suggest a geometric interpretation for the action (1.2) which is of little use for QED but is very useful for deriving the Yang–Mills action in a natural way.

Suppose we define an element of the $U(1)$ group $U(x) = e^{i\alpha(x)}$ at each point of our Minkowski space–time. One can realize it in many different ways. One such way can be obtained from another by multiplying each $U(x)$ by an element of the group $e^{i\chi(x)}$, i.e. by a gauge transformation. Mathematicians would say that a *principal fiber bundle* over R^4 with the group $U(1)$ as a fiber is defined. The function $U(x)$ is called the *section*

of this fiber bundle.* Instead of considering an element of the group itself, one could consider equally well a representation of the group on which the elements of the group act and which, in the trivial case of $U(1)$, is simply a charged field $\phi(x)$. The geometric meaning of the gauge field A_μ is clarified if the operation of *parallel transport* is defined. For each point x we define

$$U_{\Rightarrow x+dx}(x) = \Phi(x+dx,x)U(x) \equiv e^{iA_\mu(x)dx^\mu}U(x) , \tag{1.3}$$

where dx here is assumed to be small. The notation is not standard, and it probably makes sense to express it also in words: $U_{\Rightarrow x+dx}(x)$ is the function U taken at the point x and then transported parallelly with the factor Φ to an infinitesimally close point $x+dx$. If we perform a parallel transport at a finite distance along some path P parametrized as $\xi^\mu(s)$, we obtain

$$U_{\underset{\Rightarrow y}{P}}(x) = \Phi(y,x;P)U(x) \equiv e^{\int_x^y iA_\mu(\xi)d\xi^\mu}U(x) , \tag{1.4}$$

where the integral is performed along P. $U_{\underset{\Rightarrow y}{P}}(x)$ depends on the path and does not coincide, of course, neither with $U(x)$ nor with $U(y)$. What is however remarkable is that it transforms under the gauge transformations in the same way as $U(y)$. Indeed, substituting $A_\mu \to A_\mu(x) + \partial_\mu\chi(x)$, we see that the exponential $\Phi(y,x;P)$ (P marks the path) is changed as

$$\Phi(y,x;P) \to e^{i\chi(y)}\Phi(y,x;P)e^{-i\chi(x)} . \tag{1.5}$$

If substuting also $U(x) \to e^{i\chi(x)}U(x)$, we see that $U_{\Rightarrow y}(x) \to e^{i\chi(y)}U_{\Rightarrow y}(x)$. Note that for any *closed* contour C, the quantity $\Phi_C = \exp\{i\oint_C A_\mu(\xi)d\xi^\mu\}$ is gauge invariant.

To understand $F_{\mu\nu}$ in these terms, consider the operator Φ for parallel transport along the perimeter of a small square $\Box$ with the side a in, say, the (12) plane. Consider first parallel transport along one of the sides:

$$\Phi(x^\mu + a\delta^\mu_1, x^\mu) = e^{ia\int_0^1 ds A_1(x_\mu + as\delta_{\mu 1})} = e^{iaA_1(x^\mu + a\delta^\mu_1/2)} + o(a^2) .$$

*One must say here that the theory of gauge fields was developed independently in pure mathematics at roughly the same time as physicists developed it for their purposes. Later it was realized that one and the same thing was considered in different terms.

We have †

$$\begin{aligned}\Phi_{\square} &= e^{-iaA_2[x^\mu+a\delta_2^\mu/2]}e^{-iaA_1[x^\mu+a(\delta_1^\mu/2+\delta_2^\mu)]}e^{iaA_2[x^\mu+a(\delta_1^\mu+\delta_2^\mu/2)]} \\ &\qquad \times e^{iaA_1[x^\mu+a\delta_1^\mu/2]} \\ &= e^{ia^2(\partial_1 A_2-\partial_2 A_1)}+o(a^2)=e^{ia^2F_{12}}+o(a^2)\ . \end{aligned} \tag{1.6}$$

A parallel transport over a finite closed contour is

$$\Phi_C = e^{i\iint F_{\mu\nu}d\sigma^{\mu\nu}}\ ,$$

where the integral goes over any surface bounded by the contour C. This, of course, could be derived immediately from the definition of Φ_C and the Stokes theorem.

A striking analogy with general relativity is now seen. The parallel transport on our fiber bundle $U(x)$ is fully analogous to the parallel transport of vectors on a curved Riemann surface. The gauge potential A_μ plays the role of the Christoffel symbols. The tensor $F_{\mu\nu}$ appeared as the operator or parallel transport over a small closed contour and is the analog of the Riemann curvature. Finally, one can define a *covariant derivative* in the same way as in general relativity: we define

$$n^\mu \mathcal{D}_\mu\phi \equiv \lim_{\epsilon\to 0}\frac{\phi(x+\epsilon n)-\phi_{\Rightarrow x+\epsilon n}(x)}{\epsilon} = n^\mu(\partial_\mu - iA_\mu)\phi(x)\ . \tag{1.7}$$

Thus, to compare $\phi(x)$ in closed spatial points, we should first transport $\phi(x)$ with the operator $\Phi(x+\epsilon n, x)$ to the adjacent point and only after that form a difference. That is why an alias for the gauge potential A_μ used by matematicians is the *connection*, it kind of "connects" adjacent points in our fiber bundle. Under the gauge transformation, the covariant derivative transforms as

$$\mathcal{D}_\mu\phi(x) \ \to\ e^{i\chi(x)}\mathcal{D}_\mu\phi(x)\ ,$$

in contrast to the usual derivative, which does not transform in a nice way.

Let us now generalize this construction to the non-Abelian case. Suppose we have a principal fiber bundle based on an arbitrary Lie group G, and its section (or just group-valued function) $U(x)$. The gauge transformation is the transition from one section to another according to

$$U(x) \to \Omega(x)U(x)\ ,$$

†The Minkowskian conventions (N.1)–(N.5) are used.

with $\Omega \in G$. Define an infinitesimal parallel transporter

$$\Phi(x+dx,x) = e^{iA^a_\mu(x)T^a dx^\mu} \ , \tag{1.8}$$

where $A^a_\mu(x)$ is a non-Abelian version of the connection (i.e. gauge potential) belonging to the corresponding Lie algebra, and T^a are the generators of the group in some representation. We will only be interested in unitary groups and choose T^a in the fundamental representation $T^a \equiv t^a$, $\mathrm{Tr}(t^a t^b) = \frac{1}{2}\delta_{ab}$.

Let us now *require* that $U_{\Rightarrow x+dx}(x) = \Phi(x+dx,x)U(x)$ is transformed under the gauge transformations in the same way as $U(x+dx)$:

$$U_{\Rightarrow x+dx}(x) \to \Omega(x+dx)U_{\Rightarrow x+dx}(x) \ ,$$

which means that

$$\Phi(x+dx,x) \to \Omega(x+dx)\Phi(x+dx,x)\Omega^\dagger(x) \ .$$

Substituting here $\Phi(x+dx,x)$ from Eq. (1.8) and comparing the terms $\propto dx^\mu$, we *derive* the transformation law for the gauge potential:

$$\hat{A}_\mu \to \Omega(x)\hat{A}_\mu\Omega^\dagger(x) - i\left[\partial_\mu\Omega(x)\right]\Omega^\dagger(x) \ , \tag{1.9}$$

where $\hat{A}_\mu = A^a_\mu t^a$. For infinitesimal gauge transformations, $\Omega(x) = e^{i\hat{\chi}(x)} \approx 1 + i\hat{\chi}(x)$,

$$\hat{A}_\mu \to \hat{A}_\mu + \partial_\mu\hat{\chi}(x) + i[\hat{\chi}(x),\hat{A}_\mu] \ . \tag{1.10}$$

For the Abelian $U(1)$ group, we reproduce, of course, the law of transformation (1.2) of electrodynamics.

The parallel transporter (1.8) can also act on the fields in an arbitrary representation. In the full analogy with what we had earlier, we can define the covariant derivative of the field as

$$\mathcal{D}_\mu\phi(x) = (\partial_\mu - iA^a_\mu T^a)\phi(x) \ .$$

Under the gauge transformation, the covariant derivative $\mathcal{D}_\mu\phi(x)$ is multiplied from the left by the factor $e^{i\chi^a(x)T^a}$ in the same way as $\phi(x)$.

One can define also a (path-dependent) parallel transporter along a finite path P. It is given by a so-called $P-$ordered exponential which is defined as the product of an infinite number of infinitesimal transporters (1.8)

along the path:

$$\Phi(y,x;P) = P\exp\left\{i\int_x^y \hat{A}_\mu(\xi)d\xi^\mu\right\} \stackrel{def}{=} \lim_{n\to\infty}\Phi(y,\xi_n)\Phi(\xi_n,\xi_{n-1})\cdots\Phi(\xi_1,x)\ . \quad (1.11)$$

Note that the product is formed starting from the *end* point y. It is easy to see that the law of gauge transformation for $\Phi(y,x;P)$ is the same as for $\Phi(x+dx,x)$.

In the Abelian case, the parallel transporter along a closed path is gauge invariant. This is no longer true in the non-Abelian cases where it depends in general on an initial $\equiv$ final point x that we fix on the contour: $\Phi_C \to \Omega(x)\Phi_C\Omega^\dagger(x)$. However, the trace $W(C) = \mathrm{Tr}\{\Phi_C\}$ is gauge invariant. This is called the *Wilson loop.*

To define a non-Abelian field density tensor $\hat{F}_{\mu\nu}$, consider as before the operator of parallel transport along the perimeter of a small square (anticipating further lattice applications, we can call it *plaquette* right now). Substituting $A_\mu \to \hat{A}_\mu$ in (1.6) and taking into account the fact that $\hat{A}_\mu$ do not commute anymore (thereby, the order of the exponential factors is now important), we obtain

$$\Phi_{\square_{\mu\nu}} = e^{ia^2\hat{F}_{\mu\nu}} + o(a^2)\ , \quad (1.12)$$

with

$$\hat{F}_{\mu\nu} = \partial_\mu\hat{A}_\nu - \partial_\nu\hat{A}_\mu - i[\hat{A}_\mu,\hat{A}_\nu] = i[\mathcal{D}_\mu,\mathcal{D}_\nu]\ . \quad (1.13)$$

In "colored vector" notation,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu\ , \quad (1.14)$$

where f^{abc} are the structure constants of the group defined in Eq. (A.23). In the following, we will frequently go over from the matrix notation to colored vector notation and back without further comments. We will not mark matrices with hats anymore.

Under the gauge transformation,

$$F_{\mu\nu} \to \Omega(x)F_{\mu\nu}\Omega^\dagger(x)\ . \quad (1.15)$$

Thus, $F_{\mu\nu}$ has the geometric meaning of the curvature of our fiber bundle in color space.‡

Take now the trace of $\Phi_{\square_{\mu\nu}}$. To get a nontrivial result, one should expand the exponentials in the analog of Eq. (1.6) up to the order a^4. We obtain (check it!)

$$W_{\square_{\mu\nu}} = \frac{1}{N}\mathrm{Tr}\left\{\Phi_{\square_{\mu\nu}}\right\} = 1 - \frac{a^4}{2N}\mathrm{Tr}\{F_{\mu\nu}F_{\mu\nu}\} + O(a^6) \ . \tag{1.16}$$

The factor $1/N$ (for the $SU(N)$ gauge group) is introduced for convenience; no summation over μ, ν is assumed. This quantity is gauge invariant and so is the quantity $\mathrm{Tr}\{F_{\mu\nu}F_{\mu\nu}\}$ (this is seen also from the gauge transformation law (1.15) for $F_{\mu\nu}$).

The gauge and Lorentz invariant Lagrangian of pure Yang–Mills theory is defined as

$$\mathcal{L} = -\frac{1}{2g^2}\mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\} = -\frac{1}{4g^2}F^a_{\mu\nu}F^{a,\mu\nu} \ , \tag{1.17}$$

where the summation over μ, ν is now performed with the standard Minkowskian metric (N.1). This is the direct analog of the Lagrangian of pure photodynamics.

The latter is a trivial theory involving no interaction. In the Abelian case, to make things interesting, we have to include charged matter fields. In the non-Abelian case, already pure gluodynamics involves nonlinear interactions (with the coupling constant g) and is highly non-trivial. It is not difficult to also write down invariant Lagrangians describing the interactions of colored matter with non-Abelian Yang–Mills fields. Suppose we have a scalar field in the fundamental representation of the group. The corresponding Lagrangian can be written as a gauge invariant combination $[\mathcal{D}_\mu\phi(x)]^\dagger\mathcal{D}^\mu\phi(x)$. For spinor fields, we can write

$$\mathcal{L}_\psi = i\bar{\psi}\mathcal{D}_\mu\gamma^\mu\psi + \text{ optional mass term} \ . \tag{1.18}$$

QCD is a theory involving the Yang–Mills fields based on the gauge group $SU(3)$, and six different Dirac spinor fields in the fundamental representation of the group. The quanta of the gauge fields are called *gluons*), and

‡We hasten to comment that general relativity, the analogy with which we explore, is a *more* complicated theory than just a gauge theory with local Lorentz group. For sure, the local Lorentz symmetry is there, being associated with local rotations of the vierbein. But the Einstein action is invariant under a much larger symmetry: the symmetry of general coordinate transformations.

the quanta of the fermion fields quarks. The Lagrangian reads

$$\mathcal{L}_{QCD} = -\frac{1}{2g^2}\mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\} + \sum_{f=1}^{6}\bar{\psi}_f(i\not{\mathcal{D}} - m_f)\psi_f \ . \tag{1.19}$$

However, for the rest of the present lecture as well as in the following one, we will forget about the existence of the quarks and will concentrate on pure Yang–Mills dynamics.

The dynamical equations of motion are obtained by varying the Lagrangian (1.17) with respect to A_μ and have the form

$$[\mathcal{D}_\mu, F^{\mu\nu}] = \partial_\mu F^{\mu\nu} - i[A_\mu, F^{\mu\nu}] = 0 \ , \tag{1.20}$$

or

$$\partial_\mu F^{a,\mu\nu} + f^{abc}A^b_\mu F^{c,\mu\nu} = 0 \ . \tag{1.21}$$

Bearing in mind that, according to (1.13), $F_{\mu\nu}$ is a commutator of covariant derivatives and using the Jacobi identity, we derive the Bianchi identity for our color curvature

$$[\mathcal{D}_\sigma, F_{\mu\nu}] + [\mathcal{D}_\mu, F_{\nu\sigma}] + [\mathcal{D}_\nu, F_{\sigma\mu}] = 0 \ , \tag{1.22}$$

the "second pair of Maxwell equations".

Let us construct now the classical Hamiltonian of the Yang–Mills field. The canonical Hamiltonian is $\sum_i p_i\dot{q}_i - \mathcal{L}$, where q_i are generalized coordinates and p_i are their canonical momenta. Just as in electrodynamics, we meet here a difficulty that not all variables A_μ are dynamical. A_0 enters the Lagrangian (1.17) without time derivatives and has no associated canonical momentum. On the other hand, the spatial components A^a_i are dynamical and their canonical momenta P^a_i are just chromoelectric fields $P^a_i = \frac{1}{g^2}E^a_i = \frac{1}{g^2}F^a_{0i}$. Only a part of Lagrange's equations of motion, namely the equations

$$\frac{d}{dt}\frac{\delta\mathcal{L}}{\delta\dot{A}_i} - \frac{\delta\mathcal{L}}{\delta A_i} = 0 \tag{1.23}$$

describe the time evolution of the system. The equations of motion obtained by varying the Lagrangian with respect to A_0,

$$\begin{aligned} G &= [\mathcal{D}_i, E_i] = 0 \qquad\qquad \text{or} \\ G^a &= \partial_i E^a_i + f^{abc}A^b_i E^c_i = 0 \end{aligned} \tag{1.24}$$

(the Gauss law) have the meaning of *constraints* to be imposed on the initial Cauchy data for the equation system (1.23). Actually, the fact that a part of the variables is not dynamical and that we thus deal with a constrained system is very closely related to the fact that the Lagrangian (1.17) enjoys the gauge invariance (1.9). We will explore this relation in details when discussing the quantum theory. Here we just write the system of Hamilton's equations of motion corresponding to the Lagrangian (1.17). In full analogy with electrodynamics, the Hamiltonian is

$$\mathcal{H} = \frac{1}{2g^2}\int d^3x[E_i^a E_i^a + B_i^a B_i^a] = \frac{1}{2}\int d^3x \left[g^2 P_i^a P_i^a + \frac{1}{g^2} B_i^a B_i^a\right], \quad (1.25)$$

where $B_i^a = -\frac{1}{2}\epsilon_{ijk}F_{jk}^a$ is the chromomagnetic field. Dynamical Hamilton's equations, describing the phase space dynamics of the system, are

$$\begin{aligned} \frac{\delta\mathcal{H}}{\delta P_i^a} &= E_i^a = \dot{A}_i^a \ , \\ -\frac{\delta\mathcal{H}}{\delta A_i^a} &= \frac{1}{g^2}[\partial_j F_{ji}^a + f^{abc}A_j^b F_{ji}^c] = \dot{P}_i^a = \frac{1}{g^2}\dot{E}_i^a \ , \end{aligned} \quad (1.26)$$

and the constraints (1.24) have already been written. The time evolution of the system described by the Hamiltonian (1.25) is compatible with the constraints (1.24). This is guaranteed by the fact that the Poisson bracket $\{G, \mathcal{H}\}_{\text{P.B.}}$ is zero, as can be checked explicitly.

Problem. Calculate $\{G^a(\boldsymbol{x}), G^b(\boldsymbol{y})\}_{P.B.}$.
Solution. We have

$$\begin{aligned} \{G^a(\boldsymbol{x}), G^b(\boldsymbol{y})\}_{P.B.} &= \{\partial_i E_i^a(\boldsymbol{x}) + f^{acd}A_i^c(\boldsymbol{x})E_i^d(\boldsymbol{x}), \partial_j E_j^b(\boldsymbol{y}) \\ &\qquad + f^{bef}A_j^e(\boldsymbol{y})E_j^f(\boldsymbol{y})\}_{P.B.} \\ &= -f^{abc}\left[\frac{\partial}{\partial x_i}\delta(\boldsymbol{x}-\boldsymbol{y})E_i^c(\boldsymbol{y}) + \frac{\partial}{\partial y_i}\delta(\boldsymbol{x}-\boldsymbol{y})E_i^c(\boldsymbol{x})\right] \\ &\quad + f^{acd}f^{bef}\left[\delta^{ed}A_i^c(\boldsymbol{x})E_i^f(\boldsymbol{y}) - \delta^{cf}A_i^e(\boldsymbol{y})E_i^d(\boldsymbol{x})\right]\delta(\boldsymbol{x}-\boldsymbol{y}) \ . \end{aligned} \quad (1.27)$$

The first term is transformed using the identity

$$\delta'(x-y)[f(y)-f(x)] \ = \ -\delta'(x-y)f'(x)(x-y) \ = \ \delta(x-y)f'(x)$$

and gives $-f^{abc}\partial_i E_i^c(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$. The second term is transformed by use of the Jacobi identity for structure constants (A.24) and gives, respectively,

$-f^{abc}f^{cde}A_i^d(\boldsymbol{x})E_i^e(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$.

The relation obtained

$$\{G^a(\boldsymbol{x}), G^b(\boldsymbol{y})\}_{P.B.} = -f^{abc}G^c(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y}) \tag{1.28}$$

is rather important and will be used later to quantize the theory.

1.2 Classical Solutions.

We will concentrate here on the solutions in real Minkowski time. Euclidean solutions will be a subject of the next lecture.

The simplest ansatz is $A_\mu(x) = cf_\mu(x)$, where c is a colored matrix, one and the same for all components μ and for all points x. Then the commutators in $F_{\mu\nu}$ and in the equations of motion disappear and the classical dynamics of the system is the same as for the photodynamics, i.e. trivial. Any solution can be expressed as a superposition of the plane waves $A_\mu(x) = ce_\mu e^{-ikx}$. Their quantization gives gluons which will be discussed at length later.

However, the non-linear equations (1.20) admit a lot of other nontrivial solutions. Let us first discuss the solutions that our system does *not* have, namely, soliton solutions. Solitons, i.e. localized solutions of the equations of motion with finite energy, appear in many nonlinear systems. For instance, one of the first examples of nonlinear partial differential equation studied by physicists — the KdV equation $u_t + \alpha u_x + \beta u_{xxx} + \gamma u u_x = 0$ (it describes waves on the surface of shallow water and many other physical phenomena) admits soliton solutions. The well-known 't Hooft–Polyakov monopole is a soliton in a relativistic field theory. In general, a soliton may depend on $\boldsymbol{x} - \boldsymbol{v}t$, describing a quasiparticle moving with the speed $\boldsymbol{v}$. In a Lorentz-invariant theory, a soliton can always be "stopped" by a proper Lorentz transformation, so that the solution becomes static, depending only on $\boldsymbol{x}$.

Suppose the equations (1.20) admitted a soliton solution with a finite characteristic scale ρ and finite energy E. Note, however, that the equations do not involve a dimensionful constant and are thereby invariant under scale transformations $x \to \lambda x$. Thus, if a field configuration with scale ρ is a solution, so is a configuration with the same shape and scale $\lambda\rho$. By dimensional reasons, the energy of such a configuration should differ from the energy of the original configuration by the factor λ: $E \to E/\lambda$. If $\lambda > 1$,

$-f^{abc}f^{cde}A_i^d(\boldsymbol{x})E_i^e(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$.

The relation obtained

$$\{G^a(\boldsymbol{x}), G^b(\boldsymbol{y})\}_{P.B.} = -f^{abc}G^c(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y}) \tag{1.28}$$

is rather important and will be used later to quantize the theory.

1.2 Classical Solutions.

We will concentrate here on the solutions in real Minkowski time. Euclidean solutions will be a subject of the next lecture.

The simplest ansatz is $A_\mu(x) = cf_\mu(x)$, where c is a colored matrix, one and the same for all components μ and for all points x. Then the commutators in $F_{\mu\nu}$ and in the equations of motion disappear and the classical dynamics of the system is the same as for the photodynamics, i.e. trivial. Any solution can be expressed as a superposition of the plane waves $A_\mu(x) = ce_\mu e^{-ikx}$. Their quantization gives gluons which will be discussed at length later.

However, the non-linear equations (1.20) admit a lot of other nontrivial solutions. Let us first discuss the solutions that our system does *not* have, namely, soliton solutions. Solitons, i.e. localized solutions of the equations of motion with finite energy, appear in many nonlinear systems. For instance, one of the first examples of nonlinear partial differential equation studied by physicists — the KdV equation $u_t + \alpha u_x + \beta u_{xxx} + \gamma u u_x = 0$ (it describes waves on the surface of shallow water and many other physical phenomena) admits soliton solutions. The well-known 't Hooft–Polyakov monopole is a soliton in a relativistic field theory. In general, a soliton may depend on $\boldsymbol{x} - \boldsymbol{v}t$, describing a quasiparticle moving with the speed $\boldsymbol{v}$. In a Lorentz-invariant theory, a soliton can always be "stopped" by a proper Lorentz transformation, so that the solution becomes static, depending only on $\boldsymbol{x}$.

Suppose the equations (1.20) admitted a soliton solution with a finite characteristic scale ρ and finite energy E. Note, however, that the equations do not involve a dimensionful constant and are thereby invariant under scale transformations $x \to \lambda x$. Thus, if a field configuration with scale ρ is a solution, so is a configuration with the same shape and scale $\lambda\rho$. By dimensional reasons, the energy of such a configuration should differ from the energy of the original configuration by the factor λ: $E \to E/\lambda$. If $\lambda > 1$,

energy decreases. Thus, we have come to a logical contradiction. A static solution should correspond to a minimum of the energy functional. But, assuming that such a solution exists, we have just constructed configurations with lower energy, which invalidates the initial assumption.

Note that e.g. monopoles may exist because the corresponding equations *do* involve a dimensionful parameter (the Higgs vacuum expectation value). Localized static solutions may be present in pure Yang–Mills theory as well if we consider the field living on a finite spatial manifold (say, on S^3): the size of the manifold provides the scale and the above arguments do not apply. But no localized static solutions exist in flat, infinite 3-dimensional space. Physically, if an initial configuration with a finite scale is created, it will necessarily smear out as time evolves.

Nontrivial solutions of the equations (1.20) have no scale and their energy is infinite. Let us choose a rather restrictive ansatz with $A_0 = 0$ and A_i depending only on t, but not on $\boldsymbol{x}$. Let us also restrict ourselves to the $SU(2)$ case. The Lagrangian acquires the form

$$\mathcal{L} = \frac{L}{V} = \frac{1}{2g^2}(\dot{A}_i^a)^2 - \frac{1}{4g^2}[(A_i^a A_i^a)^2 - (A_i^a A_j^a)^2] , \tag{1.29}$$

where V is the spatial volume. This is just a *mechanical* system with a finite number of degrees of freedom. The equations of motion are

$$\ddot{A}_i^a + [A_i^a(A_j^b A_j^b) - A_j^a(A_j^b A_i^b)] = 0 . \tag{1.30}$$

Let us look for solutions of the form $A_1^1(t) = A_2^2(t) = A(t)$, with all other components vanishing. Then only one degree of freedom is left and the equations (1.30) are reduced to $\ddot{A} + A^3 = 0$. This equation describes an unharmonical oscillator with the potential $U = A^4/4$. The solution can be easily found if one writes down the energy integral of motion,

$$\frac{1}{g^2}\left(\dot{A}^2 + \frac{1}{2}A^4\right) = \frac{E}{V} \equiv \epsilon . \tag{1.31}$$

Equation (1.31) may be easily integrated:

$$\int \frac{dA}{\sqrt{2\epsilon g^2 - A^4}} = \frac{t - t_0}{\sqrt{2}} .$$

The integral can be expressed through elliptic functions, and we finally

obtain

$$A(t) = C_0 \,\mathrm{cn}\left[C_0(t-t_0),\ \frac{1}{\sqrt{2}}\right], \qquad C_0 = \left(2\epsilon g^2\right)^{1/4} \tag{1.32}$$

in the notation of Ryzhik and Gradshtein. The period of oscillation is inversely proportional to the amplitude: $T \propto 1/C_0$. This, of course, could be seen without solving the equation: the only scale in the Yang–Mills equations is provided by a characteristic amplitude of the field whose dimension is $[A] = m \sim 1/T$.

The solution (1.32) is a non-linear standing wave. By a Lorentz boost, one can obtain as well solutions describing nonlinear propagating waves. These solutions do not seem to have a particular physical significance, but maybe their meaning has not yet been unravelled.

An even more peculiar picture emerges if we choose a somewhat less restrictive ansatz $A_1^1(t) = A(t)$, $A_2^2(t) = B(t)$, leaving *two* dynamical degrees of freedom. For the fields of this class, the Lagrangian

$$g^2\mathcal{L} = \frac{1}{2}[\dot{A}^2 + \dot{B}^2] - \frac{A^2B^2}{2} \tag{1.33}$$

describes the motion in the (AB)-plane with a potential $U(A,B) \propto A^2B^2$. The profiles of constant potential are hyperbolae, and the problem is very much similar to the problem of the so-called Sinai billiard with concave boards. This motion is *chaotic*, i.e., for general initial conditions, the integral trajectory does not have a decent regular form, but is ergodic and densely covers the whole surface of constant energy in the phase space of the system. After few oscillations (with characteristic time $\sim 1/A_{\mathrm{char}}$), the initial conditions are completely forgotten (Lyapunov instability). This sea of chaos has, however, islands of some ordered behavior. For some special initial conditions, the phase space trajectory is closed and regular. One example is the initial conditions with $A = B, \dot{A} = \dot{B}$ in which case the solution is given by Eq. (1.32). There are many other such solutions. Some examples found by Matinian and Savvidy are shown in Fig. 1.1, which is taken from the book [12]. §

The main physical conclusion is that the Yang–Mills system is not exactly solvable (otherwise, the trajectories would have a regular rather than

§I acknowledge the courtesy of the authors who provided me with the corresponding postscript file.

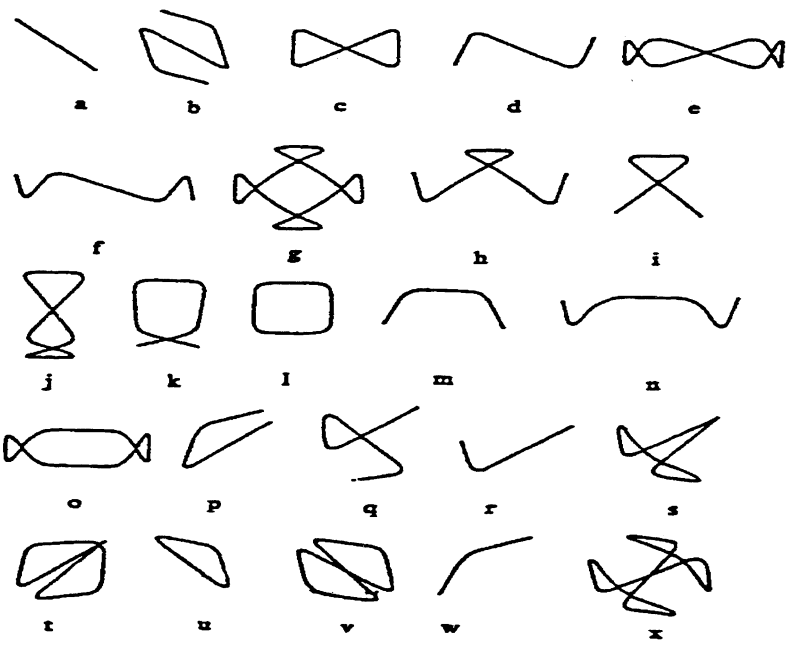


Fig. 1.1 Some closed trajectories in the (A, B) plane for the system (1.33).

chaotic form). This is in contrast to, say, the KdV equation and to some early hopes.

Lecture 2

Instantons

In this lecture we will study classical Yang–Mills theory in Euclidean space–time. We will obtain some non-trivial solutions with finite action. As we will see later, these solutions play an important role in the quantum theory.

Thus, we change $t \to -i\tau$, where τ is the Euclidean time. The Euclidean Yang–Mills action is

$$S_E \stackrel{\text{def}}{=} -iS_M(t = -i\tau) = \frac{1}{4g^2} \int d^4x F^a_{\mu\nu} F^a_{\mu\nu} , \tag{2.1}$$

where $d^4x = d^3x d\tau$ and $F^a_{\mu\nu}$ is an antisymmetric Euclidean tensor, so that S_E is positive definite.

2.1 Topological Charge.

Let us consider field configurations with a finite Euclidean action. The field strength $F_{\mu\nu}$ should fall off fast enough for the integral in Eq. (2.1) to converge. This does not necessarily mean that the vector potential A_μ vanishes at infinity as well. It suffices if it tends to a *pure gauge* potential

$$A_\mu = -i(\partial_\mu \Omega)\Omega^\dagger , \tag{2.2}$$

for which the tensor $F_{\mu\nu}$ vanishes. In general, Ω and A_μ depend on the direction n_μ ($n_\mu^2 = 1$) in which Euclidean infinity is approached. The matrix $\Omega(n_\mu)$ realizes a mapping of S^3 onto the gauge group. The space of all such mappings represents an infinite set of isolated components such that the mappings belonging to different components cannot be continuously

deformed into one another. Mathematicians call these components homotopy classes. For the mappings $S^3 \to G$, these classes are characterized by an integer number q, which we will call *topological charge.* A mathematical notation for this fact is $\pi_3(G) = \mathbb{Z}$, which is true for all simple Lie groups.

Consider the group $SU(2)$. Topologically, $SU(2) \equiv S^3$ and we are dealing with the mappings $S^3 \to S^3$. An example of a topologically nontrivial mapping with $q = 1$ is

$$\Omega(n_\mu) \equiv \Omega_0(n_\mu) = n_0 + in_i\sigma_i \ , \tag{2.3}$$

where σ_i are the Pauli matrices. Eq. (2.3) simply represents a parametrization of the distant Euclidean sphere by unitary matrices [see Eq. (A.12)]. Each $SU(2)$ matrix corresponds to a distinct point on S^3 and hence the *degree of mapping* is equal to one. An example of the mapping with $q = -1$ is $\Omega_0^\dagger(n_\mu)$. This mapping is also bijective, but the sphere is kind of "covered in the opposite direction": the corresponding Jacobian has the opposite sign. An example of a mapping with degree $q = 7$ is $\Omega_0^7(n_\mu)$ [seven different points n_μ on S^3 correspond to one and the same $SU(2)$ matrix], etc.

It is a very important fact that the topological charge of any mapping can be expressed as an integral over the distant 3-sphere. Before considering the four-dimensional theory of interest, let us discuss a very simple model: Abelian gauge theory in 2 Euclidean dimensions (the Schwinger model). Consider configurations with finite action which tend to the pure gauge $A_\mu = -i(\partial_\mu\Omega)\Omega^\dagger$, $\Omega(x) = e^{i\chi(x)}$ at Euclidean infinity. The field may depend on the direction ϕ along which infinity is approached. Now, $\Omega(\phi) = e^{ia(\phi)}$ realizes a mapping $S^1 \to U(1) \equiv S^1$. Obviously, the topological charge is the number of windings q: $\pi_1[U(1)] = Z$. For mappings of charge q, $a(\phi + 2\pi) = a(\phi) + 2\pi q$. It is easy to see now that q can be expressed in integral form,

$$q = \frac{1}{2\pi}\int_0^{2\pi} A_\phi d\phi = \frac{1}{2\pi}\oint_C \epsilon_{\mu\nu}A_\nu d\sigma_\mu \ , \tag{2.4}$$

where in the last formula $d\sigma_\mu$ is the length element of a distant contour C (not necessarily a "round" circle) multiplied by a unit normal vector pointing outside. Let us recall now that we are interested in the gauge field configurations defined on the whole two-dimensional plane. Using Gauss's

theorem, we immediately obtain

$$q = \frac{1}{2\pi}\iint d^2x\ \partial_\mu(\epsilon_{\mu\nu}A_\nu) = \frac{1}{4\pi}\iint d^2x\ \epsilon_{\mu\nu}F_{\mu\nu}\ . \tag{2.5}$$

The fact that the integral on the right side of Eq. (2.5) is quantized to integer values depends crucially on the fact that we required $\Omega(\phi)$ to be a uniquely defined function, i.e. we assumed our gauge group to be compact. We are forced to do so if our theory includes *charged* fields besides the gauge field A_μ. In pure photodynamics, the only effect of a gauge transformation is the shift of A_μ by $\partial_\mu\chi$, and the compact nature of the group is not seen. But the charged fields are multiplied by the exponential factor $e^{i\chi} \in U(1)$. The quantization condition follows from the fact that $e^{i\chi(x)}$ is uniquely defined. When going to the standard normalization for A_μ, the integrals in Eq. (2.4, 2.5) are multiplied by the electric charge of the gauge field. The quantization condition (2.5) has the same origin as the widely known condition of the quantization of the magnetic flux $\Phi = \frac{e}{2\pi}\iint dxdyH_z(x,y)$ in superconductors. Mathematicians call the integral in Eq. (2.5) the *first Chern class*. The current $j_\mu = \frac{1}{2\pi}\epsilon_{\mu\nu}A_\nu$ is called the *Chern–Simons* current.

This mathematical construction can be generalized to fiber bundles on arbitrary spaces of even dimension. We are interested in non-Abelian fiber bundles in the four-dimensional world. Let us first write the result. For nonsingular gauge fields tending to a pure gauge (2.2) at infinity, the topological charge q of a mapping $\Omega(n_\mu)$ is expressed through the integral (called *second Chern class* or *Pontryagin index*)

$$q = \frac{1}{16\pi^2}\int d^4x\ \mathrm{Tr}\{F_{\mu\nu}\tilde{F}_{\mu\nu}\}\ , \tag{2.6}$$

where $\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}$ [with the convention (N.15)]. To understand why this is correct, note that the integrand in (2.6) is a total derivative:

$$\frac{1}{16\pi^2}\mathrm{Tr}\{F_{\mu\nu}\tilde{F}_{\mu\nu}\} = \partial_\mu K_\mu = \partial_\mu\left[\frac{1}{8\pi^2}\epsilon_{\mu\nu\alpha\beta}\mathrm{Tr}\{A_\nu(\partial_\alpha A_\beta - \frac{2i}{3}A_\alpha A_\beta)\}\right] \tag{2.7}$$

(check it!). By Gauss' theorem,

$$q = \int_{S^3} d\sigma_\mu K_\mu = -\frac{1}{24\pi^2}\int d\sigma_\mu\ \epsilon_{\mu\nu\alpha\beta}\mathrm{Tr}\left\{(\partial_\nu\Omega)\Omega^\dagger(\partial_\alpha\Omega)\Omega^\dagger(\partial_\beta\Omega)\Omega^\dagger\right\} \tag{2.8}$$

[the asymptotic property $F_{\alpha\beta} \to 0$ and the form (2.2) for A_μ were used]. Substituting the particular mapping (2.3) with $q = 1$, one can explicitly check that the equality (2.8) is fulfilled, indeed. It is easy to understand why: the integrand in Eq. (2.8) is invariant with respect to the right or left multiplication of $\Omega(x)$ by an element of $SU(2)$ and is therefore nothing but the Haar measure [see Eqs. (A.6), (A.7)] on the group parametrized as in Eqs. (A.12), (2.3). The coefficient is chosen such that the volume of the group is normalized to 1. The integral (2.8) is not changed if our parametrization is deformed to a new one, $\Omega(x) \to \omega_0(x)\Omega(x)$, in such a way that a continuous family of gauge transformations $\omega(\alpha, x)$ with $\omega(0, x) = 1$, $\omega(1, x) = \omega_0(x)$ exists [i.e. when $\omega_0(x)$ is topologically trivial].* In general, $q[\Omega_1\Omega_2] = q[\Omega_1] + q[\Omega_2]$.

The instanton is a configuration of topological charge $q = 1$ with the minimal action, i.e. a solution of the equations of motion with $q = 1$. The construction of such a solution is done in several steps. First, note that the action integral can be written as

$$\begin{aligned} S &= \frac{1}{2g^2}\int d^4x \mathrm{Tr}\{F_{\mu\nu}F_{\mu\nu}\} \\ &= \frac{1}{4g^2}\int d^4x \mathrm{Tr}\{(F_{\mu\nu} - \tilde{F}_{\mu\nu})(F_{\mu\nu} - \tilde{F}_{\mu\nu})\} + \frac{1}{2g^2}\int d^4x \mathrm{Tr}\{F_{\mu\nu}\tilde{F}_{\mu\nu}\} \\ &= \frac{1}{4g^2}\int d^4x \mathrm{Tr}\{(F_{\mu\nu} - \tilde{F}_{\mu\nu})(F_{\mu\nu} - \tilde{F}_{\mu\nu})\} + \frac{8\pi^2 q}{g^2} \geq \frac{8\pi^2 q}{g^2} . \end{aligned} \tag{2.9}$$

Thus, the action of *any* field configuration with a positive q has the absolute lower bound $8\pi^2 q/g^2$. It suggests that the minimal value of the action is realized on *self-dual* configurations

$$F_{\mu\nu} = \tilde{F}_{\mu\nu} \tag{2.10}$$

and is just $8\pi^2 q/g^2$.

*One of the ways to see this is to take the variation of the integrand (2.7) and integrate it. We have

$$\delta q \sim \int d\sigma_\mu \epsilon_{\mu\nu\alpha\beta} \mathrm{Tr}\,\{\delta A_\nu(\partial_\alpha A_\beta - iA_\alpha A_\beta)\} \sim \int d\sigma_\mu \mathrm{Tr}\,\{\delta A_\nu \tilde{F}_{\mu\nu}\} = 0$$

due to the fact that our field is pure gauge at infinity.

For configurations with negative q, we can write

$$\begin{aligned} S &= \frac{1}{4g^2}\int d^4x \mathrm{Tr}\{(F_{\mu\nu}+\tilde{F}_{\mu\nu})(F_{\mu\nu}+\tilde{F}_{\mu\nu})\} - \frac{1}{2g^2}\int d^4x \mathrm{Tr}\{F_{\mu\nu}\tilde{F}_{\mu\nu}\} \\ &= \frac{1}{4g^2}\int d^4x \mathrm{Tr}\{(F_{\mu\nu}+\tilde{F}_{\mu\nu})(F_{\mu\nu}+\tilde{F}_{\mu\nu})\} - \frac{8\pi^2 q}{g^2} \geq \frac{8\pi^2|q|}{g^2} \ , \quad (2.11) \end{aligned}$$

and the minimum $8\pi^2|q|/g^2$ is realized on anti-self-dual configurations $F_{\mu\nu} = -\tilde{F}_{\mu\nu}$.

2.2 Explicit Solutions.

To prove that this is indeed the case, we should solve the self-duality equations (2.10) † explicitly. Let us do this for $q = 1$.

Note first that any self-dual antisymmetric tensor $S^{\text{s.d.}}_{\mu\nu}$ can be put into the form

$$(S_{\mu\nu})^{\text{s.d.}} = \eta^a_{\mu\nu} c^a \ ,$$

where $\eta^a_{\mu\nu}$ are the so-called 't Hooft symbols

$$\eta^a_{00} = 0 \ , \quad \eta^a_{ij} = \epsilon_{aij} \ , \quad \eta^a_{0i} = -\delta_{ai} \ , \quad \eta^a_{i0} = \delta_{ai} \ . \qquad (2.12)$$

The objects (2.12) are not so mysterious. They can be understood if one notes that an antisymmetric self-dual tensor represents an irreducible representation $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ of the Lorentz group $\equiv SU(2) \times SU(2)$ and $\eta^a_{\mu\nu}$ are nothing but the Clebsch–Gordon coefficients for constructing $(S_{\mu\nu})^{\text{s.d.}}$ out of two 4-vectors:

$$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \otimes \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \oplus \cdots \ .$$

Likewise, an anti-self-dual antisymmetric tensor can be written as

$$(A_{\mu\nu})^{\text{a.s.d.}} = \bar{\eta}^a_{\mu\nu} c^a \ ,$$

where $\bar{\eta}^a_{\mu\nu}$ differ from $\eta^a_{\mu\nu}$ by the sign of the components $\bar{\eta}^a_{0i} = -\bar{\eta}^a_{i0} = \delta_{ai}$. There are many beautiful relations involving the symbols $\eta^a_{\mu\nu}$. Here we will

†They are first-order and are thus much simpler than the general equations of motion (1.20).

For configurations with negative q, we can write

$$\begin{aligned} S &= \frac{1}{4g^2}\int d^4x \mathrm{Tr}\{(F_{\mu\nu}+\tilde{F}_{\mu\nu})(F_{\mu\nu}+\tilde{F}_{\mu\nu})\} - \frac{1}{2g^2}\int d^4x \mathrm{Tr}\{F_{\mu\nu}\tilde{F}_{\mu\nu}\} \\ &= \frac{1}{4g^2}\int d^4x \mathrm{Tr}\{(F_{\mu\nu}+\tilde{F}_{\mu\nu})(F_{\mu\nu}+\tilde{F}_{\mu\nu})\} - \frac{8\pi^2 q}{g^2} \geq \frac{8\pi^2|q|}{g^2} \ , \quad (2.11) \end{aligned}$$

and the minimum $8\pi^2|q|/g^2$ is realized on anti-self-dual configurations $F_{\mu\nu} = -\tilde{F}_{\mu\nu}$.

2.2 Explicit Solutions.

To prove that this is indeed the case, we should solve the self-duality equations (2.10) † explicitly. Let us do this for $q = 1$.

Note first that any self-dual antisymmetric tensor $S^{\text{s.d.}}_{\mu\nu}$ can be put into the form

$$(S_{\mu\nu})^{\text{s.d.}} = \eta^a_{\mu\nu} c^a \ ,$$

where $\eta^a_{\mu\nu}$ are the so-called 't Hooft symbols

$$\eta^a_{00} = 0 \ , \quad \eta^a_{ij} = \epsilon_{aij} \ , \quad \eta^a_{0i} = -\delta_{ai} \ , \quad \eta^a_{i0} = \delta_{ai} \ . \qquad (2.12)$$

The objects (2.12) are not so mysterious. They can be understood if one notes that an antisymmetric self-dual tensor represents an irreducible representation $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ of the Lorentz group $\equiv SU(2) \times SU(2)$ and $\eta^a_{\mu\nu}$ are nothing but the Clebsch–Gordon coefficients for constructing $(S_{\mu\nu})^{\text{s.d.}}$ out of two 4-vectors:

$$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \otimes \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \oplus \cdots \ .$$

Likewise, an anti-self-dual antisymmetric tensor can be written as

$$(A_{\mu\nu})^{\text{a.s.d.}} = \bar{\eta}^a_{\mu\nu} c^a \ ,$$

where $\bar{\eta}^a_{\mu\nu}$ differ from $\eta^a_{\mu\nu}$ by the sign of the components $\bar{\eta}^a_{0i} = -\bar{\eta}^a_{i0} = \delta_{ai}$. There are many beautiful relations involving the symbols $\eta^a_{\mu\nu}$. Here we will

†They are first-order and are thus much simpler than the general equations of motion (1.20).

need two of them:

$$\begin{aligned} \epsilon^{abc}\eta^b_{\mu\rho}\eta^c_{\nu\sigma} &= \delta_{\mu\nu}\eta^a_{\rho\sigma} - \delta_{\mu\sigma}\eta^a_{\rho\nu} + \delta_{\rho\sigma}\eta^a_{\mu\nu} - \delta_{\rho\nu}\eta^a_{\mu\sigma} \ , \\ \eta^a_{\mu\nu}\eta^a_{\mu\nu} &= 12 \ . \end{aligned} \tag{2.13}$$

The simplest ansatz for a self-dual field density tensor $F^a_{\mu\nu}(x)$ is

$$F^a_{\mu\nu}(x) \ = \ \eta^a_{\mu\nu}B(x) \ ,$$

where $B(x)$ is a scalar function. However, not any such tensor can play the role of a field density tensor expressible via potentials according to Equation (1.14). Let us adopt the ansatz

$$A^a_\mu(x) \ = \ \eta^a_{\mu\nu}\partial_\nu \ln\phi(x) \tag{2.14}$$

and try to find out under what conditions the corresponding field strength tensor would be self-dual. Using the definition (1.14) and the first identity in (2.13), we obtain

$$F^a_{\mu\nu} = \eta^a_{\nu\rho}[\partial_\mu\partial_\rho \ln\phi + (\partial_\mu \ln\phi)(\partial_\rho \ln\phi)] - (\mu \leftrightarrow \nu) + \eta^a_{\mu\nu}(\partial_\rho \ln\phi)^2 \ . \tag{2.15}$$

In general, only the last term has a nice self-dual form. A sufficient condition for the whole expression to be self-dual is

$$[\ldots]_{\mu\rho} = \frac{\partial_\mu\partial_\rho\phi}{\phi} \propto \delta_{\mu\rho}$$

for the square bracket in Eq. (2.15).

The simplest way to fulfill this condition is to assume $\phi(x) = x^2/\rho^2$. If we, however, substitute such a $\phi(x)$ in Eq. (2.14) and calculate the corresponding $F^a_{\mu\nu}$, we find that $F^a_{\mu\nu} = 0$. This is not the solution we were looking for, but the corresponding field configuration,

$$A^a_\mu = 2\eta^a_{\mu\nu}x_\nu/x^2 \ , \tag{2.16}$$

is also meaningful. It is a topologically nontrivial pure gauge configuration (2.2) with $\Omega = \Omega_0$ defined in Eq. (2.3), $n_\mu \equiv x_\mu/\sqrt{x^2}$.

The instanton solution is given by the ansatz (2.14) with

$$\phi(x) = 1 + x^2/\rho^2 \ . \tag{2.17}$$

It has the form

$$A_\mu^a = \frac{2\eta_{\mu\nu}^a x_\nu}{x^2+\rho^2} \ . \tag{2.18}$$

The field strength is

$$F_{\mu\nu}^a = -\frac{4\rho^2\eta_{\mu\nu}^a}{(x^2+\rho^2)^2} \ . \tag{2.19}$$

The explicit calculation of the action integral [using the second relation in Eq. (2.13)] gives the value $8\pi^2/g^2$ for the action as it should be.

The solution (2.18, 2.19) is localized in Euclidean space and thus represents an Euclidean soliton. At first sight, this is a counterexample to the no-go theorem proved in the previous lecture. But it is not the case. We have proven the theorem excluding the existence of *static* localized solutions in Yang–Mills theory. Such solutions would be characterized by an energy which is inversely proportional to the characteristic scale. On the other hand, Euclidean solutions are characterized by their *action*, which is dimensionless and *does* not change when varying the scale. The action of an instanton is always $8\pi^2/g^2$ independently of its characteristic size ρ, which is a free parameter of the solution. Another free parameter is the position of the instanton center: the solution (2.18) is centered at $x = 0$, but the problem is translationally invariant and, shifting $x \to x - x_0$, one can equally well obtain a solution centered at $x = x_0$.

The solution (2.18) is written in a particular gauge. Many other forms of the solution can be obtained by applying local gauge transformations to (2.18). Actually, one can prove that *all* the solutions of the Euclidean equations of motion in the topological sector $q = 1$ can be represented in the form (2.18), up to rescalings, translations, and gauge transformations.

All our reasoning can be transferred without essential changes to the case $q = -1$. One should only substitute $\bar\eta_{\mu\nu}^a$ for $\eta_{\mu\nu}^a$ in the ansatz (2.14). The dual 't Hooft symbols satisfy the same relations (2.13). Thus, the choice (2.17) provides for an anti-self-dual field strength tensor. We obtain the same solutions (2.18, 2.19) up to the change $\eta \to \bar\eta$. At large $\sqrt{x^2}$, the anti-instanton solution tends to a pure gauge configuration (2.2) with $\Omega = \Omega_0^+ = n_0 - in_i\sigma_i, \ \ n_\mu = x_\mu/\sqrt{x^2}$.

The equations (2.10) can also be solved for higher q. This was done by Atiyah, Drinfeld, Hitchin, and Manin. The general solution looks complicated, but a restricted class of solutions with higher topological charge q,

found by 't Hooft, has a very simple form (see **Problem 2**).

A final comment is that instanton solutions exist also for higher unitary (and not only unitary) groups. They all can be obtained, however, by gauge group rotations from an $su(2)$ solution with a trivial natural embedding $su(2) \subset su(N)$.‡

Problem 1. Apply a gauge transformation with $\Omega = \Omega_0^\dagger$ to the configuration (2.18). Such a transformation unwinds the singularity associated with a nontrivial mapping at infinity, but creates it at $x = 0$. Show that the gauge transformed field has the form

$$A_\mu^a = \frac{2\rho^2 \bar{\eta}_{\mu\nu}^a x_\nu}{x^2(x^2+\rho^2)} \tag{2.20}$$

with the dual 't Hooft symbol.

Solution. Let us rewrite the instanton solution (2.18) in the form

$$\begin{aligned} A_\mu &= A_\mu^a t^a = \frac{x^2}{x^2+\rho^2}[-i(\partial_\mu \Omega_0)\Omega_0^\dagger] \\ &= \frac{i\rho^2}{x^2+\rho^2}(\partial_\mu \Omega_0)\Omega_0^\dagger - i(\partial_\mu \Omega_0)\Omega_0^\dagger \, . \end{aligned} \tag{2.21}$$

We have

$$\begin{aligned} A_\mu^{\Omega_0^\dagger} &= \Omega_0^\dagger A_\mu \Omega_0 - i(\partial_\mu \Omega_0^\dagger)\Omega_0 = \frac{i\rho^2}{x^2+\rho^2}\Omega_0^\dagger \partial_\mu \Omega_0 + 0 \\ &= -\frac{i\rho^2}{x^2+\rho^2}(\partial_\mu \Omega_0^\dagger)\Omega_0 \, , \end{aligned} \tag{2.22}$$

which coincides with (2.20).

‡We have used here lower case letters to denote the Lie algebras (where the gauge fields A_μ belong) as opposed to the Lie groups, to make mathematicians happy. By a "natural" embedding, we understand e.g. the embedding realized by the generators t^1, t^2, t^3 living in the 2×2 block in the left upper corner of $su(N)$ [cf. Eq. (A.16)]. There are many other embeddings. For example, one of the embeddings $su(2) \subset su(3)$ is realized by the matrices t^2, t^5, t^7. An $su(3)$ field configuration (2.18) with $a = 2, 5, 7$ is a perfect Euclidean solution, but with $q = 4$ rather than $q = 1$. Also, other nontrivial embeddings lead to solutions with some higher q.

Problem 2. Show that the configuration

$$A_\mu^a = -\bar{\eta}_{\mu\nu}^a \partial_\nu \ln\left[1 + \sum_{i=1}^{q} \frac{\rho_i^2}{(x - x_{i0})^2}\right] \tag{2.23}$$

is a solution of the equations of motion (though not the most general one) with the (positive) topological charge q. (*'t Hooft*).

Solution. Let $A_\mu^a = -\bar{\eta}_{\mu\nu}^a \partial_\nu \ln \phi(x)$. Then

$$F_{\mu\nu}^a = \bar{\eta}_{\nu\rho}^a [-\partial_\mu \partial_\rho \ln\phi + (\partial_\mu \ln\phi)(\partial_\rho \ln\phi)] - (\mu \leftrightarrow \nu) + \bar{\eta}_{\mu\nu}^a (\partial_\rho \ln\phi)^2 \ . \tag{2.24}$$

Using the relations (2.13) and the relation

$$\epsilon_{\mu\nu\alpha\beta}\bar{\eta}_{\beta\gamma}^a = \delta_{\alpha\gamma}\bar{\eta}_{\mu\nu}^a + \delta_{\mu\gamma}\bar{\eta}_{\nu\alpha}^a + \delta_{\nu\gamma}\bar{\eta}_{\alpha\mu}^a \ , \tag{2.25}$$

we find

$$\tilde{F}_{\mu\nu}^a = \bar{\eta}_{\nu\rho}^a [-\partial_\mu \partial_\rho \ln\phi + (\partial_\mu \ln\phi)(\partial_\rho \ln\phi)] - (\mu \leftrightarrow \nu) - \bar{\eta}_{\mu\nu}^a \Box \ln\phi \ . \tag{2.26}$$

Thereby, the field is self-dual if the condition

$$\Box \ln\phi + (\partial_\rho \ln\phi)^2 = 0 \tag{2.27}$$

is fulfilled. Eq. (2.27) implies $\Box\phi = 0$, and we see that the argument of the logarithm in Eq. (2.23) satisfies this equation everywhere except the points $x = x_{i0}$ where the gauge potential (but not the field strength!) is singular.

Let us calculate the Pontryagin index (2.6) for the field configuration (2.23). As before, we are integrating a total derivative so that the only contributions come from the boundary. The important distinction, compared to the regular gauge solution, is that the boundary does not now include a large sphere S^3 at Euclidean infinity: the gauge potential (2.23) falls off at infinity pretty fast, $\Omega^{\text{sing. gauge}}(\infty) = 1$, and the integral (2.8) is just zero in this case. Nonvanishing boundary contributions come, however, from small spheres surrounding the singularities at $x = x_{i0}$. At the vicinity of each such singularity, the configuration (2.23) behaves as a pure gauge configuration (2.2) with $\Omega = \Omega_0^\dagger$, $n_\mu = (x - x_{i0})_\mu / \sqrt{(x - x_{i0})^2}$. The contribution of each such small sphere to the Pontryagin index is 1 [the integral like in Eq. (2.8) with $\Omega = \Omega_0^\dagger$ would give -1, but one should bear in mind that the

normal vector $d\sigma_\mu$ is now directed not outwards, but inwards]. We have q singular points, and the total Pontryagin index is equal to q.

One can think of 't Hooft's solution (2.23) as a combination of q instantons with the same orientation in group space (this is especially transparent when the instantons are far from each other). A general ADHM solution with topological charge q is also a combination of q instantons, but with arbitrary group orientations.

Lecture 3

Path Integral in Quantum Mechanics

3.1 Conventional Approach.

Quantum Chromodynamics is a quantum field theory. Quantum field theory is a quantum mechanical system with an infinite continuum number of degrees of freedom. The latter circumstance makes the theory complicated, but the basic ideology of QFT is the same as in conventional quantum mechanics: we have a Hilbert space of quantum states, observables, and an Hamiltonian operator which determines the time evolution of the wave function.* Solving the stationary Schrödinger equation, one could *in principle* find the spectrum of our QFT system (provided it exists and is well defined; as we will see later, not all classical field theories have consistent quantum counterparts.), etc.

This language is, however, not convenient for applications. Wave functions are actually not functions but functionals depending on an infinite number of variables. Also, the main object of interest in QFT applications is not the spectrum (though it is also important), but a set of scattering amplitudes, i.e. the S-matrix. The latter can be conveniently calculated in the framework of path integral techniques. In this lecture, we will develop this first for quantum mechanical systems, having a finite number of degrees of freedom. We will generalize to quantum field theories and in particular to QCD in subsequent lectures.

*Unfortunately, this approach is not possible in quantum gravity, because there time cannot be treated as a flat *independent* variable, but is intertwined with the other (dynamic) variables. This *is* the reason why quantum gravity does not yet exist today in the form of a consistent theory. This is the major challenge for the theorists today.

Classical mechanics can be described in the Hamiltonian or in the Lagrangian language. The former involves Hamilton's equations of motion, describing the dynamics of the system in phase space and implying conservation of energy. The latter involves Lagrange's equations, describing the dynamics in configuration space. Its underlying basic axiom is the minimal action principle. These two ways of description are equivalent to each other, and both can be generalized to the quantum case. The conventional description of the quantum mechanics by means of the Schrödinger equation represents the quantum version of the Hamiltonian description. The path integral formulation of quantum mechanics, developed originally by Feynman, is the quantum version of the Lagrange description. Like the minimal action principle, the Feynman path integral can be postulated as the basic axiom of the theory. Here we will derive it starting from the Schrödinger formulation.

Let us consider a quantum system involving just one dynamical degree of freedom q with the Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2} + V(q) \ , \tag{3.1}$$

where $\hat{p} = -id/dq$ is the momentum operator (from now on, hats will indicate operators). The wave function of the system satisfies the Schrödinger equation,

$$i\frac{d\psi}{dt} = \hat{H}\psi \ . \tag{3.2}$$

If we fix the initial conditions $\psi(q, t_i)$, the formal solution of the equation (3.2) reads $\psi(q,t) = \hat{U}(t - t_i)\psi(q, t_i)$, where

$$\hat{U}(t - t_i) \ = \ \exp\{-i(t - t_i)\hat{H}\} \tag{3.3}$$

is the evolution operator. Let us consider the *kernel* of the evolution operator, i.e. the matrix element

$$\mathcal{K}(q_f, q_i; t_f - t_i) \ = \ \langle q_f|\hat{U}(t_f - t_i)|q_i\rangle \ \equiv \ \langle q_f, t_f|q_i, t_i\rangle \ . \tag{3.4}$$

It describes the probability amplitude that the system will find itself at point q_f at $t = t_f$ provided it was located at $q = q_i$ at $t = t_i$. The quantity $\mathcal{K}$ plays a fundamental role in the whole approach. The spectral

decomposition of the kernel of the evolution operator reads:

$$\mathcal{K}(q_f, q_i; t_f - t_i) = \sum_k \psi_k(q_f)\psi_k^*(q_i)e^{-i\mathcal{E}_k(t_f - t_i)} , \tag{3.5}$$

where $\psi_k(q)$ are the eigenstates of $\hat{H}$, with eigenvalues $\mathcal{E}_k$. Let $t_f - t_i = \Delta t$ be small. We have

$$\langle q_f|\hat{U}(\Delta t)|q_i\rangle = \langle q_f|1 - i\hat{H}\Delta t|q_i\rangle = \int dp\langle q_f|p\rangle\langle p|1 - i\hat{H}\Delta t|q_i\rangle .$$

Here $|p\rangle$ is a state with definite momentum p and wave function $\psi_p(q) = \frac{1}{\sqrt{2\pi}}e^{ipq}$. By definition,

$$\langle q|p\rangle = \frac{1}{\sqrt{2\pi}}e^{ipq}, \ \langle p|q\rangle = \frac{1}{\sqrt{2\pi}}e^{-ipq} . \tag{3.6}$$

If the Hamiltonian has the simple form (3.1), we can also write

$$\langle p|\hat{H}|q\rangle = \frac{1}{\sqrt{2\pi}}e^{-ipq}H(p,q) ,$$

where $H(p,q)$ is just the right side of Eq. (3.1), where p is now a c-number.[†] We have

$$\mathcal{K}(q_f, q_i; \Delta t) = \frac{1}{2\pi}\int dp \exp\{ip(q_f - q_i) - iH(p, q_i)\Delta t\} . \tag{3.7}$$

To find the evolution operator over a finite time interval, divide the latter into a very large number n of very small intervals Δt. The evolution operator is then expressed as a convolution of an infinite number of infinitesimal

[†]If the Hamiltonian has a more intricate form involving terms of the type $\propto p^2q^2$ mixing momenta and coordinates, the situation is more difficult. To perform quantization, one should resolve the ordering ambiguity problem. As the Hamiltonian of QCD does not mix generalized coordinates and generalized momenta, we do not have this problem here and need not worry about it. Just note the following: *(i)* a path integral can be defined for *any* Hamiltonian no matter what particular way of the operator ordering is chosen *(ii)* very frequently, the latter can be fixed uniquely due to symmetry constraints.

evolution operators

$$\mathcal{K}(q_f, q_i; t_f - t_i) =$$
$$\lim_{n\to\infty} \int dq_{n-1} \cdots \int dq_1 \ \langle q_f|\hat{U}(\Delta t)|q_{n-1}\rangle \cdots \langle q_1|\hat{U}(\Delta t)|q_i\rangle$$
$$= \lim_{n\to\infty} \int \exp\{ip_n(q_f - q_{n-1}) + \ldots + ip_1(q_1 - q_i) \ -$$
$$i\Delta t[H(p_n, q_{n-1}) + \ldots + H(p_1, q_i)]\} \frac{dp_n}{2\pi} \frac{dp_{n-1}dq_{n-1}}{2\pi} \cdots \frac{dp_1 dq_1}{2\pi} \ . \quad (3.8)$$

This can be formally written as

$$\mathcal{K}(q_f, q_i; t_f - t_i) \ = \ \int \exp\left\{ i \int_{t_i}^{t_f} [p\dot{q} - H(p,q)]dt \right\} \prod_t \frac{dp(t)dq(t)}{2\pi} , \quad (3.9)$$

with the boundary conditions

$$q(t_i) = q_i; \ \ q(t_f) = q_f \ , \quad (3.10)$$

while the function $p(t)$ is quite arbitrary and is integrated over. For the Hamiltonian (3.1), the integral over all dp_n in Eq. (3.8) is Gaussian and can be done easily:

$$\mathcal{K}(q_f, q_i; t_f - t_i) =$$
$$\mathcal{N} \lim_{n\to\infty} \int \exp\left\{ i \left[\frac{(q_f - q_{n-1})^2}{2\Delta t} + \cdots + \frac{(q_1 - q_i)^2}{2\Delta t} \right.\right.$$
$$\left.\left. - \ \Delta t V(q_{n-1}) - \cdots \Delta t V(q_i)]\right\} \prod_{j=1}^{n} dq_j$$
$$\sim \int \prod_t dq(t) \exp\left\{ i \int_{t_i}^{t_f} \left[\frac{\dot{q}^2}{2} - V(q) \right] dt \right\} \ . \quad (3.11)$$

The infinite normalization factor $\mathcal{N}$ plays no role for our purposes and will be dropped. Eq. (3.11) is the original Feynman form of the path integral, where the integral is done over trajectories in configuration space rather than in phase space as it is the case in Eq. (3.9). The boundary conditions (3.10) are assumed. The integral in the exponent is just the action on a trajectory $q(t)$ connecting the points q_i, t_i and q_f, t_f.

The (semi-)classical limit of quantum theory corresponds to large action (in units of $\hbar$). In this case, the main contribution to the integral (3.11)

stems from the stationary points (minima) of the action, where many trajectories are added up with nearly identical phase factors.

All the results we have just derived are easily generalized to systems with several or many dynamical degrees of freedom. A slightly less trivial generalization, which we will need for some applications below, concerns the systems where the Hamiltonian represents not a scalar, but a matrix function on phase space. An example is the Pauli Hamiltonian $H = [\sigma_i(p_i - eA_i)]^2/(2m)$ for a spinning particle in an external magnetic field. The eigenfunctions $[\psi_k(q)]_p$ for such a system carry a spinor index p and the evolution operator also represents a matrix,

$$\mathcal{K}_{lp}(q_f, q_i; \Delta t) = \sum_k [\psi_k(q_f)]_l [\psi_k^*(q_i)]_p e^{-i\mathcal{E}_k \Delta t} . \tag{3.12}$$

Repeating all the steps of the derivation above, we can express (3.12) by a path integral which has the same form as Eq. (3.9), except that H and $\mathcal{K}$ are now assumed to be matrices and we have a *T-ordered* exponential,

$$T \exp\left\{ i \int_{t_i}^{t_f} [p\dot{q} - H(p,q)] dt \right\} ,$$

instead of an ordinary one.

A word of caution is in order here. Eqs. (3.8, 3.11) do not yet provide us with an operational definition of the quantum path integral, because the integrals over $\prod_j dq_j$ on the left side are not well defined: the integrand rapidly oscillates at large $|q_j|$ and does not fall off. To make the integrals convergent, we define the evolution operator through

$$\mathcal{K}(q_f, q_i;\ t_f - t_i) = \lim_{\epsilon \to +0} \mathcal{K}(q_f, q_i;\ t_f - t_i - i\epsilon) ,$$

i.e. as a convolution of a large number of infinitesimal operators over time $\widetilde{\Delta t} = (t_f - t_i - i\epsilon)/n$. For finite ϵ, the integrals converge and the only problem is to prove that the limit $\epsilon \to 0$ exists. As far as I understand, a mathematical proof of this fact is still absent (and that is why the very notion of Minkowski path integrals does not exist, for mathematicians, to date!). However, the existence of the limit follows from physical considerations. After all, we know that the finite time evolution operator (3.5) is well defined and, if the principles of quantum mechanics are correct, it *should* be represented as a convolution of many infinitesimal ones.

3.2 Euclidean Path Integral.

In the previous section we took the time variable to be slightly imaginary in order to make the integrals convergent. But we can equally well define and study the path integral assuming time to be *purely imaginary*. The Euclidean path integral thus defined is very much important in many quantum field theory applications.

Consider the evolution operator over an imaginary time interval $-i\beta$ ($t_i = 0,\ t_f = -i\beta$). According to Eq. (3.5), it is

$$\mathcal{K}(q_f, q_i; -i\beta) \ = \ \sum_k \psi_k(q_f)\psi_k^*(q_i)e^{-\beta\mathcal{E}_k} \ . \tag{3.13}$$

Note that when β is large, the main contribution to the sum comes from the ground (i.e. vacuum) state. After the transformation $t \to -i\tau$ in Eq. (3.11), we obtain

$$\mathcal{K}(q_f, q_i; -i\beta) \ \sim \ \int \prod_\tau dq(\tau) \exp\left\{ -\int_0^\beta \left[\frac{\dot{q}^2}{2} + V(q) \right] d\tau \right\} \ , \tag{3.14}$$

with $q(0) = q_i,\ q(\beta) = q_f$. Introduce the quantity

$$Z(\beta) \ = \ \int dq \ \mathcal{K}(q, q; -i\beta) = \sum_k e^{-\beta\mathcal{E}_k} \ . \tag{3.15}$$

This is nothing but the *partition function* of the system with temperature $T = 1/\beta$. Thereby, we have obtained a remarkable result: the partition function of a thermal system can be written as an Euclidean path integral over a finite imaginary time related to temperature.

In quantum mechanics and, especially, in quantum field theory, the ground state expectation values (vacuum averages) of various operators present a considerable interest. They can be expressed in the form

$$\langle 0|\hat{O}|0\rangle \ = \ \lim_{\beta\to\infty} \frac{\sum_k e^{-\beta\mathcal{E}_k}\langle k|O|k\rangle}{\sum_k e^{-\beta\mathcal{E}_k}} \ . \tag{3.16}$$

This can be written as a ratio of two path integrals

$$\begin{aligned} \langle O\rangle_0 \ &= \ \lim_{\beta\to\infty} Z^{-1}(\beta) \int dq \ \mathcal{K}(q, q; -i\beta) O(q) \\ &= \ \lim_{\beta\to\infty} Z^{-1}(\beta) \int \prod_\tau dq(\tau) O[q(0)] e^{-\int_0^\beta d\tau \mathcal{L}_E[q(\tau)]} \end{aligned} \tag{3.17}$$

with periodic boundary condition $q(\beta) = q(0)$. Here the integral is performed over all closed trajectories. The boundary values are not fixed as before, but instead are integrated over, with the only restriction that $q(0)$ and $q(\beta)$ coincide. We have limited ourselves to the case where the operator O depends only on q. A path integral representation can also be found for the vacuum average of an arbitrary operator $O(\hat{p}, q)$, but it would be a phase space path integral like in Eq. (3.9).

It is often convenient to introduce a *generating functional*

$$Z[\eta(\tau)] = \int \prod_{\tau} dq(\tau) e^{-\int_0^\beta d\tau \{\mathcal{L}_E[q(\tau)] - \eta(\tau) q(\tau)\}} . \tag{3.18}$$

Then

$$\langle O \rangle = Z^{-1}[0] O \left[\frac{\delta}{\delta \eta(0)}\right] Z[\eta(\tau)]\Big|_{\eta=0} . \tag{3.19}$$

Generating functionals can be equally well defined for Minkowski path integrals.

The developed formalism works for all quantum mechanical problems. As far as quantum field theory is concerned, it is not yet complete for two reasons. The first reason is that later on we are going to construct perturbation theory, where we will be interested in transition amplitudes connecting *in* and *out* asymptotic states describing free particles. The transition amplitude $\mathcal{K}_A = \langle A_f(\boldsymbol{x}), t_f | A_i(\boldsymbol{x}), t_i \rangle$ connecting states with definite values of the fields $A(\boldsymbol{x})$, which is the immediate analog of (3.4), does not have a direct relevance to QFT applications. One-particle states are created from the vacuum by creation operators. They are described by non-trivial wave function $\Psi[A(\boldsymbol{x})]$, which the amplitude $\mathcal{K}_A$ should be convoluted with. It turns out that a convenient formalism can be developed which allows one to calculate path integrals for the amplitudes $\langle out | in \rangle$ directly, without the necessity to handle $\mathcal{K}_A$ on intermediate stages.

The second and most important reason is that quantum field theories (and, in particular, QCD) involve fermions, while standard quantum mechanics involves only bosonic degrees of freedom. We have got to learn how to write path integrals with fermionic dynamical variables. The adequate technique for this is the so-called *holomorphic representation.* Before turning to fermions, let us see how this technique works for purely bosonic systems.

3.3 Holomorphic Representation.

This is a very concise presentation. See e.g. the books [4; 13] for more details.

The Hamiltonian of a system of free fields (describing the asymptotic states) is reduced to the Hamiltonian of an infinite number of oscillators. The relevant formalism can be introduced by simply studying the problem of a quantum oscillator with only one dynamical degree of freedom. Introduce the *classical* variables

$$a^* = \frac{1}{\sqrt{2\omega}}(\omega q - ip) \ , \quad a = \frac{1}{\sqrt{2\omega}}(\omega q + ip) \ , \tag{3.20}$$

so that the classical Hamiltonian acquires the form

$$H^{\rm cl}(p,q) \ = \ \frac{p^2}{2} + \frac{\omega^2 q^2}{2} \ = \ \omega a^* a \ . \tag{3.21}$$

The classical equations of motion are just $i\dot{a}^* = -\partial H^{\rm cl}/\partial a = -\omega a^*$ (and its complex conjugate) with the trivial solution $a^*(t) = a_0^* e^{i\omega t}$.

In quantum mechanics, a^* and a are upgraded to the creation and annihilation operators $\hat{a}^\dagger$, $\hat{a}$, with standard commutation relation

$$[\hat{a}, \hat{a}^\dagger] = 1 \ . \tag{3.22}$$

The quantum Hamiltonian is $\hat{H}^{\rm qu} = \omega \hat{a}^\dagger \hat{a}$ (with the zero-point energy $\omega/2$ subtracted). The oscillator eigenstates with energies $\mathcal{E}_k = \omega k$ are

$$|k\rangle \ = \ \frac{1}{\sqrt{k!}}(\hat{a}^\dagger)^k |0\rangle \ .$$

Usually, we write wave functions either in the coordinate or the momentum representation. Let us write them in the *holomorphic* representation, namely as a linear superposition of the wave functions

$$\psi_k(a^*) \ = \ \frac{1}{\sqrt{k!}}(a^*)^k \tag{3.23}$$

of the oscillator eigenstates. A generic wave function would be an entire (or holomorphic) function $\psi(a^*)$ of the variable a^*. The operator $\hat{a}^\dagger$ acting on $\psi(a^*)$ just multiplies it by a^* and

$$\hat{a}\psi \stackrel{\rm def}{=} d\psi(a^*)/da^* \ .$$

The operators $\hat{a}^\dagger$ and $\hat{a}$ thus defined satisfy the commutation relation (3.22). To be able to calculate the matrix elements of various operators, we have to define the scalar product in the Hilbert space with basis (3.23) in such a way that the states (3.23) are orthonormal $\langle n|m\rangle = \delta_{nm}$. One can check that the definition

$$(f,g) \stackrel{\text{def}}{=} \int [f(a^*)]^* g(a^*) e^{-a^* a} \frac{da^* da}{2\pi i} \tag{3.24}$$

satisfies this criterium (do this by introducing polar variables $a = re^{i\phi}$, and using $da^* da = 2irdrd\phi$). The action of various operators is defined as

$$(\hat{A}f)(a^*) \;=\; \int A(a^*, b) f(b^*) e^{-b^* b} \frac{db^* db}{2\pi i} \;. \tag{3.25}$$

The function $A(a^*, b)$ is called the *kernel* of the operator $\hat{A}$. The kernel of the product of two operators is the convolution of their kernels:

$$(\hat{A}\hat{B})(a^*, c) \;=\; \int A(a^*, b) B(b^*, c) e^{-b^* b} \frac{db^* db}{2\pi i} \;. \tag{3.26}$$

Using the identity

$$\int (bb^*)^n e^{-b^* b} \frac{db^* db}{2\pi i} \;=\; n! \;,$$

it is not difficult to check that *(i)* the kernel of the unit operator is $e^{a^* b}$, *(ii)* the kernel of the Hamiltonian is $e^{a^* b} \omega a^* b$, and, in general, the kernel $A(a^*, b)$ of any operator equals $e^{a^* b}$ times its *normal symbol*, i.e. the function obtained from the normal-ordered quantum operator by substituting $\hat{a}^\dagger \to a^*$, $\hat{a} \to b$. The holomorphic kernel of the evolution operator for the oscillator is

$$U(a^*, b; t) = \exp\{a^* b e^{-i\omega t}\} \;. \tag{3.27}$$

Indeed, using the definitions (3.23, 3.25), it is easy to check that

$$\left[\hat{U}(t)\psi_k\right](a^*) = \psi_k(a^*) e^{-i\mathcal{E}_k t} \;,$$

as one expects.

The holomorphic representation is particularly transparent and simple for the harmonic oscillator (a simple problem in any approach), but it can be used for *any* dynamical system. For a general system, the finite

time evolution operator is complicated. But the kernel of the infinitesimal evolution operator is simple:

$$U(a^*, b; t) = \exp\{a^* b - i\mathcal{H}(a^*, b)\Delta t\} \ , \tag{3.28}$$

where $\mathcal{H}(a^*, b)$ is the normal symbol of the Hamiltonian. As before, the kernel of the finite time evolution operator is the convolution of many infinitesimal kernels:

$$\begin{aligned} &U(a^*, b; t_f - t_i) = \\ &\lim_{n\to\infty} \int e^{-c_n^* c_n} \frac{dc_n^* dc_n}{2\pi i} \cdots \int e^{-c_1^* c_1} \frac{dc_1^* dc_1}{2\pi i} U(a^*, c_n; \Delta t) \cdots U(c_1^*, b; \Delta t) \\ &\quad = \lim_{n\to\infty} \int \exp\{[a^* c_n - c_n^* c_n + \ldots - c_1^* c_1 + c_1^* b] \\ &\qquad - i[\mathcal{H}(a^*, c_n) + \ldots + \mathcal{H}(c_1^*, b)]\Delta t\} \prod_{s=1}^{n} \frac{dc_s^* dc_s}{2\pi i} \ . \end{aligned} \tag{3.29}$$

Adding and subtracting $b^* b$ in the exponent, this can be formally written as

$$\begin{aligned} &U(a^*, b; t_f - t_i) = \\ &\int \exp\{c^*(t_i) c(t_i)\} \exp\left\{\int_{t_i}^{t_f} dt[\dot{c}^* c - i\mathcal{H}(c^*, c)]\right\} \prod_t \frac{dc^*(t) dc(t)}{2\pi i}, \end{aligned} \tag{3.30}$$

with the boundary conditions $c^*(t_f) = a^*$, $c(t_i) = b$. Note that $c(t_f)$ and $c^*(t_i)$ are *not* fixed. They are completely arbitrary variables one should integrate over.

The spectral representation for the evolution operator (3.30) is quite analogous to (3.5):

$$U(a^*, b; t_f - t_i) \ = \ \sum_k \psi_k(a^*)[\psi_k(b^*)]^* e^{-i\mathcal{E}_k(t_f - t_i)} \ . \tag{3.31}$$

3.4 Grassmann Dynamic Variables.

Consider the quantum mechanical problem of a *fermionic oscillator*. The spectrum of the system includes just two states: the ground (or vacuum) state $|0\rangle$ and the excited state $|1\rangle = \hat{a}^\dagger|0\rangle$, where $\hat{a}^\dagger$ is the fermion creation operator. The fermionic creation and annihilation operators $\hat{a}^\dagger$ and $\hat{a}$ satisfy

time evolution operator is complicated. But the kernel of the infinitesimal evolution operator is simple:

$$U(a^*, b; t) = \exp\{a^* b - i\mathcal{H}(a^*, b)\Delta t\} \ , \tag{3.28}$$

where $\mathcal{H}(a^*, b)$ is the normal symbol of the Hamiltonian. As before, the kernel of the finite time evolution operator is the convolution of many infinitesimal kernels:

$$\begin{aligned} &U(a^*, b; t_f - t_i) = \\ &\lim_{n\to\infty} \int e^{-c_n^* c_n} \frac{dc_n^* dc_n}{2\pi i} \cdots \int e^{-c_1^* c_1} \frac{dc_1^* dc_1}{2\pi i} U(a^*, c_n; \Delta t) \cdots U(c_1^*, b; \Delta t) \\ &\quad = \lim_{n\to\infty} \int \exp\{[a^* c_n - c_n^* c_n + \ldots - c_1^* c_1 + c_1^* b] \\ &\qquad - i[\mathcal{H}(a^*, c_n) + \ldots + \mathcal{H}(c_1^*, b)]\Delta t\} \prod_{s=1}^{n} \frac{dc_s^* dc_s}{2\pi i} \ . \end{aligned} \tag{3.29}$$

Adding and subtracting $b^* b$ in the exponent, this can be formally written as

$$\begin{aligned} &U(a^*, b; t_f - t_i) = \\ &\int \exp\{c^*(t_i) c(t_i)\} \exp\left\{\int_{t_i}^{t_f} dt[\dot{c}^* c - i\mathcal{H}(c^*, c)]\right\} \prod_t \frac{dc^*(t) dc(t)}{2\pi i}, \end{aligned} \tag{3.30}$$

with the boundary conditions $c^*(t_f) = a^*$, $c(t_i) = b$. Note that $c(t_f)$ and $c^*(t_i)$ are *not* fixed. They are completely arbitrary variables one should integrate over.

The spectral representation for the evolution operator (3.30) is quite analogous to (3.5):

$$U(a^*, b; t_f - t_i) \ = \ \sum_k \psi_k(a^*)[\psi_k(b^*)]^* e^{-i\mathcal{E}_k(t_f - t_i)} \ . \tag{3.31}$$

3.4 Grassmann Dynamic Variables.

Consider the quantum mechanical problem of a *fermionic oscillator*. The spectrum of the system includes just two states: the ground (or vacuum) state $|0\rangle$ and the excited state $|1\rangle = \hat{a}^\dagger|0\rangle$, where $\hat{a}^\dagger$ is the fermion creation operator. The fermionic creation and annihilation operators $\hat{a}^\dagger$ and $\hat{a}$ satisfy

the algebra

$$\hat{a}^\dagger \hat{a} + \hat{a}\hat{a}^\dagger = 1, \quad \hat{a}^{\dagger 2} = \hat{a}^2 = 0 \, . \tag{3.32}$$

The Hamiltonian of the system is $\hat{H} = \omega \hat{a}^\dagger \hat{a}$ (with positive ω). Then the energy of the ground state is zero and the energy of the excited state is ω.

A conventional realization of the algebra (3.32) is $\hat{a} \equiv \sigma_-$, $\hat{a}^\dagger \equiv \sigma_+$. Then

$$\hat{H} = \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix}, \quad |0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \, .$$

We will use another realization based on the holomorphic representation of the wave functions. It is constructed in full analogy with the holomorphic representation for the bosonic oscillator, with the important difference that the holomorphic variables a^* and a are not ordinary numbers as in the bosonic case, but *Grassmann* anticommuting numbers: $aa^* + a^*a = a^2 = a^{*2} = 0$.

The use of Grassmann variables is the only known consistent way to define a quantum theory of interacting fermionic fields beyond the perturbative expansion. This method was developed by Berezin and described in details in his book [14]. The main definitions are the following:

- Let $\{a_i\}$ be a set of n basic anticommuting variables: $a_i a_j + a_j a_i = 0$. The elements of the Grassmann algebra can be written as the functions $f(a_i) = c_0 + c_i a_i + c_{ij} a_i a_j + \ldots$, where the coefficents $c_0, c_i, \ldots$ are ordinary (real or complex) numbers. The series terminates at the n-th term: the $(n+1)$-th term of this series would involve a product of two identical anticommuting variables (say, a_1^2), which is zero. Note that even though a Grassmann number like $a_1 a_2$ commutes with any other number, it still cannot be treated as a normal number but instead represents "an even element of the Grassmann algebra". There are, of course, also odd anticommuting elements.
- One can add functions $(f(a_i) + g(a_i) = c_0 + d_0 + (c_i + d_i)a_i + \ldots)$ as well as multiply them. For example, $(1 + a_1 + a_2)(1 + a_1 - a_2) = 1 + 2a_1 - 2a_1 a_2$ (the anticommutation property of a_i was used).
- One can also differentiate the functions $f(a_i)$ with respect to Grassmann variables: $d/da_i \; 1 \stackrel{\text{def}}{=} 0$, $d/da_i \; a_j \stackrel{\text{def}}{=} \delta_{ij}$, and the derivative of a product of two functions satisfies the Leibnitz rule, except that

the operator d/da_i should be thought of as a Grassmann variable, which sometimes leads to the sign change when d/da_i is pulled in to annihilate a_i in the product. For example, $d/da_1(a_2a_3a_1) = a_2a_3$, but $d/da_1(a_2a_1a_3) = -a_2a_3$.

- Finally, one can integrate over Grassmann variables. In contrast to the usual bosonic case, the integral cannot be obtained here as a limit of integral sums, cannot be calculated numerically through "finite limits" (this makes no sense for Grassmann numbers) with the Simpson method, etc. What one can do, however, is to integrate over a Grassmann variable in the "whole range" ("from $-\infty$ to ∞", if you wish, though this is, again, meaningless). The definition is $\int da_j f(a_i) \stackrel{\text{def}}{=} d/da_j f(a_i)$.
- If the Grassmann algebra involves an even number of generators $\{a_i\}$, one can introduce an operation of involution, which we will associate with the complex conjugation. In our fermionic oscillator example, we have a pair of variables a^*, a, which are exchanged under complex conjugation. We will further assume that ordinary numbers c_0, etc. are replaced by their complex conjugates under such transformations:

$$f(a) = c_0 + c_i a_i + d_i a_i^* + \ldots \longrightarrow f^*(a) = \bar{c}_0 + \bar{c}_i a_i^* + \bar{d}_i a_i + \ldots$$

(even though complex conjugation of ordinary numbers and involution of Grassmann numbers could be defined independently). It is convenient to assume $[f(a_i)g(a_i)]^* = g^*(a_i)f^*(a_i)$ like for Hermitian conjugation. Then the product a^*a is real.

Let us return to the fermionic oscillator. The classical Hamiltonian is $\omega a^* a$. Like earlier, the classical equations of motion are $i\dot{a}^* = -\omega a^*$ with the solution $a^*(t) \propto e^{i\omega t}$ (needless to say, in this case we cannot plot the solution on a graph or numerically solve the equations of motion).

The quantum operators $\hat{a}^\dagger$ and $\hat{a}$ are represented by $\hat{a}^\dagger \equiv a^*$ and $\hat{a} \equiv d/da^*$. The wave functions of the system are holomorphic functions $\psi(a^*) = c_0 + c_1 a^*$. The functions "1" and "a^*" are the eigenstates $|0\rangle$, $|1\rangle$ of the quantum Hamiltonian. One can check that these states are orthonormal if the scalar product on our Hilbert space is defined in the following way:

$$(f, g) \stackrel{\text{def}}{=} \int [f(a^*)]^* g(a^*) e^{-a^* a} da^* da \ . \tag{3.33}$$

The analogy with the bosonic case is pretty close. We can again define the holomorphic kernels of the operators such that

$$(\hat{A}f)(a^*) \;=\; \int A(a^*,b)f(b^*)e^{-b^*b}db^*db\;, \tag{3.34}$$

and calculate the kernel of the finite time evolution operator. For a single pair of fermionic variables, the Hamiltonian is bound to have the oscillator form $\omega a^*a \,+\, E_0$ (all higher products are zero, so that unharmonic interactions are not allowed). As in the bosonic case, the finite time evolution operator has the form (3.27). If the system involves several pairs of fermionic variables a_r^*, a_r, unharmonic terms in the Hamiltonian are allowed. Making the same transformations as earlier, the finite time evolution operator can be expressed as a convolution of a large number of infinitesimal evolution operators $U(\{a_r^*\},\{b_r\};\Delta t) = \exp\{\sum_r a_r^*b_r - i\mathcal{H}(\{a_r^*\},\{b_r\})\Delta t\}$, giving rise to the path integral

$$\begin{aligned} U(\{a_r^*\},\{b_r\};t_f-t_i) \;&=\; \\ \int \exp\left\{\sum_r c_r^*(t_i)c_r(t_i)\right\} &\exp\left\{\int_{t_i}^{t_f} dt\left[\sum_r \dot{c}_r^*c_r - i\mathcal{H}(c_r^*,c_r)\right]\right\} \\ &\prod_{rt} dc_r^*(t)dc_r(t) \end{aligned} \tag{3.35}$$

with the boundary conditions $c_r(t_i) = b_r,\;\; c_r^*(t_f) = a_r^*$. As in the bosonic case, $c_r^*(t_i)$ and $c_r(t_f)$ are not fixed. Also, at intermediate times $t_i < t < t_f$, $c_r(t)$ and $c_r^*(t)$ are *independent* Grassmann variables to be integrated over.

Let us now consider the Euclidean evolution operator. Its spectral representation is

$$U(a^*,b;-i\beta) \;=\; \sum_k \psi_k(a^*)[\psi_k(b^*)]^* e^{-\beta\mathcal{E}_k} \;=\; 1 + a^*be^{-\beta\omega} \tag{3.36}$$

[confer Eq. (3.31); for simplicity, we restricted ourselves to just one fermionic degree of freedom]. Let us try to construct the partition function of the system $Z = 1 + e^{-\beta\omega}$ in the same way as we did it for the bosonic case. A first guess is

$$Z \stackrel{?}{=} \int U(a^*,a;-i\beta)e^{-a^*a}da^*da\;. \tag{3.37}$$

However, the explicit calculation of the integral gives $1 - e^{-\beta\omega}$, which is not what we want. The minus sign in the second term appeared due to the

anticommutative nature of the Grassmann numbers. For a general system with several fermionic degrees of freedom, the integral generalizing (3.37) gives

$$\sum_n (-1)^F e^{-\beta \mathcal{E}_n} \ , \tag{3.38}$$

where F if the fermion number of the eigenstate $|n\rangle$, i.e. the number of Grassmann factors in the holomorphic wave function of the eigenstate $|n\rangle$.[‡] The true answer for the standard partition function of a fermionic system is found to be

$$Z = \int U(\{a_r^*\}, \{-a_r\}; -i\beta) e^{-\sum_r a_r^* a_r} \prod_r da_r^* da_r \ . \tag{3.39}$$

It can be written as a path integral

$$\begin{aligned} Z \ = \ & \int \exp\left\{ \sum_r c_r^*(0) c_r(0) \right\} \prod_{r\tau} dc_r^*(\tau) dc_r(\tau) \\ & \exp\left\{ \int_0^\beta d\tau \left[\sum_r \dot{c}_r^* c_r - \mathcal{H}(c_r^*, c_r) \right] \right\} \ , \end{aligned} \tag{3.40}$$

where the trajectories $c_r(\tau)$, $c_r^*(\tau)$ are antiperiodic in a sense that $c_r^*(\beta) = -[c_r(0)]^*$.

Let us briefly comment on the practical way to calculate fermionic path integrals. In theories like QCD the action is quadratic in the fermionic fields and the fermionic path integral is always Gaussian. The integral can be done using the very important identity (prove it!)

$$\int \prod_r da_r^* da_r \exp\{-A_{ij} a_i^* a_j\} \ = \ \det \|A\| \ . \tag{3.41}$$

Compare this to

$$\int \prod_r \frac{da_r^* da_r}{2\pi i} \exp\{-A_{ij} a_i^* a_j\} \ = \ \frac{1}{\det \|A\|} \tag{3.42}$$

for the bosonic case.

[‡]The sum (3.38) has a name: *Witten index*. It plays an important role for supersymmetric theories, but we do not need it.

Lecture 4

Quantization of Gauge Theories

4.1 Dirac Quantization Procedure.

The technique we have developed so far allows one to perform the quantization of an ordinary (non-gauge) field theory like $\lambda\phi^4$ or Yukawa theory. What we need, however, is to learn how to quantize gauge theory. This problem has a number of important subtleties, and some additional methods need to be elaborated.

We will briefly touch on some general aspects of the theory of gauge (or, in another language, *constrained*) systems and will test general methods in the simple example of a quantum mechanical gauge system where the physics and the results are most transparent. We finally apply these methods to the Yang–Mills system of interest.

We start with the classical theory. Suppose, we have a Hamiltonian $\mathcal{H}(p_i, q_i)$ depending on the phase space variables (p_i, q_i), $i = 1, \ldots, n$, and subject to a number of constraints $\phi^a(p_i, q_i) = 0$, $a = 1, \ldots, k < n$. Thus, we are searching for the classical trajectories satisfying the following system of equations:

$$\begin{aligned} \dot{p}_i + \frac{\delta\mathcal{H}}{\delta q_i} &= 0 \ , \\ \dot{q}_i - \frac{\delta\mathcal{H}}{\delta p_i} &= 0 \ , \\ \phi^a(p_i, q_i) &= 0 \ . \end{aligned} \tag{4.1}$$

Here the first two equations are the dynamical equations of motion and the last one expresses the constraints. The equations (4.1) are compatible,

provided that, for any trajectory with initial conditions satisfying $\phi^a = 0$, this holds also at later times. This implies $\dot{\phi}^a = -\{\phi^a, \mathcal{H}\}_{\text{P.B.}} = 0$, which is true if

$$\{\phi^a, \mathcal{H}\}_{\text{P.B.}} = \sum_b C_{ab}\phi^b \ . \tag{4.2}$$

We want to quantize the theory in both the canonical and in the path integral formalism. For the latter, the action functional needs to be defined. A sufficient condition for the latter to exist is that the ϕ^a satisfy also the property

$$\{\phi^a, \phi^b\}_{\text{P.B.}} = \sum_c C_{abc}\phi^c \ . \tag{4.3}$$

Indeed, the constrained action is defined as

$$S = \int \left[\sum_i p_i \dot{q}_i - \mathcal{H}(p_i, q_i) - \sum_a \lambda^a \phi^a(p_i, q_i) \right] dt \ . \tag{4.4}$$

The action (4.4) depends on extra variables λ^a (the Lagrange multipliers) but not on their time derivatives. Variations of the functional S with respect to $\lambda^a(t)$ reproduce the constraints. Variations of S with respect to $p_i(t)$ and $q_i(t)$ give the dynamical equations

$$\begin{aligned} \dot{p}_i + \frac{\delta \mathcal{H}}{\delta q_i} + \lambda^a \frac{\delta \phi^a}{\delta q_i} = 0 \ , \\ \dot{q}_i - \frac{\delta \mathcal{H}}{\delta p_i} - \lambda^a \frac{\delta \phi^a}{\delta p_i} = 0 \ . \end{aligned} \tag{4.5}$$

This system is compatible with the constraints provided

$$\dot{\phi}^a = (\delta\phi^a/\delta p_i)\dot{p}_i + (\delta\phi^a/\delta q_i)\dot{q}_i = -\{\phi^a, \mathcal{H}\}_{\text{P.B.}} - \sum_b \lambda^b \{\phi^a, \phi^b\}_{\text{P.B.}} = 0 \ .$$

If we want this to be true for any λ^a, the conditions (4.2, 4.3) follow. Expressing the momenta $p_i(t)$ via $q_i(t)$ and $\dot{q}_i(t)$ by use of the second equation in Eq. (4.5) and substituting it into the action (4.4), we arrive at the standard Lagrange formulation.

A system with constraints satisfying the criteria (4.2, 4.3) is called the Hamiltonian system with constraints of the first kind. As was shown before, the Yang–Mills system with the Hamiltonian (1.25) and the constraints

(1.24) enjoys the properties (1.28) and $\{G^a, \mathcal{H}\}_{\text{P.B.}} = 0$ and hence belongs to this class.*

Note that the Lagrange multipliers $\lambda^a(t)$ entering the equations (4.5) are completely arbitrary: for any $\lambda^a(t)$ the solutions for $p_i(t)$ and $q_i(t)$ can be found. The freedom of choice for the $\lambda^a(t)$ is precisely a gauge invariance! A distinguished choice is $\lambda^a = 0$ where the equations of motion (4.5) coincide with original Hamilton's equations (4.1).

It is easier to understand the general theory in a concrete example, which we choose to be very simple (may be, the simplest possible one). Consider a two-dimensional oscillator with the Hamiltonian

$$\mathcal{H} = \frac{1}{2}(p_x^2 + p_y^2) + \frac{\omega^2}{2}(x^2 + y^2) \ . \tag{4.6}$$

Impose the constraint

$$p_\phi = xp_y - yp_x \equiv \epsilon_{ij} x_i p_j = 0 \tag{4.7}$$

$[x_i \equiv (x, y),\ p_i \equiv (p_x, p_y)]$. Obviously, $\{p_\phi, \mathcal{H}\}_{\text{P.B.}} = 0$ and we have a constrained system of the first kind. The canonical Lagrangian is

$$\mathcal{L} = p_i \dot{x}_i - \mathcal{H} - \lambda p_\phi \ . \tag{4.8}$$

The equations of motion (4.5) in phase space are

$$\begin{aligned} \dot{p}_i + \omega^2 x_i + \lambda \epsilon_{ij} p_j &= 0 \ , \\ \dot{x}_i - p_i + \lambda \epsilon_{ij} x_j &= 0 \ . \end{aligned} \tag{4.9}$$

Solving for the momenta from the lower pair of equations in Eqs. (4.9), and substituting them in Eq. (4.8), we obtain the canonical Lagrangian

$$\mathcal{L} = \frac{1}{2}(\dot{x}_i + \lambda \epsilon_{ij} x_j)^2 - \frac{\omega^2}{2} x_i^2 \ . \tag{4.10}$$

We see that different $\lambda(t)$ physically correspond to different choices for the reference frame, which may rotate with a time-dependent angular velocity $\lambda(t)$. The freedom of choice of $\lambda(t)$ is the gauge invariance. Explicitly, the Lagrangian (4.10) is invariant under the transformations

$$\begin{cases} x_i' = O_{ij}(\chi) x_j \\ \lambda' = \lambda + \dot{\chi} \end{cases} \ , \tag{4.11}$$

*This also follows from the fact that the Yang–Mills system has a Lagrangian formulation which was our starting point. It *implies* that the conditions (4.2, 4.3) are satisfied.

with

$$O_{ij} = \begin{pmatrix} \cos\chi, & -\sin\chi \\ \sin\chi, & \cos\chi \end{pmatrix} .$$

The best way to handle our system is, of course, to realize that the constraint $p_\phi = 0$ physically means that the system is not allowed to move in the angular direction. Then we simply can go to polar coordinates, set $\phi = 0$ and obtain the equations of motion $\dot{p}_r + \omega^2 r = 0$, $\dot{r} - p_r = 0$ or $\ddot{r} + \omega^2 r = 0$ describing radial motion of the system. Imposing the additional constraint $\phi = 0$ on top of the initial constraint $p_\phi = 0$ is called *fixing the gauge*. The constraint $\phi = 0$ picks out one representative from each *gauge orbit* defined as a set of points in the configuration space $\{q_i\}$ related to each other by a time-independent gauge transformation. In our case, the gauge orbits are just the circles of constant r.

In the general case, fixing the gauge amounts to imposing, on top of the constraints $\phi^a = 0$, also additional constraints $\chi^a = 0$ [the functions $\chi^a(p_i, q_i)$ are not quite arbitrary and should satisfy some natural requirements, which we will not discuss here]. After explicitly resolving, whenever it can be done practically, the equation system $\phi^a = 0$, $\chi^a = 0$ in terms of a restricted set of *gauge invariant* phase space variables p_i^* and q_i^*, with $i = 1, \ldots, n-k$ (p_r and r in our case), the Hamiltonian $\mathcal{H}(p^*, q^*)$ describes an ordinary unconstrained system, which may be handled classically and, if needed, quantized in the usual way. Note that *two* phase space degrees of freedom are eliminated for each constraint $\phi^a = 0$ originally imposed. In our case, both p_ϕ and ϕ are eliminated.

The analogy with gauge field theories is now seen. Consider an Abelian theory without matter fields (pure photodynamics). The dictionary of correspondences is

$$\lambda \equiv A_0/2 \ , \quad p_\phi \equiv \partial_i E_i \ , \quad \phi \equiv \partial_i A_i \ .$$

Imposing the constraint $\partial_i A_i = 0$ amounts to going over to Coulomb gauge. The Hamiltonian $\mathcal{H}_{\text{photons}} = 1/2 \int dx (E_i^2 + B_i^2)$ depends then only on the two transverse field polarizations $\boldsymbol{A}_\perp$ and their canonical momenta $\boldsymbol{E}_\perp$.

However, it is not always practically possible to explicitly solve the constraints. It is difficult and practically inconvenient for QED, which includes apart from photons also charged particles. A complication is that the constraints $\partial_i E_i = \rho_{\text{matter}}$, $\partial_i A_i = 0$ dictate in this case the appearance of a nonvanishing $A_0 = -\Delta^{-1}\rho$, and the Hamiltonian with resolved constraints

involves a nonlocal term $\propto \rho\Delta^{-1}\rho$. For non-Abelian Yang–Mills theory, it is hopeless. Coulomb gauge does not really work here due to the presence of the so-called *Gribov copies* to be discussed later. We are forced to learn how to construct and handle quantum gauge theory without explicitly resolving constraints at the classical level.

There are two ways to quantize: *(i)* via the Schrödinger equation and *(ii)* via the path integral. Let us first describe the first way, i.e. the *Dirac quantization procedure.* We are going to solve the equation system

$$\begin{aligned} \mathcal{H}(\hat{p}_i, q_i)\Psi_k(q_i) &= \mathcal{E}_k\Psi_k(q_i) \ , \\ \phi^a(\hat{p}_i, q_i)\Psi_k(q_i) &= 0 \ , \end{aligned} \tag{4.12}$$

where $\hat{p}_i = -i\partial/\partial q_i$, and $\mathcal{H}, \phi^a$ are the same functions as before (we disregard here possible ordering ambiguities as this problem does not arise for the Yang–Mills Hamiltonian and constraints). The conditions (4.2, 4.3) are promoted to

$$[\hat{\phi}^a, \hat{\mathcal{H}}] = \sum_b C_{ab}\hat{\phi}^b, \quad [\hat{\phi}^a, \ \hat{\phi}^b] = \sum_c C_{abc}\hat{\phi}^c \ . \tag{4.13}$$

The conditions (4.13) render the system (4.12) self-consistent, i.e. the Hilbert space of definite energy eigenstates and with definite (zero!) eigenvalues for all operators $\hat{\phi}^a$ is not empty.† The procedure (4.12) *defines* our quantum gauge system.

In the toy model (4.6, 4.7), the situation is very simple and transparent. The spectrum of the unconstrained system is $\mathcal{E}_{n_r,m} = \omega(1 + |m| + 2n_r)$, where m is the eigenvalue of the momentum operator $\hat{p}_\phi$ and n_r is the radial quantum number. The Hilbert space (4.12) involves only the states with zero angular momentum. The spectrum is

$$\mathcal{E}_k = \omega(1 + 2n_r) \ . \tag{4.14}$$

It coincides, up to an irrelevant overall energy shift, with the spectrum of the Hamiltonian of a one-dimensional oscillator describing radial motion on the half-line $0 \leq r < \infty$. Note that the wave functions of our system do

†These conditions are sufficient but not really necessary; there are some other self-consistent gauge systems with weaker restrictions: the constrained systems of the second kind, etc. We are not going to discuss these complications here as for QCD the conditions (4.13) are duly fulfilled.

not depend on the gauge variable ϕ. In other words, the wave functions are invariant under the gauge transformation (4.11) (with constant χ).

The Hilbert space of the quantum Yang–Mills field system is defined through

$$\begin{aligned} \mathcal{H}[\hat{E}_i^a(\boldsymbol{x}), A_i^a(\boldsymbol{x})]\Psi_k[A_i^a(\boldsymbol{x})] &= \mathcal{E}_k\Psi_k[A_i^a(\boldsymbol{x})] \ , \\ G^c[\hat{E}_i^a(\boldsymbol{x}), A_i^a(\boldsymbol{x})]\Psi_k[A_i^a(\boldsymbol{x})] &= 0 \ , \end{aligned} \tag{4.15}$$

where $\hat{E}_i^a = -ig^2\delta/\delta A_i^a(\boldsymbol{x})$. The system (4.15) has non-trivial solutions due to the commutation relations

$$[\hat{G}^a, \hat{\mathcal{H}}] = 0 \ , \quad [\hat{G}^a(\boldsymbol{x}), \hat{G}^b(\boldsymbol{y})] = if^{abc}\hat{G}^c(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y}) \ . \tag{4.16}$$

Just like in our toy quantum mechanical model, the constraint $\hat{G}^a\Psi = 0$ requires for the wave functions to be invariant under the gauge transformation (1.9) of the dynamical variables A_i^a, with time-independent $\Omega(\boldsymbol{x})$.

We will see a bit later that, in the Yang–Mills case, the Hilbert space defined in (4.15) is actually too large. One needs, on top of the constraint $G^a\Psi = 0$, to impose an *additional* constraint called *superselection rule*, which is closely associated with the instanton Euclidean field configurations. Let us, however, postpone the discussion of this issue until we learn another approach to the quantization of gauge theories, the path integral approach.

Let us concentrate on the path integral representation of the constrained partition function

$$Z_{\text{cons.}}(\beta) = \sum_k P_k e^{-\beta\mathcal{E}_k} \tag{4.17}$$

(P_k is the projector: $P_k = 1$ for gauge-invariant states and $P_k = 0$ for noninvariant ones) and again first discuss our toy oscillator model. The partition function of the unconstrained two-dimensional oscillator can be expressed via its Euclidean evolution operator (3.13) according to Eq. (3.15). Now we want, however, to include in the sum only the gauge (rotationally) invariant states. This can be done if one defines

$$Z_{\text{cons.}}(\beta) = \int_0^{2\pi} \frac{d\chi_0}{2\pi} \int d\boldsymbol{x} \ \mathcal{K}(\boldsymbol{x}^{\chi_0}, \boldsymbol{x}; -i\beta) \ , \tag{4.18}$$

where $\mathcal{K}$ is still the unconstrained evolution operator, but the final and initial points do not coincide anymore. $\boldsymbol{x}^{\chi_0}$ differs from $\boldsymbol{x}$ by rotation by the angle χ_0, which is to be integrated over. Indeed, all states in the sum

(3.13) with non-zero angular momenta give zero after integration over χ_0, and only gauge invariant states survive.

Substitute now into Eq. (4.18) the path integral representation of the evolution operator $\mathcal{K}$. We obtain

$$Z_{\text{cons.}}(\beta) = \int_0^{2\pi} \frac{d\chi_0}{2\pi} \int \prod_\tau d\boldsymbol{x}(\tau)\, e^{-\int_0^\beta d\tau \mathcal{L}^E_{\text{uncons.}}[\boldsymbol{x}(\tau)]} , \tag{4.19}$$

with $\mathcal{L}^E_{\text{uncons.}} = (\dot{\boldsymbol{x}})^2/2 + \omega^2\boldsymbol{x}^2/2$. The inner integral is done over all the trajectories with the "twisted" boundary conditions $x_i(\beta) = O_{ij}(\chi_0)x_j(0)$. Let us now change variables in the integral (4.19):

$$x_i(\tau) = O_{ij}[\chi(\tau)]x'_j(\tau) , \tag{4.20}$$

where the *only* condition imposed on the otherwise arbitrary function $\chi(\tau)$ is

$$\chi(0) = 0, \quad \chi(\beta) = \chi_0 . \tag{4.21}$$

Equation (4.20) corresponds to changing to a reference frame which rotates in Euclidean time with variable angular velocity $\lambda(\tau) = \dot{\chi}(\tau)$. The condition (4.21) requires that the overall rotation angle is fixed. However, as we also have to integrate over χ_0, *nothing* is in fact fixed and $\chi(\tau)$ is quite arbitrary. The variables $x'_i(\tau)$ are now periodic. Let us substitute Eq. (4.20) into the integral (4.19) and upgrade the integration over χ_0 to a path integral over $\prod_\tau d\chi(\tau) \sim \prod_\tau d\lambda(\tau)$. As the integral does not depend on the choice of variables, this only gives an irrelevant (albeit infinite) overall factor. The final result is very natural:

$$Z_{\text{cons.}}(\beta) \sim \int \prod_\tau d\boldsymbol{x}'(\tau) d\lambda(\tau)\, e^{-\int_0^\beta d\tau \mathcal{L}^E[\boldsymbol{x}'(\tau),\lambda(\tau)]} , \tag{4.22}$$

where $\mathcal{L}^E[\boldsymbol{x}'(\tau), \lambda(\tau)]$ is the Lagrangian (4.10) with the sign of the potential reversed. As was mentioned, $\boldsymbol{x}'(\tau)$ is periodic in τ. Periodicity of $\lambda(\tau)$ is not necessary, but, as imposing this condition is convenient and does not change the result, we will assume it. As the Lagrangian (4.10) is invariant under the gauge transformations (4.11), this is also true for the partition function (4.22).

The result (4.22) can also be obtained in a different, somewhat heuristic way. Let us write the Minkowski path integral in the Hamiltonian form,

implementing the constraints by inserting the delta-function $\delta(p_\phi)$ into the integrand:

$$\int \prod_t \frac{d\boldsymbol{p}(t)d\boldsymbol{x}(t)}{2\pi}\delta(p_\phi)\exp\left\{i\int dt[\boldsymbol{p}\cdot\dot{\boldsymbol{x}}-\mathcal{H}(\boldsymbol{p},\boldsymbol{x})]\right\} = \int \prod_t \frac{d\boldsymbol{p}(t)d\boldsymbol{x}(t)}{2\pi}\frac{d\lambda(t)}{2\pi}\exp\left\{i\int dt[\boldsymbol{p}\cdot\dot{\boldsymbol{x}}-\mathcal{H}(\boldsymbol{p},\boldsymbol{x})-\lambda(t)p_\phi(t)]\right\}. \quad (4.23)$$

The integral over momenta is Gaussian and can be done, after which the momenta are replaced by their saddle point values, which are found as the solution of Hamilton's equations of motion [the second equation in Eq. (4.9)]. We arrive at the Minkowskian version of the integral (4.22). The Euclidean partition function (4.22) may be obtained by analytic continuation.

Everything we have done for the oscillator can be applied without essential change to any system with the constraints of the first kind. In particular, in full analogy with Eq. (4.22), the partition function for Yang–Mills theory can be expressed as

$$Z = \int \prod_{\boldsymbol{x},\tau} dA^a_\mu(\boldsymbol{x},\tau)\exp\left\{-\frac{1}{2g^2}\int_0^\beta d\tau\int d\boldsymbol{x}\ \mathrm{Tr}\{F_{\mu\nu}F_{\mu\nu}\}\right\}. \quad (4.24)$$

This expression is gauge invariant. In addition, the $O(4)$ invariance is manifest. The Minkowski path integral is obtained from Eq. (4.24) by inserting the factor i in the exponent:

$$Z_M = \int \prod_x dA^a_\mu(x)\ \exp\left\{-i\frac{1}{2g^2}\int d^4x\ \mathrm{Tr}\{F_{\mu\nu}F^{\mu\nu}\}\right\}. \quad (4.25)$$

Consider now the theory with quarks. It involves also the path integrals over Grassmann field variables. In Minkowski space, we have

$$\int \prod_x d\bar{\psi}_f(x)d\psi_f(x)\ \exp\left\{i\int d^4x\ \bar{\psi}_f(i\not{\mathcal{D}}-m_f)\psi_f\right\} \quad (4.26)$$

for each quark flavor. In Euclidean space, the expression (4.26) turns into

$$\int \prod_x d\bar{\psi}_f(x)d\psi_f(x)\exp\left\{\int d^4x\ \bar{\psi}_f(i\not{\mathcal{D}}^E-m_f)\psi_f\right\}, \quad (4.27)$$

where $\not{\mathcal{D}}^E=\gamma^E_\mu\mathcal{D}_\mu$ is Hermitian with the convention (N.13).

This is the right place to discuss the rather confusing issue of *fermions in Euclidean space.* After the integration over fermionic variables in Eqs.(4.26) and (4.27) is done, we obtain, up to an irrelevant factor i^{∞}, $\det\| -i\not{p}+m_f\|$ for the integral in Eq. (4.26) and $\det\| -i\not{p}^E + m_f\|$ for the integral in Eq. (4.27). Thus, if we are interested in the partition function only, we can use the following rule of analytic continuation

$$\det\| -i\not{p} + m_f\| \longrightarrow \det\| -i\not{p}^E + m_f\| \tag{4.28}$$

and forget about fermionic variables alltogether. The recipe (4.28) alone is not sufficient, however, if we want to calculate also certain Green's functions involving fermionic fields. To calculate Euclidean Green's functions, Euclidean fermionic fields need to be properly defined.

Our main message is that the symbols $\bar{\psi}_f$ in Eq. (4.26) and Eq. (4.27) have different meaning and, if you will, a different status. In Minkowski space, there exists an alternative operator approach to quantization. The Heisenberg field operators $\hat{\bar{\psi}}$ and $\hat{\psi}$ are related to each other by Dirac conjugation: $\hat{\bar{\psi}} = \hat{\psi}^{\dagger}\gamma^0$. Speaking of the path integral (4.26), ψ and $\bar{\psi}$ are independent Grassmann variables. It is convenient, however, to assume that they are also related to each other by Dirac conjugation. Then the Lagrangian in Eq. (4.26) is a real Lorentz scalar. Indeed, $\delta^{\text{Lorentz}}\psi = \omega_{\mu\nu}\sigma_{\mu\nu}\psi$ $[\sigma_{\mu\nu} = \frac{1}{2}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)]$ and hence $\delta^{\text{Lorentz}}\bar{\psi} = \omega_{\mu\nu}\bar{\psi}\gamma^0\sigma^{+}_{\mu\nu}\gamma^0 = -\omega_{\mu\nu}\bar{\psi}\sigma_{\mu\nu}$. It follows then that $\bar{\psi}\psi$ is a (real) Lorentz scalar and $\bar{\psi}\gamma_\mu\psi$ is a (real) Lorentz vector.

In Euclidean space the situation is different. First, in the operator language, the very notion of Hermiticity ceases to be useful: the Euclidean Heisenberg operators

$$\hat{A}(\tau) \;=\; \exp\{\hat{H}\tau\}\hat{A}(0)\exp\{-\hat{H}\tau\} \tag{4.29}$$

cannot be rendered Hermitian for all τ. Second, in the path intergal formalism, the transformation law of the fermionic variables under $SO(4)$ rotations is $\delta^{SO(4)}\psi = \omega_{\mu\nu}\sigma^E_{\mu\nu}\psi$ with $\left(\sigma^E_{\mu\nu}\right)^{\dagger} = -\sigma^E_{\mu\nu}$ and, for $\bar{\psi}\psi$ to be an $SO(4)$ scalar, $\bar{\psi}$ should transform as a hermitially conjugated spinor $\psi^{\dagger}$ (without the extra γ^0 factor). We could introduce the notation $\psi^{\dagger}$ instead of $\bar{\psi}$ in Eq. (4.27), but prefer not to do so and not to assume that $\bar{\psi}$ is hermitially conjugated to ψ.[‡] The proper way to think of the relationship

[‡]We will see in Lecture 13 that, to define a viable lattice approximation to the Euclidean fermionic path integral, keeping ψ and $\bar{\psi}$ independent is not only an option, but a

between $\bar{\psi}(x)$ and $\psi(x)$ is to write

$$\begin{cases} \psi(x) = \sum_k c_k u_k(x) \\ \bar{\psi}(x) = \sum_k \bar{c}_k u_k^\dagger(x) \end{cases} , \tag{4.30}$$

where $\{u_k(x)\}$ is a complete basis in the corresponding Hilbert space. It is conveniently choosen as a set of eigenfunctions of the Dirac operator: $\not{D}^E u_k(x) = \lambda_k u_k(x)$; c_k, $\bar{c}_k$ are independent Grassmann integration variables.

In contrast to the Minkowskian fermionic action, the Euclidean action in the exponent of Eq. (4.27) cannot be rendered real even if the variables $\bar{\psi}$ and ψ are assumed to be conjugate to each other; the term $i\bar{\psi}_f \not{D}^E \psi_f$ changes sign under complex conjugation while the mass term does not. But this actually does not cause a trouble, because the full fermionic path integral

$$\det \| - i\not{D}^E + m \| = \prod_k (-i\lambda_k + m) \tag{4.31}$$

(the product is performed over all eigenmodes of the massless Dirac operator in a given gauge field background) is real! To see this, let us assume that the theory is somehow regularized both in the infrared (so that the spectrum of the Dirac operator is discrete) and in the ultraviolet [so that large eigenvalues λ_k do not contribute to the product (4.31)].[§] Note now that the spectrum of the massless Dirac operator enjoys the following symmetry: for any eigenfunction u_k of the operator $\not{D}^E$ with eigenvalue λ_k, the function $u'_k = \gamma^5 u_k$ is also an eigenfunction with the eigenvalue $-\lambda_k$ ($\not{D}^E$ and γ^5 anticommute). Therefore,

$$\prod_k (-i\lambda_k + m) = m^q \prod_k{}' (\lambda_k^2 + m^2) , \tag{4.32}$$

where q is the number of possible exact zero modes [we will show in Lecture 12 that q coincides with the absolute value of the topological charge (2.6) of the gauge field] and the product $\prod'_k$ is performed over non-zero eigenvalues only. We see that the expression (4.32) is real and the Euclidean partition function is real, too.

necessity.

[§]There are many such regularization schemes and we will begin the discussion of this very important issue in the next section.

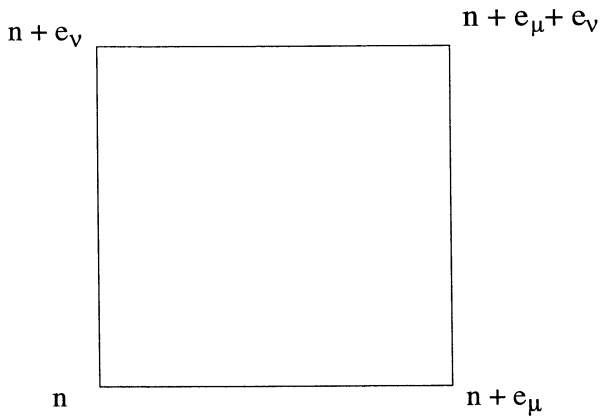


Fig. 4.1 An elementary plaquette.

4.2 Path Integral on the Lattice.

The path integrals (4.24–4.27) are symbols which require an operational definition. For ordinary quantum mechanical systems, we divided a finite time interval into $n+1$ small intervals and defined the path integral as the limit of a finite dimensional integral (3.11), with n tending to ∞. For field theories, we need to discretize not only time, but also space. Furthermore, for gauge theories, we want to do this in a way that preserves gauge invariance at intermediate steps. The corresponding gauge invariant discretization procedure for Yang–Mills theory was worked out by Wilson. We will discuss it for the Euclidean version of the theory only. A discretized gauge invariant path integral can also be defined in Minkowski spacetime, but there is not much use of it: to date, only Euclidean path integrals are feasible for numerical calculations.

Let us introduce a four-dimensional hyper-cubic lattice. The nodes of the lattice are labelled by an integer 4-vector n. Let us define on each *link* of the lattice a unitary matrix $U_{n+e_\mu,n} \in SU(N)$, where $e_1 = (1,0,0,0)$, etc. For each *plaquette* (or two-dimensional face) of the lattice labelled by its corner n_α with all the components less or equal than the corresponding components for other three corners, and by two directional vectors e_μ, e_ν (see Fig. 4.1), we define

$$W_{n,\mu\nu} = \frac{1}{N}\mathrm{Re\ Tr}\left\{U_{n,n+e_\nu}U_{n+e_\nu,n+e_\mu+e_\nu}U_{n+e_\mu+e_\nu,n+e_\mu}U_{n+e_\mu,n}\right\}, \quad (4.33)$$

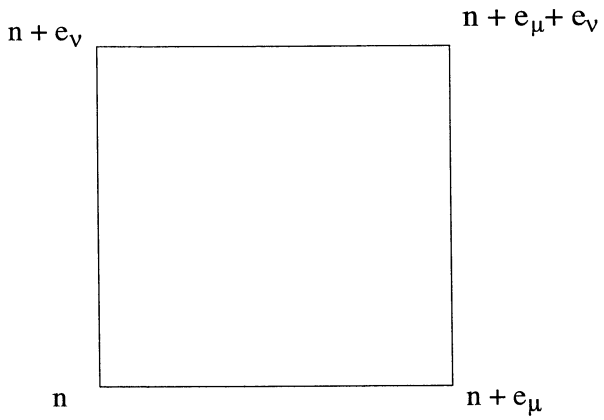


Fig. 4.1 An elementary plaquette.

4.2 Path Integral on the Lattice.

The path integrals (4.24–4.27) are symbols which require an operational definition. For ordinary quantum mechanical systems, we divided a finite time interval into $n+1$ small intervals and defined the path integral as the limit of a finite dimensional integral (3.11), with n tending to ∞. For field theories, we need to discretize not only time, but also space. Furthermore, for gauge theories, we want to do this in a way that preserves gauge invariance at intermediate steps. The corresponding gauge invariant discretization procedure for Yang–Mills theory was worked out by Wilson. We will discuss it for the Euclidean version of the theory only. A discretized gauge invariant path integral can also be defined in Minkowski spacetime, but there is not much use of it: to date, only Euclidean path integrals are feasible for numerical calculations.

Let us introduce a four-dimensional hyper-cubic lattice. The nodes of the lattice are labelled by an integer 4-vector n. Let us define on each *link* of the lattice a unitary matrix $U_{n+e_\mu,n} \in SU(N)$, where $e_1 = (1,0,0,0)$, etc. For each *plaquette* (or two-dimensional face) of the lattice labelled by its corner n_α with all the components less or equal than the corresponding components for other three corners, and by two directional vectors e_μ, e_ν (see Fig. 4.1), we define

$$W_{n,\mu\nu} = \frac{1}{N}\mathrm{Re\ Tr}\left\{U_{n,n+e_\nu}U_{n+e_\nu,n+e_\mu+e_\nu}U_{n+e_\mu+e_\nu,n+e_\mu}U_{n+e_\mu,n}\right\}, \quad (4.33)$$

with $U_{n,n+e_\nu} \equiv U^\dagger_{n+e_\nu,n}$. For $SU(2)$, we need not take the real part in Eq. (4.33) as the trace of any $SU(2)$ matrix is real. Consider the integral

$$Z = \int \prod_{\text{links}} \mathcal{D}U_{\text{link}} \exp\left\{ -\frac{2N}{g^2} \sum_{\text{plaq}} (1 - W_{\text{plaq}}) \right\} , \tag{4.34}$$

where $\mathcal{D}U$ is the Haar measure on the group. We are going to show that the exponent in Eq. (4.34) is a correct discrete approximation to the Euclidean action of continuum Yang–Mills theory and hence the integral (4.34) is a reasonable discrete approximation of the path integral (4.24).

We assume the actual spacing of our Euclidean lattice a be small compared to all characteristic scales of the theory.¶ Let us associate $U_{n+e_\mu,n}$ with the parallel transporter (1.8) along the corresponding link: $U_{n+e_\mu,n} \equiv \exp\{iaA_\mu e_\mu\}$, where the A_μ are defined at the middle of the link. Then the quantity under the trace in Eq. (4.33) is the operator of the parallel transport along the plaquette and $W_{n,\mu\nu}$ is nothing but a Wilson loop operator, which for small boxes can be expressed as in Eq. (1.16). Thus, we see that, indeed,

$$S_{\text{lat}} \to \frac{2N}{g^2} a^4 \sum_n \frac{1}{2} \sum_{\mu\nu} \frac{1}{2N} \text{Tr}\{F^2_{\mu\nu}\} + O(a^6) \to \int d^4x \, \frac{1}{2g^2} \text{Tr}\{F^2_{\mu\nu}\}$$

in the continuum limit (the identity $\sum_{\text{plaq}} = \sum_n \sum_{\mu>\nu} = \frac{1}{2} \sum_n \sum_{\mu\nu}$ was used). The integral (4.34) is gauge invariant as was desired. Indeed, the action (and also the Haar measure) are invariant under the transformations

$$U_{n+e_\mu,n} \to \Omega_{n+e_\mu} U_{n+e_\mu,n} \Omega^\dagger_n , \tag{4.35}$$

where $\{\Omega_n\}$ is a set of unitary matrices defined in the nodes of the lattice. The definition (4.34) is not the only possible one. Other gauge-invariant lattice actions can be constructed. They all reproduce the structure $\text{Tr}\{F^2_{\mu\nu}\}$ to leading order in a, but differ by terms $\propto a^6$. A very probable (but not rigorously proven) statement is that all lattice path integrals with various such Yang–Mills actions give the *same* physical results in the continuum limit, which means that the continuum theory is well defined. Note that,

¶As the coupling constant g in the Yang–Mills action is dimensionless, we do not understand yet what the characteristic scale is. That will be discussed at length in Lecture 9.

in the theories like QED or $\lambda\phi^4$ theory, continuum limit exists, but corresponds to a *free* theory, with interactions being switched out. This is the known zero charge problem. As we will see and discuss at length later, this problem is *absent* in QCD.

The integral in (4.34) involves a discrete but still infinite number of variables. To make it finite, our lattice should have finite size in both spatial and Euclidean time directions. In practice, it is convenient to implement this by imposing periodic boundary conditions on the matrix link variables:

$$U_{n+L_\alpha e_\alpha + e_\mu, n+L_\alpha e_\alpha} = U_{n+e_\mu, n} \ , \tag{4.36}$$

where the set of integers $L_\alpha = \{L_x, L_y, L_z, L_\tau\}$ characterizes the size of the lattice (the number of nodes in the corresponding direction); no summation over α is assumed. With the conditions (4.36), the theory is effectively defined on a discrete four-dimensional torus.

Toroidal boundary conditions are easier to handle than boundary conditions with rigid walls: finite size effects, which are present in all practical numerical calculations are less prominent for the torus. For these boundary effects not to be important, the physical length of the torus aL_μ should be larger than the characteristic scale of the theory, while a should be kept much smaller. In practical calculations, "much" usually means at most 4–5 times. Calculating the path integral (4.34) numerically on an asymmetric lattice with $L_x = L_y = L_z \gg L_\tau$ in the regime where boundary effects due to finite Euclidean time extension are important, while effects due to a finite spatial size are not, one finds the partition function of the system at finite temperature $T = 1/(aL_\tau)$. The properties of the vacuum wave functional are explored in calculations on large symmetric lattices.

Up to now, we discussed only pure Yang–Mills theory without fermions. QCD involves quarks and, to define the path integral in QCD, we need to handle fermionic fields on the lattice. Let us define to this end Grassmann variables $\bar{\psi}_n$, ψ_n in the nodes of the lattice for each quark flavor (color and Lorentz indices are not displayed). As a first and natural guess, let us write the extra term in the action as follows,

$$S^{\text{ferm.lat.}} = -\frac{ia^3}{2} \sum_{n,\mu} [\bar{\psi}_n U_{n,n+e_\mu} \gamma_\mu^E \psi_{n+e_\mu} - \bar{\psi}_n U_{n,n-e_\mu} \gamma_\mu^E \psi_{n-e_\mu}] + ma^4 \sum_n \bar{\psi}_n \psi_n \ . \tag{4.37}$$

We see that the action (4.37) reproduces the action

$$S^E_{\text{ferm}} = -\int d^4x[i\bar{\psi}\gamma^E_\mu(\partial_\mu - iA_\mu)\psi - m\bar{\psi}\psi] \tag{4.38}$$

in the continuum limit. Indeed, for free fermions

$$-\frac{ia^3}{2}\sum_{n,\mu}\bar{\psi}_n\gamma^E_\mu[\psi_{n+e_\mu} - \psi_{n-e_\mu}] \;\to\; -i\int d^4x\bar{\psi}\gamma^E_\mu\partial_\mu\psi\ .$$

Expanding $U_{n,n+e_\mu} \equiv 1 - iaA_\mu + O(a^2)$, we also restore the interaction term, and the last term in Eq. (4.37) turns into the continuum mass term. The action (4.37) is invariant under gauge transformations when the U are transformed according to Eq. (4.35), while $\psi_n \to \Omega_n\psi_n$.

Equation (4.37) is called the "naïve lattice fermion action", and I have to say that if the reader was convinced by the above reasoning that, in the continuum limit, it goes over to Eq. (4.38), he was naïve, too. Our implicit assumption was that the fermion fields ψ_n depend on the lattice node n in a smooth manner, so that the finite difference $\psi_{n+e_\mu} - \psi_{n-e_\mu}$ goes over to the continuum derivative. It turns out, however, that fermion field configurations which behave as $\psi_n \sim (-1)^{n_1}$ or $\psi_n \sim (-1)^{n_2+n_4}$ and change significantly at the microscopic lattice scale are equally important. After carefully performing continuum limit, these wildly oscillating modes give rise to 15 extra light fermion species with the same mass, the so-called *doublers*.

Simple-minded modifications of this naïve action, which leave only one light fermion for each flavor in the continuum limit, do not respect the chiral invariance of the theory — an issue which we did not yet discuss. We postpone a detailed discussion of these problems to Lecture 13. We will see that all these difficulties can be overcome, and a good consistent definition of the QCD path integral exists.

It is important that the action (4.37), as well as its sophistications, are bilinear in $\bar{\psi}_n$, ψ_n. The fermionic part of the path integral has the form

$$\int\prod_i d\bar{\psi}_i d\psi_i \exp\{-\mathfrak{D}_{ij}\bar{\psi}_i\psi_j\}\ , \tag{4.39}$$

where $i \equiv (n,\alpha)$, with α marking both the color and Lorentz spinor index, and $\mathfrak{D}_{ij}$ is a matrix turning into the Euclidean Dirac operator $i\not{\partial}^E - m$ in the continuum limit. We have learned how to perform integrals like (4.39)

in the previous lecture. According to Eq. (3.41), the answer is just $\det \|\mathfrak{D}\|$. Thus, the full path integral for QCD has the form

$$Z = \int \prod_{\text{links}} \mathcal{D}U_{\text{link}} \prod_{f} \det \|\mathfrak{D}_f\| \exp\left\{ -\frac{2N}{g^2} \sum_{\text{plaq}} (1 - W_{\text{plaq}}) \right\} , \quad (4.40)$$

where $\prod_f$ runs over all quark flavours.

Numerical calculations of the integrals like (4.40) are technically very difficult: not only one has to do a multidimensional integral, but also the *integrand* becomes very complicated involving the determinant of a large matrix. But they are possible, and this problem is being tackled now.

Lecture 5

θ–Vacuum

It was already mentioned that, for Yang–Mills theory, the Dirac quantization procedure giving the equation system (4.15) is not yet the full story. An additional constraint called *superselection rule* is required. To understand it better, consider first a simple quantum mechanical model.

5.1 Quantum Pendulum.

The classical Hamiltonian of a pendulum is

$$H = \frac{p_\phi^2}{2} + \omega^2(1 - \cos\phi) \tag{5.1}$$

($\omega^2 = g/l,\ m = l = 1$). Motion is finite $0 \leq \phi \leq 2\pi$, with the points 0 and 2π identified.

We want now to quantize the system. The easiest way to do it is to substitute $-i\partial/\partial\phi$ for p_ϕ into the Hamiltonian (5.1) and to impose the boundary condition $\Psi(2\pi) = \Psi(0)$ on the wave functions. Then the spectrum of the system is discrete. When there is no gravity and $\omega = 0$, the solution is especially simple:

$$\Psi_k(\phi) = \frac{1}{\sqrt{2\pi}} e^{ik\phi}, \qquad \mathcal{E}_k = \frac{k^2}{2} \tag{5.2}$$

with integer k. However, one is not forced to impose the requirement of *strict* periodicity on the wave function. The boundary condition

$$\Psi(\phi + 2\pi) = e^{i\theta}\Psi(\phi) \tag{5.3}$$

with any $\theta \in [0, 2\pi)$ is admissible as well. The condition (5.3) selects a different set of states which is, again, discrete. When $\omega = 0$, we have

$$\Psi_k(\phi) = \frac{1}{\sqrt{2\pi}} e^{i(k+\theta/2\pi)\phi}, \qquad \mathcal{E}_k = \frac{(k+\theta/2\pi)^2}{2} . \tag{5.4}$$

Anticipating field theoretic generalizations, let us express this transparent result in more general terms. In our case, the operator rotating the pendulum by one turn counterclockwise,

$$\hat{U}\Psi(\phi) = \Psi(\phi + 2\pi) , \tag{5.5}$$

commutes with the Hamiltonian. This is true in the free case, and also when a periodic in ϕ potential term is present. We can then simultaneously diagonalize the Hamiltonian and the operator $\hat{U}$. Eigenvalues of $\hat{U}$ are certain complex numbers z lying on the unit circle: $z \equiv e^{i\theta}$ [this condition follows from unitarity: the operator (5.5) preserves the normalization of the wave function]. Bearing all this in mind, we can subdivide the *large* Hilbert space of the Hamiltonian (5.1), including all the states without any particular boundary condition on the wave function imposed*, into the sectors characterized by definite eigenvalues $e^{i\theta}$ of the operator $\hat{U}$. A small Hilbert space (or the θ-sector) for the free pendulum is described by Eq. (5.4). The spectrum (5.4) is drawn in Fig. 5.1 as a function of θ. Thick lines mark the ground state of the system (the vacuum). We see that the dependence $\mathcal{E}_{\rm vac}(\theta)$ is discontinuous at $\theta = \pi(2k+1)$. When $\omega \neq 0$, the functions $\Psi_n^\theta(\phi)$ and $\mathcal{E}_n(\theta)$ are more complicated, but the picture is qualitatively the same.

The quantity θ and the picture in Fig. 5.1 have a very natural physical interpretation. We need to solve the problem of quantum motion in a periodic potential $V(\phi + 2\pi) = V(\phi)$. The problem has been solved at the dawn of quantum mechanics by Bloch who was interested in the quantum behaviour of electrons in crystals. Consulting standard textbooks, one finds that an alias for the quantity θ entering Eq. (5.3) is *quasi-momentum*. The analog of the spectrum (5.4) for a periodic potential in interest gives a zone structure of our "crystal". Each such zone corresponds to the interval $0 \leq \theta \leq 2\pi$ with fixed k. In the free case, the zones cover the whole range

*A basis of this large Hilbert space is formed by the plane waves $\Psi(\phi) = (1/\sqrt{2\pi})e^{ip_\phi \phi}$ with arbitrary p_ϕ.

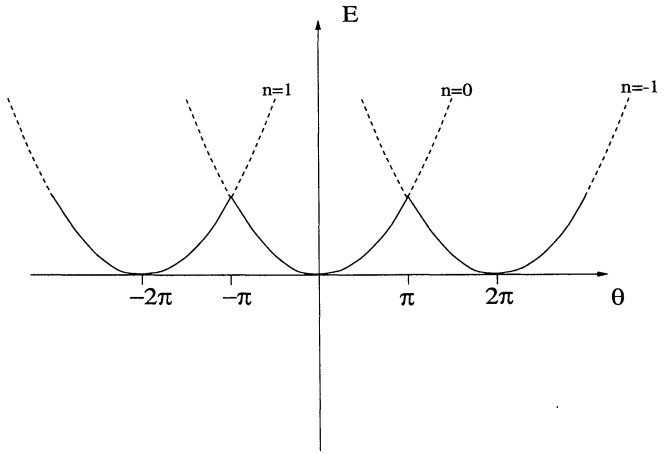


Fig. 5.1 Spectrum of the quantum pendulum as a function of θ. The solid line is the true ground state energy $\mathcal{E}_{\rm vac}(\theta) = \min_n\{\mathcal{E}_n(\theta)\}$. It is periodic in θ with the period 2π.

of energy, but when $\omega \neq 0$ and, especially, when ω is large, the zones are separated by gaps.

Let us find now the "effective mass" of our "Bloch electron"

$$\chi = m_{\rm eff}^{-1} = \frac{\partial^2 \mathcal{E}_{\rm vac}(\theta)}{\partial\theta^2}\,|_{\theta=0} \tag{5.6}$$

for a true pendulum in a gravity field ($\omega \neq 0$). Heuristically, the larger ω, the more difficult it is for the pendulum to overcome the potential barrier, and to perform a rotation by 2π. Thus, $m_{\rm eff}$, measuring the inertia of the system with respect to such rotations, should be large (and χ should be small) for large ω. We will find χ analytically in the region $\omega \gg 1$, using path integral methods. This will teach us many important lessons, which we will later use for QCD.

As a first step, substitute the definition

$$\mathcal{E}_{\rm vac}(\theta) = -\lim_{\beta\to\infty}\frac{1}{\beta}\ln Z(\theta)$$

into Eq. (5.6). Noting that $\partial Z(\theta)/\partial\theta|_{\theta=0} = 0$ due to the symmetry $Z(\theta) = Z(-\theta)$, we obtain

$$\chi = -\lim_{\beta\to\infty}\frac{1}{\beta Z(0)}\frac{\partial^2 Z(\theta)}{\partial\theta^2}\,|_{\theta=0}\ . \tag{5.7}$$

Next, we have to express

$$Z(\theta) = \sum_k{}^{\theta}\, e^{-\beta \mathcal{E}_k(\theta)} \tag{5.8}$$

(here the sum $\sum^{\theta}$ is performed only over states from sectors with a given value of θ) as a path integral. This is similar to what we have done in the previous lecture, when we projected onto the subspace of gauge-invariant wave functions:†

$$Z(\theta) \;=\; \frac{1}{2\pi} \sum_{q=-\infty}^{\infty} e^{-iq\theta} \int_0^{2\pi} d\phi\; \mathcal{K}(\phi + 2\pi q,\; \phi;\; -i\beta)\;, \tag{5.9}$$

where $\mathcal{K}$ is the Euclidean evolution operator of the "unconstrained" system defined in the large Hilbert space, which includes all eigenstates of the Hamiltonian without any boundary condition on the wave function specified. The evolution operator $\mathcal{K}(\phi + 2\pi q,\; \phi;\; -i\beta)$ describes the amplitude for the system to rotate counterclockwise q turns ($|q|$ turns clockwise for negative q) over an Euclidean time interval β. The equivalence of Eq. (5.8) and Eq. (5.9) can be explicitly seen if one substitutes into Eq. (5.9) the spectral decomposition (3.13) for the Euclidean evolution operator, and uses the identity

$$\sum_{q=-\infty}^{\infty} e^{-iq(\theta-\theta')} \;=\; 2\pi\delta(\theta - \theta')\;. \tag{5.10}$$

Thirdly, the evolution operator can be represented as a path integral according to Eq. (3.14). Substituting it into Eq. (5.9), we obtain

$$Z(\theta) \sim \sum_{q=-\infty}^{\infty} e^{-iq\theta} \int \prod_{\tau} d\phi(\tau) \exp\left\{ -\int_0^{\beta} \left[\frac{\dot{\phi}^2}{2} + \omega^2(1 - \cos\phi) \right] d\tau \right\}\;, \tag{5.11}$$

where each term of the sum involves the path integral over all the trajectories with boundary conditions $\phi(\beta) = \phi(0) + 2\pi q$. As actually $\phi \equiv \phi + 2\pi$, one should rather speak of *topologically nontrivial* closed trajectories describing rotation of the pendulum by a specific number of turns.

The fourth and final step is to calculate the path integrals entering the sum (5.11). When $\omega \gg 1$, the action of all topologically non-trivial

†We will soon see that the similarity here is far from being accidental.

trajectories is large. Hence the path integrals Z_q with $q \neq 0$ are suppressed, compared to the topologically trivial integral with $q = 0$. If we kept in the sum only the latter, we would not detect any dependence of the partition function on θ. The susceptibility $\chi^{q=0}$ would vanish, and the effective mass would be infinite. To get a non-zero χ, we have to take into account small nonzero Z_q. For large ω, the integral can be estimated by the saddle point (semi-classical) approximation: $Z_q \propto \exp\{-S_q\}$, where S_q is the *minimal* possible action in the class of all topologically non-trivial Euclidean trajectories with q turns, i.e. the instanton action.

The reader has already realized, of course, that we are using now the terminology of Lecture 2, where topologically non-trivial Euclidean configurations and solutions in Yang–Mills field theory were discussed. The analogy is, indeed, very precise, as we will shortly see. But let us first finish our calculation.

We are interested in the limit $\beta \to \infty$. Before taking it, it is convenient to rewrite the integrals in Eq. (5.11) in the symmetric form, $\int_{-\beta/2}^{\beta/2} \to \int_{-\infty}^{\infty}$. A saddle point in the integral with, say, $q = 1$ is the solution of the Euclidean equations of motion with boundary conditions

$$\phi(-\infty) = 0 \ , \quad \phi(\infty) = 2\pi \ . \tag{5.12}$$

The equations admit an integral of motion (normally, it has the meaning of energy, cf. Eq. (1.31), but here we are solving the equations in Euclidean time and such an interpretation is absent) which should be zero due to the boundary conditions (5.12):

$$\frac{\dot{\phi}^2}{2} - \omega^2(1 - \cos\phi) \ = \ 0 \ .$$

This equation is easily solved:

$$\phi \ = \ 4 \arctan e^{\omega(\tau - \tau_O)} \ . \tag{5.13}$$

This tunneling trajectory is the quantum mechanical analog of an instanton. τ_0 is the position of the instanton center in Euclidean time. The action of the solution (5.13) is ‡

$$S_I \ = \ 4\omega^2 \int_{-\infty}^{\infty} \frac{d\tau}{\cosh^2(\omega\tau)} \ = \ 8\omega \ . \tag{5.14}$$

‡Note in passing that the solution we have just derived describes also a soliton in the two-dimensional Sine–Gordon model.

That means that the full contribution to the partition function from the topologically nontrivial trajectories with unit winding q is

$$Z_1 \propto (\beta\omega)\exp\{-S_I\} = (\beta\omega)\exp\{-8\omega\}\ , \tag{5.15}$$

where the large factor β comes from the integral over the collective coordinate τ_0, on which the action does not depend. The contribution from trajectories with $q = -1$ is exactly the same. Consider now the contribution from the sector with $q = 2$. The functional integral is saturated by the classical solution with boundary conditions $\phi(-\infty) = 0$, $\phi(\infty) = 4\pi$. It is not difficult to realize that this is actually a combination of two solutions (5.14) separated by a large Euclidean time interval. The corresponding contribution to the action is $Z_2 \propto (\beta\omega)^2\exp\{-16\omega\}$. We see that, as far as the exponential factors are concerned, $Z_2 = Z_{-2}$ is suppressed compared to Z_1. On the other hand, Z_2 involves the extra factor β, which we agreed to keep large. Let it first be large, but not exponentially large so that the "fugacity"

$$\kappa \ = \ \frac{Z_1}{Z_0} \ = C\beta\omega\exp\{-8\omega\} \tag{5.16}$$

is much less than 1. In this case, all terms with higher q in the sum (5.11) may be neglected:

$$Z(\theta) \ \approx \ Z_0[1 + 2\kappa\cos\theta]\ , \tag{5.17}$$

and we derive for the topological susceptibility (5.7)

$$\chi \ = \ \frac{2\kappa}{\beta} \ = 2C\omega\exp\{-8\omega\}\ . \tag{5.18}$$

The effective mass is inversely proportional to the amplitude of quantum jump, which is the exponential of the action of the classical tunneling trajectory: a very natural and transparent result. The pre-exponential constant C in Eq. (5.18) can in principle be determined by calculating the functional integral in the Gaussian approximation. An analogous calculation for the instantons in QCD was done by 't Hooft.

Methodologically, it is very interesting to study also the opposite limit $\kappa \gg 1$. For large κ, all the terms in the series (5.11) are relevant. And not only that. Even in the topologically trivial sector with $\phi(2\pi) = \phi(0)$, a characteristic contribution to the path integral involves a certain number of instantons n_+ compensated by an equal amount of anti-instantons n_-:

our pendulum may make a number of counterclockwise turns and the same number of clockwise turns. Likewise, in the sector with nonvanishing net winding number q, the functional integral has the form of a double sum over n_+ and n_- with the restriction $n_+ - n_- = q$.

The leading contribution comes from the region where the instantons and anti-instantons are very well separated from each other. Then the functional integral presents roughly a product of the factors associated with individual (anti-)instantons, and only a combinatorial factor appears:

$$Z_q = \sum_{n_+ - n_- = q} \frac{\kappa^{n_+ + n_-}}{n_+! n_-!} = I_{|q|}(2\kappa) \; . \tag{5.19}$$

(The Bessel function $I_\nu(x)$ grows exponentially at large x.) The full partition function is

$$Z(\theta) = \sum_{q=-\infty}^{\infty} Z_q e^{-iq\theta} = \sum_{q=-\infty}^{\infty} I_{|q|}(2\kappa) e^{-iq\theta} = e^{2\kappa \cos\theta} \; . \tag{5.20}$$

Sunstituting it into (5.7), we again obtain (5.18), which universally holds both for small and large κ, with the only condition that the action $S_E = 8\omega$ is large and that β should be large compared to the characteristic oscillation period of the pendulum: $\beta \gg \omega^{-1}$.

This exercise was not idle. It teaches a lot concerning vacuum dynamics of QCD: the *instanton gas model* for the QCD vacuum state is introduced in a very similar manner. We will return to the discussion of this question in Lecture 15. Before abandoning this beautiful toy model, let us do one more exercise and rewrite the path integral (5.11) as

$$Z(\theta) \sim \int \prod_\tau d\phi(\tau) \exp\left\{ - \int_0^\beta \left[L_E + i \frac{\theta}{2\pi} \dot{\phi} \right] d\tau \right\} \; , \tag{5.21}$$

where the integral is done over all closed, topologically trivial as well as nontrivial trajectories. The factor $\exp\{-iq\theta\}$ is traded here for an extra term in the Lagrangian called *θ-term*. Usually, adding a total time derivative to the Lagrangian changes nothing, but in quantum theory this remains true only if no topologically nontrivial trajectories are involved. If $q \neq 0$, also $\int_0^\beta \dot{\phi} \equiv 2\pi q \neq 0$. The functional integral is modified and a total derivative leads to quite observable effects like a modified spectrum, etc. The analytic continuation of the integral (5.21) to a real time gives a standard Minkowski

path integral with $L_M \to L_M - (\theta/2\pi)\dot{\phi}$; the extra term in the Minkowski Lagrangian is real.

5.2 Large Gauge Transformations in Non-Abelian Theory.

Wave functionals representing the solutions of the equation system (4.15) are annihilated by the constraints $\hat{G}^a$, which means that they stay invariant under the action of the operator

$$\exp\left\{-i\int d^3x\ \chi^a(\boldsymbol{x})\hat{G}^a(\boldsymbol{x})\right\}$$

with small $\chi^a(\boldsymbol{x})$. This is tantamount to saying that the wave functionals are invariant under infinitesimal gauge transformations (1.10) of their arguments, $A_i \to A_i^\chi$, and hence also under finite *topologically trivial* gauge transformations (1.9), i.e. the transformations which can be reduced to a continuous series of infinitesimal gauge transformations (1.10). However, not all gauge transformations are topologically trivial. An example of a topologically nontrivial one is

$$\Omega_*(\boldsymbol{x}) \;=\; \exp\left\{\frac{i\pi x^a\sigma^a}{\sqrt{\boldsymbol{x}^2+\rho^2}}\right\} \tag{5.22}$$

with an arbitrary finite parameter ρ. We have $\Omega_*(\mathbf{0}) = 1$. As $|\boldsymbol{x}| \to \infty$, $\Omega_*(\boldsymbol{x}) \to -1$ irrespectively of the direction $\boldsymbol{x}/|\boldsymbol{x}|$ along which infinity is approached. For any other $|\boldsymbol{x}|$, the function $\Omega_*(\boldsymbol{x})$ gives an element of $SU(2)$, with every such element appearing only once. Therefore the function (5.22) represents a topologically nontrivial mapping of S^3 (R^3 compactified at infinity) onto the gauge group $SU(2)$. Indeed, the Chern–Simons number of the pure gauge configuration $A_i = -i\partial_i\Omega_*\Omega_*^\dagger$,

$$\begin{aligned} N_{C.S.} \;&=\; \int K_0 d^3x \\ &=\; \frac{1}{24\pi^2}\int d\boldsymbol{x}\ \epsilon_{ijk}\,\mathrm{Tr}\left\{(\partial_i\Omega_*)\Omega_*^\dagger(\partial_j\Omega_*)\Omega_*^\dagger(\partial_k\Omega_*)\Omega_*^\dagger\right\}, \end{aligned} \tag{5.23}$$

with K_μ being defined in Eq. (2.7), is equal to one. The gauge transformation (5.22) is topologically nontrivial, which means that one cannot find a family of gauge transformations $\Omega(\alpha, \boldsymbol{x})$ continuous in α such that

path integral with $L_M \to L_M - (\theta/2\pi)\dot{\phi}$; the extra term in the Minkowski Lagrangian is real.

5.2 Large Gauge Transformations in Non-Abelian Theory.

Wave functionals representing the solutions of the equation system (4.15) are annihilated by the constraints $\hat{G}^a$, which means that they stay invariant under the action of the operator

$$\exp\left\{-i\int d^3x\ \chi^a(\boldsymbol{x})\hat{G}^a(\boldsymbol{x})\right\}$$

with small $\chi^a(\boldsymbol{x})$. This is tantamount to saying that the wave functionals are invariant under infinitesimal gauge transformations (1.10) of their arguments, $A_i \to A_i^\chi$, and hence also under finite *topologically trivial* gauge transformations (1.9), i.e. the transformations which can be reduced to a continuous series of infinitesimal gauge transformations (1.10). However, not all gauge transformations are topologically trivial. An example of a topologically nontrivial one is

$$\Omega_*(\boldsymbol{x}) \;=\; \exp\left\{\frac{i\pi x^a\sigma^a}{\sqrt{\boldsymbol{x}^2+\rho^2}}\right\} \tag{5.22}$$

with an arbitrary finite parameter ρ. We have $\Omega_*(\mathbf{0}) = 1$. As $|\boldsymbol{x}| \to \infty$, $\Omega_*(\boldsymbol{x}) \to -1$ irrespectively of the direction $\boldsymbol{x}/|\boldsymbol{x}|$ along which infinity is approached. For any other $|\boldsymbol{x}|$, the function $\Omega_*(\boldsymbol{x})$ gives an element of $SU(2)$, with every such element appearing only once. Therefore the function (5.22) represents a topologically nontrivial mapping of S^3 (R^3 compactified at infinity) onto the gauge group $SU(2)$. Indeed, the Chern–Simons number of the pure gauge configuration $A_i = -i\partial_i\Omega_*\Omega_*^\dagger$,

$$\begin{aligned} N_{C.S.} \;&=\; \int K_0 d^3x \\ &=\; \frac{1}{24\pi^2}\int d\boldsymbol{x}\ \epsilon_{ijk}\,\mathrm{Tr}\left\{(\partial_i\Omega_*)\Omega_*^\dagger(\partial_j\Omega_*)\Omega_*^\dagger(\partial_k\Omega_*)\Omega_*^\dagger\right\}, \end{aligned} \tag{5.23}$$

with K_μ being defined in Eq. (2.7), is equal to one. The gauge transformation (5.22) is topologically nontrivial, which means that one cannot find a family of gauge transformations $\Omega(\alpha, \boldsymbol{x})$ continuous in α such that

each $\Omega(\alpha, \boldsymbol{x})$ represents a smooth function on S^3 (the smoothness requirement means in particular that the limiting value of $\Omega(\alpha, \boldsymbol{x})$ at $|\boldsymbol{x}| \to \infty$ exists and is uniquely defined; the necessity of this restriction will be seen a little bit later), $\Omega(0, \boldsymbol{x}) = 1$, and $\Omega(1, \boldsymbol{x}) = \Omega_*(\boldsymbol{x})$. $\Omega_*(\boldsymbol{x})$ may be called a "large" gauge transformation. The Gauss law constraint says nothing about what happens with the wave functional after its arguments are transformed by $\Omega_*(\boldsymbol{x})$.

Along with $\Omega_*(\boldsymbol{x})$, there are further distinct topologically nontrivial transformations $\Omega^{(q)}(\boldsymbol{x})$ belonging to the same homotopy class as (i.e. continuously deformable to) $[\Omega_*(\boldsymbol{x})]^q$ with integer q. Two $\Omega^{(q)}(\boldsymbol{x})$ with different q cannot be smoothly transformed into one another. The classes $\{\Omega^{(q)}(\boldsymbol{x})\}$ form what is called a homotopy group $\pi_3[G] = \mathbb{Z}$.

Note now that the following large gauge transformation operator,

$$\hat{U}\Psi[A_i(\boldsymbol{x})] = \Psi[A_i^{\Omega_*}(\boldsymbol{x})] \,, \tag{5.24}$$

commutes with the Hamiltonian and with all other physical gauge-invariant operators. This means that the operator (5.24) has the same meaning and can be handled in the same way as the operator (5.5) of the rotation by 2π for the pendulum. Thus, we diagonalize the Hamiltonian and the operator of the large gauge transformation (5.24) simultaneously, and divide the large Hilbert space spanned by all the eigenstates of the Hamiltonian satisfying the Gauss law constraints into sectors with a definite eigenvalue of $\hat{U}$. The latter is $e^{i\theta}$, as a large gauge transformation does not change the norm of the wave function. We are allowed and, moreover, are forced to consider only one such distinct sector because no physical operator has nonvanishing matrix elements between the states with different θ. Once finding ourselves in a state characterized by some θ, we can go over to a different state in the same sector as a result of some perturbation, but the value of θ *can*not be changed. It is a fundamental constant of our World, fixed once and forever. The dynamics of quantum Yang–Mills theory is simply not specified until the value of θ is given.

The analog of Eq. (5.11) reads

$$Z(\theta) \sim \sum_{q=-\infty}^{\infty} e^{-iq\theta} \int \prod_{\tau,\boldsymbol{x}} dA_\mu^a(\tau, \boldsymbol{x}) \exp\left\{ -\frac{1}{4g^2} \int_0^\beta d\tau \int d\boldsymbol{x} \left[(F_{\mu\nu}^a)^2 \right] \right\} , \tag{5.25}$$

with the boundary conditions

$$A_\mu(\boldsymbol{x},\beta) = [A_\mu(\boldsymbol{x},0)]^{[\Omega_*^q(\boldsymbol{x})]} \tag{5.26}$$

for the $q^{\underline{th}}$ term in the sum. The expression (5.25) defines the thermal partition function. Tending $\beta \to \infty$, and substituting $\int_0^\beta \equiv \int_{-\beta/2}^{\beta/2} \to \int_{-\infty}^{\infty}$, we obtain the partition function concentrated on the vacuum state.

Let us look at the $q = 1$ term in Eq. (5.25). The important fact is that the Pontryagin index (2.6) of Euclidean field configurations in the corresponding path integral is equal to 1. To see that, consider the configuration $A_i(-\infty,\boldsymbol{x}) = 0$, $A_i(\infty,\boldsymbol{x}) = -i\partial_i\Omega_*\Omega_*^\dagger$. Present our Euclidean space as the cylinder $S^3 \times R$ where R is Euclidean time, and S^3 is 3-dimensional space with the point at infinity added. Recall that the Pontryagin density $\sim F\tilde{F}$ is a total derivative, and that the 4-volume integral of $F\tilde{F}$ can be written as a surface integral according to Eq. (2.8). As the base S^3 of our cylinder has no boundary, the surface of $S^3 \times R$ consists of two spheres S^3 at $\tau = -\infty$ and $\tau = \infty$. The relation (2.7) is reduced to

$$\int d^4x \, \frac{1}{32\pi^2} F^a_{\mu\nu}\tilde{F}^a_{\mu\nu} \;=\; N_{C.S.}(\tau=\infty) - N_{C.S.}(\tau=-\infty)\;. \tag{5.27}$$

Using compact S^3 instead of R^3 as a base is quite appropriate in our case, but may seem at first a bit confusing. Alternatively, one could consider a disk D^3 with large, finite radius, and an integral of the Pontryagin density over $D^3 \times R$. One can then show that the integral of the Chern–Simons current in Eq. (2.7) over the side surface of the cylinder vanishes if the radius of the disk is much larger than the characteristic scale ρ.

When the coupling constant g is small (for the time being, we do not yet understand under what conditions this is true, this will be explained in Lecture 9, devoted to asymptotic freedom), the action of all topologically non-trivial configurations is large. The integral can be done semi-classically and $Z_1 = Z_{-1} \sim \exp\{-8\pi^2/g^2\}$, where $8\pi^2/g^2$ is the action of the instanton solution studied in Lecture 2.

Thus, we understand now the physical meaning of the instanton. It is a semi-classical tunneling trajectory connecting the classical, trivial vacuum $A_i = 0$ and the configuration $-i\partial_i\Omega_*\Omega_*^\dagger$ obtained from the trivial vacuum when applying a large gauge transformation. When the coupling constant g is small, the tunneling amplitude is suppressed as $\exp\{-S_I\}$, where $S_I = 8\pi^2/g^2$ is the instanton action. Everything is the same as for quantum pendulum and the physical interpretation of θ — the quasi-momentum

describing the drift of the system between the degenerate vacua characterized by different definite Chern–Simons numbers — is also the same.

We are now able to explain why the requirement that $\Omega(|\boldsymbol{x}| \to \infty)$ be uniquely defined was imposed in the first place. One could, of course, consider gauge transformations which tend to different values when $\boldsymbol{x} \to \infty$ along different directions. For example, so-called Gribov copies with $\tilde{\Omega}(\boldsymbol{x}) \sim \sqrt{\Omega_*(\boldsymbol{x})}$ have this property (see Lecture 7 for more details). It turns out, however, that the action on a tunneling trajectory interpolating between the trivial vacuum $A_i = 0$ and the configuration $-i\partial_i\tilde{\Omega}(\boldsymbol{x})\tilde{\Omega}^\dagger(\boldsymbol{x})$ has an *infinite* action and therefore does not contribute to the path integral. In fact, it is just the presence of this infinite barrier which justifies the topological classification.

This tunneling trajectory (the so-called *meron* solution) may still present some interest, however (see Lecture 7 for further discussion). For reference purposes, let us write here the explicit form of the meron solution in a covariant gauge:

$$A^a_\mu \;=\; \eta^a_{\mu\nu}x_\nu/x^2 \;. \tag{5.28}$$

It differs from the pure gauge field configuration (2.16) by the absense of the overall factor 2 and, in contrast to (2.16), has a nonzero field strength

$$F^a_{\mu\nu} \;=\; \frac{x_\alpha(x_\mu\eta^a_{\alpha\nu} - x_\nu\eta^a_{\alpha\mu}) - x^2\eta^a_{\mu\nu}}{x^4} \tag{5.29}$$

and action density. The latter is proportional to $1/x^4$, and the action integral diverges logarithmically both at small and large $|x|$.

In the same way as we did it for the pendulum, one can rewrite the integral (5.25) in the form

$$Z(\theta) \;\sim\; \int \prod_{\tau,\boldsymbol{x}} dA^a_\mu(\tau,\boldsymbol{x}) \exp\left\{-\int_0^\beta d\tau \int d\boldsymbol{x}\left[\frac{1}{4g^2}F^a_{\mu\nu}F^a_{\mu\nu} + i\frac{\theta}{32\pi^2}F^a_{\mu\nu}\tilde{F}^a_{\mu\nu}\right]\right\}, \tag{5.30}$$

where the integral is performed over all field configurations which are periodic in τ up to a large gauge transformation. The second term in the integrand is a total derivative and is relevant when topologically nontrivial configurations come into play. It is the famous θ-term of QCD. Analytical

continuation of the path integral to Minkowski spacetime gives

$$\mathcal{L}_M \;\to\; \mathcal{L}_M + \frac{\theta}{32\pi^2} F^a_{\mu\nu}\tilde{F}^{a\;\mu\nu}\;, \tag{5.31}$$

and the θ-term is real.

What is the value of θ in the World we live? It is experiment rather than theory which can answer this question. Note that the θ-term breaks both P and T invariance. (Indeed, $F^a_{\mu\nu}\tilde{F}^{a\;\mu\nu}$ is the same as $\boldsymbol{E}^a \cdot \boldsymbol{B}^a$, where $\boldsymbol{E}^a$ is chromoelectric and $\boldsymbol{B}^a$ is chromomagnetic field. Now, $\boldsymbol{E}^a$ is a vector, which is odd under time reflection, while $\boldsymbol{B}^a$ is a pseudo-vector, even under time reflection.) We know, however, that neither parity nor T invariance are broken in strong interactions. This means that θ is either just zero or very small. The best experimental restriction comes from the measurements on the electric dipole moment d_n of the neutron. We know that $d_n \lesssim 10^{-25} e \cdot cm$ which means that $\theta \lesssim 10^{-9}$.[§]

The fact that θ is zero or very small has no explanation in the framework of QCD. Maybe the problem will be eventually solved when we understand the nature of the sought unified theory of all interactions, incorporating also QCD. The status of this problem is roughly the same as the status of the cosmological term problem: experiment tells us that the cosmological term is zero or very small, but we do not understand why.[¶]

[§]There is a subtlety here. All physical effects noninvariant with respect to P and T transformations are actually proportional to $\sin\theta$. Thus, the symmetry requirements alone restrict θ to be close to *either* 0 *or* π. We simply note here that the second possibility *is* excluded by what we know from experiments on the properties of light pseudoscalar mesons. This can be shown in the framework of the *effective chiral Lagrangian* describing light meson dynamics to be introduced in Lecture 12. We will discuss more of the physics of the imaginary world with $\theta \approx \pi$ in Lecture 16.

[¶]Probably, the analogy between θ and the cosmological term is even deeper. In the Ogievetsky–Sokachev formulation of supergravity, the cosmological term is also written as a total derivative [15]. But this question is far beyond the scope of this book.