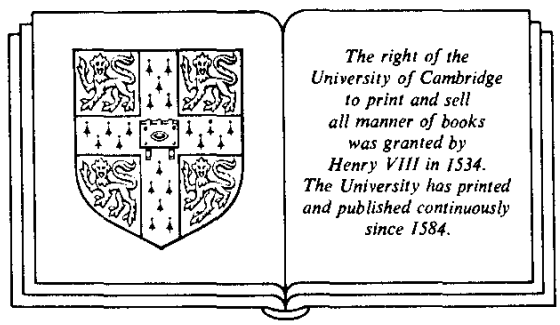
RENORMALIZATION

An introduction to renormalization, the renormalization group, and the operator-product expansion

JOHN C. COLLINS
Illinois Institute of Technology



CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON NEW YORK NEW ROCHELLE
MELBOURNE SYDNEY

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

© Cambridge University Press 1984

First published 1984

Printed in Great Britain at the University Press, Cambridge

Library of Congress catalogue card number: 83-23927

British Library cataloguing in publication data

Collins, John C.
Renormalization.—(Cambridge monographs on mathematical physics)
1. Renormalization (Physics) 2. Quantum field theory
I. Title
530.1'43 QC174.45

ISBN 0 521 24261 4

This work was supported in part by the US Department of Energy
under contract number DE-AC02-80ER10712.

TM

Contents

1	**Introduction**	1
2	**Quantum field theory**	4
2.1	Scalar field theory	4
2.2	Functional-integral solution	6
2.3	Renormalization	9
2.4	Ultra-violet regulators	12
2.5	Equations of motion for Green's functions	13
2.6	Symmetries	15
2.7	Ward identities	19
2.8	Perturbation theory	21
2.9	Spontaneously broken symmetry	25
2.10	Fermions	27
2.11	Gauge theories	28
2.12	Quantizing gauge theories	29
2.13	BRS invariance and Slavnov–Taylor identities	32
2.14	Feynman rules for gauge theories	34
2.15	Other symmetries of (2.11.7)	35
2.16	Model field theories	36
3	**Basic examples**	38
3.1	One-loop self-energy in ϕ^3 theory	38
3.1.1	Wick rotation	39
3.1.2	Lattice	40
3.1.3	Interpretation of divergence	41
3.1.4	Computation	43
3.2	Higher order	44
3.3	Degree of divergence	45
3.3.1	ϕ^3 at $d = 6$	47
3.3.2	Why may Z be zero and yet contain divergences?	48
3.3.3	Renormalizability and non-renormalizability	49
3.4	Renormalization group	50
3.4.1	Arbitrariness in a renormalized graph	50
3.4.2	Renormalization prescriptions	52
3.5	Dimensional regularization	53
3.6	Minimal subtraction	56

3.6.1	Definition	56
3.6.2	$d = 6$	57
3.6.3	Renormalization group and minimal subtraction	58
3.6.4	Massless theories	59
3.7	Coordinate space	59
4	**Dimensional regularization**	62
4.1	Definition and axioms	64
4.2	Continuation to small d	68
4.3	Properties	73
4.4	Formulae for Minkowski space	81
4.4.1	Schwinger parameters	81
4.4.2	Feynman parameters	83
4.5	Dirac matrices	83
4.6	γ_5 and $\varepsilon_{\kappa\lambda\mu\nu}$	86
5	**Renormalization**	88
5.1	Divergences and subdivergences	89
5.2	Two-loop self-energy in $(\phi^3)_6$	93
5.2.1	Fig. 5.1.2	94
5.2.2	Differentiation with respect to external momenta	97
5.2.3	Fig. 5.1.3	99
5.3	Renormalization of Feynman graphs	101
5.3.1	One-particle-irreducible graph with no subdivergences	102
5.3.2	General case	103
5.3.3	Application of general formulae	105
5.3.4	Summary	106
5.4	Three-loop example	106
5.5	Forest formula	109
5.5.1	Formula	109
5.5.2	Proof	110
5.6	Relation to $\mathscr{L}$	112
5.7	Renormalizability	116
5.7.1	Renormalizability and non-renormalizability	116
5.7.2	Cosmological term	118
5.7.3	Degrees of renormalizability	118
5.7.4	Non-renormalizability	120
5.7.5	Relation of renormalizability to dimension of coupling	121
5.7.6	Non-renormalizable theories of physics	122
5.8	Proof of locality of counterterms; Weinberg's theorem	125
5.8.1	Degree of counterterms equals degree of divergence	126
5.8.2	$\bar{R}(G)$ is finite if $\delta(G)$ is negative	127
5.8.3	Asymptotic behavior	129
5.9	Oversubtractions	130
5.9.1	Mass-shell renormalization and oversubtraction	130

5.9.2	Remarks	131
5.9.3	Oversubtraction on 1PR graphs	131
5.10	Renormalization without regulators: the BPHZ scheme	133
5.11	Minimal subtraction	135
5.11.1	Definition	135
5.11.2	$\overline{\text{MS}}$ renormalization	137
5.11.3	Minimal subtraction with other regulators	137
6	**Composite operators**	138
6.1	Operator-product expansion	139
6.2	Renormalization of composite operators: examples	142
6.2.1	Renormalization of ϕ^2	142
6.2.2	Renormalization of $\phi^2(x)\phi^2(y)$	145
6.3	Definitions	146
6.4	Operator mixing	149
6.5	Tensors and minimal subtraction	150
6.6	Properties	152
6.7	Differentiation with respect to parameters in $\mathscr{L}$	163
6.8	Relation of renormalizations of ϕ^2 and m^2	166
7	**Renormalization group**	168
7.1	Change of renormalization prescription	169
7.1.1	Change of parametrization	169
7.1.2	Renormalization-prescription dependence	171
7.1.3	Low-order examples	172
7.2	Proof of RG invariance	176
7.3	Renormalization-group equation	180
7.3.1	Renormalization-group coefficients	180
7.3.2	RG equation	183
7.3.3	Solution	184
7.4	Large-momentum behavior of Green's functions	185
7.4.1	Generalizations	187
7.5	Varieties of high- and low-energy behavior	187
7.5.1	Asymptotic freedom	187
7.5.2	Maximum accuracy in an asymptotically free theory	189
7.5.3	Fixed point theories	190
7.5.4	Low-energy behavior of massless theory	191
7.6	Leading logarithms, etc.	193
7.6.1	Renormalization-group logarithms	193
7.6.2	Non-renormalization-group logarithms	196
7.6.3	Landau ghost	196
7.7	Other theories	197
7.8	Other renormalization prescriptions	200
7.9	Dimensional transmutation	203
7.10	Choice of cut-off procedure	206

7.10.1	Example: ϕ^4 self-energy	208
7.10.2	RG coefficients	208
7.10.3	Computation of g_0 and Z: asymptotically free case	209
7.10.4	Accuracy needed for g_0	213
7.10.5	m_0^2	216
7.10.6	Non-asymptotically free case	216
7.11	Computing renormalization factors using dimensional regularization	217
7.12	Renormalization group for composite operators	219
8	**Large-mass expansion**	222
8.1	A model	224
8.2	Power-counting	225
8.2.1	Tree graphs	225
8.2.2	Finite graphs with heavy loops	226
8.2.3	Divergent one-loop graphs	226
8.2.4	More than one loop	229
8.3	General ideas	230
8.3.1	Renormalization prescriptions with manifest decoupling	231
8.3.2	Dominant regions	232
8.4	Proof of decoupling	233
8.4.1	Renormalization prescription R^* with manifest decoupling	233
8.4.2	Definition of R^*	236
8.4.3	IR finiteness of $C^*(\Gamma)$	237
8.4.4	Manifest decoupling for R^*	239
8.4.5	Decoupling theorem	239
8.5	Renormalization-group analysis	240
8.5.1	Sample calculation	242
8.5.2	Accuracy	243
9	**Global symmetries**	244
9.1	Unbroken symmetry	245
9.2	Spontaneously broken symmetry	247
9.2.1	Proof of invariance of counterterms	249
9.2.2	Renormalization of the current	251
9.2.3	Infra-red divergences	252
9.3	Renormalization methods	253
9.3.1	Generation of renormalization conditions by Ward identities	254
10	**Operator-product expansion**	257
10.1	Examples	258
10.1.1	Cases with no divergences	258
10.1.2	Divergent example	260
10.1.3	Momentum space	262
10.1.4	Fig. 10.1.3 inside bigger graph	262

10.2	Strategy of proof	263
10.3	Proof	266
10.3.1	Construction of remainder	266
10.3.2	Absence of infra-red and ultra-violet divergences in $r(\Gamma)$	269
10.3.3	$R(\Gamma) - r(\Gamma)$ is the Wilson expansion	270
10.3.4	Formula for $\bar{C}(U)$	272
10.4	General case	273
10.5	Renormalization group	274
11	**Coordinate space**	277
11.1	Short-distance singularities of free propagator	278
11.1.1	Zero temperature	278
11.1.2	Non-zero temperature	280
11.2	Construction of counterterms in low-order graphs	281
11.3	Flat-space renormalization	286
11.4	External fields	291
12	**Renormalization of gauge theories**	293
12.1	Statement of results	296
12.2	Proof of renormalizability	298
12.2.1	Preliminaries	298
12.2.2	Choice of counterterms	300
12.2.3	Graphs with external derivatives	301
12.2.4	Graphs finite by equations of motion	302
12.2.5	Gluon self-energy	302
12.2.6	BRS transformation of A^a_μ	303
12.2.7	Gluon self-interaction	304
12.2.8	$\delta_R c$	304
12.2.9	Quark–gluon interaction, $\delta\psi$, $\delta\bar{\psi}$; introduction of $\mathscr{B}_{a\mu}$	305
12.3	More general theories	307
12.3.1	Bigger gauge group	307
12.3.2	Scalar matter	307
12.3.3	Spontaneous symmetry breaking	308
12.4	Gauge dependence of counterterms	309
12.4.1	Change of ξ	310
12.4.2	Change of F_a	313
12.5	R_ξ-gauge	314
12.6	Renormalization of gauge-invariant operators	316
12.6.1	Caveat	319
12.7	Renormalization-group equation	319
12.8	Operator-product expansion	321
12.9	Abelian theories: with and without photon mass	322
12.9.1	BRS treatment of massive photon	323
12.9.2	Elementary treatment of abelian theory with photon mass	324
12.9.3	Ward identities	325

12.9.4	Counterterms proportional to A^2 and ∂A^2	326
12.9.5	Relation between c_0 and Z_1	327
12.9.6	Gauge dependence	327
12.9.7	Renormalization-group equation	329
12.10	Unitary gauge for massive photon	330
13	**Anomalies**	331
13.1	Chiral transformations	333
13.2	Definition of γ_5	334
13.3	Properties of axial currents	338
13.3.1	Non-anomalous currents	338
13.3.2	Anomalous currents	338
13.3.3	Chiral gauge theories	339
13.3.4	Supersymmetric theories	339
13.4	Ward identity for bare axial current	339
13.4.1	Renormalization of operators in Ward identities	340
13.5	One-loop calculations	342
13.6	Non-singlet axial current has no anomaly	346
13.6.1	Reduction of anomaly	346
13.6.2	Renormalized current has no anomalous dimension	347
13.7	Three-current Ward identity; the triangle anomaly	349
13.7.1	General form of anomaly	349
13.7.2	One-loop value	351
13.7.3	Higher orders	352
14	**Deep-inelastic scattering**	354
14.1	Kinematics, etc.	355
14.2	Parton model	358
14.3	Dispersion relations and moments	359
14.4	Expansion for scalar current	361
14.5	Calculation of Wilson coefficients	364
14.5.1	Lowest-order Wilson coefficients	364
14.5.2	Anomalous dimensions	365
14.5.3	Solution of RG equation – non-singlet	367
14.6	OPE for vector and axial currents	367
14.6.1	Wilson coefficients – electromagnetic case	368
14.7	Parton interpretation of Wilson expansion	369
14.8	W_4 and W_5	371
	References	372
	Index	377

1

Introduction

The structure of a quantum field theory often simplifies when one considers processes involving large momenta or short distances. These simplifications are important in improving one's ability to calculate predictions from the theory, and in essence form the subject of this book.

The first simplification to be considered involves the very existence of the theory. The problem is that there are usually ultra-violet divergences caused by large fluctuations of the field(s) on short distance scales. These manifest themselves in Feynman graphs as divergences when loop momenta go to infinity with the external momenta fixed. The simplification is that the divergences can be cancelled by renormalizations of the parameters of the action. Consequently our first task will be to treat the ultra-violet renormalizations. Renormalization is essential, for otherwise most field theories do not exist.

We will then expose the methods needed to handle high-energy/short-distance problems. The aim is to be able to make testable predictions from a strong interaction theory, or to improve the rate of convergence of the perturbation expansion in a weakly coupled theory. The simplifications generally take the form of a factorization of a cross-section or of an amplitude, each factor containing the dependence of the process on phenomena that happen on one particular distance scale. Such a factorization is useful, because the coefficients of the perturbation expansion for a process are large when the process involves widely different distance scales.

The industry called 'perturbative QCD' consists of deriving such factorization theorems for strong interactions (Mueller (1981)) and exploring their phenomenological consequences. We will only study the earliest of these factorizations, the operator product expansion of Wilson (1969). We will also discuss the theorems that describe the behavior of a theory when the masses of its fields get large (Appelquist & Carazzone (1975) and Witten (1976)). These large-mass theorems have their main uses in weak interaction theories.

The presence of ultra-violet divergences, even though they are cancelled by renormalization counterterms, means that in any process there are

contributions from quantum fluctuations on every distance scale. This is both a complication and an opportunity to find interesting physics. The complication is that the derivation of factorization theorems is made difficult. The opportunity is given by the observable phenomena that directly result from the existence of the divergences. A standard example is given by the scaling violations in deep-inelastic scattering.

It is the renormalization group (Stueckelberg & Petermann (1953) and Gell-Mann & Low (1954)) that is the key technique in disentangling the complications. The infinite parts of the counterterms are determined by the requirement that they cancel the divergences, but the finite parts are not so determined. In fact, the partition of a bare coupling g_0 into the sum of a finite renormalized coupling g_R and a singular counterterm Δg is arbitrary. One can reparametrize the theory by transferring a finite amount from g_R to Δg without changing the physics: the theory is renormalization-group invariant.

This trivial-sounding observation is in practice very useful, and far from trivial. Suppose one has some graph whose renormalized value is large (so that it is inadequate to use a few low orders of the perturbation theory to compute the corresponding quantity). Then in appropriate circumstances it is possible to adjust the partition of g_0 (viz., $g_0 = g_R + \Delta g$) so that the counterterm Δg cancels not only the divergence but also the excessively large piece of the graph's finite part. The large piece is now in the lowest order instead of higher orders. Construction of factorization theorems of the sort reviewed by Mueller (1981) provides many circumstances where this trick is applicable. Without it the factorization theorems would be almost powerless.

We see that the subjects of renormalization, the renormalization group, and the operator product expansion are intimately linked, and we will treat them all in this book. The aim will be to explain the general methods that are applicable not only to the examples we will examine but in many other situations. We will not aim at complete rigor. However there are many pitfalls and traps ready to ensnare an unwary physicist. Thus a precise set of concepts and notations is necessary, for many of the dangers are essentially combinatorial. The appropriate basis is then that of Zimmermann (1969, 1970, 1973a, 1973b).

One other problem is that of choice of an ultra-violet cut off. From a fundamental point of view, the lattice cut-off seems best as it appears in non-perturbative treatments using the functional integral (e.g., Glimm & Jaffe (1981)). In perturbation theory one can arrange to use no regulator whatsoever (e.g., Piguet & Rouet (1981)). In practice, dimensional reg-

ularization has deservedly become very popular. This consists of replacing the physical space-time dimensionality 4 by an arbitrary complex number d. The main attraction of this method is that virtually no violence is done to the structure of a Feynman graph; a second attraction is that it also regulates infra-red divergences. The disadvantage is that the method has not been formulated outside of perturbation theory (at least not yet). Much of the treatment in this book, especially the examples, will be based on the use of dimensional regularization. However it cannot be emphasized too strongly that none of the fundamental results depend on this choice.

2

Quantum field theory

Most of the work in this book will be strictly perturbative. However it is important not to consider perturbation theory as the be-all and end-all of field theory. Rather, it must be looked on only as a systematic method of approximating a complete quantum field theory, with the errors under control. So in this chapter we will review the foundations of quantum field theory starting from the functional integral.

The purpose of this review is partly to set out the results on which the rest of the book is based. It will also introduce our notation. We will also list a number of standard field theories which will be used throughout the book. Some examples are physical theories of the real world; others are simpler theories whose only purpose will be to illustrate methods in the absence of complications.

The use of functional integration is not absolutely essential. Its use is to provide a systematic basis for the rest of our work: the functional integral gives an explicit solution of any given field theory. Our task will be to investigate a certain class of properties of the solution.

For more details the reader should consult a standard textbook on field theory. Of these, probably the most complete and up-to-date is by Itzykson & Zuber (1980); this includes a treatment of the functional integral method. Other useful references include: Bjorken & Drell (1966), Bogoliubov & Shirkov (1980), Lurié (1968), and Ramond (1981).

2.1 Scalar field theory

The simplest quantum field theory is that of a single real scalar field $\phi(x^\mu)$. The theory is defined by canonically quantizing a classical field theory. This classical theory is specified by a Lagrangian density:

$$\mathscr{L} = (\partial\phi)^2/2 - P(\phi), \tag{2.1.1}$$

from which follows the equation of motion

$$\Box\phi + P'(\phi) = 0. \tag{2.1.2}$$

Here $P(\phi)$ is a function of $\phi(x)$, which we generally take to be a polynomial like $P(\phi) = m^2\phi^2/2 + g\phi^4/4!$, and $P'(\phi) = \mathrm{d}P/\mathrm{d}\phi$. (Note that we use units with $\hbar = c = 1$.)

In the Hamiltonian formulation of the same theory, we define a canonical momentum field:

$$\pi(x) \equiv \partial\mathscr{L}/\partial\dot{\phi} = \dot{\phi} = \partial\phi/\partial t, \tag{2.1.3}$$

and the Hamiltonian

$$H = \int \mathrm{d}^3x(\pi^2/2 + \vec{\nabla}\phi^2/2 + P(\phi)). \tag{2.1.4}$$

Physically, we require that a theory have a lowest energy state. If it does not then all states are unstable against decay into a lower energy state plus a collection of particles. If the function $P(\phi)$ has no minimum, then the formula (2.1.4) implies that just such a catastrophic situation exists (Baym (1960)). Thus we require the function $P(\phi)$ to be bounded below.

Quantization proceeds in the Heisenberg picture by reinterpreting $\phi(x)$ as a hermitian operator on a Hilbert space satisfying the canonical equal-time commutation relations, i.e.,

$$\left.\begin{aligned}[\pi(x), \phi(y)] &= -\mathrm{i}\delta^{(3)}(\vec{x} - \vec{y}),\\ [\phi(x), \phi(y)] &= [\pi(x), \pi(y)] = 0\end{aligned}\right\} \quad \text{if} \quad x^0 = y^0. \tag{2.1.5}$$

The Hamiltonian is still given by (2.1.4) so the equation of motion (2.1.2) follows from the Heisenberg equation of motion

$$\mathrm{i}\partial\phi/\partial t = [\phi, H]. \tag{2.1.6}$$

A solution to the theory is specified by stating what the space of states is and by giving the manner in which ϕ acts on the states. We will construct a solution by use of the functional integral. It should be noted that $\phi(x)$ is in general not a well-behaved operator, but rather it is an operator-valued distribution. Physically that means that one cannot measure $\phi(x)$ at a single point, but only averages of $\phi(x)$ over a space-time region. That is,

$$\phi_f \equiv \int \phi(x) f(x) \mathrm{d}^4x, \tag{2.1.7}$$

for any complex-valued function $f(x)$, is an operator. Now, products of distributions do not always make sense (e.g., $\delta(x)^2$). In particular, the Hamiltonian H involves products of fields at the same point. Some care is needed to define these products properly; this is, in fact, the subject of renormalization, to be treated shortly.

The following properties of the theory are standard:

(1) The theory has a Poincaré-invariant ground state $|0\rangle$, called the vacuum.

(2) The states and the action of ϕ on them can be reconstructed from the time-ordered Green's functions

$$G_N(x_1,\ldots,x_N) = \langle 0|T\phi(x_1)\ldots\phi(x_N)|0\rangle. \tag{2.1.8}$$

The T-ordering symbol means that the fields are written in order of increasing time from right to left.

(3) The Green's functions have appropriate causality properties, etc., so that they are the Green's functions of a physically sensible theory. Mathematically, these properties are summarized by the Wightman axioms (Streater & Wightman (1978)).

Bose symmetry of the ϕ-field means that the Green's functions are symmetric under interchange of any of the x's. From the equations of motion of ϕ and from the commutation relations can be derived equations of motion for the Green's functions. The simplest example is

$$\Box_y G_2(y,x) + \langle 0|TP'(\phi(y))\phi(x)|0\rangle = -i\delta^{(4)}(x-y). \tag{2.1.9}$$

For a general $(N+1)$-point Green's function, we have N δ-functions on the right:

$$\begin{aligned}\Box_y G_{N+1}(y,x_1,\ldots,x_N) &+ \langle 0|TP'(\phi(y))\phi(x_1)\ldots\phi(x_N)|0\rangle \\ &= -i\sum_{j=1}^{N}\delta^{(4)}(y-x_j)G_{N-1}(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_N).\end{aligned} \tag{2.1.10}$$

This equation summarizes both the equations of motion and the commutation relations. Solving the theory for the Green's functions means in essence solving this set of coupled equations. It is in fact the Green's functions that are the easiest objects to compute. All other properties of the theory can be calculated once the Green's functions are known.

2.2 Functional-integral solution

The solution of a quantum field theory is a non-trivial problem in consistency. Only two cases are elementary: free field theory ($P = m^2\phi^2/2$), and the case of one space-time dimension, $d = 1$. The case $d = 1$ is a rather trivial field theory, for it is just the quantum mechanics of a particle with Heisenberg position operator $\phi(t)$ in a potential $P(\phi)$. (In Section 2.1, we explained the case $d = 4$. It is easy to go back and change the formulae to be valid for a general value of d.)

For the case of ϕ^4 theory, with

$$P(\phi) = m^2\phi^2/2 + g\phi^4/4!, \tag{2.2.1}$$

solutions are rigorously known to exist if $d = 2$ or 3 (Glimm & Jaffe (1981)). If $d > 4$ then no non-trivial solution exists (Aizenman (1981)). The case $d = 4$ is difficult; the difficulty is to perform renormalization of the ultra-violet divergences beyond perturbation theory. As we will see the theory at $d = 4$ is 'exactly renormalizable' in perturbation theory; this is the most interesting case. For the most part we will ignore the difficulties in going beyond perturbation theory. We will return to this problem in Section 7.10 when we discuss the application of the renormalization group outside of perturbation theory.

If we ignore, temporarily, the renormalization problem, then a solution for the theory can be found in terms of a functional integral. The formula for the Green's functions is written as

$$G_N(x_1, \ldots, x_N) = \mathcal{N} \int [\mathrm{d}A] \mathrm{e}^{\mathrm{i}S[A]} A(x_1) \ldots A(x_N). \tag{2.2.2}$$

(See Chapter 9 of Itzykson & Zuber (1980), or see Glimm & Jaffe (1981).) On the right-hand side of this equation $A(x)$ represents a classical field, and the integration is over the value of $A(x)$ at every space-time point. The result of the integral in (2.2.2) is the N-point Green's function for the corresponding quantum field, ϕ. In the integrand appears the classical action, which is

$$S[A] = \int \mathrm{d}^4x \mathscr{L}. \tag{2.2.3}$$

The normalization factor $\mathcal{N}$ is to give $\langle 0|0\rangle = 1$, so that

$$\mathcal{N} = \left\{ \int [\mathrm{d}A] \mathrm{e}^{\mathrm{i}S[A]} \right\}^{-1}. \tag{2.2.4}$$

Equivalent to (2.2.2) is the integral for the generating functional of Green's functions:

$$Z[J] = \mathcal{N} \int [\mathrm{d}A] \exp\left\{ \mathrm{i}S[A] + \int \mathrm{d}^4x J(x)A(x) \right\}, \tag{2.2.5}$$

where $J(x)$ is an arbitrary function. Functionally differentiating with respect to $J(x)$ gives the Green's functions, e.g.,

$$\langle 0|T\phi(x)\phi(y)|0\rangle = \frac{1}{Z[0]} \frac{\delta^2}{\delta J(x)\delta J(y)} Z[J]\Big|_{(J=0)}. \tag{2.2.6}$$

It is somewhat delicate to make precise the definition of the integration over A. The principal steps are:

(1) 'Wick-rotate' time to imaginary values: $t = -i\tau$, so that space-time is Euclidean. The exponent in the integral is then:

$$-S_{\text{Eucl}}[A] = -\int d\tau\, d^3x[-\partial A^2/2 + P(A)]. \tag{2.2.7}$$

With our metric, we have $\partial A^2 = -(\partial A/\partial\tau)^2 - \vec{\nabla}A^2$. We may subtract out from $\mathscr{L}$ the minimum value of $P(A)$; this subtraction gives an overall factor in the functional integral, and it cancels between the integral and the normalization factor (2.2.4). Therefore the Euclidean action S_{Eucl} is positive definite. The factor $\exp(-S_{\text{Eucl}})$ gives much better convergence for large A and for rapidly varying A than does $\exp(iS)$ in Minkowski space.

(2) Replace space-time by a finite lattice. We may choose a cubic lattice with spacing a. Its points are then

$$x^\mu = n^\mu a.$$

where the n^μ's are integers. They are bounded to keep $\vec{x}$ inside a spatial box of volume V and to keep τ within a range $-T/2$ to $+T/2$. The integral

$$\int [dA] A(x_1)\ldots A(x_N)\exp(-S_{\text{Eucl}}[A]) \tag{2.2.8}$$

is now an absolutely convergent ordinary integral over a finite number of variables. The action S_{Eucl} is given its obvious discrete approximation.

(3) Take the continuum limit $a \to 0$, and the limits of infinite volume V and infinite time T.

(4) Analytically continue back to Minkowski space-time.

The difficulties occur at step 3. Taking the limits of infinite T and V gives divergences of exactly the sort associated with taking the thermodynamic limit of a partition function – see below. Further divergences occur when the continuum limit $a \to 0$ is taken. In addition, the canonical derivation of (2.2.2) gives an overall normalization factor which goes to infinity as $a \to 0$ or as the number of space-time points goes to infinity; this factor is absorbed by the normalization $\mathscr{N}$.

The limits of infinite volume and time are under good control. They are literally thermodynamic limits of a classical statistical mechanical system in four spatial dimensions. Recall, for example, that in ϕ^4 theory one can write

$$S[A] = g^{-1}\int d^4x(\partial\hat{A}^2/2 - m^2\hat{A}^2/2 - \hat{A}^4/4!) \tag{2.2.9}$$

where $\hat{A} = g^{1/2}A$. Thus the integral $\int [dA] \exp(-S[A])$ is proportional to

$$\int [d\hat{A}] \exp\{-(1/g)S[A \to \hat{A}, g \to 1]\}. \tag{2.2.10}$$

This is the partition function of a classical system at temperature $1/g$, when the phase space is spanned by the field $\hat{A}$, and when the energy of a given configuration is

$$\hat{S}_{\text{Eucl}}[\hat{A}] = \int d^4x(-\partial\hat{A}^2/2 + m^2\hat{A}^2/2 + \hat{A}^4/4!).$$

The identity between Euclidean field theory and certain classical statistical mechanics systems has been fruitful both in working out the rigorous mathematical treatment of quantum field theory (Glimm & Jaffe (1981)) and in finding new ways to treat thermodynamic problems (Wilson & Kogut (1974)). As is particularly emphasized in Wilson's work, there is a lot of cross-fertilization between field theory and the theory of phase transitions. The methods of the renormalization group are common to both fields, and the continuum limit in field theory can be usefully regarded as a particular type of second-order phase transition.

The thermodynamic limit gives a factor $\exp(-\rho TV)$, where ρ is the ground state energy-density. This factor is clearly cancelled by $\mathcal{N}$. All the remaining divergences are associated with the continuum limit $a \to 0$. These are the divergences that form the subject of renormalization. They are called the ultra-violet (UV) divergences.

One notational change needs to be made now. In more complicated theories, there will be several fields, and the functional-integral solution of such a theory involves an integral over the values of a classical field for each quantum field. It is convenient to have a symbol for each classical field that is clearly related to the corresponding quantum field. The standard notation is to use the same symbol. Thus we change the integration variable in (2.2.2) from $A(x)$ to $\phi(x)$, with the result that

$$\langle 0|T\phi(x_1)\ldots\phi(x_N)|0\rangle = \mathcal{N}\int [d\phi] e^{iS[\phi]}\phi(x_1)\ldots\phi(x_N). \tag{2.2.11}$$

This is somewhat of an abuse of notation. However, it is usually obvious whether one is using ϕ to mean the quantum field, as on the left-hand side, or to mean the corresponding classical field, as on the right-hand side.

2.3 Renormalization

The difficult limit is the continuum limit $a \to 0$. There are divergences in this limit; this has been known from the earliest days of quantum elec-

trodynamics (e.g., Oppenheimer (1930)). It is possible to say that the UV divergences mean that the theory makes no physical sense, and that the subject of interacting quantum field theories is full of nonsense (Dirac (1981)). Luckily we can do better, for our ultimate aim need not be to construct a field theory literally satisfying (2.1.2)–(2.1.5). Rather, our aim is to construct a relativistic quantum theory with a local field as its basic observable. These requirements are satisfied if we construct a collection of Green's functions satisfying sensible physical properties (for example, as formulated in the Osterwalder–Schrader axioms – see Glimm & Jaffe (1981)). We may further ask that we find a theory that is close in some sense to satisfying the defining equations (2.1.2)–(2.1.5). Combining the functional integral with suitable renormalizations of the parameters of the theory satisfies these requirements.

The basic idea of renormalization comes from the observation that in one-loop graphs the divergences amount to shifts in the parameters of the action. For example, they change the mass of the particles described by $\phi(x)$ from the value m to some other effective value, which is infinite if m is finite. Renormalization is then the procedure of cancelling the divergences by adjusting the parameters in the action. To be precise, let us consider the ϕ^4 theory with

$$\mathscr{L} = (\partial A_0)^2/2 - m_0^2 A_0^2/2 - g A_0^4/4! + \Lambda_0. \tag{2.3.1}$$

The subscript zero is here used to indicate so-called bare quantities, i.e., those that appear in the Lagrangian when the $(\partial A_0)^2/2$ term has unit coefficient. (We also introduce a constant term. It will be used to cancel a UV divergence in the energy density of the vacuum.) Then we rescale the field by writing

$$A_0 = Z^{1/2} A, \tag{2.3.2}$$

so that, in terms of the 'renormalized field' A, the Lagrangian is

$$\begin{aligned} \mathscr{L} &= Z\partial A^2/2 - m_0^2 Z A^2/2 - g_0 Z^2 A^4/4! \\ &= Z\partial A^2/2 - m_{\mathrm{B}}^2 A^2/2 - g_{\mathrm{B}} A^4/4!. \end{aligned} \tag{2.3.3}$$

We have dropped Λ_0 from $\mathscr{L}$ since it has no effect on the Green's functions.

The Green's functions of the quantum field ϕ are now obtained by using (2.3.3) as the Lagrangian in the functional integral (2.2.2). We let Z, m_0, and g_0 be functions of the lattice spacing a, and we choose these functions (if possible) so that the Green's functions of ϕ are finite as $a \to 0$. If this can be done, then we have succeeded in constructing a continuum field theory, and it is termed 'renormalizable'. The theory may be considered close to solving

(2.1.2) (2.1.5). This is because the theory is obtained by taking a discrete (i.e., lattice) version of the equations and then taking a somewhat odd continuum limit.

We will call m_0 the bare mass, and g_0 the bare coupling, and we will call Z the wave-function, or field-strength, renormalization. It is also common to call m_B and g_B the bare mass and coupling; but for the sake of consistency we will not do this in this book.

Another way of viewing the renormalization is to write (2.3.3) as

$$\begin{aligned}\mathscr{L} &= \partial A^2/2 - m^2 A^2/2 - gA^4/4! \\ &\quad + \delta Z \partial A^2/2 - \delta m^2 A^2/2 - \delta g A^4/4!. \end{aligned} \tag{2.3.4}$$

We will call the first three terms the basic Lagrangian and the last three the counterterm Lagrangian. The renormalized mass m and the renormalized coupling g are finite quantities held fixed as $a \to 0$. The counterterms $\delta Z = Z - 1$, $\delta m^2 = m_B^2 - m^2$, and $\delta g = g_B - g$ are adjusted to cancel the divergences as $a \to 0$. This form of the Lagrangian is useful in doing perturbation theory; we treat $\partial A^2/2 - m^2 A^2/2$ as the free Lagrangian and the remainder as interaction. The expansion is in powers of the renormalized coupling g. The counterterms are expanded in infinite series, each term cancelling the divergences of one specific graph.

The form (2.3.4) for $\mathscr{L}$ also exhibits the fact that the theory has two independent parameters, m and g. The counterterms are functions of m, g, and of a.

We will discuss these issues in much greater depth in the succeeding chapters. For the moment it is important to grasp the basic ideas:

(1) The self-interactions of the field create, among other things, dynamical contributions to the mass of the particle, to the potential between particles, and to the coupling of the field to the single particle state. Thus the measured values of these parameters are renormalized relative to the values appearing in the Lagrangian.
(2) These contributions, or renormalizations, are infinite, in many cases. The most important theorem of renormalization theory is that they are the only infinities, in the class of theories called 'renormalizable'.
(3) The infinities are cancelled by wave-function, mass, and coupling counterterms, so that the net effect of the interactions is finite.
(4) To make quantitative the sizes of the infinities, the theory is constructed as the continuum limit of a lattice theory. The infinities appear as divergences when the lattice spacing goes to zero.

2.4 Ultra-violet regulators

In the last sections we showed how to construct field theories by defining the functional integral as the continuum limit of a lattice theory. Ultra-violet divergences appear as divergences when the lattice spacing, a, goes to zero, and are removed by renormalization counterterms. The lattice therefore is a regulator, or cut-off, for the UV divergences.

To be able to discuss the divergences quantitively and to construct a theory involving infinite renormalizations, it is necessary to use some kind of UV cut-off. Then the theory is obtained as an appropriate limit when the cut-off is removed. There are many possible ways of introducing a cut-off, of which going to a lattice is only one example. The lattice appears to be very natural when working with the functional integral. But it is cumbersome to use within perturbation theory, especially because of the loss of Poincaré invariance. There are two other very standard methods of making an ultra-violet cut-off: the Pauli–Villars method, and dimensional regularization.

The Pauli–Villars (1949) method is very traditional. In its simplest version it consists of replacing the free propagator $\mathrm{i}/(p^2 - m^2)$ in a scalar field theory by

$$\begin{aligned} S_F(p, m; M) &= \frac{\mathrm{i}}{p^2 - m^2} - \frac{\mathrm{i}}{p^2 - M^2} \\ &= \frac{\mathrm{i}}{(p^2 - m^2)} \frac{(m^2 - M^2)}{(p^2 - M^2)}. \end{aligned} \tag{2.4.1}$$

As $M \to \infty$, this approaches the original propagator. The behavior for large p has clearly been improved. Thus the degree of divergence of the Feynman graphs in the theory has been reduced. All graphs in the ϕ^4 theory, except for the one-loop self-energy are in fact made finite. In the ϕ^4 theory it is necessary to use a more general form in order to make all graphs finite:

$$\begin{aligned} S_F(p, m; M_1, M_2) = &\frac{\mathrm{i}}{(p^2 - m^2)} - \frac{\mathrm{i}}{(p^2 - M_1^2)} \frac{(m^2 - M_2^2)}{(M_1^2 - M_2^2)} \\ &- \frac{\mathrm{i}}{(p^2 - M_2^2)} \frac{(m^2 - M_1^2)}{(M_2^2 - M_1^2)} \\ = &\frac{\mathrm{i}}{(p^2 - m^2)} \frac{(M_1^2 - m^2)}{(p^2 - M_1^2)} \frac{(M_2^2 - m^2)}{(p^2 - M_2^2)}. \end{aligned} \tag{2.4.2}$$

It is usually convenient to set $M_1 = M_2$.

Now the regulated propagator has extra poles at $p^2 = M^2$, or at $p^2 = M_1^2$ and $p^2 = M_2^2$. Since one of the extra poles has a residue of the opposite sign to the pole at $p^2 = m^2$, the regulated theory cannot be completely physical.

It is normally true that a theory with an ultra-violet cut-off has some unphysical features.

Perhaps the most convenient regulator for practical calculations is dimensional regularization. There it is observed that the UV divergences are removed by going to a low enough space-time dimension d, so d is treated as a continuous variable. In perturbation theory this can be done consistently (Wilson (1973)), as we will see when we give a full treatment of dimensional regularization in Chapter 4. However it has not been possible to make it work non-perturbatively, so it cannot at present be regarded as a fundamental method.

Since it is only the renormalized theory with no cut-off that is of true interest, the precise method of cut-off is irrelevant. In fact, all methods of ultra-violet cut-off are equivalent, at least in perturbation theory. The differences are mainly a matter of practical convenience (or of personal taste). Thus dimensional regularization is very useful for perturbation theory. But the lattice method is maybe most powerful when working beyond perturbation theory; it is possible, for example, to compute the functional integral numerically by Monte-Carlo methods (Creutz (1980, 1983), and Creutz & Moriarty (1982)).

Within perturbation theory one need not even use a cut-off. Zimmermann (1970, 1973a) has shown how to apply the renormalization procedure to the integrands rather than to the integrals for Feynman graphs. The lack of fundamental dependence on the procedure of cut-off is thereby made manifest. The application of this procedure to gauge theories, especially, is regarded by most people as cumbersome.

2.5 Equations of motion for Green's functions

We have defined a collection of Green's functions by the functional integral (2.2.2). (Implicit in the definition are a certain number of limiting procedures, as listed below (2.2.6).) This definition we will take as the basis for the rest of our work. First we must check that it in fact gives a solution of the theory. This means, in particular, that we are to derive the equations of motion (2.1.10) for the Green's functions, thus ensuring that both the operator equation of motion (2.1.2) and the commutation relations (2.1.5) hold. (For the remainder of this chapter we will not specify the details of how renormalization affects these results.)

It is convenient to work with the generating functional (2.2.5). We make the change of variable $A(x) \to A(x) + \varepsilon f(x)$, where ε is a small number, and $f(x)$ is an arbitrary function of x^μ. Since the integration measure is invariant

under this shift, the value of the integral is unchanged:

$$\int [\mathrm{d}A] \exp\left\{ \mathrm{i}S[A+\varepsilon f] + \int (A+\varepsilon f)J \right\} = \int [\mathrm{d}A] \exp\left\{ \mathrm{i}S[A] + \int AJ \right\}. \tag{2.5.1}$$

Picking out the terms of order ε gives

$$\int \mathrm{d}^4 y f(y) \int [\mathrm{d}A] \exp\left\{ \mathrm{i}S[A] + \int AJ \right\} \left[\mathrm{i}\frac{\delta S}{\delta A(y)} + J(y) \right] = 0, \tag{2.5.2}$$

where, as usual, we define the functional derivative

$$\frac{\delta S}{\delta A(y)} = -\Box A - \frac{\mathrm{d}P}{\mathrm{d}A}. \tag{2.5.3}$$

Since $f(y)$ is arbitrary, we get

$$\int [\mathrm{d}A] \exp\left\{ \mathrm{i}S[A] + \int AJ \right\} \left[\mathrm{i}\frac{\delta S}{\delta A(y)} + J(y) \right] = 0. \tag{2.5.4}$$

Functionally differentiating N times with respect to J, followed by setting $J = 0$, gives the equation of motion (2.1.10). For example,

$$\begin{aligned}
0 &= \mathcal{N} \frac{\delta}{\delta J(x)} [\text{left-hand side of (2.5.4)}]_{J=0} \\
&= \mathcal{N} \int [\mathrm{d}A] \mathrm{e}^{\mathrm{i}S[A]} \left[A(x) \mathrm{i} \frac{\delta S}{\delta A(y)} + \delta^{(4)}(x-y) \right] \\
&= \mathcal{N} \int [\mathrm{d}A] \mathrm{e}^{\mathrm{i}S[A]} \{ \mathrm{i}A(x)[-\Box A(y) - P'(A(y))] + \delta^{(4)}(x-y) \} \\
&= -\mathcal{N} \mathrm{i} \Box_y \int [\mathrm{d}A] \mathrm{e}^{\mathrm{i}S[A]} A(x) A(y) \\
&\quad - \mathrm{i}\mathcal{N} \int [\mathrm{d}A] \mathrm{e}^{\mathrm{i}S[A]} A(x) P'(A(y)) + \delta^{(4)}(x-y) \\
&= -\mathrm{i}\Box_y \langle 0 | T\phi(x)\phi(y) | 0 \rangle - \mathrm{i} \langle 0 | T\phi(x) P'(\phi(y)) | 0 \rangle + \delta^{(4)}(x-y) \\
&= \mathrm{i} \langle 0 | T\phi(x) \frac{\delta S}{\delta \phi(y)} | 0 \rangle + \delta^{(4)}(x-y),
\end{aligned} \tag{2.5.5}$$

which is equivalent to (2.1.9). Note that in the fourth line we have exchanged the order of integration and of differentiation for the $\Box_y$ term. We have also used the normalization condition (2.2.4). It is important that the derivative of the quantum field (next-to-last line) is *outside* the time ordering, and $\delta S/\delta\phi(y)$ in the last line is defined to be a shorthand for the combination of operators in the previous line.

This is somewhat paradoxical since we have the operator equation of motion:

$$0 = \frac{\delta S}{\delta \phi} = -\Box\phi - P'(\phi),$$

from which it is tempting to deduce that the Green's function $\langle 0|T\phi(x)\delta S/\delta\phi(y)|0\rangle$ should be zero. However, in view of the work above it is convenient to define this Green's function by the functional-integral formula

$$\langle 0|T\phi\frac{\delta S}{\delta\phi(y)}|0\rangle = \mathcal{N}\int[\mathrm{d}A]\mathrm{e}^{\mathrm{i}S[A]}A(x)\frac{\delta S}{\delta A(y)}.$$

Then, as we have seen, the $\Box_y$ is implicitly outside the time-ordering. Bringing it inside the time-ordering gives a commutator, so that we get the δ-function term in (2.1.9) or (2.5.5).

The momentum-space version of the equation of motion (2.1.10) is often useful. We define the momentum-space Green's functions

$$\begin{aligned}\bar{G}_N(p_1,\ldots,p_N) &= \int \mathrm{d}^4x_1\ldots\mathrm{d}^4x_N \exp\{\mathrm{i}(p_1\cdot x_1+\cdots+p_N\cdot x_N)\}G_N(x_1,\ldots,x_N)\\ &= \bar{G}_N(p_1,\ldots,p_N)(2\pi)^4\delta^{(4)}(p_1+\cdots+p_N).\end{aligned} \tag{2.5.6}$$

The momenta p_j are to be regarded as flowing out of the Green's functions. Translation invariance of the theory implies the δ-function for momentum conservation that is explicitly factored out in the last line of (2.5.6). A convenient notation (which we will use often) is to write

$$\tilde{G}_N(p_1,\ldots,p_N) = \langle 0|T\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle. \tag{2.5.7}$$

Implicit in this formula is the definition that the integrals over x defining the momentum-space field $\tilde{\phi}(p)$ are all taken outside the time-ordering, as stated in (2.5.6). We will use a tilde over the symbol for any function to indicate the Fourier-transformed function.

Fourier transformation of the equation of motion (2.1.10) gives

$$\begin{aligned}&-q^2\langle 0|T\tilde{\phi}(q)\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle\\ &\quad+\langle 0|TP'(\tilde{\phi})(q)\tilde{\phi}(p_i)\ldots\tilde{\phi}(p_N)|0\rangle\\ &=-\mathrm{i}\sum_{j=1}^{N}\langle 0|T\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_{j-1})\tilde{\phi}(p_{j+1})\ldots\tilde{\phi}(p_N)|0\rangle(2\pi)^4\delta^{(4)}(q+p_j).\end{aligned} \tag{2.5.8}$$

2.6 Symmetries

We now turn to the consequences of symmetries. As we will see, there are many interesting problems in renormalization theory that stem from the

following question: If a classical field theory has certain symmetries, does the symmetry survive after quantization? Generally, it is the need for renormalization of the theory that makes this a non-trivial question.

The symmetry properties are expressed in terms of Green's functions by the Ward identities. (Historically the earliest example was found by Ward (1950) in QED.) If the symmetry is not preserved by quantization there are extra terms called anomalies. In many cases there are no anomalies, so we will derive the Ward identities in this section ignoring the subtleties that in some cases lead to anomalies. Discussion of anomalous cases is given in Chapter 13.

Consider a theory of N fields which we collectively denote by a vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$. Our discussion is general enough to include the case of fields with spin. We consider a symmetry group of the action $S[\boldsymbol{\phi}]$. This is a group of transformations on the classical fields

$$\boldsymbol{\phi} \to \mathbf{F}[\boldsymbol{\phi}; \boldsymbol{\omega}] \equiv \boldsymbol{\phi}', \tag{2.6.1}$$

which leaves the action invariant:

$$S[\boldsymbol{\phi}'] = S[\boldsymbol{\phi}]. \tag{2.6.2}$$

Here $\boldsymbol{\omega} = (\omega^\alpha)$ is a set of parameters of the group, which we assume here to be a Lie group, i.e., the ω's take on a continuous set of values. We let $\boldsymbol{\omega} = \mathbf{0}$ be the identity: $\mathbf{F}[\boldsymbol{\phi}; \mathbf{0}] = \boldsymbol{\phi}$. It is easiest to work with infinitesimal transformations:

$$\delta\phi_i \equiv \omega^\alpha \frac{\partial F_i}{\partial \omega^\alpha}[\boldsymbol{\phi}; \boldsymbol{\omega}]\bigg|_{\omega = 0} \equiv \omega^\alpha \delta_\alpha \phi_i. \tag{2.6.3}$$

(A summation convention on α is understood.)

In the quantum theory the symmetry is implemented as a unitary representation $U(\boldsymbol{\omega})$ of the group on the Hilbert space of states such that

$$U(\boldsymbol{\omega})\boldsymbol{\phi}U(\boldsymbol{\omega})^{-1} = F[\boldsymbol{\phi}; \boldsymbol{\omega}]. \tag{2.6.4}$$

Since the representation is unitary, we may parametrize the group so that

$$U(\boldsymbol{\omega}) = \exp(\mathrm{i}\omega^\alpha Q_\alpha), \tag{2.6.5}$$

where the generators Q_α are hermitian operators which represent the Lie algebra of the group:

$$[Q_\alpha, Q_\beta] = \mathrm{i}c_{\alpha\beta\gamma} Q_\gamma. \tag{2.6.6}$$

The normalizations are such that the structure constants $c_{\alpha\beta\gamma}$ are totally antisymmetric. The infinitesimal transformations are then given by:

$$\delta_\alpha \phi_i(x) = \mathrm{i}[Q_\alpha, \phi_i(x)]. \tag{2.6.7}$$

There are a number of special cases, each with its own special features:

(1) Global internal symmetry: A finite-dimensional Lie group acts on the fields at each point of space-time, with the same transformation at each point. Thus

$$\delta_\alpha \phi_i = -\mathrm{i}(t_\alpha)_i{}^j \phi_j, \tag{2.6.8}$$

where the t_α form a hermitian matrix representation of the Lie algebra:

$$[t_\alpha, t_\beta] = \mathrm{i} c_{\alpha\beta\gamma} t_\gamma. \tag{2.6.9}$$

Single-particle states carrying this representation are annihilated by ϕ_i. The Lagrangian is invariant.

(2) Global space-time symmetry: The group effectively is a transformation on space-time; the Poincaré group and its extensions are the usual cases. For a Poincaré transformation $x^\mu \to \Lambda^\mu_\nu x^\nu + a^\mu$, we have the corresponding transformation of the fields:

$$\phi_i(x) \to \phi_j(\Lambda x + a) R(\Lambda)^j{}_i. \tag{2.6.10}$$

Here R is a finite-dimensional matrix representation of the Lorentz group (never unitary if non-trivial), acting on the spin indices of $\boldsymbol{\phi}$. The Lagrangian is not invariant. It transforms as

$$\mathscr{L}[\boldsymbol{\phi}, x] \to \mathscr{L}[\phi, \Lambda x + a],$$

so that the action $S = \int \mathrm{d}^4 x \mathscr{L}$ is invariant. Infinitesimal transformations of ϕ involve the derivative of ϕ.

(3) Global chiral symmetry: This looks like a global internal symmetry but acts differently on the left- and right-handed parts of Dirac fields (which we have yet to discuss). Anomalies are often present – see Chapter 13.

(4) Supersymmetry: This is a generalized type of symmetry where Bose and Fermi fields are related (Fayet & Ferrara (1977)). The only case that we will discuss is the BRS-invariance (Becchi, Rouet & Stora (1975)) of a gauge theory.

(5) Gauge, or local, symmetry: Any of the above symmetries may be extended to a symmetry whose parameters depend on x: $\omega = \omega(x)$. In quantum theories, these are not really implemented by unitary transformations. Their treatment is rather special. The elementary examples are general coordinate invariance in General Relativity, and gauge invariance in electromagnetism.

The basic tool for discussing symmetries is Noether's theorem, which relates them to conservation laws. This theorem in its most straightforward form applies only to symmetries of the first three types.

For a global symmetry, Noether's theorem asserts that a conserved current j_α^μ exists for each generator of a symmetry. Let the Lagrangian have the infinitesimal transformation

$$\mathscr{L} \to \mathscr{L} + \omega^\alpha \delta_\alpha \mathscr{L} = \mathscr{L} + \omega^\alpha \partial_\mu Y_\alpha^\mu, \tag{2.6.11}$$

so that the action is invariant. Define

$$j_\alpha^\mu = \sum_i \delta_\alpha \phi_i \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_i)} - Y_\alpha^\mu. \tag{2.6.12}$$

Then the equations of motion imply conservation of j_α^μ, i.e.,

$$\partial j_\alpha^\mu / \partial x^\mu = 0. \tag{2.6.13}$$

The generators of the symmetry group are

$$Q_\alpha = \int \mathrm{d}^3 x j_\alpha^0. \tag{2.6.14}$$

The canonical commutation relations imply that

$$\begin{aligned} [j_\alpha^0(x), \phi_i(y)] &= -\mathrm{i}\delta_\alpha \phi_i(x) \delta^{(3)}(x-y), \quad (\text{if } x^0 = y^0), \\ [Q_\alpha, \phi_i(y)] &= -\mathrm{i}\delta_\alpha \phi_i(y), \end{aligned} \tag{2.6.15}$$

as required by (2.6.7).

We need to consider not only transformations that are symmetries of the quantum theory, but also 'broken symmetries'. There are several cases (not mutually exclusive). Let us define them, since there is a certain amount of confusion in the literature about the terminology:

(1) Explicit breaking: The classical action has a non-invariant term. If $\delta_\alpha \mathscr{L} = \partial_\mu Y_\alpha^\mu + \Delta_\alpha$ then the Noether currents are not conserved: $\partial_\mu j_\alpha^\mu = \Delta_\alpha$. An important case is where this term is small, so that it can be treated as a perturbation.

(2) Anomalous breaking: Even though the classical action is invariant, the quantum theory is not, and there is no conserved current. The classical action is important for the quantum theory, since it appears in the functional integral defining the theory. The cause of anomalous breaking is generally an ultra-violet problem: $\partial_\mu j_\alpha^\mu \neq 0$ in the UV cut-off theory, and the non-conservation does not disappear when the cut-off is removed. (Cases are conformal transformations and some chiral theories.)

(3) Spontaneous breaking: The action is invariant and the currents are conserved (in the quantum theory), but the vacuum is not invariant under the transformations.

Whether or not a symmetry is broken either spontaneously or anomalously is a dynamical question. That is, one must solve the theory, at least

partially, to find the answer. Frequently, perturbation theory is adequate to do this and lowest order or next-to-lowest order calculations suffice. Renormalization is an integral part of treating anomalous breaking (see Chapter 13), while renormalization-group methods are sometimes necessary in treating spontaneously broken symmetry (Coleman & Weinberg (1973)).

The case of spontaneous symmetry breaking that is not visible in perturbation theory is often termed dynamical (Jackiw & Johnson (1973), Cornwall & Norton (1973), and Gross (1976)). Anomalous breaking is sometimes called spontaneous, but this is a bad terminology, because it gives two very different phenomena the same name.

2.7 Ward identities

Ward identities express in terms of Green's functions the consequences of a symmetry (whether or not it is broken). One derivation applies the equation of motion (2.1.10) to the divergence of a Green's function of the current j^μ_α. There are two terms: one in which the current is differentiated, and one in which the θ-functions defining the time-ordered product are differentiated. Thus a Ward identity expresses not only conservation of its current but also the commutation relation (2.6.15), which is equivalent to the transformation law. The Ward identities are central to a discussion of the renormalization of a theory with symmetries, expecially if spontaneously broken.

Our derivation of Ward identities begins by making the following change of variable:

$$A_i(x) \to A_i(x) + f^\alpha(x)\delta_\alpha A_i(x) \tag{2.7.1}$$

in the functional integral for the generating functional $Z[\mathbf{J}]$. Here $\delta_\alpha A_i$ is, as before, the variation of the field A_i under a symmetry transformation, and $f^\alpha(x)$ is a set of arbitrary complex-valued functions that vanish rapidly as $x \to \infty$. We get

$$Z[\mathbf{J}] = \int [\mathrm{d}\mathbf{A}] \exp\{\mathrm{i}S[\mathbf{A} + f^\alpha \delta_\alpha \mathbf{A}] + J^i(A_i + f^\alpha \delta_\alpha A_i)\}. \tag{2.7.2}$$

(Here we assumed that the measure is invariant under the change of variables (2.7.1).) The terms in (2.7.2) that are linear in f^α give

$$\begin{aligned} 0 &= \int [\mathrm{d}\mathbf{A}] \exp\left(\mathrm{i}S + \int \mathbf{J}\cdot\mathbf{A}\right)\{-\delta S[\mathbf{A} + f^\alpha \delta_\alpha \mathbf{A}]/\delta f^\alpha(y) + \mathrm{i}J^i \delta_\alpha A_i(y)\} \\ &= \int [\mathrm{d}\mathbf{A}] \exp\left(\mathrm{i}S + \int \mathbf{J}\cdot\mathbf{A}\right)\{\partial_\mu j^\mu_{\mathrm{cl},\alpha}(y) + \mathrm{i}J^i \delta_\alpha A_i(y)\}. \end{aligned} \tag{2.7.3}$$

Here $j^\mu_{\mathrm{cl},\alpha}$ is the Noether current in the classical theory.

The Ward identities follow by functionally differentiating with respect to the sources $J(x)$. Thus one differentiation gives

$$\frac{\partial}{\partial y^\mu}\langle 0|Tj_\alpha^\mu(y)\phi_i(x)|0\rangle = -i\delta^{(4)}(x-y)\langle 0|\delta_\alpha\phi_i(y)|0\rangle, \tag{2.7.4}$$

while a double differentiation gives

$$\begin{aligned}\frac{\partial}{\partial y^\mu}&\langle 0|Tj_\alpha^\mu(y)\phi_i(w)\phi_j(x)|0\rangle \\ &= -i\delta^{(4)}(w-y)\langle 0|T\delta_\alpha\phi_i(y)\phi_j(x)|0\rangle \\ &\quad -i\delta^{(4)}(x-y)\langle 0|T\phi_i(w)\delta_\alpha\phi_j(y)|0\rangle.\end{aligned} \tag{2.7.5}$$

Note that, just as in our derivation of the equation of motion for Green's functions in Section 2.5, the derivative $\partial/\partial y^\mu$ is outside the time-ordering. The general case is:

$$\begin{aligned}\frac{\partial}{\partial y^\mu}&\langle 0|Tj_\alpha^\mu(y)\prod_{i=1}^{N}\phi_{n_i}|x_i)|0\rangle \\ &= -i\sum_{j=1}^{N}\delta^{(4)}(y-x_j)\langle 0|T\delta_\alpha\phi_{n_j}(y)\prod_{i\neq j}\phi_{n_i}(x_i)|0\rangle.\end{aligned} \tag{2.7.6}$$

Important consequences of these Ward identities are obtained by integrating over all $\vec{y}$ (with y^0 fixed). The spatial derivatives give a surface term, which vanishes, so that we have, for example,

$$\int d^3y\frac{\partial}{\partial y^0}\langle 0|Tj_\alpha^0(y)\phi_i(x)|0\rangle = -i\delta(x^0-y^0)\langle 0|\delta_\alpha\phi_i(x)|0\rangle.$$

The spatial integral of j^0 is just the charge Q^0. The time derivative acts either on the charge or on the δ-functions defining the time-ordering; so we find that

$$\begin{aligned}&\langle 0|T\,dQ_\alpha/dt\,\phi_i(x)|0\rangle + \langle 0|[Q_\alpha,\phi_i(x)]|0\rangle\delta(x^0-y^0) \\ &\quad = -i\delta(x^0-y^0)\langle 0|\delta_\alpha\phi_i(y)|0\rangle.\end{aligned} \tag{2.7.7}$$

In this equation and its generalizations from (2.7.6), we may choose the times of the fields $\phi_i(x_i)$ not to coincide with y^0. Therefore an arbitrary Green's function of dQ_α/dt is zero, so that the operator dQ_α/dt is zero. The remaining part of (2.7.7) therefore gives:

$$\langle 0|[Q_\alpha,\phi_i(x)]|0\rangle = -i\langle 0|\delta_\alpha\phi_i(y)|0\rangle. \tag{2.7.8}$$

From (2.7.8) and its generalizations with more fields, we find that the Q_α's have the correct commutation relations with the elementary fields ϕ_i to be the generators of the symmetry group.

Finally, another specialization of (2.7.6) is to integrate it over all y^μ and to

drop the resulting surface term. The result is that

$$0 = \sum_{j=1}^{N} \langle 0|T\delta_\alpha \phi_{n_j}(x_j) \prod_{i \neq j} \phi_{n_i}(x_i)|0\rangle$$

$$= \delta_\alpha \langle 0|T \prod_{i=1}^{N} \phi_{n_i}(x_i)|0\rangle. \tag{2.7.9}$$

All the above equations are true for the case of a completely unbroken symmetry. The derivation breaks down at the first step if we have anomalous breaking. (In Chapter 13 we will discuss the anomaly terms that must then be inserted in the Ward identities to make them correct.) For an explicitly broken symmetry, where $\partial \cdot j = \Delta \neq 0$, we must add a term

$$\langle 0|T\Delta_\alpha(y) \prod_{i=1}^{N} \phi_{n_i}(x_i)|0\rangle \tag{2.7.10}$$

to the right-hand side of (2.7.6).

In the case of a spontaneously broken theory the basic Ward identities (2.7.4)–(2.7.6) remain true – we still have an exact symmetry. But the integrated Ward identities (2.7.7) and (2.7.9) are no longer true. Equation (2.7.9) must be false if the vacuum is not invariant, and the derivation fails because the surface term is not zero. This is caused by the existence of zero-mass particles. These Nambu–Goldstone bosons (Goldstone, Salam & Weinberg (1962)) are characteristic of theories with a spontaneously broken symmetry.

2.8 Perturbation theory

As an example, consider again ϕ^4 theory, with classical Lagrangian

$$\mathscr{L} = Z(\partial A)^2/2 - m_B^2 A^2/2 - g_B A^4/4!. \tag{2.8.1}$$

We will expand the Green's functions in powers of the renormalized coupling g, for small g. To expand the functional-integral formula (2.2.2) in powers of g, we write

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1, \tag{2.8.2}$$

where $\mathscr{L}_0$ is the free Lagrangian:

$$\mathscr{L}_0 = (\partial A)^2/2 - m^2 A^2/2, \tag{2.8.3}$$

and $\mathscr{L}_I$ is the interaction Lagrangian:

$$\mathscr{L}_I = -gA^4/4! + (Z-1)(\partial A)^2/2 - (m_B^2 - m^2)A^2/2 - (g_B - g)A^4/4!. \tag{2.8.4}$$

We will expand the renormalization counterterms, $Z-1$, $m_B^2 - m^2$, and

$g_B - g$, in powers of g, so that all of the terms in $\mathscr{L}_1$ have at least one power of g. The series expansion of the Green's functions is then obtained from (2.2.2) as:

$$G_N(x_1, x_2, \ldots, x_N) = \frac{\sum_{n=0}^{\infty} (i^n/n!) \int [dA] A(x_1) \cdots A(x_N) \left[\int d^4y \mathscr{L}_I(y) \right]^n \exp(iS_0[A])}{\sum_{n=0}^{\infty} (i^n/n!) \int [dA] \left[\int d^4y \mathscr{L}_I(y) \right]^n \exp(iS_0[A])}. \quad (2.8.5)$$

Here

$$S_0[A] = \int d^4y \mathscr{L}_0 = \int d^4y (\partial A^2/2 - m^2 A^2/2)$$

is the free action.

Each of the terms in the series is a Green's function in the free-field theory (aside from a common normalization), so (2.8.5) is equivalent to the Gell-Mann–Low (1951) formula:

$$G_N(x_1, \ldots, x_N) = \frac{\sum_{n=0}^{\infty} (i^n/n!) \left(\prod_{j=1}^{n} \int d^4y_j \right) \langle 0 | T \phi_F(x_1) \ldots \phi_F(x_N) \prod_{j=1}^{n} \mathscr{L}_I(y_i) | 0 \rangle}{\sum_{n=0}^{\infty} (i^n/n!) \left(\prod_{j=1}^{n} \int d^4y_j \right) \langle 0 | T \prod_{j=1}^{n} \mathscr{L}_I(y_j) | 0 \rangle}. \quad (2.8.6)$$

Here ϕ_F is a free quantum field of mass m. It is the field generated from the free Lagrangian $\mathscr{L}_0 = (\partial \phi_F)^2/2 - m^2 \phi_F^2/2$. Then $\mathscr{L}_I$ is the quantum interaction Lagrangian, $\mathscr{L} - \mathscr{L}_0$, which is a function of the free field ϕ_F.

To compute the integrals in (2.8.5) it suffices to compute the generating functional of free-field Green's functions:

$$Z_0[J] \equiv \frac{\int [dA] \exp\left(iS_0[A] + \int JA \right)}{\int [dA] \exp(iS_0[A])}. \quad (2.8.7)$$

This is done by completing the square, i.e., by making the following change of variable:

$$A(x) \to A(x) + \int d^4y G_F(x - y) J(y). \quad (2.8.8)$$

Here, $G_F(x)$ is the Feynman propagator satisfying

$$(\Box + m^2) G_F(x) = -i\delta^{(4)}(x), \quad (2.8.9)$$

and a boundary condition that, after rotating to Euclidean space by $x^0 = -i\tau$, $G_F(x) \to 0$ as $x \to \infty$. Thus

$$G_F(x) = \int \frac{d^4k}{(2\pi)^4} e^{ik\cdot x} \frac{i}{k^2 - m^2 + i\varepsilon}. \tag{2.8.10}$$

The result is that

$$Z_0[J] = \exp\left\{\tfrac{1}{2}\int d^4x\, d^4y\, J(x) G_F(x-y) J(y)\right\}. \tag{2.8.11}$$

Green's functions of free fields are obtained by differentiating with respect to J; for example

$$\begin{aligned}\langle 0|T\phi(x)\phi(y)|0\rangle &= \frac{\delta^2 Z}{\delta J(x)\delta J(y)}\bigg|_{J=0} \\ &= G_F(x-y).\end{aligned} \tag{2.8.12}$$

We can now derive the well-known Feynman rules for the interacting theory from (2.8.6). These can be given either in momentum or coordinate space. In either case the Green's function G_N is written as a sum over all possible topologically distinct Feynman graphs. Each graph Γ consists of a number of vertices joined by lines. It has N 'external vertices', one for each $\phi(x_i)$, with one line attached, and some number, n, of interaction vertices. The interaction vertices are of several types, corresponding to the terms in the interaction Lagrangian (2.8.4). The vertex for the A^4 interaction has four lines attached and the vertices for the ∂A^2 and A^2 interactions have two lines attached. The value of the graph, denoted $I(\Gamma)$, is the integral over the position y_j of the n interaction vertices. The integrand is a product of factors:

(1) $G_F(w-z)$ for each line, where w and z are the positions of the vertices at its end.
(2) A combinatorial factor $1/S(\Gamma)$.
(3) $-ig_B$ for each A^4 interaction.
(4) $-i(m_B^2 - m^2)$ for each A^2 interaction.
(5) $-i(Z-1)\partial^2/\partial w^\mu \partial w_\mu$ for each $(\partial A)^2$ interaction: the derivatives with respect to w act on one of the propagators attached to the vertex.

For each Feynman graph a number of equal contributions arise in expanding (2.8.5). If Γ has no symmetries and if it has no counterterm vertices, then this number is $n!(4!)^n$ so that the explicit $n!$ in (2.8.5) and the $4!$ in each interaction are cancelled. Graphs with symmetries have a number of contributions smaller by a factor of the symmetry number $S(\Gamma)$. (For example, the self-energy graph Fig. 2.8.1 has $S = 6$.) The combinatorial factor is then the inverse of $S(\Gamma)$.

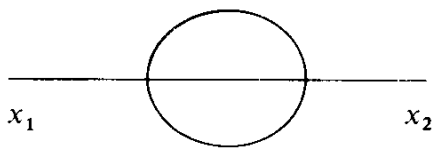


Fig. 2.8.1. A graph with symmetry factor $S = 6$.

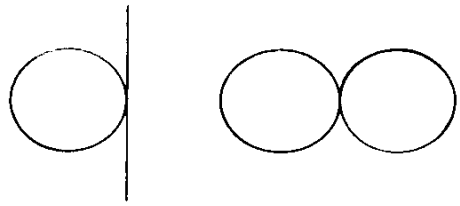


Fig. 2.8.2. A graph with a vacuum bubble.

The denominator of (2.8.5) is the sum of all graphs with no external lines. The result is to cancel all graphs in the numerator that have disconnected vacuum bubbles (like Fig. 2.8.2.).

In momentum space each line is assigned a (directed) momentum k. The Feynman rules are:

(1) A factor $i/[(2\pi)^4(k^2 - m^2 + i\varepsilon)]$ for a line with momentum k.
(2) A factor $(2\pi)^4$ times a momentum conservation δ-function for each vertex (external or interaction).
(3) An integral over the momentum of every line.
(4) A combinatorial factor $1/S(\Gamma)$.
(5) $-ig_B$ for each A^4 interaction.
(6) $-i(m_B^2 - m^2)$ for each A^2 interaction.
(7) $i(Z-1)p^2$ for each $(\partial A)^2$ interaction, where p is the momentum flowing on one of the propagators attached to the vertex.

The perturbation series in (2.8.5) need not be convergent, but only asymptotic. Let $G_{N,n}$ be the sum up to order g^n of the perturbation series for the Green's function G_N. Then it is asymptotic to G_N if for any n the error satisfies

$$|G_N - G_{N,n}| = O(g^{n+1}) \tag{2.8.13}$$

as $g \to 0$. In general, perturbation theory is asymptotic but not convergent. This is rigorously known (Glimm & Jaffe (1981)) for the ϕ^4 theory in the cases that the space-time dimension is $d = 0, 1, 2, 3$. ($d = 0$ is the case of the ordinary integral

$$\int dx \exp(-m^2x^2/2 - gx^4/4!),$$

while $d = 1$ is the quantum mechanics of the anharmonic oscillator.) Physically, the reason for non-convergence is that when $g < 0$ the energy is unbounded below and so the vacuum-state continued from $g > 0$ is unstable. (Dyson (1952) first observed this phenomenon in quantum electrodynamics.)

In later chapters we will assume (2.8.13). When we compute large-momentum behavior, it will be important to understand the maximum possible validity and accuracy of the calculations if the perturbation theory is asymptotic but not convergent.

2.9 Spontaneously broken symmetry

Consider the ϕ^4 interaction. If m^2 is positive and g is small, we have a theory of particles of mass m slightly perturbed by the interaction. This interaction is basically a repulsive δ-function potential, as can be seen by examining the Hamiltonian in the non-relativistic approximation. There is a symmetry $\phi \to -\phi$.

But if m^2 is negative this interpretation is incorrect. The true situation can be discovered by noticing that the functional integral (in Euclidean space) is dominated by classical fields with the lowest Euclidean action, which is

$$S_{\text{Eucl}}[A] = \int d^4x[-(\partial A)^2/2 + m^2A^2/2 + gA^4/4!]. \tag{2.9.1}$$

(Remember that $(\partial A)^2 = -(\partial A/\partial\tau)^2 - \vec{\nabla}A^2$ is negative.) If $m^2 > 0$, then the minimum action field is $A = 0$. But, if $m^2 < 0$, then there are two minima; these are constant fields with $P'(A) = 0$, i.e., $A = A_+ \equiv \sqrt{(-6m^2/g)}$ and $A = A_- \equiv -\sqrt{(-6m^2/g)}$.

We choose to impose the boundary condition $A(x) \to A_+$ as $x \to \infty$ in the functional integral. (The condition $A \to A_-$ gives equivalent physics, because of the $A \to -A$ symmetry of the action.) Then field configurations with A close to A_+ will dominate. We may understand this by observing that field configurations with large regions where A is not close to A_+ or A_- will give small contributions to the Euclidean functional integral (2.2.8) because their action S_{Eucl} is so big. Indeed, a constant field with A not equal to A_+ or A_- has infinitely more action than one with $A = A_+$ or A_-, and its contribution to the integral is zero. One's first inclination then is that the only configurations that contribute have $A \to A_+$ or $A \to A_-$ as $x \to \infty$. However, other configurations contribute, because there are many of them – one has to integrate over all possible fluctuations. However, one can argue – even rigorously (Glimm & Jaffe (1981) – that in general A will be close to A_+ or to A_-. A typical configuration of the classical field $A(x)$ will be close to one of these values over almost all of space-time.

Given our choice of boundary condition $A \to A_+$, even more is true: a typical configuration is close to $A = A_+$ almost everywhere, rather than to either A_- or A_+. The reason is that if it had a large region with $A(x)$ close to

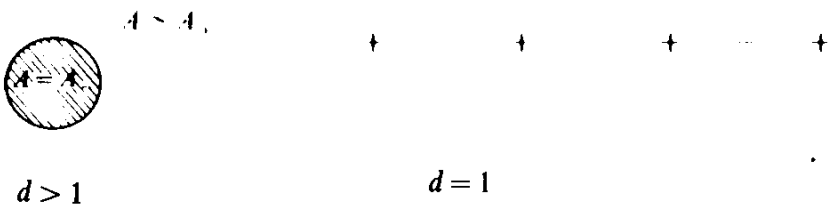


Fig. 2.9.1. Illustrating transitions between regions with fields close to different minima of the potential

A_- (Fig. 2.9.1), then there would be a contribution to the action proportional to the size of the boundary between the regions of positive $A(x)$ and of negative $A(x)$. Only if the space-time dimension is $d = 1$ will we have a finite contribution from the boundary. This special case is quantum mechanics of a particle in a potential with two wells. The particle can tunnel between the two wells.

In the case we have discussed, of a discrete rather than of a continuous symmetry, the argument that A is close to A_+ almost everywhere for the important configurations is correct in all space-time dimensions greater than one. The quantum field therefore has a vacuum expectation value close to A_+:

$$\langle 0|\phi(x)|0\rangle = \mathcal{N}\int [\mathrm{d}A]A(x)\mathrm{e}^{-S[A]} \simeq A_+ .$$

In the case of a continuous symmetry, there is a continuous series of minima of the potential. A field configuration can interpolate between different minima without going over a big hump in the potential. The only penalty comes from the gradient terms in the action. This suppresses configurations that do not stay close to one minimum, but only in more than two space-time dimensions. In one space-time dimension there is no spontaneous breaking of a continuous symmetry (Mermin & Wagner (1966), Hohenberg (1967), and Coleman (1973)).

Perturbation theory can be considered as a saddle point expansion about the minimum of the action. We write $A(x) = A'(x) + v$ where $v = A_+$. Now we treat $A'(x)$ as the independent variable. We have

$$\mathscr{L} = (\partial A')^2/2 - M^2A'^2/2 - gvA'^3/3! - gA'^4/4! + C. \qquad (2.9.2)$$

Here $C \equiv -m^2v^2/2 - gv^4/4!$, and $M^2 \equiv gv^2/2 + m^2 = -2m^2 > 0$. We now have a theory of particles of mass M with both an A'^3 and an A'^4 interaction. The symmetry is hidden; its only obvious manifestation is in the relation between the A'^3 coupling and the mass and A'^4 terms:

$$gv = M(3g)^{1/2}. \qquad (2.9.3)$$

We will show later that renormalization counterterms are correctly given by continuing in m^2 from positive m^2. The vacuum expectation value of ϕ has corrections which can be computed in perturbation theory

$$\begin{aligned}\langle 0|\phi|0\rangle &= v + \langle 0|\phi'|0\rangle \\ &= v + O(g^{1/2}).\end{aligned} \tag{2.9.4}$$

Exactly similar methods can be applied if there is a continuous symmetry. Then the Goldstone theorem tells us that there will be a massless scalar particle for each broken generator.

2.10 Fermions

The field theories obtained by functional integration as in Section 2.2 are all theories of bosons. This follows from the symmetry of the Green's functions under exchange of fields (e.g., $\langle 0|T\phi(x)\phi(y)|0\rangle = \langle 0|T\phi(y)\phi(x)|0\rangle$). In turn, this symmetry property follows from the functional-integral formula (2.2.2) because the integration variables (the values of the classical field $A(x)$) commute with each other.

To get a theory with quantized fields, it is necessary to define something like an integral over anticommuting variables. A rather small number of properties of integration are needed to derive the equations of motion for Green's functions. Requiring these properties determines the integration operation uniquely (Itzykson & Zuber (1980)).

As an example, consider the following Lagrangian for a free Dirac field:

$$\mathscr{L} = \bar{\psi}(i\partial\!\!\!/ - M)\psi. \tag{2.10.1}$$

Here ψ is a four-dimensional column vector and $\bar{\psi}$ a row vector, while $\partial\!\!\!/ = \gamma^\mu \partial_\mu$. The generating functional of Green's functions is written as:

$$Z[\eta, \bar{\eta}] = \mathscr{N} \int [d\psi\, d\bar{\psi}] \exp\left(i \int \mathscr{L} + \int \bar{\eta}\psi + \int \bar{\psi}\eta\right). \tag{2.10.2}$$

The fields and the sources $\eta(x)$ and $\bar{\eta}(x)$ take their values in the fermionic sector of a Grassmann algebra. In the lattice approximation the definition of the integration in (2.10.2) is really algebraic (Itzykson & Zuber (1980)). Green's functions are defined by differentiating with respect to the sources.

One important difference between ordinary integration and Grassmann integration will be important in treating gauge theories. The simplest case of this difference is in the integral over two variables x and $\bar{x}$ of $\exp(i\bar{x}ax)$, where a is a real number. For ordinary real variables the integral is

$$\int dx\, d\bar{x} e^{i\bar{x}ax} = 2\pi \int d\bar{x}\delta(\bar{x}a) = 2\pi/a. \tag{2.10.3}$$

For Grassmann variables we get

$$\int \mathrm{d}x\,\mathrm{d}\bar{x}\,\mathrm{e}^{\mathrm{i}\bar{x}ax} = \mathrm{i}a. \tag{2.10.4}$$

The overall normalization is irrelevant for our applications, for it is always cancelled by the overall normalization factor in the functional integral. What matters is that the a-dependence of (2.10.4) is inverse to that in (2.10.3).

2.11 Gauge theories

A gauge symmetry is an invariance under a group G where the group transformation is different at each space-time point. The earliest examples were General Relativity (where G is $GL(4)$, the group of linear transformations of the coordinate system), and electrodynamics (where G is the group of phase rotations). Yang & Mills (1954) and Shaw (1955) generalized the idea to a general group. Beg & Sirlin (1982) and Buras (1981) explain some of the uses of gauge theories as theories of physics.

Let G be a simple group and let a matter field ψ transform as

$$\psi(x) \to \exp(-\mathrm{i}g\omega^{\alpha}(x)t_{\alpha})\psi(x) = U(\omega(x))^{-1}\psi(x). \tag{2.11.1}$$

The field ψ is a column vector of components, and the hermitian matrices t_{α} form a representation of the group, with structure constants $c_{\alpha\beta\gamma}$ defined by

$$[t_{\alpha}, t_{\beta}] = \mathrm{i}c_{\alpha\beta\gamma}t_{\gamma}. \tag{2.11.2}$$

The matrices $U(\omega)$ form a representation of the group.

In order that the action be gauge invariant, we need a covariant derivative:

$$D_{\mu}\psi \equiv (\partial_{\mu} + \mathrm{i}g\mathsf{A}_{\mu})\psi. \tag{2.11.3}$$

Here we have introduced the gauge potential A_{μ}. It is a vector under Lorentz transformation. As far as its gauge symmetry properties are concerned, it can be written as a matrix A_{μ} or in terms of components A_{μ}^{α}:

$$\mathsf{A}_{\mu} = \sum_{\alpha} A_{\mu}^{\alpha} t_{\alpha}. \tag{2.11.4}$$

It transforms under the gauge group as:

$$\mathsf{A}_{\mu}(x) \to U(\omega(x)^{-1}[\mathsf{A}_{\mu}(x) - \mathrm{i}g^{-1}\partial_{\mu}]U(\omega(x)). \tag{2.11.5}$$

To build an action, we need the field-strength tensor

$$\begin{aligned} \mathsf{F}_{\mu\nu} &= \partial_{\mu}\mathsf{A}_{\nu} - \partial_{\nu}\mathsf{A}_{\mu} + \mathrm{i}g[\mathsf{A}_{\mu}, \mathsf{A}_{\nu}], \\ F_{\mu\nu}^{\alpha} &= \partial_{\mu}A_{\nu}^{\alpha} - \partial_{\nu}A_{\mu}^{\alpha} - gc_{\alpha\beta\gamma}A_{\mu}^{\beta}A_{\nu}^{\gamma}, \end{aligned} \tag{2.11.6}$$

which transforms as $\mathsf{F} \to U^{-1}\mathsf{F}U$.

A gauge-invariant Lagrangian with spin $\frac{1}{2}$ matter fields is

$$\begin{aligned}\mathscr{L}_{\text{inv}} &= -(F^{\alpha}_{\mu\nu})^2/4 + \bar{\psi}(\mathrm{i}\not{D} - M)\psi \\ &= -\operatorname{tr} \mathsf{F}_{\mu\nu}\mathsf{F}^{\mu\nu}/2 + \bar{\psi}(\mathrm{i}\not{D} - M)\psi,\end{aligned} \tag{2.11.7}$$

where we assumed the conventional normalization of the t_α's, viz., $\operatorname{tr} t_\alpha t_\beta = \delta_{\alpha\beta}/2$. In an exactly similar way, an action can be set up using scalar fields. If there are matter fields in several irreducible representations a term for each is needed in $\mathscr{L}$. The transformation (2.11.5) ensures that the coupling g is the same for all matter fields if the group is non-abelian.

The form of the infinitesimal transformations is needed:

$$\begin{aligned}\delta_\omega \psi &= -\mathrm{i}g\omega^\alpha t_\alpha \psi, \\ \delta_\omega \bar{\psi} &= \mathrm{i}g\omega^\alpha \bar{\psi} t_\alpha, \\ \delta_\omega A^\alpha_\mu &= \partial_\mu \omega^\alpha + g c_{\alpha\beta\gamma}\omega^\beta A^\gamma_\mu, \\ \delta_\omega F^\alpha_{\mu\nu} &= g c_{\alpha\beta\gamma}\omega^\beta F^\gamma_{\mu\nu}.\end{aligned} \tag{2.11.8}$$

If the group is not simple, then it is the product of several simple groups, e.g., $SU(2)\otimes SU(2)$. For each there is a gauge field and an independent coupling.

2.12 Quantizing gauge theories

A gauge theory such as the one defined by the Lagrangian (2.11.7) can be solved by the functional integral. Thus, as an example, we can write for the fermion propagator.

$$\langle 0|T\psi(x)\bar{\psi}(y)|0\rangle = \mathscr{N}_{GI}\int [\mathrm{d}A][\mathrm{d}\psi \mathrm{d}\bar{\psi}]\psi(x)\bar{\psi}(y)\exp\left(\mathrm{i}\int \mathscr{L}_{\text{inv}}\right). \tag{2.12.1}$$

In fact, a lattice approximation to the functional integral forms the basis of Monte-Carlo calculations (Creutz (1983) and Creutz & Moriarty (1982)). The only trouble with (2.12.1) is that it is exactly zero. To see this we observe that, given any field configuration, we can make a gauge transformation on it, as in (2.11.1) and (2.11.5). The new field configuration has the same action as the old field. Thus the only dependence on the gauge transformation is in the explicit $\psi(x)\bar{\psi}(y)$. Now the gauge transformation is independent at each space-time point. So (2.12.1) contains a factor

$$\begin{aligned}&\left(\prod_{\text{points } z}\int \mathrm{d}U(z)\right)U^{-1}(x)\otimes U(y) \\ &\quad = (\text{number})\left(\int \mathrm{d}U(x)U(x)^{-1}\right)\left(\int \mathrm{d}U(y)U(y)\right),\end{aligned} \tag{2.12.2}$$

which is zero. (Note that the propagator is a matrix in the representation space of the gauge group.)

The vanishing of (2.12.1) is not a fundamental problem, for we may choose only to work with Green's functions of gauge-invariant operators (e.g., $\bar{\psi}\psi$, $F^{\alpha}_{\mu\nu}F^{\alpha\mu\nu}$, the Wilson loop (Wilson (1974) and Kogut (1983))). But the vanishing is a disaster for formulating perturbation theory; for among the basic objects needed to write the Feynman rules are the propagators for the elementary fields. An elegant solution to this problem was given by Faddeev & Popov (1967). The integral over all gauge fields is written as the product of the integral over fields satisfying some given gauge condition (such as $\partial \cdot A^{\alpha} = 0$) and of the integral over all gauge transformations. Any field configuration can be obtained by gauge transforming some configuration that satisfies the gauge condition. For a gauge-invariant Green's function, the integral over gauge transformations amounts to an overall factor which cancels an inverse factor in the normalization. So the integral over gauge transformations can be consistently omitted.

The new integral over fields with the gauge-fixing condition imposed also provides a solution to the theory. But the gauge-variant Green's functions like (2.12.1) no longer vanish. It is necessary, moreover, to find the correct measure for the integral; this was the key point of the work of Faddeev and Popov.

These authors also constructed a slightly different formulation; it is this formulation that is most often used, and that we will review now. A detailed treatment and further references are to be found in Itzykson & Zuber (1980). Here we will merely summarize the argument and derive the Ward identities in the form that we will use them.

We will consider gauge conditions of the form $F_{\alpha}[A, x] = f_{\alpha}(x)$. There is one condition at each point of space-time and for each generator of the group. The functional F_{α} might be $\partial \cdot A^{\alpha}$, for example. The functions $f_{\alpha}(x)$ are any real valued functions of x.

Faddeev and Popov write an arbitrary Green's function as

$$\langle 0|TX|0\rangle = \mathcal{N}_{GI}\int [\mathrm{d}A][\mathrm{d}\psi][\mathrm{d}\bar{\psi}]X \exp(\mathrm{i}S_{\mathrm{inv}})\Delta[A]\prod_{x,\alpha}\delta(F_{\alpha}-f_{\alpha}). \tag{2.12.3}$$

Here S_{inv} is the gauge invariant action, and X is any product of fields. The factor $\Delta[A]$ is a Jacobian that arises in transforming variables to the set of fields that satisfy the gauge condition plus the set of gauge transformations. The key result is that $\Delta[A]$ is a determinant, so that it can be written as

$$\Delta = \int [\mathrm{d}c_{\alpha}][\mathrm{d}\bar{c}_{\alpha}]\exp(\mathrm{i}\mathscr{L}_{\mathrm{gc}}). \tag{2.12.4}$$

Here c_α and $\bar{c}_\alpha$ are anticommuting scalar fields, called the Faddeev Popov ghosts. The so-called gauge-compensating Lagrangian is

$$\mathscr{L}_{\rm gc} = \bar{c}^\alpha \delta_c F_\alpha[A, x], \tag{2.12.5}$$

where $\delta_c F_\alpha$ is the infinitesimal transformation of F_α with ω replaced by c. For the case $F_\alpha = \partial \cdot A^\alpha$

$$\mathscr{L}_{\rm gc} = \partial^\mu \bar{c}(\partial_\mu c_\alpha + g c_{\alpha\beta\gamma} c_\beta A_\mu^\gamma) + \text{divergence}. \tag{2.12.6}$$

We treat c and $\bar{c}$ as independent fields. They are not genuine physical fields, as they do not obey the usual spin-statistics theorem.

A convenient form of solution to the theory is obtained by averaging over all f_α's, with weight $\exp(-\xi^{-1}\int f_\alpha^2/2)$. This leaves gauge-invariant Green's functions unaltered, and gives the following formula:

$$\langle 0|TX|0\rangle = \mathscr{N}\int[\text{d fields}]X e^{iS} \tag{2.12.7}$$

with a different normalization. The integral is over all fields (A, ψ, $\bar{\psi}$, c, $\bar{c}$). The action S contains three terms:

$$S = \int d^4x(\mathscr{L}_{\rm inv} + \mathscr{L}_{\rm gf} + \mathscr{L}_{\rm gc}). \tag{2.12.8}$$

We have already defined the gauge-invariant Lagrangian by (2.11.7) and $\mathscr{L}_{\rm gc}$ by (2.12.5). The gauge-fixing term is

$$\mathscr{L}_{\rm gf} = -iF_\alpha^2/(2\xi), \tag{2.12.9}$$

where ξ is an arbitrarily chosen parameter. (If desired, it may be absorbed into a redefinition of F_α.)

The advantage of the form (2.12.8) is that Green's functions of the elementary fields are defined as in a simple non-gauge theory. For a gauge-invariant observable X the equations (2.12.3) and (2.12.7) define the same objects as

$$\langle 0|TX|0\rangle = \mathscr{N}_{GI}\int[dA d\psi d\bar{\psi}]X \exp(iS_{\rm inv}). \tag{2.12.10}$$

If X is gauge variant, then all the definitions give different results, and (2.12.7) depends on ξ. Quantities that depend on the choice of a gauge fixing are called gauge dependent, of course. We see that gauge invariance of the operators in a Green's function implies gauge independence.

It is important to distinguish the concepts of gauge invariance and gauge independence. Gauge invariance is a property of a classical quantity and is invariance under gauge transformations. Gauge independence is a property of a quantum quantity when quantization is done by fixing the gauge. It is

independence of the method of gauge fixing. Gauge invariance implies gauge independence, but only if the gauge fixing is done properly.

Gauge theories such as (2.11.7) have a dimensionless coupling if space-time is four dimensional. General results, which we will treat in later chapters, imply that the theories need renormalization. However these same results imply that many more counterterms *may* be needed than are available by renormalizing (2.11.7). In Chapter 12 we will prove that the extra couplings are absent. The tools needed are the Ward identities for the gauge symmetry. These we will prove in the next section. It is also necessary to prove that the unphysical degrees of freedom represented, for example, by the ghost fields c_α and $\bar{c}_\alpha$ do not enter unitarity relations. This proof also needs the gauge properties exhibited in the Ward identities (see Itzykson & Zuber (1980)).

2.13 BRS invariance and Slavnov–Taylor identities

After gauge fixing, the gauge invariance of a gauge theory is no longer manifest in the functional-integral solution. Slavnov (1972) and Taylor (1971) were the first to derive the generalized Ward identities that carry the consequences of gauge invariance. Their derivation was very much simplified by Becchi, Rouet & Stora (1975) through the discovery of what is now called the BRS symmetry of the action (2.12.8).

BRS symmetry is in fact a supersymmetry, that is, its transformations involve parameters that take their values in a Grassmann algebra. Let $\delta\lambda$ be a fermionic Grassmann variable. Then the BRS transformation of a matter or a gauge field is defined to be a gauge transformation with $\omega^\alpha = c^\alpha \delta\lambda$. Thus

$$\begin{aligned}
\delta_{\mathrm{BRS}}\psi &= -\mathrm{i}g(c^\alpha\delta\lambda)t_\alpha\psi = \mathrm{i}gt_\alpha c^\alpha\psi\,\delta\lambda,\\
\delta_{\mathrm{BRS}}\bar{\psi} &= \mathrm{i}g\bar{\psi}t_\alpha c^\alpha\delta\lambda,\\
\delta_{\mathrm{BRS}}A^\alpha_\mu &= (\partial_\mu c^\alpha + g c_{\alpha\beta\gamma}c^\beta A^\gamma_\mu)\delta\lambda.
\end{aligned} \tag{2.13.1}$$

Observe that $\delta\lambda$ is fermionic, so it anticommutes with fermion fields (c, $\bar{c}$, ψ, $\bar{\psi}$). The ghost fields transform as

$$\begin{aligned}
\delta_{\mathrm{BRS}}c^\alpha &= -\tfrac{1}{2}g c_{\alpha\beta\gamma}c^\beta c^\gamma\delta\lambda,\\
\delta_{\mathrm{BRS}}\bar{c}_\alpha &= F_\alpha\delta\lambda/\xi.
\end{aligned} \tag{2.13.2}$$

(Note that c^α and $\bar{c}_\alpha$ are not related by hermitian conjugation, contrary to appearances.) Since $\mathscr{L}_{\mathrm{inv}}$ is invariant under gauge transformations, it is BRS invariant. Hence

$$\begin{aligned}
\delta_{\mathrm{BRS}}\mathscr{L} &= \delta_{\mathrm{BRS}}(\mathscr{L}_{\mathrm{gf}} + \mathscr{L}_{\mathrm{gc}})\\
&= -(1/\xi)F_\alpha\delta_{c\delta\lambda}F_\alpha[A;x] - (1/\xi)F_\alpha\delta\lambda\delta_c F_\alpha - \bar{c}_\alpha\delta_{\mathrm{BRS}}(\delta_c F_\alpha)\\
&= 0.
\end{aligned} \tag{2.13.3}$$

In the second line we used $\mathscr{L}_{gc} = -c_\gamma \delta_{c\alpha} F^\gamma_{\alpha\beta}$, while to prove the last line zero we anticommuted $\delta\lambda$ and c in the first two terms of the second line. In addition we used the nilpotence of the BRS transformation:

$$\left(\frac{\delta_{\text{BRS}}}{\delta\lambda}\right)^2 (\psi \text{ or } \bar{\psi} \text{ or } A^\alpha_\mu \text{ or } c^\alpha) = 0,$$

$$\left(\frac{\delta_{\text{BRS}}}{\delta\lambda}\right)^3 \bar{c}_\alpha = 0, \tag{2.13.4}$$

which follows from anticommutativity of the c's.

By applying the Noether theorem we find a conserved current:

$$j^\mu_{\text{BRS}} = g\bar{\psi}\gamma^\mu t_\alpha \psi c^\alpha - F^{\alpha\mu\nu} D_\nu c^\alpha - (1/\xi)\partial\cdot A^\alpha D^\mu c^\alpha$$
$$-\tfrac{1}{2} g(\partial^\mu \bar{c}_\alpha) c^\beta c^\gamma c_{\alpha\beta\gamma}. \tag{2.13.5}$$

Although the BRS transformations involve Grassmann-valued parameters, the derivation of Ward identities given in Section 2.7 goes through unchanged. For our purposes, we only need the integrated Ward identity (2.7.9). A case of (2.7.9) applied to BRS invariance is called a Slavnov–Taylor identity. A simple example is

$$\begin{aligned} 0 &= \delta_{\text{BRS}} \langle 0 | T A^\alpha_\mu(x) c_\beta(y) | 0 \rangle / \delta\lambda \\ &= -\langle 0 | T(\partial_\mu c_\alpha + g c_{\alpha\delta\gamma} c^\delta A^\gamma) \bar{c}_\beta(y) | 0 \rangle \\ &\quad + (1/\xi) \langle 0 | T A^\alpha_\mu(x) \partial\cdot A^\beta(y) | 0 \rangle . \end{aligned} \tag{2.13.6}$$

We have defined the notation δ_{BRS} (quantity)$/\delta\lambda$ to mean that the $\delta\lambda$ in the BRS variation is commuted or anticommuted to the right and then deleted.

The most used cases of the Slavnov–Taylor identities are:

$$\begin{aligned} 0 &= \delta_{\text{BRS}} \langle 0 | T X \bar{c}_\alpha(x) | 0 \rangle / \delta\lambda \\ &= -\langle 0 | T(\delta_{\text{BRS}} X / \delta\lambda) \bar{c}_\alpha | 0 \rangle + (1/\xi) \langle 0 | T X \partial\cdot A^\alpha | 0 \rangle . \end{aligned} \tag{2.13.7}$$

Here X is a product of fields with total ghost number zero.

We will also need equations of motion. Let:

$$\begin{aligned} \mathscr{L}_{\bar{\psi}} &= \frac{\partial \mathscr{L}}{\partial \bar{\psi}} - \partial_\mu \frac{\partial \mathscr{L}}{\partial \partial_\mu \bar{\psi}} = (\mathrm{i}\not{D} - M)\psi, \\ \mathscr{L}_{\psi} &= \bar{\psi}(-\mathrm{i}\overleftarrow{\not{D}} - M), \\ \mathscr{L}_{A^\alpha_\mu} &= -D_\nu F^{\alpha\mu\nu} - g\bar{\psi}\gamma^\mu t_\alpha \psi + (1/\xi)\partial^\mu \partial\cdot A^\alpha + g c_{\alpha\beta\gamma}(\partial^\mu \bar{c}^\beta) c_\gamma, \\ \mathscr{L}_{\bar{c}_\alpha} &= -\partial_\mu D^\mu c^\alpha, \\ \mathscr{L}_{c^\alpha} &= -\Box \bar{c}_\alpha - g c_{\alpha\beta\gamma}(\partial_\mu \bar{c}_\beta) A^{\gamma\mu}. \end{aligned} \tag{2.13.8}$$

Then each of these is zero. Furthermore, for ϕ equal to any field in the theory we have:

$$\langle 0 | T \mathscr{L}_\phi(x) X | 0 \rangle = \mathrm{i} \langle 0 | T \delta X / \delta\phi(x) | 0 \rangle, \tag{2.13.9}$$

where time derivatives in $\mathscr{L}'_\phi$ are taken outside the time-ordering (as usual see Section 2.5).

2.14 Feynman rules for gauge theories

Feynman rules are given in Fig. 2.14.1 for the Lagrangian of (2.11.7), with the gauge-fixing term $F_\alpha = \partial \cdot A^\alpha$. Note that these agree with the figures but not the equations of Marciano & Pagels (1978). They are the rules for quantum chromodynamics (QCD), the theory of strong interactions, if the gauge group is $SU(3)$. The fermions are the quarks and consist of several triplets of $SU(3)$, each with its own mass term. (In the conventional terminology, the gauge field is called the gluon field and the gauge symmetry is called the color symmetry of strong interactions. Each irreducible representation in the quark field is called a flavor, and has a label: u, d, s, c, b, etc.)

The same Lagrangian also describes quantum electrodynamics (QED) if we change the gauge group to $U(1)$. In that case there is but one gauge field (the photon) and, since the group is abelian, the three- and four-point self-

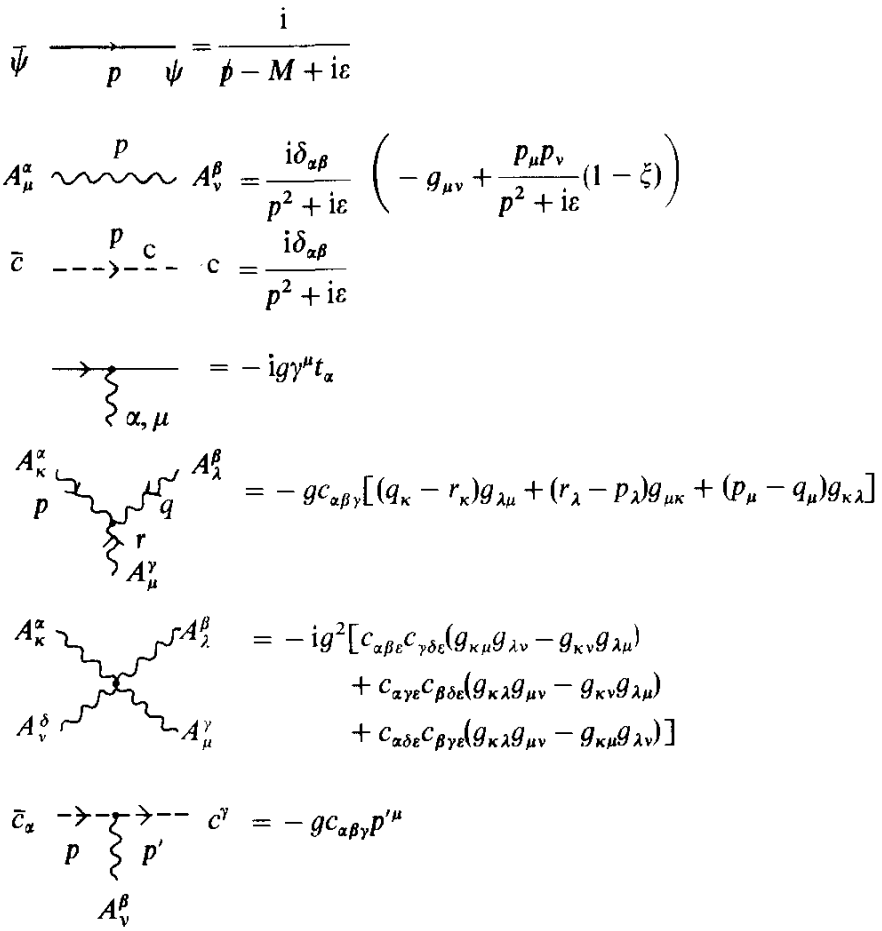


Fig. 2.14.1. Feynman rules for the gauge theory defined by the Lagrangian (2.11.7).

interactions of the gauge field vanish. Moreover, with the gauge fixing term $F[A] = \partial \cdot A$ there is no coupling to the ghost fields, so we may drop them from consideration. The transformations on the matter fields are simple phase rotations: $\psi \to e^{-iq\omega}\psi$ where q is the charge of the field (negative for the electron field). The transformation of the gauge field is $A_\mu \to A_\mu + \partial_\mu \omega$.

2.15 Other symmetries of (2.11.7)

The Lagrangian (2.11.7) is gauge invariant. After gauge fixing we get (2.12.8), which is not gauge invariant, but which has BRS symmetry. The action (2.12.8) also is invariant under global gauge transformations – those with constant ω – because we chose the gauge fixing not to break this symmetry.

There are also what in strong interactions are called flavor symmetries. These are transformations that act identically on every member of an irreducible representation of the gauge group. In this case they give conservation of the number of each of the different flavors of quark. Other flavor symmetries include the discrete symmetries of parity and time-reversal invariance.

Charge-conjugation is also an invariance of (2.12.8), and its action on the ghost fields is rather interesting. Let us define ε_α by the parity of the representation matrices under transposition:

$$t_\alpha = \varepsilon_\alpha t_\alpha^{\mathrm{T}}(\sum). \tag{2.15.1}$$

In this and the following equations, the symbol $(\sum)$ means that the summation convention on repeated indices is suspended. The fermion and gauge fields transform as usual:

$$\begin{aligned} A_\alpha^\mu &\to -A_\alpha^\mu \varepsilon_\alpha(\sum), \\ \psi &\to \eta_c(i\gamma^0\gamma^2)(\bar{\psi})^{\mathrm{T}}, \end{aligned} \tag{2.15.2}$$

where η_c is a real matrix such that

$$t_\alpha^{\mathrm{T}} = \eta_c^{\mathrm{T}} t_\alpha \eta_c. \tag{2.15.3}$$

The ghosts transform as:

$$c^\alpha \to -\varepsilon^\alpha c^\alpha, \quad \bar{c}^\alpha \to -\varepsilon^\alpha \bar{c}^\alpha. \tag{2.15.4}$$

Consider the $c^\alpha - \bar{c}^\beta - A_\mu^\gamma$ and the $A_\lambda^\alpha - A_\mu^\beta - A_\nu^\gamma$ Green's functions. They are invariant under global gauge transformations so only two couplings of the gauge indices are possible: $c_{\alpha\beta\gamma}$ which is antisymmetric, and a symmetric coupling which we can call $d_{\alpha\beta\gamma}$. Charge-conjugation invariance prohibits the symmetric coupling.

2.16 Model field theories

Although the concepts of quantum field theory are very general, we have reviewed them by examining mainly two specific models. The first was the theory of a real scalar field (2.1.1), mostly with the ϕ^4 interaction (2.2.1). The second was a gauge theory (2.11.7) with matter described by some Dirac fields. It should be clear that the general principles apply to any Lagrangian $\mathscr{L}$. A field theory is specified by listing its elementary fields and giving a formula for $\mathscr{L}$. It is solved by a functional-integral representation of its Green's functions.

The aim of physics is to describe the real world. To the extent that a field-theoretic description is the correct one, the fundamental problem in physics is to find the correct field theory. In fact the Lagrangian (2.11.7) appears to do this for strong and electromagnetic interactions if the gauge group and matter fields are correctly chosen. Weak interactions can be included by the Weinberg–Salam theory, and many speculations have been made about extensions (see the proceedings of most recent conferences on high-energy physics).

Our aim in this book would be badly served by only treating real theories. One reason is that we wish to develop techniques and concepts applicable to any field theory, for example not only to the many Grand Unified Theories currently under discussion (see Langacker (1981) and Ross (1981) for reviews), but also to the theories to be invented in the future. As is usual in the subject, we will make use of field theories that are more properly called models. The ϕ^4 theory is an obvious case. Another important reason for using models is to be able to discuss particular aspects of the methods without having other complications to clutter up the presentation.

Particular models will be introduced as needed. Some will recur often, such as the ϕ^4 theory and the simple gauge theory (2.11.7).

Another frequently used model is the ϕ^3 interaction of a real scalar field:

$$\mathscr{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - g\phi^3/3!. \tag{2.16.1}$$

This is much more unphysical than the other models. It is not even completely consistent. Because of the ϕ^3 interaction, the energy is not bounded below. This is manifest in the classical theory and true in the quantum theory (Baym (1960)). Hence any state must catastrophically decay. But the perturbation theory is well-defined, and somewhat simpler than for the ϕ^4 theory. So it proves very convenient to use the ϕ^3 model in treating the elements of the theory of renormalization within perturbation theory.

Another way of constructing models is to change the dimension, d, of space-time from its physical value 4. One motivation for this is that the renormalization problem becomes easier as d is reduced; the degree of divergence of a Feynman graph decreases. As we will see in Chapter 4, it is both useful and possible to treat d as a continuous variable, for the purpose of computing the values of terms in the perturbation expansion.

3

Basic examples

In the later chapters we will develop methods to treat large-momentum behavior. The complete treatment becomes rather intricate at times, so this chapter is devoted to exposing in their simplest form the issues we will be discussing. We will do this by examining the self-energy graph in ϕ^3 theory. This will exhibit the basic phenomena which we will later be treating in detail.

We will see that (in four-dimensional space-time) the graph is renormalized by a mass counterterm. Then the concept of 'degree of divergence' will be introduced by varying d, the dimensionality of space-time. This device will enable us to see how simple power-counting methods determine what counterterms are needed. It will also introduce us to the method of dimensional regularization.

The renormalization group will be introduced by examining the behavior of the graph as its external momentum, p, is made large. By exploiting the arbitrariness in the renormalization procedure, we can reduce the size of higher-order contributions when p^μ is large.

3.1 One-loop self-energy in ϕ^3 theory

Consider the graph shown in Fig. 3.1.1 in the ϕ^3 theory of (2.16.1). We define its contribution to the self-energy to be i times the value of the graph with the external propagators removed:

$$\Sigma_1(p^2) = \frac{\mathrm{i}}{2}\frac{g^2}{(2\pi)^4}\int \mathrm{d}^4k \frac{1}{[k^2 - m^2 + \mathrm{i}\varepsilon][(p+k)^2 - m^2 + \mathrm{i}\varepsilon]}. \quad (3.1.1)$$

The overall factor $\frac{1}{2}$ is a symmetry factor.

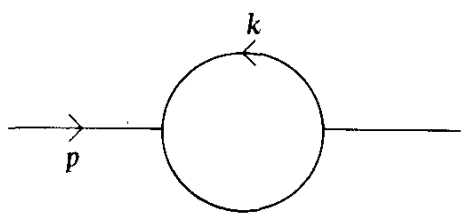


Fig. 3.1.1. One-loop self-energy graph in ϕ^3 theory.

When all components of k^μ get large, this integral diverges logarithmically. It is the simplest example of an ultra-violet divergence. As we will see, the divergence can be cancelled by a mass counterterm. But to explain the renormalization properly, we must discuss a number of other issues as well:

(1) The fact that if $|k^0| \simeq |\vec{k}|$ the divergence as k goes to infinity appears to be much worse.
(2) A precise way of formulating the statement that the divergence is cancelled by a mass counterterm.
(3) The arbitrariness inherent in the renormalization.
(4) The interpretation in coordinate space.

3.1.1 *Wick rotation*

The first of these problems is handled by recalling the Wick rotation into Euclidean space that was used to define the functional integral. This rotation determined the sign of the $i\varepsilon$ in the free propagator. The Wick rotation involved starting with imaginary time $t = -i\tau$, then performing the integral, and finally analytically continuing back to real time. In momentum space, this forces us to work with $k^0 = +i\omega$, the opposite sign appearing so that in the Fourier transformation $e^{ik\cdot x}$ is always a phase.

In the Euclidean formulation let us perform the k^0-integral first. The pole structure in the k^0-plane is shown in Fig. 3.1.2, when p^0 is imaginary. In this situation k^2 and $(p+k)^2$ are both negative, so that the integrand is positive definite. Observe that the factor i coming from the Wick rotation combines with the overall factor of i in (3.1.1) to make Σ_1 real. (We have $d^4k = id\omega d^3k$.)

Now rotate p^0 back to a real value. If $|p^0| < m$ then we have the situation shown in Fig. 3.1.3: there is no obstruction to rotating the k^0 contour to run along the real axis. It is only at this last step that there is a problem from the

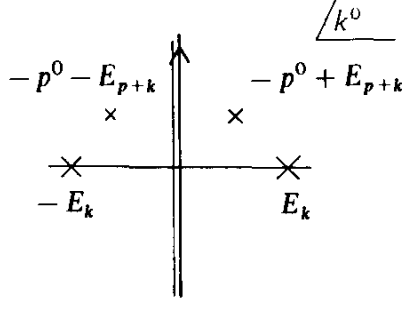


Fig. 3.1.2. The k^0-plane when p^0 is imaginary.

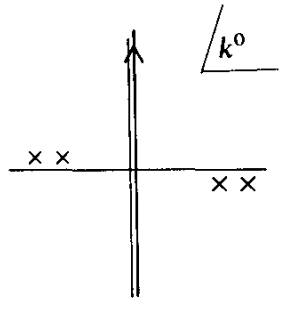


Fig. 3.1.3. The k^0-plane when p^0 is real, but $|p^0|$ is less than m.

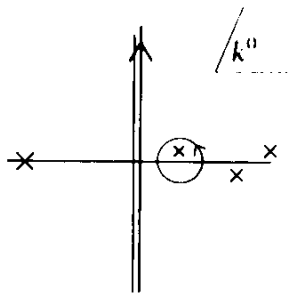


Fig. 3.1.4. The k^0-plane when p^0 is real, but $|p^0|$ is greater than m.

region of $|k^0| \simeq |\vec{k}|$. To avoid the problem we merely have to define the integral by rotating the k^0-contour to run along the imaginary axis.

Now continue p^0 to the region $|p^0| > m$. The case of negative p^0 is illustrated in Fig. 3.1.4. Again we Wick-rotate to imaginary k^0, but this time we pick up a pole term. Now the pole term occurs only when $(p^0)^2 > m^2 + (\vec{p}+\vec{k})^2 \equiv E_{p+k}^2$. Thus it contributes only in a finite region of $\vec{k}$; the UV divergence still comes from the integration over imaginary k^0.

The moral of all this is that the UV divergence is essentially Euclidean, i.e., we may regard k^0 as imaginary and $k^2 < 0$, $(p+k)^2 < 0$.

3.1.2 *Lattice*

We next need to quantify the divergence. The divergence comes from the asymptotic large-k behavior of the integrand which is $1/(k^2)^2$. Let us add and subtract a term with this behavior:

$$\Sigma_1 = \frac{ig^2}{32\pi^4}\left\{\int d^4k\left[\frac{1}{(k^2-m^2+i\varepsilon)[(p+k)^2-m^2+i\varepsilon]} - \frac{1}{(k^2-\mu^2+i\varepsilon)^2}\right]\right.$$
$$\left. + \int d^4k \frac{1}{(k^2-\mu^2+i\varepsilon)^2}\right\}. \tag{3.1.2}$$

The first integral is manifestly finite, for we have subtracted off the leading asymptotic behavior of the integrand. To avoid introducing an extra divergence at $k^2 = 0$ we have subtracted $1/(k^2-\mu^2)^2$ rather than $1/(k^2)^2$. Since we add this term back on, the value of μ is irrelevant; Σ_1 is unchanged. The second term, while divergent, is independent of p. This is the fact that will enable us to cancel the divergence by a counterterm.

Of course we are manipulating divergent integrals, so that (3.1.2), as it stands, makes no sense. We will remedy this defect by using the fact that the theory is defined initially on a lattice. The propagators will then be different functions of momentum. However, the structure of (3.1.2) will be unchanged after imposing a cut-off, as we will now show.

To define the functional integral we not only had to Wick rotate time, but also had to put the theory on a space-time lattice, of spacing a. In the lattice theory, let the free propagator of a particle of mass M be $S_F(q; M, a)$. In the limit $qa \to 0$, this is just $i/(q^2 - M^2 + i\varepsilon)$. But it is zero if $qa > 1$, since high-momentum states do not exist on the lattice. (The reason is, of course, that when one makes a Fourier transformation on a discrete space, one only uses momentum modes with wave-lengths longer than a lattice spacing.) The self-energy on the lattice is finite, and (3.1.2) now reads:

$$\Sigma_1(p, m; a) = \frac{ig^2}{32\pi^4}\left\{\int d^4k[S_F(k; m, a)S_F(p + k; m, a) - S_F(k; \mu, a)^2] + \int d^4k S_F(k; \mu, a)^2\right\}$$
$$\equiv \Sigma_{1\,\mathrm{fin}}(p, m, \mu, a) + \Sigma_{1\,\mathrm{div}}(m, \mu, a). \tag{3.1.3}$$

All the integrals are now convergent, so (3.1.3) is a correct version of (3.1.2). As the lattice spacing goes to zero, the first integral approaches the first convergent integral in (3.1.2). The second integral diverges, but is independent of p. Thus (3.1.2) is not nonsensical, provided that the propagators are implicitly replaced by lattice propagators wherever necessary.

3.1.3 *Interpretation of divergence*

No matter how it is manipulated, the self-energy diverges in the continuum limit. The use of a lattice cut-off now enables us to quantify the divergence. From (3.1.3)

$$\Sigma_1 = \frac{ig^2}{32\pi^4}\int d^4k S_F(k; \mu, a)^2 + \Sigma_{1\,\mathrm{fin}}$$
$$= \frac{-g^2}{32\pi^4}\int_{\omega^2 + \vec{k}^2 < 1/a^2} d\omega d^3k \frac{1}{(\omega^2 + \vec{k}^2 + \mu^2)^2} + \text{finite}$$
$$= \frac{-g^2}{16\pi^2}\int_0^{1/a} dk \frac{k^3}{(k^2 + \mu^2)^2} + \text{finite}$$
$$= \frac{-g^2}{16\pi^2}\ln 1/a + \text{finite} \qquad \text{as } a \to 0. \tag{3.1.4}$$

Thus we can interpret the divergence as follows: Let Σ be (i times) the sum of self-energy graphs. (As usual, the self-energy graphs are graphs for the propagator that have the external lines amputated and that cannot be split



Fig. 3.1.5. Summation of self-energy graphs into propagator.

into disconnected parts by cutting a single line.) Then the full propagator is

$$\bar{G}_2(p^2) = \mathrm{i}/(p^2 - m^2 - \Sigma + \mathrm{i}\varepsilon). \tag{3.1.5}$$

This equation is illustrated in Fig. 3.1.5; the propagator is the sum of a geometric series involving the self-energy. The actual mass m_{ph} of the particle is determined by the pole position, $p^2 = m_{\mathrm{ph}}^2$. Evidently m_{ph}^2 is not m^2 but $m^2 + \Sigma(p^2 = m_{\mathrm{ph}}^2, m^2)$. In other words, the self-energy represents the dynamical contribution to the mass coming from the interactions. The divergence (3.1.4) is independent of p^2, so it is precisely a contribution to the mass. (We ignore higher orders for now.)

Traditionally, one observes that it is convenient to parametrize the theory, not by the mass parameter m that cannot be observed directly but by the physical mass m_{ph}. One writes the mass term in $\mathscr{L}$ as

$$-m_0^2\phi^2/2 = -m_{\mathrm{ph}}^2\phi^2/2 - \delta m^2\phi^2/2. \tag{3.1.6}$$

The first term is left in the free Lagrangian, so that the free propagator is $\mathrm{i}/(p^2 - m_{\mathrm{ph}}^2 + \mathrm{i}\varepsilon)$. But the second term – called the mass counterterm – is put into the interaction Lagrangian, and adjusted so that the full propagator has a pole at $p^2 = m_{\mathrm{ph}}^2$. The counterterm exactly cancels the dynamical contribution to the particle's mass. This is the basic idea of renormalization. It is physically irrelevant that δm^2 happens to diverge.

In perturbation theory, δm^2 is determined as a power series in g. To $O(g^2)$ we have, in addition to Fig. 3.1.1, the graph of Fig. 3.1.6 corresponding to the mass counterterm in (3.1.6). The self-energy to $O(g^2)$ with the new parametrization is the sum of Figs. 3.1.1 and 3.1.6. We call it the renormalized self-energy:

$$\begin{aligned}\Sigma_{1\mathrm{R}} &= \{\Sigma_1(p^2; m_{\mathrm{ph}}^2, a) + \delta m^2\}\big|_{a\to 0} \\ &= \Sigma_{1\mathrm{fin}}(p^2; m_{\mathrm{ph}}^2; \mu, 0) + [\delta m^2 + \Sigma_{1\mathrm{div}}].\end{aligned} \tag{3.1.7}$$



Fig. 3.1.6. Counterterm graph to the self-energy.

We adjust δm^2 first of all to cancel the divergence in $\Sigma_{1\mathrm{div}}$, so that the term in square brackets is finite as $a \to 0$. Then we adjust the finite part so that

$$\delta m^2 + \Sigma_{1\mathrm{div}} = -\Sigma_{1\mathrm{fin}}(m_{\mathrm{ph}}^2, m_{\mathrm{ph}}^2, \mu). \tag{3.1.8}$$

For the renormalized self-energy we find

$$\Sigma_{1R} = \frac{ig^2}{32\pi^4}\int d^4k \left\{ \frac{1}{(k^2 - m_{ph}^2 + i\varepsilon)[(p+k)^2 - m_{ph}^2 + i\varepsilon]} - \frac{1}{(k^2 - \mu^2 + i\varepsilon)^2} \right\}$$

$$- \text{value at } p^2 = m_{ph}^2. \tag{3.1.9}$$

Note that the μ-dependence cancels.

Equation (3.1.9) gives the value of the self-energy in the continuum theory correct to $O(g^2)$.

3.1.4 *Computation*

One way to calculate Σ_{1R} is to differentiate with respect to p^2. (It is of course a Lorentz scalar.) Integrating the result gives Σ_{1R}; the constant of integration is fixed by the renormalization condition that $\Sigma_{1R}(m_{ph}^2)$ is zero.

We have:

$$\frac{\partial \Sigma_{1R}}{\partial p^2} = \frac{p^\mu}{2p^2}\frac{\partial}{\partial p^\mu}\Sigma_{1R}$$

$$= \frac{-ig^2}{32\pi^4 p^2}\int d^4k \frac{p\cdot(p+k)}{(k^2 - m^2 + i\varepsilon)[(p+k)^2 - m^2 + i\varepsilon]^2}. \tag{3.1.10}$$

This is identical to what we would have obtained from the unrenormalized expression (3.1.1) without regard to the fact that it is divergent. We could have written it down directly without going through the long explanation that we used. But then there would have been no defense to the argument that we are manipulating meaningless quantities and that therefore quantum field theory makes no sense.

Since (3.1.10) is finite it can be easily calculated by using a Feynman parameter representation and then by shifting the k-integral:

$$\frac{\partial \Sigma_{1R}}{\partial p^2} = \frac{ig^2}{16\pi^4 p^2}\int_0^1 dx\, x \int d^4k \frac{p\cdot(p+k)}{[m_{ph}^2 - p^2 x - 2p\cdot kx - k^2 - i\varepsilon]^3}$$

$$= \frac{ig^2}{16\pi^4 p^2}\int_0^1 dx\, x \int d^4k \frac{p\cdot[k + p(1-x)]}{[m_{ph}^2 - p^2 x(1-x) - k^2 - i\varepsilon]^3}$$

$$= \frac{-g^2}{32\pi^2}\int_0^1 dx \frac{x(1-x)}{[m_{ph}^2 - p^2 x(1-x)]}$$

$$= \frac{g^2}{32\pi^2}\frac{\partial}{\partial p^2}\left\{ \int_0^1 dx \ln[m_{ph}^2 - p^2 x(1-x)] \right\}. \tag{3.1.11}$$

Using the condition $\Sigma_{1R} = 0$ at $p^2 = m_{ph}^2$ now gives

$$\Sigma_{1R} = \frac{g^2}{32\pi^2} \int_0^1 dx \ln \left[\frac{m_{ph}^2 - p^2 x(1-x)}{m_{ph}^2 (1 - x + x^2)} \right]. \tag{3.1.12}$$

This integral can be worked out analytically.

3.2 Higher order

The graph of Fig. 3.1.1 is not the only divergent graph in the theory. In Chapter 5 we will discuss the general theory of renormalization and we will see how to extend the removal of divergences to all orders. In this section we will only consider a class of graphs which have divergences generated because Fig. 3.1.1 occurs as a subgraph. Examples are given in Fig. 3.2.1.

One property should be clear. This is that the divergences come from subgraphs all of whose lines are part of a loop. A general way of characterizing these subgraphs is to define the concept of a one-particle-irreducible graph or subgraph. A one-particle-irreducible (1PI) graph is one which is connected and cannot be made disconnected by cutting a single line. A graph which is not 1PI is called one-particle-reducible (1PR). The graphs in Fig. 3.2.1 are all one-particle-reducible, since they all have one or more lines that when cut leave the graph in two disconnected pieces. The self-energy subgraph of Fig. 3.1.1 consisting of the two lines in the loop is 1PI. This identical subgraph occurs several times in the graphs of Fig. 3.2.1.

We introduced a mass counterterm into the interaction, so that the

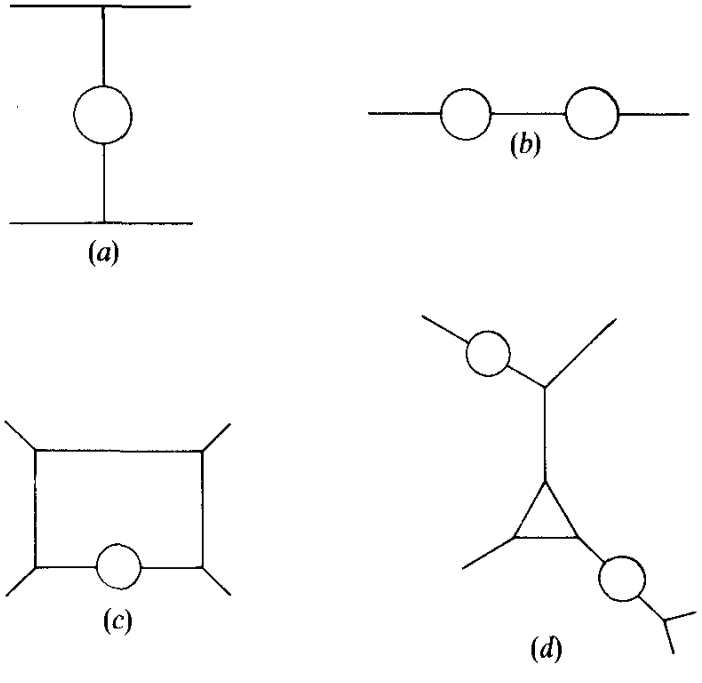


Fig. 3.2.1. Graphs containing the one-loop self-energy as a subgraph.

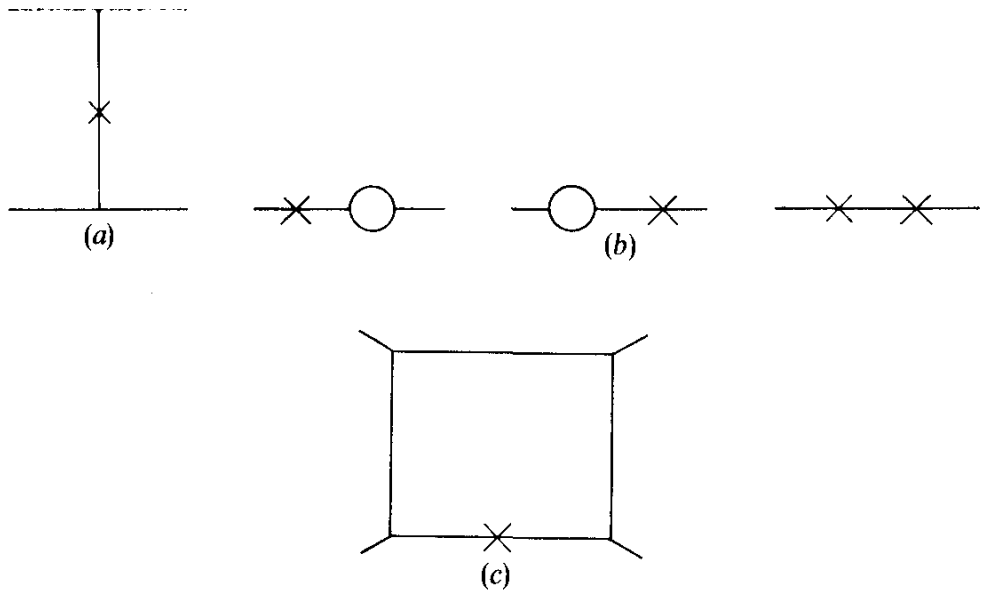


Fig. 3.2.2. Counterterm graphs to Fig. 3.2.1.

counterterm graph Fig. 3.1.6 cancels the divergence in Fig. 3.1.1. Clearly the counterterm vertex can be used as an interaction anywhere in a graph. In fact, all graphs containing it can be found as follows: (*a*) take a graph with the loop of Fig. 3.1.1 occurring as subgraph one or more times, but with no mass counterterm vertices; (*b*) replace one (or more) of the occurrences of the loop by the counterterm. The terms generated from Figs. 3.2.1 (*a*)–(*c*) are shown in Fig. 3.2.2.

Evidently, the sum of the original graph and its counterterm graph(s) is just the original graph with every occurrence of the loop replaced by its renormalized value $-i\Sigma_{1R}$. It is sensible to keep the counterterm associated with the loop, thereby considering the loop plus the counterterm as a single entity.

In the case of the graphs of Fig. 3.2.1 the result of this procedure is to make the graphs finite. The generalization to an arbitrary graph will be worked out in Chapter 5.

3.3 Degree of divergence

We saw how the Wick rotation ensured that the UV divergence of the one-loop self-energy is a purely Euclidean problem. The divergence is from the region $|k^\mu| \to \infty$, without any regard to direction. Thus, simple power-counting determined that there is a logarithmic divergence. The power-counting involves merely counting the powers of k in the integral for large k. The divergence is logarithmic as the lattice spacing, a, goes to zero.

Power-counting in this fashion works for a general graph to determine

what is called either the 'overall degree of divergence' or the 'superficial degree of divergence'. This we will discover in Chapter 5. There we will also see how the value of the degree of divergence determines the particular counterterm vertices needed for a theory. This will enable us to determine whether or not a theory is renormalizable by invoking arguments revolving around the dimension of the coupling.

In this section we will vary the dimensionality, d, of space-time in the calculation of the self-energy graph, Fig. 3.1.1. We can then explore the relation between the degree of divergence, as determined by power-counting, and the momentum dependence of the counterterm.

The integral for Fig. 3.1.1 is now

$$\Sigma_1(p^2, m^2, d) = \frac{ig^2}{2(2\pi)^d} \int d^d k \frac{1}{(k^2 - m^2 + i\varepsilon)[(p+k)^2 - m^2 + i\varepsilon]}. \quad (3.3.1)$$

The space-time has one time dimension and $d-1$ space dimensions. In the Feynman rules the factors $(2\pi)^4$ get replaced by $(2\pi)^d$, since they arise from the result

$$\int d^d x e^{ik\cdot x} = (2\pi)^d \delta^{(d)}(k). \quad (3.3.2)$$

The number of powers of k in the integral is now $d-4$; we call this the degree of divergence of the graph. If d is less than four, then the graph is convergent. But whenever d is greater than or equal to four, the graph diverges. Our discussion in the previous sections tells us we must try to renormalize it by adding a counterterm.

Now, differentiating once with respect to p^μ gives convergence if $d = 4$:

$$\frac{\partial \Sigma_1}{\partial p^\mu} = \frac{-ig^2}{(2\pi)^d} \int d^d k \frac{(p+k)_\mu}{(k^2 - m^2)[(p+k)^2 - m^2]^2}, \quad (3.3.3)$$

with the degree of divergence of the integral being reduced by one to $d-5$. If $d = 5$ the integral diverges logarithmically. However one might surmise that the divergence comes from the piece of the integrand proportional to k_μ and that symmetrizing by $k \to -k$ would kill the divergence. This is in fact true, but let us be more simple-minded.

Differentiating again with respect to p gives a result with degree of divergence $d-6$:

$$\frac{\partial^2 \Sigma_1}{\partial p^\mu \partial p^\nu} = \frac{ig^2}{(2\pi)^d} \int d^d k \frac{\{2(p+k)_\mu (p+k)_\nu - g_{\mu\nu}[(p+k)^2 - m^2]\}}{(k^2 - m^2)[(p+k)^2 - m^2]^3}. \quad (3.3.4)$$

To recover Σ_{1R} at $d = 5$ we integrate twice. There are more constants of integration that appear as an additive contribution of the form: $A + B_\mu p^\mu$.

However, we must require Lorentz invariance of Σ, so the B_μ term is eliminated, and we are left with a mass term as the only counterterm. This is the first and simplest example of the use of a symmetry argument to determine the form that we will allow for a counterterm.

3.3.1 ϕ^3 at $d = 6$

Let us now go to $d = 6$. Differentiating Σ_1 three times gives a finite result with degree of divergence $d - 7$. Integrating to obtain Σ_{1R} gives arbitrary integration constants of the form $A - Bp^2$, where again we have used Lorentz invariance. If we went to the lattice we would find divergences proportional to $1/a^2$, $m^2 \ln(a)$ and $p^2 \ln(a)$. The fact that these terms have dimension 2 corresponds to the fact that the integral for Σ_1 has degree of divergence 2. To make it finite we must not only use a mass counterterm but also a counterterm proportional to p^2; the total we will call $\delta m^2 - \delta Z p^2$. This is generated by a counterterm

$$-\delta m^2 \phi^2/2 + \delta Z(\partial\phi)^2/2 \tag{3.3.5}$$

in the Lagrangian.

Evidently the value of the degree of divergence is reflected as the maximum number of derivatives or powers of p in the counterterms. Equally, it is reflected in the integration constants that appear when we recover Σ_{1R} from the differentiated Σ_1. These two phenomena happen for a general graph, as we will see later. The method of proof will in fact be to differentiate each graph enough times with respect to its external momenta until it is finite.

The $(\partial\phi)^2$ counterterm in (3.3.5) is of course an example of the wave-function renormalization introduced in Section 2.3. We can interpret it physically by examining the propagator. The propagator for the bare field can be expressed in terms of the propagator of the renormalized field:

$$\begin{aligned}
\bar{G}_{2(0)} &\equiv \langle 0 | T\tilde{\phi}_0(p)\phi_0(0)|0\rangle \\
&= Z\langle 0 | T\tilde{\phi}(p)\phi(0)|0\rangle \\
&= iZ/(p^2 - m_{\text{ph}}^2 - \Sigma_{1(0)} - \delta m^2 + \delta Z p^2 + i\varepsilon) \\
&= iZ/(p^2 - m_{\text{ph}}^2 - \Sigma_{1R}).
\end{aligned} \tag{3.3.6}$$

Note the distinction between bare and renormalized fields. The residue of the particle pole in the propagator of an interacting field is in general not unity:

$$\bar{G}_{2(0)} = iR_{(0)}/(p^2 - m_{\text{ph}}^2 + i\varepsilon) + \text{finite}, \quad \text{as} \quad p^2 \to m_{\text{ph}}^2. \tag{3.3.7}$$

Examination of its spectral representation demonstrates that $0 \leq R(0) < 1$,

because the field ϕ_0 has canonical equal-time commutation relations. (The proof is given, for example, in Section 16.4 of Bjorken & Drell (1966).)

Now we can always change the definition of $\phi \equiv Z^{1/2}\phi_0$ by multiplying Z by a finite factor. So it is possible to adjust Z so that the renormalized propagator has a pole of unit residue at $p^2 = m_{\text{ph}}^2$:

$$\bar{G}_2 = \mathrm{i}/(p^2 - m_{\text{ph}}^2 + \mathrm{i}\varepsilon) + \text{finite}, \quad \text{as} \quad p^2 \to m_{\text{ph}}^2. \tag{3.3.8}$$

In this case we identify $R_{(0)}$ with Z. The renormalized self-energy satisfies

$$\Sigma_{1R} = \partial\Sigma_{1R}/\partial p^2 = 0 \quad \text{at} \quad p^2 = m_{\text{ph}}^2. \tag{3.3.9}$$

When integrating to obtain Σ_{1R} from the finite derivative of Σ_1, this condition enables the integration constants to be determined. It is called the mass-shell renormalization condition.

3.3.2 *Why may Z be zero and yet contain divergences?*

As a property of the exact theory we know that $0 \leq Z < 1$, if we adopt the renormalization condition (3.3.9); Z is definitely finite. However its perturbation expansion starts at $1 + g^2[C\ln(a) + \text{finite}] + \cdots$. (Here C is a constant.) This seems to be infinite as $a \to 0$, rather than finite. We resolve the contradiction by realizing that we should not expect higher-order terms to be small if the one-loop correction is large. For example, we could have the series

$$Z = \left[\frac{1}{1 - g^2[(C/D)\ln(a) + \text{const.}]}\right]^D, \tag{3.3.10}$$

where D is a positive number. Then $Z \to 0$ as $a \to 0$ even through the one-loop term goes to infinity. In any event we see that C must be positive (otherwise if we fix a and let $g \to 0$ there is a region with $Z > 1$).

It would appear impossible to derive a formula like (3.3.10) since it involves summing all orders of perturbation theory. Moreover it involves an analytic continuation from within its radius of convergence $|\ln(a)| < D/g^2C$ to $|\ln(a)| = \infty$. This seems to make no sense at all since perturbation series are in general asymptotic series rather than convergent series. However, we will see in Chapter 7, on the renormalization group, that we can find a systematic method of calculating Z and the other renormalizations in the limit $a \to 0$. It relies on the so-called asymptotic freedom of the theory. The behavior (3.3.10) will turn out to be essentially correct, in asymptotically free theories.

No matter what the truth is, it should be clear that the divergences in perturbation theory as $a \to 0$ need not be reflected as divergences in the exact theory, but only as singularities.

3.3.3 Renormalizability and non-renormalizability

Suppose we now go to a dimension $d > 6$. For example, let us set $d = 8$. Then the one-loop self-energy has a quartic divergence. The necessary counterterms contain up to four derivatives of the field:

$$-\delta m^2\phi^2/2 - \delta Z(\partial\phi)^2/2 + E(\Box\phi)^2/2. \tag{3.3.11}$$

The quartic term is not of the form of any term in the original Lagrangian, so it cannot be obtained by renormalization of the Lagrangian. When this situation occurs the theory is called non-renormalizable. Non-renormalizability is *a priori* a good reason for dropping the theory from consideration. There are possible ways to avoid this, but we will leave discussion of this until later.

There is a simple argument that helps in the determination of whether or not a theory is renormalizable. It links dimensional analysis and power counting. Consider a one-particle-irreducible graph Γ. Let its degree of divergence be $\delta(\Gamma)$, and let its mass dimension be $d(\Gamma)$. Now Γ is the product of a numerical factor, a set of couplings, and an integral. As we see from (3.3.1) the degree of divergence is the dimension of the integral. Therefore, if we let $\Delta(\Gamma)$ be the dimension of the couplings in Γ, we have

$$d(\Gamma) = \delta(\Gamma) + \Delta(\Gamma). \tag{3.3.12}$$

Now consider the counterterms to Γ. These form a polynomial of degree $\delta(\Gamma)$ in the external momenta. For each term C in the polynomial, let $\delta(C)$ be the number of derivatives and let $\Delta(C)$ be the dimension of the coefficient, so that its dimension is the same as Γ:

$$\delta(C) + \Delta(C) = d(\Gamma). \tag{3.3.13}$$

Now, the maximum number of derivatives in the counterterms is

$$\delta(\Gamma) = d(\Gamma) - \Delta(\Gamma).$$

If the couplings have negative dimension, then $\delta(\Gamma)$ can be made arbitrarily large by going to a graph of high enough order. In the absence of miraculous cancellations this tells us to expect non-renormalizability. If the couplings have zero or positive dimension, we have a finite number of counterterm vertices, since $-\Delta(\Gamma)$ is bounded above by zero and $d(\Gamma)$ decreases as the number of external lines increases.

If the couplings never have negative dimension, we observe that the coefficients of the counterterms satisfy

$$\begin{aligned}\Delta(C) &= d(\Gamma) - \delta(C)\\ &\geq d(\Gamma) - \delta(\Gamma)\\ &= \Delta(\Gamma),\end{aligned}$$

so that these coefficients also have non-negative dimension. This, in the simplest cases, is sufficient to ensure renormalizability.

For example, in the ϕ^3 theory ϕ has dimension $d/2 - 1$. (Recall that $\mathscr{L}$ has the dimension of an energy-density.) Then $\Delta(g) = 3 - d/2$. So if $d > 6$ the theory is non-renormalizable, as we saw by example. If $d \leq 6$, there are only a finite number of possible counterterms, since these are restricted to have coefficients of non-negative dimension.

3.4 Renormalization group

3.4.1 *Arbitrariness in a renormalized graph*

The infinities of a renormalizable theory amount to divergent dynamical contributions that renormalize the parameters in the Lagrangian. Traditionally one thinks of renormalization as the procedure of working with measured quantities instead of the corresponding bare quantities. The most obvious case is that of the mass m_{ph} of the particle corresponding to an elementary field. However to take the traditional view is much too restrictive.

This issue can be understood by looking at strong interactions. There we have a theory, QCD, in which free particles corresponding to the elementary fields do not appear to exist (*pace* LaRue, Phillips & Fairbank (1981)). So arises the hypothesis of quark confinement – not proved from QCD, so far – according to which quarks are never isolated particles. Even so, the theory has quark masses, which can be measured (up to considerable uncertainties for the light quarks). But one cannot identify these masses with the directly measurable masses of free quarks. One must only speak of mass parameters, measured, in this case, rather indirectly.

There is no problem in taking this point of view. For example, we write the ϕ^3 Lagrangian as a basic Lagrangian plus a counterterm Lagrangian:

$$\mathscr{L} = \partial\phi^2/2 - m^2\phi^2/2 - g\phi^3/3! - \delta Z\partial\phi^2/2 - \delta m^2\phi^2/2 - \delta g\phi^3/3!. \tag{3.4.1}$$

But we avoid identifying the renormalized mass with the mass m_{ph} of a particle. Similarly we do not identify the renormalized coupling, g, with any specific measured quantity, and we do not define Z by requiring that the residue of the propagator's pole be unity.

Consider the calculation of the one-loop self-energy at $d = 4$. We can choose $\delta Z = 0$ for this case; the only divergence is in δm^2. The renormalized self-energy (3.1.7) is

$$\Sigma_{1R} = \Sigma_{1\text{fin}} + (\delta m^2 + \Sigma_{1\text{div}}).$$

We must choose δm^2 to have a divergent part to cancel the divergence at $a = 0$ of $\Sigma_{1\text{div}}$. But the finite part of δm^2 is not determined; the arbitrariness is the same as that of the integration constant when obtaining Σ_{1R} by integrating $\partial\Sigma_1/\partial p^\mu$.

At first sight it might appear that the arbitrariness ruins the theory unless one pins down m to be the physical mass m_{ph}. This is in fact not so; the arbitrariness is more like the arbitrariness in choosing a coordinate system. Suppose one first computes the propagator with the mass-shell condition $\Sigma_{1R}(p^2 = m_{\text{ph}}^2) = 0$. Then

$$\bar{G}_2 = \mathrm{i}/[p^2 - m_{\text{ph}}^2 - \Sigma_{1\text{fin}}(p^2) + \Sigma_{1\text{fin}}(p^2 = m_{\text{ph}}^2) + O(g^4)], \quad (3.4.2a)$$

$$m_0^2 = m_{\text{ph}}^2 - \Sigma_{1\text{div}} - \Sigma_{1\text{fin}}(m_{\text{ph}}^2) + O(g^4). \quad (3.4.2b)$$

One could also compute with a different finite part to δm^2, with a result

$$\bar{G}_2 = \mathrm{i}/[p^2 - m^2 - \Sigma_{1\text{fin}}(p^2, m^2) + g^2 C + O(g^4)], \quad (3.4.3a)$$

$$m_0^2 = m^2 - \Sigma_{1\text{div}} - g^2 C + O(g^4), \quad (3.4.3b)$$

where C is any chosen number. The self-energy is now

$$\Sigma_{1R}^{(C)}(p^2, m^2) = \Sigma_{1\text{fin}} + g^2 C.$$

Evidently the two ways of renormalizing the theory give the same results if we require that the bare mass m_0^2 is the same in both of (3.4.2) and (3.4.3). In the complete solution of the theory, say by the functional integral, it is only m_0^2 that matters, not the partition into a renormalized mass squared m^2 and a counterterm $-\Sigma_{1\text{div}} - g^2 C$. Clearly we have

$$m^2 = m_{\text{ph}}^2 + g^2 C - \Sigma_{1\text{fin}}(m_{\text{ph}}^2) + O(g^4), \quad (3.4.4)$$

with the $O(g^4)$ terms depending on the renormalization of higher-order self-energy graphs.

We come then to the central idea of the renormalization group. The arbitrariness in the definition of Σ_{1R} is physically irrelevant, for a change in the arbitrary constant C can be exactly compensated by a change in m^2. A change in C merely gives a different parametrization of the set of theories that can be obtained by varying the mass parameter m. The renormalization group is the set of transformations on the parametrizations of the theory. The transformations are accomplished by moving parts of the terms in $\mathscr{L}$ from the basic Lagrangian to the counterterm Lagrangian. In the case of m it is a move from the free Lagrangian to the interaction Lagrangian. This of course gives a rearrangement of the perturbation series, which is the key to the many practical applications of the renormalization group.

It might be objected that

$$p^2 - m^2 - \Sigma_{1\text{fin}}(p^2, m^2) + g^2 C$$

is not equal to

$$p^2 - m_{\rm ph}^2 - \Sigma_{1\rm fin}(p^2, m_{\rm ph}^2) + \Sigma_{1\rm fin}(m_{\rm ph}^2, m_{\rm ph}^2),$$

since the mass parameters in $\Sigma_{1\rm fin}(p^2)$ are different, whereas the theory parametrized in either way is the same. But since $m_{\rm ph}^2 - m^2$ is $O(g^2)$ the difference in the two expressions is in fact $O(g^4)$. Thus the rearrangement of the perturbation series does not leave the p^2-dependence of the coefficients invariant. The $O(g^4)$ terms will cancel the difference (up to even higher-order terms), etc.

The utility of the renormalization group is precisely in its ability to reorganize the perturbation series. Since one effect of the interaction is to induce dynamical contributions to the mass and couplings, it is evidently a good idea to arrange that these contributions are small. The result is to reduce the values of higher-order corrections and thus improve the reliability of a perturbative calculation.

Now the effective size of the dynamical mass or coupling must be treated as dependent on the situation under consideration. This can be seen by examining Σ_{1R} given in (3.1.12) at large p^2:

$$\Sigma_{1R} \sim (g^2/32\pi^2)[\ln(-p^2/m_{\rm ph}^2) + \text{constant} + \cdots]. \tag{3.4.5}$$

If $|p^2|$ is large enough this can be large. Since the graph occurs as a subgraph of higher-order graphs, it is likely (and often is true) that higher-order graphs are as important as low-order graphs at large enough p^2. This situation is undesirable and can be remedied by a renormalization-group transformation.

We absorb the large part of Σ_{1R} into a redefinition of the renormalized mass m^2. We must examine higher-order graphs at large p^2 to demonstrate that there are no further sources of large coefficients. We will do this systematically in Chapter 7.

3.4.2 *Renormalization prescriptions*

There are infinitely many ways of resolving the ambiguity in constructing the counterterms for a given theory, each of these ways corresponding to a particular parametrization. It is essential that, whenever a particular divergent graph occurs as a subgraph of a bigger graph, the ambiguity is resolved in the same way at each occurrence, since the corresponding counterterm vertex is generated by a single term in the Lagrangian. So to perform concrete calculations one adopts some rule to resolve the ambiguity. Such a rule is called a renormalization prescription or renormalization scheme.

Of the infinitely many possible renormalization prescriptions, a few have become standard, because they are especially convenient either for practical use or for theoretical considerations. In this section we will explain two of the standard ones with the aid of the example of the one-loop graph Fig. 3.1.1.

We have already encountered the mass-shell, or physical, scheme. The renormalized mass is defined to be the physical mass, i.e., the position of the propagator pole. Wave-function renormalization is fixed by requiring the residue of the pole to be unity (see (3.3.9)). Couplings can be defined by specifying the value of a suitable S-matrix element.

A possibility that is much used in discussions of renormalization theory is the BPH or BPHZ scheme (Bogoliubov–Parasiuk–Hepp–Zimmermann), otherwise known as zero-momentum subtraction. Let Γ be a one-particle-irreducible (1PI) graph that is divergent, i.e., it has $\delta(\Gamma) \geq 0$. The prescription is that at zero external momentum its renormalized value $R(\Gamma)$ and its first $\delta(\Gamma)$ derivatives with respect to external momentum are zero. The BPHZ scheme is to implement this by subtracting off the first $\delta(\Gamma)$ terms in the Taylor expansion of the integrand (Zimmermann (1970)), that is, the renormalization is performed before the integration over loop momenta. No explicit UV cut-off is needed. In this scheme the self-energy already discussed is, at $d = 4$,

$$\Sigma_{1R}^{(BPHZ)} = \frac{ig^2}{32\pi^4}\int d^4k \left\{ \frac{1}{[(k^2 - m^2)((p+k)^2 - m^2)]} - \frac{1}{(k^2 - m^2)^2} \right\}, \quad (3.4.6)$$

while at $d = 6$, we have:

$$\Sigma_{1R}^{(BPHZ)} = \frac{ig^2}{128\pi^6}\int d^6k \left\{ \frac{1}{(k^2 - m^2)[(p+k)^2 - m^2]} - \frac{1}{(k^2 - m^2 + i\varepsilon)^2} + \frac{2p\cdot k}{(k^2 - m^2)^3} - \frac{[4p\cdot k^2 - (k^2 - m^2)p^2]}{(k^2 - m^2)^4} \right\}. \quad (3.4.7)$$

3.5 Dimensional regularization

In our initial treatment of UV divergences we used the lattice as a cut-off, or regulator. However, what we are really interested in is the renormalized theory with no cut-off. We could equally well use some other kind of regulator. For example, a Pauli–Villars type of cut-off is achieved by a higher derivative term in the Lagrangian. For example, from

$$\mathscr{L} = \partial A^2/2 - m^2A^2/2 - [(\Box + m^2)A]^2/2(M^2 - m^2) - gA^3/3! + \text{counterterms} \quad (3.5.1)$$

we obtain the free propagator (2.4.1). When the cut-off M goes to infinity, the propagator is the ordinary one, but when p^2 is much bigger than M^2, it is smaller by a factor p^2/M^2. Thus UV divergences are cut-off for the theory in six or fewer space-time dimensions; the divergences reappear when we take the limit $M \to \infty$.

If we defined the theory by a functional integral the lattice would appear as an intermediate step, but the $a \to 0$ limit would give no divergences, if M is finite. Although the Euclidean Green's functions for the cut-off theory (3.5.1) exist, the Minkowski space field theory is not physical. A symptom of this is that the pole of the free propagator at $p^2 = M^2$ has the wrong sign of residue; it implies a particle with negative metric.

A theory with no cut-off can be obtained by adding counterterms with appropriate M-dependences to cancel the divergences and then taking the $M \to \infty$ limit. As an example we showed that counterterms cancelled the divergences of the one-loop self-energy graph. Although we assumed a lattice regulator, we used no properties of the lattice propagator that are not true for the Pauli–Villars case. We assumed only that:

(1) If the cut-off is taken away (i.e., $M \to \infty$ or $a \to 0$) with p and m fixed, then the propagator goes to $i/(p^2 - m^2)$.
(2) If $p^2 \to \infty$ with fixed cut-off, then the propagator is sufficiently much smaller than $1/p^2$ that the graph is not UV divergent.
(3) In the Euclidean region there are no propagator poles.

In principle, any method of imposing a UV cut-off is good enough, but in practice some methods are more convenient than others. For most purposes dimensional regularization is the most convenient. The method starts from the observation that UV divergences are eliminated by going to a small enough space-time dimension d. We can use the space-time dimension as a regulator provided we treat d as a continuous variable (so that the cut-off can be removed by taking the limit $d \to 4$). This idea has a long history, but its popularity roughly started after the papers by Wilson (1973) in statistical mechanics and by 't Hooft & Veltman (1972a), Bollini & Giambiagi (1972), Ashmore (1972), and Cicuta & Montaldi (1972) in field theory (especially non-abelian gauge theories).

Since vector spaces of non-integer dimension do not exist as such, it is not obvious that the concept has any consistency, let alone validity, even in a purely formal sense. This we will remedy in the next chapter. For the present we will assume uniqueness and existence, and apply standard manipulations to the integral

$$\Sigma_1(p^2, m^2, d) = \frac{ig^2}{2(2\pi)^d} \int d^d k \frac{1}{(k^2 - m^2 + i\varepsilon)[(p+k)^2 - m^2 + i\varepsilon]} \tag{3.5.2}$$

until it is of a form where we can sensibly assign a value. We will (following Wilson (1973)) express (3.5.2) in terms of a standard Gaussian integral:

$$\int \mathrm{d}^d k \exp(k^2) = \mathrm{i} \int \mathrm{d}\omega \int \mathrm{d}^{d-1} k \exp(-\omega^2 - k^2). \tag{3.5.3}$$

It is sensible to give this the value $\mathrm{i}\pi^{d/2}$, which is correct if d is an integer. In the following calculation of the value of (3.5.2) with non-integer d, the assumed properties of the integration are italicized. All the manipulations are valid for any integer value of d for which the integral converges.

We use the Schwinger representation for each propagator:

$$1/(m^2 - k^2 - \mathrm{i}\varepsilon) = \int_0^\infty \mathrm{d}a \exp[-a(m^2 - k^2 - \mathrm{i}\varepsilon)]. \tag{3.5.4}$$

Observe that because of the Wick rotation we treat k^2 as negative. Then we *exchange the order of integration* to obtain

$$\Sigma_1 = \frac{\mathrm{i}g^2}{2(2\pi)^d} \int_0^\infty \mathrm{d}a \int_0^\infty \mathrm{d}b \int \mathrm{d}^d k \exp[-(a+b)m^2 + bp^2 + 2bp\cdot k + (a+b)k^2]. \tag{3.5.5}$$

We *shift* k^μ by an amount $p^\mu b/(a+b)$ and change variables to $z = a + b$, $x = a/z$ to get:

$$\Sigma_1 = \frac{\mathrm{i}g^2}{2(2\pi)^d} \int_0^1 \mathrm{d}x \int_0^\infty \mathrm{d}z\, z \int \mathrm{d}^d k \exp\{-z[m^2 - p^2 x(1-x)] + zk^2\}.$$

After *scaling* k by a factor $z^{1/2}$ we find that

$$\Sigma_1 = \frac{\mathrm{i}g^2}{2(2\pi)^d} \int_0^1 \mathrm{d}x \int_0^\infty \mathrm{d}z\, z^{1-d/2} \exp\{-z[m^2 - p^2 x(1-x)]\} \int \mathrm{d}^d k \exp(k^2). \tag{3.5.6}$$

It is this stage which brings in the dimensionality d. We have now reduced the d-dimensional integral to the form (3.5.3), which we defined to be $\mathrm{i}\pi^{d/2}$. The z-integral in (3.5.6) gives a Γ-function, so we finally obtain:

$$\Sigma_1 = \frac{-g^2}{2(4\pi)^{d/2}} \Gamma(2 - d/2) \int_0^1 \mathrm{d}x [m^2 - p^2 x(1-x)]^{d/2-2}. \tag{3.5.7}$$

This result is unique (except possibly for an overall normalization, which is universal – the same for every d-dimensional integral). The divergences now reside in the Γ-function which has simple poles at $d = 4, 6, 8, \ldots$. The residue of each pole is a polynomial in p of degree equal to the degree of divergence.

One of the main advantages of dimensional regularization is immediately apparent. Not only was the integral unchanged from its form in a theory in

integer dimensional space-time with no cut-off, but the method of calculation was unchanged. Use of the representation (3.5.4) is an efficient way of obtaining a parametric representation like (3.5.7) for a Feynman graph.

A second advantage that we will see later is that it preserves not only Poincaré invariance in the regulated theory, but also gauge symmetries. This was a main motivation for its use by 't Hooft & Veltman (1972a) and many others. Most methods of introducing a cut-off fail in this respect. (For example, gauge invariance is preserved on the lattice but full Poincaré invariance is lost.)

A third advantage – also of great importance in practice – is that a continuous space-time dimension is also a gauge-invariant cut-off for infrared divergences in theories with massless fields (Gastmans & Meuldermans (1973), Gastmans, Verwaest & Meuldermans (1976), and Marciano & Sirlin (1975)). A trivial example, but without any gauge invariance, is given by the ϕ^3 self-energy with $m=0$ at $d=2$. It is

$$\frac{ig^2}{8\pi^2}\int d^2k \frac{1}{(k^2+i\varepsilon)[(p+k)^2+i\varepsilon]},$$

which is divergent at $k=0$ and at $p=-k$. The divergence is regulated by increasing d. Care is required in using this method if UV divergences are present in the same graph, for they are regulated by reducing d.

3.6 Minimal subtraction

3.6.1 *Definition*

From the unrenormalized self-energy (3.5.7) we compute the renormalized self-energy Σ_{1R} at $d=4$ by adding a mass counterterm $\delta m^2(g,m^2,d)$ and then letting $d\to 4$. Suppose we choose m to be the physical mass. Then

$$\begin{aligned}\Sigma_{1R}^{(ph)} &= \Sigma_1(p^2,m_{ph}^2,g,d)-\Sigma_1(m_{ph}^2,m_{ph}^2,g,d)\\ &= \frac{-g^2}{2(4\pi)^{d/2}}\Gamma(2-d/2)\int_0^1 dx\{[m_{ph}^2-p^2x(1-x)]^{d/2-2}\\ &\qquad -[m_{ph}^2(1-x+x^2)]^{d/2-2}\}. \qquad (3.6.1)\end{aligned}$$

Now $\Gamma(z)$ has a pole at $z=0$:

$$\Gamma(z)=1/z-\gamma_E+O(z), \qquad (3.6.2)$$

where $\gamma_E=0.5772\ldots$ is Euler's constant. So at $d=4$

$$\Sigma_{1R}^{(ph)}=\frac{g^2}{32\pi^2}\int_0^1 dx \ln\left\{\frac{m_{ph}^2-p^2x(1-x)}{m_{ph}^2(1-x+x^2)}\right\} \qquad (3.6.3)$$

in agreement with our earlier calculation.

Since the divergence in Σ_1 amounts to a simple pole at $d = 4$, a rather obvious way of renormalizing it is to define δm^2 to cancel just the singularity, i.e., the pole ('t Hooft (1973)). This, of course, means that we are changing our renormalization prescription. By referring back to (3.5.7), we see that δm^2 in this scheme is:

$$\delta m^2 = (g^2/32\pi^2)/(2 - d/2), \tag{3.6.4}$$

from which we get Σ_{1R} by expanding Σ_1 in a power series in $d - 4$. We find

$$\Sigma_{1R} = \frac{g^2}{32\pi^2}\int_0^1 dx\left\{\ln\left[\frac{m^2 - p^2x(1-x)}{4\pi}\right] + \gamma_E\right\}. \tag{3.6.5}$$

Unfortunately this contains the logarithm of a dimensional quantity. The reason is that in the expansion in powers of $d - 4$ we did not allow for the fact that g has a dimension dependent on d. Therefore we implicitly introduced a mass scale.

To make this scale explicit, we rewrite the coupling

$$g \to \mu^{2-d/2} g, \tag{3.6.6}$$

where we have introduced a parameter μ with the dimensions of mass, called the unit of mass ('t Hooft (1973)). The redefined coupling g now has fixed dimension equal to 1, and the renormalized self-energy becomes

$$\Sigma_{1R}^{(MS)} = \frac{g^2}{32\pi^2}\int_0^1 dx\left\{\ln\left[\frac{m^2 - p^2x(1-x)}{4\pi\mu^2}\right] + \gamma_E\right\}. \tag{3.6.7}$$

We derived this by observing that

$$\begin{aligned}\mu^{2-d/2} &= e^{(2-d/2)\ln\mu} \\ &= 1 + (2 - d/2)\ln\mu + \tfrac{1}{2}(2 - d/2)^2\ln^2\mu + \cdots.\end{aligned}$$

This renormalization prescription, where counterterms are pure poles at the physical value of d, is called minimal subtraction (MS).

The unit of mass μ is entirely arbitrary. Thus the self-energy (3.6.7) now depends on three parameters instead of two. However a change of μ amounts to a change of renormalization prescription, so the change can be compensated, in this case, by a change in m. In effect minimal subtraction is a one-parameter family of renormalization prescriptions.

3.6.2 $d = 6$

We can also apply minimal subtraction to the six-dimensional theory. There we define

$$\left.\begin{aligned}\delta m^2 &= \text{poles at } d = 6, \\ \delta z &= \text{poles at } d = 6, \\ \delta g &= \mu^{3-d/2}(\text{poles at } d = 6).\end{aligned}\right\} \tag{3.6.8}$$

Note that the renormalized coupling is now dimensionless. Since

$$\Gamma(z-1) = -1/z + \gamma_E - 1 + O(z) \tag{3.6.9}$$

as $z \to 0$, we find that

$$\Sigma_{1R}^{(MS)} = \frac{-g^2}{128\pi^3}\left\{(m^2 - p/6)(\gamma_E - 1) + \int_0^1 dx[m^2 - p^2x(1-x)]\ln\left[\frac{m^2 - p^2x(1-x)}{4\pi\mu^2}\right]\right\} \tag{3.6.10}$$

and

$$\delta m^2 = \frac{g^2}{64\pi^3}\frac{m^2}{d-6} + O(g^4), \tag{3.6.11}$$

$$\delta Z = \frac{g^2}{6 \times 64\pi^3}\frac{1}{d-6} + O(g^4). \tag{3.6.12}$$

The counterterm δg for the coupling can also be calculated. From the graph of Fig. 3.6.1, we find (Macfarlane & Woo (1974))

$$\delta g = \mu^{3-d/2}\frac{g^3}{64\pi^3(d-6)} + O(g^5). \tag{3.6.13}$$

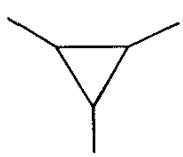


Fig. 3.6.1. One-loop vertex graph in ϕ^3 theory.

3.6.3 *Renormalization group and minimal subtraction*

When we discuss the renormalization group in Chapter 7, we will focus on one particular subgroup. The transformations in this subgroup consist of multiplying μ by a factor and making compensating changes in the renormalized coupling and mass. As a group it is trivial – being a representation of the group of positive real numbers under multiplication. What is non-trivial is the way in which it is represented in relation to the parametrization of the theory by a renormalized coupling and mass.

The renormalization group can be exploited in calculating high-energy behavior. While a full treatment will be made in Chapter 7, the basic idea can be seen by examining the one-loop self-energy. Let p^2 get large (with m and μ fixed), and consider the propagator defined using minimal subtraction at $d = 6$:

$$\bar{G}_2 \sim i\Big/\left(p^2\left\{1 - \frac{g^2}{768\pi^3}\left[\ln\left(\frac{-p^2}{\mu^2}\right) + \text{const.}\right] + O(g^4)\right\}\right). \tag{3.6.14}$$

To be able to make use of a perturbation expansion we must keep higher-order corrections small. But this is not so in (3.6.14) if p^2 is too large. The large correction can be avoided by setting μ^2 to be of order $|p^2|$. The theory is unchanged if we make suitable changes in g, in m, and in the scale of the field. We will learn how to do this in Chapter 7, with the result that the large corrections are effectively moved from higher-order terms in the perturbation series to the lowest-order graphs.

3.6.4 *Massless theories*

Let us return (for simplicity) to the self-energy of the four-dimensional theory. Consider the limit $m \to 0$. If we use mass-shell subtraction, we have (3.6.3), which diverges as $m_{\text{ph}}^2 \to 0$.

The divergence is an artifact of the mass-shell scheme, for which

$$\delta m_{\text{ph}}^2 = \frac{g^2}{2(4\pi)^{d/2}} \Gamma(2 - d/2) m_{\text{ph}}^{d-4} \int_0^1 dx (1 - x + x^2)^{d/2-2}$$

$$= \frac{g^2}{32\pi^2} \left[\frac{1}{2 - d/2} - \ln(m_{\text{ph}}^2) + \text{finite (as } m_{\text{ph}} \to 0, d \to 4) \right]. \quad (3.6.15)$$

In addition to the pole needed to cancel the UV divergence, there is a $\ln(m_{\text{ph}}^2)$ term.

Physically what happens is that in a massless theory there are long-range forces. These mean that separated particles are never completely free of each others' influence. Thus, for example, the singularity in the propagator is not a simple pole, for the self-energy (with MS subtraction) is, from (3.6.5),

$$\frac{g^2}{32\pi^2} \left[\ln\left(\frac{-p^2}{4\pi\mu^2} \right) + \gamma_{\text{E}} - 2 \right]. \quad (3.6.16)$$

The mass-shell renormalization prescription relies on the assumption of a simple propagator pole to generate counterterms, so it must fail. However, the nature of the propagator's singularity is an infra-red problem, so it is irrelevant to the question of whether an ultra-violet divergence can be renormalized. Some other renormalization scheme, like minimal subtraction, must be used in the massless theory.

3.7 Coordinate space

A good way to understand the infinite renormalizations is to work in coordinate space, as was emphasized by Bogoliubov & Shirkov (1980). This point of view is especially useful in treating field theories at finite

temperature or on a curved space-time background, as we will see in Chapter 11. There we will see why the counterterms are the same as at zero temperature in flat space-time.

For most ordinary calculations, it is cumbersome to work in coordinate space, because the propagator $S_F(x)$ for a free massive field is a Bessel function. In momentum space the propagator is simple: $i/(p^2 - m^2 + i\varepsilon)$. However the asymptotic behavior of $S_F(x)$ as $x \to 0$ is simple.

We have

$$S_F = \frac{\Gamma(d/2-1)}{4\pi^{d/2}(-x^2)^{d/2-1}} + \text{less singular as } x \to 0. \tag{3.7.1}$$

The one-loop correction to the propagator $G_2(x,y)$ is

$$\begin{aligned} G_{2,2} &= -(g^2/2)\int d^d z \int d^d w\, S_F(x-z) S_F(z-w)^2 S_F(w-y) \\ &= \int d^d z \int d^d w S_F(x-z)\hat{\Sigma}_1(z-w) S_F(w-y), \end{aligned} \tag{3.7.2}$$

where $\hat{\Sigma}_1(z-w)$ is the self-energy in coordinate space:

$$\hat{\Sigma}_1(z-w) = -(g^2/2) S_F(z-w)^2. \tag{3.7.3}$$

This is singular at $z = w$, and causes a logarithmic divergence in (3.7.2), where we integrate over all z and w. The fact that the divergence is from the region $z \simeq w$ means that it is in fact a δ-function:

$$\begin{aligned} G_{2,2} = \int d^d z S_F(x-z) S_F(z-y) \\ \cdot \left\{ \int_{w\sim 0} d^d w \left(\frac{-g^2}{32\pi^4}\right) \frac{\Gamma(d/2-1)^2}{[-(w-z)^2]^{d-2}} \right\} + \text{finite.} \end{aligned} \tag{3.7.4}$$

By Wick rotating the w^0-integral and using the following result (next chapter) for the d-dimensional integral of a Lorentz-invariant function

$$\int d^d w f(w^2) = -i[2\pi^{d/2}/\Gamma(d/2)] \int_0^\infty dw\, w^{d-1} f(w^2), \tag{3.7.5}$$

we find

$$G_{2,2} = \int d^d z S_F(x-z) S_F(z-y) \frac{ig^2}{16\pi^2(4-d)} + \text{finite}, \tag{3.7.6}$$

as $d \to 4$.

Evidently, the divergence is cancelled by adding a δ-function to $\hat{\Sigma}_1$:

$$\hat{\Sigma}_{1R} = \Sigma_1 - \frac{ig^2}{16\pi^2(4-d)} \delta^{(d)}(w-z), \tag{3.7.7}$$

which is exactly the mass counterterm computed earlier by momentum-space methods.

The important point, which is in fact true for an arbitrary graph, is that any UV divergence comes from a region in coordinate space where several interactions occur very close to each other. The divergence can then be cancelled by a counterterm which is a δ-function in the positions of these interactions. If the divergence is worse than logarithmic, then the counterterm will include derivatives of the δ-function. In any event the fact that it is a δ-function means that the counterterm can be included as a local interaction in the action. The locality means that it is a product of fields at the same point.

Since the singularity at $x = 0$ of the free propagator $S_F(x)$ is independent of the boundary conditions used to define it, we should expect, for example, that the counterterms used in thermal field theory are the same as at zero temperature. At non-zero temperature, the vacuum is replaced by a mixed state, and the boundary conditions for $S_F(x)$ change. The momentum-space proof that the counterterms are temperature-independent is therefore made difficult, but the coordinate-space proof is unchanged.

4

Dimensional regularization

We have seen how convenient it is to regulate the UV divergences of perturbation theory by continuation in the dimension of space-time. To date, no-one has shown how to use the method in the complete theory. But in perturbation theory, as we will now demonstrate, it is consistent and well-defined. Now all results obtained by this method can be obtained by other, more physical methods (say, a lattice regulator). But frequently much more labor is involved. This is not a triviality, for in complicated situations, especially in gauge theories, it enables us to handle the technicalities of renormalization in a simple way.

The idea of dimensional continuation has been used for a long time in statistical mechanics (see, for example, Fisher & Gaunt (1964)). It became very prominent when Wilson & Fisher (1972) discovered the ε-expansion and applied it to field-theoretic methods in statistical mechanics (Wilson (1973), Mack (1972), and Wilson & Kogut (1974)). In the ε-expansion one works in $4 - \varepsilon$ spatial dimensions, and expands in powers of ε. At the same time, in a purely field-theoretic context, a need arose to find a way of regulating non-abelian gauge theories that preserved gauge invariance and Poincaré invariance. This led to dimensional regularization ('t Hooft & Veltman (1972a), Bollini & Giambiagi (1972), Cicuta & Montaldi (1972), and Ashmore (1972)). Speer & Westwater (1971) had actually discovered the method earlier, but their paper is considerably more abstract, and had not attracted much attention.

Now vector spaces either have infinite dimension or a finite integer dimension. So the concept of integration on a space of finite non-integer dimension, d, cannot be taken completely literally. Either it is a set of purely formal rules for obtaining answers or it is an operation that is not literally integration in d dimensions, but only behaves in many respects as if it were integration in d dimensions. It is not sufficient to treat it only as a set of formal rules (even though that is what it becomes in practice), because one must know that the rules are consistent with one another and with the algebraic manipulations one carries out on integrals. To show that no inconsistencies can arise, we must construct an explicit definition.

There are three issues to address: (1) uniqueness, (2) existence, and (3) properties. Uniqueness is necessary, to avoid the possibility of constructing two definitions, each definition being self-consistent but giving different results from the other definition. Existence, shown by construction of an explicit definition, is necessary to prove that no inconsistencies arise. Once having seen that integration in non-integer dimension can be defined, we cannot just assume that all properties associated with ordinary integration are true; indeed they need not be.

So we also have to prove those properties which we need and which are true. We also must prove that the results agree with ordinary integration if d is an integer.

These considerations are quite non-trivial, as can be seen by considering, for example, the anomaly in the Ward identity for the axial current $j^\mu_{(5)} = \bar{\psi}\gamma^\mu\gamma_5\psi$ in the gauge model (2.11.7). If the fermion masses are zero, then a naive application of the fermion equations of motion shows that the current is conserved: $\partial_\mu j^\mu_{(5)} = 0$. In fact, the current is not conserved, as shown by Adler (1969, 1970) and Bell & Jackiw (1969). A counterterm can be added to $j^\mu_{(5)}$ to make it conserved, but only at the expense of removing its gauge invariance.

Among the objects to be extended to d dimensions are the Dirac matrices (γ^μ and γ_5). If we assumed the obvious generalization of their anticommutation relations, then for all values of d we would have

$$\left.\begin{aligned} \{\gamma^\mu, \gamma^\nu\} &= 2g^{\mu\nu}1, \\ \{\gamma_5, \gamma^\mu\} &= 0, \\ \gamma_5{}^2 &= 1. \end{aligned}\right\} \tag{4.0.1}$$

But then we would be able to derive the false result that the anomaly for the gauge-invariant axial current is zero. So there has to be an inconsistency ('t Hooft & Veltman (1972a)). More complicated problems in a similar vein arise when treating supersymmetric theories (Jones & LeVeille (1982)).

In this chapter we will start by stating the axioms for d-dimensional integration given by Wilson (1973). These are sufficient to prove uniqueness. Our calculation of a one-loop graph in Section 3.5 was in fact a realization of the uniqueness proof for one particular integral. Then we will construct an explicit definition of d-dimensional integration. The vector space on which we work is in fact infinite dimensional.

Unfortunately, the definition gives a divergent result in most cases, so we will next have to find a powerful enough extension (Section 4.2). We then prove some standard properties (Section 4.3). One particular result involves finding a definition of the metric tensor on an infinite-dimensional

space such that its trace is d rather than infinity.

Then we will be in a position to derive some useful formulae (Sections 4.4 and 4.5) for use in Feynman graph calculations. Finally we will show how to define Dirac matrices; this is obviously important if we are to be able to calculate consistently graphs containing the Adler–Bell–Jackiw anomaly.

The utility of a precise definition such as we give is that if inconsistencies arise at some stage, then one can always go back to first principles to discover the error.

4.1 Definition and axioms

Let d be a complex number. We wish to define an operation that we may regard as integration over a d-dimensional space:

$$\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}). \tag{4.1.1}$$

Here $f(\mathbf{p})$ is any given function of a vector $\mathbf{p}$, which is in the d-dimensional space. We will suppose that the space is Euclidean. (Minkowski space is regarded as a one-dimensional time together with a $(d-1)$-dimensional Euclidean space.) Following Wilson (1973) we will give an explicit definition in which the space is actually infinite dimensional; it is the integration operation that gives the dimensionality. Making d a positive integer n will effectively insert a δ-function in the integration that will force all vectors involved in defining the function $f(\mathbf{p})$ to lie in some n-dimensional subspace.

What properties must we impose on a functional of f in order to regard it as d-dimensional integration? The following properties or axioms (due to Wilson (1973)) are natural and are necessary in applications to Feynman graphs:

(1) Linearity: For any complex numbers a and b

$$\int \mathrm{d}^d\mathbf{p}[af(\mathbf{p}) + bg(\mathbf{p})] = a\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}) + b\int \mathrm{d}^d\mathbf{p} g(\mathbf{p}). \tag{4.1.2}$$

(2) Scaling: For any number s

$$\int \mathrm{d}^d\mathbf{p} f(s\mathbf{p}) = s^{-d}\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}). \tag{4.1.3}$$

(3) Translation invariance: For any vector $\mathbf{q}$

$$\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}+\mathbf{q}) = \int \mathrm{d}^d\mathbf{p} f(\mathbf{p}). \tag{4.1.4}$$

We will also require rotational covariance of our results.

Linearity is true of any integration, while translation and rotation invariance are basic properties of a Euclidean space, and the scaling property embodies the d-dimensionality.

Not only are the above three axioms necessary, but they also ensure that integration is unique, aside from an overall normalization (Wilson (1973)). In fact, they determine the usual integration measure in an integer-dimensional space (again up to normalization). The proof is simple:

Use linearity to expand $f(\mathbf{p})$ in terms of a set of basis functions. Choose a basis such as the functions

$$f_{s,\mathbf{q}}(\mathbf{p}) = \exp[-s^2(\mathbf{p}+\mathbf{q})^2]. \tag{4.1.5}$$

Then the integral of a basis function can be written in terms of the integral of one single function:

$$\int \mathrm{d}^d\mathbf{p} f_{s,\mathbf{q}}(\mathbf{p}) = s^{-d}\int \mathrm{d}^d\mathbf{p}\exp(-\mathbf{p}^2). \tag{4.1.6}$$

The integral of this one function sets the normalization. It is natural to require that the value be the usual one in integer dimensions and that we can write

$$\int \mathrm{d}^{d_1}\mathbf{p}\mathrm{d}^{d_2}\mathbf{q}\exp(-\mathbf{p}^2-\mathbf{q}^2) = \int \mathrm{d}^{d_1+d_2}\mathbf{k}\exp(-\mathbf{k}^2). \tag{4.1.7}$$

Thus the normalization is given by

$$\int \mathrm{d}^d\mathbf{p}\exp(-\mathbf{p}^2) = \pi^{d/2}. \tag{4.1.8}$$

An abstract uniqueness theorem is not sufficient for us. We also need an explicit formula so that a d-dimensional integral can be written as a sequence of ordinary integrals. This will be important in allowing us to prove standard properties of the integration. In addition it ensures that there exists a self-consistent definition. It is *a priori* possible that no consistent definition exists; the uniqueness theorem only applies if the integration operation exists.

A function $f(\mathbf{p})$ that we integrate could in principle be any function of the components of its vector argument. However, we do not, *a priori*, know the meaning of the components of, say, a vector in 3.99 dimensions. We will soon see that there are in fact infinitely many components. In practice, we will work with rotationally covariant functions. So we will assume that f is a tensor function of a finite set of vectors: $\mathbf{p}, \mathbf{q}_1, \ldots, \mathbf{q}_N$ say. For example, a scalar function is a function only of scalar products

$$f = f(p^2, \mathbf{p}\cdot\mathbf{q}_1, q_1{}^2, \ldots). \tag{4.1.9}$$

Thus f is in fact an ordinary function of scalar numbers, rather than some more complicated kind of function. Of course the values of the scalar products lie in restricted ranges. Thus:

$$\begin{aligned} \mathbf{q}_a^2 &\geq 0, \\ |\mathbf{q}_a \cdot \mathbf{q}_b| &\leq \mathbf{q}_a^2 \mathbf{q}_b^2. \end{aligned} \tag{4.1.10}$$

A tensor function is obtained by writing explicit tensors in terms of the vectors $\mathbf{p}, \mathbf{q}_1, \ldots, \mathbf{q}_N$ and of the metric tensor δ^{ij}, with scalar coefficients. For example, we might have

$$f^{ij}(\mathbf{p}, \mathbf{q}) = q^i p^j f_a(\mathbf{p}^2, \mathbf{p}\cdot\mathbf{q}, \mathbf{q}^2) + \delta^{ij} f_b(\mathbf{p}^2, \mathbf{p}\cdot\mathbf{q}, \mathbf{q}^2). \tag{4.1.11}$$

Such functions are the most general that we need to consider. (We will see later how to handle the antisymmetric tensor $\varepsilon_{\kappa\lambda\mu\nu}$ and the Dirac γ-matrices.)

To give a realization of the objects $\mathbf{p}, \mathbf{q}_1, \ldots$, we assume that they are vectors in an ordinary vector space. The space must be infinite dimensional, as we will show in a moment. So we define the vectors each to be an infinite sequence of components, $\mathbf{p} = (p^1, p^2, \ldots)$, just as we can define a three-dimensional vector $\vec{\mathbf{V}}$ as a sequence of three components (V^1, V^2, V^3). The metric is given by:

$$\mathbf{p}\cdot\mathbf{q} = p^1 q^1 + p^2 q^2 + \cdots.$$

The reason for the infinite dimensionality is that an integral with, say, $d = 3.99$ can be used not only as a regulator for a physical theory in a space-time of dimension $d_0 = 4$, but also as a regulator for a model theory in any higher dimension, e.g., $d_0 = 5$ or 6 or .... The vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots$ in (4.1.9) can be thought of as momenta of external particles, and our vector space must be large enough to accommodate d_0 linearly independent momenta. Since d_0 is arbitrary, we are forced to infinite dimension.

To define the d-dimensional integral of a scalar function, we find a finite-dimensional subspace containing all the $\mathbf{q}_j$'s. Then we write $\mathbf{p}$ as a component $\mathbf{p}_\parallel$ in this space and an orthogonal component $\mathbf{p}_\perp$:

$$\begin{aligned} \mathbf{p} &= \mathbf{p}_\parallel + \mathbf{p}_\perp \\ &= \sum_{j=1}^{J} p^j \mathbf{e}_j + \mathbf{p}_\perp. \end{aligned} \tag{4.1.12}$$

The 'parallel space', in which lie the $\mathbf{q}_j$'s, is spanned by an orthonormal basis $\mathbf{e}_j$ (with $j = 1, \ldots, J$). We define the integral over $\mathbf{p}$ to be the ordinary J-dimensional integral over $\mathbf{p}_\parallel$ performed after integration in $d - J$ dimensions over $\mathbf{p}_\perp$:

$$\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}) = \int \mathrm{d}p^1 \cdots \mathrm{d}p^J \int \mathrm{d}^{d-J}\mathbf{p}_\perp f(\mathbf{p}). \tag{4.1.13}$$

Since $f(\mathbf{p})$ does not depend on the direction of $\mathbf{p}_\top$ we now can define

$$\int \mathrm{d}^{d-J}\mathbf{p}_\top f(\mathbf{p}) = K_{d-J}\int_0^\infty \mathrm{d}p_\top p_\top^{d-J-1} f(\mathbf{p}). \tag{4.1.14}$$

Here K_ν (with $\nu = d-J$) is effectively the area of the surface of a hypersphere in ν dimensions. The value of K_ν is obtained by considering the special case where f is chosen to be a Gaussian – see (4.1.8) – with the result

$$K_\nu = \frac{2\pi^{\nu/2}}{\Gamma(\nu/2)}. \tag{4.1.15}$$

Hence we have a definition of d-dimensional integration in terms of ordinary integration:

$$\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}) = \frac{2\pi^{(d-J)/2}}{\Gamma((d-J)/2)}\int \mathrm{d}^J\mathbf{p}_\| \int_0^\infty \mathrm{d}p_\top p_\top^{d-J-1} f(\mathbf{p}). \tag{4.1.16}$$

We must check that the result is independent of the choice of the subspace of the $\mathbf{p}_\|$. We must extend the definition to handle the divergences at $\mathbf{p}_\top = 0$ when d is small, which we will do in Section 4.2. Then in Section 4.3 we will prove important properties of our definition. But first there are a couple of details to clear up.

The J-dimensional subspace of $\mathbf{p}_\|$'s is chosen subject only to the requirement that it include all $\mathbf{q}_j$'s. So it is possible to extend the space to include extra dimensions. To show this has no effect on the value of the integral we must prove

$$K_\nu \int_0^\infty \mathrm{d}p p^{\nu-1} g(p^2) = \int_{-\infty}^\infty \mathrm{d}k K_{\nu-1}\int_0^\infty \mathrm{d}p_\top p_\top^{\nu-2} g(p_\top^2 + k^2) \tag{4.1.17}$$

for any function, g, which depends on a scalar argument. This equation is true since the right-hand side is

$$\frac{2\pi^{(\nu-1)/2}}{\Gamma((\nu-1)/2)}\int_0^\infty \mathrm{d}p p^{\nu-1} g(p^2)\int_0^1 \mathrm{d}x x^{(\nu-3)/2}(1-x)^{-1/2}, \tag{4.1.18}$$

where $p_\top^2 = xp^2$ and $k^2 = (1-x)p^2$.

To show that different choices of the 'parallel' subspace have no effect on the value of the integral, we merely extend both spaces to a common larger space. The sole problem is that there may be a divergence in (4.1.18) at $x = 0$; this we will cover by Section 4.2.

Up till now we have supposed $f(\mathbf{p})$ is a scalar function. If it is a tensor $f^{ij\cdots}(\mathbf{p})$, we work component-by-component. To define, say, the component

$$\int \mathrm{d}^d\mathbf{p} f^{12}(\mathbf{p}),$$

we take the parallel space to include the 1- and 2-directions and any vectors $\mathbf{q}_j$ on which f^{ij} depends. Then we proceed as before.

For example, suppose $f^{ij}(\mathbf{p}) = p^i p^j g(\mathbf{p}^2)$, where g is a scalar function. Then

$$\int f^{12}(\mathbf{p})\mathrm{d}^d\mathbf{p} = \int \mathrm{d}p^1 \mathrm{d}p^2 \int \mathrm{d}^{d-2}\mathbf{p}_\top p^1 p^2 g[(p^1)^2 + (p^2)^2 + \mathbf{p}_\top^2] = 0,$$

while

$$\begin{aligned}\int \mathrm{d}^d\mathbf{p} f^{11}(\mathbf{p}^2) &= \int_{-\infty}^{\infty} \mathrm{d}p_\| \int \mathrm{d}^{d-1}\mathbf{p}_\top p_\|^2 g(p_\|^2 + \mathbf{p}_\top^2) \\ &= \frac{2\pi^{d/2}}{d\Gamma(d/2)} \int_0^\infty \mathrm{d}p p^{d+1} g(p^2) \\ &= \frac{1}{d}\int \mathrm{d}^d\mathbf{p} p^2 g(p^2).\end{aligned}$$

Generalizing this result, we see that

$$\int \mathrm{d}^d\mathbf{p} p^i p^j g(\mathbf{p}^2) = \frac{\delta^{ij}}{d}\int \mathrm{d}^d\mathbf{p} p^2 g(\mathbf{p}^2).$$

More general cases are treated in Section 4.3.

4.2 Continuation to small d

The convergence of the definition (4.1.16) is d-dependent at $p_\top = 0$ and $p_\top = \infty$. It improves at $p_\top = \infty$ when d gets smaller, but it improves at $p_\top = 0$ when d gets bigger. Even for a function that decreases exponentially at large p, and that is analytic for finite p, the defining integral has a divergence if the transverse space has a dimension $d - J \leq 0$; this is forced to happen if d is negative or zero. So our first task in this section is to find an explicit formula for the continuation of (4.1.16) to arbitrarily negative d. We will see that the $p_\top$-integral has poles whenever $(d - J)/2$ is zero or a negative integer, but that these are cancelled by the zeros in $1/\Gamma((d-J)/2)$.

We will then be able to adopt the resulting formula as a definition of the d-dimensional integral of a function for which (4.1.16) converges for no value of d. An example of such a function is

$$f(\mathbf{p}) = \frac{1}{(\mathbf{q}_1 + \mathbf{p})^2 + (\mathbf{q}_2 + \mathbf{p})^2 + m^2}.$$

The parallel space must be at least two-dimensional to accommodate $\mathbf{q}_1$ and $\mathbf{q}_2$, so we may set $J = 2$. Then the transverse integral converges at $p_\top = 0$ only if $d > J = 2$, while the complete integral converges at $p = \infty$ only if $d < 2$.

We will have a definition that defines $\int d^d\mathbf{p} f(\mathbf{p})$ for all small enough d. For larger values of d we define the integral by analytic continuation. In general there will be ultra-violet poles at certain values of d – just as in the Feynman graph we computed in Sections 3.5 and 3.6.

To explicitly define the continuation to small d, it is sufficient to consider a function $f(\mathbf{p}^2)$. Let us assume that $f \to 0$ rapidly enough as $p \to \infty$ that

$$\int d^d\mathbf{p} f(\mathbf{p}^2) \equiv \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty dp\, p^{d-1} f(p^2) \tag{4.2.1}$$

converges at $p \to \infty$ for some positive value of d. We also assume that $f(\mathbf{p}^2)$ is analytic at $p = 0$. Then (4.2.1) converges and is analytic in d for some range $0 < \mathrm{Re}\, d < d_{\max}$. We define the integral for all other values of d by analytic continuation in d. Explicit formulae for the continuation to smaller d's are constructed by adding and subtracting the leading behavior at $p \to 0$. For example, the following formula gives the integral in the range $-2 < \mathrm{Re}\, d < d_{\max}$:

$$\int d^d\mathbf{p} f(\mathbf{p}^2) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \left\{ \int_C^\infty dp\, p^{d-1} f(p^2) + \int_0^C dp\, p^{d-1} [f(p^2) - f(0)] + f(0) C^d/d \right\}. \tag{4.2.2}$$

This is independent of the arbitrary constant C.

When $-2 < \mathrm{Re}\, d < 0$ we may let $C \to \infty$ to obtain

$$\int d^d\mathbf{p} f(\mathbf{p}^2) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty dp\, p^{d-1} [f(p^2) - f(0)], \tag{4.2.3}$$

while at $d = 0$ the zero in $1/\Gamma(d/2)$ is cancelled by the pole term to give

$$\int d^0\mathbf{p} f(\mathbf{p}^2) = f(0). \tag{4.2.4}$$

We extend this procedure to continue to $-2l-2 < \mathrm{Re}\, d < -2l$ for any positive integer l:

$$\int d^d\mathbf{p} f(\mathbf{p}^2) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty dp\, p^{d-1} \{ f(p^2) - f(0) - p^2 f'(0) - \cdots - (p^2)^l f^{(l)}(0)/l! \},$$

$$\int d^{-2l}\mathbf{p} f(\mathbf{p}^2) = (-\pi)^{-l} f^{(l)}(0). \tag{4.2.5}$$

This equation gives us the integral when $-2l-2 < \mathrm{Re}\, d < -2l$ on the

assumption that the original formula (4.2.1) converges when d is just greater than zero. Suppose now that (4.2.1) diverges at $p = \infty$ for all positive values of d, but that f is power behaved as $p \to \infty$. Then it is sensible to adopt (4.2.5) as the definition of the integral. This particular definition is very important since we will use dimensional continuation to regulate Feynman graphs that are ultra-violet divergent at $d = 4$. The definition (4.1.16) applied to a Feynman graph frequently has a negative number $d - 4$ of transverse dimensions in order to ensure ultra-violet convergence of the complete integral. Then we may apply the definition (4.2.5) to the $p_\top$-integral with d replaced by $d - J$.

Another obstacle to continuation in d is sometimes that $f(p^2)$ is not analytic at $p^2 = 0$ but has a power-law singularity there. We may generalize the derivation of (4.2.5) to write down a formula for the continuation of the integral.

An example of the use of (4.2.5) as a definition is given by choosing

$$f(p^2) = (p^2 + A)/(p^2 + B),$$

where A and B are numbers. The definition (4.2.1) diverges for all d, but with $l = 1$, (4.2.5) gives us a definition valid for $-2 < \mathrm{Re}\, d < 0$:

$$\int \mathrm{d}^d\mathbf{p} f(\mathbf{p}) = \frac{\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty \mathrm{d}p^2 p^{d-2} \left[\frac{p^2 + A}{p^2 + B} - \frac{A}{B} \right].$$

The integral can be explicitly computed to give:

$$\int \mathrm{d}^d\mathbf{p}(\mathbf{p}^2 + A)/(\mathbf{p}^2 + B) = (\pi B)^{d/2}(A/B - 1)\Gamma(1 - d/2),$$

which can be continued to all d.

Suppose f has a power-law singularity, as for example

$$f = \frac{1}{(\mathbf{p} + \mathbf{q})^2(\mathbf{p}^2 + m^2)}.$$

The definition (4.1.16) of the integral of this function converges if $2 < d < 4$. To continue it to lower d we must subtract the power behavior at $\mathbf{p} = -\mathbf{q}$, just as we did for singularities at $p_\top = 0$, or at $p = 0$ in (4.2.2). Then we can define the integral of say

$$\frac{\mathbf{p}^2}{(\mathbf{p} + \mathbf{q})^2(\mathbf{p}^2 + m^2)}.$$

One result of all these definitions is that the integral of a power of p is

zero:

$$\int d^d\mathbf{p}(\mathbf{p}^2)^\alpha = 0 \tag{4.2.6}$$

for any value of α (integer or not). It should not be thought that there is any choice in (4.2.6). It follows from (*a*) the explicit continuation of (4.1.16) to small d, and (*b*) application of the continued formula as a definition of $\int d^d\mathbf{p}(\mathbf{p}^2)^\alpha$.

Consistency of the formalism also requires (4.2.6). For suppose that $f(\mathbf{p}) = (\mathbf{p}^2)^\alpha/(\mathbf{p}^2 + m^2)$. Then when $-2\alpha - 2 < d \leq -2\alpha$ we have

$$\int d^d\mathbf{p} f(\mathbf{p}^2) = \int d^d\mathbf{p}[f(\mathbf{p}^2) - (\mathbf{p}^2)^\alpha/m^2].$$

If linearity is to be true then we have (4.2.6).

We could subtract out the singularity differently, by a function that is not just a power of p. But then, for example, the simplification obtained in (4.2.2) by taking $C \to \infty$ would no longer occur.

Observe that if in the first of our definitions (4.1.16), we take f to be a positive-definite function, then the integral is positive. But when the integral is continued away from the region where this definition converges, then the subtraction terms mean that the integrand is no longer positive definite, so that the integral need not be positive.

At the end of Section 4.1, we proved that the value of a d-dimensional integral does not depend on how we split the integral into an ordinary integral over some integer-dimensional 'parallel space' and a spherically symmetric integral over the remaining dimensions. We let J be the dimension of the parallel space. Then the proof consists of examining what happens when J is increased by one. Ultimately we had to prove (4.1.17), which is a property of ordinary integrals. We assumed $d > J$, so that there were no subtraction terms. To generalize the result to the case that $d - J$ is not positive, we must prove that

$$\begin{aligned} K_{d-J}\int_0^\infty dp\, p^{d-J-1}\left[f(p^2) - \sum_{n=0}^{[J/2-d/2]} f^{(n)}(0)p^{2n}/n!\right] \\ = K_{d-J-1}\int_{-\infty}^{\infty} dk \int_0^\infty dp_\top p_\top^{d-J-2}\left[f(k^2+p_\top^2)\right. \\ \left. - \sum_{n=0}^{[(J+1-d)/2]} f^{(n)}(k^2)p_\top^{2n}/n!\right]. \end{aligned} \tag{4.2.7}$$

Here the symbol $[a]$ denotes the largest integer smaller than a. To prove the equation we change variables on the right-hand side to x and p, where

$p_1^2 = xp^2$ and $k^2 = (1-x)p^2$. For the right-hand side we get

$$K_{d-J-1}\int_0^1 dx\, x^{(d-J-3)/2}(1-x)^{-1/2}\int_0^{(x)} dp\, p^{d-J-1}\times$$
$$\times\left[f(p^2)-\sum_{n=0}^{[(J+1-d)/2]} f^{(n)}((1-x)p^2)x^n p^{2n}/n!\right]$$
$$=\tfrac{1}{2}K_{d-J-1}\int_0^1 dx\, x^{(d-J-3)/2}(1-x)^{-1/2}\int_0^\infty dp^2 (p^2)^{d/2-J/2-1}\times$$
$$\times\left\{\left[f(p^2)-\sum_{n=0}^{[J/2-d/2]} f^{(n)}(0)p^{2n}/n!\right]\right.$$
$$\left.-\sum_{n=0}^{[(J+1-d)/2]}\left[f^{(n)}((1-x)p^2)\frac{x^n p^{2n}}{n!}-\sum_{j=n}^{[(J-d)/2]} f^{(j)}(0)\frac{(1-x)^{j-n}x^n p^{2j}}{n!(n-j)!}\right]\right\}. \tag{4.2.8}$$

Here we have added and subtracted

$$\sum_{n=0}^{[(J-d)/2]} f^{(n)}(0)p^{2n}/n!,$$

so that the integral over p^2 of p^{d-J-2} times each square bracket term is convergent. After scaling p^2 by $(1-x)$, we get:

$$\tfrac{1}{2}K_{d-J-1}\int_0^1 dx\, x^{(d-J-3)/2}(1-x)^{-1/2}\int_0^\infty dp^2 (p^2)^{d/2-J/2-1}\times$$
$$\times\left\{\left[f(p^2)-\sum_{n=0}^{[J/2-d/2]} f^{(n)}(0)p^{2n}/n!\right]\right.$$
$$-\sum_{n=0}^{[(J+1-d)/2]}\frac{x^n(1-x)^{J/2-d/2-n}p^{2n}}{n!}$$
$$\left.\left[f^{(n)}(p^2)-\sum_{j=n}^{[(J-d)/2]} f^{(j)}(0)\frac{p^{2j}}{(n-j)!}\right]\right\}.$$

Integration by parts in the p-integral gives

$$K_{d-J-1}\int_0^1 dx\, x^{(d-J-3)/2}(1-x)^{-1/2}\times$$
$$\times\left[1-\sum_{n=0}^{[(J+1-d)/2]} x^n(1-x)^{J/2-d/2-n}\frac{(1+J/2-d/2)!}{n!(1+J/2-d/2-n)!}\right]\times$$
$$\times\int_0^\infty dp\, p^{d-J-1}\left[f(p^2)-\sum_{n=0}^{[J/2-d/2]} f^{(n)}(0)p^{2n}/n!\right].$$

An integration by parts on the x-integral is used to show that it equals $\Gamma(1/2)\Gamma[(d-J-1)/2]/\Gamma[(d-J)/2]$, from which the required result follows.

4.3 Properties

Property 1. *Axioms*: The definitions (4.1.16) and (4.2.5) satisfy Wilson's axioms (4.1.2), etc., for d-dimensional integration.

Proof. We reduced d-dimensional integration to ordinary integration so linearity follows from linearity of ordinary integration. We must choose the $\mathbf{p}_\parallel$ space to be large enough that it is the same for both functions f and g in (4.1.2). Our explicit continuation (4.2.5) to arbitrary negative d ensures that reducing the dimension of the transverse space is no problem.

Scaling and rotation covariance are explicit properties of all our definitions.

Translation invariance is valid for ordinary integration, so it follows from definition (4.1.16) provided the $\mathbf{p}_\parallel$ space is big enough to include the vector $\mathbf{q}$ used in the axiom (4.1.4).

Property 2.

$$\int \mathrm{d}^d\mathbf{p} \frac{(\mathbf{p}^2)^\alpha}{(\mathbf{p}^2+M^2)^\beta} = \pi^{d/2} M^{d+2\alpha-2\beta} \frac{\Gamma(\alpha+d/2)\Gamma(\beta-\alpha-d/2)}{\Gamma(d/2)\Gamma(\beta)}. \tag{4.3.1}$$

Proof. Immediate from (4.2.5). Note that this implies that the integral of a power of p^2 is zero, since $\Gamma(\beta) \sim 1/\beta$ as $\beta \to 0$.

Property 2*a*.

$$\int \mathrm{d}^d\mathbf{p}(\mathbf{p}^2)^\alpha = 0. \tag{4.3.1a}$$

Proof. Already done.

Property 3.

$$\frac{\partial}{\partial \mathbf{q}} \int \mathrm{d}^d\mathbf{p} f(\mathbf{p},\mathbf{q},\ldots) = \int \mathrm{d}^d\mathbf{p} \frac{\partial}{\partial \mathbf{q}} f(\mathbf{p},\mathbf{q},\cdots). \tag{4.3.2}$$

Proof. Contract with a vector $\delta\mathbf{q}$ which projects out the derivative with respect to the component of $\mathbf{q}$ in the $\delta\mathbf{q}$ direction. Then make the parallel space (of $\mathbf{p}_\parallel$'s) big enough to include $\delta\mathbf{q}$ and use (4.3.2) on ordinary integrals. This is true for all $\delta\mathbf{q}$.

Property 4.

$$\delta^{ij}\delta_{ij} = d. \tag{4.3.3}$$

Proof and definition of δ_{ij}. Now δ^{ij} is defined to be the component form of a contravariant tensor with $\delta^{ij} = 1$ if $i = j$ and zero otherwise. The obvious definition of the covariant tensor δ_{ij} is as the inverse matrix, i.e., the same thing. This gives $\delta^{ij}\delta_{ij} = \infty$. However in an infinite-dimensional space, there is space for a different definition.

A contravariant tensor may be defined by specifying its components. But a covariant tensor ω is fundamentally a linear function acting on covariant tensors: $\omega(\mathbf{T})$. We can write $\omega(\mathbf{T}) = \omega_{ij}T^{ij}$ only if the sum converges.

We need the covariant δ (which we symbolize by δ_{ij}) to be rotation invariant, and to give $\delta(\mathbf{T}) = T^{ii}$ whenever the sum exists. We would also like contraction with δ_{ij} to commute with integration. For example

$$\left.\begin{aligned} \delta_{ij}\int \mathrm{d}^d\mathbf{p}\, p^i p^j f(\mathbf{p}^2) &= \delta_{ij}\int \mathrm{d}^d\mathbf{p}\,\delta^{ij}\mathbf{p}^2 f(\mathbf{p}^2)/d \\ \text{and} \qquad & \\ \delta_{ij}\int \mathrm{d}^d\mathbf{p}\, p^i p^j f(\mathbf{p}^2) &= \int \mathrm{d}^d\mathbf{p}\,\mathbf{p}^2 f(\mathbf{p}^2). \end{aligned}\right\} \tag{4.3.4}$$

Since we have an infinite sum, we cannot immediately apply linearity to prove this equation.

Let us define

$$\delta(\mathbf{T}) = \frac{d\Gamma(d/2)}{\pi^{d/2}}\int \mathrm{d}^d\mathbf{p}\sum_{i,j} T^{ij}p_i p_j \delta(\mathbf{p}^2 - 1). \tag{4.3.5}$$

Whenever $\sum T^{ii}$ converges, this definition gives

$$\delta(\mathbf{T}) = \sum T^{ii}$$

But if the sum diverges, then it is possible to get a finite value for $\delta(\mathbf{T})$. In particular,

$$\delta(\delta^{ij}) = \frac{d\Gamma(d/2)}{\pi^{d/2}}\int \mathrm{d}^d\underline{\mathbf{p}}\,\underline{\mathbf{p}}^2\delta(\underline{\mathbf{p}}^2 - 1) = d,$$

as required. The definition is rotationally invariant. Commutation of contraction with δ_{ij} and integration will now be a consequence of commutativity of two integrals – which we will prove later.

Property 5. Integration by parts:

$$\int \mathrm{d}^d\mathbf{p}\,\partial f(\mathbf{p})/\partial p^i = 0. \tag{4.3.6}$$

Proof. Work component-by-component. Contract with an arbitrary vector k:

$$\int d^d\mathbf{p} k^i \frac{\partial}{\partial p^i} f.$$

Then, to define this integral, we must put $\mathbf{k}$ in the parallel space, and we can use the proof of (4.3.6) for ordinary integration in the space parallel to $\mathbf{k}$.

Property 6. To define integration over two (or more) variables:

$$\int d^d\mathbf{p} d^d\mathbf{k} f(\mathbf{p}, \mathbf{k}; \mathbf{q}_1, \ldots \mathbf{q}_N)$$

we must choose to calculate one integral then the other, according to the rules already stated.

For this definition to be sensible we need the result to be independent of the order of integration:

$$\int d^d\mathbf{p} \int d^d\mathbf{k} f = \int d^d\mathbf{k} \int d^d\mathbf{p} f. \tag{4.3.7}$$

We could also allow the dimensions of the $\mathbf{p}$- and $\mathbf{k}$-integrals to be different. Then exchange of order of integration $\int d^d\mathbf{p} \int d^{d'}\mathbf{k} \to \int d^{d'}\mathbf{k} \int d^d\mathbf{p}$. is allowed only if $d = d'$, or if f is independent of $\mathbf{p}\cdot\mathbf{k}$.

Proof. It is sufficient to consider the case that there are no $\mathbf{q}_i$'s, so that $f = f(\mathbf{p}^2, \mathbf{p}\cdot\mathbf{k}, \mathbf{k}^2)$. (If there are $\mathbf{q}_i$'s, then we take out a finite-dimensional integral for both $\mathbf{k}$ and $\mathbf{p}$ which spans all $\mathbf{q}_i$'s and then we apply the theorem to the remaining dimensions.)

The left-hand side of (4.3.7) is

$$\frac{4\pi^{d-1/2}}{\Gamma(d/2)\Gamma((d-1)/2)} \int_0^\infty dp \int_{-\infty}^\infty dk_1 \int_0^\infty dk_\perp p^{d-1} k_\perp^{d-2} f(p^2, k_1 p, k_1^2 + k_\perp^2), \tag{4.3.7L}$$

while the right-hand side is

$$\frac{4\pi^{d-1/2}}{\Gamma(d/2)\Gamma((d-1)/2)} \int_0^\infty dk \int_{-\infty}^\infty dp_1 \int_0^\infty dp_\perp p_\perp^{d-2} k^{d-1} f(p_1^2 + p_\perp^2, p_1 k, k^2). \tag{4.3.7R}$$

Here k_1 is the component of $\mathbf{k}$ parallel to $\mathbf{p}$, while p_1 is the component of $\mathbf{p}$ parallel to $\mathbf{k}$. Change variables to, say, $\mathbf{p}^2$, $\mathbf{k}^2$, and $z = \mathbf{p}\cdot\mathbf{k}/(pk) = p_1/\sqrt{(p_1^2 + p_\perp^2)} = k_1/\sqrt{(k_1^2 + k_\perp^2)}$, with the result that both (4.3.7L) and (4.3.7R) are equal to

$$\frac{4\pi^{d-1/2}}{\Gamma(d/2)\Gamma((d-1)/2)} \int_0^\infty dp\, p^{d-1} \int_0^\infty dk\, k^{d-1} \int_{-1}^1 dz (1-z^2)^{(d-3)/2} f(p^2, pkz, k^2). \tag{4.3.7C}$$

The theorem is thus proved in the case that (4.3.7L) and (4.3.7R) are both convergent. Note that it is not a trivial consequence of the corresponding result for integer-dimensional integration.

If the dimensions of the integrations are not the same, then let the $\mathbf{k}$ integral have dimension d'. The left-hand side gives

$$\frac{4\pi^{(d+d'-1)/2}}{\Gamma(d/2)\Gamma((d'-1)/2)}\int_0^\infty \mathrm{d}p\, p^{d-1}\int_0^\infty \mathrm{d}k\, k^{d'-1}\int_{-1}^{1}\mathrm{d}z(1-z^2)^{(d'-3)/2}f,$$

which in general is not the same as the corresponding expression for the right-hand side. But if f is independent of z, then the z-integral can be computed explicitly. The result is

$$\int \mathrm{d}^{d'}\mathbf{k}\mathrm{d}^{d}\mathbf{p}\, f(\mathbf{p}^2,\mathbf{k}^2)=\frac{4\pi^{(d+d')/2}}{\Gamma(d/2)\Gamma(d'/2)}\int_0^\infty \mathrm{d}p\, p^{d-1}\int_0^\infty \mathrm{d}k\, k^{d'-1}f(p^2,k^2). \tag{4.3.8}$$

A problem is that if d is not positive, we must make subtractions as in (4.2.5). These are clearly asymmetric between the two orders (4.3.7L) and (4.3.7R) of performing the original integral (with now $d'=d$); in practice, d will be the number of dimensions transverse to the external vectors $\mathbf{q}_1,\ldots,\mathbf{q}_N$. In applications to Feynman graphs d will therefore be negative in order to regulate UV divergences. So we must use (4.2.5) to define the integrals. Then (4.3.7) does not give (4.3.7L), (4.3.7R) and (4.3.7C).

We solve this problem by defining an auxiliary integral with a convergence factor, say

$$I(a,d)=\int \mathrm{d}^{d}\mathbf{p}\int \mathrm{d}^{d'}\mathbf{k}\, f(\mathbf{p},\mathbf{k})\exp[-a(\mathbf{p}^2+\mathbf{k}^2)]. \tag{4.3.9}$$

Assume f is power-behaved at infinity. Then for all d, (4.3.7L or R) is UV convergent. Moreover, if $d>1$ then both (4.3.7L) and (4.3.7R) are IR convergent without subtractions. The function $I(a,d)$ is analytic in a and d. Continue down to small enough d that (4.3.7) is UV convergent. Then $I(a,d)$ is given both by (4.3.7L) and (4.3.7R) with f replaced by $f\exp[-a(\mathbf{p}^2+\mathbf{k}^2)]$, and with subtractions made. Now set $a=0$ to prove the theorem (4.3.7).

Property 7.

$$\int \mathrm{d}^{d}\mathbf{k}\int \mathrm{d}^{d'}\mathbf{p}\, f(\mathbf{p}^2+\mathbf{k}^2)=\int \mathrm{d}^{d+d'}\mathbf{q}\, f(\mathbf{q}^2). \tag{4.3.10}$$

Proof. Since f is independent of $\mathbf{p}\cdot\mathbf{k}$, the previous theorem shows that the left-hand side is independent of the order of integration, even if the

dimensions of the **p**- and **k**-integrals are different. Then use (4.3.9) and change variables to $q=(\mathbf{p}^2+\mathbf{k}^2)^{1/2}$ and $x=\mathbf{p}^2/\mathbf{q}^2$.

Property 8.

$$\int \mathrm{d}^d\mathbf{p}\, p^{i_1}\cdots p^{i_t} g(\mathbf{p}^2)=\begin{cases}0, & \text{if } t \text{ is odd,}\\ T^{i_1\ldots i_t}A_t[g], & \text{if } t \text{ is even,}\end{cases} \tag{4.3.11}$$

with

$$T^{i_1\ldots i_t}=[\delta^{i_1 i_2}\delta^{i_3 i_4}\cdots\delta^{i_{t-1}i_t}+\cdots + \text{all permutations of the } i\text{'s}]/t!, \tag{4.3.12}$$

and

$$\begin{aligned}A_t(g)&=\frac{\Gamma(d/2)\Gamma(t/2+1/2)}{\Gamma(1/2)\Gamma(d/2+t/2)}\int \mathrm{d}^d\mathbf{p}(\mathbf{p}^2)^{t/2}g(\mathbf{p}^2)\\ &=\frac{2\pi^{d/2}\Gamma(t/2+1/2)}{\Gamma(1/2)\Gamma(d/2+t/2)}\int_0^\infty \mathrm{d}p\, p^{d+t-1}g(p^2).\end{aligned} \tag{4.3.13}$$

Proof. If t is odd, antisymmetry of the integrand under $\mathbf{p}\to-\mathbf{p}$ makes the integral over the 'parallel' space zero.

Antisymmetry under reversal of one component of **p**, symmetry under permutations of the i's, and rotation invariance give the general form (4.3.11) and (4.3.12). Computation of one component (say, $i_1=i_2=\cdots=i_t=1$) then gives (4.3.13).

Examples.

$$\int \mathrm{d}^d\mathbf{p}\, p^i p^j g(\mathbf{p}^2)=(1/d)\delta^{ij}\int \mathrm{d}^d\mathbf{p}\,\mathbf{p}^2 g(\mathbf{p}^2), \tag{4.3.14}$$

$$\int \mathrm{d}^d\mathbf{p}\, p^i p^j p^k p^l g(\mathbf{p}^2)=\frac{(\delta^{ij}\delta^{kl}+\delta^{ik}\delta^{jl}+\delta^{il}\delta^{jk})}{d(d+2)}\int \mathrm{d}^d\mathbf{p}\,|\mathbf{p}|^4 g(\mathbf{p}^2). \tag{4.3.15}$$

Property 9. Consider an integral

$$I(\mathbf{p}_1,\ldots,\mathbf{p}_J)=\int \mathrm{d}^d\mathbf{k}\, f(\mathbf{k},\mathbf{p}_1,\ldots,\mathbf{p}_J) \tag{4.3.16}$$

which is UV convergent by power-counting at $d=4$; that is $f=O(1/k^{4+a})$ as **k** goes to infinity in any direction, for some positive number a. Then the integral is analytic in d and in the parameters $\mathbf{p}_i$, when d is close to four, if the integrand is analytic. If the $\mathbf{p}_i$'s lie in the first four dimensions, then the integral at $d=4$ has the same value as the ordinary four-dimensional integral of f.

Example Suppose f has the form

$$f = \bar{f}(\bar{\mathbf{k}}, \bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_J) \exp(-A\hat{\mathbf{k}}^2 + 2A\hat{\mathbf{k}}\cdot\hat{\mathbf{p}}), \tag{4.3.17}$$

where for any vector $\mathbf{v}$ we let $\bar{\mathbf{v}}$ be its projection onto the first four dimensions and let $\hat{\mathbf{v}}$ be its projection onto the remaining dimensions. Then we let $\bar{I}$ be the ordinary four-dimensional integral of $\bar{f}$. By use of our definitions of d-dimensional integration, we have

$$\begin{aligned} I &= \int \mathrm{d}^d\mathbf{k} f \\ &= \bar{I} \int \mathrm{d}^{d-4}\hat{\mathbf{k}} \exp(-A\hat{\mathbf{k}}^2 + 2A\hat{\mathbf{p}}\cdot\hat{\mathbf{k}}) \\ &= \bar{I}(\pi/A)^{d/2-2} \exp(A\hat{\mathbf{p}}^2). \end{aligned} \tag{4.3.18}$$

This is manifestly analytic in d and $\hat{\mathbf{p}}$. If we set $d = 4$, the integral becomes

$$I = \bar{I} \exp(A\hat{\mathbf{p}}^2).$$

Notice that there is no restriction on $\hat{\mathbf{p}}$, even though $\hat{\mathbf{p}} = 0$ in four dimensions. However, if we let $d \to 4$ and $\hat{\mathbf{p}} \to 0$, the limit is smooth.

Proof of Property 9. The proof is easily made by examining the definition of the d-dimensional integral in terms of ordinary integrals. As usual we divide the space into a finite-dimensional parallel space big enough to contain $\mathbf{p}_1, \ldots, \mathbf{p}_J$, and into a transverse space containing the remaining dimensions. It is convenient to choose the parallel space to have an odd number $2N + 1$ of dimensions. Then:

$$I = \int_{-\infty}^{\infty} \mathrm{d}k_1 \cdots \int_{-\infty}^{\infty} \mathrm{d}k_{2N+1} \frac{\pi^{(d-1)/2-N}}{\Gamma((d-1)/2-N)} \int_0^{\infty} \mathrm{d}k_\top^2 (k_\top^2)^{(d-3)/2-N} \times$$
$$\times \left[F(k_1, k_2, \ldots, k_{2N+1}, k_\top^2) - \sum_{n=0}^{N-2} F^{(n)}(k_1, \ldots, k_{2N+1}, 0) k_\top^{2n}/n! \right]. \tag{4.3.19}$$

Here we have used F to denote f considered as a function of the first $2N + 1$ components of $\mathbf{k}$ and of $k_\top^2$. Since f is an analytic function of $\mathbf{k}$, it can be expanded in powers of $k_\top^2$.

As required by the definition, we have subtracted off a power series in $k_\top^2$, to give convergence at $k_\top = 0$. We use $F^{(n)}$ to denote the nth derivative of F with respect to $k_\top^2$. The integrand of the $k_\top^2$ integral behaves as $k_\top^{d-5}$, so we have convergence at $k_\top = 0$ if $d > 3$. Since f is analytic, there are no other singularities at finite k, and the only other possible source of a divergence is from large k. The subtractions do not introduce a divergence provided that $d < 5$. Moreover, we have assumed that $f = O(1/k^{4+a})$ as $k \to \infty$, so that there are no other large k divergences when d is close to four. Hence (4.3.19)

converges and is analytic in a neighborhood of $d = 4$.

Now the integral will depend on the $\mathbf{p}_i$'s only through the Lorentz scalars $\mathbf{p}_a \cdot \mathbf{p}_b$ (with $1 \leq a \leq b \leq J$). To determine this dependence, it is sufficient to keep the $\mathbf{p}_i$'s within some fixed J-dimensional subspace. Since (4.3.19) is a perfectly finite integral, it is an analytic function of the $\mathbf{p}_i$'s.

To determine the value of the integral when $d = 4$ and when the $\mathbf{p}_i$'s are in the first four dimensions, we use the freedom to vary the dimension of the 'parallel' space in the definition (4.3.19). Let us now make it four dimensional. We will obtain an integral of the form:

$$I = \int_{-\infty}^{\infty} dk_1 \ldots dk_4 \frac{2\pi^{d/2-2}}{\Gamma(d/2-2)} \int_0^{\infty} d\hat{k} \hat{k}^{d-5} \hat{f}(k_1, k_2, k_3, k_4, \hat{k}^2) \quad (4.3.20)$$

if $d > 4$. When we let $d \to 4$, the integral over $\hat{k}$ is singular at $\hat{k} = 0$; the resulting divergence cancels the zero of the inverse Γ-function to give

$$I(d = 4) = \int dk_1 dk_2 dk_3 dk_4 \hat{f}(k_1, k_2, k_3, k_4, 0), \quad (4.3.21)$$

as required.

We may alternatively continue from $d < 4$ using

$$I = \int dk_1 \ldots dk_4 \frac{2\pi^{d/2-2}}{\Gamma(d/2-2)} \int_0^{\infty} d\hat{k} \hat{k}^{d-5} [\hat{f}(k_1, \ldots, k_4, \hat{k}^2) - \hat{f}(k_1, \ldots, k_4, 0)]. \quad (4.3.22)$$

The singularity at $\hat{k} = 0$ is cancelled, but as $d \to 4$ we get a divergence at $\hat{k} = \infty$ which gives the same result (4.3.21).

Comment In this proof we used the freedom to alter the dimension of the parallel space. To show that the integral is well-behaved at $d = 4$, it was convenient to choose the parallel space to have an odd dimension. But to compute the actual value at $d = 4$, it was convenient to choose the parallel space to have an even dimension, specifically, four. It is instructive to see the equivalence in a simple non-trivial case. (The general case was summarized at the end of Section 4.2.)

Suppose $3 < d < 4$. Then define

$$I_1 = \frac{\pi^{d/2-2}}{\Gamma(d/2-2)} \int_0^{\infty} d\hat{k}^2 (\hat{k}^2)^{d/2-3} f(k_1, k_2, k_3, k_4, \hat{k}^2), \quad (4.3.23)$$

$$I_2 = \int_{-\infty}^{\infty} dk_5 \frac{\pi^{(d-5)/2}}{\Gamma(\frac{1}{2}(d-5))} \int_0^{\infty} dk_T^2 (k_T^2)^{(d-7)/2} [f(k_1, \ldots, k_4, k_5^2 + k_T^2) - f(k_1, \ldots, k_4, k_5^2)]. \quad (4.3.24)$$

Our definition of the d-dimensional integral of f tells us that

$$I = \int_{-\infty}^{\infty} \mathrm{d}k_1 \ldots \mathrm{d}k_4 I_1 = \int \mathrm{d}k_1 \ldots \mathrm{d}k_4 I_2, \tag{4.3.25}$$

so we must prove that $I_1 = I_2$.

To do this we change variables in I_2, by setting $k_\mp^2 = x\hat{k}^2$ and $k_5^2 = (1-x)\hat{k}^2$ to obtain:

$$\begin{aligned} I_2 = {} & \frac{\pi^{(d-5)/2}}{\Gamma(d/2-5/2)} \int_0^1 \mathrm{d}x \int_0^\infty \mathrm{d}\hat{k}^2 (\hat{k}^2)^{d/2-3} (1-x)^{-1/2} x^{d/2-7/2} \times \\ & \times [f(k_1,\ldots,k_4,\hat{k}^2) - f(k_1,\ldots,k_4,\hat{k}^2(1-x))] \\ = {} & \frac{\pi^{(d-5)/2}}{\Gamma(d/2-5/2)} \int_0^1 \mathrm{d}x \int_0^\infty \mathrm{d}\hat{k}^2 \hat{k}^{d-6} (1-x)^{-1/2} x^{d/2-7/2} \times \\ & \times \{[f(k_1,\ldots,k_4,\hat{k}^2) - f(k_1,\ldots,k_4,0)] \\ & - [f(k_1,\ldots,k_4,\hat{k}^2(1-x)) - f(k_1,\ldots,k_4,0)]\}. \end{aligned} \tag{4.3.26}$$

In the last line we subtracted and added $f(k_1,\ldots,k_4,0)$, so that we can integrate seperately each term in square brackets. In particular, we have

$$\begin{aligned} & \int \mathrm{d}\hat{k}^2 \hat{k}^{d-6} [f(k_1,\ldots,k_4,\hat{k}^2(1-x)) - f(k_1,\ldots,k_4,0)] \\ & \quad = (1-x)^{2-d/2} \int \mathrm{d}\hat{k}^2 \hat{k}^{d-6} [f(k_1,\ldots,k_4,\hat{k}^2) - f(k_1,\ldots,k_4,0)]. \end{aligned}$$

Comparison with the definition (4.3.23) of I_1 shows that

$$\begin{aligned} \frac{I_2}{I_1} = {} & \frac{\pi^{-1/2}\Gamma(d/2-2)}{\Gamma(d/2-5/2)} \int_0^1 \mathrm{d}x [(1-x)^{-1/2} x^{d/2-7/2} \\ & - x^{d/2-7/2}(1-x)^{3/2-d/2}]. \end{aligned} \tag{4.3.27}$$

The integral is in fact the analytic continuation from $d > 5$ of a beta-function, so that it equals $\Gamma(d/2-5/2)\Gamma(1/2)/\Gamma(d/2-2)$. The required result $I_1 = I_2$ follows.

Property 10. Multiple integrals are correct at $d = 4$.

Consider the integral

$$I(\mathbf{p}_1,\ldots,\mathbf{p}_N) = \int \mathrm{d}^d\mathbf{k}_1 \ldots \mathrm{d}^d\mathbf{k}_L f(\mathbf{k}_1,\ldots,\mathbf{k}_L,\mathbf{p}_1,\ldots,\mathbf{p}_N). \tag{4.3.28}$$

This might represent a Feynman graph with N external lines and L loops. Then the $\mathbf{p}_i$'s and $\mathbf{k}_i$'s represent momentum vectors. Suppose that at $d = 4$ the integral is completely convergent – in particular that there are no ultra-violet ($k \to \infty$) divergences or subdivergences. If we restrict the $\mathbf{p}_i$'s to the first four dimensions and set $d = 4$, then I is the ordinary four-dimensional integral of f.

Proof. For each vector $\mathbf{v}$, we define the projections $\bar{\mathbf{v}}$ and $\hat{\mathbf{v}}$, onto the physical and unphysical dimensions, as before. The result to be proved is that if $\mathbf{p}_i = \bar{\mathbf{p}}_i$ for each of the $\mathbf{p}_i$'s then I defined as the limit as $d \to 4$ of the dimensionally regulated integral is identical to the ordinary integral. We can split each integral over a $\mathbf{k}_i$ into an ordinary four-dimensional integral over $\bar{\mathbf{k}}_i$ and a $(d-4)$-dimensional integral over $\hat{\mathbf{k}}_i$. The result to be proved is then that

$$\lim \int \mathrm{d}^{d-4}\hat{\mathbf{k}}_1 \ldots \mathrm{d}^{d-4}\hat{\mathbf{k}}_L f(\mathbf{k}_1, \ldots, \mathbf{k}_L, \bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_N) = f(\bar{\mathbf{k}}_1, \ldots, \bar{\mathbf{k}}_L, \bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_N) \tag{4.3.29}$$

as $d \to 4$.

This formula is proved by doing all but the integral over $\hat{\mathbf{k}}_1$. Let the result be $I_{(1)}$:

$$I_{(1)}(\hat{\mathbf{k}}_1) = \int \mathrm{d}^{d-4}\hat{\mathbf{k}}_2 \ldots \mathrm{d}^{d-4}\hat{\mathbf{k}}_L f; \tag{4.3.30}$$

its only dependence on $\hat{\mathbf{k}}_1$ is via its length. We then have that the left-hand side of (4.3.29) is

$$\int \mathrm{d}^{d-4}\hat{\mathbf{k}}_1 I_{(1)} = I_{(1)}(0) = \int \mathrm{d}^{d-4}\hat{\mathbf{k}}_2 \ldots \mathrm{d}^{d-4}\hat{\mathbf{k}}_L f(\bar{\mathbf{k}}_1, \mathbf{k}_2, \ldots, \mathbf{k}_L, \bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_N),$$

by use of the Property 9. Notice that the dependence on $\mathbf{k}_1$ is on its first four dimensions. We can then repeat this process to show that

$$I_{(1)}(0) = \int \mathrm{d}^{d-4}\hat{\mathbf{k}}_3 \ldots \mathrm{d}^{d-4}\hat{\mathbf{k}}_L f(\bar{\mathbf{k}}_1, \bar{\mathbf{k}}_2, \mathbf{k}_3, \ldots).$$

Another $L-2$ repetitions give (4.3.29), from which the desired property follows.

4.4 Formulae for Minkowski space

In this section we derive a collection of results that are useful for Feynman graph calculations.

4.4.1 *Schwinger parameters*

To convert an arbitrary graph in d dimensions to a parametric integral, we first rewrite each propagator using

$$\frac{1}{(m^2 - k^2)^\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^\infty \mathrm{d}x x^{\alpha - 1} \exp[-x(m^2 - k^2)]. \tag{4.4.1}$$

Then we perform the momentum integrals. Since all Feynman graphs are of

the form of a polynomial in momenta times a product of simple scalar propagators, we only have to calculate d-dimensional integrals of the form:

$$I_n^{\mu_1\cdots\mu_n}(A,B)=\int \mathrm{d}^d k\, k^{\mu_1}\ldots k^{\mu_n}\exp\left[-(-Ak^2-2B\cdot k)\right]. \quad (4.4.2)$$

Here A depends only on the parameters introduced by (4.4.1), while B^μ depends on these parameters and also linearly on the other momenta (both external and loop momenta).

By linearity we can find I_n by differentiating I_0:

$$I_n^{\mu_1\cdots\mu_n}=\prod_{j=1}^{n}\left(\frac{1}{2}\frac{\partial}{\partial B_{\mu_j}}\right)\int \mathrm{d}^d k\exp(Ak^2+2B\cdot k). \quad (4.4.3)$$

(This uses linearity of d-dimensional integration.) We find I_0 by using the translation $k^\mu\to k^\mu-B^\mu/A$, the scaling $k\to k\,A^{-1/2}$, and Wick rotation:

$$\begin{aligned} I_0(A,B)&=\int \mathrm{d}^d k\exp(Ak^2+2B\cdot k)\\ &=\mathrm{i}(\pi/A)^{d/2}\exp(-B^2/A). \end{aligned} \quad (4.4.4)$$

Thus

$$\begin{aligned} I_1^\mu&=\int \mathrm{d}^d k\, k^\mu\exp(Ak^2+2B\cdot k)\\ &=\mathrm{i}(\pi/A)^{d/2}\exp(-B^2/A)(-B^\mu/A), \end{aligned} \quad (4.4.5)$$

$$\begin{aligned} I_2^{\mu\nu}&=\int \mathrm{d}^d k\, k^\mu k^\nu\exp(Ak^2+2B\cdot k)\\ &=\mathrm{i}(\pi/A)^{d/2}\exp(-B^2/A)(B^\mu B^\nu/A^2-\tfrac{1}{2}g^{\mu\nu}/A), \end{aligned} \quad (4.4.6)$$

$$\begin{aligned} I_3^{\lambda\mu\nu}&=\int \mathrm{d}^d k\, k^\lambda k^\mu k^\nu\exp(Ak^2+2B\cdot k)\\ &=\mathrm{i}(\pi/A)^{d/2}\mathrm{e}^{-B^2/A}\left[\frac{-B^\lambda B^\mu B^\nu}{A^3}+\frac{(B^\lambda g^{\mu\nu}+B^\mu g^{\lambda\nu}+B^\nu g^{\lambda\mu})}{2A^2}\right], \end{aligned} \quad (4.4.7)$$

$$\begin{aligned} I_4^{\kappa\lambda\mu\nu}&=\int \mathrm{d}^d k\, k^\kappa k^\lambda k^\mu k^\nu\exp(-B^2/A)\\ &=\mathrm{i}(\pi/A)^{d/2}\exp(-B^2/A)\left[\frac{B^\kappa B^\lambda B^\mu B^\nu}{A^4}-\frac{(B^\kappa B^\lambda g^{\mu\nu}+\text{five similar})}{2A^3}\right.\\ &\qquad\left.+\frac{(g^{\kappa\lambda}g^{\mu\nu}+\text{two similar})}{4A^2}\right]. \end{aligned} \quad (4.4.8)$$

Each of the loop-momentum integrals is performed in this way. At each stage the momenta only appear quadratically and linearly in the exponent.

4.4.2 Feynman parameters

It is also common to use

$$1/(AB) = \int_0^1 \mathrm{d}x/[Ax + B(1-x)]^2 \tag{4.4.9}$$

and its generalizations:

$$\frac{1}{A^\alpha B^\beta \cdots E^\varepsilon} = \frac{\Gamma(\alpha+\beta+\cdots\varepsilon)}{\Gamma(\alpha)\Gamma(\beta)\cdots\Gamma(\varepsilon)} \times \int_0^1 \mathrm{d}x\,\mathrm{d}y\cdots\mathrm{d}z\,\delta(1-x-y-\cdots z) \times \frac{x^{\alpha-1}y^{\beta-1}\cdots z^{\varepsilon-1}}{(Ax+By+\cdots Ez)^{\alpha+\beta+\cdots+\varepsilon}}. \tag{4.4.10}$$

Here $A, B, \ldots, E$ represent the denominators of the propagators of a Feynman graph. The resulting momentum integrals have the form

$$J^{\mu_1\cdots\mu_n}_{n,\alpha} = \int \mathrm{d}^d k \frac{k^{\mu_1}\cdots k^{\mu_n}}{[-k^2 - 2p\cdot k + C]^\alpha}. \tag{4.4.11}$$

Application of (4.4.1) and the results (4.4.4)–(4.4.8), etc., gives

$$J_0 \equiv \int \mathrm{d}^d k/(-k^2 - 2p\cdot k + C)^\alpha = \mathrm{i}\pi^{d/2}(C+p^2)^{d/2-\alpha}\Gamma(\alpha - d/2)/\Gamma(\alpha), \tag{4.4.12}$$

$$J_1^\mu \equiv \int \mathrm{d}^d k\, k^\mu/(-k^2 - 2p\cdot k + C)^\alpha = \mathrm{i}\pi^{d/2}(C+p^2)^{d/2-\alpha}(-p^\mu)\Gamma(\alpha - d/2)/\Gamma(\alpha), \tag{4.4.13}$$

$$J_2^{\mu\nu} \equiv \int \mathrm{d}^d k\, k^\mu k^\nu/(-k^2 - 2p\cdot k + C)^\alpha = \mathrm{i}\pi^{d/2}(C+p^2)^{d/2-\alpha} \times \times [\Gamma(\alpha - d/2)p^\mu p^\nu - \Gamma(\alpha - 1 - d/2)g^{\mu\nu}(C+p^2)/2]/\Gamma(\alpha). \tag{4.4.14}$$

4.5 Dirac matrices

The Dirac matrices satisfy the following properties:

(1) Anticommutation relation:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}1. \tag{4.5.1}$$

(2) Hermiticity:

$$\gamma^{\mu\dagger} = \gamma_\mu = \begin{cases} \gamma^\mu & \text{if } \mu = 0, \\ -\gamma^\mu & \text{if } \mu \geq 1. \end{cases} \tag{4.5.2}$$

When we use dimensional regularization, the Lorentz indices range over an infinite set of values, so we need infinite-dimensional matrices to represent the algebra (4.5.1). We will also need a trace operation:

$$\operatorname{tr} 1 = f(d),$$

so that the representation behaves as if its dimension were $f(d)$. We must require $f(d_0)$ to be the usual value at the physical space-time dimension, $d = d_0$. Usually this means $f(4) = 4$.

The trace is a linear operation on the matrices which we will define later. In an even integer dimension $d = 2\omega$, the standard representation of the γ^μ's has dimension 2^ω. However, it is not necessary to choose $f(d) = 2^{d/2}$. The variation $f(d) - f(d_0)$ is only relevant for a divergent graph, so, by Chapter 7, any change in $f(d)$ amounts to a renormalization-group transformation. It is usually convenient to set $f(d) = f(d_0)$ for all d.

To set up a formalism for dimensionally regularized γ-matrices, we must treat the following issues:

(1) We must exhibit a representation of the anticommutation relations; this will ensure consistency.
(2) The formulae for the trace of an arbitrary product of γ^μ's must be derived.
(3) While a knowledge of the γ^μ's alone is sufficient for QCD and QED, we must show how to define a γ_5 so that we can treat chiral symmetries. This will also give us a definition of the antisymmetric tensor $\varepsilon_{\kappa\lambda\mu\nu}$.

The following construction gives a representation:

Let ω be a positive integer, and suppose inductively that we have defined a 2^ω dimensional representation $\gamma^\mu_{(\omega)}$ of the algebra (4.5.1) for $0 \leq \mu \leq 2\omega - 1$. We define the infinite dimensional γ^μ for $0 \leq \mu \leq 2\omega - 1$ by having a sequence of $\gamma^\mu_{(\omega)}$'s down the diagonal, and zeros everywhere else:

$$\gamma^\mu = \begin{pmatrix} \gamma^\mu_{(\omega)} & 0 & \\ 0 & \gamma^\mu_{(\omega)} & \ddots \end{pmatrix}. \tag{4.5.3}$$

We will construct the next higher representation $\gamma^\mu_{(\omega+1)}$ of dimension $2^{\omega+1}$, with $0 \leq \mu \leq 2\omega + 1$. In order that (4.5.3) apply independently of ω, we must

choose

$$\gamma^{\mu}_{(\omega+1)} = \begin{pmatrix} \gamma^{\mu}_{(\omega)} & 0 \\ 0 & \gamma^{\mu}_{(\omega)} \end{pmatrix} \quad \text{if } 0 \le \mu \le 2\omega - 1.$$

This satisfies the anticommutation relations (4.5.1) and the hermiticity relation (4.5.2), provided that $0 \le \mu, \nu \le 2\omega - 1$. Our task then is to find $\gamma^{\mu}_{(\omega+1)}$ for $\mu = 2\omega$ and $2\omega + 1$.

Notice that the induction starts with $\omega = 1$. We can define

$$\gamma^{0}_{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^{1}_{(1)} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.5.4}$$

Given the 2^{ω}-dimensional representation $\gamma^{\mu}_{(\omega)}$ we define another matrix

$$\hat{\gamma}_{(\omega)} = \mathrm{i}^{\omega-1}\gamma^{0}_{(\omega)} \cdots \gamma^{2\omega-1}_{(\omega)}. \tag{4.5.5}$$

Observe that

$$\hat{\gamma}^{\dagger}_{(\omega)} = \hat{\gamma}_{(\omega)}, \quad \hat{\gamma}^{2}_{(\omega)} = 1, \quad \{\hat{\gamma}_{(\omega)}, \hat{\gamma}^{\mu}_{(\omega)}\} = 0. \tag{4.5.6}$$

Also, when at $\omega = 2$, we have $\hat{\gamma} = \gamma_5$, in the usual notation for Dirac matrices at $d = 4$. We define

$$\gamma^{2\omega}_{(\omega+1)} = \begin{pmatrix} 0 & \hat{\gamma}_{(\omega)} \\ -\hat{\gamma}_{(\omega)} & 0 \end{pmatrix},$$

$$\gamma^{2\omega+1}_{(\omega)} = \begin{pmatrix} 0 & \mathrm{i}\hat{\gamma}_{(\omega)} \\ \mathrm{i}\hat{\gamma}_{(\omega)} & 0 \end{pmatrix}. \tag{4.5.7}$$

It is easy to check that (4.5.1) and (4.5.2) are satisfied for $0 \le \mu, \nu \le 2\omega + 1$.

We now have an explicit representation of the Dirac matrices for any ω, and for the infinite-dimensional case, because of (4.5.3).

Standard manipulations involving the anticommutation relations are valid independently of d. Two useful results are:

$$\gamma^{\mu}\gamma_{\mu} = \tfrac{1}{2}\{\gamma^{\mu}, \gamma_{\mu}\}1 = g^{\mu}_{\mu}1 = d1, \tag{4.5.8}$$

$$\begin{aligned} \gamma^{\mu}\gamma_{\nu}\gamma_{\mu} &= 2g_{\mu\nu}\gamma^{\mu} - \gamma^{\mu}\gamma_{\mu}\gamma_{\nu} \\ &= (2-d)\gamma_{\nu}. \end{aligned} \tag{4.5.9}$$

We also need traces of γ-matrices, in graphs with fermion loops. The trace of a matrix is linear:

$$\mathrm{tr}\,(aA + bB) = a\,\mathrm{tr}\,(A) + b\,\mathrm{tr}\,(B), \tag{4.5.10}$$

and is cyclic:

$$\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA). \tag{4.5.11}$$

Here A and B are any product of γ-matrices, and a and b are any numbers.

These properties, together with the value of tr 1, define the trace of any linear combination of products of γ-matrices.

For example,

$$\begin{aligned}\operatorname{tr}(\gamma^\mu\gamma^\nu) &= \operatorname{tr}(\gamma^\nu\gamma^\mu) \quad \text{(cyclicity)}\\ &= \operatorname{tr}(-\gamma^\mu\gamma^\nu + 2g^{\mu\nu}1) \quad \text{(anticommutation)}\\ &= -\operatorname{tr}(\gamma^\mu\gamma^\nu) + 2g^{\mu\nu}\operatorname{tr} 1 \quad \text{(linearity)},\end{aligned}$$

so we have the usual result

$$\operatorname{tr}(\gamma^\mu\gamma^\nu) = g^{\mu\nu}\operatorname{tr} 1. \tag{4.5.12}$$

Similarly

$$\operatorname{tr}(\gamma^\kappa\gamma^\lambda\gamma^\mu\gamma^\nu) = (g^{\kappa\lambda}g^{\mu\nu} - g^{\kappa\mu}g^{\lambda\nu} + g^{\kappa\nu}g^{\lambda\mu})\operatorname{tr} 1. \tag{4.5.13}$$

The trace of the product of an odd number of γ-matrices is zero. For example

$$\begin{aligned}d\operatorname{tr}\gamma^\lambda &= \operatorname{tr}(\gamma^\kappa\gamma_\kappa\gamma^\lambda)\\ &= -\operatorname{tr}(\gamma^\kappa\gamma^\lambda\gamma_\kappa) + 2\operatorname{tr}\gamma^\lambda\\ &= -\operatorname{tr}(\gamma_\kappa\gamma^\kappa\gamma^\lambda) + 2\operatorname{tr}\gamma^\lambda,\end{aligned}$$

so $\operatorname{tr}\gamma^\lambda = 0$.

It should be possible to make a more constructive definition of the trace, along the lines of (4.3.5). It is necessary to check consistency. We can find a formula for the trace of any number of γ-matrices – generalizing (4.5.13). It is true for any finite-dimensional representation, $\gamma^\mu_{(\omega)}$, so it agrees with the algebraic properties. Linearity defines the trace of more general products. We must also check that contracting with $g_{\mu\nu}$ commutes with the trace. This can be checked directly.

A possible explicit definition of the trace of a matrix with components M_{ij} is

$$\operatorname{tr} M_{ij} = (\operatorname{tr} 1)\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N} M_{jj}. \tag{4.5.14}$$

This definition exploits the fact that each matrix γ^μ is an infinite set of copies of a finite-dimensional $\gamma^\mu_{(\omega)}$ strung along the diagonal. Since the γ^μ's are independent of d, the only possible d-dependence is in the choice of the value of tr 1.

4.6 γ_5 and $\varepsilon_{\kappa\lambda\mu\nu}$

In four dimensions, $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and $\varepsilon_{\kappa\lambda\mu\nu}$ is a totally antisymmetric Lorentz-invariant tensor with $\varepsilon_{0123} = 1$. We need γ_5, for example, to define

the axial current $\bar{\psi}\gamma^\mu\gamma_5\psi$. The ε-tensor comes in because $\gamma_5 = i\varepsilon_{\kappa\lambda\mu\nu}\gamma^\kappa\gamma^\lambda\gamma^\mu\gamma^\nu/4!$, and we have the trace formula:

$$\operatorname{tr}\gamma^5\gamma^\kappa\gamma^\lambda\gamma^\mu\gamma^\nu = i\varepsilon^{\kappa\lambda\mu\nu}\operatorname{tr} 1 = -i\varepsilon_{\kappa\lambda\mu\nu}\operatorname{tr} 1.$$

The appropriate definition changes when we go to two dimensions: Instead of γ_5 we have $\hat{\gamma}_{(1)} = \gamma^0\gamma^1$, and instead of $\varepsilon_{\kappa\lambda\mu\nu}$ we have $\varepsilon_{\mu\nu}$, for which $\varepsilon_{01} = 1 = -\varepsilon_{10}$, $\varepsilon_{00} = \varepsilon_{11} = 0$.

To continue dimensionally, we might expect γ_5 to satisfy

$$\{\gamma_5, \gamma^\mu\} = 0,$$

just as in four dimensions. But then, as we will see in Chapter 13, the only consistent result for γ_5 is that it has zero trace when multiplied by any string of γ^μ's. Thus we do not have a regularization involving the usual γ_5.

A consistent definition is obtained by writing

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = i\gamma^\kappa\gamma^\lambda\gamma^\mu\gamma^\nu\varepsilon_{\kappa\lambda\mu\nu}/4!, \tag{4.6.1}$$

$$\varepsilon_{\kappa\lambda\mu\nu} = \begin{cases} 1 \text{ if } (\kappa\lambda\mu\nu) \text{ is an even permutation of (0123),} \\ -1 \text{ if } (\kappa\lambda\mu\nu) \text{ is an odd permutation of (0123),} \\ 0 \text{ otherwise.} \end{cases} \tag{4.6.2}$$

This definition is not Lorentz invariant on the full space, but only on the first four dimensions. We have

$$\begin{aligned} \{\gamma_5, \gamma^\mu\} &= 0, \quad \text{if } \mu = 0, 1, 2, 3, \\ [\gamma_5, \gamma^\mu] &= 0, \quad \text{otherwise,} \\ (\gamma_5)^2 &= 1, \quad \gamma_5^\dagger = \gamma_5. \end{aligned} \tag{4.6.3}$$

The lack of full Lorentz invariance is a nuisance, but it does give the correct axial anomaly ('t Hooft & Veltman (1972a)).

5

Renormalization

In this chapter we come to the general theory of renormalization. The basic difficulty is that a graph may not only possess an overall divergence. It may have in addition many subdivergences which can be nested or can overlap in very complicated ways. Most of our effort must go to disentangling these complications.

We will begin by investigating some simple graphs. These will show us how to set up the formalism in the general case. The ultimate result is the forest formula of Zimmermann (1969). Contrary to its reputation, this is not an esoteric procedure, designed for pedantically rigorous treatments. Rather, the forest formula is merely a general way of writing down what is in fact the natural and obvious way of extracting the divergences from any integral. Its power is demonstrated by the ease of treating overlapping divergences, the handling of which is normally considered the *bête noire* of renormalization theory.

The forest formula is applied to individual Feynman graphs. It extracts the finite part of a graph by subtracting its overall divergence and its subdivergences. We will, of course, need to show that the subtractions can be implemented as actual counterterms in the Lagrangian. We will also show that the counterterms are local, i.e., polynomial in momentum.

An important advantage of using the forest formula to obtain the finite part of each graph, rather than working directly with counterterms in the Lagrangian, is that the procedure applies to more general situations. As we will see in Chapter 6, it will enable us to renormalize composite operators. A more important case is the computation of asymptotic behavior as external momenta of a Green's function get large. For this case, the forest formula permits a good derivation of Wilson's operator-product expansion, which we will discuss in Chapter 10.

Let the value of a Feynman graph be written as:

$$U(G)(p_1,\ldots,p_N) = \int \mathrm{d}^d k_1 \ldots \mathrm{d}^d k_L \, I(p_1,\ldots,p_N;k_1,\ldots k_L). \qquad (5.0.1)$$

Here $p_1,\ldots,p_N$ are the external momenta, and $k_1,\ldots,k_L$ are the loop

momenta. Renormalization is removal of that part of the large-k behavior that causes divergences. Moreover, the very same techniques can be used to extract the behavior for large p as we will see when we treat the operator product expansion in Chapter 10.

Although Weinberg's (1960) theorem tells us the power-counting applicable to either kind of asymptotic behavior, it does not tell us how to organize it. In particular it was only much later that Wilson (1969) formulated his operator product expansion, which is the important tool in computing asymptotic behavior, for example in deep-inelastic scattering – see Chapter 14. Many generalizations have been made – see Mueller (1981) for a review. These are phenomenologically very important, and the method by which they are proved is close to that for Wilson's expansion.

5.1 Divergences and subdivergences

The idea of renormalization theory is that ultra-violet divergences of a field theory are to be cancelled by renormalizations of the parameters of the theory. We propose to prove this in perturbation theory. The use of perturbation theory implies that we expand the counterterms in the action in powers of the renormalized coupling, g, thereby generating extra graphs with these counterterms as some of the vertices. To avoid superfluous technicalities, we will consider the case of ϕ^3 theory in six-dimensional space-time.

A very efficient way to understand renormalization was discovered by Bogoliubov & Shirkov (1955, 1956, 1980) and Bogoliubov & Parasiuk (1957), and we shall follow their approach. The first step is to decompose the Lagrangian as follows:

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_b + \mathscr{L}_{ct}. \tag{5.1.1}$$

Here $\mathscr{L}_0$ is the free Lagrangian used to generate the free propagator $i/(p^2 - m^2 + i\varepsilon)$ in perturbation theory:

$$\mathscr{L}_0 = (\partial\phi)^2/2 - m^2\phi^2/2, \tag{5.1.2}$$

with m being the renormalized mass. The rest of the Lagrangian, $\mathscr{L}_I = \mathscr{L}_b + \mathscr{L}_{ct}$, is the interaction, and consists of two terms. The first, which we will call the basic interaction, is

$$\mathscr{L}_b = -g\phi^3/3!, \tag{5.1.3}$$

where g is the renormalized coupling. The second term, $\mathscr{L}_{ct}$, we will call the counterterm Lagrangian.

Consider graphs generated by the basic interaction. These have UV

divergences which are to be cancelled by graphs with some of their interaction vertices taken from the counterterm Lagrangian

$$\mathscr{L}_{ct} = \delta Z(\partial\phi)^2/2 - \delta m^2\phi^2/2 - \delta g\phi^3/3! - \delta h\phi. \tag{5.1.4}$$

(The term linear in ϕ is needed to cancel tadpole graphs – see Figs. 5.1.4 and 5.1.5 below.) In order to give meaning to δZ, δm^2, δg, and δh, we must impose an ultra-violet cut-off. We will use dimensional regularization in the following sections.

The key to the method that we use is to realize that each of the three terms in the counterterm Lagrangian should not be considered as a single quantity. Rather it is to be considered as a sum of terms, each of them cancelling the overall divergence in one particular graph generated by the basic interaction. For example, the self-energy graph, Fig. 3.1.1, gives a contribution $\delta_1 Z$ to δZ, and a term $\delta_1 m^2$ to δm^2. Our calculation of this graph led to (3.5.7), so with minimal subtraction we have

$$\left.\begin{aligned} \delta_1 Z &= g^2/[384\pi^3(d-6)], \\ \delta_1 m^2 &= g^2 m^2/[64\pi^3(d-6)]. \end{aligned}\right\} \tag{5.1.5}$$

Then

$$\delta Z = \sum_{\text{graphs } G} \delta_G Z = \delta_1 Z + \cdots,$$

with similar formulae for the other counterterms.

We saw the utility of this idea by examining graphs like those in Fig. 3.2.1 and Fig. 3.2.2. Graphs like Fig. 3.2.2 contain vertices corresponding to the counterterm $\delta_1 m^2$ (and $\delta_1 Z$). Such graphs are all generated by taking graphs like Fig. 3.2.1 with no counterterms and finding where Fig. 3.1.1 occurs as a subgraph. Substitution of the counterterm for one or more of these subgraphs gives the graphs with counterterm vertices.

This leads to the idea that we consider by itself the renormalization of a single graph generated from the basic Lagrangian. We add to it a set of counterterm graphs to give a finite result. Only as a separate step do we recognize that the counterterm vertices are, in fact, generated from a piece of an interaction Lagrangian.

The graph-by-graph method is probably the most powerful approach to understanding not only the problem of ultra-violet divergences but also many other problems in asymptotic behavior. Even so, it is not at all trivial to ensure that the renormalization program can be carried out. The essential steps are:

(1) To find the regions in the space of loop momenta of a graph that give ultra-violet divergences.

(2) To show how to generate a series of counterterm graphs for a given basic graph.
(3) To show that the counterterm vertices are local (i.e., polynomial in momenta).
(4) To find the conditions under which the counterterm vertices amount only to renormalizations of the parameters of the Lagrangian.

The complications in carrying out this program arise when one treats the case of the divergence of a graph which has a divergent subgraph. To understand why there is a difficulty, we will examine the graphs of order g^4 for the full propagator – Figs. 5.1.1–5.1.3.

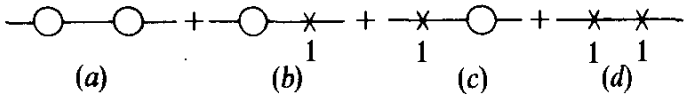


Fig. 5.1.1. A two-loop graph for the propagator in ϕ^3 theory, together with its counterterm graphs.

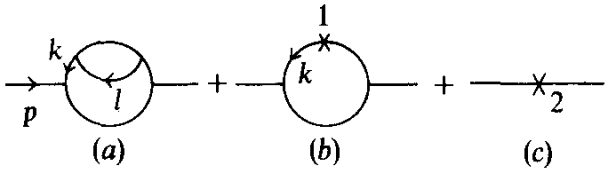


Fig. 5.1.2. A two-loop graph for the propagator in ϕ^3 theory, together with its counterterm graphs.

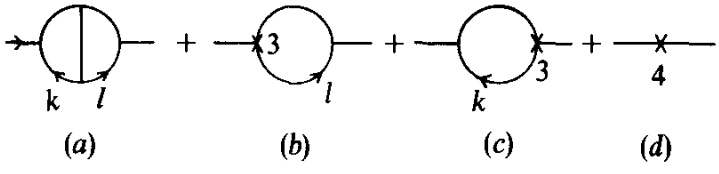


Fig. 5.1.3. A two-loop graph for the propagator in ϕ^3 theory, together with its counterterm graphs.

We ignore the graphs with tadpoles, such as Fig. 5.1.4. These are divergent and need a counterterm $\delta h\phi$. We can use a renormalization condition that $\langle 0|\phi|0\rangle$ vanishes. Then the total of the tadpole graphs is zero (e.g., Fig. 5.1.5), so we omit any graphs containing them.

Let us return to the sets of graphs listed in Figs. 5.1.1–5.1.3. In each set there is one basic graph and a set of counterterm graphs. Ultra-violet

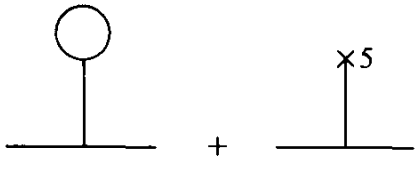


Fig. 5.1.4. A tadpole graph, together with its counterterm.

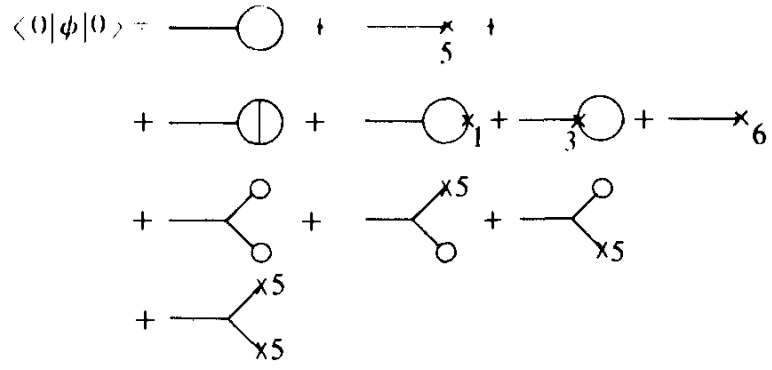


Fig. 5.1.5. Graphs to $O(g^3)$ for $\langle 0|\phi|0\rangle$.

divergences involve a loop momentum that gets large, so the divergences are always confined to one-particle-irreducible subgraphs. The simplest case is Fig. 5.1.1, where the basic graph has two insertions of the one-loop self-energy. It is made finite by adding the graphs with one or both of the self-energy subgraphs replaced by a counterterm. (We use a cross to denote a counterterm in a graph, and we use the label '1' for the counterterm of the one-loop self-energy.)

In Fig. 5.1.2, the basic graph is more complicated. We will treat it in detail in Section 5.2, and we merely summarize the results here. It has two UV divergences. The first comes from letting both loop momenta k and l go to infinity; we call this the overall divergence. But there is also a divergence where the momentum in the outer loop stays finite. This is an example of a subdivergence. Its existence, as we will see, implies that there is a term proportional to $p^2 \ln(p^2)$ in the divergence of the basic self-energy graph. This cannot be cancelled by any local counterterm. However there is also a graph with a counterterm to the subgraph. This graph is Fig. 5.1.2(*b*). After we add the two graphs, the non-local divergence cancels. The overall divergence in the result is then cancelled by a local counterterm, for which we use the label '2'. We implement this as a counterterm in the action by inserting terms $\delta_2 Z$ and $\delta_2 m^2$ into the complete counterterms δZ and δm^2.

The pattern is simple. We consider as a single finite entity one basic graph together with counterterms for its subdivergences and for the overall divergence.

Another case is shown in Fig. 5.1.3. There are two subdivergences, each corresponding to a vertex subgraph. The one-loop vertex is logarithmically divergent and is made finite by renormalizing the coupling (Fig. 3.6.1). Since the two divergent subgraphs overlap, the counterterm graphs are generated by replacing one but not both of the vertex graphs by its counterterm. The overall divergence is then local and is cancelled by counterterms $\delta_4 Z$ and $\delta_4 m^2$.

Our first task in Section 5.2 will be to verify the above statements. To generalize the argument we will then observe that power-counting as in Section 3.3 determines the strength of the overall divergence. To prove that the presence of subdivergences does not affect the form of the overall counterterm, we will differentiate with respect to external momenta to remove the overall divergence. Then we will be able to construct an inductive proof that if subdivergences have been cancelled by counterterms then the overall divergence is local and its strength is determined by simple power-counting.

We will also show how to disentangle the combinatoric problems when subdivergences are nested. Finally, we will discuss Weinberg's theorem. This theorem tells us exactly which regions of momentum we must consider. In practice one is very simple-minded about locating UV divergences. For example, we stated that the regions giving divergences for Fig. 5.1.2 are: (*a*) k and l going to infinity together, and (*b*) l going to infinity, with k fixed. In each region, all the momenta get large in a particular 1PI subgraph that is divergent by power-counting. Weinberg's theorem tells us that these are the only regions we have to consider explicitly. In the case of Fig. 5.1.2 there is another region that is important, where l goes to infinity with k also going to infinity, but much more slowly. This region interpolates between the other two, but in fact does not need to be treated as a separate case.

5.2 Two-loop self-energy in $(\phi^3)_6$

In this section we will explain the properties of overall divergences and subdivergences by computing the two-loop self-energy graphs, Figs 5.1.2 and 5.1.3, in ϕ^3 theory at space-time dimension $d = 6$. We will again use dimensional regularization, and will need the values of the one-loop counterterms in order to cancel subdivergences.

The one-loop self-energy was considered in Section 3.6.2, where we found that the counterterms needed were given by (5.1.5). We can also compute the one-loop vertex graph, Fig. 3.6.1, with the resulting counterterm being (cf., (3.6.13))

$$\delta_3 g = \mu^{3-d/2} g^3 / [64\pi^3 (d-6)]. \tag{5.2.1}$$

It is worth noting that this implies a value for the one-loop term in the bare coupling:

$$\begin{aligned} g_0 &= [\mu^{3-d/2} g + \delta_3 g + O(g^5)] Z^{-3/2} \\ &= g\mu^{3-d/2} \{1 + \tfrac{3}{4} g^2 / [64\pi^3 (d-6)] + O(g^4)\}. \end{aligned} \tag{5.2.2}$$

5.2.1 *Fig.* 5.1.2

In order to be able to compute Fig. 5.1.2 in closed form we work with the massless theory. The value of the graph is then

$$\Sigma_{2a} = \frac{-g^4\mu^{12-2d}}{(2\pi)^{2d}} \times$$

$$\times \tfrac{1}{2}\int \mathrm{d}^d k \mathrm{d}^d l \frac{1}{[(p+k)^2+\mathrm{i}\varepsilon](k^2+\mathrm{i}\varepsilon)^2(l^2+\mathrm{i}\varepsilon)[(k-l)^2+\mathrm{i}\varepsilon]}. \quad (5.2.3)$$

The inner loop is easily computed in terms of Γ-functions:

$$\int \mathrm{d}^d l \frac{1}{l^2(k-l)^2} = \mathrm{i}\pi^{d/2}\Gamma(2-d/2)\int_0^1 \mathrm{d}x[-k^2x(1-x)]^{d/2-2}$$

$$= \mathrm{i}\pi^{d/2}\Gamma(2-d/2)\frac{\Gamma(d/2-1)^2}{\Gamma(d-2)}(-k^2)^{d/2-2}. \quad (5.2.4)$$

We now have

$$\Sigma_{2a} = \frac{\mathrm{i}g^4}{2^{13}\pi^9}(16\pi^3\mu^4)^{3-d/2}\Gamma(2-d/2)\frac{\Gamma(d/2-1)^2}{\Gamma(d-2)} \times$$

$$\times \int \mathrm{d}^d k \frac{1}{(-k^2)^{4-d/2}[-(p+k)^2]}. \quad (5.2.5)$$

The denominators can be combined by a Feynman parameter:

$$\frac{1}{A^{4-d/2}B} = \frac{\Gamma(5-d/2)}{\Gamma(4-d/2)}\int_0^1 \mathrm{d}x \frac{x^{3-d/2}}{[Ax+B(1-x)]^{5-d/2}}. \quad (5.2.6)$$

after which the k-integral can be performed. The result is

$$\Sigma_{2a} = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{2}\left(\frac{-p^2}{4\pi\mu^2}\right)^{d-6} \times$$

$$\times \frac{\Gamma(2-d/2)\Gamma(5-d)\Gamma(d/2-1)^3\Gamma(d-4)}{\Gamma(d-2)\Gamma(4-d/2)\Gamma(3d/2-5)}$$

$$\equiv \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{2}\left(\frac{-p^2}{4\pi\mu^2}\right)^{d-6} \Gamma(2-d/2)\Gamma(5-d)K(d). \quad (5.2.7)$$

The overall ultra-violet divergence is contained in the factor $\Gamma(5-d)$. Observe that the argument of this Γ-function is exactly minus half times the degree of divergence. The subdivergence is contained in the factor $\Gamma(2-d/2)$; this is the same as we calculated in Chapter 3.

Before we discuss further the UV divergences we should observe that there are also infra-red divergences. These come from the existence of long-range forces in a theory with massless fields. In momentum space, they

appear as divergences when some momenta go to zero. For example, if $d \leq 2$ the integral over the momentum through any propagator has a divergence at zero momentum:

$$\int_{q \sim 0} d^d q/q^2 \simeq \text{constant}/(d-2) \text{ at } d \text{ close to } 2.$$

This accounts for the factor $\Gamma(d/2-1)^3$. When $d \leq 4$ there is also a divergence at $k=0$ with l and p fixed. Our only concern is with UV problems, so we ignore the IR divergence. If we used a massive field, the IR divergences would go away, but we would not have an explicit formula for Σ_{2a}.

Now let us expand Σ_{2a} in powers of $d-6$ to exhibit its divergences, and its dependence on p^2:

$$\Sigma_{2a} = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\frac{1}{(d-6)^2} + \frac{1}{d-6}\left[\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\right] + \right.$$
$$\left. + \tfrac{1}{2}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant} + O(d-6)\right\}. \tag{5.2.8}$$

The double pole at $d=6$ and the double logarithm in the finite part are both reflections of the fact of having a subdivergence. The p-dependence is a power of p^2 times a polynomial in $\ln(-p^2)$. This is a characteristic feature of massless theories.

The simple pole has a coefficient that is not polynomial in p. Consequently, it cannot be cancelled by any local counterterm. It is easy to see that this is caused by the presence of the subdivergence. The subdivergence comes from the region where the loop momentum of the inner loop goes to infinity while the momentum k in the outer loop remains finite. Integrating over finite k gives a logarithm of p times the divergent part of the inner loop. We have already introduced into the Lagrangian a counterterm for the inner loop, so that there is a graph, Fig. 5.1.2(*b*), in which this counterterm appears in such a way as to cancel the subdivergence.

Therefore the sum of Figs. 5.1.2(*a*) and (*b*) should have no subdivergence, but only an overall divergence. This can be cancelled by a local counterterm (i.e., a polynomial in p). We will prove this in Section 5.2.2 by differentiating three times with respect to the external momentum p^μ; this gives a result which has negative degree of divergence, i.e., there is no overall divergence. Since the subdivergence is cancelled, there is no subdivergence whatever, so the counterterm must be quadratic in p. We represent this by Fig. 5.1.2(*c*).

In Section 5.2.2 we will make explicit this proof of locality of the counterterms.

Here we will verify the above statements by explicit calculations. In our case that $m = 0$, the value of Fig. 5.1.2(b) is

$$\begin{aligned}\Sigma_{2b} &= -i\delta_1 Z \frac{g^2\mu^{6-d}}{(2\pi)^d}\int d^dk \frac{k^2}{(k^2)^2(p+k)^2} \\ &= \left(\frac{g^2}{64\pi^3}\right)^2 \left(\frac{-p^2}{6}\right)\frac{\Gamma(2-d/2)\Gamma(d/2-1)^2}{(d-6)\Gamma(d-2)}\left(\frac{-p^2}{4\pi\mu^2}\right)^{d/2-3} \\ &= \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\frac{-2}{(d-6)^2} + \frac{1}{d-6}\left[-\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\right]\right. \\ &\quad \left. -\tfrac{1}{4}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant} + O(d-6)\right\}.\end{aligned} \tag{5.2.9}$$

The non-local divergence disappears when we add this graph to Σ_{2a}, with the result

$$\begin{aligned}\Sigma_{2a} + \Sigma_{2b} = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\frac{-1}{(d-6)^2} + \frac{\text{constant}}{(d-6)}\right. \\ \left. +\tfrac{1}{4}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant} + O(d-6)\right\}.\end{aligned} \tag{5.2.10}$$

The non-local divergence has cancelled, as promised. However, the double pole and, in the finite part, the double logarithm have not cancelled, even though it is evident from our calculation that they are associated with the subdivergence nested inside the overall divergence. This is a general phenomenon. Indeed we will see in Chapter 7, where we discuss the renormalization group, that the coefficients of the double pole and of the double logarithm could have been predicted from the one-loop counterterms without any explicit two-loop calculations.

Since the non-local divergences have now cancelled, the overall divergence can be cancelled by choosing a wave-function counterterm

$$\delta_2 Z = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{1}{36}\left\{\frac{-1}{(d-6)^2} + \frac{\text{constant}}{(d-6)}\right\}. \tag{5.2.11}$$

Then we obtain at $d = 6$ a finite result, which we term the renormalized

value of Fig. 5.1.2:

$$\Sigma_{2R}^{(MS)} = \Sigma_{2a} + \Sigma_{2b} + (\Sigma_{2c} = -p^2\delta_2 Z)$$
$$= \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\tfrac{1}{4}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ \ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\right\}. \tag{5.2.12}$$

5.2.2 *Differentiation with respect to external momenta*

We saw that the graph Fig. 5.1.2 has an overall divergence which is local, but that it is local only after we have subtracted the subdivergence. In general we will need to show that the counterterm of a 1PI graph is a polynomial in its external momenta with degree equal to the degree of divergence. Our argument (following Caswell & Kennedy (1982)) depends on differentiating with respect to external momenta.

In this subsection we will apply the argument to Fig. 5.1.2, emphasizing its generality. Then in the next subsection we will apply it to Fig. 5.1.3. Even though that graph has an overlapping divergence, traditionally considered a hard problem, we will see that our method works as easily for this graph as for Fig. 5.1.2.

We first differentiate Fig. 5.1.2(*a*) three times with respect to p^μ, to make its degree of divergence negative. We represent the result pictorially by Fig. 5.2.1, where each dot indicates one differentiation with respect to p. Similarly, differentiating Fig. 5.1.2(*b*) three times gives Fig. 5.2.2. Now Fig. 5.2.2 cancels the subdivergence in Fig. 5.2.1, and there is no overall divergence, so their sum is finite. Thus the third derivative of the sum of Figs. 5.1.2(*a*) and (*b*) is finite. So the overall counterterm is quadratic in p. Lorentz invariance forces it to be of the form $A(d)p^2 + B(d)$.

We glibly asserted that Fig. 5.2.1 plus Fig. 5.2.2 is finite. This statement is not as obvious as it seems. Let us prove it. We Wick-rotate the integrations over k and l in Fig. 5.2.1, and consider regions of the integral that might give a UV divergence. If k and l go to infinity at the same rate, then there is no

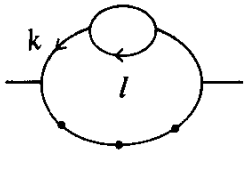


Fig. 5.2.1. Result of differentiating Fig. 5.1.2(*a*) three times with respect to its external momentum.

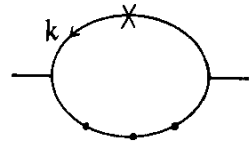


Fig. 5.2.2. Result of differentiating Fig. 5.1.2(*b*) three times with respect to its external momentum.

divergence, because the degree of divergence is negative. If l goes to infinity with k fixed, there is a divergence, but it is cancelled by the counterterm graph, Fig. 5.2.2.

The remaining significant possibility is that both k and l go to infinity, but that k is much less than l. The ratios of different components of either one of k or l are finite, so we may summarize the order of magnitude of the contribution from this region as:

$$\text{finite} \int^{\infty} \mathrm{d}k\, k^{-4} \int_{l \gg k} \mathrm{d}l\, l. \tag{5.2.13}$$

This gives a divergent contribution, if l is of order $k^{3/2}$. We must add Fig. 5.2.2, which, as we will show, cancels this new divergence. Observe that the counterterm was arranged to cancel the divergence when l goes to infinity with k fixed, rather than when k is large, as we now have.

Let us expand the inner loop of Fig. 5.2.1 in powers of k when $l \gg k$, up to its degree of divergence, which is quadratic. The coefficients of these powers are integrals over all l, restricted to $l \gg k$. The divergences in the coefficients are cancelled by Fig. 5.2.2, and we have the following estimates of the sizes of the coefficients:

$$\begin{aligned}
\text{coefficient of } k^0 &= \text{finite}\left\{\int_k^{\infty} \mathrm{d}l\, l - \text{divergence}\right\} \\
&= \text{finite}\left\{\int_1^{\infty} \mathrm{d}l\, l - \text{divergence}\right\} + \text{finite}\int_1^{k} \mathrm{d}l\, l \\
&= O(k^2), \\
\text{coefficient of } k^1 &= O(k), \\
\text{coefficient of } k^2 &= \text{finite}\left\{\int_k^{\infty} \mathrm{d}l/l - \text{divergence}\right\} \\
&= O(\ln(k)).
\end{aligned} \tag{5.2.14}$$

The sum of Figs. 5.2.1 and 5.2.2 in the region $k \to \infty$, with l possibly much bigger than k, is then of order

$$\int^{\infty} \mathrm{d}k\, k^{-2} \ln(k). \tag{5.2.15}$$

The power of k is the same as is given by the overall degree of divergence, but there is an extra logarithm. We get a finite result, as claimed. The higher-than-quadratic terms in the expansion of the loop in powers of k give no divergence at all.

What has happened? The divergence for $l \gg k \gg 1$ could only occur

because the interior loop was itself divergent. The fact that we sent k to infinity merely suppressed this divergence somewhat. Suppose we neglect k in the integral for the inside loop. Then the counterterm is in effect the negative of the integral over l of the loop from finite l to infinity. But in the region we are considering for Fig. 5.2.1, we are restricting l to be much bigger than k, which is itself getting large. So if we neglect k in the loop, then we are left with

$$\text{finite} + \int_{l<k} \mathrm{d}l \text{ (integrand of loop with } k \text{ neglected)}. \qquad (5.2.16)$$

Furthermore, we expand the loop in powers of k, to uncover the sub-leading divergences. Each extra explicit power of k in the expansion compensates for the lowering of the divergence. The quadratic term multiplies $\int \mathrm{d}l/l$, so giving an extra logarithm (but *not* a power).

The key step in the proof is to perform the integral with the larger momentum l first. We have shown that, for the purpose of determining whether or not a divergence occurs, we need only consider as distinct regions: (1) $k, l \to \infty$ at the same rate, and (2) $l \to \infty$ with k finite. (We might also try $k \to \infty$ with l finite, but the subgraph with the lines carrying the loop momentum k has negative degree of divergence, so we get no divergence from that region.) The region $k, l \to \infty$, with $k \ll l$, is schizoid: it can be considered as essentially part of either of the two regions (1) and (2) that we have just defined. As region (2), the divergence is cancelled by a counterterm when $l \to \infty$, with k large but fixed. As region (1), the final integral over k is finite, and the only sign of this intermediate case is the extra logarithm in the integrand.

5.2.3 *Fig.* 5.1.3

We conclude this section by considering the example of Fig. 5.1.3. At $d = 6$ the graph (*a*) has an overall quadratic divergence. It also has a logarithmic subdivergence when either of the loop momenta k or l gets large. The subdivergences are cancelled by vertex corrections, which are shown in graphs (*b*) and (*c*). We must prove that the overall counterterm (*d*) is quadratic in p.

Conventionally, this graph is regarded as a difficulty in the theory of renormalization, for it contains an overlapping divergence. That is to say, one of the lines is common to both subdivergences. This is seen as a problem (Bjorken & Drell (1966)) if one tries to write the graph as an insertion of a renormalized vertex, Fig. 5.2.3, in the one-loop self-energy. The graph (*b*)

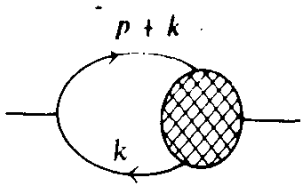


Fig. 5.2.3. Illustrating the problem of overlapping divergences, as seen in Fig. 5.1.3.

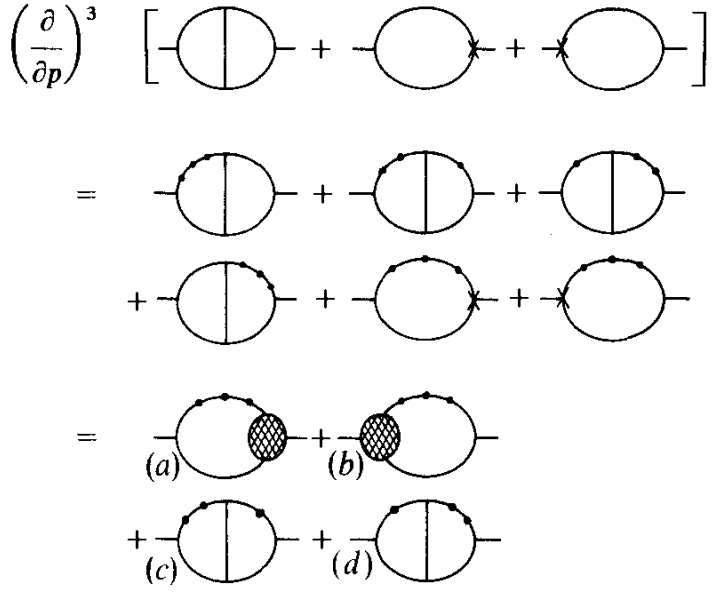


Fig. 5.2.4. Result of differentiating Fig. 5.1.3 three times with respect to its external momentum.

for the counterterm to one of the subdivergences is not of this form. The corresponding difficulty does not happen in our first example, Fig. 5.1.2.

However, our trick of differentiating three times with respect to p works as well for Fig. 5.1.3 as it did for Fig. 5.1.2. For the sum of (*a*), (*b*), and (*c*), we find Fig. 5.2.4. The point is that differentiating either of the subgraphs makes it convergent, while the counterterms for the subdivergences are independent of momenta. We get terms (*a*) and (*b*), which have renormalized subgraphs, and graphs (*c*) and (*d*), which have no subdivergences at all. None of the graphs has an overall divergence. The calculation of the overall counterterms is left as an exercise for the reader. The correct result is (Macfarlane & Woo (1974)):

$$\delta_4 Z = \left(\frac{g^2}{64\pi^3}\right)^2 \left[\frac{1}{6}\frac{1}{(d-6)^2} + \frac{1}{3}\frac{1}{(d-6)}\right],$$

$$\delta_4 m^2 = \left(\frac{g^2}{64\pi^3}\right)^2 m^2 \left[-\frac{1}{(d-6)^2} - \frac{1}{2}\frac{1}{(d-6)}\right], \tag{5.2.17}$$

if we use minimal subtraction.

5.3 Renormalization of Feynman graphs

We have seen that, in order to construct a sensible (i.e., local) counterterm for a (1PI) Feynman graph, one must first subtract off its subdivergences. This is natural since the subtractions for subdivergences are automatically generated from having the counterterms be definite pieces of the interaction Lagrangian. Without subtraction of the subdivergences, the divergence of a graph need not be local. It may even have a power of momentum greater than the degree of divergence; an obvious case of this is a graph that is finite according to naive power-counting but that has a subdivergence.

It is therefore useful to devise a procedure for starting with a basic Feynman graph G, constructing a set of counterterm graphs, and thereby obtaining a finite renormalized value $R(G)$:

$$R(G) = U(G) + S(G). \tag{5.3.1}$$

Here $U(G)$ is the 'unrenormalized' value of the basic graph (which diverges as the UV cut-off is removed), and $S(G)$ is the subtraction – the sum of the counterterm graphs.

The strategy we use to construct $S(G)$ is very general. It applies to the asymptotic behavior of any integral as one or more parameters approach a limiting value. In field theory it can be applied not only to the renormalization problem but also to the calculation of the asymptotic behavior of a Green's function as some but not all of its external momenta get large. (A standard example which we will treat in Chapter 10 is the operator product expansion of Wilson (1969)).

The procedure that we use for renormalization was first developed by Bogoliubov and Parasiuk (see Bogoliubov & Shirkov (1980)), with corrections due to Hepp (1966). Their construction was recursive and has the acronym BPH. Zimmermann (1969) showed how to solve the recursion – the result being called the forest formula. All these authors used zero-momentum subtractions. Zimmermann (1970, 1973a) showed moreover that there is then no need to use an explicit UV cut-off. He applied the algorithm for computing $R(G)$ directly to the integrand rather than to the integral; this formulation has the title BPHZ. It is not necessary to use zero-momentum subtractions. For example Speer (1974), Collins (1975b), Breitenlohner & Maison (1977a, b, c) showed how to make the same ideas work using minimal subtraction.

Our treatment will aim at showing the underlying simplicity of the methods and their power to demystify renormalization theory. We will see that the methods do not depend on use of a particular renormalization prescription, even though we will often use minimal subtraction.

We will examine the structure of the subtractions for a graph G. (A graph we define by specifying its set of vertices and lines, each line joining two vertices and each vertex attached to at least one line.) We write the graph's unrenormalized value as

$$U_G(p_1,\ldots,p_N)=\int \mathrm{d}^d k_1 \ldots \mathrm{d}^d k_L I_G(p_1,\ldots,p_N;k_1,\ldots,k_L). \tag{5.3.2}$$

Here we let L be the number of loops and N be the number of vertices. The external momenta at the vertices are p_i. In a Feynman graph for a Green's function there is an external momentum at the vertices for the external fields, but at an interaction vertex, we have $p_i = 0$.

5.3.1 *One-particle-irreducible graph with no subdivergences*

The simplest case is a one-particle-irreducible (1PI) graph with no subdivergences. Then the only possible divergence is an overall divergence where the momenta on all the lines get large simultaneously. We may renormalize the graph by subtracting an overall counterterm:

$$R(G) = U(G) - T\circ U(G). \tag{5.3.3}$$

Here T denotes some operation that extracts the divergence of $U(G)$. It implements whatever renormalization prescription that we choose to use. For example, we might use minimal subtraction. In that case T takes the Laurent expansion of $U(G)$ about $d = d_0$, and picks out the pole terms. (We let the physical space-time dimension be d_0; i.e., $d_0 = 4$ for the real world, or $d_0 = 6$ for the ϕ^3 model we used in the previous sections.) We will use either of two notations for the action of T on an unrenormalized object: $T\circ U(G)$ or $T(G)$. Both will mean the same.

We could use zero-momentum subtractions. In that case T picks out the terms up to order $\delta(G)$ in the Taylor expansion of $U(G)$ about zero momentum. Here $\delta(G)$ is, as usual, the degree of divergence. There are many other possibilities. In our work, we will use the minimal subtraction scheme.

Then, for example, the one-loop self energy in $(\phi^3)_6$ gives

$$\begin{aligned} T\circ &\left\{\frac{ig^2}{2}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\,\frac{\mu^{6-d}}{(k^2-m^2)[(p+k)^2-m^2]}\right\} \\ &= \text{pole part of}\left\{\frac{-g^2}{128\pi^3}\Gamma(2-d/2)\times\right. \\ &\quad\left.\times\int_0^1 \mathrm{d}x\left[\frac{m^2-p^2x(1-x)}{4\pi\mu^2}\right]^{d/2-3}[m^2-p^2x(1-x)]\right\} \\ &= \frac{g^2}{64\pi^3}\,\frac{(p^2/6-m^2)}{(d-6)}. \end{aligned} \tag{5.3.4}$$

The one-loop vertex, Fig. 3.6.1, gives

$$T\left\{g^3\int\frac{d^dk}{(2\pi)^d}\frac{\mu^{9-3d/2}}{(k^2-m^2)[(p+k)^2-m^2][(p+q+k)^2-m^2]}\right\}$$

$$= \text{pole part of}\left\{\frac{-ig^3\mu^{3-d/2}}{64\pi^3}\Gamma(3-d/2)\times\right.$$

$$\left.\times\int_0^1 dx\int_0^{1-x}dy\left[\frac{m^2-q^2xy-(p^2x+(p+q)^2y)(1-x-y)}{4\pi\mu^2}\right]^{d/2-3}\right\}$$

$$= i\frac{g^3\mu^{3-d/2}}{64\pi^3(d-6)}. \tag{5.3.5}$$

Observe that in this last case we define the pole to come with a factor $\mu^{3-d/2}$. This is an example of a general rule that one must define the pole part of $U(G)$ to have the same dimension as $U(G)$, for all d.

5.3.2 *General case*

In general we not only have to handle the case of an overall divergence, but also the case that subdivergences are nested within the overall divergences. Another case is exemplified by the propagator with two self-energy insertions (Fig. 5.1.1), where within one graph there are two subgraphs which can diverge independently.

As we saw from examples, we must subtract off subdivergences before finding the overall divergence. In view of the complications when the subdivergences themselves have subdivergences, etc. (*ad nauseam*), we must be extremely precise as to what is to be done. This is what we will now do. It is helpful to have a specific non-trivial example in mind, to make sense of the mathematics. Such examples are treated in subsection 5.3.3 and in Section 5.4. The reader should try to read these sections concurrently with the general treatment in this section.

First, let us define a specific divergence as being the divergence occurring when the loop momenta on a certain set of lines get big, with the momenta on other lines and the external momenta staying finite. Whether or not a given set of lines has a divergence associated with it is determined by power-counting. A divergence is thus associated with a certain subgraph. (At this stage, we do not require that the subgraph be connected.)

If a graph G diverges when all its lines get large loop momenta, it is said to have an overall divergence. A one-particle-reducible graph (like any of Fig. 3.2.1) cannot have an overall divergence – some lines are not a part of any loop. All other divergences involve a smaller subset of the lines. They, of course, are called subdivergences. Every subdivergence of a graph is the overall divergence of one of its subgraphs.

Observe that Fig. 5.1.1(*a*) has no overall divergence, but has three subdivergences. These come from the regions in the integration over loop momenta where: (1) the left-hand loop has large momentum, (2) the right-hand loop has large momentum, and (3) both loops have large momenta.

To renormalize a graph G we assume that we know how to renormalize its subdivergences, and we then let $\bar{R}(G)$ be the unrenormalized value of G with subtractions made to cancel the subdivergences. Then the only remaining divergence that is possible is an overall divergence. So we define an overall counterterm:

$$C(G) = -T\circ\bar{R}(G) \tag{5.3.6}$$

by applying to $\bar{R}(G)$ the same subtraction operator T as we discussed earlier; if there is no overall divergence (e.g., if G is one-particle-reducible) then $C(G)$ is zero. In any event the renormalized value of G is defined as

$$R(G) = \bar{R}(G) + C(G). \tag{5.3.7}$$

The definitions (5.3.6) and (5.3.7) give us $R(G)$ provided that we know how to subtract subdivergences. This is essentially a matter of renormalizing smaller graphs; we will construct $\bar{R}(G)$ in a moment. Once we have done this, we will have a recursive definition of $R(G)$: successive application of (5.3.7) to smaller and smaller subgraphs ultimately brings us to graphs with no subdivergences. These we know how to renormalize.

Now let us define $\bar{R}(G)$, which is to be $U(G)$ with subdivergences subtracted. For the case of a graph with no subdivergences we must define

$$\bar{R}(G) = U(G) \quad \text{(if } G \text{ has no subdivergences)}. \tag{5.3.8a}$$

For a larger graph we define

$$\bar{R}(G) = U(G) + \sum_{\gamma \subsetneq G} C_\gamma(G). \tag{5.3.8b}$$

We sum over all subgraphs γ of G, other than G itself, as indicated by the notation $\gamma \subsetneq G$. The other new notation $C_\gamma(G)$ means that we replace the subgraph γ by its overall counterterm, as defined by (5.3.6), i.e.,

$$C(\gamma) = \begin{cases} -T\circ\bar{R}(\gamma), & \text{if } \gamma \text{ has an overall divergence} \\ 0 & \text{if } \gamma \text{ has no overall divergence} \end{cases}. \tag{5.3.9}$$

To make a simple formula, we write the sum as being over all γ's rather than only over divergent γ's; then the $C_\gamma(G)$ is zero if γ is not overall divergent. We could of course restrict the sum only to those subgraphs that have an overall divergence.

One tricky point in the above equations arises in defining $C(\gamma)$ for a disconnected subgraph γ. An example is the subgraph of Fig. 5.1.1(*a*) consisting of the two self-energy loops. We will discuss this next.

5.3.3 Application of general formulae

Equations (5.3.6) to (5.3.9) give a definition of $R(G)$. Let us see how they apply to simple examples. For a 1PI graph with no subdivergences they just reproduce $R(G) = U(G) - T\circ U(G)$.

Next consider a graph like Fig. 3.2.1(*a*) or (*c*), whose only divergence is a subgraph with no further subdivergences. Then there is no overall divergence, so by (5.3.7)

$$R(G) = \bar{R}(G). \tag{5.3.10}$$

There is only one subdivergence, so (5.3.8) collapses to give

$$\bar{R}(G) = U(G) + C_\gamma(G), \tag{5.3.11}$$

where γ is the divergent subgraph. Here $C_\gamma(G)$ is the full graph with γ replaced by $-T\circ U(\gamma)$. We reproduce the obvious result. There is one counterterm graph like Fig. 3.2.2(*a*) or (*c*).

We now look at a graph with two or more divergent subgraphs which do not intersect and which have no subdivergences. It is sufficient to consider G to be Fig. 5.1.1(*a*). There is no overall divergence, so again $R(G) = \bar{R}(G)$. Let γ_1 and γ_2 be the self-energy bubbles. Then the subdivergences correspond to the three subgraphs γ_1, γ_2, and $\gamma_1 \cup \gamma_2$. (Here $\gamma_1 \cup \gamma_2$ means, as usual, the union of γ_1 and γ_2.) So

$$R(G) = \bar{R}(G) = U(G) + C_{\gamma_1}(G) + C_{\gamma_2}(G) + C_{\gamma_1 \cup \gamma_2}(G). \tag{5.3.12}$$

Evidently $C_{\gamma_1}(G)$ is just $U(G)$ with γ_1 replaced by its counterterm, $-T\circ U(\gamma_1)$; and similarly for γ_2. But what is $C_{\gamma_1 \cup \gamma_2}(G)$?

It corresponds to a subtraction for $\gamma_1 \cup \gamma_2$ for the region where all loop momenta are large. But we must subtract from it the counterterms for the regions where only one momentum is large:

$$C(\gamma_1 \cup \gamma_2) = -T\circ[U(\gamma_1)U(\gamma_2) + C(\gamma_1)U(\gamma_2) + U(\gamma_1)C(\gamma_2)]. \tag{5.3.13}$$

Here we used the fact that $\gamma_1 \cup \gamma_2$ is disconnected, so that

$$U(\gamma_1 \cup \gamma_2) = U(\gamma_1)U(\gamma_2). \tag{5.3.14}$$

To work out (5.3.13), we must define T when acting on a disconnected graph to act independently on its components. Thus:

$$T\circ[U(\gamma_1)U(\gamma_2)] = [T\circ U(\gamma_1)][T\circ U(\gamma_2)] = C(\gamma_1)C(\gamma_2), \tag{5.3.15}$$

$$T\circ[C(\gamma_1)U(\gamma_2)] = [T\circ C(\gamma_1)][T\circ U(\gamma_2)] = -C(\gamma_1)C(\gamma_2), \tag{5.3.16}$$

etc.

We used the property that $T\circ U(\gamma_i) = -C(\gamma_i)$. Furthermore, $T\circ[T\circ U(\gamma_i)] = T\circ U(\gamma_i)$, i.e., the pole part of a pole part is itself. We therefore find

that

$$C(\gamma_1 \cup \gamma_2) = [T_\circ U(\gamma_1)][T_\circ U(\gamma_2)], \tag{5.3.17}$$

so that we reproduce the counterterm graph Fig. 5.1.1(*d*).

The above procedure generalizes to an arbitrary graph. It may appear excessively complicated, but it allows the smoothest way of defining $R(G)$.

Finally, we observe that our definitions (5.3.6)–(5.3.10) exactly reproduce our results for the two-loop graphs like Figs. 5.1.2 and 5.1.3.

5.3.4 *Summary*

In this section we have proved very little. We have set up a series of definitions that state exactly what we mean by the renormalization of a Feynman graph. The notation we have introduced will be important in making proofs. What we will need to prove is that the overall counterterms are local and of a degree in momentum given by naive power-counting. We will also show how to solve the recursion to find an explicit formula due to Zimmermann (1969).

5.4 Three-loop example

The three-loop self energy graph of Fig. 5.4.1 in ϕ^3 theory in six dimensions is an example of a graph with nested and multiply overlapping divergences. We call it G. Its divergent subgraphs are:

$$\begin{aligned}
\gamma_1 &= \{\text{lines carrying loop momentum } k\},\\
\gamma_2 &= \{\text{lines carrying loop momentum } q\},\\
\gamma_3 &= \{\text{lines carrying loop momenta } k \text{ and/or } l\},\\
\gamma_4 &= \{\text{lines carrying loop momenta } q \text{ and/or } l\},\\
\gamma_5 &= \gamma_1 \cup \gamma_2.
\end{aligned} \tag{5.4.1}$$

The first four of these are connected 1PI vertex graphs; the last is a set of two unconnected vertex graphs. According to our definitions of Section 5.3 we have

$$\begin{aligned}
R(G) &= \bar{R}(G) + C(G)\\
&= \bar{R}(G) - T_\circ[\bar{R}(G)].
\end{aligned} \tag{5.4.2}$$

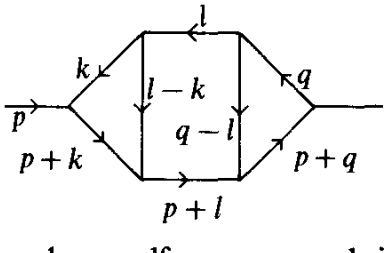


Fig. 5.4.1. Three-loop self-energy graph in ϕ^3 theory.

This equation states that we first subtract subdivergences to obtain $\bar{R}(G)$, and then take off the overall divergence.

To define $\bar{R}(G)$ we subtract subdivergences:

$$\bar{R}(G) = G + \sum_{i=1}^{5} C_{\gamma_i}(G)$$

$$= G + \sum G\big|_{\gamma_i \to C(\gamma_i)} \tag{5.4.3}$$

We represent this as Fig. 5.4.2. The notation with the vertical bar in this equation denotes that we take G and replace γ_i by the corresponding counterterm $C(\gamma_i)$. In the figure, the labels $1,\ldots,4$ signify which of the subgraphs $\gamma_1,\ldots,\gamma_4$ has been replaced by its counterterm.

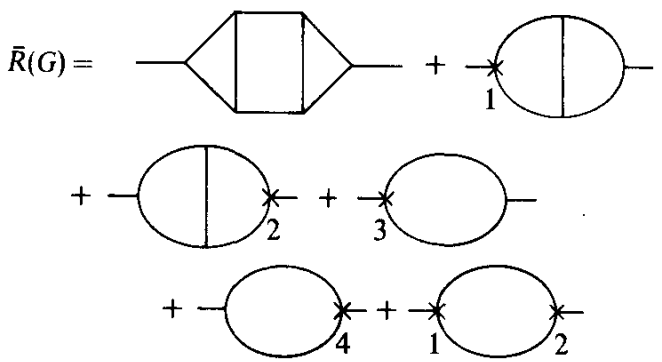


Fig. 5.4.2. Subtraction of subdivergences of Fig. 5.4.1.

Only γ_1 and γ_2 have no further subdivergences, so $C(\gamma_1)$ and $C(\gamma_2)$ are the ordinary one-loop counterterms. But we have still to define $C(\gamma_i)$ for $i = 3, 4, 5$:

$$\begin{aligned} C(\gamma_3) &= -T\circ[\gamma_3 - \gamma_3|_{\gamma_1 \to T(\gamma_1)}], \\ C(\gamma_4) &= -T\circ[\gamma_4 - \gamma_4|_{\gamma_2 \to T(\gamma_2)}], \\ C(\gamma_5) &= -T\circ[\gamma_1\gamma_2 - T(\gamma_1)\gamma_2 - \gamma_1 T(\gamma_2)] \\ &= [-T(\gamma_1)][-T(\gamma_2)]. \end{aligned} \tag{5.4.4}$$

The overall result is obtained by combining (5.4.2)–(5.4.4). If we represent the effect of applying T to a 1PI graph by enclosing it in a box, we can write $R(G)$ as shown in Fig. 5.4.3. There are sixteen terms in all. The first eight represent $U(G)$ minus its subdivergences, and the last eight form the subtraction for the overall divergence. The expansion of $R(G)$ represented in Fig. 5.4.3 is an example of the forest formula, to be discussed in the next section.

As we saw by examining two-loop graphs, the overall counterterm for a graph is non-local unless we first subtract off subdivergences. Otherwise there would be divergent contributions from where some but not all

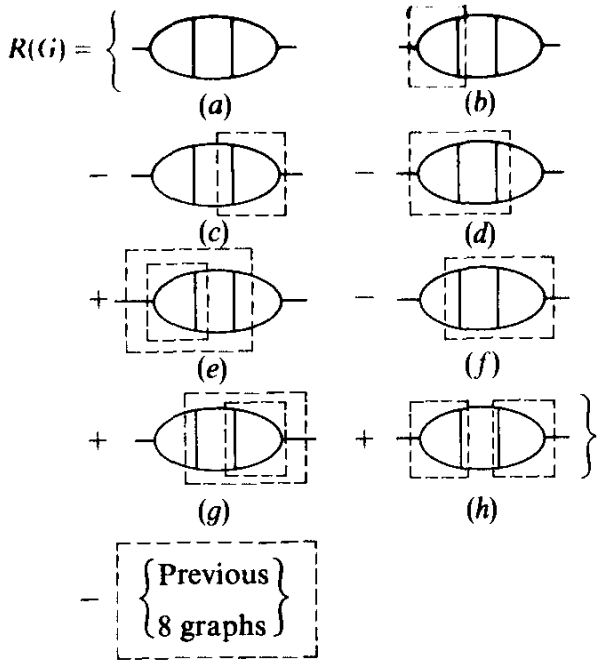


Fig. 5.4.3. Renormalization of Fig. 5.4.1.

subgraphs have large loop momenta. There would also be divergent contributions from the region where all loop momenta get large but at different rates. We can check from Fig. 5.4.3 that none of these problems occur for $R(G)$. Let us do this explicitly.

Let us show that the overall counterterm for G is local. We differentiate $\bar{R}(G)$ three times with respect to the external momentum p and show that the result is finite. Given the momentum routing of Fig. 5.4.1, there are three lines to differentiate: $p + k$, $p + l$, $p + q$. Here we have used the momentum carried by the line as a label for the line. Differentiating the original graph gives ten terms, where the three derivatives are applied to any combination of the three lines. One term is where we differentiate $p + k$ three times (Fig. 5.4.4). Although there is then no overall divergence, there remain subdivergences, so we must examine the corresponding differentiations applied to the counterterm graphs (*b*)–(*h*). We must regard the derivatives as acting on these graphs before divergences are computed (by the operation symbolized by the box).

The differentiation makes the subgraphs γ_1 and γ_3 completely finite, by removing both their overall divergence and γ_3's only subdivergence. Hence

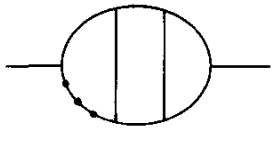


Fig. 5.4.4. One of the terms obtained by differentiating Fig. 5.4.1 three times with respect to its external momentum.

the counterterm graphs (*b*), (*d*), (*e*), and (*h*) are zero after differentiation. This leaves graphs (*c*), (*f*), and (*g*). These cancel the subdivergences of (*a*) coming from the subgraphs γ_2 and γ_4, which are unaffected by the differentiation.

We may examine the other nine terms in $\partial^3 \bar{R}(G)/\partial p^3$ similarly, and we find that in fact $\partial^3 \bar{R}(G)/\partial p^3$ is finite, as claimed.

5.5 Forest formula

5.5.1 *Formula*

Zimmermann (1969, 1970) gave an explicit solution of the recursive definition of the renormalized value $R(G)$ of a graph G. The general idea can be gleaned from the example we examined in the previous section. There the recursion generated a series of sixteen terms. One was the original graph, and the others had the subtraction operation T applied one or more times. For example, in graph (*e*) we first replace the left-most loop γ_1 by $T(\gamma_1)$, with a result we can write as $T_{\gamma_1}(G)$. We then take the subgraph equivalent to γ_3, viz. $T_{\gamma_1}(\gamma_3)$, and replace it by the result of acting with T. This gives graph (*e*). The sum of the two graphs (*d*) and (*e*) is used as the subtraction for the subdivergences of G associated with γ_3.

Each of the sixteen terms is pictured as the original graph with some number of connected 1PI subgraphs enclosed in boxes to indicate application of T to the subgraph. Each term can be specified by giving its set of boxed subgraphs. Each such set is called a *forest*. The subgraphs which form a particular forest are either disjoint or nested; they are said to be *non-overlapping*. The set of all possible forests for G is called $\mathscr{F}(G)$.

There are sixteen forests occurring in Fig. 5.4.3. The first eight are (in set theory notation):

(*a*) the empty set $\varnothing$,
(*b*) $\{\gamma_1\}$,
(*c*) $\{\gamma_2\}$,
(*d*) $\{\gamma_3\}$,
(*e*) $\{\gamma_1, \gamma_3\}$,
(*f*) $\{\gamma_4\}$,
(*g*) $\{\gamma_2, \gamma_4\}$,
(*h*) $\{\gamma_1, \gamma_2\}$.

These do not contain the whole graph; they are called the *normal* forests. The other eight forests in Fig. 5.4.3 consist of one of the above eight forests to which is added as another element the complete graph G. A forest of G

containing G is called a *full* forest. The distinction between normal and full forests is that the normal forests subtract off the subdivergences, and the full forests combine to subtract the overall divergence.

Not all forests occur in Fig. 5.4.3; for example, the forest

$$U = \{\text{subgraph consisting of lines carrying loop momentum } l\}$$

does not appear. Such forests contain at least one overall convergent subgraph as an element.

Inspection of Fig. 5.4.3 shows that

$$R(G) = \sum_{U \in \mathcal{F}(G)} \prod_{\gamma \in U} (-T_\gamma) G. \tag{5.5.1}$$

Here the sum is over all forests U of G. The operator T_γ replaces γ by $T(\gamma)$. Note that for nested γ's the T_γ's should be applied inside to outside. Equation (5.5.1) is called the forest formula; it is due to Zimmermann (1969). Suppose we compute $R(G)$ for an arbitrary graph G by using (5.5.1). Then, as we will prove shortly, the result is the same as if we used the recursive definition of $R(G)$ given in Section 5.3.

It is convenient to let the sum over forests be over all forests rather than only over those consisting of subgraphs that are superficially divergent; the extra forests give a zero contribution. The reason for doing this is that we will sometimes wish to change the definition of T so that we make subtractions for some convergent graphs, as well as for divergent graphs. For example, such a redefinition will be the key to proving the operator product expansion in Chapter 10.

We now have both a recursive and a non-recursive definition for the renormalization of a Feynman graph. It will prove very useful to have both definitions available. Different proofs will need different forms of the definition. In particular, proofs by induction on the number of loops of a graph will naturally use the recursive definition.

5.5.2 *Proof*

The proof of the forest formula is elementary, but somewhat involved. We first use (5.5.1), and the following equations:

$$\bar{R}(G) = \sum_{U \in \mathcal{F}(G)} \prod_{\gamma \in U} (-T_\gamma) \circ G; \tag{5.5.2}$$

$$C(G) = \begin{cases} -T_G \circ \bar{R}(G), & \text{if } G \text{ is 1PI,} \\ \prod [C(\gamma_i)], & \text{if } G \text{ is a disjoint union of 1PI } \gamma_i\text{'s,} \\ 0 & \text{otherwise,} \end{cases} \tag{5.5.3}$$

as definitions of $R(G)$, $\bar{R}(G)$, and $C(G)$. Here $\mathscr{F}(G)$ is the set of normal forests of G (i.e., those that do not contain G). These definitions are correct for the graph Fig. 5.4.1, as can be seen by inspection of Fig. 5.4.3.

Since the recursive definitions uniquely give $R(G)$, $\bar{R}(G)$, and $C(G)$ in terms of the operation T_γ, it suffices to show that (5.5.1)–(5.5.3) satisfy the recursion relations (5.3.6)–(5.3.9). Notice first that (5.3.6) and (5.3.9) are really the same, except for being applied to different graphs.

If $\bar{R}(G)$ given by (5.5.2) is correct, and if subgraphs are correctly renormalized, then (5.5.3) is equivalent to our original definition (5.3.6) of $C(G)$. Moreover, suppose that G is connected and one-particle-irreducible. Now each forest of G is either a normal forest, that is, a forest of which G is not an element, or it is a normal forest to which is adjoined G. Then the formula (5.3.7) for $R(G)$ is a direct consequence of (5.5.1)–(5.5.3) for such a graph. If G is not a union of 1PI graphs, then there is no overall divergence, and again (5.3.7) holds.

So it remains to prove the following:

(1) $R(G)$ is correct when G is a disjoint union of more than one 1PI graph. (Note that this case occurs in renormalizing the graph of Fig. 3.2.1(*b*), as we saw in Section 5.3.)
(2) $\bar{R}(G)$ is correct, i.e., it satisfies (5.3.8), for a general graph.

If G is a disjoint union of 1PI graphs γ_i, then each forest is a union of forests, one for each component. Then $R(G) = \prod_i R(\gamma_i)$, as we should expect. The problem is that this is not manifestly true in the recursive definition, where we make an overall subtraction for G. We dealt with this problem between (5.3.11) and (5.3.17).

Our proof of (5.3.8) is by induction on the size of G. Now a one-loop 1PI graph has no non-trivial subgraphs, so its only normal forest is the empty set. Then formula (5.5.2) collapses to $\bar{R}(G) = U(G)$, just as it should. This enables us to start the induction.

It remains to prove (5.3.8*b*). For this, observe that each forest U has a unique set of biggest subgraphs $M_1,\ldots,M_j$. Each M_i is contained in no bigger subgraph in U, and each $\gamma \in U$ is contained in some M_i. The existence and uniqueness of this set of M's is seen by considering pairs γ_i, γ_k of elements of U. Since γ_i and γ_k are non-overlapping, there are three possibilities:

(1) $\gamma_i \subset \gamma_k$, in which case remove γ_i from further consideration.
(2) $\gamma_k \subset \gamma_i$, in which case remove γ_k from further consideration.
(3) $\gamma_i \cap \gamma_k = \varnothing$, in which case leave both in.

Repeat until no further eliminations are possible; then the result is the set of M_i's.

The forest U is the union of a full forest, one for each M_i. We can write our definition (5.5.2) of $\bar{R}(G)$ as

$$\bar{R}(G) = G + \sum_{M_1,\ldots,M_n} \left\{ \prod_{i=1}^{n} (-T_{M_i}) \sum_{U_1 \in \bar{\mathscr{F}}(M_1)} \cdots \sum_{U_n \in \bar{\mathscr{F}}(M_n)} \times \right.$$
$$\left. \times \left[\prod_{\gamma_1 \in U_1} (-T_{\gamma_1}) \cdots \prod_{\gamma_n \in U_n} (-T_{\gamma_n}) \right] G \right\}. \tag{5.5.4}$$

Here the first term comes from the case in (5.5.2) that $U = \varnothing$, and the sum in the second term is over non-empty sets of disjoint 1PI graphs M_i *excepting the case that* $M_i = G$. By setting $\gamma = M_1 \cup \ldots \cup M_n$ and using (5.5.3) to determine $C(M_1 \cup \ldots \cup M_n)$, we find (5.3.8*b*).

5.6 Relation to $\mathscr{L}$

We have seen how to renormalize an individual Feynman graph by making a series of subtractions. The motivation for doing this came from consideration of examples in which the subtractions were generated by counterterms in the interaction Lagrangian. We will now show that this is true to all orders. We will assume the natural result (to be proved later) that the polynomial degree of the overall counterterm of a graph is given by its degree of divergence, just as for low-order graphs.

First, we must make precise the result that we will prove. For each 1PI graph G, we have constructed its overall counterterm $C(G)$. Since this is a polynomial in the external momenta of G, it can be written as the vertex derived from an interaction term $D(G)/N(G)$ in the Lagrangian $\mathscr{L}$. Here $N(G)$ is a symmetry factor of the same sort as the 3! that appears with the ϕ^3 interaction term in $\mathscr{L}$. Each power of a momentum entering $D(G)$ corresponds to i times a derivative of the corresponding field. If G is an n-point graph and each of its external lines corresponds to the same type of field, then $N(G)$ is $n!$. If there are a number of different fields and n_i is the number of lines of type i entering G then

$$N(G) = \prod_i n_i!. \tag{5.6.1}$$

For each graph for a Green's function, the forest formula gives a set of graphs with counterterms. We will demonstrate that the set of counterterm graphs is generated from the counterterm vertices in the interaction Lagrangian.

Consider ϕ^3 theory. As before, we write the Lagrangian as:

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_b + \mathscr{L}_{ct} = \mathscr{L}_0 + \mathscr{L}_I. \tag{5.6.2}$$

The free Lagrangian

$$\mathscr{L}_0 = (\partial\phi)^2/2 - m^2\phi^2/2 \tag{5.6.3}$$

generates the propagator, while the interaction $\mathscr{L}_1$ consists of two terms, $\mathscr{L}_b$ and $\mathscr{L}_{ct}$. The basic interaction is

$$\mathscr{L}_b = -g\mu^{3-d/2}\phi^3/3!, \tag{5.6.4}$$

and $\mathscr{L}_{ct}$ is the counterterm Lagrangian used to cancel the ultra-violet divergences:

$$\mathscr{L}_{ct} = \sum_G D(G)/N(G). \tag{5.6.5}$$

Here the sum is over all 1PI graphs. Those that have no overall divergence generate no counterterm; for these $D(G) = 0$. Each 1PI graph that has an overall divergence generates a term in (5.6.5). The formulae (5.6.2) and (5.6.5) apply in any theory.

Since (5.6.5) applies to any theory, it applies in particular to ϕ^3 in higher than six dimensions. Thus it enables us to renormalize a non-renormalizable theory. But the sum must include counterterms $D(G)$ with an arbitrarily large number of powers of momentum and with an arbitrarily large number of external lines for G. It is only in six or fewer dimensions that the counterterms have the same form as terms in the basic ϕ^3 Lagrangian $\mathscr{L}_0 + \mathscr{L}_b$.

Now that we have developed a convenient notation, the most difficult part of the proof is to ensure that the combinatorial factors come out right. We will prove that the Lagrangian defined by (5.6.2) and (5.6.5) gives the same renormalized Green's functions as those generated by our recursive definition in Section 5.3 (and therefore the identically same Green's functions as given in Section 5.5 by the forest formula). The proof will be given for ϕ^3 theory in six or fewer dimensions, but it easily generalizes.

Consider the full N-point Green's function G_N at order g^P. It is sufficient to work only with connected graphs. If the theory is renormalizable (as we will prove in Section 5.7), then the sum of counterterms has the form:

$$\mathscr{L}_{ct} = \delta Z(\partial\phi)^2/2 - \delta m^2\phi^2/2 - \delta g\phi^3/3!. \tag{5.6.6}$$

with (by (5.6.5))

$$\begin{aligned} -\delta Z &= \sum_{2\text{-point } G} [\text{Coefficient of } -ip^2 \text{ in } C(G)], \\ -\delta m^2 &= \sum_{2\text{-point } G} [\text{Coefficient of } ip^0 \text{ in } C(G)], \\ -\delta g &= \sum_{3\text{-point } G} C(G)/i. \end{aligned} \tag{5.6.7}$$

We ignore the tadpoles, yet again. The term of order g^P in the perturbation

expansion of G_N has vertices generated by the different terms in $\mathscr{L}_b + \mathscr{L}_{ct}$. There will be graphs with all of their vertices being the basic interaction $\mathscr{L}_b$. Let the set of these be called B. The other graphs will contain one or more of the counterterm vertices generated by (5.6.6). Each counterterm can then be decomposed into a sum of terms by applying (5.6.7) at each counterterm vertex. Each term has each of the counterterm vertices replaced by the overall divergence of some graph. Then in the result, each term T corresponds to a unique basic graph $b(T) \in B$.

So we have

$$G_N = \sum_G \left(G + \sum_{b(T)=G} T \right). \tag{5.6.8}$$

On the other hand, we have constructed the renormalization of each of the graphs G by writing

$$R(G) = G + \sum_{\gamma \subseteq G} C_\gamma(G). \tag{5.6.9}$$

Each of the terms T in (5.6.8) is constructed by replacing each of a set of one or more disjoint 1PI subgraphs $\gamma_1, \ldots, \gamma_j$ by its counterterm given by $iD_{\gamma_i}(G)$. On identifying γ in (5.6.9) with $\gamma_1 \cup \gamma_2 \cdots \cup \gamma_j$, we expect that

$$G_n = \sum_G R(G). \tag{5.6.10}$$

This result would be obvious, were it not that the symmetry factors do not manifestly match up.

The problem is illustrated by Fig. 5.6.1. There the basic graph is (*a*), and the complete set of subtractions needed to renormalize it consists of (*b*)–(*e*). Now the symmetry factor for (*a*) is 1/8: There is a factor 1/2 for each self-energy graph and an overall 1/2 for the top–bottom symmetry of the whole graph. Each of the subtractions (*b*) and (*c*) has a symmetry factor 1/4, since the remaining factor 1/2 goes into the counterterm for the self-energy. Both graphs (*b*) and (*c*) are equal.

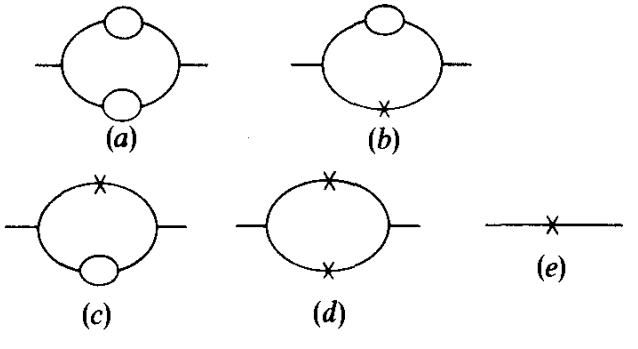


Fig. 5.6.1. Renormalization of a graph, to illustrate symmetry factors.

Considered as Feynman graphs, these are the same graph, but with symmetry factor 1/2. So they give one term with factor 1/2 in (5.6.8) (derived from the Lagrangian), while in (5.6.9) (from the recursion formula) there are two equal terms with factor 1/4. The end result is the same. We must consider (*b*) and (*c*) as distinct graphs when defining $R(G)$, since they correspond to different regions of loop-momentum space – we must take each momentum variable to be distinguishable.

To construct a general proof is tedious. The symmetry factor of a graph G is $1/N(G)$, where $N(G)$ is the dimension of the graph's symmetry group. So we write

$$G = \bar{G}/N(G), \tag{5.6.11}$$

where the overbar indicates computation ignoring all symmetry factors. Similarly we define $\overline{C(G)}$ by

$$C(G) = \overline{C(G)}/N(G). \tag{5.6.12}$$

Now the renormalized value of a graph G is

$$\begin{aligned} R(G) &= G + \sum_{\gamma \subseteq G} C_\gamma(G) \\ &= \frac{1}{N(G)}\left[\bar{G} + \sum_\gamma \overline{C_\gamma(G)}\right] \\ &= \frac{1}{N(G)}\bar{G} + \sum_\gamma \left(\frac{\prod_i N_i}{N(G)}\right)\left[\frac{1}{\prod_i N_i}\overline{C_\gamma(G)}\right]. \end{aligned} \tag{5.6.13}$$

In the last line we have observed that γ is a disjoint union of 1PI graphs $\gamma_1, \gamma_2, \ldots$. Moreover, we have explicitly indicated the symmetry factors $1/N_i \equiv 1/N(\gamma_i)$ for each γ_i which is replaced by its overall counterterm. For a given subgraph $\gamma = \cup \gamma_i$ the symmetry groups of the γ_i's are a commuting set of subgroups of the symmetry group of G. Therefore the quantity $N(G)/\prod_i N_i$ must be an integer.

Next, consider the Green's functions generated by the Lagrangian (5.6.2), as in (5.6.8),

$$G_N = \sum_G \left\{\frac{1}{N(G)}\bar{G} + \sum_{\gamma \subseteq G}{}' \frac{1}{N(G/\gamma)}\overline{C_\gamma(G)}\right\}. \tag{5.6.14}$$

Here we have observed that each graph containing one or more counterterms is generated from a basic graph by replacing some 1PI subgraphs $\gamma_1, \ldots, \gamma_j$ by counterterms. We write γ as the union of the γ_i's. Then we let G/γ be the graph resulting from substituting counterterms for the γ_i's. By the

definition of the counterterm Lagrangian, the result is the same as $C_\gamma(G)$, aside from symmetry factors. The prime on the $\sum'$ indicates that only γ's giving distinct Feynman graphs are considered. (Thus, for example, Figs. 5.6.1(*b*) and (*c*) are not counted separately.) Thus, if we define

$$K(G,\gamma) = [\text{number of graphs } \gamma' \text{ for which } G/\gamma = G/\gamma'],$$

then we must prove that

$$K(G,\gamma) = \frac{N(G)}{\left[\prod_i N_i\right][N(G/\gamma)]}. \tag{5.6.15}$$

It is easiest to couch this final step in the language of group theory. The denominator of the right-hand side of (5.6.15) is the product of dimensions of commuting subgroups of the symmetry group of G. (Note that, for example, two ϕ^3 counterterms generated by different self-energy subgraphs are counted as different.) These subgraphs generate another subgroup, of which the set of cosets in the symmetry group of G has exactly the dimension of the right-hand side of (5.6.15). But, concretely, each coset corresponds to one of the graphs counted by $K(G,\gamma)$.

5.7 Renormalizability

5.7.1 *Renormalizability and non-renormalizability*

In this section we explain the properties of renormalizability, non-renormalizability, and super-renormalizability of a field theory. We do this first for every order of perturbation theory, and then we consider to what extent the properties are true beyond perturbation theory, for the complete theory. The method in perturbation theory is power-counting and dimensional analysis.

Consider first ϕ^3 theory in a space-time of integer dimension d_0. We have seen how to renormalize it to get finite Green's functions by adding counterterms (5.6.5) to the Lagrangian. Each counterterm is a polynomial in the field ϕ and its derivatives. The theory is called renormalizable if the only counterterms needed are proportional to the terms $(\partial\phi)^2$, ϕ^2, and ϕ^3 present in the original Lagrangian $\mathscr{L}_0 + \mathscr{L}_b$. This is equivalent to saying that the Lagrangian has the form

$$\mathscr{L} = (\partial\phi_0)^2/2 - m_0^2\phi_0^2/2 - g_0\phi_0^3/3!, \tag{5.7.1}$$

where the bare field ϕ_0 is $Z^{1/2}\phi$. The bare mass m_0, the bare coupling g_0, and the field-strength renormalization Z each have singular behavior as the

ultra-violet regulator is removed. A linear term $\delta h\phi$ is needed as well. We may regard it as being present in the original Lagrangian. In any event it is only a single extra coupling. It can be ignored if we impose the renormalization condition that $\langle 0|\phi|0\rangle = 0$ to determine δh, and if we ignore tadpole graphs.

We generalize to an arbitrary theory by calling a theory renormalizable if the Green's functions of its elementary fields can be made finite by rescaling the fields (in a cut-off dependent way) and by making some suitable cut-off dependent change in the couplings and masses.

In perturbation theory we determine whether or not we have renormalizability by examining the possible values of the degree of divergence $\delta(G)$ for the 1PI graphs. For every graph G a counterterm is needed if $\delta(G) \geq 0$. As we will prove in Section 5.8 the counterterm $C(G)$ is a polynomial of degree $\delta(G)$ in the external momenta, and provided we use a scheme like dimensional regularization that preserves Poincare invariance, the counterterms are Poincare invariant.

Let us now determine whether or not ϕ^3 theory in d space-time dimensions is renormalizable. In d space-time dimensions the N-point 1PI graphs have dimension (in momentum space)

$$d(G_N) = N + d - Nd/2.$$

Then (by (3.3.12)) the degree of divergence of a graph for G_N at order g^P is

$$\delta(G_N) = d + (1 - d/2)N + (d/2 - 3)P. \tag{5.7.2}$$

Note that the minimum value of P to have a one-loop connected graph is N.

Inspection of (5.7.2) shows that if $d > 6$ then, for any value of N, there can be made N-point graphs with arbitrarily high degree of divergence by going to large enough order in g. The theory is therefore not renormalizable if $d > 6$, and the non-renormalizability is a direct consequence of the negative dimension of g.

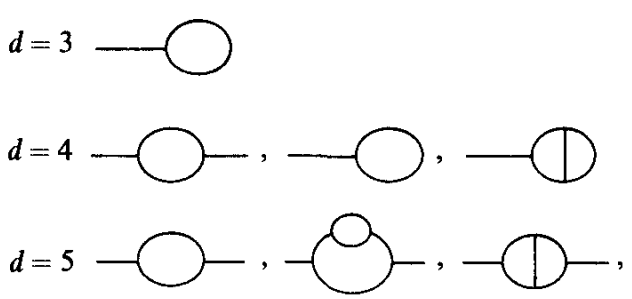


Fig. 5.7.1. All the graphs with overall divergences in ϕ^3 theory at those space-time dimensions where it is super-renormalizable.

If $d = 6$, only the one-, two-, and three-point functions are divergent, with degree of divergence 4, 2, and 0, respectively. The permissible counterterms are just terms of the form of those in $\mathscr{L}_0 + \mathscr{L}_b$, so the theory is renormalizable if $d = 6$. Moreover, there is a divergence in every order of g (except for tree graphs, of course).

If $d = 3$, 4, or 5, then only a finite set of graphs, illustrated in Fig. 5.7.1, have overall divergences, and renormalization is needed only for the mass and for the tadpole coupling. Again we have renormalizability.

5.7.2 *Cosmological term*

Strictly speaking, we should also consider Feynman graphs with no external lines. These are the vacuum bubbles. They generate the energy density of the vacuum, and normally are ignored. But in gravitational physics, they cannot be ignored. Counterterms for such graphs (present in ϕ^3 theory whenever $d \geq 2$) are proportional to the unit operator. They are a renormalization of what in General Relativity is the cosmological constant. A counterterm is even needed for free-field theory – where the divergence is conventionally removed by normal-ordering (see, for example, Bjorken & Drell (1966)). We see that normal-ordering is nothing but a primitive form of renormalization.

5.7.3 *Degrees of renormalizability*

It is convenient to distinguish three types of renormalizable theory:

(1) Finite: no counterterms needed at all.
(2) Super-renormalizable: only a finite set of graphs need overall counterterms.
(3) Strictly renormalizable: infinitely many graphs need overall counterterms. (But note that they only renormalize a finite set of terms in the basic Lagrangian, since we assumed renormalizability of the theory.)

Finiteness or super-renormalizability normally occur when all the couplings in the basic Lagrangian have positive dimension.

Note that in a super-renormalizable theory, the number of divergent basic graphs is infinite. For example, even if there is only one graph γ with an overall divergence, any graph containing γ as a subgraph is divergent. However all such graphs become finite after adding to γ its counterterm, so only one counterterm, $C(\gamma)$, appears in the Lagrangian.

Mathematical physicists (see Glimm & Jaffe (1981)) have investigated renormalizability beyond perturbation theory. This is important, since

perturbation series are in general asymptotic series rather than convergent series. Thus one cannot simply sum the perturbation series to obtain the complete theory. Even so, it has been proved for many super-renormalizable theories that perturbation theory gives an exactly correct account of the divergences. (A much investigated case is ϕ^4 theory in two and in three space-time dimensions.)

In a super-renormalizable theory, the series for a bare mass or coupling in terms of the renormalizable quantities has a finite number of terms. Therefore the series converges, and one only has to prove that (*a*) perturbation theory is asymptotic to the true theory, and (*b*) there are no terms like $\exp(-1/g)$ in the bare masses or couplings that are smaller than any power of g. The rigorous proof amounts to showing that in summing the perturbation series to a finite order, the error is correctly estimated by the first term omitted. In particular, this applies to the existence of any possible ultra-violet divergence.

Rigorous proofs are not yet available for any strictly renormalizable theory. One difficulty is obvious: the series for, say, the bare coupling, g, is an infinite series, each term of which diverges as the UV cut-off is removed. Since the series is presumably asymptotic rather than convergent, one cannot directly obtain any information about renormalization in the full theory: the error obtained in using a truncated form of the series is of the order of the first term omitted, and that is always divergent.

It might even appear that perturbation theory has no light at all to shed on the question of renormalizability of the full theory. This is in fact not so, as we will see when we discuss the renormalization group in Chapter 7. If the theory has the property called asymptotic freedom then a series of suitable redefinitions of g allows short-distance phenomena to be computed reliably. In particular the UV divergences can be computed in terms of weak coupling series without divergent coefficients. It is sensible to conjecture that a suitably refined analysis can be made to obtain rigorous bounds of the errors so that the perturbative results correctly give the divergences. Monte-Carlo studies of the functional integral (Creutz & Moriarty (1982)) support this conjecture. In four dimensions, only certain non-abelian gauge theories (including QCD) are asymptotically free (Gross (1976)).

We will also see in Chapter 7 that in non-asymptotically free theories, like ϕ^4 and QED in four dimensions, perturbation theory cannot reliably describe short-distance phenomena. There are, in fact, indications (Symanzik (1982)) that the full ϕ^4 theory is not renormalizable, contrary to the situation order-by-order in perturbation theory.

5.7.4 *Non-renormalizability*

For theories which are not renormalizable in perturbation theory, there are many possibilities. Among them are the following:

(1) There is only a finite set of 1PI Green's functions which have overall divergences. A typical case is ϕ^3 theory in six or fewer space-time dimensions when the basic Lagrangian,

$$\mathscr{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - g\phi^3/3!, \tag{5.7.3}$$

has no term linear in ϕ. The one-, two-, and three-point functions have divergences, but there is no term $h\phi$ whose coupling can be renormalized to cancel the divergence of the tadpole graphs. However, addition of such a term generates a renormalizable theory. More generally, suppose we have a finite set of overall-divergent Green's functions. A renormalizable theory is generated by adding a finite set of extra interactions.

(2) There is an infinite set of Green's functions with overall divergences. However, for all but a finite set of the Green's functions, the divergences cancel after summing over all graphs of a given order. (There are no known cases of this.)

(3) As for case 2, except that the divergences cancel only for the S-matrix, rather than for all off-shell Green's functions. An important case is a spontaneously broken gauge theory, when it is quantized in its unitary gauge.

(4) The theory is made renormalizable by going beyond perturbation theory in some systematic and sensible way. One case (as in the Gross–Neveu (1974) model – see Gross (1976)) is of a theory that is strictly renormalizable and asymptotically free for some dimension $d = d_0$, and that is considered in some dimension d slightly greater than d_0.

(5) As for case 1, except that the extra terms make physical nonsense. A case is the Yang–Mills theory with a mass term in Feynman gauge. Then the extra terms destroy unitarity ('t Hooft (1971a)).

(6) None of the above.

Roughly speaking, there are no general rules. Each case must be handled separately. Only for the last two cases (5 and 6) should a theory be called non-renormalizable. A fundamental theory should be renormalizable, for otherwise either physical quantities are actually infinite or they are finite, but an infinite set of parameters is needed to specify the finite parts of the counterterms.

Nevertheless, a statement that a particular theory is non-renormalizable

is really a statement of ignorance: nobody has found a way to construct a physically sensible version of the theory. (Cases 1 to 5 are where somebody has found a way.) In practice, when a theory is labelled non-renormalizable, what is usually meant is that the theory is not renormalizable order-by-order in perturbation theory; such a statement can be proved by calculating a finite number of graphs.

Within the usual functional-integral approach (with a lattice cut-off), not only has the complete ϕ^4 theory been proved renormalizable for $d = 2$ and 3, but it has been proved non-renormalizable for $d > 4$ (Aizenman (1981)).

5.7.5 *Relation of renormalizability to dimension of coupling*

To prove perturbative renormalizability of a theory of scalar fields, we generalize the argument leading to (5.7.2). The argument will apply when no coupling has negative dimension. Renormalizability will hold with possibly the addition of extra interactions (like the $h\phi$ term in ϕ^3 theory) whose coefficients have non-negative dimension. Our proof will easily generalize to theories with fermion and gauge fields. The problems we will encounter in gauge theories will all be to do with the question of whether these extra terms are compatible with the gauge invariance. But we will leave these questions to Chapter 12.

Let a general term in $\mathscr{L}$ or $\mathscr{L}_{ct}$ be written schematically as

$$(\text{coupling } f)(\text{derivative})^A(\text{field})^N. \tag{5.7.4}$$

The vertex generated by this term is one possible graph for the 1PI Green's function Γ_N with N external lines. Thus the dimension of Γ_N satisfies

$$d(\Gamma_N) = d(f) + A. \tag{5.7.5}$$

Since no coupling has negative dimension, the degree of divergence of any graph for Γ_N is at most $d(\Gamma_N)$, as we saw from examples in Section 3.3.3, and as we will prove in Section 5.8. That is, the degree of divergence $\delta(\Gamma_N)$ satisfies

$$\delta(\Gamma_N) \leq d(\Gamma_N), \tag{5.7.6}$$

with equality only for graphs all of whose couplings have zero dimension.

To renormalize the N-point graphs, we add counterterms of the form (5.7.4) with at most $\delta(\Gamma_N)$ derivatives. So the possible counterterms satisfy

$$d(f) = d(\Gamma_N) - A \geq \delta(\Gamma_N) - A \geq 0. \tag{5.7.7}$$

The last inequality follows since a counterterm with A derivatives is needed only if the degree of divergence is at least A. From (5.7.7) it follows that we

need no couplings of negative dimension, given that none are present in the original Lagrangian.

Some of the couplings generated as counterterms may not be present in the original Lagrangian even if it contains no couplings of negative dimension. But the number of new couplings needed is nevertheless finite, because only a finite set of counterterms satisfy (5.7.7).

5.7.6 *Non-renormalizable theories of physics*

From the discussion above, it is natural to conclude that a theory of physics should be renormalizable. In fact, the strong, electromagnetic, and weak interactions appear to be described by a renormalizable theory. This theory is a combination of quantum chromodynamics for strong interactions and the Weinberg–Salam theory for weak interactions.

Around 1970 there was a revolution in the theory of weak interactions when it was discovered that non-abelian gauge theories are renormalizable. It is precisely one of these theories that was found to be necessary to construct a renormalizable theory of weak interactions in agreement with experiment. See Beg & Sirlin (1982) for a historical review.

Unfortunately, this progress has not extended to gravity. Einstein's theory of general relativity is non-renormalizable, after quantization, and there is no very promising alternative. (This situation exists despite many significant attempts to improve it – Hawking & Israel (1979).)

It is a mistake to suppose that non-renormalizable theories should be banished from consideration. Remember that for many years weak interactions were *successfully* calculated using the 'four-fermion' theory, which is non-renormalizable. For most purposes, weak interactions could be adequately treated in the lowest order of perturbation theory, where no renormalization is needed. But the non-renormalizability of higher-order calculations raised the question of consistency of the theory: is it legitimate to calculate even an approximation to a nonsensical (i.e., non-existent) theory? Will the results of calculations mean anything? The same questions arise in gravity. There, the classical theory of general relativity is very successful, but the quantized theory is badly non-renormalizable.

We must therefore understand how and why we may use non-renormalizable theories in physics.

Now, to perform consistent calculations in any theory which contains ultra-violet divergences, we must impose an ultra-violet cut-off, M, of some sort. In the case of a renormalizable theory we can take M to infinity and obtain finite results that are insensitive to the cut-off. Another related

property of a renormalizable theory is the decoupling theorem of Appelquist and Carazzone, which we will discuss in Chapter 8. This theorem applies to a renormalizable theory which contains fields whose masses are much bigger than the energies of the scattering processes under consideration. The theorem states that the heavy fields can be deleted with only a small effect (suppressed by a power of the heavy mass) on cross-sections, etc. The hallmark of a renormalizable theory is in fact that it is complete in itself. It contains no direct indications of whether it is only part of a larger and more complete theory.

These statements are false for a non-renormalizable theory. Consider the old four-fermion theory of weak interactions. Its coupling is $G \sim 10^{-5}$ GeV^{-2}. We cannot take the UV cut-off arbitrarily large, for an nth order graph has a divergence of order $2n$; it behaves like $(M^2G)^n$ for large cut-off M. Counterterms to make the graph finite need a correspondingly large number of derivatives, but only a finite number of counterterms are available. Hence we cannot take the cut-off to infinity, and if we want insensitivity to the cut-off we must take $M \ll G^{-1/2}$. Moreover, the energy, E, of the process under consideration must be much less than M, otherwise the calculation is dominated by details of the cut-off procedure. In other words, the four-fermion interaction is a good approximation to physics only if $E \ll M \ll G^{-1/2}$. The minimum possible relative error of calculations is of the order of the maximum of M^2G and E^2/M^2.

Now, it is always possible in principle to do experiments at arbitrarily high energy. So the applicability of four-fermion theory at low energies implies that at energies rather below $G^{-1/2} \sim 300$ GeV there is new physics. That is, the four-fermion theory becomes incorrect at that energy. The last fifteen years of weak interaction physics confirms this. (See, for example, Bjorken (1982))

For gravity, the corresponding energy scale is the Planck mass, of the order of 10^{19} GeV. This is extremely far beyond the range of normal accelerator experiments. Evidence for phenomena on such an energy scale must come from much more esoteric observations. Examples might be found in certain areas of the cosmology of the early universe, or from seeing the decay of a proton (Langacker (1981)).

In any case, a non-renormalizable theory contains indications that it cannot describe all phenomena. It contains the seeds of its own destruction as a viable theory of a field of physics. So, given a successful non-renormalizable theory, one must ask the following questions: (1) 'Of which more complete theory is it a part?' (2) 'How is it related to that theory?' An example is given by the relation between the Weinberg–Salam theory,



Fig. 5.7.2. W-boson exchange gives an effective four-point interaction at low energies.

which is renormalizable, and the four-fermion theory, which is not. The four-fermion theory arises as an approximation to W-boson exchange at low energy (Fig. 5.7.2). One replaces the propagator

$$\mathrm{i}/(q^2 - m_W^2),$$

for the W-boson, by

$$\mathrm{i}/(-m_W^2).$$

The graph is suppressed by factor of at least E^2/m_W^2 compared to photon exchange. It gives an example of the decoupling theorem: the heavy particles have small effects at low energies.

The only reason we can see such effects is the high degree of symmetry of the strong and electromagnetic interactions. These interactions conserve P, C, T, and the number of each flavor of quark and of each flavor of lepton. Weak-interaction amplitudes are much smaller than strong-interaction or electromagnetic amplitudes for similar processes, and are therefore normally invisible. But there are many processes that are completely forbidden in the absence of weak interactions; for these, any weak-interaction cross-section, no matter how small, is all there is.

So one important way in which a non-renormalizable theory arises is as a low-energy approximation to a renormalizable theory in a process that is forbidden in the absence of the heavy fields. The heavy fields effectively give a cut-off on the non-renormalizable theory. Then, for example, the four-fermion coupling G is computable in terms of the underlying theory via a formula like

$$G = \text{constant } g^2/m_W^2 + \text{higher order corrections in } g.$$

Here g is the dimensionsless coupling of the Weinberg–Salam theory. One manifest characteristic of this non-renormalizable theory is the weakness of its interactions. Also note that a higher power of G is a higher inverse power of m_W^2. We are taking the leading power of m_W as m_W gets large, so it is in general incorrect to calculate in the non-renormalizable theory beyond lowest order. Higher-order calculations must be done in the full theory.

A slightly different situation arises in gravity. There one must perform calculations beyond tree approximation, since gravitationally bound states

like the solar system are formed by multiple exchange of gravitons. Counterterms are generated involving higher-derivative interactions (e.g., R^2, $R_{\mu\nu}^2$, etc.). The ambiguity in the finite parts of these counterterms gives an uncertainty in the Green's functions. However the uncertainty is a power of momentum divided by some large mass scale, and is negligible for low-momentum-transfer processes. In weak interactions, the size of the higher-order corrections is of the same order of magnitude as the intrinsic error in the calculations, but in gravity this is not so because of the zero mass of the graviton.

Another difference is that gravity is actually the strongest of the four fundamental interactions when considered on a large enough scale. In contrast, on atomic or molecular scales, it is the other three interactions that are by far the strongest. However, the strong and weak interactions have a finite range, so that they are essentially zero outside the nucleus. Particles can have both signs of electric charge, so that bulk matter, if charged, tends to attract charge of the opposite sign to it. Bulk matter is therefore generally neutral. But gravity couples to mass or energy, so it is always attractive. Hence gravity wins out as the strongest interaction for large enough assemblages of matter. However at nuclear and atomic scales, it is negligible by a factor of about 10^{40} compared to the other interactions.

Let us summarize by restating the key conclusion about the distinction between renormalizable and non-renormalizable theories. A non-renormalizable theory considered at low energy gives some indications that at high enough energies it must break down, and cannot be a complete theory. A renormalizable theory gives no such indication.

5.8 Proof of locality of counterterms; Weinberg's theorem

In our examples, we saw that the counterterm $C(G)$ of a graph G is a polynomial in its external momenta, of degree equal to its overall degree of divergence $\delta(G)$. This is a general property, as we will now prove.

The original proof of this theorem and some related results is due to Weinberg (1960); a simpler proof was given by Hahn & Zimmermann (1968). It is useful to distinguish three results:

(1) Suppose that a 1PI graph G and all its 1PI subgraphs have negative degree of divergence. Then the graph is finite. That the degrees of divergence of the graph and subgraphs are negative means that there is no divergence when all or some of the loop momenta go to infinity together, with the other momenta finite. The problem is to eliminate the possibility of a divergence from more exotic scalings.

(2) Suppose that a 1PI graph G has negative degree of divergence, but that it might have subdivergences. Then the graph is finite if we first subtract off subdivergences. More simply, if $\delta(G) < 0$ then $\bar{R}(G)$ is finite.
(3) If a 1PI graph G has degree of divergence $\delta(G)$, then its overall counterterm $C(G)$ is polynomial in the external momenta of G of degree $\delta(G)$.

Property (1) is a trivial case of (2). We will reduce (3) to (2) by the same differentiation method as we used in Section 5.2.2. The proofs will be by induction. This naturally suggests that we use the recursive definition of the renormalization $R(G)$ of G.

One generalization is useful. It is that the renormalization prescription may be chosen so that result (3) reads ('t Hooft (1973), Weinberg (1973), and Collins (1974)):

(3′) If a 1PI graph G has degree of divergence $\delta(G)$, then its overall counterterm $C(G)$ is polynomial in the external momenta of G and in the massive parameters in the Lagrangian. (The parameters in question are the masses of fermions and the squared masses of bosons.) The dimensions of the terms in the polynomial are at most $\delta(G)$.

5.8.1 *Degree of counterterms equals degree of divergence*

We first prove Property (3), that the overall counterterm $C(G)$ of a 1PI graph is polynomial in the external momenta of degree $\delta(G)$. We will do this assuming Property (2), that a graph with its subdivergences subtracted is finite if its degree of divergence is negative. Let G be a 1PI graph with degree of divergence $\delta(G) \geq 0$. We will consider $\bar{R}(G)$, which is G plus counterterms for its subdivergences. Following Caswell & Kennedy (1982), let us differentiate the graph $\delta(G)+1$ times with respect to external momenta. This produces a result that has negative degree of divergence. We differentiate not only the graph G, but also its various counterterm graphs $C_\gamma(G)$. The aim is to show that the result is actually convergent. To do this we will show that the differentiated counterterm graphs are the correct counterterm graphs for the differentiated original graph. This may sound obvious, but there are some subtleties, so we will give the details.

Let the external momenta of G be $p_1, \ldots, p_n$. Its renormalized value is

$$\begin{aligned} R(G) &= G + \sum_{\gamma \subseteq G} C_\gamma(G) \\ &= \bar{R}(G) + C(G) \\ &= \bar{R}(G) - T_G \circ \bar{R}(G). \end{aligned} \tag{5.8.1}$$

Let ∂ denote differentiation with respect to one of the external momenta, and let ∂^λ denote any λ-fold differentiation with respect to the external momenta. The property to be shown is that $\partial^\lambda \bar{R}(G)$ is finite if $\lambda > \delta(G)$. (It is clear from naive power-counting that $\delta(\partial^\lambda G) = \delta(G) - \lambda$.)

Suppose we ensure that differentiation commutes with the basic subtraction operator T_γ. This amounts to imposing a very natural relation between the finite parts of, for example, $T_\gamma(\partial G)$ and $T_\gamma(G)$. (The relation is satisfied by the pole-part subtractions, but it is possible to choose exotic renormalization prescriptions not satisfying the hypothesis.) Then for any graph γ we have

$$\partial C(\gamma) = C(\partial\gamma). \tag{5.8.2}$$

It follows that, for the original graph, we have

$$\partial^\lambda \bar{R}(G) = \bar{R}(\partial^\lambda G). \tag{5.8.3}$$

The point here is that a differentiation when acting on a graph gives a number of terms, in each of which one of the propagators or vertices is differentiated. It is a simple generalization of the argument given in Sections 5.2.2 and 5.2.3 for specific graphs that the counterterms for subgraphs of G are the correct ones after differentiation in (5.8.3).

Now $\bar{R}(\partial^\lambda G)$ is the sum of a graph $\partial^\lambda G$ that has negative degree of divergence and the counterterm graphs for its subdivergences. Hence by Property (2) it is finite, so that we may choose the subtraction operator T to give zero. Therefore the counterterm in (5.8.1) for the undifferentiated graph is polynomial of degree $\delta(G)$ in the external momenta.

The same argument (Collins (1974)) also shows that counterterms are polynomials in mass. Here it is necessary to note that differentiation with respect to m^2 does not automatically reduce the degree of divergence. This only happens if counterterms for subdivergences are polynomial. If a counterterm has a piece proportional to $\ln(m^2)$ then differentiating with respect to m^2 leaves the degree of divergence unchanged. The proof merely demonstrates that it is always possible to choose counterterms to be polynomial in m^2; it is not compulsory.

5.8.2 $\bar{R}(G)$ *is finite if* $\delta(G)$ *is negative*

It was evident in one-loop examples that a graph with degree of divergence δ is renormalized by a local counterterm of degree δ in the external momenta. To generalize the result to an arbitrary graph, we constructed a renormalization procedure which involved computing the counterterm for

a 1PI graph only after subtracting subdivergences. We differentiated the graph $\delta + 1$ times with respect to its external momenta to prove its counterterm to be local and of degree δ. This proof relied on assuming the following statement:

> If a graph has negative degree of divergence and has its subdivergences subtracted according to the rules, then it is finite. More briefly, if $\delta(\Gamma) < 0$ then $\bar{R}(\Gamma)$ is finite.

This statement sounds extremely plausible. It is nevertheless in need of proof. We have to ensure that the subtraction procedure actually accomplishes its purpose of removing the subdivergences. (That is, there are no spurious divergences induced by the procedure.) In addition, we normally only consider the divergences as arising from regions in which some loop momenta go to infinity, all at the same rate; this generates the usual power-counting. It is necessary to eliminate more exotic possibilities.

The most important problem, which is the one we will examine, is to treat the case that a collection of loop momenta go to infinity, but at different rates. In Section 5.2.2, we examined the special case of Fig. 5.2.1. The general case is very similar. Inductively, we assume that properties (1) to (3), listed at the beginning of Section 5.8, are true for all smaller graphs than the graph G under consideration. We consider regions of the integration over loop momenta where all or some momenta go to infinity, not necessarily at the same rate. We will eliminate them as possible sources of additional divergences. If all the momenta go to infinity together, then the negative overall degree of divergence means that there is no actual divergence from this region.

If some momenta stay finite while the others go to infinity (not necessarily at the same rate), then let γ be the subgraph consisting of all those lines with the large momenta. Our inductive hypothesis ensures that all the resulting divergences are cancelled by counterterms for subgraphs.

The remaining case is that *all* of the loop momenta go to infinity, but again not at the same rate. Let k denote the components of the smallest momenta, and let l denote the rest. (Our notation is meant to copy that used for Fig. 5.2.1, and so is the proof.) Let γ be the subgraph consisting of all those lines carrying the loop momenta l. It may be a single 1PI graph or a disjoint union of 1PI subgraphs. Let these 1PI subgraphs be $\gamma_1, \ldots, \gamma_L$. Expand each subgraph in powers of its external momenta up to its degree of divergence. The remainder for each subgraph is really a graph with negative overall degree of divergence; the contribution vanishes as l goes to infinity, so we should have set $l = O(k)$. The expanded terms contribute just as they

did for Fig. 5.2.1. After subtraction of divergences we have a factor in the integral over k corresponding to the dimension of the subgraph.

We gloss over here some of the important details, notably what happens to the value of a general subgraph when some of its external momenta get large. But the main lines of the argument should be apparent.

The structure of the proof is the same as in subsection 5.2.2 and the result is the same.

5.8.3 *Asymptotic behavior*

Weinberg (1960) not only proved the convergence theorem stated above with more complete rigor, but he also investigated what happens when several of the external momenta $p_1, p_2, \ldots, p_n$ of a graph γ get large in the Euclidean region. They are assumed all to be of an order Q, with the ratios p_j^μ/Q fixed as $Q \to \infty$. None of the sums of subsets of p_j^μ/Q vanish. Weinberg then states how to find the asymptotic behavior:

(1) Consider any subgraph γ connected to all the lines carrying the large momenta. Let all the loop momenta of γ be of order Q. Compute the power of Q: Q^{a_γ}.

(2) Look at all such subgraphs. Let a_γ have a maximum value a.

(*a*) If there is a unique graph with this maximum power, then $\Gamma \propto Q^a$ as $Q \to \infty$.

(*b*) If there are several subgraphs with $a_\gamma = a$, then let N be the number of such subgraphs. The asymptotic behavior is:

$$\Gamma = Q^a[A_0 B_0 + A_1 B_1 \ln Q + A_2 B_2 (\ln Q)^2 + \cdots A_{N-1} B_{N-1} (\ln Q)^{N-1}] + O(Q^{a-1}). \tag{5.8.4}$$

Here the A_i's are functions of those momenta that are fixed as $Q \to \infty$ and the B_i's are functions of the finite quantities p_j^μ/Q.

This theorem is needed inductively in the guts of the convergence theorem proved in the last subsection. Its proof is similar.

It is not obvious that this part of Weinberg's theorem is of much use for physics, other than for its part in this convergence proof, since the asymptotic behavior is of Euclidean momenta. However, in the deep-inelastic scattering of a lepton on a hadron, there is a photon or a weak interaction boson that is far off-shell. The momentum carried by the boson is effectively Euclidean, and Weinberg's theorem applies. We will see this in Chapter 14. There are also generalizations to other intrinsically Minkowskian situations (e.g. Amati, Petronzio & Veneziano (1978), Ellis *et*

al. (1979), Libby & Sterman (1978), Mueller (1978, 1981), Stirling (1978), and Buras (1981)). These are beyond the scope of the present book.

5.9 Oversubtractions

We showed how to renormalize a Feynman graph by making subtractions for the divergent subgraphs and for the overall divergence of the graph. It is possible, however, to make subtractions on graphs that are not divergent. Subtractions can also be made with a higher degree polynomial in the external momenta than called for by the degree of divergence. Either of these cases is called oversubtraction. Now, the general form of the renormalization, either by the recursive method or by the forest formula, did not specify the exact form of the subtraction operator T. So oversubtractions can be made without changing the general formalism.

There are two important uses for oversubtractions. The first is when we wish to use 'physical values' of masses or couplings as the renormalized parameters. We will discuss this in a moment. The second use is to construct operator product expansions. There, subtractions are made not only to cancel UV divergences but also to extract asymptotic behavior as some external momenta get large. We will discuss this later in Chapter 10.

5.9.1 *Mass-shell renormalization and oversubtraction*

We have considered renormalization as the procedure of removing divergences. Another point of view comes from the observation that one cannot observe directly the mass and coupling parameters that appear as coefficients in the Lagrangian. For example, consider a theory where each field has a corresponding single-particle state. Then the masses that are measured are those of the single particles, and it is often sensible to parametrize the theory in terms of these masses. Similar remarks can be applied to couplings. (Thus in QED one normally parametrizes the theory by the electron's mass and charge, defined by the long-range part of its electric field.) It can also be convenient to rescale the fields so that each propagator has a pole of unit residue.

In a simple renormalizable theory like ϕ^3 in six dimensions the renormalizations to accomplish such a mass-shell parametrization are precisely those necessary to cancel the UV divergences. Thus we may define the subtraction operator applied to a self-energy graph $\Sigma(p^2)$ to be

$$T_{(\mathrm{ph})} \circ \Sigma(p^2) = \Sigma(m_{\mathrm{ph}}^2) + (p^2 - m_{\mathrm{ph}}^2)\Sigma'(m_{\mathrm{ph}}^2), \tag{5.9.1}$$

so that the inverse propagator satisfies

$$-i[p^2 - m_{ph}^2 - \Sigma_{ph}(p^2)] = -i(p^2 - m_{ph}^2) + O(p^2 - m_{ph}^2)^2, \quad (5.9.2)$$

as $p^2 \to m_{ph}^2$. We use the subscript 'ph' to indicate renormalization according to the mass-shell scheme.

Of course, mass-shell renormalization is only one out of many renormalization prescriptions that we may use to cancel UV divergences. But we may also choose to renormalize in the absence of divergences. Consider, as an example, ϕ^3 theory again, but now in four dimensions. We may continue to use (5.9.1) and (5.9.2) for the renormalization of the propagator so we have a 'physical' parametrization. But all except the one-loop self-energy graph have no divergence, so all the wave-function counterterms are finite and all but one of the mass counterterms are finite. The combinatorics of the renormalization procedure as described earlier all work unchanged.

5.9.2 *Remarks*

One important technical problem is to check that the oversubtracted and the normally subtracted theories differ only by a reparametrization. This can be done by the methods which we will describe in Sections 7.1 and 7.2.

In the previous subsection 5.9.1, we took the point of view that renormalization is the process of reparametrizing the theory in terms of 'physical' quantities. It should be noted that this is not always a useful point of view. In the first place, other renormalization prescriptions are more convenient for handling certain types of calculation. In the second place, there may be infra-red divergences that make the mass-shell structure of a theory not what one would naively expect: thus in QED the electron's propagator does not have a simple pole. And, finally, in some theories there are many more particles and couplings than independent parameters. This is very common in gauge theories.

5.9.3 *Oversubtraction on IPR graphs*

The aim of oversubtraction, generally, is to impose some condition on Green's functions. So far, we have assumed the condition to be imposed on the 1PI graphs, since those are the ones needing counterterms for divergences. However, consider $\phi^3 + \phi^4$ theory:

$$\mathscr{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - f\phi^3/6 - g\phi^4/24 + \text{counterterms}. \quad (5.9.3)$$

Let us choose to renormalize at zero external momentum. Thus the self-

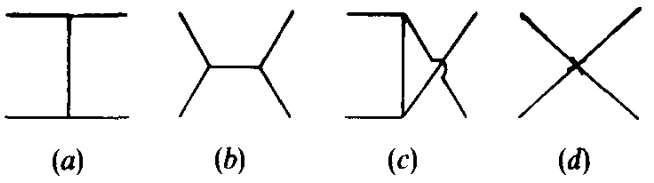


Fig. 5.9.1. Subtraction of one-particle-reducible subgraphs.

energy Σ and the three-point 1PI function $\Gamma_{(3)}$ satisfy

$$\Sigma(p^2 = 0) = \frac{d\Sigma}{dp^2}(p^2 = 0) = 0,$$

$$\Gamma_{(3)}(p_1^2 = p_2^2 = p_3^2 = 0) = \text{lowest order} = -if. \tag{5.9.4}$$

Following from our earlier work we might renormalize the four-point function $\Gamma_{(4)}$ by requiring the sum of the 1PI graphs to be equal to their lowest order value at zero external momentum. However it is also sensible to impose instead the condition on the amputated four-point function $\Gamma_{(4)}^{(a)}$. These graphs are 1PI only in the four external lines. (We should amputate the graphs since the counterterm vertex will have attached to it external propagators.) Thus in addition to the three tree graphs of Fig. 5.9.1 (*a*)–(*c*), we require the counterterm, Fig. 5.9.1(*d*):

Our general method of renormalization tells us that whenever we have a basic graph containing one of the graphs (*a*), (*b*), or (*c*) in Fig. 5.9.1 as a subgraph, there will be counterterm graphs in which this subgraph is replaced by the counterterm vertex (*d*). These counterterm graphs may be divergent even when the basic graph is finite. An example is shown in Fig. 5.9.2. In Fig. 5.9.2(*a*) if we impose the renormalization condition on the 1PI functions only the graph (*b*) occurs as counterterm; (*a*) plus (*b*) is finite. If we impose the condition on amputated graphs we immediately meet graph (*c*) where the line A is replaced by its 1/3 share of the counterterm Fig. 5.9.1(*d*).

But graph (*c*) has a divergence, so we must renormalize it by a three-point counterterm to the subgraph consisting of the line A and the loop B. This is

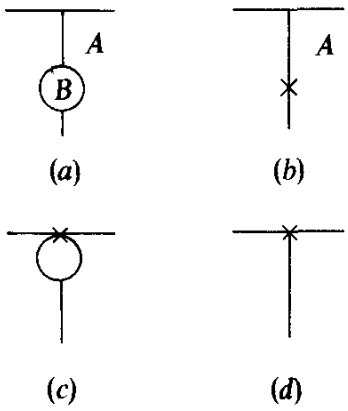


Fig. 5.9.2. The subtractions of Fig. 5.9.1, inside a bigger graph.

shown in graph (*d*). Note that the graph consisting of line *A* and loop *B* has a subdivergence, but no overall divergence. Even so, the overall counterterm in (*d*) is divergent.

It is not difficult to see that the extra counterterms needed to impose the renormalization condition on the 1PR amputated graphs do not change our results on renormalization. The instructions for renormalization in Sections 5.3 and 5.5 can be used provided only that we replace the term '1PI subgraph' by 'amputated subgraph'.

The use of subtractions on 1PR graphs is too baroque for normal use. However it is a device that is useful for discussing the large mass expansion and the operator-product expansion (Chapters 8 and 10).

5.10 Renormalization without regulators: the BPHZ scheme

In setting up the renormalization procedure in Sections 5.3 and 5.5 we were careful not to use a specific definition of the subtraction operation. This was to allow for the choice of one out of the infinitely many possible renormalization prescriptions. An obvious one is the mass-shell subtraction procedure indicated in the last section. Another is the minimal subtraction procedure to be defined precisely in Section 5.11; we have already made much use of it. In this section we will explain the method of Zimmermann (1969), in which the subtractions are applied directly to the Feynman integrand, so that no regulator need be used.

The starting point is the method due to Bogoliubov & Parasiuk (1957) and Hepp (1966), called the BPH scheme. They observed that the overall counterterm for a graph Γ is a polynomial of degree $\delta(\Gamma)$, its degree of divergence. So they defined the subtraction operator $T(\Gamma)$ to be the terms up to order $\delta(\Gamma)$ in the Taylor expansion of Γ about zero external momentum. For example, consider the one-loop self-energy graph Fig. 3.1.1 in ϕ^3 theory in six dimensions. After dimensional regularization its unrenormalized value is

$$\Sigma_a(p^2,d) = \frac{-g^2}{2(4\pi)^{d/2}}\Gamma(2-d/2)\int_0^1 \mathrm{d}x\,[m^2 - p^2x(1-x)]^{d/2-2}. \quad (5.10.1)$$

The terms up to order p^2 in its Taylor expansion about $p=0$ are

$$T\circ\Sigma_a = \frac{-g^2}{2(4\pi)^{d/2}}\Gamma(2-d/2)\times$$
$$\times\int_0^1 \mathrm{d}x m^{d-4}[1-(d/2-2)p^2x(1-x)/m^2]. \quad (5.10.2)$$

The renormalized value of the graph is $\Gamma_a - T'\Gamma_a$. So at $d=6$ this is

$$-\frac{g^2}{128\pi^3}\int_0^1 \mathrm{d}x\{[m^2-p^2x(1-x)]\ln[1-p^2x(1-x)/m^2]+p^2x(1-x)\}. \tag{5.10.3}$$

Zimmermann's (1969) achievement was to realize that this construction can be applied directly to the integrand. Subdivergences are subtracted with the aid of his forest formula. Then the result is an integral which, according to power-counting, has no UV divergences. The integral therefore has in fact no divergences (Hahn & Zimmermann (1968), Zimmermann (1968)). This method is called BPHZ renormalization.

In Section 3.4, we applied this method to the above graph, with the result (3.4.7). It can be explicitly calculated by putting all the terms over a common denominator and then using standard parametric methods. The result agrees with (5.10.3).

An example involving a subdivergence is given by Fig. 5.10.1 for ϕ^4 in four dimensions. Let the renormalized integral be

$$\Gamma_{\mathrm{BPHZ}} = \frac{ig^3}{2(2\pi)^8}\int \mathrm{d}^4k\mathrm{d}^4l I(p_1,p_2,p_3,p_4,k,l). \tag{5.10.4}$$

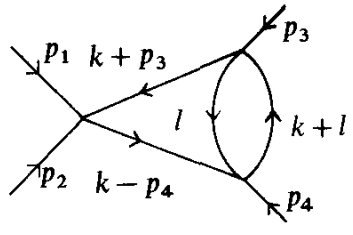


Fig. 5.10.1. A two-loop vertex graph in ϕ^4 theory.

Then we will construct the integrand I.

The unrenormalized integrand is

$$U = \frac{1}{(l^2-m^2)}\frac{1}{[(k+l)^2-m^2]}\frac{1}{[(k+p_3)^2-m^2]}\frac{1}{[(k-p_4)^2-m^2]} \tag{5.10.5}$$

Subtraction of the sole subdivergence gives

$$\begin{aligned}\tilde{R}(U) &= U - \frac{1}{(l^2-m^2)^2}\frac{1}{[(k+p_3)^2-m^2]}\frac{1}{[(k-p_4)^2-m^2]} \\ &= \frac{-2k\cdot l - k^2}{(l^2-m^2)^2[(k+l)^2-m^2][(k+p_3)^2-m^2][(k-p_4)^2-m^2]}.\end{aligned} \tag{5.10.6}$$

Then the overall divergence is subtracted to give

$$I = R(U) = \bar{R}(U) - [R(U)]|_{p_1 \cdot p_2 = p_1 \cdot p_4 = 0}$$
$$= \frac{(2k\cdot l + k^2)\{(k^2 - m^2)[p_4^2 + p_3^2 + 2k\cdot(p_3 - p_4)] + (p_3^2 + 2k\cdot p_3)(p_4^2 - 2k\cdot p_4)\}}{(l^2 - m^2)^2(k^2 - m^2)^2[(k+l)^2 - m^2][(k+p_3)^2 - m^2][(k-p_4)^2 - m^2]}. \tag{5.10.7}$$

The BPHZ scheme has a number of advantages:

(1) It is applied to the integrand and generates a convergent integral without requiring any regularization.
(2) Thus it exhibits the fact that the properties of a renormalized field theory do not depend on which UV regulator is used.
(3) Mathematically it is rather elegant. In particular there is no need to discuss directly the divergences of Feynman graphs; it is only required to have a theorem that tells us that a graph that is convergent according to the naive criteria is actually convergent.
(4) It allows a very simple proof of the operator-product expansion.

There are a number of disadvantages:

(1) It is not the best scheme for theories (especially gauge theories) with complicated symmetries, where relations between counterterms have to be preserved; the scheme does not allow direct computation of the value of a divergence.
(2) The subtractions are made at zero momentum and therefore are infrared divergent in a massless theory.
(3) When the scheme is generalized to handle massless theories, it becomes much more complicated (Lowenstein, Weinstein & Zimmermann, 1974a, b).

5.11 Minimal subtraction

5.11.1 *Definition*

It can be proved (Speer (1974) and Breitenlohner & Maison (1977a, b, c)) that, when dimensional regularization is used, the UV divergences of Feynman graphs appear as poles at isolated values of the space-time dimension d. Minimal subtraction ('t Hooft (1973)) – the MS scheme – consists of defining the counterterms to be poles at the physical value of d, $d = 4$. We have already used this scheme, in Chapter 3. Our purpose in this section is to make precise the definition of minimal subtraction.

The main complication is that bare couplings have a dimension that depends on d, so that we must introduce the unit of mass μ, as follows:

(1) Consider in turn the coefficient $g_i + \delta g_i$ of each term in $\mathscr{L}$. Let the dimension of g_i be $a_i + b_i(4-d)$. Then we replace $g_i + \delta g_i$ by $\mu^{b_i(4-d)}(g_i + \delta g_i)$. Thus the renormalized coupling g_i and the counterterm δg_i both have dimension a_i, independently of d.

(2) Let Γ be a 1PI graph to which it is desired to apply a subtraction operator T. Let the dimension of Γ be $A + B(d-4)$, and suppose the couplings all contain powers of μ as just explained. Then we define

$$T(\Gamma) = \mu^{B(d-4)}\{\text{pole part of } (\mu^{B(4-d)}\Gamma) \text{ at } d = 4\}.$$

The pole part is obtained by making a Laurent expansion about $d = 4$. We have arranged to take the pole part of a function whose dimension does not depend on d.

(3) Suppose we are talking about a theory in a different number of physical dimensions than four. For example, we might be in ϕ^3 theory in six dimensions. Then the '4' in the above formulae is replaced by the correct physical value.

For a simple graph with no subdivergences, like the one-loop self-energy in (5.10.1), this prescription amounts to subtracting the pole:

$$\begin{aligned}\Sigma_a^{(\mathrm{MS})}(d=6) = \lim_{d\to 6}\Bigg\{ & \frac{-g^2\mu^{6-d}}{2(4\pi)^{d/2}}\Gamma(2-d/2)\int_0^1 \mathrm{d}x[m^2 - p^2x(1-x)]^{d/2-2} \\ & - \left[\text{pole} = \frac{-g^2}{128\pi^3}\frac{1}{d/2-3}(m^2 - \tfrac{1}{6}p^2)\right]\Bigg\} \\ = \frac{-g^2}{128\pi^3}\Bigg\{ & [\gamma_{\mathrm{E}} - 1 - \ln(4\pi)](m^2 - \tfrac{1}{6}p^2) \\ & + \int_0^1 \mathrm{d}x[m^2 - p^2x(1-x)]\ln\left[\frac{m^2 - p^2x(1-x)}{\mu^2}\right]\Bigg\}. \end{aligned} \tag{5.11.1}$$

For graphs with subdivergences, the subdivergences must of course be subtracted before removing the overall pole.

The advantages of the scheme are:

(1) It automatically preserves complicated symmetries. The exceptions are chiral symmetries and the like, which in general cannot be preserved by quantization – see Chapter 13.

(2) It has no problems with massless theories. In fact, dimensional continuation regulates both IR and UV divergences, thus removing the need for a separate IR cut-off.

(3) Calculations are very convenient.

(4) Computation of the divergent part of a Feynman graph – needed for

renormalization-group calculations is almost trivial at the one-loop level.

Some disadvantages of minimal subtraction are:

(1) It is unphysical.
(2) The proof of the operator-product expansion is made harder than in the BPHZ scheme.

5.11.2 $\overline{MS}$ *renormalization*

The MS scheme has found much use especially in work on QCD, where it has become standard. Another disadvantage that then appears is that minimal subtraction tends to produce large coefficients in the perturbation expansion. These are primarily due to the $\ln(4\pi) - \gamma_E \sim 1.95$ term such as appears in (5.11.1). It has become conventional to work with a modified scheme, called the $\overline{\text{MS}}$ scheme (Bardeen, Buras, Duke & Muta (1978)). Here the μ of the MS scheme is written as

$$\mu = \bar{\mu}\left(\frac{e^{\gamma_E}}{4\pi}\right)^{1/2} \simeq 0{\cdot}38\bar{\mu}. \tag{5.11.2}$$

Then we have, instead of (5.11.1), the cleaner form

$$\Sigma_a^{(\overline{\text{MS}})} = \frac{-g^2}{128\pi^3}\left\{\tfrac{1}{6}p^2 - m^2 + \int_0^1 dx[m^2 - p^2x(1-x)]\ln\left[\frac{m^2 - p^2x(1-x)}{\bar{\mu}^2}\right]\right\}. \tag{5.11.3}$$

5.11.3 *Minimal subtraction with other regulators*

Minimal subtraction could also be applied with other UV cut-offs. For example, if a lattice of spacing a is used, then the singular $a \to 0$ behavior of graph of degree of divergence δ is

$$a^{-\delta}[\text{polynomial in } \ln(a)].$$

One can therefore define $T(\Gamma)$ as the singular part of Γ, with the general form

$$T(\Gamma) = \sum_{\beta=1}^{\beta_{\max}} [\ln(a\mu)]^\beta A_{0,\beta} + \sum_{\alpha=1}^{\delta} \sum_{\beta=0}^{\beta_{\max}} \frac{1}{a^\alpha}[\ln(a\mu)]^\beta A_{\alpha,\beta}. \tag{5.11.4}$$

After subtraction of subdivergences, the coefficients $A_{\alpha,\beta}$ are polynomials in masses and momenta. Note again the appearance of a unit of mass. This scheme has found little use.

6

Composite operators

In Chapter 2 we met a number of equations involving products of field operators at the same point. Examples are given by the equations of motion (2.1.10) and the Ward identities (2.7.6). These products we will call composite operators. When computed directly they have ultra-violet divergences: the product $\phi(x)\phi(y)$ makes unambiguous sense if x is not equal to y, but if x equals y then we have $\phi(x)^2$, which diverges. Since the equations of motion and the Ward identities express fundamental properties of the theory, it is useful to construct finite, renormalized composite operators with which to express these same properties.

It could be argued that there is no need to have renormalized equations of motion. One could say that one only actually needs the equations of motion in the regulated theory, where they are finite. A situation of practical importance where we actually do need renormalized composite operators is the operator-product expansion, to be discussed in Chapter 10. This is used in a phenomenological situation such as deep-inelastic scattering (Chapter 14) where we wish to compute the behavior of a Green's function when some of its external momenta get large. Equivalently, we need to know how a product of operators, like $\phi(x)\phi(y)$, behaves as $x \to y$.

This information is contained in the operator-product expansion of Wilson (1969) which has the form

$$\phi(x)\phi(y) \sim C_1(x-y)1 + C_{\phi^2}(x-y)[\phi(y)^2] + \cdots . \tag{6.0.1}$$

Here the symbol $[A(x)]$ denotes the renormalized operator corresponding to an unrenormalized composite operator $A(x)$. The coefficients $C(x-y)$ are c-numbers, and each has a subscript which labels the operator that it multiplies.

Therefore in this chapter we show how to renormalize Green's functions of composite operators, e.g.,

$$\begin{aligned} &\langle 0|T\phi(x)\phi(y)\phi^2(z)|0\rangle, \\ &\langle 0|T\phi(w)\phi(x)\phi(y)^2\phi(z)^2|0\rangle, \\ &\langle 0|T\phi(x)^2\phi(y)^2|0\rangle. \end{aligned} \tag{6.0.2}$$

We will first motivate the use of composite operators by seeing how the operator-product expansion arises in a low-order graph. Then we will examine the divergences that appear in low-order graphs for composite operators. We will see that we must expect multiplicative renormalization:

$$[\phi^2] = Z_{\phi^2}\phi^2, \tag{6.0.3}$$

where $[\phi^2]$ is finite as the UV cut-off is removed, while Z_{ϕ^2} is a divergent renormalization factor. The unrenormalized operator $\phi(x)^2$ is divergent when the cut-off is removed.

These examples will provide motivation to define renormalized composite operators by application to Feynman graphs of the same R-operation that we defined in Chapter 5. After discussion of a number of technical issues, we will derive some basic properties of the renormalized operators, including the equations of motion and the Ward identities.

6.1 Operator-product expansion

We will postpone a complete treatment of the operator-product expansion to Chapter 10. Here we merely wish to motivate our definition of composite operators with an example of their use.

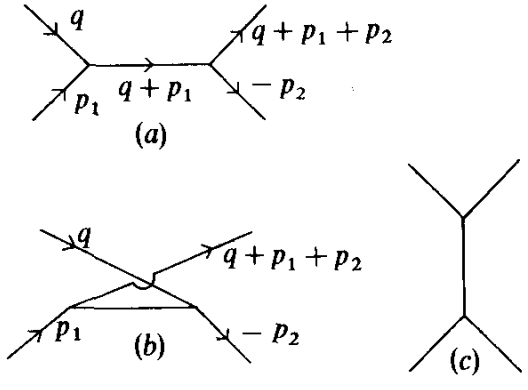


Fig. 6.1.1. Take $q \to \infty$ in these graphs to obtain the lowest-order example of the operator-product expansion.

Consider the graphs of Fig. 6.1.1 for the four-point function in ϕ^3 theory. We let q^μ go to infinity with p_1 and p_2 fixed, and with the ratios of the components of q fixed. Then $|q^2| \to \infty$. We expand the graphs in powers of q^2 to find:

$$\text{Fig. 6.1.1}(a) + (b) \sim \left[\frac{\mathrm{i}}{(p_1^2 - m^2)}\frac{\mathrm{i}}{(p_2^2 - m^2)}\right]\frac{2\mathrm{i}g^2}{(q^2)^3}, \tag{6.1.1a}$$

$$\text{Fig. 6.1.1}(c) \sim \left[\frac{-g^2}{(p_1^2-m^2)(p_2^2-m^2)((p_1+p_2)^2-m^2)}\right]\frac{ig}{(q^2)^2}\times$$
$$\times\left[1-\frac{2q\cdot(p_1+p_2)}{q^2}+\frac{(2m^2-(p_1+p_2)^2)}{q^2}+\frac{4q\cdot(p_1+p_2)^2}{(q^2)^2}\right]. \quad (6.1.1b)$$

In each term the dependence of p_1 and p_2 has factorized. We now show that this is a case of an operator-product expansion like (6.0.1) (after Fourier transformation into momentum space).

In (6.1.1a) the factor in square brackets is in fact the value of the lowest-order graph for the following Green's function:

$$\langle 0|T\tilde{\phi}(-p_1)\tilde{\phi}(-p_2)\phi^2(0)/2|0\rangle$$
$$=\int d^4x\int d^4y\exp(-ip_1\cdot x-ip_2\cdot y)\langle 0|T\phi(x)\phi(y)\phi^2(0)/2|0\rangle. \quad (6.1.2)$$

We have not Fourier transformed the $\phi^2(0)$ operator, but have set it at the origin. If we had made the Fourier transform, then we would merely pick up a momentum-conservation δ-function, which we do not have in (6.1.1a).

To understand the appearance of the operator $\phi^2(0)/2$ in (6.1.2), we may find a functional-integral formula for the Green's functions that appear in this equation. Since $\phi^2(0)$ means the product of two fields at the same space-time point, such a formula follows from our work in Section 2.2. It is

$$\langle 0|T\phi(x)\phi(y)\phi^2(0)/2|0\rangle=\mathcal{N}\int[dA]A(x)A(y)\tfrac{1}{2}A^2(0)e^{iS}. \quad (6.1.3)$$

The Feynman rules for this Green's function can then be derived. They are the usual ones for the Green's function $\langle 0|T\phi(x)\phi(y)|0\rangle$ with the addition that each graph contains exactly one special vertex for the $\phi^2(0)/2$ operator. The lowest-order graph is shown in Fig. 6.1.2, where the special $\phi^2/2$ vertex is indicated by a cross. The value of the vertex is unity, for the explicit factor 1/2 in $\phi^2/2$ gets cancelled. This happens in exactly the same way as the 1/4! in the ϕ^4 interaction or the 1/3! in the ϕ^3 interaction gets cancelled to leave a value $-ig$ for an interaction vertex.

The operator $\phi^2(0)/2$ is our first example of a composite operator (or composite field). By this term we mean, in general, a product of elementary



Fig. 6.1.2. Lowest-order graph for two-point function of ϕ^2.

fields (or their derivatives) at the same point. It is the properties of such operators that we will investigate in this chapter.

We can write

$$\text{Fig. 6.1.1}(a)+(b) \sim \frac{2ig^2}{(q^2)^3}\langle 0|T\tilde{\phi}(p_1)\tilde{\phi}(p_2)\phi(0)^2/2|0\rangle, \qquad (6.1.4a)$$

This is illustrated in diagrams by Fig. 6.1.3.

In similar fashion, we derive an operator formula for (6.1.1*b*):

$$\text{Fig. 6.1.1}(c) \sim \frac{ig^2}{(q^2)^2}\left[1-\frac{2iq^\mu}{q^2}\frac{\partial}{\partial x^\mu}+\frac{(2m^2+\Box)}{q^2}-\frac{4q^\mu q^\nu \partial_\mu \partial_\nu}{q^4}\right]\times$$
$$\times\langle 0|T\tilde{\phi}(p_1)\tilde{\phi}(p_2)\phi(x)|0\rangle|_{x=0}. \qquad (6.1.4b)$$

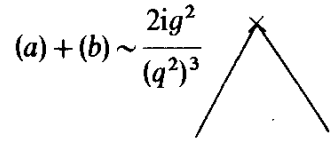


Fig. 6.1.3. Generation of terms in the operator-product expansion from the graphs of Fig. 6.1.1(*a*) and (*b*).

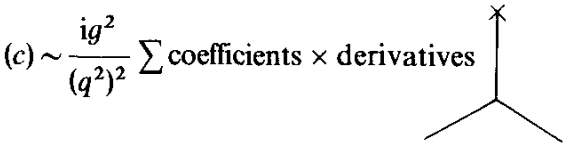


Fig. 6.1.4. Generation of terms in the operator-product expansion from the graph of Fig. 6.1.1(*c*)

Here the form of the first square-bracket factor means that we need an elementary field $\phi(x)$, rather than a composite field. The p_1's and p_2's in the numerators in the second square-bracket factor have turned themselves into derivatives with respect to x; we set $x=0$ at the end. Equation (6.1.4*b*) is illustrated in Fig. 6.1.4.

The form of (6.1.4) suggests the following formula:

$$\langle 0|T\tilde{\phi}(q)\phi(0)\tilde{\phi}(p_1)\tilde{\phi}(p_2)|0\rangle$$
$$\sim \sum_i C_i(q)\langle 0|T\mathcal{O}_i(0)\tilde{\phi}(p_1)\tilde{\phi}(p_2)|0\rangle. \qquad (6.1.5)$$

The sum is over a set of local operators $\mathcal{O}_i$. Each of these is either the elementary field ϕ, one of its derivatives, or a composite operator such as ϕ^2. The coefficients $C_i(q)$ are called the Wilson coefficients. In Chapter 10 we will generalize this result to all orders. We will have an expansion for any Green's function with large momentum on some of its external lines.

Now higher-order corrections to the Green's functions of the composite operators such as the $\phi^2(0)$ that appears in (6.1.4*a*) have ultra-violet divergences beyond those appearing in Green's functions of elementary fields. We will see this in the next section, Section 6.2. To obtain an operator product expansion, like (6.1.5), with finite coefficients, we will need to renormalize the composite operators. This particular problem will occupy most of this chapter.

6.2 Renormalization of composite operators: examples

6.2.1 *Renormalization of* ϕ^2

The Feynman rules for Green's functions of unrenormalized composite operators can be derived from the functional integral, in the presence of an ultra-violet cut-off. In coordinate space they are the usual rules, modified only by having several external fields at the same point. For example, we consider

$$\langle 0|T\phi(x)\phi(y)\phi^2(z)/2|0\rangle \tag{6.2.1}$$

in ϕ^3 theory in six-dimensional space-time. The connected graphs up to order g^2 are shown in Fig. 6.2.1. As before, the vertex for $\phi^2/2$ is denoted by a cross.

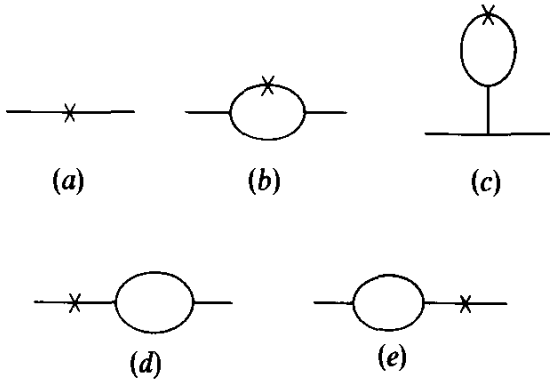


Fig. 6.2.1. Renormalization of the operator ϕ^2.

To work in momentum space, we Fourier transform, as usual, and define

$$\begin{aligned} G &= \langle 0|T\tilde{\phi}(p)\tilde{\phi}(q)\phi(z)^2/2|0\rangle \\ &= \int \mathrm{d}^d x \mathrm{d}^d y \exp[\mathrm{i}(p_1\cdot x + p_2\cdot y]\langle 0|T\phi(x)\phi(y)\phi(z)^2/2|0\rangle \end{aligned} \tag{6.2.2}$$

The lowest-order graph, Fig. 6.2.1(*a*), is equal to

$$G_a = \frac{\mathrm{i}}{(p_1^2 - m^2 + \mathrm{i}\varepsilon)} \frac{\mathrm{i}}{(p_2^2 - m^2 + \mathrm{i}\varepsilon)}. \tag{6.2.3}$$

Observe that the factor 1/2 in the operator $\phi^2/2$ is cancelled, just like the 1/3! that comes with the interaction vertices.

Let us now turn to the one-loop graphs of Fig. 6.2.1. They are all divergent: Fig. 6.2.1(*b*) is logarithmically divergent, while the remaining graphs, Fig. 6.2.1(*c*) to (*e*), are quadratically divergent (all at $d = 6$). The divergences in the last two graphs, Figs. 6.2.1(*d*) and (*e*), involve self-energy corrections only, so these divergences are cancelled by the usual wave-



(a) (b)

Fig. 6.2.2. Counterterm graphs for Fig. 6.2.1(*d*) and (*e*).

function and mass counterterms, Fig. 6.2.2. In these and other graphs in this chapter, we indicate counterterms by a heavy dot and an insertion of a composite operator by a cross.

The remaining two graphs have no counterterm from the interaction, and they are both divergent. For Fig. 6.2.1(*b*) we get

$$G_b = \frac{\mathrm{i}}{(p_1^2 - m^2 + \mathrm{i}\varepsilon)} \frac{\mathrm{i}}{(p_2^2 - m^2 + \mathrm{i}\varepsilon)} \times$$

$$\times \left\{ \frac{\mathrm{i}g^2\mu^{6-d}}{(2\pi)^d} \int \mathrm{d}^d k \frac{1}{(k^2 - m^2 + \mathrm{i}\varepsilon)[(k-p_1)^2 - m^2 + \mathrm{i}\varepsilon][(k+p_2)^2 - m^2 + \mathrm{i}\varepsilon]} \right\}$$

$$= \frac{\mathrm{i}}{(p_1^2 - m^2)} \frac{\mathrm{i}}{(p_2^2 - m^2)} \frac{g^2}{64\pi^3} \Gamma(3 - d/2) \times \int_0^1 \mathrm{d}x \times$$

$$\times \int_0^{1-x} \mathrm{d}y \left[\frac{m^2 - p_1^2 y(1-x-y) - p_2^2 x(1-x-y) - (p_1+p_2)^2 xy}{4\pi\mu^2} \right]^{d/2-3} \tag{6.2.4}$$

and for Fig. 6.2.1 (*c*) we get

$$G_c = \frac{-g\mu^{3-d/2}}{(p_1^2 - m^2)(p_2^2 - m^2)[(p_1+p_2)^2 - m^2]} \times$$

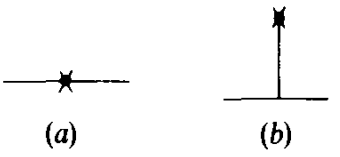
$$\times \frac{\mathrm{i}g\mu^{3-d/2}}{2(2\pi)^d} \int \mathrm{d}^d k \frac{1}{(k^2 - m^2)[(p_1+p_2+k)^2 - m^2]}$$

$$= \frac{-g\mu^{3-d/2}}{(p_1^2 - m^2)(p_2^2 - m^2)[(p_1+p_2)^2 - m^2]} \times$$

$$\times \frac{-g\mu^{d/2-3}}{128\pi^3} \Gamma(2 - d/2) \int_0^1 \mathrm{d}x \frac{[m^2 - (p_1+p_2)^2 x(1-x)]^{d/2-2}}{(4\pi\mu^2)^{d/2-3}}. \tag{6.2.5}$$

Note that there is a symmetry factor 1/2 in this last equation. The fact that the sum of (6.2.4) and (6.2.5) diverges means that the operator $\phi^2(0)$ is not finite.

For use in the operator-product expansion we do not need precisely the operator ϕ^2. Rather, we need some local operator similar to ϕ^2 that is finite. This indicates that we should define a renormalized operator by subtraction of the divergences. Let us agree to use minimal subtraction. Then the counterterm graphs are obtained by replacing each divergent loop by



(a) (b)

Fig. 6.2.3. Counterterm graphs for Fig. 6.2.1(*b*) and (*c*).

minus its pole part, as illustrated in Fig. 6.2.3. Thus the counterterm graph for Fig. 6.2.1(*b*) is

$$\frac{\mathrm{i}^2}{(p_1^2-m^2)(p_2^2-m^2)}\frac{g^2}{64\pi^3(d-6)}, \tag{6.2.6}$$

and the counterterm graph for Fig. 6.2.1(*c*) is

$$\frac{-g\mu^{3-d/2}}{(p_1^2-m^2)(p_2^2-m^2)[(p_1+p_2)^2-m^2]}\left\{\frac{g\mu^{d/2-3}}{64\pi^3(d-6)}[m^2-\tfrac{1}{6}(p_1+p_2)^2]\right\}. \tag{6.2.7}$$

The positioning of the factors of μ is such that the counterterm in curly brackets has exactly the same dimension as the loop to which it is a counterterm.

We thus find the renormalized values at $d=6$:

$$R(G_b)=\frac{\mathrm{i}^2}{(p_1^2-m^2)(p_2^2-m^2)}\frac{g^2}{64\pi^3}\left\{-\frac{\gamma}{2}-\int_0^1 \mathrm{d}x\times\right.$$
$$\left.\times\int_0^{1-x}\mathrm{d}y\ln\left[\frac{m^2-(p_1^2y+p_2^2x)(1-x-y)-(p_1+p_2)^2xy}{4\pi\mu^2}\right]\right\}, \tag{6.2.8}$$

$$R(G_c)=\frac{-g}{(p_1^2-m^2)(p_2^2-m^2)[(p_1+p_2)^2-m^2]}\left(\frac{-g}{128\pi^3}\right)\times$$
$$\times\left\{(\gamma-1)[m^2-\tfrac{1}{6}(p_1+p_2)^2]+\int_0^1\mathrm{d}x[m^2-(p_1+p_2)^2x(1-x)]\times\right.$$
$$\left.\times\ln\left[\frac{m^2-(p_1+p_2)^2x(1-x)}{4\pi\mu^2}\right]\right\}. \tag{6.2.9}$$

To interpret these renormalizations we observe that the counterterms are vertices for $\phi^2(0)$ in (6.2.6) and for $(m^2+\Box/6)\phi$ in (6.2.7). Thus

$$G_a+R(G_b)+R(G_c)+R(G_d)+R(G_e)$$
$$=\left[1+\frac{g^2}{64\pi^3(d-6)}\right]\langle 0|T\tilde{\phi}(p_1)\tilde{\phi}(p_2)\tfrac{1}{2}\phi^2(0)|0\rangle$$
$$+\frac{g\mu^{d/2-3}}{64\pi^3(d-6)}\langle 0|T\tilde{\phi}(p_1)\tilde{\phi}(p_2)(m^2+\tfrac{1}{6}\Box)\phi(0)|0\rangle+O(g^4). \tag{6.2.10}$$

So what we are computing is a Green's function of the operator

$$\tfrac{1}{2}[\phi^2] \equiv \left[1 + \frac{g^2}{64\pi^3(d-6)}\right]\tfrac{1}{2}\phi^2 + \frac{g\mu^{d/2-3}}{64\pi^3(d-6)}(m^2 + \tfrac{1}{6}\Box)\phi + \text{higher order}. \tag{6.2.11}$$

We use the square brackets on the left-hand side to denote a renormalized operator. In subsequent sections we will see that this result generalizes to all orders: a renormalized operator $[\phi^2]$ can be defined to all orders by a formula of the form:

$$\tfrac{1}{2}[\phi^2] = Z_{a\frac{1}{2}}\tfrac{1}{2}\phi^2 + \mu^{d/2-3}Z_b m^2\phi + \mu^{d/2-3}Z_c\Box\phi, \tag{6.2.12}$$

where the Z's depend only on g and d.

Such renormalized operators were defined by Zimmermann (1973a). He called them normal products and used the notation $N[\phi^2]$. His definition differed from ours only in that he used BPHZ renormalization instead of minimal subtraction.

Observe that in order to obtain a degree of divergence of at least zero, the operators that appear as counterterms in (6.2.11) or (6.2.12) have dimension less than or equal to that of the original operator ϕ^2. This is a general phenomenon. Moreover, the only operators of such dimension are those actually appearing in (6.2.12). We may write the renormalization as a matrix equation in the following form:

$$\begin{pmatrix} \frac{1}{2}[\phi^2] \\ \phi \\ \Box\phi \end{pmatrix} = \begin{pmatrix} Z_a & \mu^{d/2-3}Z_b m^2 & \mu^{d/2-3}Z_c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\phi^2 \\ \phi \\ \Box\phi \end{pmatrix}. \tag{6.2.13}$$

Here we have used the fact that ϕ and $\Box\phi$ are finite. The operators ϕ and $\Box\phi$ are said to mix with ϕ^2 under renormalization, because the off-diagonal elements Z_b and Z_c are non-zero. Moreover, no further operators are needed in the renormalizations, so ϕ^2, ϕ and $\Box\phi$ are said to form a closed set under renormalization.

6.2.2 *Renormalization of* $\phi^2(x)\phi^2(y)$

Sometimes we need Green's functions involving two or more composite operators. A simple example is

$$\langle 0|T\tfrac{1}{2}[\tilde{\phi}^2](p)\tfrac{1}{2}[\phi^2](0)|0\rangle = \int \mathrm{d}^d x e^{-ip\cdot x}\langle 0|T\tfrac{1}{2}[\phi^2](x)\tfrac{1}{2}[\phi^2](0)|0\rangle, \tag{6.2.14}$$

for which the lowest-order graph is Fig. 6.2.4. The renormalizations of

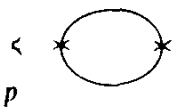


Fig. 6.2.4. Lowest-order graph for $\langle 0|T[\phi^2]/2[\phi^2]/2|0\rangle$.

(6.2.11) do not appear until the next order, so the graph has the value

$$-\frac{1}{2(2\pi)^d}\int \mathrm{d}^d k \frac{1}{(k^2-m^2)[(p+k)^2-m^2]}. \tag{6.2.15}$$

This is ultra-violet divergent, even in free-field theory. Even though the operators $\phi^2(x)$ and $\phi^2(0)$ on the right of (6.2.14) are well-defined, we have to integrate through the point $x=0$. At $x=0$ there is a singularity, which is not integrable if $d \geq 4$. We may nevertheless define a finite Green's function by adding a local counterterm:

$$\langle 0|T\tfrac{1}{2}[\phi^2](x)\tfrac{1}{2}[\phi^2](0)|0\rangle_R = \langle 0|T\tfrac{1}{2}[\phi^2](x)\tfrac{1}{2}[\phi^2](0)|0\rangle - C(x)\langle 0|1|0\rangle. \tag{6.2.16}$$

with

$$C(x) = (m^2 + \Box/6)\delta^{(d)}(x)\frac{(-\mathrm{i}\mu^{d-6})}{64\pi^3(d-6)} + O(g^2). \tag{6.2.17}$$

Once more we have used minimal subtraction at $d=6$.

6.3 Definitions

We define renormalized Green's functions of composite operators by applying the R-operation to the Feynman graphs, just as we did for Green's functions of elementary fields in Sections 5.3 and 5.5. We will need to show that the counterterms generate multiplicative renormalizations of the operators (e.g., (6.2.12)). This is similar to what we did in Section 5.6, where we showed that the counterterms in elementary Green's functions are generated by counterterms in the Lagrangian. Our motivation for starting with the graph-by-graph renormalization is again to allow a simple treatment of the problems of subdivergences. We do not need a new proof that the counterterms for operator insertions are local; our original proof suffices.

As before, we have a choice of many renormalization prescriptions. The ones that are most useful for subsequent developments are the BPHZ scheme – see Zimmermann (1973a) – and minimal subtraction – see Breitenlohner & Maison (1977a, b, c) and Collins (1975b). In any case we have a subtraction operator $T(G)$ which is applied to a graph G (after removal of subdivergences) in order to extract the divergent part of G. In the

BPHZ scheme, $T(G)$ gives the first $\delta(G)$ terms in the Taylor expansion about zero external momentum, where $\delta(G)$ is, as usual, its degree of divergence.

In the case of minimal subtraction we must state how the unit of mass μ is treated. As can be seen from the examples in Section 6.2, we must always arrange to compute the pole part of a quantity whose dimension does not vary with d. So suppose we need $T_{\overline{\text{MS}}}(G)$ for a 1PI graph G of dimension $\mu^{B(d-d_0)}$. Then we define

$$T = \mu^{B(d-d_0)}\{\text{pole part of } G\mu^{B(d_0-d)}\}, \tag{6.3.1}$$

where, as usual, d_0 is the physical space-time dimension.

The definition of a renormalized Green's function by the R-operation is rather abstract, and we will now show that it amounts to adding counterterm operators. In the one-loop examples of Section 6.2 this was rather obvious. In the general case, we start from the formula for renormalization of an arbitrary Green's function (see (5.3.7)–(5.3.9)):

$$R(G) = G + \sum_{\gamma} C_{\gamma}(G). \tag{6.3.2}$$

Here, as usual, the sum is over all subgraphs γ that consist of a set of disjoint 1PI subgraphs $\gamma_1, \ldots, \gamma_n$. Each γ_i is replaced by its overall counterterm vertex, generated as in Section 5.3.

Let us distinguish the various γ_i's that occur, according to the number of composite operator insertions that they contain. Consider, for example, Fig. 6.3.1, which illustrates the renormalization of $\langle 0|T\phi\ \phi\ \phi^2/2|0\rangle$ in the ϕ^3 theory in six dimensions. There is a one-loop subgraph γ_a for the three-point function. This is renormalized by its counterterm C_{γ_a} in the Lagrangian. There is also a two-loop subgraph γ_b which contains the composite operator vertex. (This has a subdivergence, which must be subtracted.) The counterterm C_{γ_b} can be considered as

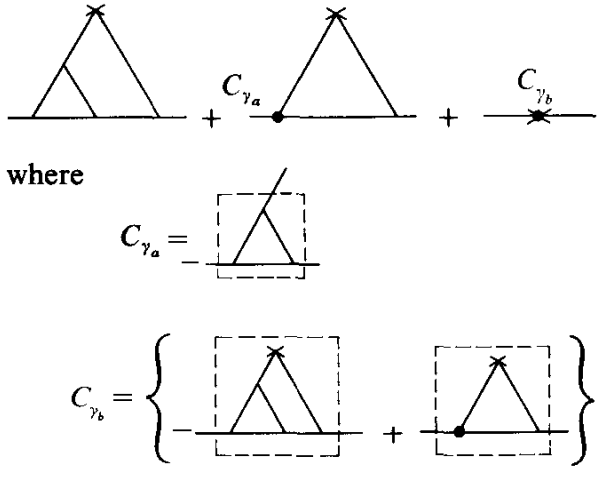


Fig. 6.3.1. Renormalization of two-loop graph with insertion of composite operator.

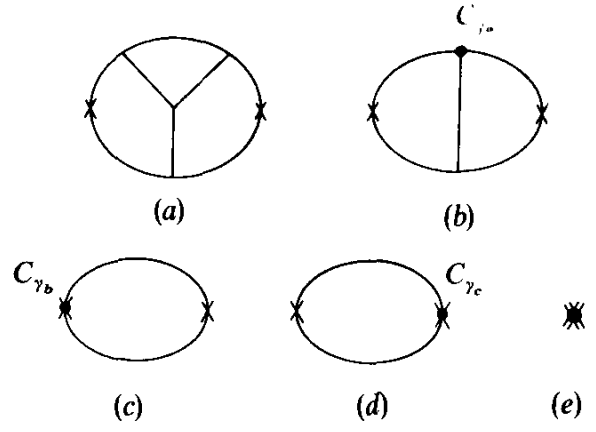


Fig. 6.3.2. Renormalization of three-loop graph for $\langle 0|T[\phi^2]/2[\phi^2]/2|0\rangle$.

generated by an $O(g^4)$ term in the renormalization factor Z_a.

Next consider Fig. 6.3.2, a graph of order g^4 for the Green's function $\langle 0|T[\phi^2]/2[\phi^2]/2|0\rangle$. In addition to an overall logarithmic divergence, it has three divergent subgraphs. One subgraph γ_a is renormalized by a counterterm in the interaction Lagrangian. The other two subgraphs γ_b and γ_c each look like γ_b of Fig. 6.3.1, and are renormalized by the same counterterm. These two counterterms are generated from

$$\langle 0|T\tfrac{1}{2}[\phi^2](x)\tfrac{1}{2}[\phi^2](y)|0\rangle = Z_a^2\langle 0|T\tfrac{1}{2}\phi^2\tfrac{1}{2}\phi^2|0\rangle + \text{other terms from } Z_b \text{ and } Z_c, \quad (6.3.3)$$

by expanding each Z_a to $O(g^4)$ and picking out the terms for γ_b. Finally, the overall counterterm is obtained. It gives a term of $O(g^4)$ in the $C(x)$ of (6.2.16).

These arguments generalize easily to arbitrary graphs. It suffices to consider a renormalized Green's function of one composite operator

$$\langle 0|T\prod_{i=1}^{N}\phi(x_i)[A(y)]|0\rangle. \quad (6.3.4)$$

We define the renormalization by the recursive formula (6.3.2), and we wish to prove that this equals

$$\sum_B Z_{AB}\langle 0|T\prod_{i=1}^{N}\phi(x_i)B(y)|0\rangle. \quad (6.3.5)$$

Here Z_{AB} is a renormalization factor whose value is given by writing, in analogy to (5.6.5),

$$[A] = \sum_B Z_{AB}B = \sum_G \frac{1}{N_G!}D(G). \quad (6.3.6)$$

Here G is any 1PI basic graph that includes a vertex for A. It has N_G external lines in addition to the vertex for A, and $D(G)$ is the operator corresponding

to the overall counterterm $C(G)$. The factor $1/N_G!$ is just like the $1/N(G)$ in (5.6.5), to organize the symmetry factors.

Consider a graph for (6.3.4). We investigate one of its counterterms $C_\gamma(G)$. The subgraph γ consists of 1PI graphs $\gamma_1, \ldots, \gamma_n$. A γ_j that does not contain the vertex for A is replaced by $C(\gamma_j)$, which corresponds to one of the counterterms in $\mathscr{L}$. A γ_j which does contain the vertex for A must be one of the G's that are summed over in (6.3.6), so the counterterm C_{γ_j} is generated by one of the counterterm operators on the right of (6.3.6).

Now sum over all graphs G for our Green's function and expand each $R(G)$ by (6.3.2). The sum over 1PI subgraphs γ_j that correspond to counterterms in $\mathscr{L}$ can be done independently of the sum over the subgraphs giving the counterterms for the operator vertex. The result is then the desired result (6.3.5).

6.4 Operator mixing

We have seen that a renormalized composite operator $[A]$ is expressed in terms of unrenormalized operators by

$$[A] = \sum_B Z_{AB} B. \tag{6.4.1}$$

In the case of $[\phi^2]$ we saw that the operators that were needed as counterterms had the same or lower dimension. Let us now demonstrate this for the general case.

The proof is essentially dimensional analysis. Let G be a 1PI graph containing a vertex for A and having the same number N_B of external lines as a particular operator B. Now B is a product of N_B fields with a certain number D_B of derivatives. A counterterm can only be generated if the degree of divergence $\delta(G)$ is at least D_B. Now in a renormalizable theory all couplings have non-negative dimension, so

$$\begin{aligned} \delta(G) &= \dim(G) - \dim(\text{couplings}) \\ &\leq \dim(G). \end{aligned} \tag{6.4.2}$$

On the other hand, since $Z_{AB}B$ is a possible counterterm we have

$$\dim(G) = D_B + \dim(Z_{AB}). \tag{6.4.3}$$

But we only need B as a counterterm if $D_B \leq \delta(G)$, so

$$\dim(Z_{AB}) \geq 0$$

for every counterterm. This means that the maximum dimension of a counterterm operator is the dimension of A, as we wished to prove.

In Section 5.8.1 we examined the dependence of counterterms in the

Lagrangian on mass parameters. Provided we used minimal subtraction this dependence was polynomial, with the mass behaving as a dimensional coupling in determining allowed counterterms. The same argument applies here. The result is that Z_{AB} is a polynomial in masses (and super-renormalizable couplings) times the inevitable power of the unit of mass μ. The coefficients of the polynomial are dimensionless functions of the dimensionless couplings and of d. A typical example of this is given by our calculation in Section 6.2 of the renormalization of $[\phi^2]$ – see (6.2.11) and (6.2.12).

6.5 Tensors and minimal subtraction

Suppose we use minimal subtraction to define $[(\partial\phi)^2]$ and $[\partial_\mu\phi\partial_\nu\phi]$. It is tempting to suppose that

$$[(\partial\phi)^2] = g^{\mu\nu}[\partial_\mu\phi\partial_\nu\phi]. \tag{6.5.1}$$

This supposition is in fact false, as we will now demonstrate. This means that the taking of a trace does not commute with taking a finite part, in general. We will explain the significance of this fact.

The lowest-order graph (in ϕ^3 theory) for either operator is Fig. 6.5.1. The 1PI part for $\partial\phi^2/2$ before renormalization is

$$\begin{aligned} G(p) &= \frac{ig\mu^{3-d/2}}{2(2\pi)^d}\int d^dk \frac{k\cdot(k+p)}{(k^2-m^2)[(p+k)^2-m^2]} \\ &= \frac{g\mu^{d/2-3}}{128\pi^3}\int_0^1 dx\left[\frac{m^2-p^2x(1-x)}{4\pi\mu^2}\right]^{d/2-3}\times \\ &\quad\times\{\Gamma(2-d/2)p^2x(1-x)[m^2-p^2x(1-x)] \\ &\quad\quad+(d/2)\Gamma(1-d/2)[m^2-p^2x(1-x)]^2\} \end{aligned} \tag{6.5.2}$$

For $\partial_\mu\phi\partial_\nu\phi/2$ we have

$$\begin{aligned} G_{\mu\nu}(p) &= \frac{ig\mu^{3-d/2}}{2(2\pi)^d}\int d^dk \frac{k_\mu(k+p)_\nu}{(k^2-m^2)[(p+k)^2-m^2]} \\ &= \frac{g\mu^{d/2-3}}{128\pi^3}\int_0^1 dx\left[\frac{m^2-p^2x(1-x)}{4\pi\mu^2}\right]^{d/2-3}\times \\ &\quad\times\{\Gamma(2-d/2)p_\mu p_\nu x(1-x)[m^2-p^2x(1-x)] \\ &\quad\quad+\tfrac{1}{2}g_{\mu\nu}\Gamma(1-d/2)[m^2-p^2x(1-x)]^2\}. \end{aligned} \tag{6.5.3}$$

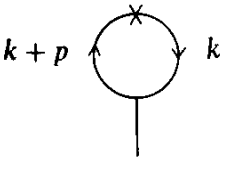


Fig. 6.5.1. One-loop graph for the operators in (6.5.1).

Manifestly

$$g^{\mu\nu}G_{\mu\nu} = G \qquad (6.5.4)$$

for the unrenormalized Green's functions. This has to be true since $\partial\phi^2 = g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$.

We can renormalize by minimal subtraction.

$$R[G(p)] = G(p) - \frac{g\mu^{d/2-3}}{128\pi^3(d-6)}[-3m^4 + \tfrac{4}{3}m^2p^2 - \tfrac{1}{6}p^4], \quad (6.5.5a)$$

$$R[G_{\mu\nu}(p)] = G_{\mu\nu}(p) - \frac{g\mu^{d/2-3}}{128\pi^3(d-6)}[p_\mu p_\nu(\tfrac{1}{3}m^2 - \tfrac{1}{15}p^2) - g_{\mu\nu}(\tfrac{1}{2}m^4 - \tfrac{1}{6}m^2p^2 + \tfrac{1}{60}p^4)]. \quad (6.5.5b)$$

Thus

$$R[G(p)] - g^{\mu\nu}R[G_{\mu\nu}(p)] = -\frac{g\mu^{d/2-3}}{128\pi^3}[\tfrac{1}{2}m^4 - \tfrac{1}{6}m^2p^2 + \tfrac{1}{60}p^4]. \quad (6.5.6)$$

The reason why contraction with $g^{\mu\nu}$ does not commute with the subtraction of the pole is simply that taking the trace introduces d-dependence. Thus:

$$g^{\mu\nu} \text{ pole part of } \frac{g_{\mu\nu}}{d-6} = \frac{d}{d-6},$$

$$\text{pole part of } \frac{g^{\mu\nu}g_{\mu\nu}}{d-6} = \text{pole } \frac{d}{d-6} = \frac{6}{d-6}.$$

We must evidently be careful to specify whether a trace is inside or outside of the renormalization. The need to do this is characteristic of dimensional regularization. Which place to put the trace depends on the problem under consideration.

The problem arises whenever we have to consider a tensor of rank at least 2. (It could also arise in connection with taking a trace of Dirac γ-matrices except that we choose the trace of the unit Dirac matrix to be independent of d.) We have discovered that our renormalized operators do not have all the properties that the bare operators do. The lack of commutativity of the trace and the finite-part operation is related to a physical effect, that there is an anomaly in the Ward identity for scale transformations – see Callan (1970), Symanzik (1970b) and Brown (1980).

If we were to use, say, zero-momentum subtractions (BPH or BPHZ), then the trace and the finite-part operation would commute – as can be checked from our example. So it might appear that zero-momentum subtraction provides a better all-purpose definition of renormalized operators than does minimal subtraction. However, some of the properties we will prove when using minimal subtraction now disappear or become more complicated.

For example, the equations of motion which we will prove in Section 6.6 are only true if the mass terms are oversubtracted. This turns out to prevent some Ward identities from being true when the simplest renormalized operators are used, whereas they are true in their simplest form when using minimal subtraction. We will prove this in Section 6.6 also. The moral is that one cannot completely eliminate the problems.

It is possible (Collins (1975b)) to construct a definition of, say, $[\partial_\mu\phi\partial_\nu\phi/2]$ which uses minimal subtraction and for which $g^{\mu\nu}[\partial_\mu\phi\partial_\nu\phi/2] = [(\partial\phi)^2]$. This is done by writing tensors in terms of Lorentz-irreducible components. Thus we write a second-rank tensor $M_{\mu\nu}$ as the sum of an antisymmetric term, a symmetric traceless term, and an invariant term:

$$M_{\mu\nu} = \tfrac{1}{2}(M_{\mu\nu} - M_{\nu\mu}) + \tfrac{1}{2}(M_{\mu\nu} + M_{\nu\mu} - (2/d)g_{\mu\nu}M^\kappa_\kappa) + (1/d)g_{\mu\nu}M^\kappa_\kappa.$$

The subtraction procedure is applied to each term separately. This definition loses other properties of the renormalized products. For example, conservation of energy and momentum is a consequence of the fact that the bare energy-momentum tensor $\theta_{\mu\nu}$ has zero divergence: $\partial^\mu\theta_{\mu\nu} = 0$. If we define a renormalized energy-momentum tensor $[\theta_{\mu\nu}]$ by our original definition of minimal subtraction, then this is the same as the bare $\theta_{\mu\nu}$ up to allowed redefinitions and it is conserved. But if we construct $[\theta_{\mu\nu}]$ by the procedure just suggested, then it is not conserved.

6.6 Properties

One of our motivations for working out the theory of renormalization of composite operators was that in Chapter 2 we had proved equations of motion and Ward identities. These results involved unrenormalized composite operators. So now that we have defined renormalized composite operators, we must prove the equations of motion and Ward identities expressed in terms of these renormalized operators. This is particularly important for the Ward identities, for these express the symmetry properties of the theory.

In this section we will derive a number of useful properties of the renormalized operators. Some properties will be purely technical, while others will be the actual equations of motion and Ward identities. Our proof will be given for the case that the operators are renormalized by minimal subtraction. A typical proof of some equation starts by observing that the corresponding equation is true for the unrenormalized operators. Renormalization is almost the same procedure applied to both sides of the unrenormalized equation, so the main problem is to find the places where the renormalization procedure is not identical for the two sides.

It is always possible to make the theorems false by changes in the renormalization prescription. The point of using minimal subtraction is that it is a universal prescription that preserves almost all of the desirable properties. (The reason is that it amounts, roughly, to defining each counterterm by the requirement 'remove exactly the singularity'.) These properties are relations between different operators.

The other standard renormalization prescription that preserves most of these relations is the BPH (or BPHZ) method of zero-momentum subtraction. In fact, the proofs were first given using the BPHZ prescription (Zimmermann (1973a, b), Lowenstein (1971) and Lam (1972)). However the use of minimal subtraction is better for gauge theories because of their infrared singularities. The proofs were given in this case by Collins (1975b) and Breitenlohner & Maison (1977a, b, c). All these works are rather technical. However, the basic ideas are simple.

Property 1. *Linearity*:

$$a[A] + b[B] = [aA + bB], \tag{6.6.1a}$$

where a and b are pure numbers, while A and B are composite operators. This equation is to be interpreted as an equation for Green's functions of the operator. That is, if X is any product of renormalized operators (elementary or composite), then

$$a\langle 0|T[A]X|0\rangle + b\langle 0|T[B]X|0\rangle = \langle 0|T[aA + bB]X|0\rangle \tag{6.6.1b}$$

Proof. This property is almost obvious. If A and B have different numbers of external legs (e.g. ϕ^2 and ϕ^4), then there is no simple way of defining the right-hand side of (6.6.1) except as being the left-hand side. But if A and B have the same fields, like $(\partial\phi)^2$ and ϕ^2, then the Feynman graphs for $\langle 0|TAX|0\rangle$, $\langle 0|TBX|0\rangle$ and $\langle 0|T(aA + bB)X|0\rangle$ are the same; the differences are only in the placement of powers of momentum. The equation corresponding to (6.6.1b) is true for the basic graphs (i.e., without counterterms). To obtain the renormalized Green's functions we apply the forest formula to each graph. The terms in the forest formula are the same, since the graphs for the three Green's functions are the same. Then (6.6.1) follows from linearity of the subtraction operators T_γ.

Comments. (1) It is necessary to be pedantic about this proof because: (a) it is a prototype for less trivial cases, and (b) it fails for the case of zero-momentum subtractions. The reason for the failure is that the T_γ operation is then not linear. For example, $[(\partial\phi)^2]$ and $[(\partial\phi)^2 + m^2\phi^2]$ need two extra subtractions

compared with $[\phi^2]$, because the degree of divergence is two higher. So we can only have

$$[(\partial\phi)^2]_{\text{BPH}} + m^2[\phi^2]_{\text{BPH}} = [(\partial\phi)^2 + m^2\phi^2]_{\text{BPH}} \tag{6.6.2}$$

if the $[\phi^2]$ operator is oversubtracted.

(2) The coefficients a and b in (6.6.1) must be independent of d, for otherwise taking a pole part is non-linear. Furthermore, we cannot use (6.6.1) to show, for example, $g^{\mu\nu}[\partial_\mu\phi\partial_\nu\phi] = [(\partial\phi)^2]$. As we saw, this equation is in fact false. The proof fails because it can only be applied to the case that we sum over a finite number of operators. It does not automatically apply to infinite summations. However, when we defined dimensional regularization in Chapter 4, we saw that our vectors and tensors have to have infinitely many components.

Property 2. *Differentiation is distributive.* Let A be the composite operator

$$A = \prod_{j=1}^{n} \phi_j(x)$$

where each ϕ_j is an elementary field or one of its derivatives. Then

$$\frac{\partial}{\partial x^\mu}[A] = \left[\frac{\partial A}{\partial x^\mu}\right] = \sum_{i=1}^{n}\left[\frac{\partial\phi_i}{\partial x^\mu}\prod_{j\neq i}\phi_i(x)\right]. \tag{6.6.3a}$$

Again this equation is to be interpreted for Green's functions:

$$\frac{\partial}{\partial x^\mu}\langle 0|T[A(x)]X|0\rangle = \langle 0|T[\partial A/\partial x^\mu]X|0\rangle. \tag{6.6.3b}$$

Proof. Let p^μ be the momentum leaving the Green's function (6.6.3*b*) at the vertex for A. Then the derivative $\partial/\partial x^\mu$ gives a factor $-ip^\mu$. The point of (6.6.3) is to state that we get the same results whether or not we take p^μ inside the finite-part operation. To prove the equation, it is enough to observe that this statement is true for the basic subtraction operator T_γ.

Comments. (1) We can contract μ with an index in A. Thus we have

$$g^{\mu\nu}\partial_\mu[\phi\partial_\nu\phi] = g^{\mu\nu}[\partial_\mu(\phi\partial_\nu\phi)] = [(\partial\phi)^2] + [\phi\,\Box\,\phi]. \tag{6.6.4}$$

Since the overall derivative merely gives a factor $-ip_\mu$ there is no possibility of introducing extra d-dependence by contracting with $g^{\mu\nu}$. This is in contrast to the case considered in Section 6.5.

(2) Note that derivatives are always implicitly taken outside of the time ordering. Thus:

$$\langle 0|T(\phi\partial_\mu\phi)(x)\tfrac{1}{2}\phi^2(y)|0\rangle = \lim_{z\to x}\frac{\partial}{\partial x^\mu}\langle 0|T\phi(z)\phi(x)\tfrac{1}{2}\phi^2(y)|0\rangle. \tag{6.6.5}$$

This gives the simplest Feynman rules in momentum space, with each derivative of a field giving a factor of momentum on the corresponding line. The lowest-order graph for the Fourier transform of (6.6.5) is

$$\int \mathrm{d}^d x \mathrm{d}^d y \mathrm{e}^{\mathrm{i}p\cdot x + \mathrm{i}q\cdot y} \langle 0 | T(\phi \partial_\mu \phi)(x) \tfrac{1}{2}\phi^2(y) | 0 \rangle$$

$$= (2\pi)^d \delta^{(d)}(p+q) \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{\mathrm{i}k_\mu}{(k^2 - m^2)[(k-p)^2 - m^2]}. \tag{6.6.6}$$

Property 3. *Simple equation of motion.* Decompose the action into a basic action and a counterterm action:

$$\mathscr{S} = \mathscr{S}_{\mathrm{b}} + \mathscr{S}_{\mathrm{ct}}, \tag{6.6.7}$$

just like the decomposition (5.1.1) of the Lagrangian, except that $\mathscr{S}_{\mathrm{b}}$ includes both the free and interaction terms: $\mathscr{S}_{\mathrm{b}} = \int \mathrm{d}^d x (\mathscr{L}_0 + \mathscr{L}_{\mathrm{b}})$. Then define functional derivatives with respect to renormalized fields

$$\mathscr{S}_\phi(x) \equiv \frac{\delta \mathscr{S}}{\delta \phi(x)} = \frac{\partial \mathscr{L}}{\partial \phi(x)} - \partial_\mu \frac{\partial \mathscr{L}}{\partial \partial_\mu \phi(x)}, \tag{6.6.8}$$

$$\mathscr{S}_\phi^{(\mathrm{b})}(x) = \frac{\delta \mathscr{S}_{\mathrm{b}}}{\delta \phi(x)} = \mathscr{S}_\phi \text{ with counterterms omitted.} \tag{6.6.9}$$

We already know the unrenormalized equation of motion (2.5.5)

$$\langle 0 | T \mathscr{S}_\phi(x) X | 0 \rangle = \mathrm{i} \frac{\delta}{\delta \phi(x)} \langle 0 | T X | 0 \rangle. \tag{6.6.10}$$

Now we wish to prove the renormalized equation

$$\langle 0 | T[\mathscr{S}_\phi^{\mathrm{b}}(x)] X | 0 \rangle = \mathrm{i} \frac{\delta}{\delta \phi(x)} \langle 0 | T X | 0 \rangle, \tag{6.6.11}$$

from which follows the operator equation

$$[\mathscr{S}_\phi^{\mathrm{b}}] = 0. \tag{6.6.12}$$

Comments. (1) The functional derivatives are to be treated in a purely formal sense.

(2) Even though $[\mathscr{S}_\phi(x)]$ is zero as an operator, its Green's functions (6.6.11) are non-zero because we define them by taking derivatives outside the time-ordering. Bringing them inside gives equal-time commutators (Section 2.5).

(3) Signs for fermions are easiest to determine by examining the derivation in Chapter 2.

Example. In the ϕ^3 theory

$$\mathscr{L} = Z(\partial\phi)^2/2 - m_0^2 Z\phi^2/2 - g_0 Z^{3/2}\phi^3/6,$$

we have

$$\mathscr{S}_\phi = -Z\Box\phi - m_0^2 Z\phi - \tfrac{1}{2}g_0 Z^{3/2}\phi^2, \tag{6.6.13}$$

$$[\mathscr{S}_\phi^{\mathrm{b}}] = -\Box\phi - m^2\phi - \tfrac{1}{2}\mu^{3-d/2}g[\phi^2]. \tag{6.6.14}$$

Then cases of (6.6.11) are

$$\begin{aligned}\langle 0|T[\mathscr{S}_\phi^{\mathrm{b}}(x)]\phi(y)|0\rangle &= (-\Box_x - m^2)\langle 0|T\phi(x)\phi(y)|0\rangle \\ &\quad - \tfrac{1}{2}\mu^{3-d/2}g\langle 0|T[\phi^2(x)]\phi(y)|0\rangle \\ &= i\delta^{(d)}(x-y),\end{aligned} \tag{6.6.15}$$

$$\begin{aligned}&\langle 0|T[\mathscr{S}_\phi^{\mathrm{b}}(x)][\phi^3(y)]\phi(z)\phi(w)|0\rangle_{\mathrm{R}} \\ &\quad = 3i\delta^{(d)}(x-y)\langle 0|T[\phi^2(y)]\phi(z)\phi(w)|0\rangle \\ &\qquad + i\delta^{(d)}(x-z)\langle 0|T[\phi^3(y)]\phi(w)|0\rangle \\ &\qquad + i\delta^{(d)}(x-w)\langle 0|T[\phi^3(y)]\phi(z)|0\rangle.\end{aligned} \tag{6.6.16}$$

Proof. A general proof of (6.6.11) is rather complicated because of the arbitrary number of fields. To show the main points it is sufficient to prove one case, (6.6.15) in ϕ^3 theory. The problem in proving the renormalized equation from the unrenormalized equation is that the Feynman graphs are different for the different terms.

We write

$$\begin{aligned}\mathscr{S}_\phi^{\mathrm{b}} &= \mathscr{S}_{0,\phi} + \mathscr{S}_{\mathrm{int},\phi}^{\mathrm{b}} \\ &= -(\Box + m^2)\phi - \tfrac{1}{2}g\mu^{3-d/2}\phi^2.\end{aligned} \tag{6.6.17}$$

Examples of low-order graphs are given in Figs. 6.6.1 and 6.6.2. The counterterms are those arising from the action.

In momentum space we evidently have

$$\langle 0|T\tilde{\mathscr{S}}_{0,\phi}(p)\tilde{\phi}(q)|0\rangle = (p^2 - m^2)\langle 0|T\tilde{\phi}(p)\tilde{\phi}(q)|0\rangle. \tag{6.6.18}$$

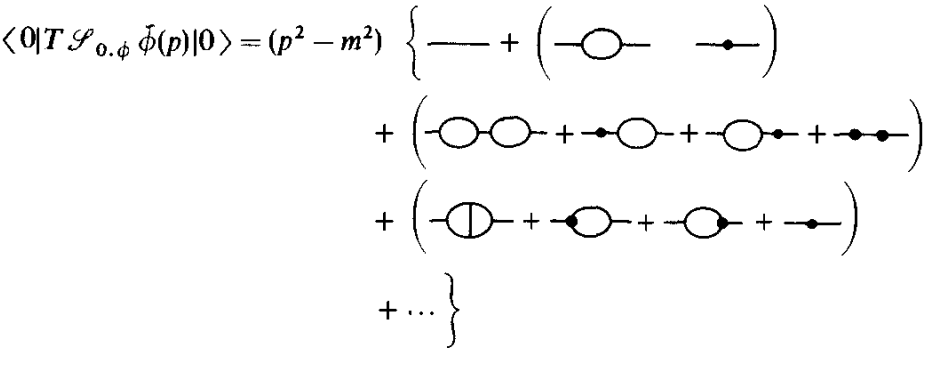


Fig. 6.6.1. Low-order graphs for (6.6.15) with free part of action.

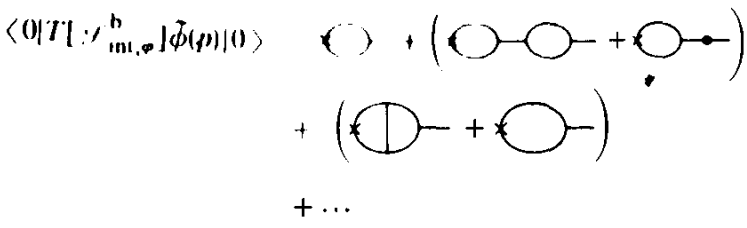


Fig. 6.6.2. Low-order graphs for (6.6.15) with interaction part of action.

The $p^2 - m^2$ multiplying the free propagator attached to $\tilde{\phi}(p)$ cancels the denominator of the propagator. Thus in the graph where the other end of the propagator is $\tilde{\phi}(q)$ we obtain the right-hand side of (6.6.15).

In all the remaining graphs the other end of the propagator is an interaction vertex, either a basic interaction or a counterterm. In fact we obtain

$$\langle 0|T\tilde{\mathscr{S}}_{\text{int},\phi}(p)\tilde{\phi}(q)|0\rangle \tag{6.6.19}$$

where

$$\begin{aligned}\mathscr{S}_{\text{int},\phi}(x) &= \mathscr{S}_{\phi} - \mathscr{S}_{0,\phi} \\ &= -(Z-1)\Box\phi - (m_0^2 Z - m^2)\phi - \tfrac{1}{2}g_0 Z^{3/2}\phi^2. \end{aligned} \tag{6.6.20}$$

This is exactly what we must obtain in order that the unrenormalized equation of motion (6.6.10) is true. Notice that because (6.6.18) is finite, so is (6.6.19). We must now prove that the counterterms in (6.6.20) are precisely those that are needed to give the operator $[S^{\text{b}}_{\text{int},\phi}] = -\frac{1}{6}g\mu^{3-d/2}[\phi^2]$ renormalized according to our standard prescription for composite operators.

Now, the renormalization prescription is precisely to add to the basic term $S^{\text{b}}_{\text{int},\phi} = -\frac{1}{6}g\mu^{3-d/2}\phi^2$ a series of counterterm operators whose coefficients are pure poles at $d = 6$, so as to make its Green's functions finite. But this is precisely (6.6.20). The relation between the counterterms can be seen from a comparison of Figs. 6.6.1 and 6.6.3.

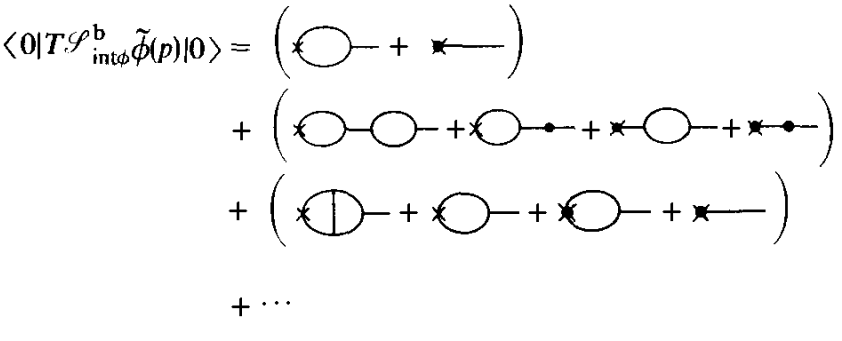


Fig. 6.6.3. Renormalization of Fig. 6.6.2.

Property 4. *Equation of motion times operator.* With the same notation as before we have

$$\langle 0|T[A\mathscr{S}_\phi(x)]X|0\rangle_R = i\langle 0|TA(x)\frac{\delta}{\delta\phi(x)}X|0\rangle_R, \tag{6.6.21}$$

where A is any product of operators at the same point. Hence

$$[A\mathscr{S}_\phi] = 0. \tag{6.6.22}$$

Comments and examples. (1) This property is crucial to proving Ward identities.

(2) All operators appearing on the right of (6.6.21) are to be renormalized.

(3) In ϕ^3 theory, cases of (6.6.21) are

$$\begin{aligned}&\langle 0|T[-\phi\Box\phi - m^2\phi^2 - \tfrac{1}{2}g\phi^3](x)\phi(y)\phi(z)|0\rangle \\ &\quad = i\delta(x-y)\langle 0|T\phi(y)\phi(z)|0\rangle + (y\leftrightarrow z),\end{aligned} \tag{6.6.23}$$

$$\begin{aligned}&\langle 0|T[-\phi^2\Box\phi - m^2\phi^3 - \tfrac{1}{2}g\phi^4](x)[\phi^2](y)\phi(z)|0\rangle_R \\ &\quad = 2i\delta(x-y)\langle 0|T[\phi^3](y)\phi(z)|0\rangle \\ &\qquad + i\delta(x-z)\langle 0|T[\phi^2](y)[\phi^2](z)|0\rangle_R.\end{aligned} \tag{6.6.24}$$

Proof. The unrenormalized version of (6.6.21) follows almost directly from the previous property in its unrenormalized version. This in turn follows from the functional-integral solution of the theory, as shown in Section 2.5. We have

$$\begin{aligned}&\langle 0|TA(x)\mathscr{S}_\phi(x)X|0\rangle \\ &\qquad = i\langle 0|TA(x)\frac{\delta X}{\delta\phi(x)}|0\rangle + i\langle 0|T\frac{\delta A(x)}{\delta\phi(x)}X|0\rangle.\end{aligned} \tag{6.6.25}$$

Then we use the fact that in dimensional regularization

$$\delta^{(d)}(0) = \int d^d p 1 = 0, \tag{6.6.26}$$

according to the results in Chapter 4. This enables us to eliminate the $\delta A(x)/\delta\phi(x)$ term.

The renormalized equation of motion (6.6.21) can be proved by generalizing the method for the previous property. It is enough to consider the example (6.6.23). Low-order graphs for the left-hand side of (6.6.23) are shown in Fig. 6.6.4. The $(-\Box - m^2)$ factor in $\mathscr{S}_{0,\phi}$ cancels an attached propagator. If the other end attaches to an external field (viz., $\phi(y)$ or $\phi(z)$), then we have a contribution to the right-hand side. If it attaches to an interaction then the negative of a contribution with $\mathscr{S}_{\text{int}}$ is obtained, such as Fig. 6.6.5. Since these manipulations do not change the one-particle-

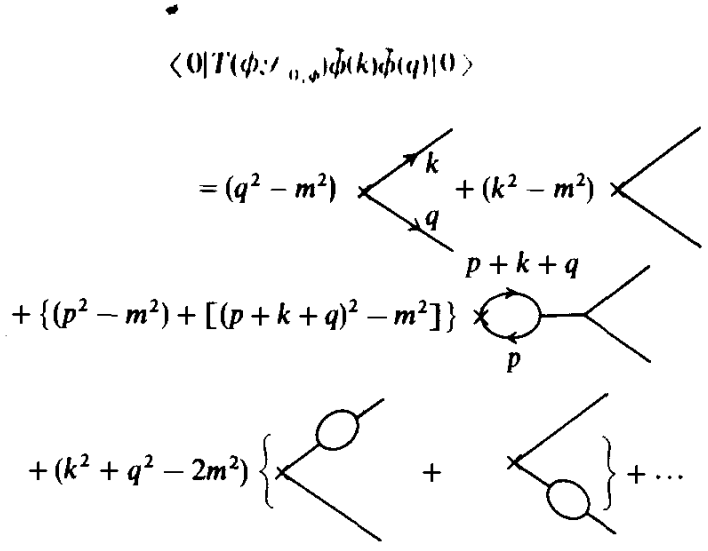


Fig. 6.6.4. Low-order graphs for the left-hand side of (6.6.23), using free action.

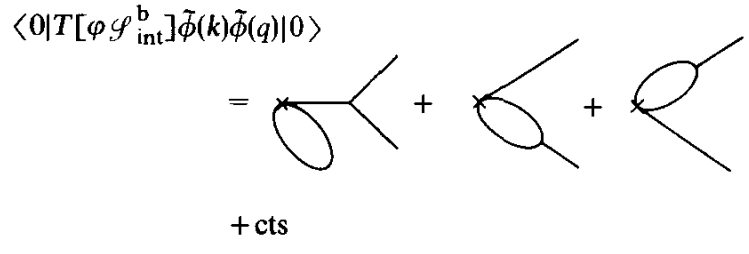


Fig. 6.6.5. Low-order graphs for the left-hand side of (6.6.23), using interaction part of action.

(ir)reducibility structure, renormalization can be performed without changing the result.

Since the $\mathscr{S}_{0,\phi}$ terms need renormalization two subtleties arise:

(1) Since $\Box\phi = g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$, the ambiguity about the placement of $g^{\mu\nu}$ is relevant. To preserve the derivation it must come before the renormalization is performed.

(2) In the BPHZ scheme $m^2\phi^2$ must be oversubtracted otherwise we cannot use linearity to combine $-\phi\Box\phi$ and $-m^2\phi^2$.

The case of (6.6.24) involves two further subtleties illustrated by Fig. 6.6.6. In the first graph (*a*) the $(-\Box - m^2)$ multiplies the line coming back to a $\phi(x)$ factor. This term gives zero after use of (6.6.26). The second graph (*b*) has two 1PI loops separated by a line. In the basic graph, the $q^2 - m^2$ factor cancels the propagator to give graph (*c*), which has a different reducibility structure. The first two counterterm graphs give the obvious counterterm graphs in Fig. 6.6.6(*d*). These correspond to the first two counterterms in Fig. 6.6.6(*b*). But the last counterterm in (*b*) has two vertices while the corresponding graph (*e*) has one vertex. Their operator structure is different: graph (*e*) is another counterterm to renormalize the $x \to y$ singularity of $\phi^4(x)\,\phi^2(y)$. Even so the

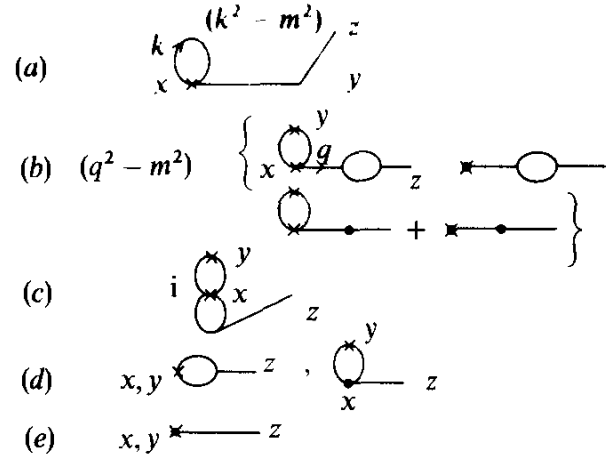


Fig. 6.6.6. Some graphs for (6.6.24).

two counterterm graphs must be equal. They are both a single free propagator times a pole part coefficient times a polynomial in momentum times the same power of the unit of mass. They both make the complete Green's function finite.

The general proof is rather tedious and can be found in Collins (1975b). This proof was given for the minimal subtraction scheme, but also works for the BPH(Z) scheme. In the original BPHZ proof, by Lam (1972), of (6.6.21), there is no treatment of this complication, that the counterterms for the two sides are not in manifest correspondence – i.e., that the forests are different.

Property 5. *Ward identities*: We will use the notation of Section 2.6 for transformations under potential symmetry operations. Let

$$\phi_j \to \phi_j + \delta\phi_j$$

be an infinitesimal transformation of the fields under which the basic Lagrangian transforms as

$$\mathscr{L}_{\text{basic}} \to \mathscr{L}_{\text{basic}} + \Delta_{\text{b}} + \partial_\mu Y^{\mu}_{\text{b}}.$$

We are restricting our attention to the transformations generated by one particular generator of a group. Thus, as compared to Section 2.6, we now drop the index 'α', which labelled the generators. The subscript 'b' on Δ_{b} and Y_{b} indicates that we are considering transformations on the basic Lagrangian (i.e., without counterterms). In the equations below, we will add in the counterterms by use of our standard renormalization scheme. Note also that in setting up the renormalized Green's functions, in Section 2.8, we defined a free Lagrangian $\mathscr{L}_0$ and a basic interaction Lagrangian $\mathscr{L}_{\text{b}}$. We now work with their sum: $\mathscr{L}_{\text{basic}} = \mathscr{L}_0 + \mathscr{L}_{\text{b}}$.

We proved earlier the unrenormalized Ward identity (2.7.6). The

renormalized Ward identity is

$$\frac{\partial}{\partial x^\mu}\langle 0|T[j^\mu_{\rm b}(x)]X|0\rangle_{\rm R}$$

$$=\langle 0|T[\Delta_{\rm b}(x)]X|0\rangle_{\rm R}-{\rm i}\langle 0|T\delta_{\rm b}\phi(x)\frac{\delta X}{\delta\phi(x)}|0\rangle_{\rm R}. \tag{6.6.27}$$

We will have to prove it. From it follows

$$\partial_\mu[j^\mu_{\rm b}]=[\Delta_{\rm b}] \tag{6.6.28}$$

and by integration over all x:

$$0=\int {\rm d}^4x\langle 0|T[\Delta^{\rm b}(x)]X|0\rangle_{\rm R}-{\rm i}\delta\langle 0|TX|0\rangle_{\rm R}. \tag{6.6.29}$$

Of course $\Delta=0$ for a symmetry. The current is

$$[j^\mu_{\rm b}]=\sum_j\left[\frac{\partial\mathscr{L}_{\rm basic}}{\partial\partial_\mu\phi_j}\delta_{\rm b}\phi_j\right]-[Y^\mu_{\rm b}]. \tag{6.6.30}$$

Proof. In defining $Y^\mu_{\rm b}$ and $\Delta_{\rm b}$, we have used the basic Lagrangian, i.e., the one with the counterterms omitted. This is because we use the operation symbolized by square brackets to generate the counterterms. The proof of (6.6.27)–(6.6.30) follows the usual proof of Noether's theorem, but using the previously proved properties to write it directly in terms of renormalized operators.

First we use linearity and distributivity to obtain

$$-[\Delta_{\rm b}]+[\partial_\mu j^\mu_{\rm b}]=\sum_j\partial_\mu\left[\left(\frac{\partial\mathscr{L}_{\rm basic}}{\partial\partial_\mu\phi_j}\right)\delta\phi_j-Y^\mu_{\rm b}\right]+[\partial_\mu Y^\mu_{\rm b}-\delta\mathscr{L}_{\rm basic}]$$

$$=\sum_j\left[\partial_\mu\frac{\partial\mathscr{L}_{\rm basic}}{\partial\partial_\mu\phi_j}\delta\phi_j\right]-\sum_j\left[\frac{\partial\mathscr{L}_{\rm basic}}{\partial\phi_j}\delta\phi_j\right]$$

$$=-\sum_j[\delta\phi_j\mathscr{S}^{\rm b}_{\phi_j}]. \tag{6.6.31}$$

From this the Ward identity (6.6.27) follows by the equation of motion (6.6.22).

We exchanged the order of renormalization and tracing over μ to write $\partial_\mu[j^\mu_{\rm b}]=[\partial_\mu j^\mu_{\rm b}]$. This is permitted – see our remarks below (6.6.4).

Comments (1) The theorem appears to give an unrestricted proof of the renormalized Ward identities. This appearance is false, since there are symmetries that can and often do have anomalous breaking – see Chapter 13. Such symmetries are dilatation and conformal symmetries and

chiral and supersymmetries. The potential for such anomalies can be seen by computing Δ^b in a regulated theory; it contains a non-zero coefficient which vanishes as $d \to 4$. Such is the case for conformal transformations and for chiral symmetries (where the transformations involve γ_5 or $\varepsilon_{\kappa\lambda\mu\nu}$ explicitly). Minimal subtraction is then not easily applicable and the properties we used in the proof are false.

(2) Corresponding problems appear with any other regulator and with any other renormalization scheme (Piguet & Rouet (1981)). Minimal subtraction confines the problems to cases with anomalies.

Property 6. *Non-renormalization of current*: Consider an exact internal symmetry (such as the symmetry that gives electric charge conservation). Compute the corresponding unrenormalized current j^μ from the complete Lagrangian. Now j^μ contains counterterms derived from the counterterm Lagrangian. We will now prove that these make j^μ finite and that

$$j^\mu = [j^\mu_b]. \tag{6.6.32}$$

Comments This theorem does not apply to space-time symmetries – see Callan, Coleman & Jackiw (1970), Freedman, Muzinich & Weinberg (1974), Collins (1976), Brown & Collins (1980) and Joglekar (1976) for the case of the energy-momentum tensor. It also cannot be extended to the case of a non-conserved current unless the breaking term has dimension below that of $\mathscr{L}$ (Symanzik (1970a)). Furthermore, the proof does not apply directly if the transformation $\delta\phi_j$ is non-linear in ϕ_j. It also needs generalization for gauge theories.

Proof. Both j^μ and $[j^\mu_b]$ consist of the basic current j^μ_b plus some minimal subtraction counterterms. The difference

$$\varepsilon^\mu = j^\mu - [j^\mu_b] \tag{6.6.33}$$

is a series of pure pole terms, and we wish to prove it vanishes. Each term has dimension 3 or less (at $d = 4$), since the currents have dimension 3.

Now both j^μ_b and $[j^\mu_b]$ satisfy the same Ward identity, so

$$\partial_\mu \langle 0 | T\varepsilon^\mu(x) X | 0 \rangle = 0. \tag{6.6.34}$$

Thus $\partial_\mu \varepsilon^\mu = 0$, without use of equations of motion; any need to use the equations of motion would give a non-zero right-hand side to (6.6.34). In the absence of gauge fields, it is impossible to construct such a term. The theorem is thus proved.

In the presence of gauge fields, such terms do exist. For example, in

quantum electrodynamics we have a counterterm to the electromagnetic current proportional to $\partial_\nu F^{\mu\nu}$, where $F_{\mu\nu}$ is the field-strength tensor $\partial_\mu A_\nu - \partial_\nu A_\mu$. With non-abelian gauge fields one might have $\varepsilon_\mu \propto \varepsilon_{\kappa\lambda\mu\nu}\partial^\kappa(A_a^\lambda A_b^\nu)$, but the presence of the $\varepsilon_{\kappa\lambda\mu\nu}$ indicates a chiral symmetry, which in any case needs special treatment. Moreover, in a non-abelian theory, we must also take account of the constraints imposed by gauge invariance, which is a subject we will not treat until Chapter 12.

The energy-momentum tensor also has possible counterterms, like $(\partial_\mu\partial_\nu - g_{\mu\nu}\Box)\phi^2$ – see the references quoted above.

6.7 Differentiation with respect to parameters in $\mathscr{L}$

Consider Green's function derived from the bare classical Lagrangian

$$\mathscr{L} = Z(\partial A)^2/2 - m_B^2 A^2/2 - g_B A^4/4!$$

by using the functional integral:

$$G_N \equiv \langle 0|T\phi(x_1)\dots\phi(x_N)|0\rangle = \frac{\int [\mathrm{d}A]A(x_1)\dots A(x_N)\mathrm{e}^{\mathrm{i}\int\mathscr{L}}}{\int [\mathrm{d}A]\mathrm{e}^{\mathrm{i}\int\mathscr{L}}}. \tag{6.7.1}$$

Differentiating with respect to g_B gives

$$\frac{\partial G_N}{\partial g_B} = \frac{\int [\mathrm{d}A]A(x_1)\dots A(x_N)\left(\frac{-\mathrm{i}}{4!}\int \mathrm{d}^4y A^4(y)\right)\mathrm{e}^{\mathrm{i}\int\mathscr{L}}}{\int [\mathrm{d}A]\mathrm{e}^{\mathrm{i}\int\mathscr{L}}} - \frac{\left\{\int [\mathrm{d}A]A(x_1)\dots A(x_N)\mathrm{e}^{\mathrm{i}\int\mathscr{L}}\right\}\left\{\int [\mathrm{d}A]\left(\frac{-\mathrm{i}}{4!}\right)\int \mathrm{d}^4y A^4(y)\mathrm{e}^{\mathrm{i}\int\mathscr{L}}\right\}}{\left\{\int [\mathrm{d}A]\mathrm{e}^{\mathrm{i}\int\mathscr{L}}\right\}^2}$$

$$= \int \mathrm{d}^4y\left(\frac{-\mathrm{i}}{4!}\right)\langle 0|T\phi(x_1)\dots\phi(x_N)(\phi^4(y) - \overset{(X)}{\langle 0|\phi^4(y)|0\rangle})|0\rangle$$

$$= \mathrm{i}\int \mathrm{d}^4y\langle 0|T\phi(x_1)\dots\phi(x_N)\left(\frac{\partial\mathscr{L}(y)}{\partial g_B} - \langle 0|\frac{\partial\mathscr{L}(y)}{\partial g_B}|0\rangle\right)|0\rangle. \tag{6.7.2}$$

Similar formulae hold for differentiation with respect to the other parameters Z or m_B^2.

The renormalized equivalents of those equations are also useful. One use

will be to show, within perturbation theory, that terms quadratic in the fields can be shifted between free and interaction Lagrangians without affecting the Green's functions.

First consider differentiation with respect to the renormalized coupling, g. We differentiate the renormalized Green's function

$$G_N(x_1,\ldots,x_N)=\sum_\Gamma\left[\Gamma+\sum_\gamma C_\gamma(\Gamma)\right]. \tag{6.7.3}$$

Applied to each basic graph Γ in this formula, the differentiation just gives

$$\frac{-\mathrm{i}}{4!}\int \mathrm{d}^d y\langle 0|T\phi(y)^4\phi(x_1)\ldots\phi(x_N)|0\rangle|_{\text{no counterterms}}. \tag{6.7.4}$$

Renormalization of (6.7.4) produces a set of counterterms isomorphic to those in (6.7.3). So

$$\frac{\partial}{\partial g}G_N=\frac{-\mathrm{i}}{4!}\int \mathrm{d}^d y\langle 0|T\{[\phi(y)^4]-\langle 0|[\phi(y)^4]|0\rangle\}\phi(x_1)\ldots\phi(x_N)|0\rangle. \tag{6.7.5}$$

The subtraction of the vacuum expectation value of $[\phi^4]$ comes about because no vacuum bubbles are used in G_N. Thus each ϕ^4 vertex in $\partial G_N/\partial g$ is connected to some external line.

Suppose we let the basic Lagrangian be

$$\mathscr{L}_{\text{basic}}=z(\partial\phi)^2/2-m^2\phi^2/2-g\phi^4/4! \tag{6.7.6}$$

and let the free and interaction Lagrangians be

$$\mathscr{L}_0=(\partial\phi)^2/2-m_1^2\phi^2/2 \tag{6.7.7a}$$

$$\mathscr{L}_{\mathrm{b}}=(z-1)(\partial\phi)^2/2-(m^2-m_1^2)\phi^2/2-g\phi^4/4!, \tag{6.7.7b}$$

with $m_1^2+m_2^2=m^2$. Notice that we have allowed the $(\partial\phi)^2$ term to have an arbitrary coefficient. We choose to put some of the terms quadratic in ϕ into the interaction, so that we can derive an equation for $\partial G_N/\partial z$ or $\partial G_N/\partial m^2$ like (6.7.5). Then we will show we can move the quadratic terms to the $\mathscr{L}_0$ without changing the Green's functions.

Differentiation with respect to z or m^2 gives

$$\frac{\partial}{\partial z}G_N=\frac{\mathrm{i}}{2}\int \mathrm{d}^d y\langle 0|T\{[(\partial\phi)^2]-\langle 0|[(\partial\phi)^2]|0\rangle\}\phi(x_1)\ldots\phi(x_N)|0\rangle, \tag{6.7.8}$$

$$\frac{\partial}{\partial m^2}G_N=\frac{-\mathrm{i}}{2}\int \mathrm{d}^d y\langle 0|T\{[\phi(y)^2]-\langle 0|[\phi(y)^2]|0\rangle\}\phi(x_1)\ldots\phi(x_N)|0\rangle. \tag{6.7.9}$$

However, we may want to put all of $(\partial\phi)^2 z/2 - m^2\phi^2/2$ into the free Lagrangian. In this case the free propagator is

$$\mathrm{i}/(zp^2 - m^2).$$

We wish to prove that (6.7.8) and (6.7.9) remain valid.

In this case $\partial/\partial m^2$ applied to an unrenormalized graph Γ gives a sum over terms in which each propagator is differentiated

$$\frac{\partial}{\partial m^2}\frac{\mathrm{i}}{zp^2 - m^2} = -\mathrm{i}\left(\frac{\mathrm{i}}{zp^2 - m^2}\right)^2. \tag{6.7.10}$$

This gives us the same result for the unrenormalized graphs as the right-hand side of (6.7.9). Next, we differentiate a counterterm graph $C_\gamma(\Gamma)$. Either a propagator is differentiated, so that the $-\mathrm{i}$ in (6.7.9) gives the basic vertex for $[\phi^2(y)]$, or a counterterm $C(\gamma_1)$ is differentiated. In this second case there is also a counterterm graph $C_\gamma(\partial\Gamma/\partial m^2)$ with a term $\partial\gamma_1/\partial m^2$

Now

$$C(\partial\gamma_1/\partial m^2) = \partial C(\gamma_1)/\partial m^2 \tag{6.7.11}$$

in the minimal subtraction scheme. (The reason is that both are defined to be pure poles times μ to a power – the same for both graphs.) We thus obtain all the counterterm graph for the right-hand side of (6.7.9).

Similarly (6.7.8) is true if $z(\partial\phi)^2$ is all in the free Lagrangian $\mathscr{L}_0$.

We thus see that, for any renormalized parameter λ in the Lagrangian $\mathscr{L}$, we have

$$\frac{\partial}{\partial\lambda}G_N = \mathrm{i}\int \mathrm{d}^d y \langle 0|T\left\{\left[\frac{\partial\mathscr{L}_{\text{basic}}}{\partial\lambda}(y)\right] - \left\langle 0\left|\left[\frac{\partial\mathscr{L}_{\text{basic}}}{\partial\lambda}(y)\right]\right|0\right\rangle\right\}\phi(x_1)\ldots\phi(x_N)|0\rangle \tag{6.7.12}$$

From this result we can see that the Lagrangian (6.7.6) is equivalent to the one with unit kinetic term

$$\mathscr{L}'_{\text{b}} = (\partial\phi')^2/2 - m'^2\phi'^2/2 - g'\phi'^4/4! \tag{6.7.13}$$

by a scaling of the field, with

$$\phi = z^{-1/2}\phi', \tag{6.7.14a}$$

$$m^2 = zm'^2, \tag{6.7.14b}$$

$$g = z^2 g'. \tag{6.7.14c}$$

The proof is to write (6.7.6) as

$$\mathscr{L}_{\text{basic}} = z(\partial\phi)^2/2 - m'^2 z\phi^2/2 - g' z^2\phi^4/4!. \tag{6.7.15}$$

Then differentiation of a renormalized Green's function G_N of ϕ with respect to z gives

$$\begin{aligned} z\mathrm{d}G_N/\mathrm{d}z &= z\partial G_N/\partial z|_{\text{fixed } m,g} \\ &\quad + m^2\partial G_N/\partial m^2|_{\text{fixed } z,g} + 2g\partial G_N/\partial g|_{\text{fixed } z,m} \\ &= \mathrm{i}\int \mathrm{d}^4y\langle 0|T\{[z(\partial\phi)^2/2 - m^2\phi^2/2 - 2g\phi^4/4!] \\ &\qquad - \text{vacuum expectation value}\}\ \phi(x_1)\ldots\phi(x_N)|0\rangle \\ &= \frac{\mathrm{i}}{2}\int \mathrm{d}^4y(\langle 0|T\{[\phi\mathscr{S}^{\mathrm{b}}_{\phi}] - \text{vacuum expectation value}\}\times \\ &\qquad \times \phi(x_1)\ldots\phi(x_N)|0\rangle). \end{aligned} \tag{6.7.16}$$

We now use the equation of motion (6.6.21) with $A = \phi$ to give

$$z\mathrm{d}G_N/\mathrm{d}z = -NG_N/2. \tag{6.7.17}$$

From this it follows that

$$G_N|_{z,g,m^2} = z^{-N/2}G_N|_{z\to 1, g\to g', m^2\to m'^2}$$

i.e.,

$$\langle 0|T\,\phi(x_1)\ldots\phi(x_N)|0\rangle = z^{-N/2}\langle 0|T\,\phi'(x_1)\ldots\phi'(x_N)|0\rangle, \tag{6.7.18}$$

exactly as we would expect. The proof is non-trivial only because we are shifting terms between the free and interaction Lagrangians. Thus we must ensure that counterterms do not go astray.

6.8 Relation of renormalizations of ϕ^2 and m^2

Observe that at order g^2 the renormalization factor Z_a for ϕ^2 in (6.2.13) and (6.2.11) is the inverse of the renormalization factor m_B^2/m^2. This relation is true to all orders, as we will now prove. (We are now back in ϕ^3 theory at $d = 6$.)

We use the renormalized formula

$$\begin{aligned} m^2\frac{\partial}{\partial m^2}G_N &= \mathrm{i}\int \mathrm{d}^4y\langle 0|T\left\{m^2\frac{\partial\mathscr{L}}{\partial m^2} - \text{vacuum expectation value}\right\}\times \\ &\qquad \times \phi(x_1)\ldots\phi(x_N)|0\rangle \\ &= \frac{-\mathrm{i}}{2}m^2\int \mathrm{d}^4y\langle 0|T\{[\phi^2] - \langle 0|[\phi^2]|0\rangle\}\phi(x_1)\ldots\phi(x_N)|0\rangle. \end{aligned}$$

But we also have (6.8.1)

$$\mathscr{L} = (\partial\phi_0)^2/2 - m^2Z_m\phi_0^2/2 - g_0\phi_0^3/6, \tag{6.8.2}$$

so that

$$m^2 \partial \mathscr{L}/\partial m^2 = -m^2 Z_m \phi_0^2/2. \tag{6.8.3}$$

Hence

$$\frac{\partial}{\partial m^2} G_N = -\frac{\mathrm{i}}{2} m^2 Z_m \int \mathrm{d}^d y \langle 0 | T\{\phi_0^2 - \langle 0|\phi_0^2|0\rangle\} \phi(x_1) \ldots \phi(x_N)|0\rangle. \tag{6.8.4}$$

Therefore

$$\int \mathrm{d}^d y m^2 [\phi^2] = \int \mathrm{d}^d y m_0^2 \phi_0^2, \tag{6.8.5}$$

from which the desired result follows.

Generalizations of this method can be found in Brown (1980) and Brown & Collins (1980).

7

Renormalization group

As we saw in Chapter 3, the renormalization procedure has considerable arbitrariness: the counterterm for a graph must cancel its divergence but may contain any amount of finite part. A rule for choosing the value of the counterterm we called a renormalization prescription. In one-loop order it was clear from the examples that a change in renormalization prescription can be cancelled by a change in the finite, renormalized couplings corresponding to each divergence. Thus a change in renormalization prescription does not change the theory but only the parametrization by renormalized coupling and mass. What is not so easy is to see that this property is true to all orders. This we will show in Section 7.1. The invariance of the theory under such transformations is called renormalization-group (RG) invariance.

A particularly useful type of change of renormalization prescription is to change the renormalization mass μ. Infinitesimal changes are conveniently described by a differential equation, called the renormalization-group equation, which is derived in Section 7.3. This leads to the concept of the effective momentum-dependent coupling. This concept is very useful in calculations of high-energy behavior, as explained in Section 7.4. The coefficients in the renormalization-group equation are called the renormalization-group coefficients and are important properties of a theory. Various developments of the formalism occupy the remaining sections.

The renormalization group was first discussed by Stueckelberg & Petermann (1953) and by Gell-Mann & Low (1954). Very similar ideas are applied in statistical physics (Wilson & Kogut (1974)). Many important recent applications arise because of the asymptotic freedom of QCD.

Results of calculations of renormalization-group coefficients can be found in many places:

(1) Gross (1976) lists many one-loop results for theories with scalars and fermions and up to two loops for gauge theories with only fermions.
(2) Cheng, Eichten & Li (1974) give the β-function for a general renormalizable theory to one-loop order.

(3) Tarasov, Vladimirov & Zharkov (1980) compute renormalization-group coefficients to three-loop order in gauge theories with fermions using minimal subtraction.
(4) Vladimirov, Kazakov & Tarasov (1979) compute to four-loop order in ϕ^4 theory.
(5) Chetyrkin, Kataev & Tkachov (1981) and Chetyrkin & Tkachov (1981) compute the anomalous dimension in ϕ^4 theory at five-loop order.
(6) Tkachov (1981) summarizes the methods used for the above calculations.
(7) Caswell & Zanon (1981) perform calculations in supersymmetric theories at three-loop order.

7.1 Change of renormalization prescription

7.1.1 *Change of parametrization*

The techniques we will describe are valid for any theory. However, to be specific, we will mainly work with the theory we have been using as a source of examples, the ϕ^3 theory in six space-time dimensions. There are three alternative, but equivalent, forms in which to write the Lagrangian. First of all, we can write it in terms of the bare field ϕ_0:

$$\mathscr{L} = (\partial\phi_0)^2/2 - m_0^2\phi_0^2/2 - g_0\phi_0^3/6. \tag{7.1.1a}$$

(As before, we ignore the term linear in ϕ.) The importance of this form is that the bare field ϕ_0 is invariant under change of renormalization prescription: its normalization is determined, because it satisfies canonical equal-time commutation relations.

When we renormalize the theory, we obtain finite Green's functions of the renormalized field $\phi = Z^{-1/2}\phi_0$. In terms of the renormalized field, the Lagrangian is

$$\begin{aligned}\mathscr{L} &= Z(\partial\phi)^2/2 - m_0^2 Z\phi^2/2 - g_0 Z^{3/2}\phi^3/6 \\ &\equiv Z(\partial\phi)^2/2 - m_B^2\phi^2/2 - g_B\phi^3/6.\end{aligned} \tag{7.1.1b}$$

This is the second of the three forms.

In the perturbative theory of renormalization, we wrote the Lagrangian as the sum of a free Lagrangian, a basic interaction Lagrangian, and a counterterm Lagrangian:

$$\begin{aligned}\mathscr{L} &= \mathscr{L}_0 + \mathscr{L}_b + \mathscr{L}_{ct} \\ &= \mathscr{L}_{basic} + \mathscr{L}_{ct}.\end{aligned} \tag{7.1.1c}$$

This is the third form of the Lagrangian. Here, we have chosen to define a

basic Lagrangian

$$\mathscr{L}_{\text{basic}} = (\partial\phi)^2/2 - m^2\phi^2/2 - \mu^{3-d/2} g\phi^3/6, \tag{7.1.2}$$

where m and g are the renormalized mass and coupling. Since we will mostly use minimal subtraction, it is sensible to define g to be dimensionless, and therefore to introduce the unit of mass μ. The counterterm Lagrangian is

$$\mathscr{L}_{\text{ct}} = \delta Z(\partial\phi)^2/2 - \delta m^2\phi^2/2 - \delta g\phi^3/6, \tag{7.1.3}$$

and the counterterms δZ, δm^2, and δg are computed as definite functions of g, m and μ with the aid of some renormalization prescription.

To be concrete, let us use minimal subtraction, so that

$$\delta Z = \sum_{j=1}^{\infty} (6-d)^{-j} c_j(g, m, \mu), \tag{7.1.4a}$$

$$\delta m^2 = m^2 \sum_{j=1}^{\infty} (6-d)^{-j} b_j(g, m, \mu), \tag{7.1.4b}$$

$$\delta g = \mu^{3-d/2} \sum_{j=1}^{\infty} (6-d)^{-j} a_j(g, m, \mu). \tag{7.1.4c}$$

We saw, in Section 5.8, that in fact the coefficients a_i, b_i, c_i are independent of m and μ; they are functions of the dimensionless coupling g only. However we will not use this fact at the moment.

The three forms of the Lagrangian listed in (7.1.1) are equivalent – if we use any of them in the functional integral, then the same Green's functions will result. The coefficients Z, m_0^2, and g_0 will be singular when d approaches 6 with g, m, and μ fixed. The singularities will be just such as to give finite Green's functions of ϕ at $d = 6$.

The parametrization of the Green's functions by g, m, and μ is rather arbitrary. Suppose that we change variables to g', m', and μ', which are some given functions of g, m, and μ. These functions may even depend on the regulator, d, provided that the change of variable remains non-singular at $d = 6$. Then we get the same theory, for the collection of Green's functions G_N is unchanged. It is just the numerical values of the renormalized mass and coupling and of the unit of mass that have changed.

We may even change the scale of the renormalized field by writing $\phi' = \zeta\phi$, where ζ is finite. The Green's functions are now different:

$$G'_N = \zeta^N G_N.$$

But observe that the value of ζ is irrelevant for a physical observable like the S-matrix. For example, consider an S-matrix element involving N particles. It is obtained from G_N by (1) dividing out an external propagator $G_2(p_i)$ for

each external line and (2) multiplying by $\hat{z}^{1/2}$ for each external line, where $i\hat{z}$ is the residue of the pole of $G_2(p)$. Finally we let the momenta p_i go on-shell. Thus

$$S = \lim_{p^2 \to m_{\text{ph}}^2} \left\{ G_N \hat{z}^{N/2} \Big/ \prod_{i=1}^{N} G_2 \right\}$$
$$= \lim \{ G'_N (\zeta \hat{z}^{1/2})^N / \textstyle\prod G'_2 \}. \tag{7.1.5}$$

But

$$G_2 = \zeta^{-2} G'_2, \tag{7.1.6}$$

so the particle pole is at the same position in G'_2 as in G_2 and the new residue is

$$\hat{z}' = \zeta^2 \hat{z}. \tag{7.1.7}$$

Hence

$$S = \lim \{ G'_N (\hat{z}'^{1/2})^N / \textstyle\prod G'_2 \}. \tag{7.1.8}$$

and the S-matrix is invariant, as claimed.

7.1.2 *Renormalization-prescription dependence*

In the bare Lagrangian (7.1.1*a*), there are only two parameters. So there should be only a two-parameter collection of physical theories obtained from it. As we have just seen, the freedom to vary the scale of the renormalized field ϕ in the second form of $\mathscr{L}$, viz. (7.1.1*b*), does not introduce a third real parameter into the physics.

Unfortunately, we appear to have introduced a large and indefinite number of parameters by having to choose one out of the infinitely many possible renormalization prescriptions. One might suppose that in different renormalization prescriptions, the singular behavior of m_0 and g_0 as $d \to 6$ could be different in such a way that one picks up different phases of the theory. In fact, this is not so. We will show that a change of renormalization prescription is one of the reparametrizations discussed in the previous subsection 7.1.1. This is the property we have defined as renormalization-group invariance.

Even within a single renormalization prescription, we have introduced a third parameter, the unit of mass, μ. Notice that the basic Lagrangian does not depend on μ and g separately, but only on the combination $\mu^{3-d/2} g$; but notice also that this property is not true for, say, the renormalized Green's functions at *one-loop* order. However, a change of μ is in effect a change in renormalization prescription. Indeed, we could include in our definition of the renormalization prescription the requirement that μ and m have a fixed

ratio; we would still have two free parameters. A change in this ratio is then a change in renormalization prescription. Our proof of renormalization-group invariance will, in fact, only explicitly cover the case of a change in μ. The more general case will be essentially the same.

We will prove that if we change μ to μ', then the physics is unchanged, provided that we choose suitable new values, g' and m', for the renormalized coupling and mass. That is, an S-matrix element $S(g, m, \mu)$ satisfies

$$S(g, m, \mu) = S(g', m', \mu').$$

The bare mass and coupling m_0 and g_0 are similarly invariant.

The new renormalized field ϕ' with the new values (g', m', μ') of the parameters is not the same as with the old values but is related by

$$\phi' = \zeta\phi,$$

where ζ is a finite function of g, m, μ and μ'. Hence the renormalized Green's functions satisfy

$$\begin{aligned} G_N(p_1, \ldots, p_N; g, m, \mu) &= \zeta^{-N} G_N(p_1, \ldots, p_N; g', m', \mu') \\ &= \zeta^{-N} G'_N. \end{aligned} \tag{7.1.9}$$

It will be convenient, in our proof, to compute Green's functions of the original field ϕ, but with the new value μ' of the unit of mass. Now, in terms of the new field ϕ', the Lagrangian is

$$\mathscr{L} = Z'\partial\phi'^2/2 - m_B'^2\phi'^2/2 - g'_B\phi'^3/6. \tag{7.1.10}$$

Here we write

$$Z' = Z(g', m'/\mu', d), \tag{7.1.11a}$$

$$m_B'^2 = m_B^2(g', m', \mu', d), \tag{7.1.11b}$$

$$g'_B = g_B(g', \mu', d). \tag{7.1.11c}$$

These are the same functions of the new renormalized parameters m', g', and μ' as the original bare parameters were of the old renormalized parameters m, g, and μ. To get the Green's functions of the original field ϕ but with the new value of the unit of mass, we substitute $\zeta\phi$ for ϕ' to obtain

$$\begin{aligned} \mathscr{L} = {} & Z(g', m'/\mu', d)\zeta^2\partial\phi^2/2 - m_B^2(g', m'/\mu', d)\zeta^2\phi^2 \\ & - g_B(g', m', \mu', d)\zeta^3\phi^3/6. \end{aligned} \tag{7.1.12}$$

7.1.3 *Low-order examples*

Let us remind ourselves how the changes in g and m are obtained at one-loop order. We must start at tree approximation, where we ignore all counterterms. In order that the two formulae (7.1.1*b*) and (7.1.12) for $\mathscr{L}$ be

the same in tree approximation when we change μ to μ', we must write

$$g' = g_{\text{new}} = (\mu/\mu')^{3-d/2} g,$$
$$m' = m,$$
$$\zeta = 1. \text{ (All tree approximation.)} \tag{7.1.13}$$

Note that at $d=6$ we have $g'=g$ to lowest order. We distinguish g' and g_{new}: g_{new} is defined to be exactly the value of g' in tree approximation. In higher-order calculations, g' gets corrections, but g_{new} will be defined to be $(\mu/\mu')^{3-d/2}g$ always.

To treat higher-order corrections, we write the Lagrangian (7.1.12) as

$$\mathscr{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - \mu'^{3-d/2}[g(\mu/\mu')^{3-d/2}]\phi^3/6 + \delta' Z(\partial\phi)^2/2 - \delta' m^2\phi^2/2 - \delta' g\phi^3/6. \tag{7.1.14}$$

Our strategy will be to express the counterterms in (7.1.14) as minimal subtraction counterterms plus some new finite pieces. The finite pieces will accomplish the change of parametrization.

First we consider the one-loop graph for the self-energy, Fig. 3.1.1. The unrenormalized value is

$$\Gamma_1 = \frac{ig^2\Gamma(2-d/2)}{128\pi^3}\int_0^1 dx\,[m^2 - p^2x(1-x)]^{d/2-2}(4\pi\mu^2)^{3-d/2}. \tag{7.1.15}$$

This is invariant under the transformation $(\mu,\ m,\ g)\to(\mu',\ m,\ g(\mu/\mu')^{3-d/2})$. If we use unit of mass μ, then the counterterm is

$$C(\Gamma_1) = -\text{pole}(\Gamma_1) = \frac{-ig^2}{64\pi^3(d-6)}(m^2 - \tfrac{1}{6}p^2). \tag{7.1.16}$$

Next we use μ' and $g_{\text{new}} = (\mu/\mu')^{3-d/2}g$ instead of μ and g. The counterterm changes to

$$C'(\Gamma_1) = \frac{-ig_{\text{new}}^2}{64\pi^3(d-6)}(m^2 - \tfrac{1}{6}p^2) = \frac{-ig^2}{64\pi^3(d-6)}(m^2 - \tfrac{1}{6}p^2)\left(\frac{\mu}{\mu'}\right)^{6-d}. \tag{7.1.17}$$

Notice that we define $C'(\Gamma_1)$ to be the negative of the pole part of Γ_1, with the d-dependence of g_{new} ignored. That is, we consider the function $\Gamma_1 = \Gamma_1(p, g_{\text{new}}, m, \mu')$ and extract its pole at $d=6$ with g_{new} fixed. This prescription ensures that we may later replace g_{new} by its value at $d=6$ without changing the renormalized value of the graph.

Since both $C(\Gamma_1)$ and $C'(\Gamma_1)$ cancel the divergence, their difference is finite. We may therefore obtain the same value for the graph plus

counterterm by letting the counterterm coefficients in (7.1.14), namely $\delta' Z$ and $\delta' m^2$, each be a sum of two terms:

$$\delta' Z = \frac{g_{\text{new}}^2}{384\pi^3(d-6)} + \frac{g_{\text{new}}^2}{384\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right], \tag{7.1.18}$$

$$\delta' m^2 = \frac{m^2 g_{\text{new}}^2}{64\pi^3(d-6)} + \frac{m^2 g_{\text{new}}^2}{64\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right]. \tag{7.1.19}$$

The first term in each equation is the minimal subtraction counterterm for the new coupling, while the second term is finite as $d \to 6$. Using the formula (7.1.12) for the Lagrangian, we may regroup these terms to give

$$\begin{aligned}\zeta^2 &= 1 + \frac{g_{\text{new}}^2}{384\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right] \\ &= 1 + \frac{g^2}{384\pi^3}\left[\frac{1-(\mu/\mu')^{6-d}}{d-6}\right],\end{aligned} \tag{7.1.20}$$

$$m'^2\zeta^2 = m^2 + \frac{m^2 g^2}{64\pi^3}\left[\frac{1-(\mu/\mu')^{6-d}}{d-6}\right]. \tag{7.1.21}$$

Hence we obtain the value of m':

$$\begin{aligned}m'^2 &= m^2\left\{1 + \frac{5g^2}{384\pi^3}\left[\frac{1-(\mu/\mu')^{6-d}}{d-6}\right] + O(g^4)\right\} \\ &\to m^2\left[1 + \frac{5g^2}{384\pi^3}\ln(\mu/\mu') + O(g^4)\right] \quad \text{as } d \to 6.\end{aligned} \tag{7.1.22}$$

Also at $d = 6$ we have

$$\zeta(d=6) = 1 + \frac{g^2}{768\pi^3}\ln(\mu/\mu') + O(g^4). \tag{7.1.23}$$

We can apply the same procedure to the vertex graph, Fig. 3.6.1, whose value is the factor in curly brackets in (5.3.5). The counterterm $\delta' g$ is written as

$$\frac{g_{\text{new}}^3 \mu'^{3-d/2}}{64\pi^3(d-6)} + \frac{g_{\text{new}}^3 \mu'^{3-d/2}}{64\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right], \tag{7.1.24}$$

with the result that

$$\begin{aligned}\zeta^3 g' \mu'^{3-d/2} &= g_{\text{new}} \mu'^{3-d/2}\left\{1 + \frac{g_{\text{new}}^2}{64\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right]\right\} \\ &= g\mu^{3-d/2}\left\{1 + \frac{g^2}{64\pi^3}\left[\frac{1-(\mu/\mu')^{6-d}}{d-6}\right]\right\}.\end{aligned} \tag{7.1.25}$$

It follows that

$$g' = g(\mu/\mu')^{3-d/2}\left\{1 + \frac{3g^2}{256\pi^3}\left[\frac{1-(\mu/\mu')^{6-d}}{d-6}\right] + O(g^4)\right\}$$

$$\to g\left[1 + \frac{3g^2}{256\pi^3}\ln(\mu/\mu') + O(g^4)\right] \quad \text{as } d \to 6. \tag{7.1.26}$$

Our strategy for understanding the effect of a change in renormalization prescription is to absorb the difference into a finite counterterm. The counterterm will itself generate divergent counterterms when we insert it into a bigger graph. Finally we reorganize the Lagrangian by putting all the finite counterterms into the basic Lagrangian. Then we see that the change in renormalization prescription is exactly compensated by a change in the parameters of the theory.

What happens when we go to higher order? An example is given by the two-loop self-energy graph of Fig. 7.1.1(*a*). Graphs (*b*), (*c*), and (*d*) renormalize its subdivergences and its overall divergence. We write the renormalized graph with the original value of μ as

$$\Sigma_1(p, g, m, \mu) = \Sigma_a + 2\Sigma_b + \Sigma_d. \tag{7.1.27}$$

(We used the fact that $\Sigma_b = \Sigma_c$.) The unrenormalized graph is unchanged if we replace μ by μ' and g by g_{new}. But the vertex counterterm is treated exactly as at (7.1.24), so that:

$$\Sigma_b(p, g, m, \mu) = \Sigma_b(p, g_{\text{new}}, m, \mu') + [\Sigma_b(p, g, m, \mu) - \Sigma_b(p, g_{\text{new}}, m, \mu')]. \tag{7.1.28}$$

The first term contains the counterterm for one of the subgraphs, computed by minimal subtraction with unit of mass μ'. The second term exactly compensates the difference. It has the counterterm replaced by the finite

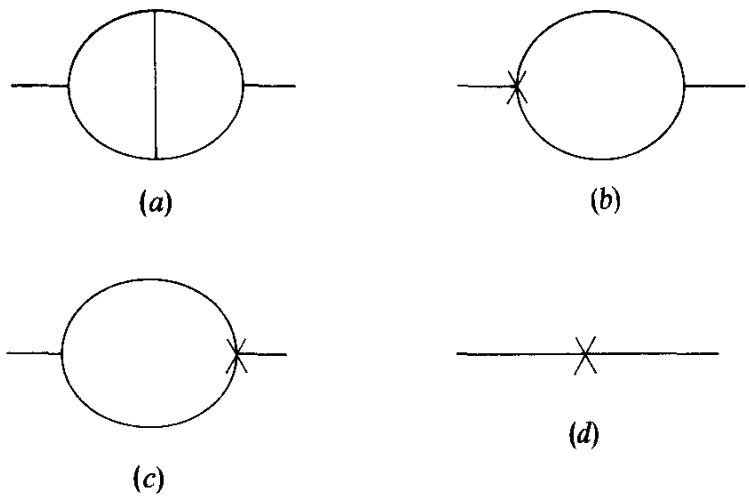


Fig. 7.1.1. Self-energy graph with counterterm graphs.

part in (7.1.24), viz.

$$\Delta g = \frac{g_{\text{new}}^3 \mu'^{3-d/2}}{64\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right]. \tag{7.1.29}$$

We now write

$$\Sigma_1 = \Sigma_a(p, g_{\text{new}}, m, \mu') + 2\Sigma_b(p, g_{\text{new}}, m, \mu') + \Sigma_d(p, g_{\text{new}}, m, \mu') + \left\{2\Gamma_1 \frac{\Delta g}{g_{\text{new}}\mu'^{3-d/2}} + \Sigma_d(p, g, m, \mu) - \Sigma_d(p, g_{\text{new}}, m, \mu')\right\}. \tag{7.1.30}$$

Here we wrote out the first three terms as the minimal renormalization of Σ_1 with unit of mass μ'. The remainder is finite, since Σ_1 is finite. The term $2\Gamma_1 \Delta g/(g_{\text{new}}\mu'^{3-d/2})$ is the one-loop self-energy graph with one of its couplings replaced by Δg. It has a divergence which can be cancelled by a minimal counterterm:

$$C = \frac{-\mathrm{i} g_{\text{new}} \Delta g}{64\pi^3(d-6)}(m^2 - \tfrac{1}{6}p^2). \tag{7.1.31}$$

Hence the term in curly brackets is

$$\left[\frac{2\Gamma_1 \Delta g}{g_{\text{new}}\mu'^{3-d/2}} + 2C\right] + [\Sigma_d(p, g, m, \mu) - \Sigma_d(p, g_{\text{new}}, m, \mu') - 2C]. \tag{7.1.32}$$

The second term is finite, since there are no remaining divergences. It is of the form

$$\mathrm{i}(-\Delta m^2 + \Delta Z p^2),$$

and so gives rise to another finite contribution to m'^2 and to ζ.

7.2 Proof of RG invariance

To show to all orders of perturbation theory that g', m' and ζ can be chosen so that (7.1.9) holds, we generalize from our treatment of the examples. We write

$$\begin{aligned} \zeta^3 g' \mu'^{3-d/2} &= g\mu^{3-d/2} + \Delta g, \\ \zeta^2 &= 1 + \Delta\zeta^2, \\ \zeta^2 m'^2 &= m^2 + \Delta m^2. \end{aligned} \tag{7.2.1}$$

The original version of the theory has counterterms computed with unit of mass μ:

$$\mathscr{L} = \mathscr{L}_{\text{basic}} + \mathscr{L}_{\text{ct}}(\mu). \tag{7.2.2}$$

We now change to unit of mass μ' and wish to show that identically the

same Green's functions and total Lagrangian are obtained if we make changes of the form (7.2.1).

The Lagrangian is written in the form

$$\mathscr{L} = \mathscr{L}_{\text{basic}} + \mathscr{L}_{\text{c}} + \mathscr{L}'_{\text{ct}}, \tag{7.2.3}$$

with the basic Lagrangian the same as before, but written as

$$\mathscr{L}_{\text{basic}} = (\partial\phi)^2/2 - m^2\phi^2/2 - g_{\text{new}}\mu'^{3-d/2}\phi^3/6. \tag{7.2.4}$$

(As before, we define $g_{\text{new}} = (\mu/\mu')^{3-d/2}g$.) The term $\mathscr{L}_{\text{c}}$ is a 'compensating Lagrangian' of the form

$$\mathscr{L}_{\text{c}} = \Delta\zeta^2\partial\phi^2/2 - \Delta m^2\phi^2/2 - \Delta g\phi^3/6. \tag{7.2.5}$$

The counterterm Lagrangian $\mathscr{L}'_{\text{ct}}$ in (7.2.3) is computed using minimal subtraction with unit of mass μ'.

We may later reorganize (7.2.3) so that the basic Lagrangian is taken as $\mathscr{L}_{\text{basic}} + \mathscr{L}_{\text{c}}$. We may drop the d-dependence of the finite counterterms Δg, $\Delta\zeta^2$, and Δm^2, since renormalized Green's functions are finite functions of renormalized quantities. Finally we may rescale the fields to give (7.1.10).

But a proof is most easily given with the form (7.2.3). Each of the finite counterterms is computed as a sum of terms, one for each 1PI graph contributing to the relevant Green's function:

$$\Delta g = \sum_{\Gamma} \Delta_{\Gamma}(g), \text{ etc.} \tag{7.2.6}$$

Particular cases are given by the examples in Section 7.1. Thus

$$\Delta_{3.1.1}\zeta^2 = \frac{g_{\text{new}}^2}{384\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right],$$

$$\Delta_{3.1.1}m^2 = \frac{m^2 g_{\text{new}}^2}{64\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right],$$

$$\Delta_{3.6.1}g = \mu'^{3-d/2}\frac{g_{\text{new}}^3}{64\pi^3}\left[\frac{(\mu'/\mu)^{6-d}-1}{d-6}\right],$$

$$\Delta_{7.1.1}(\zeta^2 p^2 - m^2) = -\mathrm{i}[\Sigma_d(p, g, m, \mu) - \Sigma_d(p, g_{\text{new}}, m, \mu') - 2C],$$

where the label on Δ indicates the number of the figure depicting the basic graph.

The general proof is by induction on the size of a graph. Consider a graph G contributing to some Green's function. Using the basic interaction $-g\mu^{3-d/2}\phi^3/6$ we renormalize it with unit of mass μ using the method of Section 5.11. The renormalized value of G is then

$$R(G) = G + \sum_{\gamma \subseteq G} C_{\gamma}(G). \tag{7.2.7}$$

Here the sum is over subgraphs of G that consist of one or more disjoint 1PI graphs, and $C_\gamma(G)$ denotes the replacement of each of these 1PI graphs by a counterterm with unit of mass μ. Similarly we can renormalize G with a different unit of mass μ' but with the same basic interaction $-g_{\text{new}}\mu'^{3-d/2}\phi^3/6$ to get

$$R'(G) = G + \sum_{\gamma \subseteq G} C'_\gamma(G). \tag{7.2.8}$$

Here, we use the prime to denote use of the unit of mass μ' instead of μ.

We will now derive a series of new basic graphs containing *finite* counterterms. These counterterms will be used to generate the compensating Lagrangian $\mathscr{L}_c$ in (7.2.3). The new graphs will need renormalization, and we will arrange them so that when they are added to $R'(G)$, we get back the original value $R(G)$. The aim will be to have a finite counterterm Δ_γ for every 1PI graph γ that is a vertex or self-energy graph. We will arrange the Δ_γ's so that

$$R(G) = R'\left[G + \sum_{\gamma \subsetneq G} \Delta_\gamma(G) + \Delta(G) \right]. \tag{7.2.9}$$

Here the sum over γ is over products of 1PI graphs. The overall counterterm for a graph is computed using minimal subtraction, but with the d-dependence of g_{new} and of the finite counterterms Δ_γ ignored (see our remarks below (7.1.17)). That is, the counterterms in R' are a series of poles at $d = 6$ with their coefficients being power series in g_{new}. In the case of a vertex subgraph, there is also the usual factor $\mu'^{3-d/2}$. The finite subtraction $\Delta(G)$ for the complete graph in (7.2.9) is only non-zero if G is a 1PI vertex or a self-energy graph.

We will prove the following relation between counterterms for a 1PI graph

$$C(\gamma) = C'\left(\gamma + \sum_{\delta \subsetneq \gamma} \Delta_\delta(\gamma) \right) + \Delta(\gamma). \tag{7.2.10}$$

The above equations (7.2.9) and (7.2.10) are trivially true for tree graphs, where no counterterms are needed, for we can set $\Delta(\text{tree graph}) = 0$. So let us assume they are true for all graphs smaller than a given graph G, with all the Δ_γ's finite. There are two cases: (*a*) G is not overall divergent; then we must prove (7.2.9) with no counterterm $\Delta(G)$. (*b*) G is 1PI and overall divergent; then we use (7.2.10) with γ replaced by G to define $\Delta(G)$. We must prove $\Delta(G)$ finite and prove (7.2.9). We must also prove $\Delta(G)$ is polynomial in the external momenta of G, with degree equal to the degree of divergence

of G; we will assume inductively that this is true with G replaced by any smaller graph.

Consider the terms in (7.2.7). For each we identify a contribution in (7.2.9). The term G is the same as G in $R'(G)$. Next let γ be a 1PI subgraph of G (G itself being excluded). Decompose $C_\gamma(G)$ by (7.2.10):

(1) The $C'(\gamma)$ term occurs in $R'(G)$ as $C'_\gamma(G)$.
(2) The term $\Delta(\gamma)$ occurs as $\Delta_\gamma(G)$.
(3) The term $C'_\gamma(\Delta_\delta(\gamma))$ occurs as a counterterm in the renormalization $R'(\Delta_\delta(G))$.

If G is not overall divergent, these exhaust all of the terms in $R(G)$ and on the right-hand side of (7.2.9). But if G is 1PI and overall divergent then there remains $C(G)$ in (7.2.7) and the terms

$$C'\left[G + \sum_{\delta \subsetneq G} \Delta_\delta(G)\right] + \Delta(G)$$

in (7.2.9). We are therefore forced to define $\Delta(G)$ by (7.2.10), and it remains to prove $\Delta(G)$ finite. This is now easy, since

$$R(G), R'(G), \text{ and } R'\left[\sum_{\delta \subsetneq G} \Delta_\delta(G)\right] \tag{7.2.11}$$

are all finite, while we have

$$R'(\Delta(G)) = \Delta(G).$$

Moreover all the terms in (7.2.10), except possibly $\Delta(G)$, are ordinary minimal-subtraction counterterms. So they are polynomial in the external momenta of G, with degree equal to the degree of divergence of G. So $\Delta(G)$ is polynomial, of the same degree. (Note that the replacement of a subgraph γ by Δ_γ does not change the overall degree of divergence of any graph Γ satisfying $\gamma \subseteq \Gamma \subseteq G$. This is because of our inductive assumption on the polynomial degree of Δ_γ.)

The theorem is also true if R and R' stand not for renormalization with different units of mass, but for any two renormalization prescriptions. It is important that the d-dependence of Δ_γ is taken outside of the extraction of pole parts when computing a counterterm like $C_\Gamma(\Delta_\gamma)$. This can be seen from the example of Fig. 7.1.1, at (7.1.31). The primed counterterms must be a particular function of g_{new}, μ', m', and d. The fact that g' itself is a function of other variables is ignored.

Equation (7.2.10) expresses the counterterm $C(\gamma)$ for a graph γ (with unit of mass μ) in terms of renormalization counterterms with unit of mass μ' and a finite counterterm $\Delta(\gamma)$. We can use the finite counterterms to generate

the compensating Lagrangian $\mathscr{L}'_c$ in (7.2.3), and the primed counterterms to generate $\mathscr{L}'_{ct}$. It should be evident that the new Lagrangian is the same as the original one (with unit of mass μ), considered as a function of ϕ and $\partial\phi$. If we set $\phi_0 = Z^{1/2}\phi$, with $Z = Z(g, m, \mu)$, then we can deduce that the bare parameters m_0 and g_0 are renormalization-group invariant:

$$m_0(g, m, \mu) = m_0(g', m', \mu'),$$
$$g_0(g, m, \mu) = g_0(g', m', \mu'). \tag{7.2.12}$$

7.3 Renormalization-group equation

We saw in Sections 7.1 and 7.2 that a change in the unit of mass μ accompanied by suitable changes in coupling and mass does not change the theory, while the Green's functions satisfy

$$G_N(x_1, \ldots, x_N; g, m, \mu) = \zeta^{-N} G_N(x_1, \ldots, x_N; g', m', \mu'). \tag{7.3.1}$$

We wish to compute g', m'^2 and ζ as functions of g, m^2, μ, and μ'. If the ratio μ/μ' is large, it is not sufficient to use lowest-order perturbation theory, since, for example, in (7.1.26) the coefficient of g^2 may be large.

An important device is to consider a large change in μ as being made up of a sequence of very small changes, so that g', m'^2, and ζ are obtained as solutions of differential equations. This is the subject to which we now turn.

The consequence of our work in Sections 7.1 and 7.2 is that for a given physical theory, we have for each value of μ a definite value of the coupling $g(\mu)$ and mass $m(\mu)$. These are called the effective (or running) coupling and mass. We will derive differential equations for $g(\mu)$ and $m(\mu)$.

The easiest way to derive the results is to look at the Green's functions and the Lagrangian expressed in terms of the bare field ϕ_0. The important point is that the Green's functions of ϕ_0 are invariant under our change of parametrization $(\mu, g, m) \to (\mu', g', m')$. (This is because the mass and coupling in the bare Lagrangian are invariant.)

7.3.1 *Renormalization-group coefficients*

Physical quantities like the S-matrix are invariant under the change of variable $(\mu, g(\mu), m(\mu)) \to (\mu', g(\mu'), m(\mu'))$. This invariance is conveniently expressed by considering a small change in μ, accompanied by the corresponding changes in g and m. We write the result as

$$\mu \mathrm{d}S/\mathrm{d}\mu = 0. \tag{7.3.2}$$

The total derivative with respect to μ can be written as

$$\mu \mathrm{d}/\mathrm{d}\mu = \mu\partial/\partial\mu + \beta\partial/\partial g - \gamma_m m^2 \partial/\partial m^2. \tag{7.3.3}$$

On the right-hand side the partial derivatives with respect to μ, g or m are taken with the other two fixed, and the coefficients β and γ_m give the variations of $g(\mu)$ and $m^2(\mu)$ when μ is varied. The sign of the $m^2\partial/\partial m^2$ term is the usual convention (Weinberg (1973)). We have

$$\beta = \mu \mathrm{d}g(\mu)/\mathrm{d}\mu, \tag{7.3.4a}$$

$$\gamma_m = -m^2 \mu \mathrm{d}m^2(\mu)/\mathrm{d}\mu. \tag{7.3.4b}$$

The coefficients β and γ_m are called renormalization-group coefficients. As we will see they are easy to calculate in terms of the counterterms, as functions of g, m, μ, and d. If we use minimal subtraction, they have no mass dependence. This means that (7.3.4) can be readily solved as differential equations for $g(\mu)$ and $m(\mu)$. Indeed this is the easiest way in practice to compute the effective coupling and mass.

To compute β and γ_m it is convenient to consider the Lagrangian expressed in terms of the bare field ϕ_0 – see (7.1.1*a*). We saw in Section 7.2 that m_0^2 and g_0 are renormalization-group invariant:

$$\begin{aligned} \mu \mathrm{d}g_0/\mathrm{d}\mu &= 0, \\ \mu \mathrm{d}m_0^2/\mathrm{d}\mu &= 0. \end{aligned} \tag{7.3.5}$$

Suppose we have computed g_0 and m_0 to some order in g. Then (7.3.5) can be solved to give β and γ_m.

An example of such a calculation comes from our results on ϕ^3 theory, where

$$\begin{aligned} g_0 &= \mu^{3-d/2} g\left[1 + \frac{3g^2}{256\pi^3(d-6)} + O(g^4)\right], \\ m_0^2 &= m^2\left[1 + \frac{5g^2}{384\pi^3(d-6)} + O(g^4)\right]. \end{aligned} \tag{7.3.6}$$

Thus

$$0 = \mu \frac{\mathrm{d}g_0}{\mathrm{d}\mu} = \mu^{3-d/2}\left\{g(3-d/2) - \frac{3g^3}{512\pi^3} + O(g^5) + \beta(g)\left[1 + \frac{9g^2}{256\pi^3(d-6)} + O(g^4)\right]\right\},$$

so

$$\begin{aligned} \beta(g) &= (d/2-3)g - \frac{3g^3}{256\pi^3} + O(g^5) \\ &\to -\frac{3g^3}{256\pi^3} + O(g^5) \text{ at } d = 6. \end{aligned} \tag{7.3.7}$$

Similarly

$$0 = \mu \frac{\mathrm{d}m_0^2}{\mathrm{d}\mu}$$

$$= -m^2\gamma_m[1 + O(g^2)] + m^2\beta(g)\left[\frac{5g}{192\pi^3(d-6)} + O(g^3)\right],$$

so that

$$\gamma_m = \frac{5}{6}\frac{g^2}{64\pi^3} + O(g^4). \tag{7.3.8}$$

Observe that the $(d/2-3)g$ term in β is important in these derivations, even though the term disappears at $d=6$. Observe also that, even though the coefficients in g_0 and m_0 diverge at $d=6$, it is crucial to expand strictly in powers of g. A phenomenon true to all orders is that β and γ_m are independent of m and μ, provided that we renormalize by minimal subtraction.

The general calculation of β and γ_m can be organized with the aid of (7.1.4) for the counterterms. Since we use minimal subtraction, the m- and μ-dependence of the bare parameters is simple:

$$g_0 = \mu^{3-d/2}\bar{g}_0(g,d),$$
$$m_0^2 = m^2 Z_m(g,d). \tag{7.3.9}$$

Then

$$\beta(g,d) = (d/2-3)\bar{g}_0 \Big/ \frac{\partial \bar{g}_0}{\partial g},$$

$$\gamma_m(g) = \beta(g,d)\frac{\partial \ln Z_m}{\partial g}. \tag{7.3.10}$$

The expressions (7.3.9) are to be expanded in powers of g with the aid of (7.1.4). Now

$$g_0 = (\mu^{3-d/2}g + \delta g)Z^{-3/2}$$
$$= \mu^{3-d/2}\left[g + \frac{a_1(g) - \frac{3}{2}c_1(g)}{6-d} + \text{higher poles}\right],$$

$$Z_m = (1 + \delta m^2/m^2)Z^{-1}$$
$$= \left[1 + \frac{b_1(g) - c_1(g)}{6-d} + \text{higher poles}\right], \tag{7.3.11}$$

where we have picked out the single poles in the series expansion of the counterterms (7.1.4). (These are all that will be relevant.) Then (7.3.10)

becomes

$$\beta(g,d)=\left(\frac{d}{2}-3\right)\frac{\left[g+\dfrac{a_1(g)-\frac{3}{2}c_1(g)}{6-d}+\text{higher poles}\right]}{\left[1+\dfrac{a_1'(g)-\frac{3}{2}c_1'(g)}{6-d}+\text{higher poles}\right]}$$

$$=(d/2-3)g+\tfrac{1}{2}\left(1-g\frac{\partial}{\partial g}\right)[\tfrac{3}{2}c_1(g)-a_1(g)]+\text{poles}$$

$$\equiv(d/2-3)g+\bar{\beta}(g), \tag{7.3.12a}$$

$$\gamma_m(g)=[(d/2-3)+\bar{\beta}(g)]\frac{\partial}{\partial g}\left[\frac{b_1(g)-c_1(g)}{6-d}+\text{higher poles}\right]$$

$$=\tfrac{1}{2}g\frac{\partial}{\partial g}[c_1(g)-b_1(g)]. \tag{7.3.12b}$$

These manipulations are made by expanding in powers of g.

In the last line of each of (7.3.12) we have used the fact that although pole terms are in principle present, they must cancel in order that β and γ_m be finite as $d\to 6$. Notice that β and γ_m are independent of m and μ, and that the only d-dependence is the $(d/2-3)g$ term in β. Only the single-pole terms are needed for the calculation. There is a series of relations between these and the higher poles that ensures that the poles cancel in (7.3.12); we will investigate these later.

7.3.2 *RG equation*

The RG coefficients β and γ_m are computed from two out of three combinations of the counterterms. The differential equations (7.3.4) then enable g' and m' in (7.3.1) to be computed. To complete the calculation we need ζ. This is related to the wave-function renormalization. It is easiest to obtain by observing that the bare Green's functions $G_N^{(0)}=Z^{N/2}G_N$ are renormalization-group invariant, so that

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}G_N=\mu\frac{\mathrm{d}}{\mathrm{d}\mu}(G_N^{(0)}Z^{-N/2})$$

$$=-\frac{N}{2}\mu\frac{\mathrm{d}}{\mathrm{d}\mu}(\ln Z)G_N. \tag{7.3.13}$$

Let us define the finite coefficient

$$\gamma=\frac{\mathrm{d}}{\mathrm{d}\mu}\ln Z; \tag{7.3.14}$$

then G_N satisfies the following renormalization group equation (e.g. Weinberg (1973)):

$$\left[\mu\frac{\mathrm{d}}{\mathrm{d}\mu}+\frac{N}{2}\gamma\right]G_N=\left[\mu\frac{\mathrm{d}}{\mathrm{d}\mu}+\beta\frac{\partial}{\partial g}-\gamma_m m^2\frac{\partial}{\partial m^2}+\frac{N}{2}\gamma\right]G_N=0. \quad (7.3.15)$$

In the minimal subtraction scheme for ϕ^3 we find

$$\begin{aligned}\gamma&=[(d/2-3)g+\bar{\beta}(g)]\frac{\partial}{\partial g}\left[\frac{c_1(g)}{6-d}+\text{higher poles}\right]\\&=-\tfrac{1}{2}g\frac{\mathrm{d}}{\mathrm{d}g}c_1(g)\\&=\frac{g^2}{384\pi^3}+O(g^4).\end{aligned} \quad (7.3.16)$$

7.3.3 *Solution*

We wish to solve the RG equations to find $g(\mu')$, $m(\mu')$ and $\zeta(\mu',\mu)$ in (7.3.1), given that $g(\mu)=g$ and $m(\mu)=m$. The RG equation tells us that

$$\begin{aligned}\mu'\frac{\mathrm{d}}{\mathrm{d}\mu'}\ln\zeta&=\frac{1}{N}\mu'\frac{\mathrm{d}}{\mathrm{d}\mu'}\ln[G_N(\mu')/G_N(\mu)]\\&=-\tfrac{1}{2}\gamma[g(\mu')].\end{aligned} \quad (7.3.17)$$

So we must solve this equation together with (7.3.4) for g and m. The boundary conditions are

$$\begin{aligned}g(\mu)&=g,\\m(\mu)&=m,\\\zeta(\mu,\mu)&=1.\end{aligned} \quad (7.3.18)$$

Explicit solutions can be written:

$$\begin{aligned}\ln(\mu'/\mu)&=\int_{g(\mu)}^{g(\mu')}\frac{\mathrm{d}g'}{\beta(g')},\\m^2(\mu')&=m^2(\mu)\exp\left\{-\int_\mu^{\mu'}\frac{\mathrm{d}\bar{\mu}}{\bar{\mu}}\gamma_m(g(\bar{\mu}))\right\}\\&=m^2(\mu)\exp\left\{-\int_{g(\mu)}^{g(\mu')}\mathrm{d}g'\frac{\gamma_m(g')}{\beta(g')}\right\},\\\zeta(\mu',\mu)&=\exp\left\{-\tfrac{1}{2}\int_\mu^{\mu'}\frac{\mathrm{d}\bar{\mu}}{\bar{\mu}}\gamma(g(\bar{\mu}))\right\}\\&=\exp\left\{-\tfrac{1}{2}\int_{g(\mu)}^{g(\mu')}\mathrm{d}g'\frac{\gamma(g')}{\beta(g')}\right\}.\end{aligned} \quad (7.3.19)$$

Approximations can be made by taking a finite number of terms in the perturbation series for β, γ_m and γ. For example:

$$\ln(\mu'/\mu) = -\frac{256\pi^3}{3}\int_{g(\mu)}^{g(\mu')}\frac{dg}{g^3}[1+O(g^2)]$$
$$= \frac{128\pi^3}{3}\left[\frac{1}{g(\mu')^2}-\frac{1}{g(\mu)^2}\right]+O\left[\ln\left(\frac{g(\mu')}{g(\mu)}\right)\right]. \quad (7.3.20)$$

This is accurate if $g(\mu')$ and $g(\mu)$ are small. Notice that $g(\mu')\to 0$ as $\mu'\to\infty$. This is the property called asymptotic freedom. It is determined by the negative sign of the first term in β at $d=6$.

The full solution to the RG equation is

$$G_N(x;g,m,\mu) = \exp\left[-\frac{N}{2}\int_{\mu'}^{\mu}\frac{d\bar{\mu}}{\bar{\mu}}\gamma(g(\bar{\mu}))\right]G_N(x;g(\mu'),m(\mu'),\mu'). \quad (7.3.21)$$

7.4 Large-momentum behavior of Green's functions

The most important application of the renormalization group is to compute large-momentum behavior. In this section we treat the simplest case, that of a Green's function $\tilde{G}_N(p_1,\ldots,p_N)$ all of whose external momenta are made large. (Notice that we have used our standard notation, where the tilde indicates Fourier transformation into momentum space.)

Let us suppose initially that all the Lorentz invariants formed from the momenta are non-zero. Then we scale all the momenta by a factor κ: $p_i\to\kappa p_i$, and let κ get large. Thus all the Lorentz invariants $p_i\cdot p_j$ are scaled by a factor κ^2 and become large. Under these conditions, Weinberg's theorem tells us that at least in a renormalizable theory all internal lines of graphs for $\tilde{G}_N$ carry large momenta, and that graphs for $\tilde{G}_N$ have the asymptotic behavior

$$\kappa^{\dim \tilde{G}}(\text{logarithms of } \kappa). \quad (7.4.1)$$

Since all propagator denominators are large, we should be able to neglect masses and make only an error a power of κ smaller than the leading behavior (7.4.1).

For example consider the propagator in ϕ^3 theory at $d=6$. The tree graph goes like i/p^2 at large p^2, while the one-loop correction is

$$\frac{i}{p^2}\frac{g^2}{128\pi^3}\left\{\tfrac{1}{6}(\gamma-1)+\int_0^1 dx\,x(1-x)\ln\left[\frac{-p^2x(1-x)}{4\pi\mu^2}\right]\right.$$
$$\left.+O\left[\frac{m^2}{p^2}\ln\left(\frac{m^2}{p^2}\right)\right]\right\}. \quad (7.4.2)$$

(We used (3.6.10) for the self-energy graph to derive this equation.)

Now there is a term proportional to $\ln(-p^2/\mu^2)/p^2$ that gets large relative to the tree graph if p^2 is large enough. Thus the perturbation series has large coefficients and is not directly useful. However we may use the renormalization group to set $\mu^2 = O(p^2)$. This makes the coefficient small again. So we use the following strategy to compute $\tilde{G}_N(\kappa p)$ at large κ:

(1) Set $\mu' = \kappa\mu$, and use the solution of the RG equation to write

$$\tilde{G}_N(\kappa p, g, m, \mu) = \zeta(\kappa\mu, \mu)^{-N/2}\tilde{G}_N(\kappa p, g(\kappa\mu), m(\kappa\mu), \kappa\mu). \tag{7.4.3}$$

(2) Neglect m (if $m(\kappa\mu)$ does not get too large). Then use dimensional analysis to give

$$\begin{aligned}\tilde{G}_N(\kappa p, g, m, \mu) &\sim \zeta^{-N/2}\tilde{G}_N(\kappa p, g(\kappa\mu), 0, \kappa\mu)\\ &= \kappa^{\dim(\tilde{G}_N)}\zeta(\kappa\mu, \mu)^{-N/2}\tilde{G}_N(p, g(\kappa\mu), 0, \mu).\end{aligned} \tag{7.4.4}$$

(3) Large κ-dependent coefficients, as in (7.4.2), are now removed, so if $g(\kappa\mu)$ is small, a low-order calculation suffices.

This procedure makes it evident that the coupling that is relevant is the effective coupling at the scale of the momenta involved.

It should be noted that we have related the large-momentum behavior of $\tilde{G}_N$ to the finite-momentum behavior of the zero-mass theory. It is therefore crucial that the zero-mass limit exists. However this limit does not always exist: if we use mass-shell or zero-momentum subtractions, then we see from, for example, (3.4.7), that the same self-energy as considered in (7.4.2) diverges as $m \to 0$. Now Weinberg's theorem tells us that, in the dominant momentum region for a graph without counterterms, all lines are far off-shell; hence masses can be neglected. So the problem must be that with mass-shell or zero-momentum renormalization prescriptions, the counterterms diverge as $m \to 0$. This is easily checked from our explicit calculations (see (5.10.2)).

We can now see the practical importance of the theorem whose proof was summarized at the end of Section 5.8, that the counterterms may be *chosen* polynomial in masses. It ensures that the zero-mass limit may be taken directly and used to compute large momentum behavior. The minimal subtraction scheme is one way of ensuring that counterterms are polynomial in mass.

If one uses, say, zero-momentum subtractions, large-momentum behavior may be computed by changing renormalization prescription to, say, minimal subtraction. Another approach is to observe that the logarithms of p^2 break a possible symmetry of the theory under scaling transformations. The consequences of this point of view were worked out by Callan (1970)

and Symanzik (1970b). They derived the Ward identity for scaling transformations. It is called the Callan–Symanzik equation and looks similar to the RG equation. This equation may also be used to discuss high-energy behavior.

7.4.1 *Generalizations*

The behavior of $\tilde{G}_N$ when all momenta are scaled by a large factor κ is not normally experimentally relevant, for all the external momenta are then far off-shell. In coordinate space the corresponding region is the short-distance limit of $G_N(x_1, \ldots x_N)$, where every $x_j - x_k$ is made small: $x_j - x_k \to (x_j - x_k)/\kappa$.

This means that we should be able to use RG methods to discuss the renormalization of the theory, for renormalization is a purely short-distance phenomenon. We will work out the details in Section 7.10. On the other hand, physical experiments involve long distances. To get results for high-energy experiments we need the so-called factorization theorems. The simplest of these is the operator-product expansion which we will treat in Chapter 10. These theorems typically give a cross-section as a product of a factor which can be computed by pure short-distance methods and of simple factors related to wave-functions of the incoming and/or outgoing particles.

We could also use RG methods to discuss the infra-red limit $\kappa \to 0$. This is only useful if masses can be neglected. Certainly this is true in a purely massless theory, as we will see in Section 7.5.4, and then IR behavior is computable if and only if the theory is not asymptotically free. But in a massive theory, it is not useful to take μ much less than a typical mass, for one obtains logarithms of m/μ, and these prevent a simple use of perturbative methods when $\mu \ll m$.

7.5 Varieties of high- and low-energy behavior

7.5.1 *Asymptotic freedom*

In solving the RG equation to obtain high-energy behavior, we find two cases according to whether β is positive or negative. In this section we discuss the asymptotically free case, when β is negative. Suppose the effective coupling $g(\mu)$ is small for one value of μ, so that β is well approximated by its first term. Then the evolution equation (7.3.4*a*) shows that g becomes even smaller at larger values of μ and in fact goes to zero as $\mu \to \infty$. Thus perturbation theory is reliable for computing high-energy

behavior. From (7.3.7) we see that ϕ^3 in six dimensions is asymptotically free.

It is instructive to compute the behavior of g from the first two terms in β. Let us define

$$\beta = -A_1 g^3 - A_2 g^5 + O(g^7). \tag{7.5.1}$$

We have the equation for the evolution of the effective coupling:

$$\mu \frac{\mathrm{d}g}{\mathrm{d}\mu} = \beta(g). \tag{7.5.2}$$

Equations of the same form as (7.5.1) and (7.5.2) hold in any renormalizable theory, for example in QCD, though A_1 and A_2 are not necessarily positive in the general case.

The solution of (7.5.2) is

$$\begin{aligned} \ln \mu &= \text{constant} + \int^{g(\mu)} \mathrm{d}g'/\beta(g') \\ &= \text{constant} + \int^{g(\mu)} \mathrm{d}g' \left[\frac{-1}{A_1 g'^3} + \frac{A_2}{A_1^2 g'} + O(g') \right], \\ &= \text{constant} + \frac{1}{2A_1 g(\mu)^2} + \frac{A_2}{A_1^2} \ln [g(\mu)] + O(g^2). \end{aligned} \tag{7.5.3}$$

The constant can be computed from a knowledge of $g(\mu)$ at one value of μ. It is conventional (Buras, Floratos, Ross & Sachrajda (1977)) to write the constant in the form $\ln \Lambda + \frac{1}{2} A_2 A_1^{-2} \ln (A_1)$, where Λ is a parameter with the dimensions of mass. Then

$$\ln (\mu^2/\Lambda^2) = \frac{1}{A_1 g(\mu)^2} + \frac{A_2}{A_1^2} \ln [A_1 g(\mu)^2] + O(g(\mu)^2), \tag{7.5.4}$$

so that

$$g(\mu)^2 = \frac{1}{A_1 \ln (\mu^2/\Lambda^2)} - \frac{A_2 \ln [\ln (\mu^2/\Lambda^2)]}{A_1^3 \ln^2 (\mu^2/\Lambda^2)} + O\left\{ \frac{\ln^2 [\ln (\mu/\Lambda)]}{\ln^3 (\mu/\Lambda)} \right\}. \tag{7.5.5}$$

A specification of $g(\mu)$ at one value of μ is exactly equivalent to a specification of Λ. The precise choice of the scale of Λ is that in (7.5.5) the omitted terms are of the order shown rather than of order $1/\ln^2(\mu/\Lambda)$. The expansion (7.5.5) is much used in QCD. A higher-order calculation of g can be made from the following form of the solution

$$\begin{aligned} \ln (\mu^2/\Lambda^2) = &\frac{1}{A_1 g(\mu)^2} + \frac{A_2}{A_1^2} \ln [A_1 g(\mu)^2] \\ &+ 2 \int_0^{g(\mu)} \mathrm{d}g' \left[\frac{1}{\beta(g')} + \frac{1}{A_1 g'^3} - \frac{A_2}{A_1^2 g'} \right]. \end{aligned} \tag{7.5.6}$$

It is necessary to go to two-loop order to obtain both the terms on the right of (7.5.6) that diverge as $g \to 0$.

The values of $m(\mu)$ and ζ may be similarly calculated. For example, if

$$\gamma = C_1 g^2 + \cdots, \tag{7.5.7}$$

then

$$\begin{aligned}\zeta(\mu', \mu) &= \exp\left[\frac{1}{2}\int_{g(\mu)}^{g(\mu')} \mathrm{d}g \, \frac{C_1}{g} \frac{}{A_1} + O(g^2)\right] \\ &= \exp\left[\frac{C_1}{2A_1} \ln \frac{g(\mu')}{g(\mu)} + O(g^2)\right] \\ &= \left[\frac{g(\mu')}{g(\mu)}\right]^{C_1/2A_1} [1 + O(g^2)] \\ &\propto [\ln(\mu'/\mu)]^{-C_1/4A_1}, \quad \text{as } \mu' \to \infty. \end{aligned} \tag{7.5.8}$$

7.5.2 *Maximum accuracy in an asymptotically free theory*

The results above enable calculations of Green's functions to be made at high energy. By taking more and more terms in the series, we may improve the predictions. However, in general, perturbation series are asymptotic, not convergent. A trivial example is the ordinary integral

$$I(g, m) = m(2\pi)^{-1/2} \int_{-\infty}^{\infty} \mathrm{d}z \exp(-m^2 z^2/2 - g z^4/4!)$$

This can be considered to be a functional integral in Euclidean ϕ^4 field theory at zero space-time dimension with the normalization chosen to give $I = 1$ when $g = 0$. The perturbation expansion is

$$\begin{aligned} I &\sim m(2\pi)^{-1/2} \sum_{N=0}^{\infty} \int_{-\infty}^{\infty} \mathrm{d}z \exp(-m^2 z^2/2) \left(\frac{-g z^4}{4!}\right)^N \frac{1}{N!} \\ &= \sum_{N=0}^{\infty} \left(\frac{-g}{6m^4}\right)^N \pi^{-1/2} \frac{\Gamma(2N + 1/2)}{\Gamma(N+1)} \\ &\equiv \sum_{N=0}^{\infty} \left(\frac{g}{m^4}\right)^N I_N. \end{aligned} \tag{7.5.9}$$

Now

$$I_N = N^N \left(\frac{-2}{3\mathrm{e}}\right)^N (\pi N)^{-1/2} [1 + O(1/N)], \quad \text{as } N \to \infty, \tag{7.5.10}$$

so the series is divergent. The divergence is associated with the fact that the defining integral diverges if g is negative, so that I is not analytic at $g = 0$. The corresponding property of ϕ^4 theory is that the Hamiltonian is unbounded below (i.e., the vacuum does not exist) if g is negative.

What we can say is that we can approximate I by truncating the series:

$$I = \sum_{N=0}^{N_m} (g/m^4)^N I_N + O[(g/m^4)^{N+1}]. \tag{7.5.11}$$

The error is estimated by the first term omitted, i.e.,

$$O[I_{N_m+1}(g/m^4)^{N_m+1}].$$

These results are standard in the theory of asymptotic expansions for simple integrals. All experience, together with rigorous theorems for quantum mechanics and super-renormalizable field theories (Glimm & Jaffe (1981) and references therein), indicate that this behavior is typical for functional integrals in a non-trivial dimension (i.e., $d > 0$).

Next, let us suppose we wish to compute some quantity in an asymptotically free theory by truncating its perturbation expansion

$$\sum_{N=0}^{N_m} g(\mu)^{2N} I_N.$$

We assume that the quantity depends on some momentum p, and that we set the unit of mass μ to be of order p. A case would be the propagator with $\mu = O(|p|)$. Suppose that the coefficients in the expansion behave like

$$I_N \sim N^N b^N N^a d[1 + O(1/N)], \tag{7.5.12}$$

for large N. What is the best accuracy with which we can calculate the quantity? This is given by the minimum error, i.e., the minimum of $I_N g^{2N}$ as N varies. The result is that the minimum possible error in a perturbative calculation is of order

$$\text{constant}\,|p|^{-2A_1/eb}(\ln|p|)^a.$$

This means that beyond a certain level, power-law corrections to the asymptotic behavior computed in perturbation theory are meaningless since they are smaller than the irreducible error in using perturbation theory. Power-law corrections are those that are a power of p^2 smaller than the leading term.

7.5.3 *Fixed point theories*

In four dimensions, the only theories that are asymptotically free are non-abelian gauge theories with a small enough number of matter fields – see Coleman & Gross (1973) and Gross (1976). Other theories, like ϕ^4 and QED, have an effective coupling that increases with energy. Thus, in such theories it is impossible to compute the true high-energy behavior by perturbation theory. (Note however that the coupling in QED is α/π

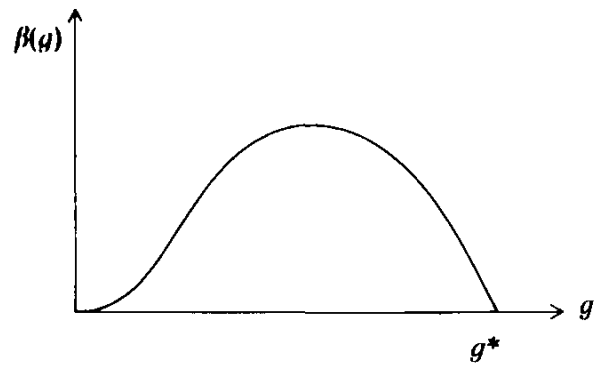


Fig. 7.5.1. $\beta(g)$ in a non-asymptotically free theory with a fixed point at $g = g^*$.

$\sim 1/430$. This is very small, so the non-perturbative region in QED does not occur until very many orders of magnitude beyond experimentally accessible energies.)

An interesting possibility is that $\beta(g)$ has the form shown in Fig. 7.5.1, with a zero at $g = g^*$. Then $g(\mu)$ approaches the 'fixed point' g^* as $\mu \to \infty$. At large momentum Green's functions behave like

$$\tilde{G}_N(\kappa p_1, \ldots, \kappa p_N, g, m, \mu) \sim \text{const}\, \kappa^{\dim \tilde{G}_N + N\gamma(g^*)/2} \tilde{G}_N(p_1, \ldots, p_N, g^*, 0, \mu). \tag{7.5.13}$$

This behavior is as if ϕ had an extra term $\gamma(g^*)/2$ in its dimension. Consequently, the function $\gamma(g)/2$ is called the anomalous dimension of the field ϕ.

7.5.4 *Low-energy behavior of massless theory*

If $m = 0$, then the renormalization group can be used to compute infra-red behavior. The calculability is the opposite of that for the UV behavior. Consider first an asymptotically free theory. There, the effective coupling $g(\mu)$ goes to zero when μ goes to infinity, so that short-distance behavior is computable perturbatively, as we saw in Section 7.5.1. But, when μ is small, $g(\mu)$ is large, so the infra-red behavior cannot be computed reliably by perturbation theory. (This is the case for strong interactions, according to QCD.)

Let us now consider a non-asymptotically free theory. For large μ, the effective coupling is large, so the short-distance behavior is not perturbatively computable. (For example, a perturbative calculation in low order of the position of the fixed point, g^*, in Fig. 7.5.1 and of the value of $\beta(g^*)$ is subject to large errors from higher-order uncalculated corrections.) But when μ goes to zero, so does the effective coupling. We are assuming the absence of a mass term for the field, so there are no large logarithms of m/μ as μ goes to zero. Hence, we can compute IR behavior in such a theory, just

as we computed UV behavior in an asymptotically free theory. We will now do this.

Now, for almost any graph there are large logarithms of p^2/μ^2 as $p^2 \to 0$ in a massless theory, just as in the ultra-violet. In the case of the propagator, these logarithms mean that the propagator's singularity is not a pole, at least order-by-order.

To investigate this singularity let us again write

$$\gamma(g) = C_1 g^2, \quad \beta(g) = -A_1 g^3, \tag{7.5.14}$$

using the same notation as before, but now with $A_1 < 0$. We assume that g is below the first non-zero fixed point g^*, if there is one. The propagator is

$$\begin{aligned}\tilde{G}_2(\kappa p; g(\mu), \mu) &= \kappa^{-2} \tilde{G}_2(p; g(\kappa\mu), \mu) \exp \int_{g(\mu)}^{g(\kappa\mu)} \mathrm{d}g \gamma(g)/\beta(g) \\ &= \mathrm{i}/(\kappa^2 p^2)[1 + O(g(\kappa\mu)^2)] \exp\left\{ -\int \mathrm{d}g (C_1/A_1 g)[1 + O(g^2)] \right\} \\ &\sim \mathrm{i}/(\kappa^2 p^2) \cdot \text{constant} \cdot [\ln(1/\kappa)]^{C_1/2A_1},\end{aligned} \tag{7.5.15}$$

as $\kappa \to 0$. Hence if C_1 is non-zero, then $p^2 \tilde{G}_2(p)$ does not have a finite non-zero limit as $p \to 0$; the singularity of $\tilde{G}_2(p^2)$ at $p^2 = 0$ does not correspond to a *simple* single-particle pole. The massless particle that gives rise to the singularity cannot be treated as an ordinary particle, because its long-range interactions are too strong. Positivity of the metric of the state vectors constrains C_1 to be positive, and we assumed a theory with A_1 negative, so κ^2 times the right-hand side of (7.5.15) goes to zero. (The positivity argument is the one given in the textbooks (e.g., Bjorken & Drell (1966)) that the residue of the pole in a propagator is less than unity if the propagator is of the canonical field. Application of this argument in the theory with an ultra-violet cut-off shows that the divergence of the self-energy must be such that C_1 is positive.)

Notice that if $C_1 = 0$, then the propagator does have a finite residue at $p = 0$:

$$-\mathrm{i}\kappa^2 p^2 \tilde{G}_2(\kappa p, g(\mu), \mu) \to \exp\left\{ -\int_0^{g(\mu)} \mathrm{d}g \gamma(g)/\beta(g) \right\}. \tag{7.5.16}$$

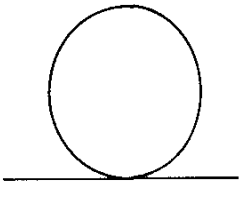


Fig. 7.5.2. Lowest-order self-energy graph in ϕ^4 theory.

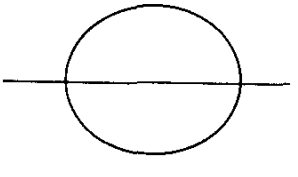


Fig. 7.5.3. Lowest-order self-energy graph that contributes to the anomalous dimension in ϕ^4 theory.

Observe that only the first term in γ is relevant to the finiteness of the limit, and that the limit does not exist order-by-order. A case where $C_1 = 0$ is the ϕ^4 theory, because the one-loop graph Fig. 7.5.2 is independent of p. The two-loop graph Fig. 7.5.3 provides the lowest-order term in the anomalous dimension.

7.6 Leading logarithms, etc.

7.6.1 *Renormalization-group logarithms*

We saw in Section 7.4 how to compute the large momentum behavior of a Green's function by approximating it by a Green's function with $m = 0$. Then we used the renormalization group to reorganize the perturbation series into a form with small coefficients. It is of interest to examine how the complete result can be obtained by a systematic resummation of the perturbation expansion.

For concreteness, let us examine the propagator $\tilde{G}_2(p, g, \mu)$ in massless ϕ^3 theory in six dimensions. We write its perturbation expansion as

$$\tilde{G}_2 = (\mathrm{i}/p^2) \sum_{n=0}^{\infty} g^{2n} T_n(-p^2/\mu^2), \tag{7.6.1}$$

where the lowest coefficient is $T_0 = 1$. We will prove that each T_n is a polynomial in $\ln \mu$ (and hence in $\ln(-p^2/\mu^2)$) of degree at most n, with n being the number of loops. To do this we will regard the RG equation (7.3.15) not as an equation to give the variation of $\tilde{G}_2$ when μ is changed with g set equal to the effective coupling $g(\mu)$ (thus keeping the theory fixed), but as an equation for the μ-dependence of $\tilde{G}_2$ with g fixed. Picking out the order g^{2n} term gives

$$\frac{\partial}{\partial \ln \mu} T_n = \left\{ \left[-\gamma(g) - \beta(g) \frac{\partial}{\partial g} \right] \sum_{n'=0}^{n-1} T_{n'} g^{2n'} \right\}_{\text{coefficient of } g^{2n}}. \tag{7.6.2}$$

Since γ is $O(g^2)$ and β is $O(g^3)$, this equation determines T_n in terms of lower-order $T_{n'}$'s:

$$T_n = \text{constant} + \left\{ \sum_{n'<n} -\left[\gamma(g) + \beta(g) \frac{\partial}{\partial g} \right] g^{2n'} \int_0^{\ln \mu} \mathrm{d} \ln \mu' \, T_{n'}(-p^2/\mu'^2) \right\}_{\text{coefficient of } g^{2n}}. \tag{7.6.3}$$

Iteration of this procedure another $n-1$ times gives T_n in terms of T_0 and n constants of integration. Evidently T_n is a polynomial of degree n in $\ln \mu$ as

claimed:

$$T_n = \sum_{l=0}^{n} T_{n,n-l}[\ln(-p^2/\mu^2)]^l. \tag{7.6.4}$$

All but the constant term are determined in terms of lower-order coefficients.

A convenient way of organizing the series is to define for each term

$$L = -\text{number of logarithms} + \text{number of powers of } g^2 = -l + n. \tag{7.6.5}$$

This is non-negative. The sum of the terms with $L=0$ (viz., $T_{n0}[\ln(-p^2/\mu^2)]^n$) gives what is called the leading logarithm approximation to $\tilde{G}_2$. Application of $\mu\partial/\partial\mu$, $\beta\partial/\partial g$ or γ to $\tilde{G}_2$ (with one-loop values for β and γ) increases L by 1. All the non-leading logarithms give even higher values of L. So the leading logarithm series exactly satisfies the one-loop approximation to the RG equation. Hence we may sum the leading logarithm series by solving this approximation to the RG equation

$$\begin{aligned}\tilde{G}_2(p^2;g,\mu)_{L=0} &= \left\{\exp\left[\int_{\mu}^{\sqrt{-p^2}} \frac{d\mu'}{\mu'}\gamma(g(\mu'))\right]\tilde{G}_2(p^2;g(\sqrt{-p^2}),\sqrt{-p^2})\right\}_{L=0} \\ &= [1 + A_1 g^2 \ln(-p^2/\mu^2)]^{C_1/2A_1} i/p^2. \end{aligned} \tag{7.6.6}$$

Here we used the same notation as in Section 7.5.1. This equation reproduces the approximation derived at (7.5.8).

Another way of treating both the leading and non-leading logarithms is to use the RG equation (7.6.2) to give a recursion relation for the $T_{n,L}$'s:

$$\begin{aligned}&\sum_{L=0}^{n-1} 2(L-n)T_{n,L}[\ln(-p^2/\mu^2)]^{n-L-1} \\ &\quad = [2A_1(n-1) - C_1]\sum_{L=0}^{n-1} T_{n-1,L}[\ln(-p^2/\mu^2)]^{n-1-L} \\ &\qquad + [2A_2(n-2) - C_2]\sum_{L=0}^{n-2} T_{n-2,L}[\ln(-p^2/\mu^2)]^{n-2-L} + \cdots, \end{aligned} \tag{7.6.7}$$

where $\gamma = C_1 g^2 + C_2 g^4 + \cdots$, and $\beta = -A_1 g^3 - A_2 g^5 + \cdots$.

The leading logarithm part of this equation is

$$-2nT_{n,0} = [2A_1(n-1) - C_1]T_{n-1,0}. \tag{7.6.8}$$

This equation determines the leading logarithm series in terms of $T_{00} = 1$, and of A_1 and C_1; this series sums to (7.6.6).

Equation (7.6.7) also determines the non-leading logarithms. For example the next-to-leading terms are

$$-2(n-1)T_{n,1} = [2A_1(n-1) - C_1]T_{n-1,1} + [2A_2(n-2) - C_2]T_{n-2,0}. \tag{7.6.9}$$

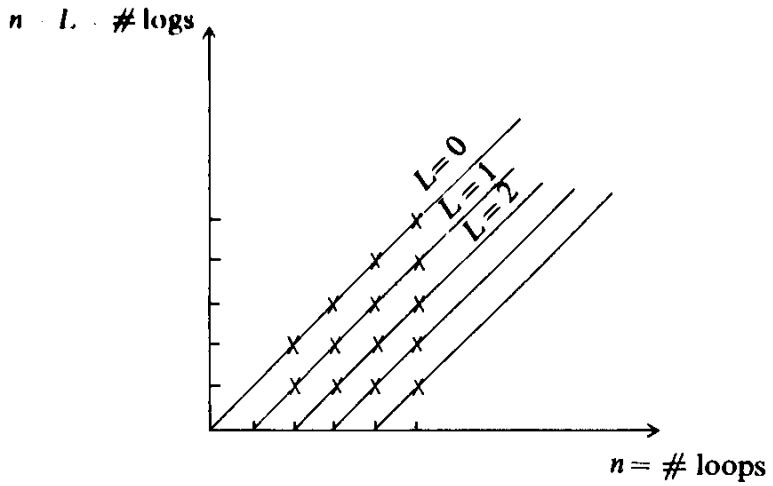


Fig. 7.6.1. Illustrating the leading logarithms and non-leading logarithms of a Green's function.

Again a convergent series results. Its sum is equally accurate as the result of using the two-loop approximation to β and γ in the solution (7.3.21). There μ is set equal to $(-p^2)^{1/2}$ and we take the one-loop approximation to $\tilde{G}_2(p;\bar{g}(\sqrt{-p^2}), m=0, \sqrt{-p^2})$, i.e., $(\mathrm{i}/p^2)\,(1+g^2 T_{11})$.

The series for larger L may be similarly determined. In Fig. 7.6.1 we illustrate the structure of the calculations. The diagonal lines are lines of constant L, and the recursion relation (7.6.7) determines a coefficient $T_{n,L}$ in terms of lower-order terms on its diagonal and on the higher diagonals.

Suppose we have computed perturbation theory to $n-1$ loops for $\tilde{G}_2$ and wish to compute the n-loop term. In this term the coefficients of all but $\ln(-p^2/\mu^2)$ and the constant are fixed by the lower-order calculations. Thus the new information is in the nth order coefficient C_n for γ and in the terms with one and no logarithms, i.e., in $T_{n,n-1}$ and $T_{n,n}$. The $(\ln)^0$ term in (7.6.7) is

$$-2T_{n,n-1} = -C_n + \sum_{j=1}^{n-1} [2A_{n-j}j - C_{n-j}]T_{jj}. \tag{7.6.10}$$

This shows that knowing the coefficient, $T_{n,n-1}$, of the singly logarithmic term in T_n is equivalent to knowing the n-loop coefficient C_n in $\gamma(g)$.

Exactly the same procedure may be applied to any Green's function $\tilde{G}_N$. The only difference is that there are several external momenta. It is enough to consider the connected graphs. Let us write

$$\tilde{G}_N(\text{connected}) = (p^2)^{\dim(G)/2} g^{N-2} \sum_{n=0}^{\infty} \sum_{L=0}^{n} G_{n,L}^{(N)} \ln^{n-L}(-p^2/\mu^2) g^{2n}, \tag{7.6.11}$$

where p is one of the external momenta and we have factored out g^{N-2}, which is the power of g appearing in the tree approximation. The

coefficients $G^{(N)}_{n,L}$ are now functions of the dimensionless ratios of the Lorentz invariants formed from the external momenta.

The leading logarithm series and all the non-leading series are convergent, so they can be summed. The $n!$ behavior of large orders only appears when we consider the single log and constant terms, and thus in the sum over L.

7.6.2 *Non-renormalization-group logarithms*

In all of the above cases there was one logarithm of the large momentum per loop. There are more complicated situations where not all invariants get large. A simple standard example is the form-factor of the electron in QED with a massive photon (Fig. 7.6.2). Here $q^2 = (p_1 - p_2)^2$ gets large but p_1^2 and p_2^2 are fixed. It turns out (Sudakov (1956), Jackiw (1968)) that there are *two* logarithms per loop. These must be in the coefficients $G^{(N)}_{n,L}$ in (7.6.11), since the power of the logarithms is too high for them to be the explicit logarithms in (7.6.11).

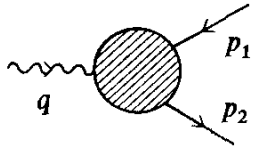


Fig. 7.6.2. The electron's form factor in QED.

One would like to find the large-momentum behavior in such situations. A much-used technique is to sum the leading logarithms, which are often relatively easy to compute. For the on-shell form-factor this gives a convergent series which sums to (Jackiw (1968))

$$F \sim \exp[-(e^2/16\pi^2)\ln^2 q^2]. \tag{7.6.12}$$

(See also Mueller (1981). For the simple cases that we considered earlier, with all momenta large, the leading logarithm approximation is justified by renormalization-group methods, as we have seen. For cases like the present one of the form-factor, it may be a bad approximation (Collins & Soper (1981, 1982b)). However methods are available to obtain large-momentum behavior in some of these situations. See Mueller (1979, 1981) and Collins (1980) for the electron form-factor and Collins & Soper (1981) for cases in strong interactions.

7.6.3 *Landau ghost*

The leading logarithmic approximation (7.6.6) for the propagator has a

singularity when

$$p^2 = -\mu^2 \exp[-1/(A_1 g^2)]. \tag{7.6.13}$$

This singularity, if present in the true propagator, would signal a state of this value of mass squared. Since the residue has the opposite sign to that for a normal propagator pole, this would be a state with unphysical properties. It is called the Landau ghost (Landau & Pomeranchuk (1955)). In a non-asymptotically free theory like QED, it occurs at very large energies and in an asymptotically free theory like QCD, it occurs at low energies. In either case it occurs where perturbation theory is inapplicable and so where the leading logarithmic approximation is a bad approximation.

7.7 Other theories

We restricted our attention in deriving the renormalization group equation to a theory with one field, one coupling, and one mass parameter. However we may treat, by exactly the same method, a theory with several fields $\phi_1, \ldots, \phi_A$ (each may be Bose or Fermi), several couplings $g_1, \ldots, g_B$, and several masses. A change in the unit of mass μ is compensated by a change in each of the parameters and in the scale of the fields. The main problem is a proliferation of indices. It is easiest to treat couplings and masses on the same footing. So we have a collection $g_1, \ldots, g_C$ of renormalized parameters, with C being the total number of couplings and masses. Then we must write

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} g_j(\mu) = \beta_j(\mathbf{g}), \tag{7.7.1}$$

each function β_j being, *a priori*, a function of all the parameters. For the case of a theory with a single coupling, and a single mass, would have $(g_1, g_2) = (g, m^2)$, and $\beta_1 = \beta(g)$ and $\beta_2 = -\gamma_m(g)m^2$.

Given a function f of the renormalized parameters and of μ, we have

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} f(\mathbf{g}, \mu) = \left(\mu \frac{\partial}{\partial \mu} + \sum_j \beta_j \frac{\partial}{\partial g_j} \right) f. \tag{7.7.2}$$

The RG coefficients can be determined by noting that the bare couplings $g_{j(0)}$ are invariant:

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} g_{j(0)}(\mathbf{g}, \mu, d) = 0. \tag{7.7.3}$$

These form C equations for C unknowns.

The RG equations for Green's functions are complicated by the possibility that fields of the same quantum numbers may mix under

renormalization. Writing

$$\phi_{(0)i} = \sum_j \zeta_{ij}(\mathbf{g}, d)\phi_j,$$

we find, for example, that 'kinetic energy' terms in $\mathscr{L}$ are of the form

$$\tfrac{1}{2}\sum_i \partial\phi_{(0)i}^2 = \tfrac{1}{2}\sum_{j,l} \partial\phi_j \partial\phi_l \sum_i \zeta_{ij}\zeta_{il} = \tfrac{1}{2}\sum_{j,l} \partial\phi_j \partial\phi_l Z_{jl}.$$

Hence, we have a matrix counterterm for the field-strength renormalization

$$Z_{ij}(\mathbf{g}, d) = \sum_l \zeta_{li}\zeta_{lj} = (\zeta^{\mathrm{T}}\zeta)_{ij}, \tag{7.7.4}$$

where ^{T} denotes transpose. (Note that ζ has a different meaning here than in Sections 7.1 and 7.2.)

If we define a matrix anomalous dimension by

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\phi_i = -\tfrac{1}{2}\sum_j \gamma_{ij}\phi_j, \tag{7.7.5}$$

then invariance of the bare fields gives

$$0 = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}\phi_{(0)i} = \left(\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\zeta - \tfrac{1}{2}\zeta\gamma\right)\phi, \tag{7.7.6a}$$

i.e.,

$$\tfrac{1}{2}\zeta\gamma = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}\zeta,$$

or

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\mathbf{Z} = -\tfrac{1}{2}\{\mathbf{Z}, \gamma\}. \tag{7.7.6b}$$

If $\mathbf{Z}$ is diagonal (as is the case in most theories we consider): $Z_{ij} = \delta_{ij}Z_i$, then (7.7.6) reduces to an anomalous dimension for each field

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\phi_i = -\tfrac{1}{2}\gamma_i\phi_i, \tag{7.7.7}$$

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln Z_i = \gamma_i.$$

In the case of a diagonal $\mathbf{Z}$, the renormalization group equation for an N-point Green's function is

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}G_N \equiv \left[\mu\frac{\partial}{\partial\mu} + \sum_{j=1}^{B}\beta_j(\mathbf{g})\frac{\partial}{\partial g_i}\right]G_N$$

$$= -\sum_{\alpha=1}^{N}\tfrac{1}{2}\gamma_{i_\alpha}(\mathbf{g})G_N. \tag{7.7.8}$$

Here γ_{i_α} is the anomalous dimension function for the αth external field of G_N. If the renormalization matrix $\mathbf{Z}$ is not diagonal, then we have a similar, but more complicated, set of equations for the Green's functions.

The equations for the evolution of the couplings are coupled, so their solution is in general complicated. Considerable simplification can be achieved by using our knowledge of the dependence of the counterterms on massive couplings. Let the couplings $g_1, \ldots, g_A$ be dimensionless and let the corresponding bare couplings be

$$\mu^{(4-d)\rho_i} g_{(0)i}(\mathbf{g}, d),$$

where the $g_{(0)i}$ depend only on the renormalized dimensionless couplings and on the UV cut-off. For the sake of definiteness, we assume that the physical dimension of space-time in the theory is $d = 4$. The wave-function renormalizations Z_i also only depend on the dimensionless couplings and on d. Let the other parameters (masses and super-renormalizable couplings) be denoted by f_s, and let the dimension of f_s be $(4-d)\tau_s + \sigma_s$. If f_s is the mass of a fermion, then $\sigma_s = 1, \tau_s = 0$. If it is a boson mass squared, then $\sigma_s = 2$ and $\tau_s = 0$, while for a super-renormalizable coupling $\tau_s \neq 0$ and $\sigma_s > 0$. Then (by Section 5.8) the bare quantity corresponding to f_s has the form

$$f_{(0)s} = \sum_{d_X = \sigma_s} \mu^{(4-d)\tau_s} X F_{sX}(\mathbf{g}, d). \tag{7.7.9}$$

Here X is any product of the dimensional couplings with dimension (at $d = 4$) equal to the dimension σ_s of f_s.

Requiring invariance of the bare couplings gives

$$0 = (4-d)\rho_i g_{(0)i} + \sum_{j=1}^{A} \beta_j \frac{\partial}{\partial g_j} g_{(0)i}, \tag{7.7.10a}$$

$$\gamma_j = \sum_{j=1}^{A} \beta_j \frac{\partial}{\partial g_j} \ln Z_j, \tag{7.7.10b}$$

$$0 = (4-d)\tau_s \sum_X X F_{sX} + \sum_{X,j} X \beta_j \frac{\partial}{\partial g_j} F_{sX} + \sum_{t,X} \beta_t \frac{\partial X}{\partial f_t} F_{s,X}. \tag{7.7.10c}$$

There is then a triangular structure to the evolution equation: the evolution of a coupling depends only on couplings of the same and lower dimensions. In these equations, the index j runs over the values 1 to A, i.e., over those values that correspond to the dimensionless couplings, while the indices s and t run over the labels for the dimensionful couplings.

If we use minimal subtraction, the calculation of the coefficients is rather easy. Let $\hat{G}_i(g)$, $\hat{Z}_i(g)$, and $\hat{F}_{sX}(g)$ be the coefficients of single poles in $g_{(0)i}$, Z_i and F_{sX}. Then we have

$$\begin{aligned}\beta_j(\mathbf{g},d) &= (d-4)\rho_j g_j + \bar{\beta}_j(\mathbf{g}),\\ \beta_s(\mathbf{g},\mathbf{f},d) &= (d-4)\tau_s f_s + \bar{\beta}_s(\mathbf{g},\mathbf{f}),\end{aligned} \tag{7.7.11}$$

with γ_j, $\bar{\beta}_j$ and $\bar{\beta}_s$ satisfying

$$\begin{aligned}\bar{\beta}_i &= \rho_i \hat{G}_i - \sum_{j=1}^{A} \rho_j g_j \frac{\partial}{\partial g_j}\hat{G}_i,\\ \gamma_i &= \sum_{j=1}^{A} \rho_j g_j \frac{\partial}{\partial g_j}\hat{Z}_i,\\ \bar{\beta}_s &= \sum_X \left\{\left(\tau_s - \sum_t \tau_t f_t \frac{\partial}{\partial f_t}\right)(X\hat{F}_{sX}) - X\sum_{j=1}^{A}\rho_j g_j \frac{\partial}{\partial g_j}\hat{F}_{sX}\right\}.\end{aligned} \tag{7.7.12}$$

7.8 Other renormalization prescriptions

It was only for the sake of simplicity that we restricted our attention to the minimal subtraction procedure. The proof in Section 7.2 in fact shows that any change in renormalization prescription can be compensated by a change in renormalized parameters and a change in the scale of the renormalized field. Let us examine what happens in more general schemes. It is sufficient to restrict our attention to a theory with a single coupling and mass, like ϕ^3 theory in six dimensions.

If we choose a renormalization scheme with an extra mass μ, which might be a renormalization point, then renormalization-group coefficients can still be defined and computed by (7.3.4), (7.3.5), and (7.3.14). What we lose in general are:

(1) the simple formulae (7.3.12),
(2) the lack of dependence of β, γ_m and γ on the masses.

In order to discuss UV limits, it is sensible to choose a scheme in which the limit $m \to 0$ exists. This means that β, γ_m, γ are finite, order-by-order, as $m \to 0$.

Now, different renormalization schemes are related by finite renormalizations of the parameters. So we may relate the RG coefficients in different schemes by looking at the theory in the physical space-time dimension and then computing $\mu \mathrm{d}/\mathrm{d}\mu$ in one scheme in terms of $\mu \mathrm{d}/\mathrm{d}\mu$ in another scheme with the aid of the chain rule.

Suppose we have a second scheme in which the new mass, coupling and

field are m', g' and ϕ':

$$\left.\begin{aligned} g' &= g'(g, m^2/\mu^2), \\ m'^2 &= m^2 z_m(g, m^2/\mu^2), \\ \phi' &= \phi\zeta(g, m^2/\mu^2). \end{aligned}\right\} \tag{7.8.1}$$

Then the Green's functions in the new scheme are

$$G'_N(p; g', m'^2, \mu) = \zeta^N G_N(p; g, m^2, \mu). \tag{7.8.2}$$

The renormalization-group coefficients in the new scheme are:

$$\beta'(g', m'^2/\mu^2) \equiv \mu\frac{\mathrm{d}}{\mathrm{d}\mu}g' = \left(\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} - \gamma_m m^2\frac{\partial}{\partial m^2}\right)g'(g, m^2/\mu^2), \tag{7.8.3}$$

$$\begin{aligned}\gamma'_m(g', m'^2/\mu^2) &\equiv -\frac{\mu}{m'^2}\frac{\mathrm{d}}{\mathrm{d}\mu}m'^2 \\ &= \gamma_m - \left(\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} - \gamma_m m^2\frac{\partial}{\partial m^2}\right)\ln[z_m(g, m^2/\mu^2)].\end{aligned} \tag{7.8.4}$$

Our definition of the total derivative $\mathrm{d}/\mathrm{d}\mu$ is as the derivative with respect to μ when the bare coupling g_0 and bare mass m_0 are held fixed. Therefore, it is the same in both schemes. Notice that there are two steps in computing β' or γ'_m: First, compute the right-hand side expressed in terms of g, m, and μ; second, change variables to the new coupling and mass.

The anomalous dimension of ϕ' is obtained as

$$\begin{aligned}\gamma'(g', m'^2/\mu^2) &= -\frac{2}{N}\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln G'_N \\ &= -2\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln\zeta + \gamma \\ &= \gamma - 2\left(\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} - \gamma_m m^2\frac{\partial}{\partial m^2}\right)\ln\zeta(g, m^2/\mu^2).\end{aligned} \tag{7.8.5}$$

A considerable simplification occurs in relating mass-independent schemes. Then g', ζ and z_m are functions of g alone, so that the renormalization-group coefficients in the new scheme satisfy

$$\left.\begin{aligned} \beta'(g') &= \beta(g)\frac{\partial}{\partial g}g'(g), \\ \gamma'_m(g') &= \gamma_m - \beta\frac{\partial}{\partial g}\ln z_m(g), \\ \gamma'(g') &= \gamma - 2\beta\frac{\partial}{\partial g}\ln\zeta(g). \end{aligned}\right\} \tag{7.8.6}$$

In this case, let g', z_m and ζ have perturbation expansions

$$\left.\begin{aligned} g' &= g + a_1 g^3 + a_2 g^5 + \cdots, \\ z_m &= 1 + b_1 g^2 + b_2 g^4 + \cdots, \\ \zeta &= 1 + c_1 g^2 + c_2 g^4 + \cdots, \end{aligned}\right\} \tag{7.8.7}$$

and let the expansions of β, γ_m, and γ be

$$\left.\begin{aligned} \beta(g) &= -A_1 g^3 - A_2 g^5 - A_3 g^7 - \cdots, \\ \gamma_m(g) &= B_1 g^2 + B_2 g^4 + \cdots, \\ \gamma(g) &= C_1 g^2 + C_2 g^4 + \cdots \end{aligned}\right\} \tag{7.8.8}$$

The expansions of β', γ'_m, and γ' are written similarly with all quantities primed. Then we can express them in terms of g by using (7.8.7). For example:

$$\begin{aligned} \beta'(g'(g)) &= -A'_1 g^3 (1 + a_1 g^2 + a_2 g^4 + \cdots)^3 - A'_2 g^5 (1 + a_1 g^2 + \cdots)^5 - A'_3 g^7 + \cdots \\ &= -A'_1 g^3 - g^5 (A'_2 + 3a_1 A'_1) - g^7 (A'_3 + 5A'_2 a_1 + 3A'_1 a_2 + 3A'_1 a_1^2) + \cdots. \end{aligned} \tag{7.8.9}$$

This must agree with the perturbation expansion of the right-hand side of (7.8.6)

$$\beta \partial g'/\partial g = -A_1 g^3 - g^5 (A_2 + 3a_1 A_1) - g^7 (A_3 + 3A_2 a_1 + 5A_1 a_2) + \cdots. \tag{7.8.10}$$

From these equations we see that the first two coefficients in β do not change when the renormalization prescription is changed, i.e., $A_1 = A'_1$, $A_2 = A'_2$. By generalizing the above equations to all orders we also see that, by adjusting the terms in the expansion of $g'(g)$, we may choose the terms beyond the second in β' to be whatever we want. In similar fashion we see that only the $O(g^2)$ terms in γ_m and γ are invariant.

Note that if the one-loop term in γ or γ_m is zero then the whole of γ (or γ_m respectively) may be made zero by a choice of renormalization prescription. This privilege does not extend to β: if the first non-vanishing term in β is at n-loop order ($n > 1$) then that term is RG invariant (but not the $(n+1)$-loop term).

In a theory with more than one dimensionless coupling we may try to apply the same methods. This is left as an exercise. It will be found that only the first term in each β is now invariant, except in the case that the one-loop β-function does not mix the different couplings.

The invariance of these coefficients only applies within mass-independent renormalization prescriptions. If one were to use, say, on-shell subtractions, then the parameter μ would not appear, so all derivatives with respect to μ are zero. Then we have $\beta = \gamma = \gamma_m = 0$. (The asymptotic

behavior that we extract by varying μ can no longer be computed by renormalization-group methods, if we stay within this renormalization prescription. In this case the Callan Symanzik equation must be used instead - see Callan (1970) and Symanzik (1970b).)

7.9 Dimensional transmutation

Consider a renormalizable field theory with one dimensionless coupling g and no masses. A physically important case is QCD with several flavors of massless quark; with two or three flavors this is an approximation to actual strong interactions.

Since the basic theory has no masses we must use a renormalization prescription with an arbitrary renormalization mass μ. Although the theory apparently has two parameters, g and μ, we saw that this is not so: a change in μ can be compensated by a change in g. In fact, as Coleman & Weinberg (1973) pointed out, the theory really has no parameters at all. The point is simple but somewhat elusive, so we explain it at length.

A physically measurable quantity must be renormalization-group invariant. For example, let $M(g,\mu)$ be a particle mass. By dimensional analysis, it is μ times a function of g alone. So

$$0 = \mu \frac{\mathrm{d}}{\mathrm{d}\mu} M = M + \beta \frac{\partial}{\partial g} M. \tag{7.9.1}$$

Hence

$$\begin{aligned} M &= \mu \cdot \text{constant} \cdot \exp\left[-\int^{g} \frac{\mathrm{d}g'}{\beta(g')} \right] \\ &= \mu C \exp\left\{ -\frac{1}{2A_1 g^2} - \frac{A_2}{A_1^2} \ln(g) - \int_0^{g} \mathrm{d}g' \frac{[A_1^2 g'^3 + \beta(g')(A_1 - A_2 g'^2)]}{A_1^2 g'^3 \beta(g')} \right\} \\ &= \mu C \exp\left\{ -\frac{1}{2A_1 g^2} - \frac{A_2}{A_1^2} \ln(g) + O(g^2) \right\}. \end{aligned} \tag{7.9.2}$$

Here C is a constant and we have written $\beta(g) = -A_1 g^3 - A_2 g^5 + \cdots$, as usual.

Note that the Green's functions are not renormalization-group invariant: to measure a Green's function, one must define the field operators. This definition has an arbitrariness, which is the freedom to vary its scale.

The formula (7.9.2) has a number of consequences:

(1) Non-zero particle masses cannot be computed in ordinary perturbation theory (in a theory with no mass in the Lagrangian). For to avoid large logarithms one must set μ to be of order M, where M is the particle mass

being computed. Then (7.9.2) tells us that $g(M)$ is not a free parameter; it is a number of order unity.

(2) In a non-asymptotically free theory ($A_1 < 0$), suppose we have a small value for $g(\mu)$. Then $\mu \ll M$, where M is the value of the mass of any given massive particle. Perturbation theory is therefore only useful for Green's functions when the external momenta are much below the threshold for producing any of the massive particles.

(3) In an asymptotically free theory ($A_1 > 0$), we have $\mu \gg M$ whenever $g(\mu)$ is small. Perturbation theory is useful in such a theory only when momenta are much bigger than particle masses.

(4) Since the g-dependence of (7.9.2) is universal, i.e., the same for all particles, ratios of particle masses are pure numbers independent of g and μ.

Let us emphasize once more that these results are true when there are no explicit mass terms in the Lagrangian.

The observation of Coleman & Weinberg (1973) comes from asking what can be measured in the theory. Suppose we start with $\mu = \mu_1$ and $g = g_1$ and ask how the theory changes when we work with the theory with a different value of g, $g = g_2$. (We suppose g_1 and g_2 are between $g = 0$ and $g = g^*$, the first fixed point of β.) Each version of the theory has an effective coupling satisfying

$$g_{\text{vers }1}(\mu_1) = g_1, \quad g_{\text{vers }2}(\mu_1) = g_2.$$

Now evolve $g(\mu)$ in the second version to the value of μ where

$$g_{\text{vers }2}(\mu_2) = g_1.$$

Then the second version of the theory is just the first version with all momenta scaled by a factor μ_2/μ_1. For example let σ be a cross-section depending on momenta $p_1, \ldots, p_N$. Then RG invariance and dimensional analysis give us

$$\begin{aligned}\sigma(p_1, \ldots, p_N; g_2, \mu_1) &= \sigma(p_1, \ldots, p_N; g_1, \mu_2) \\ &= \left(\frac{\mu_2}{\mu_1}\right)^{\dim \sigma} \sigma\left(\frac{\mu_1}{\mu_2} p_1, \ldots, \frac{\mu_1}{\mu_2} p_N; g_1, \mu_1\right). \qquad (7.9.3)\end{aligned}$$

The last factor is the cross-section in version 1 of the theory, with its momenta scaled.

We see that changing the dimensionless coupling in a massless theory does not basically change the theory, but only its mass scale. This is called dimensional transmutation.

There are many ways of specifying the scale of the theory: in QCD one might give the proton mass. For perturbative purposes it is better to use

something that can be directly used in perturbation theory, for example the value of μ at which $g(\mu)$ has some given value (e.g., 0.1) in one's chosen renormalization prescription. One standard way is to notice that for large μ, $g(\mu)$ has its asymptotic behavior given by

$$g^2(\mu) = \frac{1}{A_1 \ln(\mu^2/\mu_0^2)} - \frac{A_2 \ln(\ln(\mu^2/\mu_0^2))}{A_1^3 \ln^2(\mu^2/\mu_0^2)} + \frac{\text{constant}}{\ln^2(\mu/\mu_0)} + O\left[\frac{\ln^2(\ln(\mu/\mu_0))}{\ln^3(\mu/\mu_0)}\right]. \tag{7.9.4}$$

Here μ_0 is a reference value of μ. If μ_0 is changed then the series is reorganized; only the first two terms are unchanged. As is conventional (Buras, Floratos, Ross & Sachrajda (1977)), we define the scale Λ of strong interactions as the value of μ_0 for which the $1/\ln^2(\mu^2/\mu_0^2)$ term is zero. This gives (7.5.5).

If we change from, say, minimal subtraction to momentum-space subtraction, then the theory is unchanged provided the coupling is adjusted. This may be done in perturbation theory. For example, we might find that g in the MS scheme and in the momentum-space subtraction scheme are related by

$$g_{\text{MS}} = g_{\text{mom}} + a_1 g_{\text{mom}}^3 \cdots. \tag{7.9.5}$$

Now let g_{MS} be given by (7.5.5) with $\Lambda = \Lambda_{\text{MS}}$, and let g_{mom} be given by (7.5.5) with $\Lambda = \Lambda_{\text{mom}}$. (We already know that A_1 and A_2 are the same in both schemes.) Substituting these expansions into (7.5.5) and requiring consistency gives

$$\Lambda_{\text{MS}} = \Lambda_{\text{mom}} \exp(a_1/A_1). \tag{7.9.6}$$

Notice that both A_1 and a_1 are obtained from one-loop calculations and that there are no higher-order corrections whatever (Celmaster & Gonsalves (1979)).

An amusing consequence is obtained by substituting (7.5.5) for g in (7.9.2). Since M is independent of μ we may let $\mu \to \infty$. The higher-order terms all go away and leave

$$M = C\Lambda(A_1)^{A_2/2A_1^2}. \tag{7.9.7}$$

This equation is not very useful for performing perturbative calculations.

If the theory is a complete theory of physics, then measurements of σ and the p's in (7.9.3) will be in terms of a standard of mass. This we may take to be the mass M of some particle (say, the proton). Let us now change the theory by changing the coupling from g_1 to g_2, just as we did earlier. Then the standard of mass is multiplied by μ_2/μ_1. So if we do experiments in

version 2 with numerical values of momenta equal to those in version 1, the momenta are actually increased by a factor μ_2/μ_1. Therefore, (7.9.3) tells us that σ gets multiplied by a factor $(\mu_2/\mu_1)^{\dim\sigma}$. But its unit of measurement increases by the same factor, so the numerical value is unchanged. In this sense massless theories with different values of the coupling (or different values of Λ) are indistinguishable. This is perhaps the most important result of Coleman & Weinberg (1973).

However, there are many experiments that claim to *measure* Λ. There are even some that give (without qualification) a single measured value of g. How can this be? The second problem is easy to dispose of. What is being measured is the effective coupling g in some renormalization scheme with μ set to a value of the order of the energy of the experiment (typically in e^+e^--annihilation). Strictly one should specify not only the value of g but also the scheme and the value of μ. Now the experiments are at around 10 to 30 GeV, and Λ is at most a few 100 MeV. The variation of g over this range and the variations between the usual renormalization schemes are often no more than the size of experimental errors. So it is possible to talk loosely.

However, we just asserted that massless QCD with different values of Λ is the same theory. The sense of a measurement of Λ is that we measure the numerical value of the ratio of Λ (defined by (7.5.5)) to a standard of mass. For the purposes of the argument, we may regard the standard as being that the nucleon mass is 939 MeV. In terms of dimensionless quantities the measurement is of the constant C in (7.9.7) when M is the nucleon mass. (In the MS scheme, we find that C is between about 5 and 20.) The non-zero masses of the quarks make a relatively small perturbation of the above argument.

Notice that if we play God and double the size of Λ, then the size of the standard mass also doubles, so that numerical results of experiments are unchanged.

In QED the situation is different. The electron has a mass, and its Coulomb field is classical at large distances. A mass-shell renormalization scheme is natural. Since there is a very important mass-scale, an unqualified statement of a measurement of the QED coupling, viz., $e=(4\pi/137)^{1/2}$, makes good sense. QED with a different value of e is a different theory, unlike QCD in the absence of quark masses.

7.10 Choice of cut-off procedure

It is very convenient to use dimensional continuation as an ultra-violet cut-off in perturbation theory. However, there is no known construction of a

complete theory in an arbitrary complex dimension, so one must beware of assigning too much physical significance to use of dimensional continuation. This is especially true when we use minimal subtraction, which is a procedure that exploits the form of the cut-off dependence of the theory. However, the renormalized theory with the cut-off removed does not depend on the form of the cut-off. We saw this in our one-loop calculations. In general the fact is easiest to see by using BPHZ renormalization, in which an integrand is constructed that gives a manifestly convergent integral. The only freedom left is a change of renormalization prescription, otherwise known as a change of parametrization.

In this section we will examine the renormalization-group properties when a different UV cut-off is used. For definiteness we cut off the theory by using a lattice, with spacing a. We consider any theory with a single dimensionless coupling g and a single mass m. It is, of course, possible to generalize to any cut-off procedure and to any theory. In general we will need a renormalization mass μ, in order that we can take the massless limit. The bare coupling g_0, bare mass m_0, and the field-strength renormalization Z are written as functions of the finite parameters g, m and μ, and of the cut-off a. Then the renormalized Green's functions are written in terms of the bare Green's functions

$$G_N(x_1, \ldots, x_N; g, m, \mu, a) = Z^{-N/2}(g, m, \mu, a) G_N^{(0)}(x_1, \ldots, x_N; g_0, m_0, a), \quad (7.10.1)$$

and for them the limit $a \to 0$ exists.

The renormalization-group structure is essentially unchanged. Let us again choose a mass-independent renormalization prescription, so that g_0, Z, and m_0 have the forms:

$$\left.\begin{aligned} g_0 &= g_0(g, \mu a), \\ Z &= Z(g, \mu a), \\ m_0^2 &= m^2 Z_m(g, \mu a) + a^{-2} Y(g, \mu a). \end{aligned}\right\} \quad (7.10.2)$$

The massless theory has $m = 0$, and, as before, g_0 and Z are independent of mass. But now the cut-off parameter is dimensional, so g_0 and Z have explicit dependence on μ as shown. But the dimension of g_0 is fixed at zero, so the d-dependent power of μ is not used.

The m^2-dependence of the bare mass squared is again linear. But it is no longer true that $m_0 = 0$ when $m = 0$. In the case of dimensional regularization the only remaining dimensional parameter is μ, and it is not possible (Collins (1974)) to generate by minimal subtraction a counterterm $\mu^2 Y(g, d)$. But with a lattice cut-off a term $a^{-2} Y$ is both possible and needed, as we will now verify by computing a low-order graph.

7.10.1 *Example*: ϕ^4 *self-energy*

The simplest example that shows the existence of the Y-term in (7.10.2) is the self-energy graph of Fig. 7.10.1, not in the ϕ^3 theory that we have been using, but in the ϕ^4 theory in four space-time dimensions (with $m = 0$). The Lagrangian is (2.3.1). With dimensional regularization the value of the graph is

$$\tfrac{1}{2}g_0(2\pi)^{-d}\int \mathrm{d}^d k/k^2 = 0,$$

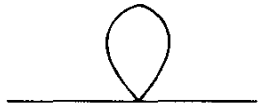


Fig. 7.10.1. Lowest-order self-energy graph in ϕ^4 theory.

but with a lattice cut-off we find

$$-\mathrm{i}g_0(32\pi^4)^{-1}\int_{|k^\mu|<\pi/a} \mathrm{d}^3k\,\mathrm{d}\omega\, D(\omega, k; a). \tag{7.10.3}$$

Here the Euclidean lattice propagator is $1/(\omega^2 + k^2)$ if ω and $|k|$ are much smaller than $1/a$. For general values of k^μ, it is

$$a^2 \Big/ \Big\{4\sum_{\mu=1}^{4} \sin^2(k^\mu a/2)\Big\},$$

which is positive definite, so that the integral (7.10.3) is non-zero and diverges to a number of order $1/a^2$ as $a \to 0$.

A similar divergence occurs in the self-energy of a scalar field in any theory.

7.10.2 *RG coefficients*

We now continue our general discussion of the renormalization group when a lattice cut-off is used. As in the treatment using dimensional regularization we define a renormalization-group operator

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu} = \mu\frac{\partial}{\partial\mu} + \bar{\beta}(g, \mu a)\frac{\partial}{\partial g} - \bar{\gamma}_m(g, \mu a)m^2\frac{\partial}{\partial m^2}. \tag{7.10.4}$$

We have changed our notation slightly, and used an overbar to indicate renormalization-group coefficients in the cut-off theory. These coefficients $\bar{\beta}$ and $\bar{\gamma}_m$ have finite limits, $\beta(g)$ and $\gamma_m(g)$, as $a \to 0$. In our later work it will be rather important to distinguish the coefficients before and after the

cut-off is removed. The coefficients can be computed from

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}g_0 = 0 = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}m_0^2,$$

with (7.10.4) used for $\mu\mathrm{d}/\mathrm{d}\mu$. We also have the anomalous dimension $\bar{\gamma} = \mu\mathrm{d}\ln Z/\mathrm{d}\mu$, just as with dimensional regularization. This all results in

$$\bar{\beta} = \frac{-(\mu a)\dfrac{\partial}{\partial(\mu a)}g_0(g,\mu a)}{\dfrac{\partial}{\partial g}g_0(g,\mu a)},$$

$$\bar{\gamma}_m = \left[\mu a\frac{\partial}{\partial(\mu a)} + \bar{\beta}\frac{\partial}{\partial g}\right]\ln Z_m(g,\mu a), \tag{7.10.5}$$

$$\bar{\gamma} = \left[\mu a\frac{\partial}{\partial(\mu a)} + \bar{\beta}\frac{\partial}{\partial g}\right]\ln Z(g,\mu a).$$

In addition there is the constraint

$$0 = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}(a^{-2}Y) = \frac{1}{a^2}\left[\mu a\frac{\partial}{\partial(\mu a)} + \bar{\beta}\frac{\partial}{\partial g}\right]Y(g,\mu a). \tag{7.10.6}$$

The information on the divergences is all contained in the finite functions $\bar{\beta}$, $\bar{\gamma}_m$ and $\bar{\gamma}$. If desired, we can use minimal subtraction with the form

$$g_0 = g + g^3G_{11}\ln(a\mu) + g^5[G_{22}\ln^2(a\mu) + G_{21}\ln(a\mu)] + \cdots,$$

so that $\bar{\beta} = -g^3G_{11} - g^5G_{21} - \cdots$ is a function of g alone. In order that $\bar{\beta}$ be finite as $a \to 0$ all the logarithms of $a\mu$ must cancel in $\bar{\beta}$. This implies a set of relations for the counterterms, the first of which is $2G_{22} = 3G_{11}^2$. An analogous set of relations occurs when we use dimensional regularization, as can be seen from (7.3.12). These we will discuss further in Section 7.11. Note that for minimal subtraction with the lattice cut-off we have $\beta = \bar{\beta}$, $\gamma_m = \bar{\gamma}_m$, etc.

7.10.3 *Computation of g_0 and Z; asymptotically free case*

If we were to compute the exact theory, rather than a perturbative approximation, we would need to know how $g_0(g,\mu a)$ depends on a as $a \to 0$ with g and μ fixed. A low-order calculation is not sufficient, for g_0 has large logarithms in its perturbative expansion. Provided the theory is asymptotically free, we can remedy this by using the renormalization group to improve the calculation, just as we did for the large-momentum behavior of

Green's functions. The starting point is the equation

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}g_0 \equiv \mu\frac{\partial}{\partial\mu}g_0 + \bar{\beta}\frac{\partial}{\partial g}g_0 = 0, \tag{7.10.7}$$

which is in effect the renormalization-group equation for g_0. We may solve it just as for the Green's functions.

Ultimately, we will let a approach zero while holding g and μ fixed. But first let us keep a non-zero. Then we can define an effective coupling $\bar{g}(\mu)$ by

$$\mu'\mathrm{d}\bar{g}(\mu')/\mathrm{d}\mu' = \bar{\beta}(\bar{g}(\mu')), \tag{7.10.8a}$$

with the boundary condition

$$\bar{g}(\mu) = g. \tag{7.10.8b}$$

We will also need the effective coupling at $a=0$. For the moment, let us denote it by the symbol $\hat{g}(\mu')$. It satisfies

$$\mu'\mathrm{d}\hat{g}(\mu')/\mathrm{d}\mu' = \beta(\hat{g}(\mu')) = \bar{\beta}(\hat{g}(\mu'); \mu' a = 0),$$
$$\hat{g}(\mu) = g.$$

Implicitly there is dependence of $\hat{g}$ on μ and g, and of $\bar{g}$ on μ, g and a:

$$\hat{g} = \hat{g}(\mu'; \mu, g), \quad \bar{g} = \bar{g}(\mu'; \mu, a, g).$$

Of course, $\bar{g}(\mu') \to \hat{g}(\mu')$ as $a \to 0$.

We can solve the renormalization-group equation (7.10.7) for g_0 to find

$$g_0 = g_0(g, a\mu) = g_0(\bar{g}(1/a), 1).$$

Now, when a is small, it might appear that we can replace $\bar{g}(1/a)$ by $\hat{g}(1/a)$, and that g_0 is well approximated by the first term in its perturbation expansion (since $\bar{g}(1/a)$ is small). That is,

$$g_0 = \hat{g}(1/a) + \text{negligible error.}$$

These suppositions are actually false, for two reasons. First, $\bar{\beta}(g, a\mu)$ in general depends on $a\mu$, so we cannot just replace $\bar{g}$ at $\mu = 1/a$ by $\hat{g}(1/a)$ computed in the $a = 0$ theory. Secondly, we cannot simply drop the higher-order terms in g_0, since the dependence of renormalized Green's functions on g_0 is singular. Thus small errors in g_0 may give rise to large errors in a Green's function computed as a function of bare quantities.

To derive the correct formula we must examine the a-dependence of $\bar{\beta}$ more closely. So we write the perturbation expansion of g_0 in the form:

$$g_0 = g + g^3[G_{11}\ln(a\mu) + G_{10} + \bar{G}_1(a\mu)]$$
$$+ g^5[G_{22}\ln^2(a\mu) + G_{21}\ln(a\mu) + G_{20} + \bar{G}_2(a\mu)] + \cdots. \tag{7.10.9}$$

Here we have not specified the renormalization prescription, so in addition to the logarithms we need finite functions $\bar{G}_1(a\mu)$, etc. We have explicit constant terms G_{i0}, so we define $\bar{G}_i(a\mu)$ to be zero at $a\mu = 0$. Once divergences and subdivergences have been subtracted from Feynman graphs, the remainders converge with power-law convergence in momentum. This is a consequence of our treatment of Weinberg's theorem, and is further treated in Weinberg (1960). Therefore we can say

$$\bar{G}_i(a\mu) = O((a\mu)^{c_i})$$

as $a\mu \to 0$, for some positive number c_i. In general, $\bar{G}_i$ equals $a\mu$ times logarithms of $a\mu$, so we can safely set $c_i = 1/2$.

First we compute the $\bar{\beta}$-function:

$$\bar{\beta} = -\mu \frac{\partial}{\partial \mu} g_0 \Big/ \frac{\partial g_0}{\partial g}$$

$$= -\frac{\left\{ g^3 \left[G_{11} + \frac{\partial}{\partial \ln(a\mu)} \bar{G}_1 \right] + g^5 \left[2G_{22} \ln(a\mu) + G_{21} + \frac{\partial}{\partial \ln(a\mu)} \bar{G}_2 \right] + \cdots \right\}}{\{1 + 3g^2 [G_{11} \ln(a\mu) + G_{10} + \bar{G}_1] + \cdots\}}$$

$$= -g^3 \left[G_{11} + \mu a \frac{\partial}{\partial(\mu a)} \bar{G}_1 \right]$$

$$- g^5 \left\{ G_{21} - 3G_{10}G_{11} + \frac{\partial}{\partial \ln(\mu a)} [\bar{G}_2 - \tfrac{3}{2}\bar{G}_1^2 - 3G_{11}\bar{G}_1 \ln(a\mu) - 3G_{10}\bar{G}_1] \right\} + \cdots. \quad (7.10.10)$$

The relation $2G_{22} = 3G_{11}^2$ must hold in order that $\bar{\beta}$ is finite as $a \to 0$. The limit $a \to 0$ gives

$$\beta(g) = \bar{\beta}(g, 0) = -g^3 G_{11} - g^5 (G_{21} - 3G_{10}G_{11}) + \cdots, \quad (7.10.11)$$

so that with our usual notation $A_1 = G_{11}$ and $A_2 = G_{21} - 3G_{11}G_{10}$. Since $\bar{G}_1(a\mu)$ and $\bar{G}_2(a\mu)$ go to zero like a power of $a\mu$ (times logarithms) when $a\mu \to 0$, their logarithmic derivatives $a\mu \partial \bar{G}_j / \partial(a\mu)$ also go to zero like a power.

The first step in our calculation of g_0 is to observe that the RG invariance of g_0 implies that

$$g_0(g, \mu a) = g_0(\bar{g}(\mu'), \mu' a)$$
$$= g_0(\bar{g}(1/a), 1). \quad (7.10.12)$$

The next step is to examine the size of the error that is made in replacing $\bar{g}(1/a)$ by the effective coupling $\hat{g}(1/a)$ in the $a = 0$ theory. Finally, we will find the accuracy to which $g_0(\bar{g}(1/a), 1)$ must be computed in order to obtain

the correct renormalized Green's functions at $a = 0$.

The difference between the two effective couplings $\hat{g}(1/a)$ and $\bar{g}(1/a)$ will turn out to be of order $\hat{g}(1/a)^3$ when a is small. So let us define the fractional error $\hat{g}^2\Delta$ by

$$\bar{g}(\mu';\mu,a,g) = \hat{g}(\mu';\mu,g)[1 + \hat{g}(\mu';\mu,g)^2\Delta(\mu';\mu,a,g)]. \quad (7.10.13)$$

We will now show that Δ is finite when $\mu' = 1/a$ and $a \to 0$.

From (7.10.13) and the definitions of $\bar{\beta}$ and β we find that

$$\mu'\frac{\partial}{\partial\mu'}\Delta(\mu') = \hat{g}^{-3}[\bar{\beta}((1+\hat{g}^2\Delta)\hat{g}, \mu' a) - (1 + 3\hat{g}^2\Delta)\beta(\hat{g})]$$

$$= -\mu' a\frac{\partial}{\partial(\mu' a)}\bar{G}_1 + O(\hat{g}^2(\mu' a)^{1/2}) + O(\hat{g}^4\Delta). \quad (7.10.14)$$

Now $\hat{g}^2(\mu') \sim 1/A_1 \ln(\mu'/\Lambda)$ as $\mu' \to \infty$, so this equation tells us that $\Delta(\mu', \mu, a, g)$ is finite when $a \to 0$ and $\mu \le \mu' \le 1/a$. In fact it implies that

$$\Delta(\mu') = -\bar{G}_1(a\mu') + O(1/\ln(1/a)). \quad (7.10.15)$$

We now compute g_0. It is convenient to write a formula for its square:

$$g_0(g, \mu a)^2 = g_0(\bar{g}(1/a), 1)^2$$

$$= \{\hat{g}(1/a)(1 + \hat{g}^2\Delta) + \hat{g}^3(1 + \hat{g}^2\Delta)^3[G_{10} + \bar{G}_1(1)] + O(\hat{g}^5)\}^2$$

$$= \frac{1}{[A_1 \ln(1/a^2\Lambda^2)]} - \frac{A_2 \ln[\ln(1/a^2\Lambda^2)]}{[A_1^3 \ln^2(1/a^2\Lambda^2)]}$$

$$+ \frac{2G_{10}}{[A_1^2 \ln^2(1/a^2\Lambda^2)]} + O\{\ln^2[\ln(1/a)]/\ln^3(1/a)\}. \quad (7.10.16)$$

Here we used the formula for $\hat{g}(\mu)$ in terms of μ and Λ – (7.5.5). The formula for g_0 in terms of $1/a$ and Λ is the same as (7.5.5) except for an additional $1/\ln^2$ term. Observe that it was essential to keep the a-dependence in $\bar{\beta}(g, a\mu)$; the $-\bar{G}_1(1)$ term in (7.10.15) canceled the $\bar{G}_1(1)$ in the two-loop coefficient in g_0.

Finally we express (7.10.16) in terms of g, a and μ:

$$g_0(g, a\mu)^2 = \frac{1}{A_1 \ln(1/a^2\mu^2)} - \frac{A_2 \ln(\ln(1/a^2\mu^2))}{A_1^3 \ln^2(1/a^2\mu^2)}$$

$$+ \frac{1}{A_1^2 \ln^2(1/a^2\mu^2)}\left[2G_{10} - \frac{1}{g^2} - \frac{A_2}{A_1}\ln(A_1 g^2) - 2A_1 f(g)\right]$$

$$+ O\{\ln^2[\ln(1/a\mu)]/\ln^3(1/a\mu)\}, \quad (7.10.17)$$

where

$$f(g) = \int_0^g dg'[1/\beta(g') + 1/(A_1 g'^3) - A_2/(A_1^2 g')]. \quad (7.10.18)$$

Similar formulae hold for Z_ψ, for Z_m, and for Y. Thus

$$Z = [A_1 g^2 \ln(1/a^2\mu^2)]^{-C_1/2A_1} \times \\ \times \exp\left\{\int_0^g dg' \left[\frac{\gamma(g')}{\beta(g')} + \frac{C_1}{A_1 g'}\right]\right\}\left\{1 + O\left[\frac{\ln(\ln(a\mu))}{\ln(a\mu)}\right]\right\}, \tag{7.10.19}$$

where $\gamma = C_1 g^2 + O(g^4)$.

7.10.4 *Accuracy needed for* g_0

Let us now suppose we compute the renormalized Green's functions:

$$G_N(x_1, \ldots, x_N; g, m, \mu; a) = Z^{-N/2} G_{(0)N}(x_1, \ldots, x_N; g_0, m_0; a). \tag{7.10.20}$$

We must now let a approach zero, and ask how accurately we need to compute g_0 and Z. In (7.10.17) and (7.10.19) we gave formulae for g_0 and Z, with explicit estimates for the errors coming from uncalculated corrections. These equations tell us the value of $g_0(g, a\mu)$ when we let $a \to 0$ while keeping g and μ fixed. Since the bare Green's functions have singular dependence on g_0, the uncalculated corrections might affect the values of the renormalized Green's functions. In fact these terms do not affect the renormalized Green's functions in the continuum limit, as we will now show.

The key observation is that the renormalized Green's functions are finite functions of the renormalized parameters. Thus we do not need to hold the renormalized coupling and mass fixed while taking the continuum limit $a \to 0$. We may in fact let them vary continuously, provided only that their values at $a = 0$ are the same as before. Now examine (7.10.17). It is evident that we may absorb the whole of the correction term into just such a variation of g. In fact the necessary change in g is of order $\ln^2(\ln(1/a))/\ln(1/a)$ as $a \to 0$. So we may choose the bare coupling to be

$$g_0^2 = \frac{1}{A_1 \ln(1/a^2\mu^2)} - \frac{A_2 \ln[\ln(1/a^2\mu^2)]}{A_1^3 \ln^2(1/a^2\mu^2)} \\ + \frac{1}{A_1^2 \ln^2(1/a^2\mu^2)}\left[2G_{10} - \frac{1}{g^2} - \frac{A_2}{A_1}\ln(A_1 g^2) - 2A_1 f(g)\right]. \tag{7.10.21a}$$

Hence in (7.10.16), we can also drop the $O\{\ln^2[\ln(1/a^2)]/\ln^3(1/a)\}$ terms. So we have the following formula for g_0 in terms of $a\,\Lambda$ alone:

$$g_0^2 = \frac{1}{A_1 \ln(1/a^2\Lambda^2)} - \frac{A_2 \ln[\ln(1/a^2\Lambda^2)]}{A_1^3 \ln^2(1/a^2\Lambda^2)} + \frac{2G_{10}}{A_1^2 \ln^2(1/a^2\Lambda^2)}. \tag{7.10.21b}$$

In the case of the wave-function renormalization Z, the uncalculated corrections can be absorbed into a factor ζ^N multiplying the Green's function G_N. This factor must approach unity in the continuum limit. Hence we may use

$$Z = \left[A_1 g^2 \ln\left(\frac{1}{a^2 M^2}\right)\right]^{-C_1/2A_1} \exp\left\{\int_0^g \mathrm{d}g' \left[\frac{\gamma(g')}{\beta(g')} + \frac{C_1}{A_1 g'}\right]\right\}, \tag{7.10.22}$$

where M is an arbitrary mass that is irrelevant when $a \to 0$. Notice that for the coupling we had a form (7.10.21*b*) that had dependence on Λ, but not on μ or on g. This was because g_0 is renormalization-group invariant: we may take μ arbitrarily large without affecting g_0, provided that we also set g equal to the effective coupling at μ. When μ is very big, g is very small, and the higher-order corrections contained in $f(g)$ go to zero. But Z is not invariant; it must depend on g. What we can say is that any dependence on a of the form

$$Z = \text{finite} \cdot [\ln(1/a)]^{-C_1/2A_1}$$

will produce finite Green's functions.

Notice that if the one-loop divergence in Z vanishes, then we may let Z be finite:

$$Z = \exp\left[\int_0^g \mathrm{d}g' \gamma(g')/\beta(g')\right].$$

There will in general be divergences in the self-energy graphs in higher orders. What we have proved is that they must sum to something finite.

In the case of g_0, any a-dependence of the form

$$g_0^2 = 1/A_1 \ln(1/a^2) - A_2 \ln[\ln(1/a)]/[A_1^3 \ln^2(1/a^2)] + \text{finite}/\ln^2 a$$

will give finite renormalized Green's functions. Only knowledge of A_2 and A_1 is necessary for this. They are obtained from one- and two-loop calculations. The coefficient of the $1/\ln^2 a$ determines the value of g.

The formula (7.10.21*b*) shows the fundamental significance of the Λ-parameter. In a renormalizable field theory, there are divergences, so one cannot simply specify a single number as the bare coupling constant. Rather, one must construct the theory as the continuum limit of some lattice theory, with g_0 depending on the lattice spacing, a. Equation (7.10.21*b*) gives g_0 as a definite numerical function of a.

Unfortunately, there is a certain arbitrariness in precisely how one constructs a lattice approximation to a continuum theory. This arbitrariness is physically irrelevant (although some particular approximation may

be superior when it is used for a numerical calculation). So (7.10.21*b*) is also important because it expresses the bare coupling in terms of quantities (Λ, A_1, and A_2), which have direct meaning in the continuum theory, and in terms of one number G_{10}, which depends on the lattice approximation, but which can actually be computed analytically (Hasenfratz & Hasenfratz (1980)). The result of such a lattice calculation is normally given as the ratio of a Λ_{lattice} to the value of Λ in some standard continuum renormalization scheme. The definition of Λ_{lattice} is the value for which

$$g_0^2 = 1/[A_1 \ln(1/a^2\Lambda_{\text{lattice}}^2)] - A_2 \ln[\ln(1/a^2\Lambda_{\text{lattice}}^2)]/[A_1^3 \ln^2(1/a^2\Lambda_{\text{lattice}}^2)] \tag{7.10.23}$$

gives the same continuum limit as (7.10.21). It is easily checked that this is

$$\Lambda_{\text{lattice}} = \Lambda \exp(G_{10}/A_1). \tag{7.10.24}$$

Despite the fundamental significance of Λ, there is a convention dependence in its definition. In specifying a theory by its value of Λ, one must specify these conventions. This is analogous in its effect to the need for specifying a system of units in electromagnetism. The main convention is that of the renormalization prescription. The other convention is the one implicit in the choice of the constant in (7.5.3). It is sensible to follow the usual convention, to avoid confusion.

We have seen that higher-order corrections (beyond two loops) do not enter into our formula for g_0 in terms of Λ. This is in contrast to (7.10.21*a*), which expresses g_0 in terms of g and μ. So it is sensible to treat Λ as a fundamental parameter of the theory – say in strong interactions. But practical considerations intervene if one tries to measure Λ. A typical measurement consists of measuring a quantity for which a useful perturbation expansion exists (for example, a jet cross-section in $e^+ - e^-$ annihilation). The experiment therefore measures the effective coupling $g(\mu)$ at some value of μ which is of the order of the energy of the experiment. There are errors in this measurement caused by uncalculated higher-order terms in the theoretical calculation of the cross-section, not to mention non-perturbative corrections. We can then deduce Λ from (7.5.6), with further errors due to corrections in β.

Since g is more directly related to the size of the cross-section, it is perhaps correct to argue that experiments should quote their results as a value of g. But to give the value of Λ is equally valid. However, a small fractional error in g corresponds to a much larger fractional error in Λ. This can be seen from (7.5.6). If we change g and Λ while holding μ fixed, then

$$|\mathrm{d}\Lambda/\Lambda| = |\mathrm{d}g/g|[2A_1 g^2]^{-1}[1 + O(g^2)].$$

If one could do a real calculation of the mass of, say, the proton in QCD, then it is the error in the value of Λ that would determine the error in the mass. As we saw in Section 7.9, when we discussed dimensional transmutation, the mass is proportional to Λ.

7.10.5 m_0^2

Unfortunately m_0^2 has a $1/a^2$ term, but the variation of m_0 with m^2 depends on a $[\ln(1/a)]^{-b_{11}/A_1}$ term. So we need the coefficient of $1/a^2$ to very high accuracy. Any slight error (say of order $1/a$) will be equivalent to making the renormalized mass diverge like $1/a$ as $a \to 0$. The resulting need to be very accurate in m_0 leads many people to consider scalar field theories unnatural.

In the case of fermion theories there is a symmetry under $\psi \to \gamma_5 \psi$ when $m = 0$, so the Y term is absent and we have

$$m_0 = mZ_m \simeq m \cdot \text{constant}\,[\ln(1/a)]^{-B_1/2A_1}.$$

7.10.6 *Non-asymptotically free case*

The values of g_0, etc., as $a \to 0$ are not perturbatively computable unless the theory is asymptotically free. However, if we suppose that β in a non-asymptotically free theory has a fixed point, then we may write

$$\begin{aligned} g_0(g, a\mu) &= g_0(\bar{g}(1/a), 1) \\ &\to g_0(g^*, 1) \text{ as } a \to 0. \end{aligned} \tag{7.10.25}$$

Note that $g_0(\bar{g}(1/a), 1)$ is a finite function of $\bar{g}$, so the limit exists. However the same value is obtained for g_0 at $a = 0$ for any value of $g(\mu)$. So the way in which the limit is approached determines the value of g.

An example is easily constructed. Suppose we have a theory in which

$$\bar{\beta}(g) = \sin^2(g^2)/(2g), \tag{7.10.26}$$

and

$$g_0(g, 1) = g. \tag{7.10.27}$$

Then the effective coupling has the form

$$g = [\arctan(1/\ln(\Lambda/\mu))]^{1/2}.$$

There is a fixed point $g^* = \pi^{1/2}$. We therefore find that the bare coupling as $a \to 0$ must be

$$\begin{aligned} g_0(g, a\mu) &= [\arctan(1/\ln(a\Lambda))]^{1/2} \\ &= \pi^{1/2} - \frac{1}{2\pi^{1/2}\ln(1/a\Lambda)} + O(1/\ln^2(a\Lambda)). \end{aligned} \tag{7.10.28}$$

It is necessary to know how $g_0(g, a\mu)$ approaches its limit $g_0(g^*, 1)$ in order to determine the value of Λ.

7.11 Computing renormalization factors using dimensional regularization

In the previous section, Section 7.10, we computed how the bare coupling g_0 should behave as a function of the lattice spacing a. In this section we present the corresponding argument using dimensional continuation as the cut-off. We do this by treating the defining equation (7.3.10) of β as a differential equation to compute g_0. Our argument will be valid in any asymptotically free theory, like ϕ^3 theory in six dimensions or QCD in four dimensions. If we let d_0 be the physical space-time dimension, then we regularize by going to a lower dimension $d = d_0 - \varepsilon$.

First we compute the relations between lower and higher poles in the renormalization. Now we write

$$g_0 = \mu^{\varepsilon/2}\left[g + \sum_{j=1}^{\infty} d_j(g)\, \varepsilon^{-j} \right], \tag{7.11.1}$$

and we have the definition of $\bar{\beta}$:

$$\varepsilon g_0/2 + \bar{\beta}(g, d)\partial g_0/\partial g = 0. \tag{7.11.2}$$

Let us expand (7.11.2) in powers of ε. The terms proportional to ε and ε^0 give us:

$$\bar{\beta} = -\varepsilon g/2 + \beta(g) = -\varepsilon g/2 + \tfrac{1}{2}(g\partial/\partial g - 1)d_1(g).$$

We have changed notation from our original definitions to correspond to the definitions that we used in Section 7.10 for the lattice cut-off. There we defined $\bar{\beta}$ to be RG coefficient in the cut-off theory, while we defined β as the limit of $\bar{\beta}$ as the cut-off is removed.

Now, the coefficient of the pole ε^{-j} in (7.11.2) is

$$\tfrac{1}{2}(1 - g\partial/\partial g)d_{j+1}(g) + \beta(g)\partial d_j/\partial g = 0. \tag{7.11.3}$$

This is a differential equation which, when solved using the boundary condition $d_j(0) = 0$, gives all the higher coefficients $d_j(g)$ in terms of the single pole $d_1(g)$.

Similar relations ('t Hooft (1973)) hold for all renormalization counterterms. The structure is similar to the leading logarithm expansion. They show that in each order of perturbation theory the only new information in the counterterm in a given order of perturbation theory is in the single pole.

A convenient way of solving these relations is to work out the solution of the differential equation (7.11.2). This gives

$$\ln[g_0(g, \mu, d)\mu^{-\varepsilon/2}] = \int_0^g \mathrm{d}g' \left[\frac{1}{g' - 2\beta(g')/\varepsilon} - \frac{1}{g'} \right] + \ln g, \tag{7.11.4}$$

i.e.,

$$g_0 = \mu^{\varepsilon/2} g \exp\left\{\int_0^g \mathrm{d}g' \frac{2\beta(g')}{g'^2 \varepsilon} \frac{1}{[1 - 2\beta(g')/(g'\varepsilon)]}\right\}. \tag{7.11.5}$$

The boundary condition $g_0/g \to \mu^{\varepsilon/2}$ as $g \to 0$ has been used.

We now ask how g_0 must behave as $\varepsilon \equiv d_0 - d \to 0$, with g (and μ) fixed. If the theory is not asymptotically free (so that in $\beta(g) = -A_1 g^3 + \cdots$, A_1 is negative), then the integrand has a pole at

$$g'^2 = -(d_0 - d)/2A_1 + O(d - d_0)^2.$$

The solution (7.11.4) only unambiguously exists if g^2 is less than this value, which is zero when $d = d_0$. To get to the $d = d_0$ theory with g non-zero we must continue g_0 so that the integration avoids the pole. The result is that g_0 has an imaginary part. This, among other things, suggests that the theory is unphysical (see Wilson (1973), Gross (1976)). Recently, evidence has accumulated that the lattice ϕ^4 theory does not have a non-trivial continuum limit – see Symanzik (1982) for a review.

If the theory is asymptotically free then we may continue (7.11.4) to $d = d_0$, i.e., $\varepsilon = 0$. The integrand becomes singular when $\varepsilon = 0$, and we examine the singularity by expanding in powers of g':

$$\begin{aligned}
\ln g_0 = {} & \ln g + \frac{\varepsilon}{2} \ln \mu \\
& - \frac{\varepsilon}{2} \int_0^g \mathrm{d}g' \left\{ \frac{1}{[-\varepsilon g'/2 + \beta(g')]} - \frac{1}{[-\varepsilon g'/2 - A_1 g'^3]} \right. \\
& \qquad \left. - \frac{A_2 g'^5}{[-\varepsilon g'/2 - A_1 g'^3]^2} \right\} \\
& - \frac{\varepsilon}{2} \int_0^g \mathrm{d}g' \left\{ \frac{1}{[-\varepsilon g'/2 - A_1 g'^3]} + \frac{A_2 g'^5}{[-\varepsilon g'/2 - A_1 g'^3]^2} + \frac{2}{\varepsilon g'} \right\}.
\end{aligned} \tag{7.11.6}$$

In the first integral we may set $\varepsilon = 0$ and have errors that are $o(1)$. So

$$\begin{aligned}
\ln g_0 = {} & \tfrac{1}{2} \ln\left(\frac{\varepsilon}{2A_1}\right) \\
& + \frac{\varepsilon}{2}\left[\ln \mu - \frac{1}{2A_1 g^2} - \frac{A_2}{2A_1^2} \ln\left(\frac{2A_1 g^2}{\varepsilon}\right) + \frac{A_2}{2A_1^2} - f(g) \right] + o(\varepsilon),
\end{aligned} \tag{7.11.7}$$

where $f(g)$ is defined by (7.10.18). Thus

$$g_0^2 = \Lambda^\varepsilon \left(\frac{\varepsilon}{2A_1}\right) \left\{ 1 + \frac{\varepsilon A_2}{2A_1^2} \left[1 - \ln\left(\frac{2}{\varepsilon}\right) \right] + o(\varepsilon) \right\}. \tag{7.11.8}$$

7.12 Renormalization group for composite operators

We have seen how a change in renormalization prescription for the interactions of a theory can be compensated by a change in the values of the renormalized parameters. The same property holds for the composite operators we defined in Chapter 6.

For example, consider the renormalized $[\phi^2]$ operator in ϕ^3 theory at $d = 6$. In Section 6.2 we calculated it in the one-loop approximation:

$$\langle 0|T\phi(x)\phi(y)[\phi^2](z)/2|0\rangle =$$
$$= \text{tree graph} + \{\text{one-loop graph} + \text{counterterm graphs}\} + \cdots \quad (7.12.1)$$

A change in renormalization prescription amounts to a finite change in the counterterm graphs. Since the counterterms are of the form

$$\tfrac{1}{2}[\phi^2] = \tfrac{1}{2}\phi^2 + \tfrac{1}{2}\delta Z_a \phi^2 + \delta Z_b m^2 \phi + \delta Z_c \Box\phi + \text{higher order}, \quad (7.12.2)$$

we have

$$\tfrac{1}{2}[\phi^2]_{\text{newRP}} = \tfrac{1}{2}[\phi^2]_{\text{oldRP}} + \tfrac{1}{2}a[\phi^2] + b\phi + c\Box\phi. \quad (7.12.3)$$

Here a, b, and c are finite quantities of the same order in the coupling as the one-loop counterterms. The equation (7.12.3) is, so far, only derived at the one-loop order – so the finite counterterms are to be used with their operators inserted in tree graphs.

Let us examine the situation we expect to all orders. We will use minimal subtraction. Then the renormalization in the notation of (6.2.12) is

$$\tfrac{1}{2}[\phi^2] = \tfrac{1}{2}Z_a Z^{-1}\phi_0^2 + \mu^{d/2-3}Z_b Z^{-1/2} m^2 \phi_0 + \mu^{d/2-3} Z_c Z^{-1/2}\Box\phi_0. \quad (7.12.4)$$

Now the bare field is independent of μ, so we may write

$$\begin{aligned}
\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\tfrac{1}{2}[\phi^2] &= Z_a Z^{-1}\tfrac{1}{2}\phi_0^2 \mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(Z_a/Z) \\
&\quad + \mu^{d/2-3}Z_b Z^{-1/2} m^2\phi_0 \mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(\mu^{d/2-3}Z_b Z^{-1/2}m^2) \\
&\quad + \mu^{d/2-3}Z_c Z^{-1/2}\Box\phi_0 \mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(\mu^{d/2-3}Z_c Z^{-1/2}) \\
&= \tfrac{1}{2}[\phi^2]\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(Z_a/Z) \\
&\quad + \mu^{d/2-3}Z_b m^2\phi\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(\mu^{d/2-3}Z_b Z^{1/2}Z_a^{-1}m^2) \\
&\quad + \mu^{d/2-3}Z_c\Box\phi\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(\mu^{d/2-3}Z_c Z^{1/2}Z_a^{-1}), \quad (7.12.5)
\end{aligned}$$

which has the form

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\tfrac{1}{2}[\phi^2] = \gamma_a\tfrac{1}{2}[\phi^2] + \gamma_b m^2\mu^{d/2-3}\phi + \gamma_c\mu^{d/2-3}\Box\phi. \tag{7.12.6}$$

We can formulate this as a matrix equation:

$$\Phi \equiv \begin{pmatrix} \frac{1}{2}[\phi^2] \\ \phi \\ \Box\phi \end{pmatrix} = \begin{pmatrix} Z_a Z^{-1} & m^2 Z_b Z^{-1/2}\mu^{d/2-3} & Z_c Z^{-1/2}\mu^{d/2-3} \\ 0 & Z^{-1/2} & 0 \\ 0 & 0 & Z^{-1/2} \end{pmatrix} \begin{pmatrix} \frac{1}{2}\phi_0^2 \\ \phi_0 \\ \Box\phi_0 \end{pmatrix} \equiv M\Phi \tag{7.12.7}$$

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\Phi = \mu\frac{\mathrm{d}M}{\mathrm{d}\mu}M^{-1}\Phi = \gamma_\Phi\Phi. \tag{7.12.8}$$

The coefficients γ_a, γ_b, and γ_c are finite at $d=6$. From our calculations in Section 6.2, we have

$$\left.\begin{aligned} \gamma_a &= \frac{5g^2}{384\pi^3} = \frac{5}{6}\lambda, \\ \gamma_b &= \frac{g}{64\pi^3} = \lambda/g, \\ \gamma_c &= \frac{g}{384\pi^3} = \frac{1}{6}\lambda/g, \end{aligned}\right\} \tag{7.12.9}$$

where $\lambda = g^2/64\pi^3$. Thus

$$\gamma_\Phi = \begin{pmatrix} \frac{5}{6}\lambda & \dfrac{\lambda m^2\mu^{d/2-3}}{g} & \dfrac{\lambda\mu^{d/2-3}}{6g} \\ 0 & \frac{1}{6}\lambda & 0 \\ 0 & 0 & \frac{1}{6}g \end{pmatrix} + \text{higher order}. \tag{7.12.10}$$

Observe that γ_a, γ_b, and γ_c are all independent of μ and m. This follows from the same arguments that we used to prove the same property for γ_m and γ.

From the RG equation (7.12.7) we prove renormalization-group equations for Green's functions of the composite operators. For example:

$$\begin{aligned} \left(\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} - \gamma_m m^2\frac{\partial}{\partial m^2}\right)&\langle 0|T\phi(x)\phi(y)\tfrac{1}{2}[\phi^2(z)]|0\rangle \\ &= (\gamma_a - \gamma)\langle 0|T\phi\phi\tfrac{1}{2}[\phi^2]|0\rangle \\ &\quad + (\gamma_b m^2 + \gamma_c\Box)\langle 0|T\phi(x)\phi(y)\phi(z)|0\rangle, \end{aligned} \tag{7.12.11}$$

where we have used $\mu\mathrm{d}\phi/\mathrm{d}\mu = -\gamma\phi/2$, and we have set $d=6$, thus eliminating the $\mu^{d/2-3}$ factors.

We must prove (7.12.8) both to all orders for the $[\phi^2]$ operator and in its generalizations to deal with any operators. Since bare operators are automatically RG invariant, the only question is whether the anomalous dimensions are finite. This is handled by a simple generalization of the proof given in Section 7.2 for the ordinary Green's functions. We will not spell out the details – for that is just a mathematical exercise.

8

Large-mass expansion

A common situation in physics is that in investigating phenomena on a certain distance scale, one sees no hint of those phenomena that happen at much shorter distance scales. In a classical situation this observation seems evident. For example, one can treat fluid dynamics without any knowledge of the atomic physics that generates the actual properties of the fluids. However, in a quantum field theory this decoupling of short-distance phenomena from long-distance phenomena is not self-evident at all.

Consider an $e^+ - e^-$ annihilation experiment at a center-of-mass energy well below 10 GeV, the threshold for making hadrons containing the b-quark. There is, for practical (or experimental) purposes, no trace of the existence of this quark. However, the quark *is* present in Feynman graphs as a virtual particle, and can have an apparently significant effect on cross-sections. Our task in this chapter is therefore to prove what is known as the decoupling theorem. This states that a Feynman graph containing a propagator for a field whose mass is much greater than the external momenta of the graph is in fact suppressed by a power of the heavy mass. The physics at low energy is described by an effective low-energy theory that is obtained by deleting all heavy fields from the original theory.

The decoupling of heavy particles is not absolutely universal. One important and typical exception is that of weak interactions. Let us consider the interactions of hadrons at energies of a few GeV. The effective low-energy theory, in the sense just described, consists of strong and electromagnetic interactions alone, without weak interactions. So weak interactions should be ignorable at low energies. However, it is well known that there are in fact many observed processes, particularly decays, that are due entirely to weak interactions. The point is that, in the absence of weak interactions, these processes are exactly forbidden by symmetries, such as parity, charge-conjugation, and strangeness conservation. Weak-interaction amplitudes for the processes in question would be power-law corrections – suppressed by a power of energy divided by the mass of the W-boson – were it not that they are corrections to zero. The consequence of this particular situation is that, at low energies, weak interactions are

described by a *non*-renormalizable theory, viz., the four-fermion interaction. Efforts to find a renormalizable theory led to gauge theories, and a prediction of the W- and Z-bosons from phenomena at energies much lower than their masses: low-energy phenomena have indeed provided clues as to what might happen at much higher energy.

In this chapter, we will treat the cases where decoupling occurs. The theorem that tells us to expect decoupling to occur in many theories was formalized by Appelquist & Carazzone (1975) and Symanzik (1973). They work with a renormalizable theory in which some fields have masses very large compared with the others. They then consider Green's functions of the low-mass fields at momenta much less than the large masses. The theorem is that the Green's functions are the same as those in an effective low-energy theory obtained by deleting all of the heavy fields. Corrections are smaller by a power of momentum divided by a heavy mass. The sole effect of loops of heavy particles is that the couplings of the low-energy theory can have different values from those in the complete theory.

Since the renormalized couplings have no particular *a priori* value, the heavy particles are unobservable until close to threshold. The practical importance of the theorem is that one can understand low-energy physics without having a complete Lagrangian for all phenomena.

We will also show how the renormalization group can be applied in the computation of the relation between the couplings of the low-energy effective theory and those of the full theory.

There are many ramifications of the decoupling theorem, but we will not treat these. One of these is the detailed application of the decoupling theorem to gauge theories (see, for example, Kazama & Yao (1982)). Another is the large-mass expansion of Witten (1976) – where Green's functions of the heavy fields are computed; this expansion is used is deep-inelastic scattering.

We will also not treat the exceptions to the decoupling theorem. These can be treated by the same techniques as those used to prove the decoupling theorem itself. We have already mentioned weak interactions as one of the typical exceptions. Let us just note two other main classes of exception:

(1) In theories with spontaneous symmetry breaking, a mass is often made large by increasing a dimensionless coupling (Veltman (1977) and Toussaint (1978)). The decoupling theorem assumes that a mass is made large by increasing dimensional parameters.
(2) Some dimensionless couplings needed by power-counting violate renormalizability of the low-energy theory (see Collins, Wilczek & Zee (1978)).

In any event the effective low-energy theory is non-renormalizable.

It might be supposed that since General Relativity is non-renormalizable in perturbation theory, it contains some clues to phenomena at very high energies (see, for example, Hawking & Israel (1979)).

8.1 A model

We will restrict our attention to a very simple model. It is a ϕ^3 theory with two fields in six space-time dimensions:

$$\mathscr{L} = (\partial\phi_1)^2/2 + (\partial\phi_h)^2/2 - m^2\phi_1^2/2 - M^2\phi_h^2/2 - \mu^{3-d/2}[g_1\phi_1^3/6 + g_2\phi_1\phi_h^2/2] - \mu^{d/2-3}f\phi_1 + \text{counterterms}. \quad (8.1.1)$$

Symmetry under $\phi_h \to -\phi_h$ has been imposed to cut down the number of possible couplings. Then (8.1.1) contains all couplings necessary for renormalizability. We assume that the renormalized mass, M, of the heavy field is made large while all other parameters are held finite. The factors of the unit of mass μ needed with dimensional regularization are explicitly indicated.

All our techniques can be readily extended to treat more complicated (realistic) theories.

As usual we have introduced a linear term in the Lagrangian to cancel the tadpole graphs. This is determined by the renormalization condition that $\langle 0|\phi_1|0\rangle = 0$.

The remaining counterterms can be put in the form

$$\begin{aligned}\mathscr{L}_{ct} = {} & (Z_1 - 1)\partial\phi_1^2/2 + (Z_h - 1)\partial\phi_h^2/2 \\ & - [m^2(Z_m - 1) + M^2 Z_{mM}]\phi_1^2/2 - [M^2(Z_M - 1) + m^2 Z_{Mm}]\phi_h^2/2 \\ & - \mu^{3-d/2}[(g_{1B} - g_1)\phi_1^3/6 + (g_{2B} - g_2)\phi_1\phi_h^2/2] \\ & - \mu^{d/2-3}(f_B - f)\phi_1. \end{aligned} \quad (8.1.2)$$

As usual, we may choose the dimensionless renormalizations (viz., the Z's and the g_B's) to be independent of the dimensional parameters m^2, M^2, and f.

The decoupling theorem asserts that phenomena on energy scales much less than M are described by an effective low-energy theory whose Lagrangian has the form

$$\begin{aligned}\mathscr{L}_{eff} &= z\partial\phi_1^2/2 - m^{*2}z\phi_1^2/2 - \mu^{3-d/2}g^* z^{3/2}\phi_1^3/6 - \mu^{d/2-3}z^{1/2}f^*\phi_1 \\ &\quad + \text{counterterms} \\ &= \partial\phi^{*2}/2 - m^{*2}\phi^{*2}/2 - \mu^{3-d/2}g^*\phi^{*3}/6 - \mu^{d/2-3}f^*\phi^* \\ &\quad + \text{counterterms}. \end{aligned} \quad (8.1.3)$$

Here we have defined a scaled field $\phi^* = z^{1/2}\phi_1$. We will prove that g^*, m^* and the coefficient z can be chosen so that Green's functions of ϕ_1 obtained from $\mathscr{L}'_{\text{eff}}$ differ from those obtained from the full Lagrangian (8.1.1) by terms which are of the order of a power of external momenta divided by M.

It is usually convenient to work with the scaled field ϕ^* which has unit coefficient for its kinetic term in the basic Lagrangian. Then Green's functions in the full theory are related to Green's functions in the low-energy theory by

$$\begin{aligned}\tilde{G}_N(p_1,\ldots,p_N;g_1,g_h,m,M,\mu) &= \langle 0|T\tilde{\phi}_1(p_1)\ldots\tilde{\phi}_1(p_N)|0\rangle_{\text{full theory}} \\ &= z^{-N/2}\tilde{G}^*_N(p_1,\ldots,p_N;g^*,m^*,\mu)[1+O(1/M^a)] \\ &= z^{-N/2}\langle 0|T\tilde{\phi}^*(p_1)\ldots\tilde{\phi}^*(p_N)|0\rangle[1+O(1/M^a)], \quad (8.1.4)\end{aligned}$$

as $M\to\infty$ with $p_1,\ldots,p_N$ fixed. The fractional errors go to zero as a power of M times logarithms; the power is typically M^{-2}. We can therefore use $1/M^a$, with a slightly less than two, to bound the error. As is our convention, the tilde signs over the fields and Green's functions indicate a Fourier transform into momentum space.

8.2 Power-counting

In this section, we will establish the rules for finding the leading power of M in the value of a graph as $M\to\infty$. These form a simple generalization of Weinberg's theorem, and will involve us in understanding which regions of momentum space are important. We will mostly be interested in graphs for the Green's functions of the light field ϕ_1. Our aim will be to find those graphs that contain lines for the heavy field and that do not vanish as M goes to infinity.

8.2.1 *Tree graphs*

Because we choose to impose the symmetry $\phi_h\to-\phi_h$, the only tree graphs containing lines for the heavy field have heavy external lines. An example is Fig. 8.2.1. Since all momenta on the lines are fixed, and since the free ϕ_h-propagator is

$$i/(p^2-M^2)\sim -i/M^2,$$

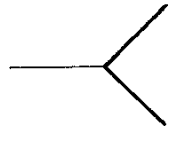


Fig. 8.2.1. A tree graph with a heavy line.

the behavior of any given tree graph as $M \to \infty$ is

$$M^{-2H}, \tag{8.2.1}$$

where H is the number of heavy lines. (We use the natural terminology of calling a line of a Feynman graph heavy or light according to whether its free propagator is for ϕ_h or ϕ_l respectively. Our graphical notation is that heavy lines are thicker than light lines.)

8.2.2 *Finite graphs with heavy loops*

Consider now a graph that has one or more loops but no ultra-violet divergences or subdivergences, and that has some heavy internal lines. The lowest-order example is Fig. 8.2.2(*a*), for the four-point function. Its external momenta (if small) may evidently be neglected on the lines of the loop, whose value is then

$$\Gamma_2 = \frac{g_2^4}{(2\pi)^6}\int d^6k \frac{1}{(k^2 - M^2)^4} = \frac{ig_2^4}{384\pi^3 M^2}. \tag{8.2.2}$$

(We label the symbol Γ by the figure number.)

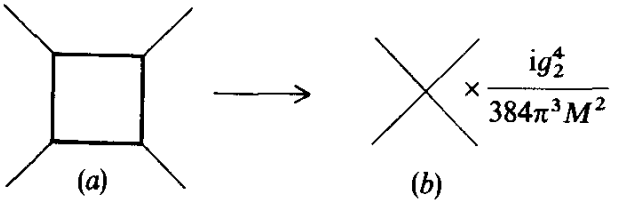


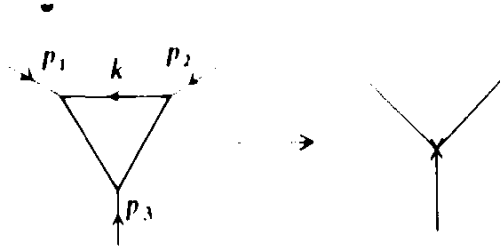
Fig. 8.2.2. Large-mass behavior of graph without an ultra-violet divergence.

The graph vanishes as $M \to \infty$. The precise power of M can be obtained by considering the possible regions of momentum space (after Wick rotation), as follows. Any region of k that is finite as $M \to \infty$ gives a contribution of order M^{-6}. Since the graph is UV finite, the only other possibility is $k = O(M)$. Simple power-counting gives M^{-2}, as found in (8.2.2). This power-counting is the same as for the UV degree of divergence.

The graph is negligible (by a power of M^2) compared to graphs with no heavy lines. If, nevertheless, we wanted its leading contribution, then it would be effectively the local four-point vertex symbolized in Fig. 8.2.2(*b*). The non-renormalizability of this coupling (when the space-time dimension is six) is tied to the negative power of M^2.

8.2.3 *Divergent one-loop graphs*

Consider the logarithmically divergent vertex graph in Fig. 8.2.3(*a*). After



(a) (b)

Fig. 8.2.3. Large-mass behavior of graph with an ultra-violet divergence.

minimal subtraction the loop gives

$$R(\Gamma_3)=\frac{ig_2^3}{64\pi^3}\left\{\tfrac{1}{2}\gamma+\int_0^1 dx\int_0^{1-x} dy\times\right.$$
$$\left.\times\ln\left[\frac{M^2-(p_1^2x+p_2^2y)(1-x-y)-p_3^2xy}{4\pi\mu^2}\right]\right\}$$
$$=\frac{ig_2^3}{128\pi^3}\left[\gamma+\ln\left(\frac{M^2}{4\pi\mu^2}\right)+O\left(\frac{p_i^2}{M^2}\right)\right]. \tag{8.2.3}$$

The same power-counting as for Fig. 8.2.2 confirms the power M^0 for the leading behavior as $M\to\infty$. There is also a logarithm. This occurs because two regions contribute to the leading behavior: the first is where the loop momentum k is of order M. The second region is the UV region where $k\to\infty$. After subtraction of the ultra-violet divergence a finite contribution remains.

Evidently the graph gives a contribution that increases with M. Fortunately the non-vanishing part of the loop is independent of the external momenta. So for large M, the loop is effectively a three-point vertex, as shown in Fig. 8.2.3(*b*). A proof which generalizes to higher order is to differentiate with respect to any external momentum. Since the differentiated graph is finite, it vanishes when $M\to\infty$, like a power of M.

Recall our statement of the decoupling theorem, that at low energies, we could calculate Green's functions from the effective low-energy Lagrangian (8.1.3). The result of our calculation of Fig. 8.2.3 is that the graph generates an extra piece in the ϕ_1^3 coupling of the low-energy theory. Let us therefore write

$$z^{3/2}g^*=g_1-\frac{g_2^3}{128\pi^3}\left[\gamma+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right]+O(g^5). \tag{8.2.4}$$

We may drop the graph Fig. 8.2.3 and replace it by the order g^3 term on the right of (8.2.4). The loop has been replaced by a local vertex where all the lines come to a single point. This corresponds to the fact that the internal line is far off-shell and can only exist for a short time.



Fig. 8.2.4. Large-mass behavior of graph with a quadratic ultra-violet divergence.

The self-energy graph, Fig. 8.2.4, gives a leading term of order M^2. The value of the loop is

$$R(\Gamma_4) = \frac{ig_2^2}{128\pi^3}\left\{(\gamma-1)(M^2-\tfrac{1}{6}p^2) + \int_0^1 dx[M^2-p^2x(1-x)]\ln\left[\frac{M^2-p^2x(1-x)}{4\pi\mu^2}\right]\right\}$$
$$= \frac{ig_2^2}{128\pi^3}\left\{M^2\left[\gamma-1+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right]-\frac{p^2}{6}\left[\gamma+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right] + O\left(\frac{p^2}{M^2}\right)\right\}. \tag{8.2.5}$$

Again the loop momentum k can be either UV or of order M to contribute, so there will be at most a single logarithm of M^2/μ^2. Since we have to differentiate three times with respect to p_μ before obtaining a convergent graph, the non-vanishing terms, as $M\to\infty$, are quadratic in p. From the effective Lagrangian (8.1.3), we see that the graph may be replaced by a contribution to the basic self-energy vertex $i[(z-1)p^2-(m^{*2}z-m^2)]$ in the low-energy theory, with

$$z = 1-\frac{g_2^2}{768\pi^3}\left[\gamma+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right]+O(g^4), \tag{8.2.6}$$

$$zm^{*2} = m^2-\frac{g_2^2M^2}{128\pi^3}\left[\gamma-1+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right]+O(g^4). \tag{8.2.7}$$

We can now compute g^* and m^*:

$$g^* = g_1-\frac{g_2^2(g_2-\frac{1}{4}g_1)}{128\pi^3}\left[\gamma+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right]+O(g^5), \tag{8.2.8}$$

$$m^{*2} = m^2-\frac{g_2^2}{128\pi^3}\left\{M^2\left[\gamma-1+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right] - \tfrac{1}{6}m^2\left[\gamma+\ln\left(\frac{M^2}{4\pi\mu^2}\right)\right]\right\}+O(g^4). \tag{8.2.9}$$

Notice that there is a contribution of order M^2 to the self-energy and

hence to m^{*2}. In order to keep the physical mass of ϕ_1 finite and hence keep m^* finite as $M \to \infty$, we must let m^2 have a term proportional to $g_2^2 M^2$ (with higher-order corrections):

$$m^2 = \text{finite} + g_2^2 \frac{M^2}{128\pi^3}\left[\gamma - 1 + \ln\left(\frac{M^2}{4\pi\mu^2}\right)\right] + \text{higher order}. \quad (8.2.10)$$

On expanding m^{*2} in powers of coupling, we find

$$m^{*2} = \text{finite term in } m^2 + \tfrac{1}{6}\left(\frac{g_2^2 m^2}{128\pi^3}\right)\left[\gamma + \ln\left(\frac{M^2}{4\pi\mu^2}\right)\right] + \text{higher order}. \quad (8.2.11)$$

Since m is the mass parameter for the light field, it is generally considered unnatural to have to fine-tune it within a fractional accuracy of m^{*2}/M^2, as is required by (8.2.10), to obtain a finite value of m^* when $M \to \infty$. In the context of grand unified theories this is called the gauge hierarchy problem (Weinberg (1974, 1976), Gildener & Weinberg (1976)). It is hoped to solve it by finding a phenomenologically sensible theory with no need for fine-tuning.

8.2.4 *More than one loop*

We may have one of the divergent one-loop graphs occurring inside a larger superficially convergent graph. A typical example is Fig. 8.2.5. When

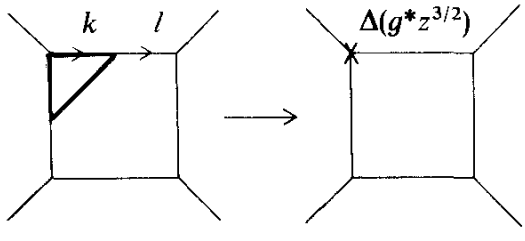


Fig. 8.2.5. Large-mass behavior of two-loop graph with an ultra-violet divergence.

$M \to \infty$ with the external momenta fixed, the only region of the loop momenta that gives a non-zero contribution is where the outer-loop momentum l is finite and the inner-loop momentum k is of order M or larger. So the heavy loop can be replaced by its effective low-energy vertex computed at (8.2.3). This procedure does not change the overall degree of divergence.

The situation at higher order or with overall-divergent graphs is more subtle as we will now see. The graph of Fig. 8.2.6 is typical. Now, it contains a subgraph, consisting of the heavy loop, which we have already considered.

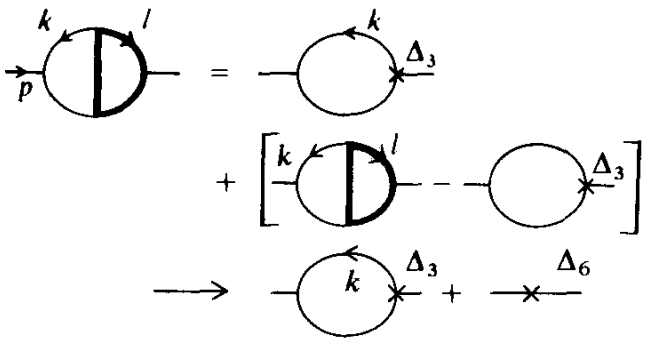


Fig. 8.2.6. Large-mass behavior of another two-loop graph with an ultra-violet divergence.

Therefore, the low-energy theory contains a graph where the heavy loop is replaced by a vertex using (8.2.4) for $g^*z^{3/2}-g$. This graph exactly reproduces the region where k is finite and l is large for Fig. 8.2.6. We add and subtract this graph from the original graph as indicated in the figure. The subtracted term (in square brackets) has a vanishing contribution from finite k (as $M\to\infty$). So we replace it by an effective vertex Δ_6. The same arguments as we used for one-loop self-energy, Fig. 8.2.4, show that it has three terms, proportional to p^2, m^2, and M^2, with coefficients polynomial in $\ln(M^2/\mu^2)$.

In this and in other graphs there are UV divergences for the whole graph and for subgraphs. Implicitly, the counterterm graphs are to be included. Provided we use mass-independent renormalization we are guaranteed that the counterterm graphs satisfy the same power-counting as the original graphs. In particular they are polynomial in the light masses. Thus the counterterm graphs do not change the power-counting and differentiation arguments that are crucial to our work.

8.3 General ideas

Structurally, the arguments in the last section appear similar to those we used in Chapter 7 to show that renormalization-prescription dependence can be compensated by finite counterterms. In fact, as we will see in the next section, Section 8.4, a proof of the decoupling theorem can be constructed exactly by changing the renormalization prescription. We will show that a renormalization prescription can be chosen to have a number of convenient properties, the most important of which is that the low-energy theory is constructed simply by deleting all heavy fields without changing the couplings and masses of the light fields. This property is called manifest decoupling, and we will explain it with the aid of an example in subsection 8.3.1.

Our approach follows the method given by Appelquist & Carazzone (1975) and Witten (1976). This approach generates the effective theory as a series of subtractions. The simplest non-trivial case is given in Fig. 8.2.6.

There is another approach due to Weinberg (1980) in which the decoupling is considered by first integrating over the heavy fields in the functional integral. (See also Ovrut & Schnitzer (1980).) This method is less convenient for treating graphs like Fig. 8.2.6, so we do not use it.

8.3.1 *Renormalization prescriptions with manifest decoupling*

Suppose we used BPH(Z) renormalization instead of minimal subtraction. Then the renormalization condition is that the terms up to $p^{\delta(\Gamma)}$ are zero in the Taylor expansion of a graph Γ about zero external momentum. Here $\delta(\Gamma)$ is the degree of divergence. For a graph with a single loop, consisting of a heavy line, these terms are precisely those that are non-vanishing as $M \to \infty$. Examples are given by the graphs of Figs. 8.2.3 and 8.2.4.

In fact, for a general graph, the effective low-energy theory in this renormalization presciption is obtained merely by deleting all graphs containing heavy lines, together with all their counterterm graphs. The values of the couplings and masses are not changed. Therefore the BPH(Z) prescription has the property we called 'manifest decoupling'. It might appear sensible always to use a renormalization prescription that has this property. However, for many purposes it is useful to use other renormalization prescriptions, e.g. minimal subtraction and its relatives. Particular cases are theories containing massless fields, especially non-abelian gauge theories, and theories with spontaneous symmetry breaking. In any case, it is good to have a direct method of proof of decoupling that can work with any prescription. Furthermore, a prescription like minimal subtraction is more convenient if one also wishes to compute high-energy behavior (Section 7.4) with the aid of the renormalization group. In fact, the method we will use will start from a mass-independent renormalization prescription defined for both the full theory and for the effective low-energy theory. Then the renormalization of the low-energy theory is extended to a renormalization prescription of the full theory in such a way as to satisfy manifest decoupling. This method was first stated by Collins, Wilczek & Zee (1978).

One renormalization prescription that gives manifest decoupling at low energies and that allows the use of renormalization-group methods at high energies is due originally to Gell-Mann & Low (1954). In this scheme, one makes subtractions at some arbitrarily chosen value of momentum. This

scheme was applied to the large-mass problem by Georgi & Politzer (1976). The disadvantage of this scheme, compared with the scheme that we will actually use, is that renormalization-group coefficients are explicitly functions of M/μ, and of m/μ:

$$\beta_1 = \beta_1(g_1, g_2; M/\mu, m/\mu).$$

This makes calculations complicated. Furthermore, this scheme obscures some symmetries.

8.3.2 *Dominant regions*

Before actually constructing a proof of the decoupling theorem, let us give a precise statement of the regions that give unsuppressed contributions (i.e. not suppressed by a power of M^2). We consider each graph in the full theory together with the set of subtraction graphs needed to cancel its divergences and subdivergences. We do not consider the subtraction graphs separately.

First of all, any graph with no heavy lines at all contributes without suppression.

A graph with one or more heavy lines cannot give a contribution unless at least one loop momentum is of order M. The contribution to a graph when M is large can be considered as the sum of contributions from various possible regions of momentum space. The regions can be specified by the sizes of the loop momenta. For our purposes, it is enough to classify a momentum as either finite or large. 'Large' we define to mean 'of order M or bigger'. We can do power-counting for each region in the obvious way. For the loops carrying large momenta, counting powers of M is the same as for the ultra-violet degree of divergence. This gives a factor M^δ, where δ is the ultra-violet degree of the lines carrying large momenta. A heavy line carrying finite momentum counts as M^{-2}. A light line carrying finite momentum counts as M^0.

The leading power of M for a graph is obtained as the maximum of the powers for the possible regions. Let us define $\delta_M(\Gamma)$ to be this highest power. In general there will be logarithmic enhancements. But Weinberg's theorem guarantees that the power $\delta_M(\Gamma)$ is correctly given by considering only the regions we have listed. The graphs treated in Section 8.2 provide examples of this procedure. (The subscript 'M' is to distinguish $\delta_M(\Gamma)$ from the ultra-violet degree of divergence.)

A region contributing to a leading power that is M^0 or bigger is symbolized by contracting to a point both the heavy lines and the lines carrying large momentum. The points represent vertices in the effective low-energy theory; we have already used this notation in Section 8.2, in the

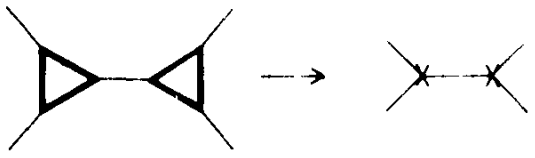


Fig. 8.3.1. Graph with two contracted subgraphs.



Fig. 8.3.2. One-particle reducible subgraphs may have to be contracted.

figures. In general (see Fig. 8.3.1) the contractions will result in several vertices. We include in our definition the restriction that a subgraph is only contracted to a point if it contains at least one heavy line. A contracted subgraph is 1PI in the light lines; for if it can be split into two parts by cutting a light line then that line is not carrying a large loop momentum. However, the contracted graph may be 1PR in the heavy lines. For example, in a theory where the symmetry $\phi_h \to -\phi_h$ is not valid, a graph like Fig. 8.3.2 gives a leading power M^0; the self-energy gives a power M^2 which cancels the $1/M^2$ in the propagator.

A subgraph that is contracted to a single vertex gives the same power of M as its UV power-counting. So

$$\dim(\text{subgraph}) = \text{power of } M + \dim(\text{couplings}).$$

Hence in a renormalizable theory (where couplings have non-negative dimension) the only contracted graphs that have a non-vanishing value as $M \to \infty$ correspond to vertices whose couplings have non-negative dimension. These vertices give the difference between $g^* z^{3/2}$ and g, etc. Thus the couplings in the effective low-energy theory satisfy the dimensional criterion for renormalizability. In a scalar theory, this implies actual renormalizability, provided all the couplings are used that have non-negative dimension and that obey the symmetries of the full theory.

8.4 Proof of decoupling

8.4.1 *Renormalization prescription R^* with manifest decoupling*

Let us work with the theory defined by (8.1.1). We choose to renormalize it according to a mass-independent prescription, which we will denote by a symbol R. For definiteness we choose this to be minimal subtraction. By deleting all heavy fields from (8.1.1) and by changing the values of the couplings we obtain the form of expected low-energy theory (8.1.3). We choose to renormalize the low-energy theory by a mass-independent prescription R^*, which we also take to be minimal subtraction.

Our proof will consist of extending R^* to a renormalization of the full

theory. The extension will satisfy manifest decoupling. Since different renormalization prescriptions differ only by a reparametrization, the statement (8.1.4) of the decoupling theorem will hold. The structure of R^* in the full theory will give a 'mass-independent' form for g^*, m^{*2} and z:

$$\left.\begin{aligned} g^* &= g^*(g_1, g_2, M/\mu), \\ z &= z(g_1, g_2, M/\mu), \\ m^{*2} &= m^2 z_m(g_1, g_2, M/\mu) + M^2 z_{mM}(g_1, g_2, M/\mu). \end{aligned}\right\} \tag{8.4.1}$$

Mass independence means independence of the light mass. As before, to save notational complication we choose to renormalize the linear coupling $f\phi_1$ by the prescription that $\langle 0|\phi_1|0\rangle = 0$. We then ignore both the linear coupling and the tadpole graphs.

The reason we use mass-independent renormalization prescriptions for all the couplings other than the term linear in ϕ is that we can thereby make very clear the decoupling of phenomena at small mass scales from large mass scales. In addition, the renormalization-group equations for (8.4.1) are much simpler to work with than they would otherwise be.

It is convenient to define two concepts:

(1) A heavy graph is one containing at least one heavy line (i.e. a line for the heavy field ϕ_h).
(2) A light graph is one that contains no heavy lines.

For each basic graph Γ in the full theory we have a series of counterterm graphs that are used to cancel its divergences. If Γ is a heavy graph, then we also consider its counterterm graphs to be heavy graphs, even though they may contain no explicit heavy lines.

We have chosen a renormalization prescription R^* for the low-energy theory. This defines the renormalized value of any graph in the low-energy theory, and therefore of any light graph in the full theory. We now wish to extend this prescription to heavy graphs, in such a way that it satisfies manifest decoupling. That is, the renormalized value $R^*(\Gamma)$ of a heavy graph goes to zero as $M \to \infty$. The basic idea is to subtract such graphs at zero momentum. That this is a sensible procedure is easily seen by examining a few of the graphs from Section 8.2.

For example, we saw that Fig. 8.2.3 diverges logarithmically when $M \to \infty$, if we use minimal subtraction. But with zero-momentum subtraction we have

$$\begin{aligned} R^*(\Gamma_3) &= \frac{ig_2^3}{64\pi^3} \int_0^1 dx \int_0^{1-x} dy \ln\left[1 - \left(\frac{p_1^2 x}{M^2} + \frac{p_2^2 y}{M^2}\right)(1-x-y) - \frac{p_3^2 xy}{M^2}\right] \\ &= O(p^2/M^2) \quad \text{as } M \to \infty. \end{aligned} \tag{8.4.2}$$

Clearly, the difference between the two renormalizations is just the difference given by (8.2.4) for $g^* z^{1/2} - g$. At high energy we would use minimal subtraction – so M can be neglected compared with momenta – but at low energy we would use the R^* prescription – so that we can simplify calculations by dropping heavy graphs.



Fig. 8.4.1.

Consider next Fig. 8.4.1 for the self-energy of the heavy field. This graph contains both light and heavy lines. It would behave like $M^2 \ln(M)$ for large M, if we used minimal subtraction. Instead, let us define the subtraction by

$$R^*(\Gamma_{4.1}) = \frac{ig_2^3}{64\pi^3} \frac{\Gamma(2-d/2)}{(4\pi\mu^2)^{d/2-3}} \times$$

$$\times \int_0^1 dx \{ [M^2 x + m^2(1-x) - p^2 x(1-x)]^{d/2-2} - (M^2 x)^{d/2-2}$$

$$- (d/2-2)(M^2 x)^{d/2-3} [m^2(1-x) - p^2 x(1-x)] \}$$

$$\xrightarrow[(d\to 6)]{} \frac{ig_2^3}{64\pi^3} \int_0^1 dx \left\{ [M^2 x + m^2(1-x) - p^2 x(1-x)] \times \right.$$

$$\left. \times \ln\left[1 + \frac{m^2(1-x)}{M^2 x} - \frac{p^2(1-x)}{M^2} \right] - m^2(1-x) + p^2 x(1-x) \right\}$$

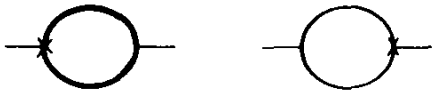
$$= O(1/M^2) \quad \text{as } M \to \infty. \tag{8.4.3}$$

Here, we observed that when $M \to \infty$ the dependence of the unrenormalized graph is linear in m^2 and p^2. So we expanded about $p = m = 0$ and subtracted the terms up to quadratic in m and p. This means that the counterterms are polynomial in m, i.e., 'mass-independence' holds good. Normally subtractions at zero mass and momentum have infra-red divergences, but the presence of a heavy line prevents this here.

As a final example let us examine the two-loop graph of Fig. 8.2.6. The unrenormalized graph cannot be expanded about $m = p = 0$ to give a counterterm, because there are two light lines. At $m, p \sim 0$ they give m- and p-dependence of the form

$$\int_{k \sim m \sim 0} d^6 k \frac{(\text{value of heavy loop at } p = k = 0)}{(k^2 - m^2)[(p+k)^2 - m^2]} \sim (p^2 + m^2) \ln(p^2 + m^2). \tag{8.4.4}$$

The right-hand side of this equation is schematic and symbolizes the



(a) (b)

Fig. 8.4.2. Counterterms for Fig. 8.2.6.

maximum powers and logarithms of m and of p that occur. However, to obtain the renormalized value of the graph we must first subtract subdivergences by the counterterms Fig. 8.4.2, constructed by the R^*-scheme. Now, the counterterm graph (*b*) has infra-red behavior exactly equal and opposite to that of (8.4.4), because the counterterm is minus the value of the heavy loop at $p = k = 0$. Thus the sum of the two graphs has the extra convergence we need. The overall counterterm is then linear in p^2 and m^2. There are no logarithms of m as $M \to \infty$.

8.4.2 *Definition of* R^*

To define the renormalization prescription R^* in general, we simply summarize and generalize what we have just done for particular graphs.

We define the renormalization prescription R^* in the full theory to be the same as our chosen prescription for the low-energy theory whenever it acts on a purely light graph. For a heavy graph Γ, we assume inductively that we have defined the quantity $\bar{R}^*(\Gamma)$ in the usual way to be the unrenormalized value of Γ plus counterterms in the R^*-scheme to cancel its subdivergences. If Γ has degree of divergence $\delta(\Gamma) > 0$, then its overall counterterm is defined by subtraction at $m = p = 0$. The renormalized value of Γ is $R^*(\Gamma) = \bar{R}^*(\Gamma) + C^*(\Gamma)$, as usual.

To define $C^*(\Gamma)$ precisely, we first expand $\bar{R}^*(\Gamma)$ in a Taylor series about the point where its external momenta and the light mass m are zero. Pick out the terms where momenta and m^2 occur with dimension up to $\delta(\Gamma)$, and let the counterterm $C^*(\Gamma)$ be the negative of these terms. Our examples tell us to expect that with such a counterterm:

(1) the leading $M \to \infty$ behavior is canceled,
(2) there are no IR singularities in the counterterm.

We must prove these statements in general. The proof will generalize from the simplest non-trivial case, Fig. 8.2.6. There, the UV divergent unrenormalized graph is not polynomial in m and p, but after subtraction of subdivergences by the R^*-scheme, it becomes polynomial. Then the R^*-prescription can legitimately generate the overall counterterm. Moreover, after subtraction of the subdivergences, the leading large-M behavior is also

polynomial in m and p with degree equal to the degree of divergence, so that it is cancelled by the overall counterterm.

Even with the subtractions for subgraphs, there are in general IR singularities in the Taylor expansion of a graph. For example, consider Fig. 8.4.1 and expand its integrand – see (8.4.3) – in powers of m^2 and p^2. All the terms beyond the second give divergences at $x=0$; it is only the terms needed to cancel the UV divergence that are non-singular.

8.4.3 *IR finiteness of* $C^*(\Gamma)$

Suppose Γ is a heavy graph, 1PI in its light lines. Potential infra-red divergences in $C^*(\Gamma)$ arise when m and the external momenta are made small. They come from regions where some or all of the loop momenta are of order m. The simplest case is where all the internal momenta are of order m.

If Γ were a light graph, we would obtain a contribution of order $m^{\delta(\Gamma)}$, where $\delta(\Gamma)$ is the UV degree of divergence. So let us call $-\delta(\Gamma)$ the canonical IR degree of divergence of Γ. If $\delta(\Gamma)=0$, this is a logarithmic divergence. If $\delta(\Gamma)>0$, then the graph is finite as $m\to 0$. But to get the coefficients of the polynomial counterterms we differentiate up to $\delta(\Gamma)$ times with respect to m and the external momenta. The highest terms in the polynomial are therefore always logarithmically IR divergent, for a light graph.

However, Γ is actually a heavy graph. So at least one of its propagators counts as $1/M^2$ instead of $1/m^2$. Thus all the counterterms have an IR finite contribution from this region, where all its loop momenta are small.

This discussion is sufficient for all one-loop graphs. But multi-loop graphs have IR divergences coming from regions where only some loops have small momenta. For example, Fig. 8.2.6 has a divergence from the region where p and k are small, i.e., order m, and l is finite or large. This corresponds to IR degree -2, and is given by (8.4.4). As we saw, the IR divergence is canceled by the graph with a counterterm for the heavy loop.

The general case is that some light lines carry momenta of order m and

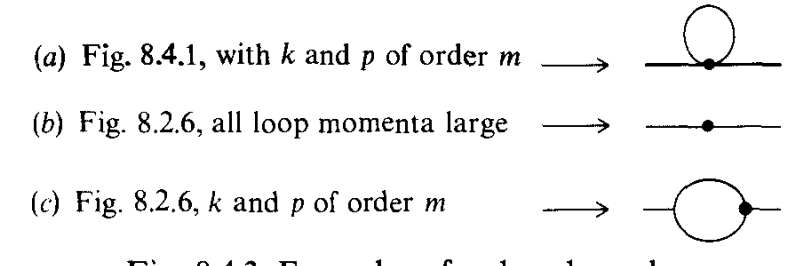


Fig. 8.4.3. Examples of reduced graphs.

the remainder of the lines either are heavy or carry large momentum. Each such region is symbolized by a reduced graph in which the subgraphs consisting of the lines with large momenta and of the heavy lines are contracted to points. Examples of reduced graphs are shown in Fig. 8.4.3. Note:

(1) Counterterm graphs can also have infra-red divergences. The counterterms are inside the vertices of the reduced graphs.
(2) All lines of reduced graphs are light, so at least one vertex of a reduced graph corresponds to a heavy subgraph.

We can write the infra-red degree of divergence for the region corresponding to a particular reduced graph γ as

$$\delta_{\mathrm{IR}}(\Gamma;\gamma) = -\delta(\Gamma) + \sum_{\substack{\text{reduced} \\ \text{vertices } V}} [\delta(V) + \text{IR degree of } R^*(V)]. \quad (8.4.5)$$

The meaning of this equation can be seen from an example. Consider Fig. 8.2.6 when k and p are of order m. If the graph were purely light, we would have IR degree equal to -2, which is the negative of the UV degree. This would imply that the m^2 and p^2 terms in the expansion about $m = p = 0$ would be divergent. However, the single reduced vertex – as illustrated in Fig. 8.4.3(c) – has a counterterm. This counterterm ensures that the vertex's value is of order m^2/M^2 instead of m^0. The IR degree for the whole graph is thereby decreased by 2. The second term in (8.4.5), where the sum is over this single vertex, indicates this reduction. The degree for the region is then -4; we can therefore expand up to order m^2 and p^2 without an infra-red divergence. The terms of order m^4, p^4, etc., are infra-red divergent, but they are not needed for ultra-violet renormalization.

In the general case of (8.4.5), each reduced vertex V would contribute $-\delta(V)$ if it were light and all its internal lines had momenta of order m. But it actually contributes what we will now prove is a smaller amount. Remember that counterterm graphs also contribute, and we assume that counterterm vertices are included inside reduced vertices. The IR degree of $R^*(V)$ is its power as its external momenta are scaled like m. The possible cases for V are:

(1) If V is overall convergent and contains a heavy line, then its infra-red degree is greater than its ultra-violet degree. Fig. 8.4.3(a) has a vertex with UV degree -2 and IR degree zero.
(2) If V is overall divergent and contains a heavy line then ordinarily we would expect it to behave as m^0 when $m \to 0$ with fixed M. But we make

subtractions by the R^* scheme so that its behavior is actually $m^{\delta(\Gamma)+2}$. By induction we may assume its subtractions have no IR divergence.

Hence, in every region of momenta a heavy graph Γ always has at least one mechanism to reduce its IR degree below $-\delta(\Gamma)$ and none to increase it. Thus the overall counterterm $C^*(\Gamma)$ is IR finite. It is crucial to our inductive proof that we first subtract subdivergences by the R^* scheme.

8.4.4 *Manifest decoupling for R^**

A purely light graph is a graph in both the full theory and in the low-energy theory. It survives unaltered when we let $M \to \infty$. We will now prove that all the heavy graphs vanish when $M \to \infty$, given that we renormalize them by the R^* scheme.

To do this, decompose each heavy graph into its skeleton, i.e., a series of 1PI graphs connected by lines that are not part of any loop. Since a heavy line that is outside a loop vanishes as $M \to \infty$, all heavy graphs vanish as $M \to \infty$, if the 1PI graphs vanish.

The $M \to \infty$ limit of a 1PI graph can be related to an IR limit by scaling all masses and momenta:

$$M \to 1, \quad p \to p/M, \quad m \to m/M.$$

Then

$$\Gamma(p, m, M) = M^{d(\Gamma)}\Gamma(p/M, m/M, 1), \tag{8.4.6}$$

where $d(\Gamma)$ is the dimension of Γ. So Γ vanishes as $M \to \infty$ provided the infra-red behavior is less singular than $m^{-d(\Gamma)}$. But this is what we showed in the proof of IR finiteness of the counterterms. (Note that the dimension of a graph is greater than or equal to its UV degree of divergence.)

8.4.5 *Decoupling theorem*

We have constructed two renormalization prescriptions, labelled R and R^*, for the theory under consideration. The Green's functions in the schemes R and R^* are equal provided we make appropriate changes in the parameters:

$$\left.\begin{aligned} g_1 &\to g^*, \\ g_h &\to g_h^*, \\ \text{coefficient of } \partial\phi_1^2/2 &\to z, \\ \text{coefficient of } \partial\phi_h^2/2 &\to z_h, \\ m^2 &\to m^{*2}, \\ M^2 &\to M^{*2}, \end{aligned}\right\} \tag{8.4.7}$$

This is just a particular case of a renormalization-group transformation, and is proved by Section 7.2. When $M \to \infty$ we may drop all heavy graphs in the R^* scheme (so also g_h^*, z_h, M^* drop out of consideration). This then gives (8.1.4), which is the decoupling theorem.

Mass-independence is true because we have arranged all counterterms to be polynomials in the light mass of the appropriate degree.

8.5 Renormalization-group analysis

When one computes a graph containing lines for fields with widely different masses, one finds, in general, that its value gets large as a power of the logarithm of the mass ratio. Such large coefficients are undesirable in a perturbation expansion, for they mean that the reliability of using a few low-order terms is worsened. This situation arises in both strong- and weak-interaction physics. We will now show how to combine the decoupling theorem and the renormalization group to do calculations without their being made unreliable by the large logarithms.

A convenient method is to use a mass-independent scheme (specifically minimal subtraction) for high-momentum calculations, where one often wishes to neglect all masses, and to use the R^* scheme, as defined in Section 8.4, at low momenta, where one wishes to neglect heavy graphs. An advantage of this method is a simplification of many of the calculations needed to match high-energy and low-energy calculations. One needs only the pole parts of graphs and the values at zero external momentum.

We will explain how to use this scheme in the toy theory (8.1.1). First let us write the RG equations for the Green's functions. For a Green's function of N_l light and N_h heavy fields, we have

$$\left(\mu\frac{\mathrm{d}}{\mathrm{d}\mu} + \tfrac{1}{2}N_l\gamma_l + \tfrac{1}{2}N_h\gamma_h\right)G_{N_l,N_h} = 0, \tag{8.5.1}$$

where

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu} = \mu\frac{\partial}{\partial\mu} + \beta_1(g_1,g_2)\frac{\partial}{\partial g_1} + \beta_2(g_1,g_2)\frac{\partial}{\partial g_2} - (M^2\gamma_M + m^2\gamma_{mM})\frac{\partial}{\partial M^2} - (m^2\gamma_m + M^2\gamma_{Mm})\frac{\partial}{\partial m^2}. \tag{8.5.2}$$

The RG coefficients are obtained from the renormalization counterterms as usual. Their lowest-order values are

$$\beta_1 = -\frac{1}{128\pi^3}(\tfrac{3}{2}g_1^3 + g_1g_2^2 + g_2^3) + \cdots,$$

$$\beta_2 = -\frac{1}{128\pi^3}(\tfrac{1}{4}g_2^3 + g_1 g_2^2 - \tfrac{1}{6}g_1^2 g_2) + \cdots,$$

$$\gamma_l = \frac{1}{192\pi^3}g_2^2 + \cdots,$$

$$\gamma_h = \frac{1}{384\pi^3}(g_1^2 + g_2^2)\ldots. \tag{8.5.3}$$

In the effective low-energy theory, the RG equation is

$$\left[\mu\frac{d^*}{d^*\mu} + \tfrac{1}{2}N\gamma^*\right]G_N^* = 0, \tag{8.5.4}$$

with

$$\left.\begin{aligned} \mu\frac{d^*}{d^*\mu} &= \mu\frac{\partial}{\partial\mu} + \beta^*\frac{\partial}{\partial g^*} - \gamma^* m^{*2}\frac{\partial}{\partial m^{*2}}, \\ \beta^* &= -\frac{3}{464\pi^3}g^{*3} + \cdots, \\ \gamma^* &= \frac{g^{*2}}{384\pi^3} + \cdots. \end{aligned}\right\} \tag{8.5.5}$$

To compare the low-energy theory and the full theory, we extended the renormalization scheme of the low-energy theory to a renormalization scheme R^* for the full theory. In this scheme the RG operator has the form

$$\mu\frac{d^*}{d^*\mu} = \mu\frac{\partial}{\partial\mu} + \beta^*\frac{\partial}{\partial g^*} + \beta_2^*\frac{\partial}{\partial g_2^*} - (M^{*2}\gamma_M^* + m^{*2}\gamma_{mM}^*)\frac{\partial}{\partial M^{*2}} - (m^{*2}\gamma_m^* + M^{*2}\gamma_{Mm}^*)\frac{\partial}{\partial m^{*2}}, \tag{8.5.6}$$

and the anomalous dimensions of the fields are γ^* and γ_h^*. In fact β^*, γ_m^*, and γ^* are identical to those in the low-energy theory (see (8.5.4) and (8.5.5)), while $\beta_2^* = \gamma_{mM}^* = \gamma_{Mm}^* = \gamma_M^* = \gamma_h^* = 0$. This is easily seen by examining the Green's functions which provide the normalization conditions for the renormalizations.

For example, consider the inverse of the heavy propagator when both p^2 and m^2 are much less than M^2:

$$1/G_{0,2} = -i[p^2 - M^{*2} + O(p^4, m^2p^2, m^4)] \tag{8.5.7}$$

which satisfies

$$\mu\frac{d^*}{d^*\mu}G_{0,2}^{-1} = \gamma_h^* G_{0,2}^{-1}. \tag{8.5.8}$$

This is only consistent if $\gamma_h^* = \gamma_M^* = \gamma_{mM}^* = 0$.

8.5.1 Sample calculation

We wish to start with the full theory renormalized by minimal subtraction. In that version of the theory, we know the evolution of the couplings. Our aim is to compute Green's functions in the low-energy theory and the values of the mass and coupling. The low-energy effective couplings are

$$\left.\begin{aligned} g^{*2}(\mu) &= \frac{128\pi^3}{3\ln(\mu/\Lambda^*)} + \cdots, \\ g_2^* &= \text{fixed}, \\ m^{*2} &= \text{constant}\,[\ln(\mu/\Lambda^*)]^{-1/9}, \\ M^{*2} &= \text{fixed}. \end{aligned}\right\} \tag{8.5.9}$$

The effective couplings for the full theory with minimal subtraction are more complicated because they solve a coupled equation for two variables.

To make the transition between the schemes we compute the lowest-order divergent graphs. We equate the self-energy for ϕ_1 in the two schemes, with use of the Lagrangian (8.1.3) for the low-energy theory. This gives

Fig. 8.2.4 + pole counterterm

= Fig. 8.2.4 + zero-momentum counterterm + $i(z-1)p^2 - i(m^{*2}z - m^2)$.

We thus obtain z and m^{*2} as given by (8.2.6) and (8.2.7). To keep m^{*2} finite, and not of order M^2, we must replace m^2 by

$$m^2 + \frac{1}{128\pi^3} M^2 g_2^2 [\gamma - 1 + \ln(M^2/4\pi\mu^2)]. \tag{8.5.10}$$

Notice the presence of logarithms of M/μ. If they are large enough, they invalidate the use of perturbation theory to compute g^*, m^{*2}, and z. However, the equations we write are valid at any value of μ, so we may perform the calculations with μ of order M. After computing g^*, m^*, and z in terms of g_1, g_2, m, and M, we can evolve them to the value of μ that we wish to use for calculations in the low-energy theory.

A convenient point to do the matching is where $g_1 = g^*$, i.e., at $\mu^2 = \mu_0^2$, where $\mu_0^2 = M^2 e^{\gamma}/4\pi$. Then for a general value of μ we have

$$g^{*2}(\mu) = \frac{1}{1/g_1^2(\mu_0) + (3/128\pi^3)\ln(\mu/\mu_0) + \cdots}. \tag{8.5.11a}$$

A similiar equation holds for m^*:

$$m^{*2}(\mu) = m^2[(128\pi^3/3g_1^2) + \ln(\mu/\mu_0) + \cdots]^{-1/9}. \tag{8.5.11b}$$

The solution for z is more complicated since the renormalization-group equation for both renormalization prescriptions is needed.

8.5.2 Accuracy

We compute $g^*(\mu)$ in the low-energy theory by matching to the full theory at some μ_0 of order M and then evolving to an arbitrary renormalization mass μ from μ_0. Given the accuracy in $g^*(\mu)$ that we need for a particular calculation, we will find the order to which we must perform the matching and to which we must know β.

The RG equation for $g^*(\mu)$ gives

$$\ln(\mu/\mu_0) = \int_{g(\mu_0)}^{g(\mu)} \mathrm{d}g'/\beta(g'). \tag{8.5.12}$$

So if there are small errors $\Delta g(\mu_0)$ and $\Delta(1/\beta)$ in $g(\mu_0)$ and $1/\beta$ then the error in $g(\mu)$ is

$$\begin{aligned}\Delta g(\mu) &\sim \beta[g(\mu)]\left\{\frac{\Delta g(\mu_0)}{\beta[g(\mu_0)]} - \int \mathrm{d}g' \Delta\left[\frac{1}{\beta(g')}\right]\right\} \\ &= O[g(\mu)^3]\left\{\frac{\Delta g(\mu_0)}{O[g(\mu_0)^3]} - \int \mathrm{d}g' \Delta\left[\frac{1}{\beta(g')}\right]\right\}.\end{aligned}$$

Suppose we perform matching up to n_m-loop order; then the error in $g(\mu_0)$ is of order $g(\mu_0)^{2n_m+3}$. Suppose β is computed to n_β-loop order; then the error in $1/\beta(g)$ is of order $g^{2n_\beta-3}$. These translate to errors in $g(\mu)$ of order

$$g(\mu)^3 g(\mu_0)^{2n_m}$$

and

$$g(\mu)^3 \ln[g(\mu)/g(\mu_0)] \text{ if } n_\beta = 1,$$

or

$$O[g(\mu)^{1+2n_\beta}] + O[g(\mu)^3 g(\mu_0)^{2(n_\beta-1)}] \text{ if } n_\beta \geq 2.$$

For example, if we wish to perform reliable two-loop calculations, then we need Δg to be much smaller than g^3. This means that we need to do the matching correct to one loop and that the β-function is needed to two loops. This is the minimum accuracy needed to correspond to a fractional error on Λ (defined in Chapter 7) which is much less than unity.

9

Global symmetries

In this chapter we consider the impact of global symmetries of a field theory on its renormalization. As an example consider the theory of a charged scalar field:

$$\mathscr{L} = \partial\phi^\dagger \partial\phi - m^2\phi^\dagger\phi - g(\phi^\dagger\phi)^2/4. \tag{9.0.1}$$

This classical Lagrangian is invariant under the transformation $\phi \to e^{-i\omega}\phi$. The quantum theory is also invariant. For this particular theory, the quantum invariance is not a very deep statement. However, symmetries do not always survive quantization, as we will see in Chapter 13. Thus it is useful to examine the consequences of the symmetry in this theory. One consequence is that only invariant counterterms are needed; for example, we do not need to use non-invariant counterterms proportional to

$$\phi^2 + \phi^{\dagger 2} \quad \text{or} \quad i(\phi^2 - \phi^{\dagger 2}).$$

Other consequences are the Ward identities, which characterize the action of the symmetry at the level of Green's functions.

The main step in proving the statements is to impose an ultra-violet cut-off. If this is done by putting the theory on a lattice or by using dimensional regularization, the symmetry is preserved. The arguments given in Section 2.7 are sufficient to prove Ward identities in the bare theory. From the invariance of Green's functions follows invariance of the counterterms. As we will see in Section 9.1 we can then write renormalized Ward identities in the renormalized theory, which therefore exhibits the symmetry.

In more general cases this simple procedure fails.

One case is that the UV cut-off breaks the symmetry. For example, putting the theory on a lattice breaks Poincaré invariance. Luckily, other regulators, like dimensional continuation, preserve this invariance, and the renormalized theory with no cut-off is Poincaré invariant. Some symmetries cannot be preserved after quantization. It must be true that no regulator can preserve them. An example, to be treated in Chapter 13, is the chiral invariance of QCD.

Another case, which we will treat later in this chapter, is of spontaneous symmetry breaking, typified by the theory given by (9.0.1) with m^2 replaced by $-m^2$. This is called the Goldstone model. In this case the ground-state – the vacuum – is not invariant under the symmetry, and the field acquires a vacuum expectation value:

$$\langle 0|\phi|0\rangle = [2|m^2|/g]^{1/2} + \text{higher order}.$$

If we use an invariant regulator, like dimensional continuation, we will still be able to prove Ward identities. Hence, we will be able to prove that only symmetric counterterms are needed, so that the symmetry is preserved. From the Ward identities follows Goldstone's theorem, that there is a massless boson for each generator of a broken symmetry.

9.1 Unbroken symmetry

We first consider a totally unbroken internal symmetry. The fields carry a matrix representation of the generators. Thus:

$$\delta_\alpha \phi_i = -\mathrm{i}(t_\alpha)_i{}^j \phi_j, \tag{9.1.1}$$

in the notation of Section 2.6.

The proof that the symmetry can be preserved under quantization is elementary. We spell out the steps so that we can see what needs to be done in less trivial cases:

(1) Regulate in a way that preserves the symmetry. Lattice and dimensional regularization both do this since the symmetry commutes with all space-time transformations.

(2) Include in $\mathscr{L}$ all possible invariant counterterms up to the appropriate dimension. Thus $\delta_\alpha \mathscr{L} = 0$. For the model (9.0.1) we replace $\mathscr{L}$ by

$$\mathscr{L} = Z\partial\phi^\dagger\partial\phi - m_B^2\phi^\dagger\phi - g_B(\phi^\dagger\phi)^2/4. \tag{9.1.2}$$

(3) To do perturbation theory, let the free Lagrangian be invariant: $\delta_\alpha \mathscr{L}_0 = 0$. Then the interaction Lagrangian is also invariant.

(4) At each order, choose the counterterms to cancel the divergences in 1PI Green's functions. Since the free propagators and the interactions are all invariant under the symmetry, the divergences are symmetric and non-invariant counterterms are not needed.

(5) Remove the UV cut-off. The Green's functions are symmetric:

$$\delta_\alpha \langle 0|T\phi_{j_1}(y_1)\ldots\phi_{j_N}(y_N)|0\rangle = 0. \tag{9.1.3}$$

In the case of the model (9.0.1) the propagator for the charged field carries an arrow indicating the direction of flow of charge. All vertices have

equal numbers of ingoing and outgoing lines. In (9.1.3) we have $\delta\phi = -i\phi$ and $\delta\phi^\dagger = i\phi^\dagger$, so this equation is literally a statement of charge conservation.

The current for a symmetry is defined by Noether's theorem (Section 2.6):

$$j^\mu_\alpha = \sum_i \delta_\alpha \phi_i \frac{\partial \mathscr{L}}{\partial \partial_\mu \phi_i}. \tag{9.1.4}$$

In the case of the simple model (9.0.1) there is a single current

$$j^\mu = iZ\phi^\dagger \overleftrightarrow{\partial^\mu} \phi. \tag{9.1.5}$$

We derived the Ward identities of the bare theory (2.7.6). For the theory (9.0.1) these are

$$\frac{\partial}{\partial x^\mu} \langle 0 | T j^\mu(x) \phi(y_1) \cdots \phi(y_N) \phi^\dagger(z_1) \cdots \phi^\dagger(z_N) | 0 \rangle$$
$$= i \sum_{j=1}^{N} [\delta(x - y_j) - \delta(x - z_j)] \langle 0 | T \phi(y_1) \cdots \phi(y_N) \phi^\dagger(z_1) \cdots \phi^\dagger(z_N) | 0 \rangle. \tag{9.1.6}$$

We showed in Section 6.6 that the current is in fact finite; no extra renormalization counterterms are needed beyond those implied by the factor Z in (9.1.5).

It is of interest to see how the divergences that are present get cancelled by the factor Z. For the two-point function of j^μ we have the 1PI graphs of Fig. 9.1.1, up to order g^2. Since $Z = 1 + O(g^2)$ in this theory, we may replace Z by 1 everywhere except in the tree graph (*a*). Graph (*b*) could be logarithmically divergent by power-counting, but is in fact zero, so no counterterm is needed at order g. Graph (*e*) is also zero. Graphs (*c*) and (*d*) are finite after their subdivergences are cancelled by a counterterm; they also cancel each other. These cancellations arise since these graphs have a

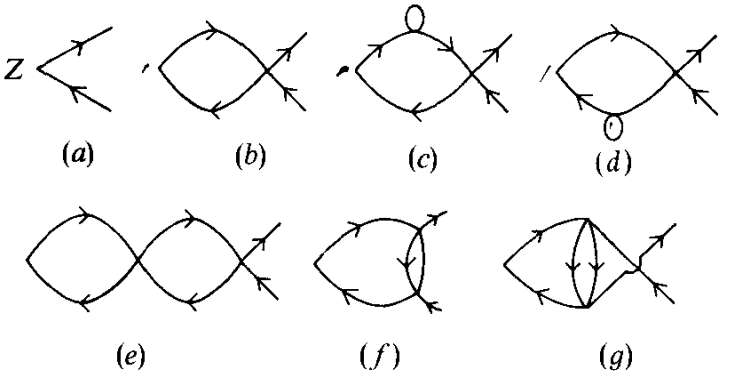


Fig. 9.1.1. Graphs up to order g^2 for the two-point function of j^μ.

subgraph which is a graph for

$$\langle 0|\, T j^{\mu}(x)\, \phi^{\dagger}\phi(y)|0\rangle .$$

In momentum space this is of the form $q^{\mu} f(q^2)$. The Ward identity implies that its divergence is zero:

$$\frac{\partial}{\partial x^{\mu}}\langle 0|\, T j^{\mu}(x)\, \phi^{\dagger}\phi(y)|0\rangle = \mathrm{i}\delta(x-y)\langle 0|\delta(\phi^{\dagger}\phi)|0\rangle = 0,$$

so that $q^2 f(q^2) = 0$.

Graphs (f) and (g) each have a subdivergence which is cancelled by a graph of the form (b), which is zero. Their overall divergence must be cancelled by using the order g^2 term in Z in graph (a).

9.2 Spontaneously broken symmetry

To explain the renormalization of theories with spontaneously broken symmetry it will be sufficient to consider the case of the Goldstone model:

$$\begin{aligned}\mathscr{L} &= Z\partial_{\mu}\phi^{\dagger}\partial^{\mu}\phi + m^2\phi^{\dagger}\phi - g(\phi^{\dagger}\phi)^2/4 + \delta m^2\phi^{\dagger}\phi - \delta g(\phi^{\dagger}\phi)^2/4 \\ &= (\partial\phi_1)^2/2 + (\partial\phi_2)^2/2 + m^2(\phi_1^2+\phi_2^2)/2 - g(\phi_1^2+\phi_2^2)^2/16 \\ &\quad + \text{counterterms}. \end{aligned} \tag{9.2.1}$$

Here we have written the complex scalar field in terms of real fields: $\phi = (\phi_1 + \mathrm{i}\phi_2)2^{-1/2}$. The mass term is of the 'wrong sign'. This will result in spontaneous breaking of the symmetry under $\phi \to \phi \mathrm{e}^{-\mathrm{i}\omega}$. The Noether current for this symmetry is

$$j^{\mu} = \mathrm{i}Z\phi^{\dagger}\overleftrightarrow{\partial^{\mu}}\phi = Z(\phi_1\partial^{\mu}\phi_2 - \phi_2\partial^{\mu}\phi_1). \tag{9.2.2}$$

For small couplings the Euclidean functional integral is dominated by fields close to the minimum of the potential in (9.2.1). This is at

$$|\phi| = 2m/g^{1/2}. \tag{9.2.3}$$

The perturbation expansion amounts to a saddle point expansion about the minimum. It is set up by making the substitution

$$\phi_1 = \phi_1' + 2m/g^{1/2}, \tag{9.2.4}$$

to give

$$\begin{aligned}\mathscr{L} &= (\partial\phi_1')^2/2 + (\partial\phi_2)^2/2 - m^2\phi_1'^2 - g(\phi_1'^2+\phi_2^2)^2/16 \\ &\quad - mg^{1/2}\phi_1'(\phi_1'^2+\phi_2^2)/2 + \mathscr{L}_{\mathrm{ct}}, \end{aligned} \tag{9.2.5a}$$

where

$$\begin{aligned}\mathscr{L}_{\mathrm{ct}} &= -\delta g(\phi_1'^2+\phi_2^2)^2/16 - \delta g m g^{-1/2}\phi_1'(\phi_1'^2+\phi_2^2)/2 \\ &\quad - \phi_1'^2(3m^2\delta g/g - \delta m^2)/2 - \phi_2^2(m^2\delta g/g - \delta m^2)/2 \\ &\quad - 2\phi_1' m g^{-1/2}(m^2\delta g/g - \delta m^2) + (Z-1)(\partial\phi_1'^2 + \partial\phi_2^2)/2. \end{aligned} \tag{9.2.5b}$$

The idea of making this perturbation expansion is that in the functional integral we impose a boundary condition that fixes the phase of the field at ∞. By the symmetry we may make this phase real, without loss of generality. In three or more space-time dimensions, fields that have a different phase over a large region have an action so much larger that quantum fluctuations cannot destroy the boundary condition. Then ϕ_1 is forced to have a real vacuum expectation value close to $2m/g^{1/2}$.

In setting up the perturbation expansion we have tadpole graphs like Fig. 9.2.1. These generate a vacuum expectation value for ϕ_1'

$$\langle 0|\phi_1'|0\rangle = \delta v$$

that starts at order $g^{1/2}$. It means that ϕ_1 has vacuum expectation value $2mg^{-1/2} + \delta v$. There are then graphs like Fig. 9.2.2, where the tadpoles appear as subgraphs. It is possible to recast the Feynman rules by writing $\phi_1 = \phi_1'' + 2mg^{-1/2} + \delta v$ and requiring ϕ_1'' to have zero vacuum expectation value. A better practical approach is to impose $\delta v = 0$ as a renormalization condition on δm^2.

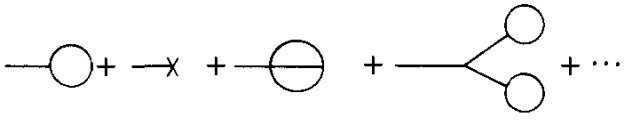


Fig. 9.2.1. Graphs for $\langle 0|\phi_1'|0\rangle$.

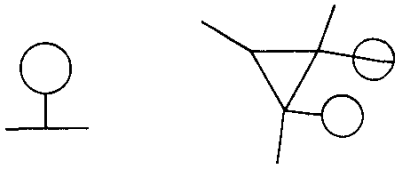


Fig. 9.2.2. Graphs containing tadpoles as subgraphs.

If we start with the theory (9.0.1) without spontaneous symmetry breaking and vary m^2 until it is negative, then we should pass through a phase transition and thereby reach the Goldstone model (9.2.1). There must be an actual phase transition because $\langle 0|\phi|0\rangle$ is exactly zero in the phase with unbroken symmetry. Since this expectation value is non-zero in the Goldstone phase there must be non-analyticity of the theory as a function of m^2.

Now, we must renormalize the theory: the continuation in the renormalized mass m^2 is sensible only if the counterterms are the same functions of m^2 and g in the two phases. It is sensible to use a mass-independent renormalization prescription, for then the dependence on m^2 of the counterterms is the simplest possible. We will prove the following:

(1) Renormalization of the Goldstone phase is accomplished by using only symmetric counterterms in the Lagrangian (9.2.1).
(2) The dimensionless counterterms $Z-1$, δg, and $\delta m^2/m^2$ can be chosen to be the same as in the phase of unbroken symmetry (the so-called Wigner phase).
(3) The current given by (9.2.2) is finite just as it is in the Wigner phase. Since the bare Lagrangian is invariant under $\phi \to \phi e^{-i\omega}$, Ward identities are valid and from them Goldstone's theorem follows, that the physical mass of ϕ_2 is exactly zero.

We must also discuss the choice of a practical renormalization prescription.

9.2.1 *Proof of invariance of counterterms*

We will do perturbation theory by choosing the free Lagrangian

$$\mathscr{L}_0 = \partial\phi_1'^2/2 + \partial\phi_2^2/2 - m^2\phi_1'^2, \tag{9.2.6}$$

and the basic interaction

$$\mathscr{L}_b = -g(\phi_1'^2 + \phi_2^2)/16 - mg^{1/2}\phi_1'(\phi_1'^2 + \phi_2^2)/2. \tag{9.2.7}$$

The counterterms are given the form (9.2.5*b*) and δg, $\delta m^2/m^2$, and Z are given the same values as in the unbroken theory with a mass-independent renormalization scheme. We will prove that these counterterms are sufficient to make the broken-symmetry theory finite.

Some of the interaction vertices are the same as in the Wigner phase. The others are obtained by substituting $2m/g^{1/2}$ for ϕ_1. Therefore graphs involving the extra vertices are obtained by erasing external ϕ_1 lines on symmetric graphs. Examples are shown in Fig. 9.2.3. The only complication is that mass terms generated from the basic interaction go into the free rather than the interaction Lagrangian. This is the sole source of complications in our proof.

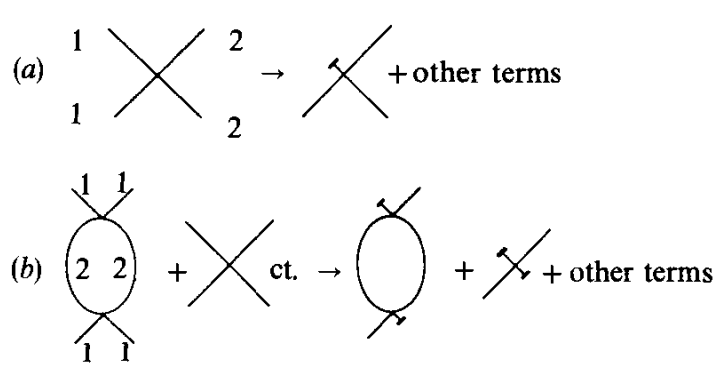


Fig. 9.2.3. Generation of graphs in theory with spontaneously broken symmetry from graphs in the symmetric theory.

To relate the counterterms to those in the unbroken theory let us write the free propagators as follows:

$$\phi_1' : \frac{i}{p^2 - 2m^2} = \frac{i}{p^2 - M^2} + \frac{i(2m^2 - M^2)}{(p^2 - M^2)^2} + \frac{i(2m^2 - M^2)^2}{(p^2 - M^2)^2(p^2 - 2m^2)},$$

$$\phi_2 : \frac{i}{p^2} = \frac{i}{p^2 - M^2} + \frac{-iM^2}{(p^2 - M^2)^2} + \frac{iM^4}{(p^2 - M^2)^2 p^2}. \tag{9.2.8}$$

Here M^2 is a arbitrary parameter. We substitute (9.2.8) for every line in a graph.

Suppose we substitute the first term on the right of (9.2.8) for every line of a basic graph which has only four-point basic vertices. Then we obtain a graph in the symmetric theory with mass M.

The difference between these symmetric graphs and the true theory is given by:

(1) graphs with one or more three-point vertices,
(2) graphs with the second or third term on the right of (9.2.8) substituted for one or more propagators.

In either case the degree of divergence is reduced. Now the maximum degree of divergence is two. So substitution of the third term in (9.2.8) always makes a graph overall convergent. We are allowed at most one substitution of the second term.

Let us now suppose that all graphs with fewer than N loops are successfully renormalized by our symmetric counterterms. We will prove inductively that all N-loop graphs are renormalized. The induction starts because tree graphs need no renormalization. We decompose the mass counterterm in $\mathscr{L}_{ct}$ as

$$-\tfrac{1}{2}\phi_1'^2[3m^2\delta g/g + (Z_m - 1)(-m^2 - M^2) + (Z_m - 1)M^2]$$
$$-\tfrac{1}{2}\phi_2^2[m^2\delta g/g + (Z_m - 1)(-m^2 - M^2) + (Z_m - 1)M^2]. \tag{9.2.9}$$

Here $Z_m = (m^2 + \delta m^2)/m^2$ is the mass renormalization factor.

After substitution of (9.2.8) for each propagator in a basic 1PI graph with N-loop all subdivergences are cancelled by counterterms of lower order, according to the inductive hypothesis. We are left with the following overall divergences:

(1) Logarithmically divergent graphs for the four-point function with all propagators set to $i/(p^2 - M^2)$ and with only four-point vertices. Such graphs have an overall divergence independent of M which is removed by counterterms in δg for the symmetric theory. No other 1PI graph for the four-point function has an overall divergence.

(2) Self-energy graphs with four-point vertices only and with propagators $i/(p^2 - M^2)$. Field renormalization and the $(Z_m - 1)M^2$ terms in (9.2.9) renormalize these, again exactly as in the symmetric theory.

(3) Self-energy graphs as in (2) but with one propagator replaced by a second term in (9.2.8). In the numerators of (9.2.8) we write

$$2m^2 - M^2 = 3m^2 - (m^2 + M^2) = (2m/g^{1/2})^2(3g/4) - (m^2 + M^2),$$
$$-M^2 = m^2 - (m^2 + M^2) = (2m/g^{1/2})^2(g/4) - (m^2 + M^2). \qquad (9.2.10)$$

The terms with $-(m^2 + M^2)$ are renormalized by the $(Z_m - 1)$ $(-m^2 - M^2)$ parts of the mass counterterms. They correspond to the effect of differentiating the self-energy graphs with respect to M^2. The other terms in (9.2.10) we will regard as an insertion of a four-point vertex on a line when two ϕ_1 fields are replaced by $2m/g^{1/2}$. These terms are considered under (4).

(4) Graphs of classes (1) and (2) in which one or more external ϕ_1 fields are deleted and replaced by $2m/g^{1/2}$. Examples are Fig. 9.2.3(*b*) and Fig. 9.2.4. The same replacement generates the counterterm Lagrangian (9.2.5*b*) from the symmetric theory, so we have counterterms for them.

This completes the proof.

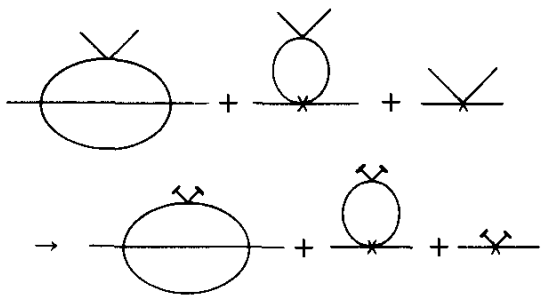


Fig. 9.2.4. Generation of graphs with loops in theory with spontaneously broken symmetry from graphs in the symmetric theory.

9.2.2 *Renormalization of the current*

The same procedure shows that the current

$$j^\mu = Z(\phi_1'\partial^\mu\phi_2 - \phi_2\partial^\mu\phi_1') + 2Zmg^{-1/2}\partial^\mu\phi_2 \qquad (9.2.11)$$

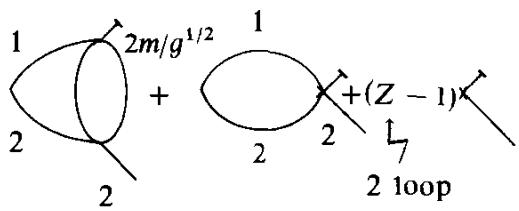


Fig. 9.2.5. Renormalization of current in spontaneously broken theory.

has finite Green's functions. Note that the term $2Zmg^{1/2}\partial^\mu\phi_2$ contains the counterterms that renormalize graphs like Fig. 9.2.5.

The Ward identities are then true and involve finite quantities. A typical case is

$$\begin{aligned}\partial_\mu\langle 0|Tj^\mu(x)\phi_2(y)|0\rangle &= i\langle 0|\delta\phi_2(y)|0\rangle \\ &= -i\langle 0|\phi_1(y)|0\rangle \\ &= -i(2m/g^{1/2}+\delta v). \end{aligned} \tag{9.2.12}$$

By multiplying by the inverse propagator for ϕ_2 and going to momentum space, we find

$$p_\mu\Gamma^\mu_{j,2}(p) = (2m/g^{1/2}+\delta v)(i/G_{22}(p^2)), \tag{9.2.13}$$

where $\Gamma^\mu_{j,2}$ is the set of graphs for $\langle 0|Tj^\mu\phi_2|0\rangle$ that are 1PI in ϕ_2. This is illustrated in Fig. 9.2.6. Since $\Gamma^\mu \propto p^\mu$ as $p^2\to 0$, (9.2.13) implies that G_{22}^{-1} has a zero at $p^2=0$, in other words that ϕ_2 is massless to all orders of perturbation theory. This is the Goldstone theorem (Goldstone, Salam & Weinberg (1962)).

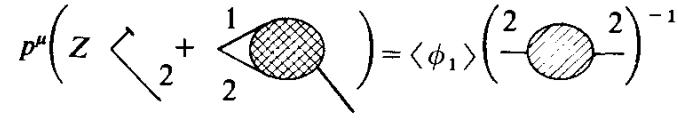


Fig. 9.2.6. The Ward identity that implies Goldstone's theorem.

9.2.3 *Infra-red divergences*

Individual graphs with a self-energy insertion on a ϕ_2 line have infra-red divergences. Such a graph is illustrated in Fig. 9.2.7, and the divergence comes from the region where the momentum k on the ϕ_2 line is close to zero:

$$\int_{k\sim 0} d^4k \frac{1}{(k^2)^2}.$$

If uncancelled, this divergence indicates that the self-energy shifts the mass to a value other than zero. But the Goldstone theorem tells us that the self-energy is zero at $k=0$. So the infra-red divergence cancels against divergences in other graphs of the same order.

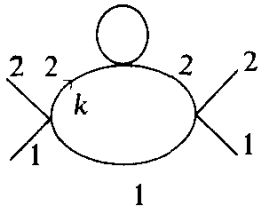


Fig. 9.2.7. Graph with infra-red divergence.

9.3 Renormalization methods

One of the practical problems that arises in making calculations in a theory with spontaneous symmetry breaking is to find the most convenient renormalization prescription. Fundamentally, there is no problem, for all renormalization prescriptions are related by renormalization-group transformations, and are therefore equally good. But, in practice, choice of one prescription over another can save some labor. The problems become particularly acute in gauge theories of weak interactions (Beg & Sirlin (1982)).

Among the issues to be considered in choosing a renormalization prescription are:

(1) If we ignore higher-order corrections, then some parameters are equal to quantities, like particle masses, that are easily measurable. It is often convenient to impose exact equality as a renormalization condition.
(2) One must treat tadpole graphs. Their effect is to provide an additional shift δv in the vacuum expectation value of the field. Leaving these graphs as they are considerably increases the number of graphs contributing to a given Green's function. Shifting the field by δv gives many extra terms in the formulae for the coefficients in (9.2.5*a*) and (9.2.5*b*). One can impose $\delta v = 0$ as a renormalization condition, at the expense of removing the simple connection between the phases of broken and unbroken symmetry (as was exploited in Section 9.2).
(3) It is necessary to relate calculations done by different people. Direct comparisons can be made only if the same renormalizations are used. It is evidently useful to agree on a standard.
(4) If the coupling is not very small or if there occur very large ratios of masses and momenta, then one must choose a renormalization prescription with the ability to remove the large logarithms.

One approach is to use dimensional regularization with minimal subtraction. Graphs can be renormalized by the forest formula. At one-loop order this amounts to subtraction of the pole part from each 1PI graph. We can do this without regard to the symmetry relations between counterterms for different Green's functions. Since the counterterms have the pure-pole form, these relations are automatically satisfied.

Another approach is to compute Z, δg, and δm^2 by three renormalization conditions imposed on some of the 1PI Green's functions in the broken-symmetry phase. Then the values of counterterms for other Green's functions are computed from (9.2.5*b*). It is convenient to determine $m^2\delta g/g - \delta m^2$ by requiring $\delta v = 0$, i.e., $\langle 0|\phi_1|0\rangle = 2mg^{-1/2}$ exactly. Then

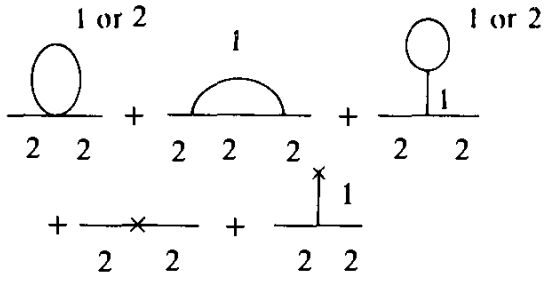


Fig. 9.3.1. Renormalization of self-energy at one-loop order.

the mass counterterm for ϕ_2 forces the 1PI self-energy of ϕ_2 to be exactly zero when $k^2 = 0$. The one-loop graphs are shown in Fig. 9.3.1.

Both of these approaches require explicit computation of the values of Z, δg, and δm^2 to find the values of counterterms for the various Green's functions. It is also possible (Symanzik (1970a)) to use the Ward identities to generate renormalization conditions for all divergent 1PI Green's functions from the three basic conditions. These conditions are simple if the three basic conditions are imposed at zero external momentum.

9.3.1 *Generation of renormalization conditions by Ward identities*

The general Ward identity is

$$\frac{\partial}{\partial x^\mu}\langle 0|Tj^\mu(x)\prod_{j=1}^{N}\phi_{n_j}(y_j)|0\rangle$$
$$= -\mathrm{i}\prod_{j=1}^{N}\delta(x-y_j)\langle 0|T\delta\phi_{n_j}(x)\prod_{i\neq j}\phi_{n_i}(y_i)|0\rangle. \tag{9.3.1}$$

In the Goldstone model, the labels n_i take the values $1'$ or 2, and we have $\delta\phi_1' = \phi_2$ and $\delta\phi_2 = -\phi_1 = -(\phi_1' + 2m/g^{1/2})$. For simplicity we impose the condition $\langle 0|\phi_1'|0\rangle = 0$. Suppose we have obtained renormalization conditions valid up to $l-1$ loops. We will now find the appropriate conditions for l-loop graphs.

The case $N = 1$ was given in (9.2.12) and (9.2.13), and in Fig. 9.2.6. We saw that one renormalization condition on the self-energy of ϕ_2 is that it is zero at $p^2 = 0$. Another condition on the derivative can be chosen arbitrarily, corresponding to the freedom to multiply Z by a finite factor. Suppose we choose to make the residue of the Goldstone pole equal to unity, and we choose to make $\delta v = 0$. Then we also obtain the renormalization condition on the Green's function $\langle 0|Tj^\mu(x)\phi_2(y)|0\rangle$ of the current with ϕ_2. The condition is that it is equal to its lowest-order value at $p = 0$. This condition is equivalent to making the counterterm equal to

$$2mg^{-1/2}(Z-1)\partial^\mu\phi_2,$$

as required if j^μ is to be the Noether current.

Similarly we may treat the case $N = 2$:

$$\begin{aligned}\partial_\mu \langle 0|Tj^\mu(x)\phi_1'(y)\phi_2(z)|0\rangle &= -i\delta(x-y)\langle 0|T\phi_2(y)\phi_2(z)|0\rangle \\ &\quad + i\delta(x-z)\langle 0|T\phi_1'(y)\phi_1'(z)|0\rangle,\end{aligned} \tag{9.3.2}$$

where we used $\langle 0|\phi_1'|0\rangle = 0$. In terms of 1PI graphs in momentum space this gives Fig. 9.3.2. After use of Fig. 9.2.6 we find Fig. 9.3.3.

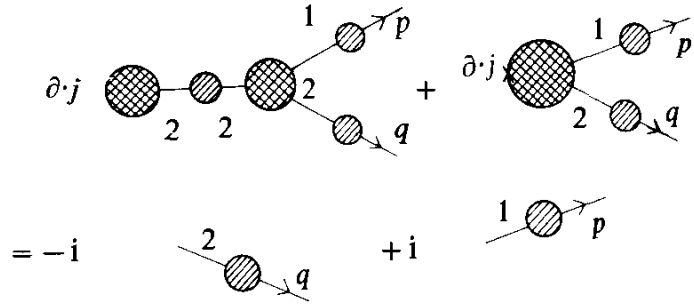


Fig. 9.3.2. Ward identity for two-point function of j^μ.

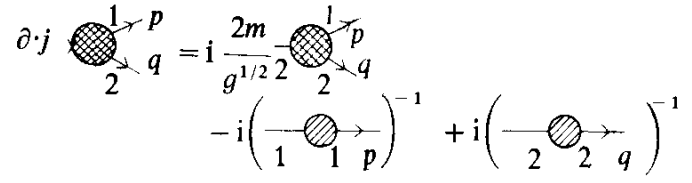


Fig. 9.3.3. Result of multiplying Fig. 9.3.2 by two inverse propagators.

We now set $p + q = 0$ to eliminate the left-hand side. This gives

$$\begin{aligned}0 &= \frac{2m}{g^{1/2}} i\Gamma_{212}(0,p,-p) - [p^2 - 2m^2 - \Sigma_1(p^2)] + [p^2 - \Sigma_2(p^2)] \\ &= \frac{2m}{g^{1/2}} i\Gamma_{212}(0,p,-p) + 2m^2 - \Sigma_1(p^2) - \Sigma_2(p^2).\end{aligned} \tag{9.3.3}$$

Here Σ_1 and Σ_2 are self-energies and Γ_{212} is the 1PI Green's function for two ϕ_2 fields and one ϕ_1'. We choose a mass renormalization condition for Σ_1, say $\Sigma_1(0) = 0$. Since we already know that $\Sigma_2(0) = 0$, this tells us that the renormalization condition on Γ_{212} is

$$\Gamma_{212}(0,0,0) = -img^{-1/2} = \text{lowest-order value}. \tag{9.3.4}$$

Since graphs for Γ_{212} are at worst logarithmically divergent we know $\Gamma_{212}(p,q,r)$ completely at this order. From (9.3.3) we can now determine $\Sigma_1(p^2)$. But the calculation of $\Sigma_1(p^2)$ from its graphs is already fixed except for a renormalization condition that determines the value of the field-

strength counterterm. So (9.3.3) gives us the renormalization in such a way that the counterterm is $-(Z-1)p^2$.

Similar arguments may be applied to give renormalization conditions for all the remaining 1PI Green's functions that have overall divergences. They are easiest to express in terms of Ward identities for 1PI Green's functions. (See Lee (1976), and references therein.)

The structure of these arguments generalizes what would be done in the unbroken phase. For example, (9.3.2) integrated over x would give $\Sigma_1 = \Sigma_2$ in this phase. This condition would say that the counterterms for Σ_1 and Σ_2 are equal. But in the Goldstone phase this is not so. Integrating over x is equivalent to setting the momentum at the vertex for the current to zero. The derivative with respect to x gives a factor of this momentum, but since there is a pole $1/p^2$ the right-hand side of (9.3.2) is not zero. The argument that we had to use is more complicated.

10

Operator-product expansion

In this chapter we will investigate two closely related problems. We work with ϕ^4 theory in $d = 4$ dimensions and consider a time-ordered product of two fields, $T\phi(x)\phi(0)$, together with its Fourier transform

$$T\tilde{\phi}(q)\phi(0) = \int \mathrm{d}^4 x \mathrm{e}^{\mathrm{i}q\cdot x} T\phi(x)\phi(0).$$

(It is easiest to work with time-ordered products. The methods work with any pair of operators $TA(x)\,B(0)$ in any theory.)

The first problem is to ask how $T\phi(x)\phi(0)$ behaves as $x^\mu \to 0$. If the theory were totally finite then the result would just be $\phi^2(0)$. However, there are ultra-violet divergences that prevent the product from existing, so the limit does not exist. It was the idea of Wilson (1969) that $\phi(x)\phi(0)$ should behave like a singular function of x times the renormalized $[\phi^2]$ operator, as $x \to 0$. The full result is that we have an expansion of the form

$$T\phi(x)\phi(0) \sim \sum_{\mathcal{O}} C_{\mathcal{O}}(x^\mu)[\mathcal{O}(0)] \tag{10.0.1}$$

as $x \to 0$. Here the sum is over a set of local renormalized composite fields $[\mathcal{O}]$ and the $C_{\mathcal{O}}(x)$'s are c-number functions. This formula, or one of its generalizations, is called an operator product expansion (OPE), and the coefficients $C_{\mathcal{O}}$ are often called Wilson coefficients. Corrections to (10.0.1) are smaller by a power of x^2 than the terms given.

The second problem we wish to treat is the behavior of $T\tilde{\phi}(q)\phi(0)$ as $|q^2| \to \infty$. More precisely we will consider the momentum-space Green's function

$$\tilde{G}_{N+2} = \langle 0|T\tilde{\phi}(q)\phi(0)\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle, \tag{10.0.2}$$

when $q^\mu \to \infty$ along a fixed direction with $p_1, \ldots, p_N$ fixed. In other words we scale the invariants $q^2 \to \kappa^2 q^2$, $p_i \cdot q \to \kappa p_i \cdot q$. There is an operator product expansion

$$\tilde{G}_{N+2} \sim \sum_{\mathcal{O}} \tilde{C}_{\mathcal{O}}(q)\langle 0|T\mathcal{O}(0)\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle. \tag{10.0.3}$$

The relation between the coordinate-space and momentum-space expan-

sions is elementary. Let us take the Fourier transform from x to q of a momentum-space Green's function of $T\phi(x)\phi(0)$. Then the large-q behavior is dominated by the singularities in x-space. The only relevant singularity is at $x = 0$. So $\tilde{C}_\phi(q)$ is the large-q part of the Fourier transform of $C_\phi(x)$. Conversely if one Fourier transforms $\tilde{G}_{N+2}$ to get

$$\langle 0|T\phi(x)\phi(0)\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle = \int \frac{d^4q}{(2\pi)^4} e^{-iq\cdot x}\tilde{G}_{N+2}, \qquad (10.0.4)$$

then the limit $x \to 0$ fails to exist if $\tilde{G}_{N+2}$ falls only as $1/q^4$ or slower as $q \to \infty$.

Thus knowing the large-q behavior is equivalent to knowing the singular small-x behavior, but the coordinate-space expansion also includes information on the leading non-singular part of the small-x region.

These expansions have a number of uses, particularly in an asymptotically free theory. There the perturbation theory when improved by the renormalization group gives an effective method of computing the Wilson coefficients. Among the uses are the following:

(1) The expansion (10.0.1) in coordinate space provides a definition of renormalized composite operators that does not involve any regularization (Brandt (1967)).

(2) Although there is no physically important process which directly uses the limit taken in the momentum-space expansion (10.0.3), it is used indirectly for deep-inelastic scattering of a lepton on a hadron. This involves a matrix element of the form

$$\langle p|\tilde{j}(q)j(0)|p\rangle.$$

Here $q^2 \to -\infty$, but with the ratio $q^2/q\cdot p$ fixed instead of $q^2/q\cdot p^2$ fixed. A dispersion relation relates this case to the limit used in (10.0.3), so the OPE is used indirectly, as we will see in Chapter 14.

(3) The form and the method of proof of short-distance operator-product expansion can be generalized to handle many interesting high-energy scattering processes. (See Buras (1981) and Mueller (1981) for a review.) The results in the present chapter form a prototype for these other results.

10.1 Examples

10.1.1 *Cases with no divergences*

We will mainly restrict our attention to Green's functions of $\phi(x)\phi(0)$ in which both $\phi(x)$ and $\phi(0)$ are connected to other external lines. This is the

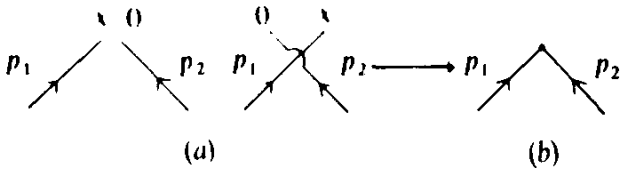


Fig. 10.1.1. Lowest-order graphs for operator-product expansion of $\phi(x)\phi(0)$.

case that is relevant to most applications. Our theory will be ϕ^4 theory

$$\mathscr{L} = \partial\phi^2/2 - m^2\phi^2/2 - g\phi^4/24 + \text{counterterms}. \tag{10.1.1}$$

First consider the tree graphs Fig. 10.1.1(a) for

$$\langle 0|T\phi(x)\phi(0)\tilde{\phi}(p_1)\tilde{\phi}(p_2)|0\rangle. \tag{10.1.2}$$

These give

$$\frac{\mathrm{i}}{p_1^2 - m^2}\frac{\mathrm{i}}{p_2^2 - m^2}[\exp(-\mathrm{i}p_1\cdot x) + \exp(-\mathrm{i}p_2\cdot x)]. \tag{10.1.3}$$

Expansion in a power series about $x = 0$ gives

$$\frac{\mathrm{i}}{p_1^2 - m^2}\frac{\mathrm{i}}{p_2^2 - m^2}[2 - \mathrm{i}(p_1 + p_2)\cdot x - (p_1\cdot x^2 + p_2\cdot x^2)/2 + \cdots]. \tag{10.1.4}$$

This is equivalent to the replacement

$$T\phi(x)\phi(0) = \phi^2(0) + \tfrac{1}{2}x^\mu\partial_\mu\phi^2 + \tfrac{1}{2}x^\mu x^\nu\phi\partial_\mu\partial_\nu\phi + \cdots, \tag{10.1.5}$$

as illustrated in Fig. 10.1.1(b). This equation has the form of the operator-product expansion (10.0.1).

Thus the operator-product expansion in this case (free-field theory) is really a Taylor expansion of $\phi(x)$ about $x = 0$. The power of x in each term is just such that no dimensional coefficients are needed:

$$C_{\mathscr{O}}(x) = \text{constant} \times |x|^a, \text{ with } a = \dim(\mathscr{O}) - \dim[\phi(x)\phi(0)]. \tag{10.1.6}$$

This result also correctly gives the power-law behavior in the presence of renormalizable interactions, as we will see. But there will also be logarithmic corrections.

A feature which does not appear to survive inclusion of interactions is that the series on the right of (10.1.4) is convergent and sums to give $T\phi(x)\phi(0)$.

Consider next the graphs of Fig. 10.1.2 for the four-point function of $T\phi(x)\phi(0)$. The important factor comes from the lines carrying momentum q:

$$\mathrm{i}g^2\int\frac{\mathrm{d}^4q}{(2\pi)^4}\frac{\{\exp(-\mathrm{i}q\cdot x) + \exp[\mathrm{i}(q - p_A - p_B)\cdot x]\}}{(q^2 - m^2)[(p_A - q)^2 - m^2][(q - p_A - p_B)^2 - m^2]}. \tag{10.1.7}$$

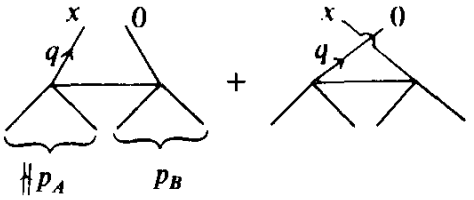


Fig. 10.1.2. Higher-order graphs for operator-product expansion of $\phi(x)\phi(0)$.

We now expand the integrand in powers of x to obtain

$$(10.1.7) \sim ig^2 \int \frac{d^4q}{(2\pi)^4} \frac{2 - i(p_A + p_B)\cdot x}{(q^2 - m^2)[(p_A - q)^2 - m^2][(q - p_A - p_B)^2 - m^2]} + \cdots. \tag{10.1.8}$$

These first two terms are just those we would expect from (10.1.5). But the higher terms have at least two extra powers of q in the numerator and are therefore ultra-violet divergent. The divergences are those of the Green's function of the composite operators. They indicate that modification of the higher terms of the expansion is needed. For example, the behavior of the coefficient of $\phi\partial_\mu\partial_\nu\phi$ is modified by a logarithm of x.

Similar modifications will be needed for the coefficient of ϕ^2, when we consider higher-order corrections. Therefore we will find it convenient just to restrict our attention to the leading-power behavior, corresponding to the ϕ^2 term in (10.1.5).

10.1.2 *Divergent example*

Aside from trivial propagator corrections the contribution of order g to the two-point function of $T\phi(x)\phi(0)$ is given by Fig. 10.1.3(*a*), which gives

$$\frac{i^2}{(p_1^2 - m^2)(p_2^2 - m^2)} ig \int \frac{d^4q}{(2\pi)^4} \frac{e^{-iq\cdot x}}{(q^2 - m^2)[(q - p_1 - p_2)^2 - m^2]}. \tag{10.1.9}$$

When $x \to 0$ the integral diverges logarithmically. This is a symptom of the fact that there are two important regions of q that contribute. The first is

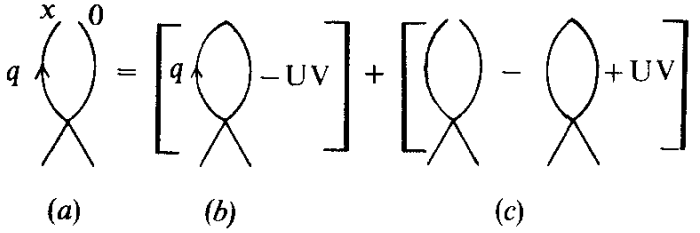


Fig. 10.1.3. Graph for operator-product expansion of $\phi(x)\phi(0)$ with divergences for the operator.

where q is finite as $x \to 0$; the contribution is correctly given by replacing $T\phi(x)\phi(0)$ by $\phi^2(0)$ in the graph. The second region is where q becomes large, up to $O(1/x)$ as $x \to 0$; in this region the interaction vertex in coordinate space is close to x and 0.

In the second region the loop is confined to a small region in coordinate space. From the point of view of p_1 and p_2 the loop is a point. So we should be able to represent the contribution of this region by an extra term in the Wilson coefficient of ϕ^2:

$$\left.\begin{aligned} T\phi(x)\phi(0) &\sim C_{\phi^2}(x)[\phi^2], \\ C_{\phi^2} &= 1 + (g/16\pi^2)c_1(x^2). \end{aligned}\right\} \tag{10.1.10}$$

Let us now calculate $c_1(x^2)$.

Now the contribution of the first region is given by replacing $T\phi(x)\phi(0)$ by $\phi^2(0)$. However, this operator has an ultra-violet divergence. So let us add and subtract the renormalized Green's function of $[\phi^2]$, i.e.,

$$\langle 0|T[\phi^2(0)]\tilde{\phi}(p_1)\tilde{\phi}(p_2)|0\rangle, \tag{10.1.11}$$

to give the equation depicted by Fig. 10.1.3. The contribution of order 1 from the first region is entirely contained in the term (b) representing (10.1.11):

$$\frac{\mathrm{i}^2}{(p_1^2 - m^2)(p_2^2 - m^2)} \frac{g}{16\pi^2} \left\{ \int_0^1 \mathrm{d}x \ln \left[\frac{m^2 - (p_1 + p_2)^2 x(1-x)}{4\pi\mu^2} \right] + \gamma \right\}. \tag{10.1.12}$$

Here we have used minimal subtraction. The remainder, term (c), is

$$\frac{1}{(p_1^2 - m^2)(p_2^2 - m^2)} \frac{-\mathrm{i}g}{(2\pi)^4} \times$$
$$\times \left\{ \int \mathrm{d}^4 q \frac{\mathrm{e}^{\mathrm{i}q\cdot x} - 1}{(q^2 - m^2)[(q - p_1 - p_2)^2 - m^2]} \quad - \text{UV divergence} \right\}. \tag{10.1.13}$$

When $x \to 0$ the contribution from finite q is of order $|x|$. But there is a contribution of order 1 from large q: this is the contribution to the original graph minus whatever is taken care of by graph (b).

We can identify $c_1(x^2)$ in (10.1.10) as the $x \to 0$ behavior of the curly bracket factor of (10.1.13) (aside from a normalization factor), since to lowest order

$$\langle 0|T\phi^2\tilde{\phi}(p_1)\tilde{\phi}(p_2)|0\rangle = \frac{-2}{(p_1^2 - m^2)(p_2^2 - m^2)}.$$

Now the leading-power behavior of the curly-bracket factor is independent of p_1, p_2 and m. This is easily seen by differentiating with respect to any of

these variables. The result is a convergent integral which goes to zero like a power of x when $x \to 0$. So we may define $c_1(x)$ by setting $m = p_1 = p_2 = 0$:

$$c_1(x) = \frac{1}{2\pi^2}\left\{\frac{\mathrm{i}}{(2\pi\mu)^{d-4}}\int \mathrm{d}^d q \frac{(\mathrm{e}^{\mathrm{i}q\cdot x} - 1)}{(q^2)^2} + \frac{2}{d-4}\right\}. \tag{10.1.14}$$

The integral is easily done by using

$$1/(q^2)^2 = \int_0^\infty \mathrm{d}z z \mathrm{e}^{-z(-q^2)}$$

to turn it into a Gaussian form, with the result

$$c_1(x) = \tfrac{1}{2}[\gamma + \ln(-\pi^2\mu^2x^2)]. \tag{10.1.15}$$

10.1.3 *Momentum space*

We now Fourier transform Fig. 10.1.3 to obtain the $O(g)$ contribution to

$$\langle 0|T\tilde{\phi}(q)\phi(0)\tilde{\phi}(p_1)\tilde{\phi}(p_2)|0\rangle.$$

As $q^2 \to \infty$ we find

$$\text{Fig. } 10.1.3(a) \sim \frac{-1}{(p_1^2 - m^2)(p_2^2 - m^2)}\frac{\mathrm{i}g}{(q^2)^2} + O[1/(q^2)^3]. \tag{10.1.16}$$

This gives a contribution to the term in the operator-product expansion (10.0.3) with $\mathcal{O} = \phi^2$. The coefficient is

$$\tilde{C}_{\phi^2}(q^2) = \frac{\mathrm{i}g}{2(q^2)^2} + O(g^2), \tag{10.1.17}$$

which is just the Fourier transform of the order g term in the coordinate-space expansion, $gc_1(x)/(16\pi^2)$. The g^0 term in the coordinate-space coefficient is independent of x, so that it gives a $\delta^{(4)}(q)$ in momentum space, and hence nothing at large q^2.

10.1.4 *Fig.* 10.1.3 *inside bigger graph*

The expansion (10.0.1) or (10.0.3) indicates that the same asymptotic behavior as $x \to 0$ (or as $q \to \infty$) is obtained independently of the Green's function considered. This happens because graphs like Fig. 10.1.3 can occur

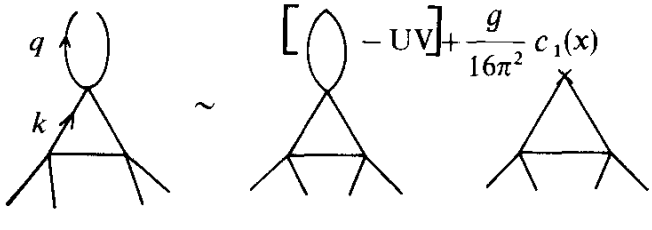


Fig. 10.1.4. Even higher-order graph for operator-product expansion of $\phi(x)\phi(0)$.

as subgraphs of graphs with more external lines. An example is given in Fig. 10.1.4.

10.2 Strategy of proof

First we will make precise the limits in which the operator-product expansion applies. If we are in Euclidean space (i.e., with imaginary time and energy) then there is really only one way in which we can take x^μ to zero or q^μ to infinity. However, in Minkowski space we can let $x^2 \to 0$ without each component going to zero, and we can let components of q^μ go to infinity without $q^2 \to \infty$. These cases are interesting physically. For example, the $q^\mu \to \infty$ limit with finite q^2 is the case of high-energy scattering. Much is known about these limits, but they are beyond the scope of this book.

We will prove the coordinate-space expansion (10.0.1) in the case that all components of x^μ go to zero with their ratios fixed. The corresponding momentum-space expansion (10.0.3) we will prove in the limit that all components of q^μ go to infinity with a fixed ratio, and with q^μ not light-like, so that $q^2 \to \infty$. These limits are essentially Euclidean.

Our proof will be in perturbation theory. The first step is to identify the regions of loop-momentum space that give the leading-power behavior in the $x \to 0$ or $q \to \infty$ limits. Then we generalize the arguments of the previous section, which applied to specific graphs.

The region of large q which we investigate in the momentum-space expansion (10.0.3) is precisely the one to which Weinberg's (1960) theorem applies. The theorem tells us to consider all subgraphs connected to the $\tilde{\phi}(q)$ and $\phi(0)$ in (10.0.2). For each such subgraph we let all its loop momentum be of order q, and count powers just as we did for UV divergences. The subgraph(s) with the largest power of q^2 correspond to the dominant regions of momentum space. Then as $q^2 \to \infty$, the complete graph is proportional to this power of q^2 times possible logarithms of q^2. Corrections are smaller by a power of q^2. Although Weinberg's result also tells us the highest power of $\ln(q^2)$ that appears, it is easier to determine this by first constructing the operator-product expansion and then applying renormalization-group methods (as in Section 10.5 below) to the coefficients.

Essentially the same method can be applied to obtain the short-distance behavior, (10.0.1). For example, in Fig. 10.1.4 we have leading contributions with q finite or with q large (of order $1/x$), but always with the lower loop momentum, k, finite. The leading power is $(x^2)^0$. The logarithm of x in

(10.1.15) for the corresponding Wilson coefficient comes from integrating over momenta intermediate between these two regions.

Suppose we have large momentum confined to a subgraph Γ. Then the power of q (for the momentum-space expansion) is exactly the dimension of Γ, since our theory is renormalizable, with dimensionless couplings. The leading power of q^2 comes from subgraphs with the largest dimension, i.e., subgraphs with the smallest possible number of external lines. This number is two (beyond $\tilde{\phi}(q)$ and $\phi(0)$), so that the leading power is $1/(q^2)^2$. The subgraphs have the form of the subgraph U in Fig. 10.2.1. In the ultra-violet subgraph U, all lines carry momentum of order q. This subgraph is 1PI in its lower two lines. All momenta in the infra-red subgraph I are finite.

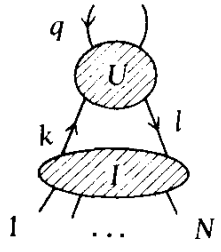


Fig. 10.2.1. General structure of leading regions of momentum space for N-point function of $\phi(x)\phi(0)$.

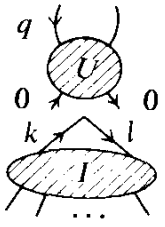


Fig. 10.2.2. Factorization of Fig. 10.2.1.

Now, to the leading power of q^2, the ultra-violet graph U is independent of the external momenta k and l flowing into it. Thus we may replace U by its value when $k = l = 0$ (and we may set the mass $m = 0$). We may also replace the infra-red subgraph I by an insertion of a vertex for $\phi^2/2$ in an N-point Green's function. This is illustrated in Fig. 10.2.2. This has the same structure as the operator-product expansion. But it should be emphasized that we are supposing that loop momenta are restricted to certain regions. These regions are not defined very precisely, and it is one of the tasks of the proof to remedy the impreciseness.

Schematically we have

$$\sum_{\text{graphs}} \tilde{G}_{N+2}(q, p_1, \ldots, p_N)$$

$$\sim \sum_U U(q, k=0, l=0) \sum_I \int_{\substack{\text{small } k \\ \text{small } l}} \mathrm{d}^4k\, \mathrm{d}^4l\, \delta(k + l - \textstyle\sum p_i) I(k, l, p_1, \ldots, p_N). \tag{10.2.1}$$

To construct the expansion we generalize the technique that we applied to Fig. 10.1.3. We consider each graph U that could appear in Fig. 10.2.1, but we do not restrict its momenta. It can occur as a subgraph of some graph for the complete Green's function. If all momenta in U are of order q and if all

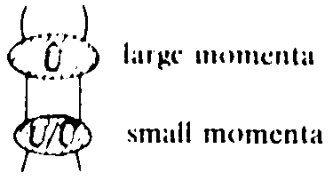


Fig. 10.2.3. A possible leading region for a subgraph of the form of U in Fig. 10.2.1.

momenta outside are small, then we get a leading contribution to the cross-section. We can also have a leading contribution where the large momenta occur inside a proper subgraph of U – as in Fig. 10.2.3. Suppose we subtract off all of these contributions. Then we integrate over all loop momenta of U and find that the result only gives a leading contribution when all its momenta are large. We therefore define the contribution of U to the Wilson coefficient as

$$C(U) = U - \text{subtractions for regions of form Fig. 10.2.3} \quad (10.2.2)$$

all evaluated at $k = l = m = 0$.

The resulting formula for $C(U)$ is very similar to that of the formula for renormalizing the ultra-violet divergences of a graph. In fact, as Zimmermann (1970, 1973b) explains, a good way to prove the operator-product expansion is to treat it exactly as a problem in renormalization. His method, used in the next section, is not to compute directly the Wilson coefficients but to define first a quantity which is a Green's function minus the leading terms in its operator-product expansion:

$$\tilde{G}_{N+2} - \sum_{\mathcal{O}} C_{\mathcal{O}} \tilde{G}_{\mathcal{O},N}.$$

This is constructed as a sum over graphs Γ for $\tilde{G}_{N+2}$. Each graph has subtracted from it not only counterterms to remove ultra-violet divergences but also counterterms to cancel the large-Q (or small-x) behavior. The result we call $R_W(\Gamma)$.

Now, the subtractions that remove the large-Q behavior are a sort of oversubtraction. So $R_W(\Gamma)$ is simply related to $R(\Gamma)$ in the style of a renormalization-group transformation. This transformation can then be written as the Wilson expansion, as we will see.

A disadvantage of Zimmermann's proof is that it uses BPHZ renormalization. He takes advantage of certain short-cuts available through the use of zero-momentum subtractions. We will choose not to take these shortcuts so that minimal subtraction can be applied to ultra-violet divergences. Our method of proof will be essentially the same as the one used for problems with large masses (Chapter 8).

The same techniques apply to the coordinate-space expansion. Here, the

momentum q is integrated over. When we do power-counting, large momenta are regarded as of order $1/x$ and small momenta as finite when $x \to 0$. The leading behavior again comes from graphs of the form of U in Fig. 10.2.1, and the power is $(x^2)^0$. There is a difference in the form of the possible graphs U that carry large momentum. All the graphs that we use in the momentum-space case are also used in coordinate space. But we can also have the graph consisting of the vertices for $\phi(x)$ and $\phi(0)$ and of nothing else.

10.3 Proof

We must now prove the operator product expansion. In ϕ^4 theory we consider the part of the Green's function,

$$\bar{G}_{N+2}(x, p_1, \ldots, p_N) = \langle 0 | T\phi(x)\phi(0)\tilde{\phi}(p_1)\cdots\tilde{\phi}(p_N)|0\rangle, \qquad (10.3.1)$$

in which each of $\phi(x)$ and $\phi(0)$ is connected to some of the other external lines. We will now scale x by a factor κ and construct a decomposition of the form:

$$\bar{G}_{N+2}(\kappa x, p_1, \ldots, p_N) = C(\kappa^2 x^2)\langle 0 | T_{\frac{1}{2}}[\phi^2](0)\tilde{\phi}(p_1)\cdots\tilde{\phi}(p_N)|0\rangle + r_{N+2}(\kappa x, p_1, \ldots, p_N). \qquad (10.3.2)$$

In every order of perturbation theory, the coefficient $C(\kappa^2 x^2)$ behaves like $(\kappa^2)^0$ times logarithms, when the scaling parameter κ goes to zero, while the remainder goes to zero like a power of κ.

Fourier transformation on x gives the result

$$\tilde{G}_{N+2}(q/\kappa, p_1, \ldots, p_N) = \tilde{C}(q^2/\kappa^2)\langle 0 | T_{\frac{1}{2}}[\phi^2]\tilde{\phi}(p_1)\cdots\tilde{\phi}(p_N)|0\rangle + \tilde{r}_{N+2}(q/\kappa, p_1, \ldots, p_N). \qquad (10.3.3)$$

When $q/\kappa \to \infty$, $\tilde{C}(q^2/\kappa^2)$ behaves like κ^4/q^4 times logarithms, while $\tilde{r}_{N+2}(q/\kappa, p_1, \ldots)$ is smaller by a power.

Our proof is given for a specific Green's function in a specific theory. However, it can easily be generalized. Features specific to a gauge theory will be pointed out in Chapter 12. The particular case of QCD with the application to deep-inelastic scattering will be treated in Chapter 14.

10.3.1 *Construction of remainder*

We consider the set of graphs for G_{N+2}. For each graph Γ, we will construct its contribution $r(\Gamma)$ to the remainder $\tilde{r}_{N+2}$. Each graph Γ we consider to be

an unrenormalized (but regulated) graph containing only basic interaction vertices. These are derived from

$$\mathscr{L}_{\text{basic}} = \partial\phi^2/2 - m^2\phi^2/2 - \mu^{4-d}g\phi^4/24,$$

where g and m are the renormalized coupling and mass.

As usual the renormalized value of the graph is $R(\Gamma)$, which is Γ, plus a series of counterterm graphs to cancel its ultra-violet divergences. For our proof we will use a renormalization prescription in which the theory is finite when the renormalized mass m is set to zero. The renormalized value $R(\Gamma)$ is then the contribution of Γ to the Green's function $\tilde{G}_{N+2}$.

The remainder term $r(\Gamma)$ is equal to Γ plus a somewhat different series of counterterm graphs. These counterterms will be constructed so that they cancel not only the ultra-violet divergences but also the leading $x \to 0$ or $q \to \infty$ behavior of Γ.

Now $r(\Gamma)$ is in effect a oversubtracted form of Γ. The oversubtractions are of the form of an insertion of the operator $[\phi^2]$ times a coefficient. Thus $R(\Gamma) - r(\Gamma)$ is the Wilson expansion, i.e., the first term on the right of (10.3.2) or (10.3.3).

The coefficient $C(x^2)$ (or $\tilde{C}(q^2)$) depends on the coupling g and on the renormalization mass μ. It must be independent of all the momenta. In order to be able to neglect m in the ultra-violet limit $x \to 0$ or $q \to \infty$, we must use a renormalization prescription in which the counterterms do not become infinite when $m \to 0$ (with fixed regulator). For concreteness we will use minimal subtraction in what follows.

In order to define $r(\Gamma)$, let us recall the definition of the ordinary renormalization $R(\Gamma)$. This starts from the fact that the divergences of Γ come from regions of loop momenta where all lines in some set of 1PI subgraphs carry a momentum that approaches infinity. We label each region by the subgraph consisting of all the lines carrying large momentum. Then

$$R(\Gamma) = \Gamma - \sum_{\gamma} C_{\gamma}(\Gamma). \tag{10.3.4}$$

The sum is over all subgraphs γ of Γ, and $C_{\gamma}(\Gamma)$ is essentially Γ with the subgraph γ replaced by its large-momentum divergence. We define $C_{\gamma}(\Gamma)$ to be non-zero only if γ is a disjoint union of one or more 1PI graphs $\gamma_1, \ldots, \gamma_n$. In that case each γ_i is replaced by a counterterm vertex $C(\gamma_i)$, which is the divergent part of γ_i. To avoid double-counting, the subdivergences are subtracted off first:

$$C(\gamma_i) = \mathscr{P}\left[\gamma_i - \sum_{\delta \subsetneq \gamma} C_{\delta}(\gamma_i)\right]. \tag{10.3.5}$$

Here '$\mathscr{P}$' denotes 'pole part at $d=4$' and the sum is over all proper subgraphs δ of γ_i. As usual, if we use some other renormalization prescription than minimal subtraction the operator $\mathscr{P}$ is replaced by the operator appropriate to the renormalization prescription.

The definition of the remainder $r(\Gamma)$ is almost identical to the definition of $R(\Gamma)$. Now, the leading short-distance (i.e., $x \to 0$) behavior of Γ comes from the following regions:

(1) where the momentum q is finite,
(2) where q gets large and the momenta in a graph of the form of U in Fig. 10.2.1 also get large.

Further leading contributions come from regions where in addition momenta get infinite in some set of divergent 1PI graphs. These extra contributions correspond to the ultra-violet divergences. In the momentum-space expansion (10.3.3) the same regions are leading except for the region of finite q.

We define $r(\Gamma)$ to be Γ with all ultra-violet divergences subtracted and then with all the leading small-x behavior subtracted:

$$r(\Gamma) = R(\Gamma) - \sum_{\delta} \sum_{\gamma} L_{\gamma \cup \delta}(\Gamma). \tag{10.3.6}$$

Here the sum over δ is over all graphs of the form of U in Fig. 10.2.1 and the sum over γ is over all subgraphs γ of Γ that do not intersect δ. We use $L_{\gamma \cup \delta}$ to symbolize a subtraction operation defined below. It is used to extract the contribution that comes from the region where the momenta in graph δ are of order $1/x$ and the momenta in γ go to infinity. The case of finite q in Fig. 10.2.1 is covered by the case that δ consists of the vertices for $\phi(x)$ and $\phi(0)$ only.

We define the subtraction $L_{\gamma \cup \delta}(\Gamma)$ to be zero unless γ is a disjoint union of 1PI $\gamma_1, \ldots, \gamma_n$. In that case each γ_i is replaced by its overall counterterm $C(\gamma_i)$ defined in (10.3.5) while the graph δ is replaced by a quantity $L(\delta)$. $L(\delta)$ is to contain the leading behavior of δ when all internal lines have large momenta. This is the same idea as that $C(\gamma_i)$ is the overall divergence of γ_i. Now there are regions where a subgraph δ' of δ carries momenta of order q and other lines in δ carry small momenta. To avoid double-counting we subtract them first. So we write:

$$L(\delta) = T\left[\delta - \sum_{\delta' \subsetneq \delta} \sum_{\gamma \cap \delta = \varnothing} L_{\gamma \cup \delta'}(\delta)\right]. \tag{10.3.7}$$

Here T is to be an operator that picks out the leading $x \to 0$ behavior of its argument. Now this behavior is independent of m, and of the finite external

momenta. (We prove this by differentiating with respect to the variables m and p_i^μ.) So we define T to set the values of m and the finite external momenta to zero.

We now have a complete definition of $r(\Gamma)$.

10.3.2 *Absence of infra-red and ultra-violet divergences in* $r(\Gamma)$.

In Fig. 10.1.1(a) the only possible graph of the form of U is the one that consists of $\phi(x)\phi(0)$ alone. We call it δ_1. Its value is $e^{-ip_1\cdot x}+e^{-ip_2\cdot x}$. It has no subgraphs so that

$$L(\delta_1)=(e^{-ip_1\cdot x}+e^{-ip_2\cdot x})_{p_1=p_2=0}=2. \tag{10.3.8}$$

The remainder is therefore

$$r(\text{Fig. } 10.1.1)=\text{Fig. } 10.1.1-\text{Fig. } 10.1.1(b). \tag{10.3.9}$$

Here we regarded $L(\delta_1)$ as a $[\phi^2]$ vertex. It is manifest that this remainder is exactly the graph minus its Wilson expansion.

Now we turn to Fig. 10.1.3(a). It has δ_1 as a subgraph of form U and also the loop, which we call δ_2. Then by (10.3.7)

$$\begin{aligned}L(\delta_2)&=\{\delta_2-L_{\delta_1}(\delta_2)\}_{m=p_1=p_2=0}\\&=\left\{\frac{\mu^{4-d}ig}{(2\pi)^d}\int d^dq\frac{e^{-iq\cdot x}-1}{(q^2-m^2)[(p_1+p_2-q)^2-m^2]}\right\}_{m=p_1=p_2=0}\\&=\frac{ig\mu^{4-d}}{(2\pi)^d}\int d^dq\frac{(e^{-iq\cdot x}-1)}{(q^2+i\varepsilon)^2}.\end{aligned} \tag{10.3.10}$$

Hence

$$\begin{aligned}r(\Gamma)&=R(\Gamma)-L_{\delta_1}(\Gamma)-L_{\delta_2}(\Gamma)\\&=\frac{i^2}{(p_1^2-m^2)(p_2^2-m^2)}\frac{ig\mu^{4-d}}{(2\pi)^d}\left\{\int d^dq\frac{e^{-iq\cdot x}}{(q^2-m^2)[(p_1+p_2-q)^2-m^2]}\right.\\&\quad\left.-\int d^dq\frac{1}{(q^2-m^2)[(p_1+p_2-q)^2-m^2]}-\int d^dq\frac{(e^{-iq\cdot x}-1)}{(q^2)^2}\right\},\end{aligned} \tag{10.3.11}$$

where we use Γ to denote Fig. 10.1.3(a).

The following properties hold:

(1) $L_{\delta_2}(\Gamma)$ is infra-red convergent even though it has zero mass and zero external momentum. Although δ_2 has an infra-red divergence when m, p_1, and p_2 approach zero, the subtraction term $L_{\delta_1}(\delta_2)$ exactly cancels the divergence.

(2) $L_{\delta_1}(\Gamma)$ is ultra-violet divergent, since in replacing the vertex $\delta_1 = e^{-iq\cdot x}$ by 1 we remove the ultra-violet cut off. However, $L_{\delta_2}(\Gamma)$ contains all the large-q behavior of Γ and of subtractions for subgraphs of δ_2. Thus $L_{\delta_2}(\Gamma)$ cancels the ultra-violet divergence of $L_{\delta_1}(\Gamma)$.

(3) When m, p_1, p_2 approach zero, we find that

$$\begin{aligned} r(\delta_2) &\equiv R(\delta_2) - L_{\delta_1}(\delta_2) - L(\delta_2) \\ &= \delta_2 - L_{\delta_1}(\delta_2) - L(\delta_2) \\ &\to 0. \end{aligned}$$

This is just the statement that $L(\delta_2)$ is the value of δ_2 at $m = p_1 = p_2 = 0$, after subtractions on subgraphs are made.

The explanations of these properties are convoluted, but with the aim of demonstrating that they are true in general. Refer now to the general definition of $r(\Gamma)$, viz., (10.3.6), and refer to Fig. 10.2.1 instead of Fig. 10.1.3. Then the above properties get replaced by:

(1) $L(\delta)$ is infra-red convergent for any graph of form U in Fig. 10.2.1: the only regions that could give infra-red problems are cancelled by subtractions.

(2) $r(\Gamma)$ is ultra-violet and infra-red convergent if m, p_1^2 and p_2^2 are non-zero. The individual terms $L_{\gamma\cup\delta}(\Gamma)$ are IR finite. The subtractions remove all ultra-violet behavior.

(3) $r(\delta) = 0$ when $m = p_1 = p_2 = 0$.

10.3.3 $R(\Gamma) - r(\Gamma)$ *is the Wilson expansion*

Although

$$W(\Gamma) \equiv R(\Gamma) - r(\Gamma) = \sum_{\delta} L_{\delta}\Big(\sum_{\substack{\gamma \\ \gamma\cap\delta=\varnothing}} C_{\gamma}(\Gamma) \Big) \tag{10.3.12}$$

contains the leading $x \to 0$ behavior of $R(\Gamma)$, it is not yet in the form of the operator-product expansion, which is

$$W(\Gamma) = \sum_{\delta} \bar{C}(\delta) R(\Gamma/\delta). \tag{10.3.13}$$

Here $\bar{C}(\delta)$ is the contribution of a subgraph δ (of form U) to the Wilson coefficient, while Γ/δ is Γ with δ contracted to a point, i.e., replaced by a ϕ^2 vertex. $R(\Gamma/\delta)$ will now include pole-part subtractions for the divergent Green's functions of ϕ^2.

Summing (10.3.13) over Γ can be done by independently summing over δ and Γ/δ. This gives the operator-product expansion (first term of the right-

hand side of (10.0.1)) with

$$C_{\phi^2}(x^2) = \sum_{\delta}' \bar{C}(\delta). \tag{10.3.14}$$

(We have used the same symbol for the complete Wilson coefficient $C_{\phi^2}(x)$ as for the contribution from a particular graph.) Hence to prove the expansion (10.0.1) or (10.0.3), we have to prove that $W(\Gamma)$ as defined by (10.3.12) equals the right-hand side of (10.3.13).

Any graph Γ for the Green's function G_{N+2} can be written in the form of Fig. 10.2.1. In general there will be several possibilities for the upper subgraph U. For the following argument we will choose U to be the biggest possible graph. Then, in (10.3.12) for $W(\Gamma)$, all the subgraphs δ are contained in U. So we have

$$\begin{aligned} W(\Gamma) &= \int \frac{d^4k d^4l}{(2\pi)^8} W(U) R(I) \\ &\equiv W(U) \otimes R(I), \end{aligned} \tag{10.3.15}$$

since no divergent 1PI subgraphs include parts of both U and I. Suppose we prove (10.3.13) when Γ is replaced by U. Then from (10.3.15) it follows that

$$\begin{aligned} W(\Gamma) &= \sum_{\delta} \bar{C}(\delta) R(U/\delta) \otimes R(I) \\ &= \sum_{\delta} \bar{C}(\delta) R(\Gamma/\delta). \end{aligned} \tag{10.3.16}$$

The last line follows because (*a*) the graphs for Γ/δ are of the form U/δ times I, and (*b*) ultra-violet divergent 1PI subgraphs are entirely within U/δ or within I. Notice that the extra divergences for Γ/δ as compared with Γ are due to the presence of the ϕ^2 vertex. These divergences are confined to graphs of the form of U/δ.

Rather than prove (10.3.13) for the case $\Gamma = U$, we use it as a *definition* of $\bar{C}(U)$, recursive in terms of $\bar{C}(\delta)$ for smaller δ:

$$2\bar{C}(U) = W(U) - \sum_{\delta \subsetneq U} \bar{C}(\delta) R(U/\delta). \tag{10.3.17}$$

The factor 2 multiplying $\bar{C}(U)$ is the lowest-order 1PI Green's function of the operator ϕ^2. It is evidently much too good to be true that we have reduced what ought to be a deep proof to a mere definition. The important physics was at the previous step, (10.3.16). By implicit use of power-counting arguments, to restrict ultra-violet divergences to within U/δ, we proved (10.3.13) given its truth for the case $\Gamma = U$. The Wilson coefficient $C_{\phi^2}(x)$ as it occurs in the expansion (10.0.1) is then the same no matter what Green's function we take of $T\phi(x)\phi(0)$. This universality is an important feature of the Wilson expansion.

The other important feature of the expansion is that the Wilson coefficient is a purely ultra-violet object. This will enable renormalization-group techniques to be useful in its calculation. For this purpose we must now prove that $\bar{C}(U)$ is independent of m, k and l.

The proof is done by differentiating the right-hand side of (10.3.17) with respect to the mass m or with respect to one of the external momenta k or l. Let Δ represent this operation. It is applied in turn to each propagator in $W(U)$ and in $R(U/\delta)$. Inductively, we assume $\bar{C}(\delta)$ satisfies $\Delta\bar{C}(\delta) = 0$. This is true for the lowest-order graph, δ_1, for which $\bar{C}(\delta_1) = 1$ follows from (10.3.8). In the general case $\Delta\bar{C}(U)$ is therefore given by

$$2\Delta\bar{C}(U) = \Delta W(U) - \sum_{\delta \subsetneq U} \bar{C}(\delta)\Delta R(U/\delta), \tag{10.3.18}$$

as illustrated in Fig. 10.3.1. We have a series of terms in each of which one particular propagator of U is differentiated. The differentiation improves the ultra-violet behavior, so the differentiated propagator cannot be a part of a leading large-momentum region connected to the vertices $\phi(x)\phi(0)$. We therefore factor out a maximal two-particle graph δ', just as we factored out U in (10.3.15).

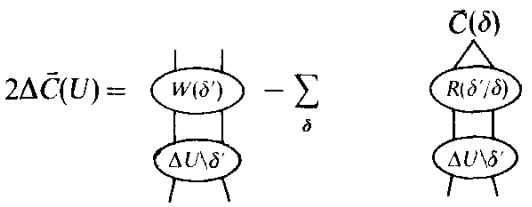


Fig. 10.3.1. Differentiation of Wilson coefficient with respect to a mass or an external momentum.

Now we can use the result

$$W(\delta') = \sum_{\delta''} \bar{C}(\delta'')R(\delta'/\delta'')$$

to give zero for $\Delta\bar{C}(U)$. (This argument is analogous to (10.3.16).) The operator-product expansion is now proved.

10.3.4 *Formula for* $\bar{C}(U)$

From the above results we may deduce

$$\bar{C}(U) = L\left[R(U) - \sum_{\delta \subsetneq U} \bar{C}(\delta)R(U/\delta) \right]. \tag{10.3.19}$$

Heuristically this says that $\bar{C}(U)$ is that part of U arising when all its internal

lines have momenta of order $1/x$, less contributions taken care of by terms $\bar{C}(\delta)$ in the operator-product expansion with smaller graphs δ contained in U. The term in square brackets is the unrenormalized U minus (*a*) all contributions involving momenta not all of order q, and (*b*) contributions with all momenta of order q taken care of by the subtractions under (*a*). This equation reproduces exactly the calculation performed in Section 10.1 of the $O(g)$ Wilson coefficient, viz.,

$$\bar{C}(\delta_2) = [\delta_2 - \bar{C}(\delta_1)R(\delta_2/\delta_1)]_{m=p_1=p_2=0}.$$

10.4 General case

The most fundamental form of the operator-product expansion is (10.0.1), which is proved by obtaining its matrix elements from the Green's functions (10.0.2). We proved the Green's function expansion restricted to connected graphs in ϕ^4 theory, and restricted to the leading power of x^2. The only operator that then contributes is $\mathcal{O} = \phi^2$. Our proof generalizes. We may take the full Green's functions in any theory and include non-leading powers of x.

In the general case the operator L extracts the appropriate number of terms in a Taylor expansion in the mass m^2 and in the external momenta. The result is that each Wilson coefficient $C_{\mathcal{O}}(x)$ behaves in each order of perturbation theory like a power of x^2 times logarithms of $(x^2\mu^2)$ times a polynomial in x^μ and m^2 with dimensionless coefficients. If $\mathcal{O}$ is a tensor operator then the coefficient is also a tensor, as illustrated by the lowest-order example (10.1.5).

The leading operator for ϕ^4 theory is in fact the unit operator since it has lowest dimension. We have

$$T\phi(x)\phi(0) = C_1(x^2)1 + C_{\phi^2}(x^2)[\phi^2] + O(x^\mu \cdot \text{logs}). \qquad (10.4.1)$$

Operators linear in ϕ are prohibited by the $\phi \to -\phi$ symmetry.

To compute the coefficient, $C_1(x^2)$, of the unit operator to lowest order, we use the graph of Fig. 10.4.1. We will extract terms up to $O(x^0)$, so that we

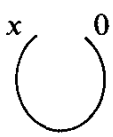


Fig. 10.4.1. Lowest-order term for $\langle 0|T\phi(x)\phi(0)|0\rangle$.

find:

$$\begin{aligned} C_1(x^2) &= C_1(x^2)\langle 0|1|0\rangle \\ &= \{\text{Fig. } 10.4.1 - C_{\phi^2}(x^2)\langle 0|[\phi^2(0)]|0\rangle\}_{\substack{\text{expanded in powers} \\ \text{of } m^2 \text{ to order } m^2}} \\ &= \left\{ \int \frac{d^d q}{(2\pi)^d} \frac{ie^{iq\cdot x}}{q^2 - m^2 + i\varepsilon} - \left[\int \frac{d^d q}{(2\pi)^d} \frac{i}{q^2 - m^2 + i\varepsilon} - \frac{\mu^{d-4} m^2}{8\pi^2 (d-4)} \right] \right\}_{\text{expanded}} \\ &= \int \frac{d^d q}{(2\pi)^d} i(e^{iq\cdot x} - 1)\left[\frac{1}{q^2} + \frac{m^2}{(q^2)^2} \right] + \frac{\mu^{d-4} m^2}{8\pi^2 (d-4)} \\ &= -\frac{1}{4\pi^2 x^2} + \frac{m^2}{16\pi^2}[\gamma + \ln(-\pi^2 \mu^2 x^2)] \quad \text{at } d = 4. \end{aligned} \tag{10.4.2}$$

The Wilson coefficient is obtained by expanding in powers of mass and external momentum up to an appropriate degree, which is two here. We have no external momenta, so the expansion is in powers of m^2.

In the next section, when we apply the renormalization group to compute the Wilson coefficients we will find that each coefficient is given as a series in the effective coupling with the renormalization mass μ set to $(-x^2)^{1/2}$. The main application of these methods will be to asymptotically free theories. If we truncate the perturbation expansion, then the error will be of the order of the first omitted term, and hence the fractional error is of order

$$1/[\ln(-x^2)]^{p+1},$$

where p is the number of loops in the highest-order graph. This error dominates any positive power of x.

Consequently, it is difficult to use the power-law corrections to the leading power in a Wilson coefficient. This would suggest we cannot properly use anything but the coefficient of the unit operator. However, in applications we will normally work with connected Green's functions. For these the leading coefficient is of the two-particle operator ϕ^2.

10.5 Renormalization group

To make calculations for a Wilson coefficient, we must use the renormalization group to obtain maximum information from a low-order calculation in the way we will now explain.

The coefficient has the functional dependence

$$C(x^2, g, \mu).$$

To use this we must set μ^2 to be of order $|1/x^2|$ to avoid large logarithms, just as in calculating Green's functions when all their external momenta get large (Chapter 7). The RG equations are simple, since, to the leading power, there is no mass dependence. (If we use non-leading powers of x then we have polynomial dependence on masses.) The renormalization-group equation for the Wilson coefficient can be derived most easily from the renormalization-group equations for Green's functions in the following fashion.

In the ϕ^4 theory we have the renormalization-group operator:

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}=\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}-\gamma_m m^2\frac{\partial}{\partial m^2}. \tag{10.5.1}$$

The renormalization-group equations we need are

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}G_{N+2}^{(\mathrm{conn})}=-(N/2+1)\gamma G_{N+2}^{(\mathrm{conn})}, \tag{10.5.2}$$

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}G_N([\phi^2],\ldots)^{(\mathrm{conn})}=(\gamma_m-\tfrac{1}{2}N\gamma)G_N([\phi^2],\ldots)^{(\mathrm{conn})}. \tag{10.5.3}$$

Here, $G_{N+2}^{(\mathrm{conn})}$ denotes the Green's function of $N+2$ external ϕ-fields, restricted to connected graphs. $G_N([\phi^2],\ldots)^{(\mathrm{conn})}$ denotes the connected Green's function of the renormalized $[\phi^2]$ operator with N ϕ-fields. To derive (10.5.3) we used the fact that

$$m^2[\phi^2]=m_0^2\phi_0^2+\text{coefficient times 1},$$

and $m_0^2=Z_m m^2$.

We apply the operator $\mu\mathrm{d}/\mathrm{d}\mu+(N/2+1)\gamma$ to both sides of

$$\begin{aligned}&\langle 0|T\phi(x)\phi(0)\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle\\&\quad=C_{\phi^2}(x)\langle 0|T[\phi^2]\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle+O(x^2),\end{aligned} \tag{10.5.4}$$

to obtain

$$0=\left[\mu\frac{\mathrm{d}C_{\phi^2}}{\mathrm{d}\mu}+(\gamma+\gamma_m)C_{\phi^2}\right]\langle 0|T[\phi^2]\tilde{\phi}(p_1)\ldots\tilde{\phi}(p_N)|0\rangle+O(x^2). \tag{10.5.5}$$

Since the $O(x^2)$ terms are order x^2 independently of g, m, and μ, they remain of order x^2 after $\mu\mathrm{d}/\mathrm{d}\mu$ is applied. (As usual, 'order x^2' means order x^2 modulo possible logarithms.) Immediately we obtain

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}C_{\phi^2}=(\gamma_m+\gamma)C_{\phi^2}. \tag{10.5.6}$$

This equation can be solved – it is effectively in the $m = 0$ theory. The anomalous dimension of C_{ϕ^2} is the anomalous dimension of $\phi(x)\phi(0)$ minus the anomalous dimension of $[\phi^2]$. This is the same relation as for the engineering dimensions. The signs in (10.5.6) arise from peculiarities of our definitions of γ_m and γ.

11

Coordinate space

In the previous chapters we set up renormalization theory in momentum space. In this chapter, we will give a treatment in coordinate space. Now, the utility of a momentum-space description, such as we gave in the earlier chapters, comes from the translation invariance of a problem. However, the momentum-space formulation rather obscures the fact that UV divergences arise from purely short-distance phenomena. Thus a coordinate-space treatment is useful from a fundamental point of view. There are also a number of situations, essentially external field problems, where a coordinate-space treatment is the most appropriate from a more practical point of view. A particular advantage is that the coordinate-space method makes it easy to see that the counterterms are the same as with no external field.

An important case, which we will treat in detail in this chapter, is that of thermal field theory at temperature T (Fetter & Walecka (1971), Bernard (1974), and references therein). There one works with imaginary time using periodic boundary conditions (period $1/T$).

It is first necessary to work out the short-distance singularities of the free propagator. This we will do in Section 11.1. A number of forms of the propagator will be given, whose usefulness will become apparent when we treat some examples in Section 11.2. The reader should probably skip to this section first and refer back to Section 11.1 as the need for various properties of the propagator arises. We will explicitly show that the particular counterterms that we compute are independent of temperature. Our arguments will be of a form that will readily generalize not only to higher orders but also to other situations.

In Section 11.3 we will show to all orders of perturbation theory that counterterms at finite temperature are independent of T. We will do this by constructing counterterms in a theory in flat space-time in such a way as to make manifest the fact that only the short-distance singularities of the propagator affect the counterterms.

Another case, which we will treat in Section 11.4, is flat space-time with,

say, an external electromagnetic field. If the field is weak it can be treated as a perturbation, i.e., as part of the interaction. But if the field is strong, one must put it in the free Lagrangian. Thus the free electron propagator satisfies

$$(\mathrm{i}\partial\!\!\!/ + \mathrm{e}A\!\!\!/ - M)S_{\mathrm{F}}(x, y; eA) = \mathrm{i}\delta(x - y), \tag{11.0.1}$$

which cannot in general be solved by expanding in a series in the field (Schwinger (1951)).

Another common external field problem is that of quantizing a quantum field theory in a curved space-time, where there is no remnant of a global Poincaré symmetry. Momentum space is then an inappropriate tool. One wishes to show that the UV counterterms can be kept the same as in flat space-time. We will not discuss this particular case here. But an extension of the techniques described should enable a fairly simple treatment to be given.

11.1 Short-distance singularities of free propagator

11.1.1 *Zero temperature*

The UV counterterm of a 1PI graph is ultimately determined by the short-distance singularities of the free propagators that make up the graph. So we need to obtain these singularities; and we will do this in this section.

It is sufficient to consider the scalar propagator

$$S_{\mathrm{F}}(x^2; d, m) = \int \frac{\mathrm{d}^d q}{(2\pi)^d} \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}q\cdot x}}{(q^2 - m^2 + \mathrm{i}\varepsilon)}. \tag{11.1.1}$$

Propagators for fields with spin can be obtained by differentiating with respect to x. The propagator satisfies the equation

$$(\Box + m^2)S_{\mathrm{F}} = -\mathrm{i}\delta^{(d)}(x), \tag{11.1.2}$$

together with appropriate boundary conditions. When x is non-zero this equation reduces to

$$(4z\partial^2/\partial z^2 + 2d\partial/\partial z - m^2)S_{\mathrm{F}} = 0, \tag{11.1.3}$$

where $z = -x^2$. Thus $S_{\mathrm{F}} = z^{1/2 - d/4} w(mz^{1/2})$, where w satisfies the modified Bessel equation of order $\nu = d/2 - 1$. The particular solution we need is determined by requiring that the δ-function in (11.1.2) be obtained at $x^2 = 0$ and that $S_{\mathrm{F}} \to 0$ as $x^2 \to -\infty$.

A standard method of solving (11.1.3) is to expand in powers of z:

$$S_F = z^a \sum_{n=0}^{\infty} (zm^2)^n w_n$$
$$= (-x^2)^a \sum (-m^2x^2)^n w_n. \tag{11.1.4}$$

There are two independent solutions. One is analytic at $z=0$, i.e., it has $a=0$; the other solution is singular, with $a=1-d/2$. S_F is a linear combination of these solutions. The quickest way to compute the singularity is to observe that, if $m=0$, then (11.1.1) gives

$$S_F(m=0) = \Gamma(d/2-1)/[4\pi^{d/2}(-x^2)^{d/2-1}]. \tag{11.1.5}$$

This normalizes the coefficient of the singular solution. The normalization of the regular solution is obtained from the properties of Bessel functions, by requiring that $S_F \to 0$ as $x^2 \to -\infty$. Then we find that

$$S_F = \frac{1}{4\pi^{d/2}(-x^2)^{d/2-1}} \sum_{n=0}^{\infty} \left(\frac{m^2x^2}{4}\right)^n \frac{\Gamma(d/2-1-n)}{n!}$$
$$+ m^{d-2}(4\pi)^{-d/2} \sum_{n=0}^{\infty} (m^2x^2/4)^n \Gamma(1-n-d/2)/n!$$
$$= S_{\text{Fsing}} + S_{\text{Fana}}. \tag{11.1.6}$$

Here, we have used the series expansion of the Bessel functions. The coefficient of the regular solution can be obtained quickly by explicitly computing, from (11.1.1), that $S_F = \Gamma(1-d/2)m^{d-2}/(4\pi)^{d/2}$ at $x=0$, if d is less than 2.

For the purposes of renormalization we will need the singularities of S_F. These are of two types: the singularities of S_{Fsing} as $x \to 0$, and the singularities of S_{Fana} as d approaches an even integer. The fact that most of the Γ-functions have poles when d is an even integer $d=2\omega$, reflects the fact that the ansatz (11.1.4) is then incorrect for the singular solution. However, the correct result is obtained by expanding in powers of $d-2\omega$ and then letting $d \to 2\omega$. The limits $x \to 0$ and $d \to 2\omega$ are non-uniform. Application to graphs like the ϕ^3 self-energy need $x \neq 0$, but tadpole graphs (e.g., Fig. 11.1.1 in ϕ^4 theory) have a propagator with $x=0$ for all values of d. So

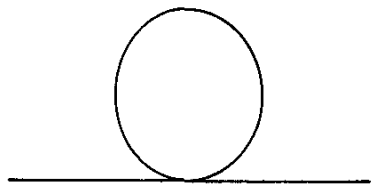


Fig. 11.1.1. Graph with a tadpole.

it is convenient to extract all the singular behavior of S_F close to $d = 2\omega$ and $x = 0$ by writing

$$S_F = \bar{S}_{\text{Fsing}} + \bar{S}_{\text{Fana}}$$

$$\bar{S}_{\text{Fsing}} = \frac{1}{4\pi^\omega(-x^2)^{\omega-1}}\left\{\sum_{n=0}^{\omega-2}\left(\frac{m^2x^2}{4}\right)^n \frac{\Gamma(d/2-1-n)}{n!(-\pi x^2)^{d/2-\omega}} + \sum_{n=\omega-1}^{\infty}\left(\frac{m^2x^2}{4}\right)^n \frac{1}{n!}\left[\frac{\Gamma(d/2-1-n)}{(-\pi x^2)^{d/2-\omega}} + \frac{(-1)^{n+\omega}\mu^{d-2\omega}}{(d/2-\omega)(n+1-\omega)!}\right]\right\},$$

$$\bar{S}_{\text{Fana}} = \frac{m^{2\omega-2}}{(4\pi)^\omega}\sum_{n=0}^{\infty}\left(\frac{m^2x^2}{4}\right)^n \frac{1}{n!} \times \times\left\{\Gamma(1-n-d/2)\left(\frac{m^2}{4\pi}\right)^{d/2-\omega} - \frac{(-1)^{n+\omega}\mu^{d-2\omega}}{(d/2-\omega)(\omega+n-1)!}\right\}. \quad (11.1.7)$$

Neither term is a solution of the equation for S_F. The analytic part now has a finite limit as $d \to 2\omega$, with no singular behavior at $x = 0$. The singular term also has a finite limit as $d \to 2\omega$ if $x \neq 0$. It contains the $x \to 0$ singularities. But if we need $S_F(x = 0)$ then we first take $d < 2$, then $x \to 0$, and finally $d \to 2\omega$. Then the only singular contribution is from the $n = \omega - 1$ term in $\bar{S}_{\text{Fsing}}$:

$$\bar{S}_{\text{Fsing}}(x = 0) = \frac{(-1)^\omega m^{2\omega-2}\mu^{d-2\omega}}{(d/2-\omega)(4\pi)^\omega(\omega-1)!}. \quad (11.1.8)$$

This result will be needed in evaluating tadpole graphs.

11.1.2 *Non-zero temperature*

Thermal Green's functions at inverse temperature $\beta = 1/T$ are obtained by Wick-rotating time $t \to -i\tau$ and then by imposing periodic boundary conditions (Fetter & Walecka (1971)). Thus the free propagator $S_F(x - y; \beta)$ satisfies

$$(-\partial^2/\partial\tau^2 - \vec{\nabla}^2 + m^2)S_F(\tau, \vec{x}) = \delta(\tau)\delta^{(3)}(\vec{x}), \quad (11.1.9a)$$

$$S_F(\tau, \vec{x}) \to 0, \quad \text{as } \vec{x} \to \infty, \quad (11.1.9b)$$

$$S_F(\tau + \beta, \vec{x}) = S_F(\tau, \vec{x}). \quad (11.1.9c)$$

All integrals over space-time are restricted to the time range between 0 and β. The propagator is obtained from the zero-temperature propagator $S_F(x; \infty)$ by writing

$$S_F(x; \beta) = \sum_{n=-\infty}^{\infty} S_F(\tau + n\beta, \vec{x}; \infty). \quad (11.1.10)$$

(Note that in $S_F(x-y;\beta)$ both τ_x and τ_y are between 0 and β, so when we apply $-\partial^2/\partial\tau^2 - \nabla^2 + m^2$ to it the only δ-function in (11.1.9*a*) arises from the $n=0$ term.)

The only singularities in $S_F(x;\beta)$ are from the $n=0$ term when $x^\mu = 0$, so we have

$$S_F(x;\beta) = \bar{S}_{\text{Fsing}}(x) + \bar{S}_{\text{Fana}}(x;\beta), \tag{11.1.11}$$

where $\bar{S}_{\text{Fsing}}$ is the same as in (11.1.7), with $-x^2 = \tau^2 + \mathbf{x}^2$. The analytic term is different. It differs from $\bar{S}_{\text{Fana}}(x;\infty)$ at zero temperature by

$$\Delta S_F = \sum_{n\neq 0} S_F(\tau + n\beta, x; \infty),$$

which, in the range $0 < t < \beta$, is a solution of the homogeneous equation $(\Box + m^2)S_F = 0$. Notice that it is not a function of x^2 alone, i.e., it is not Poincaré invariant.

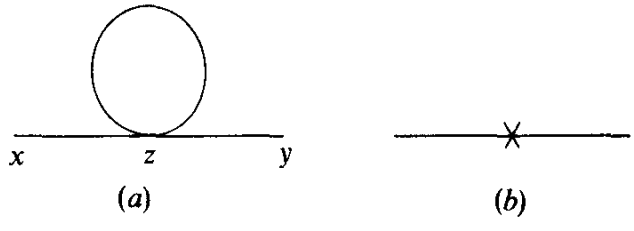
Another, direct, way of deriving the form (11.1.11) is to expand $S_F(x;\beta)$ in a power series about $x^\mu = 0$, and then to require the differential equation (11.1.9*a*) to be satisfied. The δ-function on the right-hand side ensures that $S_F = S_{\text{Fsing}}(x) + S_{\text{Fana}}(x;\beta)$, where S_{Fsing} is the same singular solution as in (11.1.6). Then S_{Fana} is a solution of the homogeneous equation analytic at $x = 0$, and for which $S_{\text{Fsing}} + S_{\text{Fana}}$ satisfies periodic boundary conditions. We then add and subtract the poles at $d = 2\omega$ to obtain (11.1.11). The coefficient of the pole in S_{Fana} is the same as at zero temperature, because it must cancel the pole in S_{Fsing}, which is independent of temperature.

11.2 Construction of counterterms in low-order graphs

To explain how renormalization works in coordinate space we consider the graphs shown in Figs. 11.2.1 to 11.2.4 for ϕ^4 theory in four space-time dimensions. These are sufficient to show the various complications that occur.

Our treatment works as well at any temperature.

The simplest example is the one-loop correction to the propagator,



(*a*) (*b*)

Fig. 11.2.1. Self-energy graph and counterterm.

Fig. 11.2.1. Its contribution to the full propagator is:

$$M_1 = -\frac{ig}{2}\mu^{4-d}\int \mathrm{d}^d z S_F(x-z,d;1/T)S_F(z-y,d;1/T)S_F(0,d;1/T). \tag{11.2.1}$$

The functions $S_F(x-z)$ and $S_F(y-z)$ are singular at $z=x$ and at $z=y$, but the singularities are integrable, so that they do not contribute any divergence. The factor $S_F(0,d;1/T)$ is divergent, so we use (11.1.8) and (11.1.11) to write M_1 as

$$M_1 = -\frac{ig}{2}\int \mathrm{d}^4 z S_F(x-z)S_F(z-y)\bar{S}_{\text{Fana}}(0,d=4;1/T)$$
$$-\frac{ig}{2}\int \mathrm{d}^d z S_F(x-z)S_F(z-y)\frac{m^2}{8\pi^2(d-4)} + O(d-4). \tag{11.2.2}$$

The divergent term is evidently exactly cancelled by the mass-renormalization counterterm, Fig. 11.2.1(*b*). The result for the renormalized propagator at order g is

$$-\frac{ig}{2}\int \mathrm{d}^4 z S_F(x-z)S_F(z-y)\bar{S}_{\text{Fana}}(0,4,1/T). \tag{11.2.3}$$

Notice that the renormalization only involved the singular term in S_F. Thus we have shown that the counterterm is the same at any temperature.

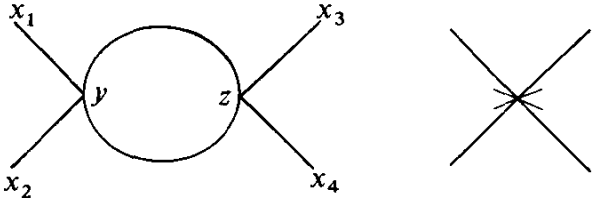
The unrenormalized graph of Fig. 11.2.2 for the four-point function is

$$M_2 = \tfrac{1}{2}(-ig\mu^{4-d})^2\int \mathrm{d}^d y \mathrm{d}^d z S_F(x_1-y)S_F(x_2-y)S_F(y-z)^2 \times$$
$$\times S_F(z-x_3)S_F(z-x_4). \tag{11.2.4}$$

The only divergence as d approaches 4 comes from the singularity of $S_F(y-z)^2$ at $y=z$; it is logarithmic. Let us write

$$M_2 = g^2\int \mathrm{d}^d y \mathrm{d}^d z g S_F(y-z)^2 f(y,z,\{x_i\}). \tag{11.2.5}$$

Since the divergence is logarithmic, it is governed by the leading behavior of



(*a*) (*b*)

Fig. 11.2.2. Vertex graph and counterterm.

the integrand of (11.2.5) as $y \to z$. Thus we can generate a counterterm, illustrated in Fig. 11.2.2(*b*),

$$M_{2b} = -\int \mathrm{d}^d y \mathrm{d}^d z g^2 f(y, y, \{x_i\})[\text{pole part of} \int_{z \simeq y} \mathrm{d}^d z S_{\mathrm{F}1}(y-z)^2]. \tag{11.2.6}$$

The singularity is entirely given by the first term $S_{\mathrm{F}1}$ in the expansion (11.1.7) of $\bar{S}_{\mathrm{Fsing}}$

$$S_{\mathrm{F}1} \equiv \Gamma(d/2-1)/[4\pi^{d/2}(-x^2)^{d/2-1}].$$

The pole-part factor in (11.2.6) is then

$$\begin{aligned} -\operatorname{pole}\left\{\int_{z\simeq y} \mathrm{d}^d z S_{\mathrm{F}1}(y-z)^2\right\} &= -\operatorname{pole}\left\{\frac{\Gamma(d/2-1)^2}{16\pi^d}\int_{z\simeq 0} \mathrm{d}^d z(-z^2)^{2-d}\right\} \\ &= \frac{\mathrm{i}}{16\pi^4}\operatorname{pole}\left\{\frac{2\pi^{d/2}}{\Gamma(d/2)}\int_0 \mathrm{d}z z^{3-d}\right\} \\ &= \frac{\mathrm{i}}{8\pi^2}\frac{\mu^{d-4}}{4-d}. \end{aligned} \tag{11.2.7}$$

In the first line we shifted z to $z+y$. In the second line we Wick rotated z^0, and the notation $\int_{z\simeq 0}\mathrm{d}z$ indicates that only the region near $z=0$ matters. The factor μ^{d-4} in the last line is needed, as usual, to preserve explicit dimensional correctness.

We thus find that the renormalized M_2 is

$$\begin{aligned} M_2 = \lim_{d\to 4}\int \mathrm{d}^d z \mathrm{d}^d y\left[g^2 S_{\mathrm{F}}(y-z)^2 + \frac{\mathrm{i}g^2\mu^{4-d}}{8\pi^2(4-d)}\delta^{(d)}(y-z)\right]\times \\ \times f(y,z;\{x_i\}). \end{aligned} \tag{11.2.8}$$

The factor in square brackets is a well-defined distribution at $d=4$. Again observe that we only used the singular part $\bar{S}_{\mathrm{Fsing}}$ of S_{F} in obtaining the counterterm, so that the counterterm is temperature-independent.

The graph Fig. 11.2.3 has a logarithmic subdivergence as well as an

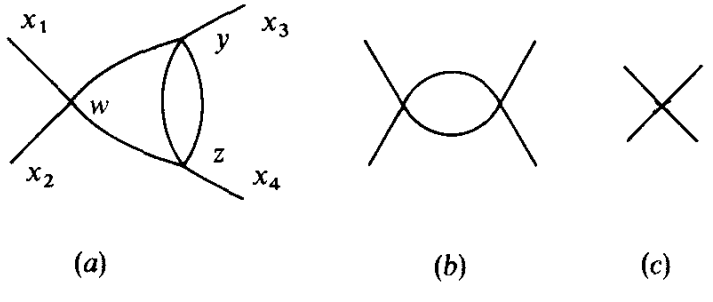


Fig. 11.2.3. Two-loop vertex graph and counterterms.

overall divergence. Its unrenormalized value is

$$M_3 = \tfrac{1}{2}(-ig\mu^{4-d})^3 \int d^dw\, d^dy\, d^dz\, S_F(x_1 - w) S_F(x_2 - w) S_F(x_3 - y) \times$$
$$\times S_F(x_4 - z) S_F(w - y) S_F(w - z) S_F(y - z)^2$$
$$\equiv g^3 \int d^dw\, d^dy\, d^dz f^{(3)}(w, y, z, \{x_i\}) S_F(w - y) S_F(w - z) S_F(y - z)^2. \tag{11.2.9}$$

If w is not close to y or z there is a logarithmic divergence when y approaches z. This is identical to the divergence in Fig. 11.2.2. There is also an overall divergence when all of the interaction points w, y, and z approach each other.

We first subtract the subdivergence by the counterterm (11.2.7):

$$\bar{R}(M_3) = g^3 \int f^{(3)}(w, y, z, \{x_i\}) S_F(w - y) S_F(w - z) \times$$
$$\times \left[S_F(y - z)^2 + \frac{i}{8\pi^2} \frac{\mu^{d-4}}{(4-d)} \delta^{(d)}(y - z) \right]. \tag{11.2.10}$$

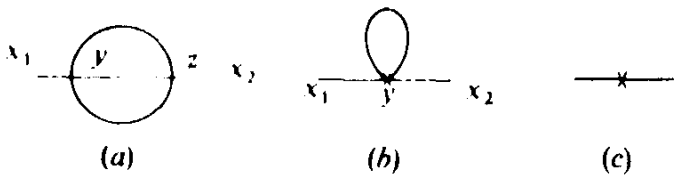
The distribution in square brackets is singular at $y = z$. However, it is integrable; that is, it produces a finite result when d goes to 4, if it is integrated with a test function continuous at $y = z$. The function $f^{(3)}(w, y, z) S_F(w - y) S_F(w - z)$ is continuous at $y = z$, unless also $w = y = z$. The remaining divergence in $\bar{R}(M_3)$ comes from the region $w \sim y \sim z$, where $f^{(3)}$ is not continuous. (There is also a singularity if y and z are both equal to x_3 or x_4; however, this region is again integrable.) The divergence at $w = y = z$ is logarithmic, so again it is determined entirely by the leading singular terms of the propagators. The counterterm is therefore

$$C = -\int d^dw\, g^3 f^{(3)}(w, w, w, \{x_i\})\, \text{pole} \left\{ \int d^dy\, d^dz\, S_{F1}(y) S_{F1}(z) \times \right.$$
$$\left. \times \left[S_F(y - z)^2 + \frac{i}{8\pi^2} \frac{\mu^{d-4}}{(4-d)} \delta^{(d)}(y - z) \right] \right\}, \tag{11.2.11}$$

as shown in Fig. 11.2.3(*c*). Recall that the factor in square brackets is singular but integrable at $y = z$. So the only divergent behavior comes when $S_F(y) S_F(z)$ is singular, i.e., at $y = z = 0$. It is easily checked that the counterterm C then reduces to

$$C = \int d^dw f^{(3)}(w, w, w, \{x_i\}) \left(\frac{-g^3 \mu^{2d-8}}{(16\pi^2)^2} \right) \left[\frac{2}{(4-d)^2} - \frac{1}{4-d} \right] \tag{11.2.12}$$

just as in momentum space (Vladimirov, Kazakov & Tarasov (1979)).



(a) (b) (c)

Fig. 11.2.4. Two-loop self-energy graph and counterterms.

Our final graph, Fig. 11.2.4, is a little more difficult. Its unrenormalized contribution to the propagator is

$$M_4 = \tfrac{1}{6}(-ig\mu^{4-d})^2 \int d^d y d^d z S_F(x_1 - y) S_F(y - z)^3 S_F(z - x_2)$$

$$\equiv g^2 \int d^d y d^d z f^{(4)}(y, z; \{x_i\}) S_F(y - z)^3. \tag{11.2.13}$$

The difficulty is that in momentum space there is not only the quadratic overall divergence, but also three logarithmic subdivergences. However (11.2.13) appears to have only an overall divergence, when y goes to z; this involves all three propagators. Furthermore, the counterterm, calculated in momentum space, for a subdivergence from two of the lines gives a graph of the form of Fig. 11.2.1(*a*). There the third propagator is at $y = z$, whereas the propagators in (11.2.13) are also used at $y \neq z$.

The correct way to handle this case is to write

$$S_F = S_{F1} + S_{F2} + S_{rem}, \tag{11.2.14}$$

where S_{F1} is the same leading term as before, and S_{F2} is the second term in $\bar{S}_{Fsing}$:

$$S_{F2}(x) = -\frac{m^2}{16\pi^2}\left[\frac{\Gamma(d/2-2)}{(-\pi x^2)^{d/2-2}} - \frac{\mu^{d-4}}{d/2-2}\right]. \tag{11.2.15}$$

The remainder $S_{rem}(y - z)$ is finite as $d \to 4$ and/or $y \to z$. We now substitute (11.2.14) for each factor of $S_F(y - z)$ in (11.2.13) and obtain

$$M_4 = g^2\left\{\int d^d y d^d z f^{(4)}(y, z) S_{F1}(y - z)^3 + 3\int f^{(4)}(y, z) S_{F1}^2 S_{F2}\right.$$

$$\left. + 3\int f^{(4)}(y, z) S_{F1}^2 S_{rem}\right\} + \text{finite}. \tag{11.2.16}$$

The term with all S_{F1}'s has the quadratic divergence, while the terms with two S_{F1}'s are logarithmically divergent. The other terms are finite. The term with $S_{F1}^2 S_{rem}$ has a divergence coming from the two S_{F1}'s, and its factor 3 reflects the fact that there are three divergent subgraphs. The graph with the

counterterm for the subdivergences is Fig. 11.2.4(*b*):

$$M_{4b} = \frac{3ig^2\mu^{4-d}}{8\pi^2(4-d)}\int d^d y f^{(4)}(y,y;\{x_i\})S_F(0)$$
$$= \frac{3ig^2\mu^{4-d}}{8\pi^2(4-d)}\int d^d y f^{(4)}(y,y;\{x_i\})[S_{F2}(0) + S_{rem}(0)]. \quad (11.2.17)$$

Here we used the value we calculated at (11.2.7) for the counterterm vertex. The $S_{rem}(0)$ term here cancels the corresponding divergence in the third term in (11.2.16), while the S_{F2} term combines with the second term in (11.2.16) to produce a logarithmic divergence from

$$3g^2\int d^d y f^{(4)}(y,y)\left\{\int_{z\simeq y} d^d z\, S_{F1}(y-z)^2 S_{F2}(y-z)\mu^{8-2d} - \frac{\mu^{4-d}}{8\pi^2(4-d)}S_{F2}(0)\right\}. \quad (11.2.18)$$

Let us return to (11.2.16). The divergence in the first term is sensitive to $f^{(4)}$ and its second derivative at $y=z$:

$$g^2\int d^d z\int_{y\simeq z} d^d y S_{F1}(y-z)^3[f^{(4)}(z,z) + (y-z)^\mu(\partial f^{(4)}/\partial y^\mu)|_{y=z} + \tfrac{1}{2}(y-z)^\mu(y-z)^\nu(\partial^2 f^{(4)}/\partial y^\mu \partial y^\nu)|_{y=z}]$$
$$= \text{finite} + [2ig^2\pi^{d/2}/\Gamma(d/2)] \times \int d^d z\int_0 dy\, y^{d-1}\frac{\Gamma(d/2-1)^3}{(4\pi^{d/2})^3 y^{3d-6}}\left[f^{(4)}(z,z) - \frac{y^2}{2d}\frac{\partial^2 f^{(4)}}{\partial y^\mu \partial y_\mu}\bigg|_{y=z}\right]$$
$$= \text{finite} - i\left(\frac{g}{16\pi^2}\right)^2\int d^d y d^d z f^{(4)}(y,z)\,\Box\delta^{(d)}(y-z)/(8-2d). \quad (11.2.19)$$

Note that the term with $f^{(4)}(z,z)$ gives a pole at $d=2$ but not at $d=4$.

11.3 Flat-space renormalization

In the last section we computed counterterms for some low-order graphs. Our method was not a good method for computing the finite parts. But it made very explicit the fact that the counterterms depend only on the short-distance structure of the free propagator. In particular it made it obvious that the counterterms are independent of temperature. We must now spell out how to generalize the results to an arbitrary graph in an arbitrary theory.

We define the renormalization of a graph G by the same structure that we had in momentum space (i.e., by the recursion method or by the forest

formula). A counterterm is needed for the overall divergence of every 1PI subgraph. The overall degree of divergence of a 1PI subgraph is obtained by first counting powers of position as all of its vertices approach each other, and by then taking the negative of the result (since it is an $x \to 0$ divergence). The integrations over relative positions of the vertices are included in the power-counting. The result coincides with the usual momentum-space definition for a 1PI subgraph. The value of G is obtained by integrating over positions of its interaction vertices, and the divergence and subdivergences come from regions where some of these positions approach each other. To each region corresponds a subgraph consisting of the vertices that go to the same point, together with all the lines joining them.

It might appear that these graphs should be in one-to-one correspondence with the 1PI subgraphs that in momentum space are divergent when all their internal loop momenta are large. However this is not so, for the following reasons:

(1) There are graphs overall divergent in coordinate space that are not 1PI.
(2) There are divergent subgraphs in momentum space that are not obtained directly in coordinate space; these are when some but not all lines connecting vertices of a 1PI graph are in the subgraph.

An example of the first case is Fig. 11.3.1, where there is a logarithmically divergent graph consisting of the vertices at w, y, z, and of the lines joining them. In momentum space the counterterms for the self-energy subgraph cancel all the divergences. But in coordinate space $S_F(w-y)$ is singular, so we have to justify the momentum-space result. Other graphs are divergent in momentum space but do not occur directly as divergences in coordinate space. A typical example is given by the one-loop subgraphs of Fig. 11.2.4. The divergence comes from the region $y \to z$, and it appears to involve all three propagators, never just two propagators. As we saw in Section 11.2, the trick to handling this problem is to make a decomposition of the propagator, as in (11.2.14).

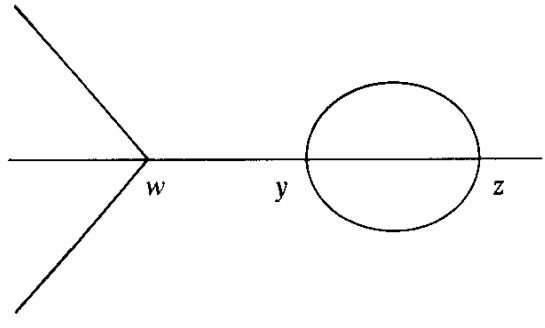


Fig. 11.3.1. The w–y–z subgraph is one-particle-reducible and apparently has an overall divergence in coordinate space.

First, however, let us consider the problem of divergent 1PR graphs like the (w, y, z) subgraph of Fig. 11.3.1. After subtraction of the self-energy counterterms, Fig. 11.3.1 has the form

$$\int \mathrm{d}^d w \mathrm{d}^d y \mathrm{d}^d z f(w) S_F(w-y) \times$$
$$\times [S_F(y-z)^3 + A\delta(y-z) + B\Box\delta(y-z)] S_F(z-x_2). \quad (11.3.1)$$

Here $f(w)$ is a function non-singular at $w = y$. The factor in square brackets is the self-energy plus counterterms. At $d = 4$, we have $S_F(w-y) \sim 1/(w-y)^2$ and $S_F(y-z)^3 \sim 1/(y-z)^6$, so it might appear that there is a logarithmic divergence when w and y both approach z. A counterterm for this divergence would be a four-point vertex and therefore allowed. However, the momentum-space result gives no such counterterm, so it is important to derive the same result directly in coordinate space.

We first integrate over w, and see that

$$\int \mathrm{d}^d w f(w) S_F(w-y)$$

is finite and non-singular as a function of y. Then we can perform the y-integral. Although $S_F(y-z)$ is singular at $y = z$, it is integrated with a function with no singularity there, so the counterterms are sufficient to cancel the divergence. Then, finally, we integrate over z. The crucial point is that we find an order of integration with the following property: – each integral kills the singularity on precisely one propagator without introducing new singularities. If we replaced Fig. 11.3.1 by Fig. 11.2.3, say, we would have a not-quite similar integral:

$$\int \mathrm{d}^d w \mathrm{d}^d y \mathrm{d}^d z S_F(w-y) S_F(w-z)[S_F(y-z)^2 - \text{counterterm}].$$

The integral over w, for example, involves two propagators $S_F(w-y)$, $S_F(w-z)$. The result becomes singular at $y = z$.

So far we have chosen to define renormalization by the forest formula or by the recursive method. We have seen that, after subtraction, only 1PI graphs have divergences. Some subgraphs that need counterterms in momentum space (like Fig. 11.2.4) do not appear explicitly as divergences in coordinate space, since we defined the subgraph corresponding to a coordinate-space divergence to include all the lines connecting its vertices. Even so, we saw in Section 11.2 that we have counterterms for such subgraphs – as in Fig. 11.2.4(*b*). To show how these counterterms arise in general and to show that the counterterms are independent of temperature,

we decompose S_F as

$$S_F = S_{F1} + \cdots + S_{FN} + S_{\text{rem}}. \tag{11.3.2}$$

Here S_{F1} to S_{FN} are the first N terms in the singular part $\bar{S}_{\text{Fsing}}$ of S_F. We choose N so that $2N$ is larger than the degree of divergence of any subgraph of the graph G that is being renormalized. Thus when S_F is replaced by S_{rem}, any (sub)graph γ containing it becomes overall convergent.

When we substitute (11.3.2) for every propagator in a graph, we obtain a sum of graphs, in each of which every propagator is replaced by one of $S_{F1}, \ldots, S_{FN}$, or S_{rem}. We renormalize each of these new graphs separately. A subgraph needs a counterterm if its overall degree of divergence is a positive integer or zero, and we make counterterms for all 1PI subgraphs that are divergent.

The rules for computing the degree of divergence are essentially obvious. Any propagator with S_F replaced by S_{F1} contributes as in the original graph, and a replacement by S_{FN} reduces the degree of divergence by $2N-2$. The only subtlety is that replacement of S_{F1} by $S_{\text{rem}}(x-y)$ is considered as contributing $-2N$ to the degree of divergence, even though it does not behave as $(x-y)^{2-d+2N}$ when $x-y \to 0$: Its analytic part behaves as $(x-y)^0$. However, the analytic part never actually contributes to any overall divergence, and the singular part of S_{rem} does go like $(x-y)^{2-d+2N}$ relative to S_{F1}.

A typical case is the third term on the right of (11.2.16). Before subtraction of the subdivergence, this term is proportional to

$$\int f^{(4)}(y,z) S_{\text{rem}}(y-z) S_{F1}(y-z)^2 \mathrm{d}^d y \mathrm{d}^d z$$

Now $S_{F1}(y-z)^2 \sim (y-z)^{4-2d}$ as $y \to z$, while $S_{\text{rem}} \to \bar{S}_{\text{Fana}}(0) \neq 0$. So there is a logarithmic divergence at $y=z$. But after subtraction, we have:

$$\int f^{(4)}(y,z) S_{\text{rem}}(y-z) \left[S_{F1}(y-z)^2 + \frac{\mathrm{i}\mu^{d-4}}{8\pi^2(4-d)} \delta(y-z) \right] \mathrm{d}^d y \mathrm{d}^d z.$$

When integrated with an analytic function like $f^{(4)}(y,z)\bar{S}_{\text{Fana}}(y-z)$, the factor in square brackets gives a finite result. So any divergence comes from the singularity in S_{rem}. But the singularity is of order $(y-z)^{6-d}$, so no divergence occurs.

We now generalize our treatment of Fig. 11.2.4. We examine structures of the form of Fig. 11.3.2(*a*). The subgraph A is 1PI and overall divergent. The lines $l_1, \ldots, l_a$ join vertices of A. A graph that consists of A and some or all of the l_i's is 1PI and is overall divergent. In coordinate space, these divergences come from the region where all vertices of A approach the same point; it

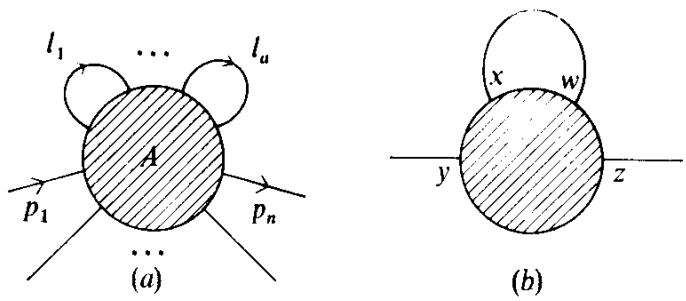


Fig. 11.3.2. General structure where subgraphs are divergent in momentum space, but appear to have no corresponding divergence in coordinate space.

appears that the lines $l_1,\ldots,l_a$ all participate in every one of the divergences.

In order to preserve the correct counterterm structure we must include counterterms for graphs of form A in our definition of renormalization. This ensures that counterterms are given by terms in the action. For simplicity consider the case in ϕ^4 theory of one line l with two external lines p_1, p_2. Schematically we have

$$\int \mathrm{d}^d w \mathrm{d}^d x \mathrm{d}^d y \mathrm{d}^d z S_F(x-w) A(w,x,y,z) f(y,z). \tag{11.3.3}$$

Here $S_F(x-w)$ is the propagator of line l, while $f(y,z)$ is analytic at $y=z$ and represents the rest of the graph. $A(w,x,y,z)$ is the value of graph A. We subtract off all subdivergences of the graph $A\cup\{l\}$. Among these is the overall counterterm for A, which is of the form

$$C_A(w,x,y,z;d,g,\mu) = C(d,g,\mu)\delta^{(d)}(w-z)\delta^{(d)}(x-z)\delta^{(d)}(y-z), \tag{11.3.4}$$

since A gives a logarithmic divergence. The result of replacing A by its counterterm is then

$$C(d,g,\mu)\int \mathrm{d}^d z f(z,z) S_F(0). \tag{11.3.5}$$

This counterterm is necessarily generated by the Feynman rules, since when A occurs in a four-point Green's function without additional loops it needs an overall counterterm. We wish to show how this counterterm is needed to cancel certain parts of the divergences of (11.3.3) as $y\to z$.

Let us now substitute each propagator by (11.3.2) with $N=2$. The propagator for line l may be replaced by S_{F1}, S_{F2}, or S_{rem}. The terms with S_{F1} and S_{F2} both contribute to a divergence at $y=z$, whether in the original graph (11.3.3) or in (11.3.5). The divergence from (11.3.5) with $S_F = S_{F1}$ or S_{F2} is local and temperature-independent, so that it can be cancelled by overall wave-function and mass counterterms. The terms with S_{rem} also give divergences, but cancel in the sum of (11.3.3) and (11.3.5), just as in the simplest case of Fig. 11.2.4.

The argument presented above is considerably more cumbersome than needed for the particular case considered. However it was presented so as to emphasize its form as a special case of a general argument. This case happens to be the only one present in Green's functions. However in the vacuum bubbles (used to compute ground-state energies) and in the presence of composite operators the degree of divergence can be higher, with the consequence of needing a more general proof.

In any event, the moral is that if we choose counterterms to cancel divergent subgraphs including those of form A in Fig. 11.3.2, then the counterterms need only depend on the $S_{F1}, S_{F2}, \ldots$. In particular, the value of $\bar{S}_{\text{Fana}}$ is irrelevant. It is only $\bar{S}_{\text{Fana}}$ that knows the boundary conditions, so only it knows the temperature. Note that the S_{Fi}'s are monomials in mass. Hence our counterterms are polynomials in mass parameters, as we saw by a totally different method in momentum space.

11.4 External fields

Consider as an example QED in the external electromagnetic field generated by a classical source J^μ. We have (in covariant gauge)

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}F_{\mu\nu}^2 + \bar{\psi}(\mathrm{i}\partial\!\!\!/ + eA\!\!\!/ - M)\psi - J^\mu A_\mu \\ &-\tfrac{1}{4}(Z_3 - 1)F_{\mu\nu}^2 + (Z_2 - 1)\bar{\psi}(\mathrm{i}\partial\!\!\!/ + eA\!\!\!/)\psi - \bar{\psi}\psi(Z_2 M_0 - M) \\ &-(1/2\xi)\partial\cdot A^2.\end{aligned} \tag{11.4.1}$$

To separate the classical and quantum parts of the electromagnetic field, we let $\mathscr{A}_\mu$ be a c-number potential that satisfies the classical Maxwell equations with source J^μ. Then we replace A_μ in (11.4.1) by $A_\mu + \mathscr{A}_\mu$, with the result

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}F_{\mu\nu}^2 - (1/2\xi)\partial\cdot A^2 + \bar{\psi}(\mathrm{i}\partial\!\!\!/ + e\mathscr{A}\!\!\!/ - M)\psi + e\bar{\psi}A\!\!\!/\psi \\ &-\tfrac{1}{4}(Z_3 - 1)F_{\mu\nu}^2 + (Z_2 - 1)\bar{\psi}(\mathrm{i}\partial\!\!\!/ + eA\!\!\!/)\psi - (Z_2 M_0 - M)\bar{\psi}\psi \\ &+\tfrac{1}{2}\mathscr{A}_\mu J^\mu + \tfrac{1}{2}(Z_3 - 1)\mathscr{A}_\mu J^\mu + (Z_3 - 1)A_\nu J^\nu \\ &+(Z_2 - 1)\bar{\psi}e\mathscr{A}\!\!\!/\psi + \text{total derivative}.\end{aligned} \tag{11.4.2}$$

We assume the gauge condition $\partial\cdot\mathscr{A} = 0$. The total electromagnetic field is $A_\mu + \mathscr{A}_\mu$, with the classical c-number part satisfying

$$\begin{aligned}\partial\cdot\mathscr{A} &= 0 \\ \Box\mathscr{A}^\mu &= J^\mu.\end{aligned} \tag{11.4.3}$$

If the field $\mathscr{A}_\mu$ is large then we are not allowed to expand in powers of $\mathscr{A}_\mu$. Indeed, as Schwinger (1951) pointed out, the electron propagator $S_F(x, y; e\mathscr{A})$ in the external field is not necessarily analytic at $\mathscr{A} = 0$.

Moreover, he showed that it is the non-analytic part that is relevant for pair production in an electric field. We will therefore do perturbation theory in e without assuming that $e\mathscr{A}_\mu$ is small. The lowest-order propagator (in powers of e) therefore satisfies the equation:

$$(\mathrm{i}\gamma^\mu\partial/\partial x^\mu + e\gamma^\mu\mathscr{A}_\mu - M)S_\mathrm{F}(x, y; e\mathscr{A}) = \mathrm{i}\delta^{(4)}(x-y)\mathbf{1}. \qquad (11.4.4)$$

Since S_F is no longer simple in momentum space, it is not clear that the renormalizations are the same as with $\mathscr{A}_\mu = 0$. Let us work in coordinate space. We construct a power series in $x - y$ to solve (11.4.4). The solution is a series singular at $x = y$, plus a series analytic there. We only need explicitly the first few singular terms, and to prove renormalizability we decompose S_F as in (11.3.2). It is important not to use an infinite series for S_F at that stage, because the series will not converge for general values of x, y and $\mathscr{A}$. The singular series is obtained by treating both e and M in (11.4.4) as perturbations and expanding in powers. This is similar to the way in which the singular part of the scalar propagator is an expansion in powers of m^2.

We obtain the leading power of $x - y$ by solving

$$\mathrm{i}\gamma^\mu\partial/\partial x^\mu S_{\mathrm{F}1} = \mathrm{i}\delta(x-y)\mathbf{1}. \qquad (11.4.5)$$

Each non-leading term $S_{\mathrm{F}n}$ is obtained in terms of earlier terms by solving at $x \neq y$ the equation

$$\mathrm{i}\gamma^\mu\partial/\partial x^\mu S_{\mathrm{F}n} = (M - e\mathscr{A}_\mu\gamma^\mu)\sum_{j=1}^{n-1} S_{\mathrm{F}j}. \qquad (11.4.6)$$

To obtain $S_{\mathrm{F}n}$ uniquely, we require it to be a singular power of $x - y$ times a function of y. The operation $\mathrm{i}\gamma^\mu\partial/\partial x^\mu$ makes $S_{\mathrm{F}n}$ more singular while multiplication by M or by $e\gamma^\mu\mathscr{A}_\mu$ leaves the degree of singularity unchanged.

After this work we see that the renormalizations are correctly given by treating as a perturbation the term $J^\mu A_\mu$ in the Lagrangian (11.4.1) before the shift of the field to $A_\mu + \mathscr{A}_\mu$. Since A_μ is the renormalized field, this term cannot affect the divergences – it simply tells us to integrate $J^\mu(x)$ separately with each of one or more external photon fields of an ordinary Green's function.

We see that, after the shift to $A_\mu + \mathscr{A}_\mu$, the divergences are correctly treated by expanding in powers of $\mathscr{A}$ and then taking the first few terms. The non-analyticity is entirely confined to the remainder term in S_F; this contributes to important physics, but not to the divergences.

12

Renormalization of gauge theories

It is important to show that renormalization of a gauge theory can be accomplished without violating its gauge invariance. Gauge invariance is physically important; among other things it is used (via the Ward identities) to show that the unphysical states decouple ('t Hooft (1971a)).

In Chapter 9 we considered the case that the basic Lagrangian of a theory is invariant under a global symmetry, as opposed to a gauge symmetry, such as we will be investigating in this chapter. We showed that the counterterm Lagrangian is also invariant under the symmetry. Suppose now that the basic Lagrangian is invariant under a gauge symmetry. One might suppose that the counterterms are also invariant under the symmetry, just as for a global symmetry. This is not true, however, since the introduction of gauge fixing (as explained in Sections 2.12 and 2.13) destroys manifest gauge invariance of the Lagrangian. One might instead point out that the theory with gauge fixing is BRS invariant and deduce that the counterterms are BRS invariant. This deduction is false. To see this, we recall that an ordinary internal symmetry relates Green's functions with certain external fields to other Green's functions differing only by change of symmetry labels. However, BRS symmetry relates a field to a composite field (2.13.1). This wrecks the proof of BRS invariance of counterterms except in an abelian theory, where the Faddeev–Popov ghost is a free field.

Before treating the non-abelian case, let us examine an abelian theory, QED. The Lagrangian is

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}F_{\mu\nu}^2 + \bar{\psi}(\mathrm{i}\partial\!\!\!/ + e_{\mathrm{R}}A\!\!\!/ - M)\psi \\ &-\tfrac{1}{4}(Z_3 - 1)F_{\mu\nu}^2 + (Z_2 - 1)\bar{\psi}(\mathrm{i}\partial\!\!\!/ + e_{\mathrm{R}}A\!\!\!/ - M)\psi \\ &- Z_2\bar{\psi}(M_0 - M)\psi - (1/2\xi)\partial\cdot A^2. \end{aligned} \tag{12.0.1}$$

Here e_{R} and M are the renormalized coupling and mass, while A_μ, ψ and $\bar{\psi}$ are the renormalized fields. We have chosen to include only counterterms invariant under the gauge transformation

$$\left.\begin{aligned}\psi &\to \mathrm{e}^{\mathrm{i}e_{\mathrm{R}}\omega}\psi, \\ \bar{\psi} &\to \mathrm{e}^{-\mathrm{i}e_{\mathrm{R}}\omega}\bar{\psi}, \\ A_\mu &\to A_\mu + \partial_\mu\omega. \end{aligned}\right\} \tag{12.0.2}$$

A special case of our later results will show that coefficients of the possible gauge non-invariant counterterms actually vanish. (We will write $e_R = \mu^{2-d/2}e$ if dimensional regularization is used, with e being the dimensionless renormalized charge.) In terms of unrenormalized quantities $A_\mu^0 \equiv Z_3^{1/2}A_\mu$, $\psi_0 \equiv Z_2^{1/2}\psi$, $e_0 \equiv Z_3^{-1/2}e_R$, and $\xi_0 \equiv \xi Z_3$, we have

$$\mathscr{L} = -\tfrac{1}{4}(\partial_\mu A_{0\nu} - \partial_\nu A_{0\mu})^2 + \bar{\psi}_0(i\partial\!\!\!/ + e_0 A\!\!\!/_0 - M_0)\psi_0 - (1/2\xi_0)\partial\cdot A_0^2. \tag{12.0.3}$$

The counterterm for the photon–electron vertex is, from (12.0.1),

$$(Z_2 - 1)e_R\bar{\psi}A\!\!\!/\psi, \tag{12.0.4}$$

which has a coefficient proportional to the wave-function renormalization. It is easy to verify that to one-loop order this result is correct. Our later results prove it to all orders.

If the non-abelian theory of (2.11.7) and (2.11.8) were renormalized by gauge-invariant counterterms, then the same relation of the $\bar{\psi}A\!\!\!/\psi$ counterterm to the quark field strength renormalization Z_2 would hold. It is easily verified by explicit calculation that this is false. As we pointed out above, the reason that the counterterms need not be BRS invariant is that the BRS invariance is not a symmetry that relates elementary fields to elementary fields. Even so, the counterterms are such that the Lagrangian is BRS invariant after renormalization, but under a renormalized BRS symmetry, as we will show in this chapter.

One result will be that whereas the gauge transformation of a fermion field in QED is given by $\delta\psi = ie_R\omega\psi$, with e_R being the renormalized charge, the BRS transformation of a fermion field in QCD is $\delta\psi = ig_R c^a t^a X\psi\delta\lambda$, where X is a divergent renormalization factor. The composite operator $\delta\psi$ is finite.

There are a number of strategies for proving renormalizability. Before explaining them, let us remark that the aim is to show that a finite theory exists which has gauge-invariance properties. The gauge invariance is exhibited by Ward identities. It is possible to choose counterterms in such a way that gauge invariance does not hold. For example, we could add a term $(A^\mu A_\mu)^2$ to the QED Lagrangian. Since the coefficient of this term is dimensionless, we obtain a finite theory by adding appropriate extra infinite counterterms. But this theory is not gauge invariant. This implies, for example, that negative metric states do not decouple from physical processes, and the theory is unphysical. The proofs state that it is possible to choose counterterms so that gauge invariance holds.

Several approaches can be distinguished:

1. *Invariant regulator.* Use an ultra-violet regulator that does not break gauge symmetry, for example dimensional regularization. Then Ward identities are true when the Lagrangian is given by (2.11.7), (2.12.5), and (2.12.9). Allow all parameters and fields to get renormalized. The theorems to be proved are that this is sufficient to make the theory finite when the regulator is removed. This is the traditional approach. The advantages center around the manifest preservation of gauge invariance. The disadvantages are that chiral symmetries cannot be regulated gauge invariantly; this symptomizes the fact that not all chiral symmetries can be preserved after quantization – see Chapter 13.

2. *Gauge invariant regulator + MS.* Let us again use dimensional regularization (or another gauge-invariant cut-off). But now let us choose to renormalize each separate graph by the forest formula or by the recursive method (as given in Chapter 5). We do not explicitly constrain the counterterms to satisfy gauge invariance; so in general we have violated gauge invariance. But if we use minimal subtraction then the counterterms are gauge invariant. The reason is simple: since we will prove that we can renormalize the theory gauge invariantly, the lowest-order counterterm that is not gauge invariant must be finite, after summing over all graphs of this order. But in minimal subtraction the only finite counterterm is zero. This method is of great use when renormalizing the complicated non-local operators that appear in generalized operator product expansions (Collins & Soper (1981)). We can renormalize the graphs without explicitly investigating the Ward identities. The disadvantage is that the method is closely tied to a specific renormalization prescription.

3. *Non-invariant regulator plus non-invariant counterterms.* One can use any regulator and adjust overall counterterms, if possible, to satisfy all the Ward identities (Piguet & Rouet (1981), Symanzik (1970a), 't Hooft (1971a), and Piguet & Sibold (1982a, b, c)).

There are two different forms of the Ward identities, either of which may be used. There are the Ward identities derived in Section 2.13 for Green's functions, and there are the ones for the 1PI graphs, as derived by Lee (1976). Our approach will use the BRS identities for Green's functions together with a combination of approaches 1 and 2. It is based on the treatment of Brandt (1976). Most other treatments have used the identities for the 1PI graphs.

We will restrict our attention to the simplest theories, like QCD. More general cases – with chiral or supersymmetries – are not treated here. See Chapter 13 for references.

12.1 Statement of results

For simplicity we will mainly treat one case: a theory of a gauge field coupled to a Dirac field, with the gauge-fixing term being the usual one, $-\sum_a(\partial\cdot A^a)^2/(2\xi)$. We assume that the gauge group is simple (in the mathematical sense); physically, this implies that there is only one independent gauge coupling. With a $U(1)$ gauge group and one or more Dirac fields, this theory is quantum electrodynamics. If the gauge group is $SU(3)$ and the matter fields are in the triplet representation, then we have quantum chromodynamics.

The result to be proved is that the Green's functions are made finite by renormalizing the values of all the parameters in the basic Lagrangian (2.11.7). These parameters are the gauge coupling, the fermion masses M, the field strength renormalizations, and the gauge-fixing parameter.

The resulting Lagrangian expressed in terms of renormalized fields is

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}Z_3 G_{\mu\nu}^{a2} + \sum_i Z_2^{(i)}\bar{\psi}_i(\mathrm{i}\not{D} - M_0^{(i)})\psi_i \\ &-(1/2\xi)\partial\cdot A^{a2} + \tilde{Z}\partial_\mu\bar{c}^a D^\mu C_a. \end{aligned} \tag{12.1.1}$$

We have allowed the fermion fields to be in several irreducible representations of the gauge group labelled by i. There are separate field-strength renormalizations $Z_2^{(i)}$ and bare masses $M_0^{(i)}$ for each representation. The covariant derivative is

$$\begin{aligned} D_\mu\psi &= (\partial_\mu + \mathrm{i}g_0 A_{0\mu}^a t^a)\psi \\ &= (\partial_\mu + \mathrm{i}g_R X\tilde{Z}^{-1}t^a A_\mu^a)\psi, \end{aligned} \tag{12.1.2}$$

where g_0 and g_R are the bare and renormalized couplings. (With dimensional regularization we write $g_R = \mu^{2-d/2}g$, with g dimensionless.) Following Lee (1976) we write the bare coupling as

$$g_0 = \frac{X}{\tilde{Z}Z_3^{1/2}}g_R, \tag{12.1.3}$$

so that the coupling of the gauge field to the ghost is Xg_R:

$$\tilde{Z}\partial_\mu\bar{c}^a D^\mu c_a = \tilde{Z}\partial_\mu\bar{c}^a\partial^\mu c_a + g_R X c_{abc}(\partial_\mu\bar{c}^a)c^b A_c^\mu. \tag{12.1.4}$$

The field strength tensor is

$$\begin{aligned} G_{\mu\nu}^a &= Z_3^{-1/2}G_{(0)\mu\nu}^a \\ &\equiv Z_3^{-1/2}(\partial_\mu A_{0\nu}^a - \partial_\nu A_{0\mu}^a - g_0 c_{abc}A_{0\mu}^b A_{0\nu}^c) \\ &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_R X\tilde{Z}^{-1}c_{abc}A_\mu^b A_\nu^c. \end{aligned} \tag{12.1.5}$$

Observe that in accordance with the results to be proved, the coefficient of the gauge-fixing term $(\partial\cdot A)^2$ is finite, when the renormalized field is used.

This gives a gauge-fixing term $(\partial \cdot A_0)^2/(2\xi_0)$, when expressed in terms of the bare field, with

$$\xi_0 = Z_3 \xi. \tag{12.1.6}$$

The main theorem to be proved is:

Theorem 1. The renormalizations

$$M_0^{(i)}, X, \tilde{Z}, Z_2^{(i)}, \text{and } Z_3 \tag{12.1.7}$$

can be chosen so that Green's functions of A, ψ, $\bar{\psi}$, c, and $\bar{c}$ are finite.

To prove this result we will use the Ward identities for BRS invariance. These involve a number of composite fields, which we also need to prove finite. The counterterms for these composite operators are related to the basic counterterms (12.1.7). We will prove:

Theorem 2.

$$\delta_{\text{BRS}}(\text{renormalized field})/\delta\lambda_{\text{R}}$$

is finite. That is, its Green's functions with any number of renormalized fields are finite. We define the renormalized BRS transformation $\delta_{\text{BRS}}/\delta\lambda_{\text{R}}$ as follows:

(1) Let the BRS transformation δ_{BRS} be defined by (2.13.1), (2.13.2) with g and ξ replaced by g_0 and ξ_0 and with the fields replaced by unrenormalized fields (i.e., $A \to A_0$, etc). Then the Lagrangian (12.1.1) is BRS invariant.

(2) Define $\delta\lambda_{\text{R}} = \delta\lambda Z_3^{-1/2}\tilde{Z}^{-1/2}$. Then

$$\left.\begin{aligned}
\delta_{\text{R}}\psi &= \delta_{\text{BRS}}\psi/\delta\lambda_{\text{R}} = -\mathrm{i}g_{\text{R}} X t^a \psi c_a, \\
\delta_{\text{R}}A_\mu^a &= \delta_{\text{BRS}}A_\mu^a/\delta\lambda_{\text{R}} = \partial_\mu c_a \tilde{Z} + g_{\text{R}} c_{abc} c^b A_\mu^c X, \\
\delta_{\text{R}}c^a &= \delta_{\text{BRS}}c^a/\delta\lambda_{\text{R}} = -\tfrac{1}{2}g_{\text{R}} X c_{abc} c^b c^c = -\tfrac{1}{2}g_{\text{R}} X c \wedge c, \\
\delta_{\text{R}}\bar{c}^a &= \delta_{\text{BRS}}\bar{c}^a/\delta\lambda_{\text{R}} = \frac{1}{\xi}\partial \cdot A^a.
\end{aligned}\right\} \tag{12.1.8}$$

Renormalized Ward identities follow from the unrenormalized ones by multiplication by $\delta\lambda/\delta\lambda_{\text{R}}$. The operators appearing in them are finite, because of Theorem 2.

Certain auxiliary operators are useful for reasons which only become apparent in proving Theorems 1 and 2.

$$\mathcal{O}_a = \tilde{Z}\Box c^a + \partial^\mu (c^b A_\mu^c) g_{\text{R}} X c_{abc}, \tag{12.1.9}$$

$$\mathcal{B}_{a\mu} = (\tilde{Z}/X - 1)\partial_\mu \bar{c}^a/g_{\text{R}} + c_{abc}\bar{c}_b A_\mu^c, \tag{12.1.10}$$

$$\bar{c} \wedge \bar{c} = c_{abc}\bar{c}_b\bar{c}_c. \tag{12.1.11}$$

The operator $\mathcal{O}_a$ is zero by the ghost equations of motion. We will prove:

Theorem 3. Green's functions of $\mathcal{O}_a$, $\mathcal{B}_{a\mu}$, and $\bar{c} \wedge \bar{c}$ with any number of basic fields are finite.

Theorem 4. Green's functions of $\mathcal{B}_{a\mu}$ with $\delta_R\phi$ and any number (greater than zero) of basic fields are finite, where ϕ is any basic field.

Theorem 5. Green's functions of $\bar{c} \wedge \bar{c}$ with one or two $\delta_R\phi$'s and any number (bigger than zero) of basic fields are finite.

The last few results have no intuitive appeal. They will be needed as part of an inductive proof of the important Theorem 1. We will also find it convenient to use CPT invariance of the theory (after dimensional regularization). Now reversal of one time and one space component is equivalent to reversal of components 0, 1, 2, and 3 with a spatial rotation. So to obtain the TP part of CPT, we need only consider reversal of the 0 and 1 components only. Therefore, we define

$$\theta_\nu^\mu = \begin{cases} -1, & \text{if } \mu = \nu = 0 \text{ or } 1, \\ +1, & \text{if } \mu = \nu \geq 2, \\ 0, & \text{otherwise.} \end{cases} \tag{12.1.12}$$

Let the fields transform under CPT as

$$\left.\begin{aligned} \psi(x) &\to \gamma^1 \bar{\psi}^{\mathrm{T}}(\theta x), \\ A^a(x) &\to A^a(\theta x), \\ c^a(x) &\to c^a(\theta x), \\ \bar{c}^a(x) &\to \bar{c}^a(\theta x). \end{aligned}\right\} \tag{12.1.13}$$

Then the theory is CPT invariant. (We use γ-matrices in which $\gamma^0 = \gamma^{0\mathrm{T}} = \gamma^{0*}$, and $\gamma^{\mu\dagger} = \gamma_\mu$.) Notice that the ghost field, c_a, transforms to itself, rather than to the antighost field, $\bar{c}^a$, even though these fields might be regarded as complex conjugate fields.

12.2 Proof of renormalizability

12.2.1 *Preliminaries*

The Ward identities of a gauge theory provide relations between different Green's functions. However, the identities mostly relate Green's functions of elementary fields to Green's functions containing the composite fields listed in (12.1.8)–(12.1.11). However, to prove renormalizability, we actually need relations between Green's functions of elementary fields only. Consequently proofs tend to be rather indirect and long.

The following references contain a representative selection of the proofs

in the literature: 't Hooft (1971a, b), Taylor (1971), Slavnov (1972), 't Hooft & Veltman (1972b), Lee & Zinn-Justin (1972), Becchi, Rouet & Stora (1975), Lee (1976), Itzykson & Zuber (1980), and Piguet & Rouet (1981). A great simplification was introduced by the discovery by Becchi *et al.* of their symmetry. However, the proofs still are mostly rather inexplicit. The proof to be given in this section gives all the steps needed to go from the basic Ward identities to the relations between the counterterms. The method follows that given by Brandt (1976) and Cvitanovic (1977). A point at which many proofs became rather inexplicit turns out in this method to be the point at which the operator $\mathscr{B}_{a\mu}$ (defined in (12.1.10)) makes its appearance. It is an unobvious operator to use, but its use is essential to completing the proof that the gauge coupling to matter fields is the same as the self-coupling of the gauge field.

The proof is by induction on the number, N, of loops. We assume that all graphs with less than N loops have been successfully renormalized by counterterms of the form implied by the Lagrangian (12.1.1). We also require that Green's functions of the composite operators considered in Theorems 2 to 5 are also finite up to $N-1$ loops if the indicated counterterms are used. The induction starts with tree graphs, which need no counterterms.

Our strategy is as follows:

(1) At each order below N loops we have values of the five independent renormalizations $\tilde{Z}$, Z_2, Z_3, M_0, and X. For each 1PI Green's function with an overall divergence a value of the overall counterterm is hence computed at each order less than N. Partition this counterterm into a set of counterterms to cancel the overall divergences of the individual graphs for the Green's function.
(2) Compute N-loop contributions to $\tilde{Z}$, Z_2, Z_3, M_0, and X by imposing renormalization conditions on certain Green's functions. The Ward identities will be true, but it is not immediate that the many other 1PI graphs are finite. This is done at step (3).
(3) Using these Ward identities show that Theorems 1 to 5 hold for N-loop graphs.

Step (1) is technical but important. Its use is that, in order to say that the only divergence of an N-loop graph is the overall divergence, we must have subtracted off its subdivergences. However, for individual graphs the constraints imposed by gauge invariance do not hold. Consider, for example, the two-loop graph, Fig. 12.2.1. Its self-energy subgraph needs a counterterm of the form $A_1 g_{\mu\nu} k^2 - B_1 k_\mu k_\nu$.

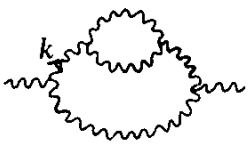


Fig. 12.2.1.

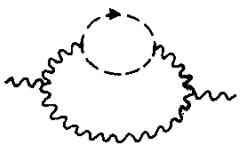


Fig. 12.2.2.

Similarly the subgraph of Fig. 12.2.2 has a counterterm $A_2 g_{\mu\nu} k^2 - B_2 k_\mu k_\nu$. As we will see, the Ward identities imply that the total self-energy counterterm is obtained from the term $-\frac{1}{4}(Z_3 - 1)(\partial_\mu A_\nu - \partial_\nu A_\mu)^2$ in $\mathscr{L}$. It follows that $A_1 + A_2 = B_1 + B_2$. However, the relation is false for the separate graphs, i.e., we have $A_1 \neq B_1$, $A_2 \neq B_2$.

At steps (2) and (3), where we discuss the N-loop counterterms, we then know that the only divergences are overall ones. Moreover, we know this without having to check on intricate series of cancellations between different graphs. But for the purposes of finding the constraints imposed by the Ward identities on the N-loop counterterms, it is convenient to consider a single overall counterterm for the sum of all N-loop graphs for a given 1PI Green's function. Having obtained these constraints, we decompose the N-loop counterterms into individual portions for each graph; this enables us to continue the induction at the next order.

If we use minimal subtraction, the counterterms can be obtained graph-by-graph without worrying about the constraints imposed by Ward identities. These constraints will be satisfied automatically. For example, in a general renormalization prescription the counterterms to the subgraphs in Figs. 12.2.1 and 12.2.2 have the form

$$A_i = g^2[a_i/(d-4) + a_i'],$$
$$B_i = g^2[b_i/(d-4) + b_i'],$$

where a_i' and b_i' are finite and depend on the renormalization prescription. The Ward identities tell us that we can renormalize the divergences by a transverse counterterm

$$(a_1 + a_2)/(d-4) + a_1' + a_2' = (b_1 + b_2)/(d-4) + b_1' + b_2'.$$

Evidently $a_1 + a_2 = b_1 + b_2$, and $a_1' + a_2' = b_1' + b_2'$. The first equation must always be satisfied, the second must be imposed by choice of renormalization prescription. Minimal subtraction with $a_1' = a_2' = b_1' = b_2' = 0$ always satisfies these equations.

12.2.2 *Choice of counterterms*

We now assume that step (1) has been carried out. The next step is to pick a set of 1PI Green's functions to fix $\tilde{Z}$, Z_2, Z_3, M_0, and X. This is somewhat

arbitrary, but our choice will determine the form of the rest of the proof. We choose the following set:

(1) The fermion self-energy has divergences proportional to $\not{p}$ and to 1. These are cancelled by counterterms in Z_2 and M_0.
(2) $\tilde{Z}$ is chosen to cancel the p^2 divergence in the ghost self-energy.
(3) X is chosen to make the ghost–gluon coupling finite as far as the c_{abc} part is concerned.
(4) Z_3 is chosen to cancel the part of the divergence of the gluon's self-energy that is proportional to $-g_{\mu\nu}k^2 + k_\mu k_\nu$.

Next we will examine the Green's functions used in Theorems 1 to 5 to check for possible divergences at N-loop order. Since all divergences at lower order are cancelled, the possible remaining divergences are overall divergences of N-loop 1PI Green's functions. These are Green's functions with either elementary external lines or with insertions of the various composite operators we use. The dimension of a 1PI Green's function must be zero or greater in order that it have a non-negative degree of divergence and thus be potentially divergent. The contributions to such a Green's function are (*a*) N-loop basic graphs, (*b*) graphs with counterterms to cancel subdivergences, (*c*) an overall N-loop counterterm derived from our knowledge of $\tilde{Z}$, Z_2, Z_3, M_0, and X. We must prove that the sum of these contributions is finite.

12.2.3 *Graphs with external derivatives*

There is a derivative on a ghost line where it exits from an interaction. Thus the 1PI graphs of Fig. 12.2.3 have negative degree of divergence even

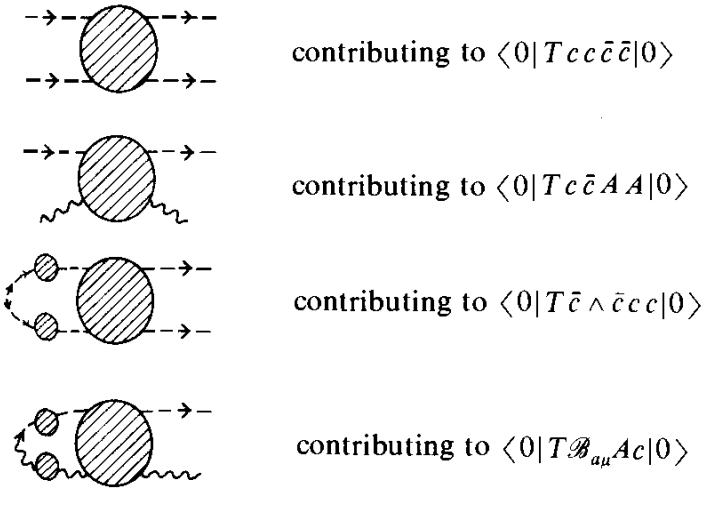


Fig. 12.2.3. Graphs with negative degree of divergence and non-negative dimension

though their dimension is zero. Subdivergences are cancelled and no overall counterterms are present, so the corresponding 1PI Green's functions are finite.

The same argument shows that the ghost self-energy needs no mass counterterm, but only a field-strength renormalization $\tilde{Z}$.

12.2.4 *Graphs finite by equations of motion*

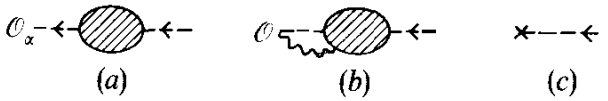
Consider Green's functions of $\mathcal{O}_a$ with basic fields. These are finite. For example,

$$\begin{aligned}\langle 0|T\mathcal{O}_a(x)\bar{c}^b(y)|0\rangle &= \tilde{Z}\Box_x\langle 0|Tc_a(x)\bar{c}^b(y)|0\rangle \\ &\quad + g_{\mathrm{R}}Xc_{adc}\frac{\partial}{\partial x^\mu}\langle 0|Tc_d(x)A^{c\mu}(x)\bar{c}^b(y)|0\rangle \\ &= -\mathrm{i}\delta^{ab}\delta^{(d)}(x-y). \qquad (12.2.1)\end{aligned}$$

The only graphs for the left-hand side with an N-loop 1PI subgraph are of the form Fig. 12.2.4(*a*) and (*b*). The graph (*a*) has a ghost self-energy made finite by its wave-function counterterm. Graph (*b*) needs a counterterm in $\mathcal{O}_a$ proportional to $\Box c_a$. Such a counterterm is graph (*c*), which has the N-loop contribution to the $\tilde{Z}\Box c_a$ term in $\mathcal{O}_a$. Finiteness of (12.2.1) shows that this is the correct counterterm. Similarly, the other potentially divergent Green's function of $\mathcal{O}_a$, viz.,

$$\langle 0|T\mathcal{O}_a\bar{c}^bA^c|0\rangle$$

is finite.



(*a*) (*b*) (*c*)

Fig. 12.2.4. Overall-divergent graphs for Green's functions of $\mathcal{O}_a$

12.2.5 *Gluon self-energy*

We have the Ward identity

$$\begin{aligned}0 &= \delta\langle 0|T\partial\cdot A^a(x)\bar{c}^b(y)|0\rangle/\delta\lambda_{\mathrm{R}} \\ &= (1/\xi)\langle 0|T\partial\cdot A^a(x)\partial\cdot A^b(y)|0\rangle - \langle 0|T\mathcal{O}^a(x)\bar{c}^b(y)|0\rangle \\ &= \frac{1}{\xi}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial y^\nu}\langle 0|TA^{a\mu}(x)A^{b\nu}(y)|0\rangle + \mathrm{i}\delta^{(d)}(x-y). \qquad (12.2.2)\end{aligned}$$

The second term involves $\mathcal{O}_a$ because of our definition of $\delta_{\mathrm{R}}A^a_\mu$. In the last line we used (12.2.1) on $\mathcal{O}_a$, and remembered that Green's functions of $\partial\cdot A$ have the derivatives outside the time-ordering, by definition.



Fig. 12.2.5. Gluon self-energy.

The only possible divergences in $\langle 0|TA_\mu^a(x)A_\nu^b(y)|0\rangle$ are in the N-loop self-energy graphs (Fig. 12.2.5), and they consist of terms proportional to $g_{\mu\nu}M^2$ or to $k_\mu k_\nu$. Remember that we used Z_3 to cancel any divergence proportional to $g_{\mu\nu}k^2 - k_\mu k_\nu$. Both the remaining divergences are absent, by (12.2.2). Thus no renormalization of the gauge parameter is needed, and no gluon mass is needed. The form of (12.2.2) is just as in QED. It implies that the gluon self-energy is purely transverse, so that the longitudinal part of the propagator is unchanged by higher-order correction. Hence, we can write:

$$\int \mathrm{d}^d x \mathrm{e}^{\mathrm{i}q\cdot x}\langle 0|TA_\mu^a(x)A_\nu^b(0)|0\rangle = \frac{\mathrm{i}\delta_{ab}}{q^2+\mathrm{i}\varepsilon}\left[\frac{-g_{\mu\nu}+q_\mu q_\nu/q^2}{1+\Pi(q^2)} - \frac{\xi q_\mu q_\nu}{q^2}\right]. \tag{12.2.3}$$

12.2.6 *BRS transformation of* A_μ^a.

Consider the Ward identity

$$\begin{aligned} 0 &= \delta\langle 0|TA_\mu^a(x)\bar{c}^b(y)|0\rangle/\delta\lambda_{\mathrm{R}} \\ &= (1/\xi)\langle 0|TA_\mu^a(x)\partial\cdot A^b(y)|0\rangle - \langle 0|T\delta_{\mathrm{R}}A^a(x)\bar{c}^b(y)|0\rangle. \end{aligned} \tag{12.2.4}$$

The first term we have just proved to be finite. The only potentially divergent graphs for the second term are shown in Fig. 12.2.6. Equation (12.2.4) shows their sum to be finite.

The $\bar{c}(y)A(z)$ Green's function (Fig. 12.2.7) of $\delta_{\mathrm{R}}A$ is also possibly divergent (logarithmically). But we have:

$$\begin{aligned} &\frac{\partial}{\partial x^\mu}\langle 0|T\delta_{\mathrm{R}}A^{a\mu}(x)\bar{c}^b(y)A_\nu^c(z)|0\rangle \\ &\quad = \langle 0|T\mathcal{O}^a(x)\bar{c}^b(y)A_\nu^c(z)|0\rangle \\ &\quad = -\mathrm{i}\delta^{(d)}(x-y)\delta^{ab}\langle 0|A_\nu^c(z)|0\rangle \\ &\quad = 0. \end{aligned} \tag{12.2.5}$$

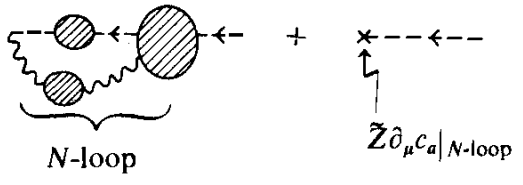


Fig. 12.2.6. Divergent graphs for right-hand side of (12.2.4).

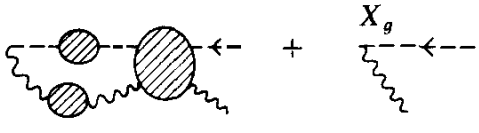


Fig. 12.2.7. Graphs for $\langle 0|T\delta_{\mathrm{R}}A^{a\mu}(x)\bar{c}^b(y)A_\nu^c(z)|\,0\rangle$.

Applying $\partial/\partial x^\mu$ is equivalent to multiplying by k_μ in momentum space. Since the divergence is at most logarithmic, (12.2.5) shows that there is no divergence at all.

12.2.7 *Gluon self-interaction*

Now

$$\begin{aligned} 0 &= \delta\langle 0|TA^a_\mu(x)A^b_\nu(y)\bar{c}^c(z)|0\rangle/\delta\lambda_R \\ &= \langle 0|T\delta_R A^a_\mu(x)A^b_\nu(y)\bar{c}^c(z)|0\rangle \\ &\quad - \langle 0|TA^a_\mu(x)\delta_R A^b_\nu(y)\bar{c}^c(z)|0\rangle \\ &\quad - \langle 0|TA^a_\mu(x)A^b_\nu(y)\delta_R\bar{c}^c(z)|0\rangle \\ &= \text{finite} + \frac{1}{\xi}\frac{\partial}{\partial z^\lambda}\langle 0|TA^a_\mu A^b_\nu A^{c\lambda}|0\rangle, \end{aligned} \tag{12.2.6}$$

where we used the previous result. The three-gluon vertex is linearly divergent, and we have a counterterm equal to $(Z_3 X\tilde{Z}^{-1} - 1)$ times the lowest-order vertex. There is no possible left-over divergence that satisfies (12.2.6).

Similarly

$$\frac{\partial}{\partial z^\nu}\langle 0|TA_\kappa(w)A_\lambda(x)A_\mu(y)A^\nu(z)|0\rangle = \text{finite}. \tag{12.2.7}$$

Here the only potentially divergent N-loop 1PI subgraphs are as in Fig. 12.2.8. We have just seen that the divergences in Fig. 12.2.8(*b*) are cancelled by the counterterm for the triple gluon coupling. Since the divergence in graph (*a*) is logarithmic, (12.2.7) proves that it is exactly cancelled by the counterterm in the four-gluon interaction.

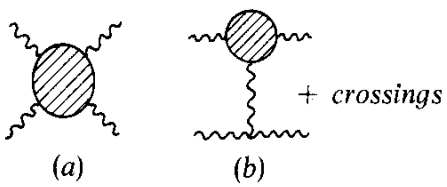


Fig. 12.2.8. Potentially divergent graphs for four-point Green's function of gluon.

12.2.8 $\delta_R c$

The only Green's function of $\delta_R c$ that could be divergent is

$$\langle 0|T\delta_R c\,\bar{c}\,\bar{c}|0\rangle. \tag{12.2.8}$$

But

$$\begin{aligned} 0 &= \delta\langle 0|T c^a(x)c^b(y)\bar{c}^c(z)|0\rangle \\ &= \langle 0|T\delta_R c^a(x)c^b(y)\bar{c}^c(z)|0\rangle \\ &\quad -(1/\xi)\langle 0|T c^a(x)\partial\cdot A^b(y)\bar{c}^c(z)|0\rangle \\ &\quad +(1/\xi)<0|T c^a(x)\bar{c}^b(y)\partial\cdot A^c(z)|0\rangle, \end{aligned} \tag{12.2.9}$$

so finiteness of (12.2.8) follows from finiteness of the ghost-gluon vertex, which is a renormalization condition.

12.2.9 *Quark–gluon interaction, $\delta\psi$, $\delta\bar{\psi}$; introduction of $\mathscr{B}_{a\mu}$*

Consider the Ward identity

$$\begin{aligned} 0 &= \delta\langle 0|T\psi(x)\bar{\psi}(y)\bar{c}(z)|0\rangle/\delta\lambda_R \\ &= \langle 0|T\delta_R\psi\bar{\psi}\bar{c}|0\rangle - \langle 0|T\psi\delta_R\bar{\psi}\bar{c}|0\rangle + \langle 0|T\psi\bar{\psi}(\partial\cdot A/\xi)|0\rangle. \end{aligned}$$

Finiteness of the last term follows if we can prove $\delta_R\psi = -ig_R X t^a\psi c_a$ finite. (Note that $\delta_R\bar{\psi}$ is related to $\delta_R\psi$ by the CPT transformation of (12.1.13).) Now X was defined by requiring the ghost-gluon vertex to be finite. An explicit proof, which we now give, brings in all the remaining operators listed in (12.1.9)–(12.1.11), and in particular $\mathscr{B}_{a\mu}$.

In Fig. 12.2.9 we list all Green's functions still to be proved finite. Observe that $\delta_R\bar{c} = \partial\cdot A/\xi$, so that it is finite if the gauge field is. Also Green's functions of $\mathscr{B}_{a\mu}$ or $\bar{c}\wedge\bar{c}$ with $\delta_R\phi$ and any number greater than zero of basic fields are finite, if the Green's functions of Figs. 12.2.7 to 12.2.9 are finite. (Here $\delta_R\phi$ is the BRS variation of any elementary field ϕ.) This proves Theorems 4 and 5, so it remains to prove finiteness of the Green's functions illustrated in Fig. 12.2.9.

The idea behind the proof is to examine the right-most vertex on the ghost line in Fig 12.2.9(a). This comes from the following term in the interaction Lagrangian:

$$\int c_{abc}g_R(\partial^\mu\bar{c}^a)c^b A^c_\mu d^4x. \tag{12.2.10}$$

(*a*) , logarithmically divergent, contributing to $\langle 0|T\delta_R\psi\bar{\psi}\bar{c}|0\rangle$

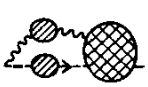
(*b*) , logarithmically divergent, contributing to $\langle 0|T\mathscr{B}_{a\mu}c|0\rangle$

Fig. 12.2.9. Green's functions not yet proved finite.

The derivative is on the line entering the vertex graph. By integrating by parts, we see that the vertex equals

$$-g_{\mathrm{R}}c_{abc}\int \mathrm{d}^4x[\bar{c}^a(\partial_\mu c^b)A^{c\mu}+\bar{c}^a c^b\partial\cdot A^c]. \tag{12.2.11}$$

In the first term the derivative is outside the loop-momentum integrals, so the degree of divergence is reduced by one. Factoring out the field $\partial_\mu c^b$ on the external line gives the basic vertex $c_{abc}\bar{c}^b A^c_\mu$ in the operator $\mathscr{B}_{a\mu}$. The second term in (12.2.11) contains $\partial\cdot A$ which we shall relate to something else by use of Ward identities.

We first prove finiteness of $\mathscr{B}_{a\mu}$, by formalizing the argument leading to (12.2.11). This is done in an unobvious way:

$$\begin{aligned}
&\Box_z\langle 0|T\delta_{\mathrm{R}}c^a(x)\bar{c}^b(y)\bar{c}^c(z)|0\rangle\\
&\quad=(1/X)\{\tilde{Z}\,\Box_z\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b\bar{c}^c|0\rangle\\
&\qquad\qquad+g_{\mathrm{R}}Xc_{cde}\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b(A^e_\mu\partial^\mu\bar{c}^d)|0\rangle\}\\
&\qquad+g_{\mathrm{R}}\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b(c_{cde}\bar{c}^d\partial\cdot A^e)|0\rangle\\
&\qquad-g_{\mathrm{R}}\frac{\partial}{\partial z^\mu}\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b\mathscr{B}^\mu_c(z)|0\rangle\\
&\quad=\mathrm{i}g_{\mathrm{R}}c_{acd}\langle 0|Tc^d(x)\bar{c}^b(y)|0\rangle\delta^{(d)}(x-z)\\
&\qquad+\tfrac{1}{2}g_{\mathrm{R}}\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b\delta_{\mathrm{R}}(c_{cde}\bar{c}^d\bar{c}^e)|0\rangle\\
&\qquad-g_{\mathrm{R}}\frac{\partial}{\partial z^\mu}\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b\mathscr{B}^\mu_c(z)|0\rangle\\
&\quad=\text{finite}-\tfrac{1}{2}g_{\mathrm{R}}\langle 0|T\delta_{\mathrm{R}}c^a\partial\cdot A^b(\bar{c}\wedge\bar{c})^c|0\rangle\\
&\qquad-g_{\mathrm{R}}\frac{\partial}{\partial z^\mu}\langle 0|T\delta_{\mathrm{R}}c^a\bar{c}^b\mathscr{B}^\mu_c(z)|0\rangle.
\end{aligned}\tag{12.2.12}$$

The next-to-last line follows by the antighost equation of motion. The last line uses a Ward identity plus the nilpotence property

$$\delta_{\mathrm{R}}[\delta_{\mathrm{R}}c^a]=0. \tag{12.2.13}$$

Now the second term on the last line is finite (by Theorem 5), and the left-hand side of (12.2.12) is finite. So the last term on the right is finite. The only possible uncancelled divergence is of the form of Fig. 12.2.10, from which finiteness of $\mathscr{B}_{a\mu}$ follows.

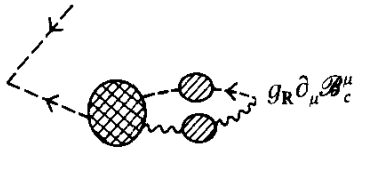


Fig. 12.2.10. Only possibly divergent graph for (12.2.12).

Finiteness of $\delta_R\psi$ follow by the same manipulations applied to $\langle 0|T\delta_R\psi\bar{\psi}\bar{c}|0\rangle$:

$$\begin{aligned}\Box_z\langle 0|T\delta_R\psi(x)\bar{\psi}(y)\bar{c}^a(z)|0\rangle &= g_R\langle 0|Tt^a\psi(x)\bar{\psi}(y)|0\rangle\delta^{(d)}(x-z)\\ &\quad -\tfrac{1}{2}g_R\langle 0|T\delta_R\psi\,\delta_R\bar{\psi}(\bar{c}\wedge\bar{c})|0\rangle\\ &\quad -g_R\frac{\partial}{\partial z^\mu}\langle 0|T\delta_R\psi(x)\bar{\psi}(y)\mathscr{B}_a^\mu(z)|0\rangle.\end{aligned} \tag{12.2.14}$$

12.3 More general theories

In the last section we proved renormalizable the simplest gauge theories: that is, those with a gauge group with a single component and with fermion matter fields. More general cases can easily be treated by the same methods. The general result is that renormalizations are needed for each independent coupling in the basic Lagrangian and for the field strength for each irreducible field multiplet. Let us examine some specific generalizations.

12.3.1 *Bigger gauge group*

The gauge group can in general be a product of several components: $G=\prod_{i=1}^n G_i\otimes U(1)^\nu$. Here each G_i is a simple compact non-abelian group (like $SU(N)$), and there are ν abelian $U(1)$'s. For each G_i and for each $U(1)$ factor there is an independent coupling g_i. When we perform gauge fixing there will be a multiplet of ghost fields for each component of the gauge group. The proof of renormalizability will need no change. It will relate the renormalizations of the couplings within each multiplet. Thus for each of the $n+\nu$ components of the gauge group there are renormalization factors X_i, Z_{3i}, and $\tilde{Z}_i$, for the coupling, the gauge field and the ghost field. In addition there are the usual renormalizations for the matter fields.

There are some special features of the abelian case which we will treat in Section 12.9.

12.3.2 *Scalar matter*

The part of the Lagrangian for a scalar field coupled to gauge fields is

$$D^\mu\phi^\dagger D_\mu\phi - V(\phi^\dagger,\phi). \tag{12.3.1}$$

Here V is a function of ϕ and $\phi^\dagger$ that is invariant under the gauge group, and D_μ is the usual covariant derivative.

For example, consider an $SU(2)$ gauge theory in which ϕ is a doublet

under the gauge group. Then the most general renormalizable form of V is

$$V = m^2\phi^\dagger\phi + \tfrac{1}{4}\lambda(\phi^\dagger\phi)^2, \tag{12.3.2}$$

while

$$\begin{aligned} D_\mu\phi^\dagger D\phi &= |\partial_\mu\phi|^2 - igA^{a\mu}\phi^\dagger t^a\overleftrightarrow{\partial}_\mu\phi + g^2A_\mu^aA^{b\mu}\phi^\dagger t^at^b\phi \\ &= |\partial_\mu\phi|^2 - igA_\mu^a\phi^\dagger t^a\overleftrightarrow{\partial}^\mu\phi + \tfrac{1}{4}g^2A^{a2}\phi^\dagger\phi. \end{aligned} \tag{12.3.3}$$

Wave-function and mass renormalization are used to make the propagator finite, and the $\phi^\dagger\partial\phi A$ coupling is proved finite as for a fermion. The self-coupling of the ϕ-field is made finite by a renormalization of the $\lambda\phi^\dagger\phi^2$ term, which is the only four-point coupling invariant under global $SU(2)$ transformations. One further Green's function, viz., $\langle 0|TAA\phi^\dagger\phi|0\rangle$ has a potential logarithmic divergence. It is proved finite by the Ward identity:

$$\begin{aligned} 0 &= \delta\langle 0|T\bar{c}A\phi^\dagger\phi|0\rangle \\ &= (1/\xi)\langle 0|T\partial\cdot AA\phi^+\phi|0\rangle + \langle 0|T\bar{c}\delta A\phi^\dagger\phi|0\rangle \\ &\quad + \langle 0|T\bar{c}A\delta\phi^\dagger\phi|0\rangle + \langle 0|T\bar{c}A\phi^\dagger\delta\phi|0\rangle. \end{aligned} \tag{12.3.4}$$

12.3.3 *Spontaneous symmetry breaking*

Consider the abelian Higgs model as an easy example. Its basic Lagrangian is

$$\mathscr{L}_{\text{basic}} = -\tfrac{1}{4}F_{\mu\nu}^2 + (\partial_\mu - ieA_\mu)\phi^\dagger(\partial^\mu + ieA^\mu)\phi - \tfrac{1}{4}\lambda^2(\phi^\dagger\phi - f^2/2\lambda^2)^2. \tag{12.3.5}$$

The symmetry is spontaneously broken with $\langle 0|\phi|0\rangle = f/(\lambda\sqrt{2})$, in lowest order. So we write $\phi = [f/\lambda + (\phi_1 + i\phi_2)]/\sqrt{2}$:

$$\begin{aligned} \mathscr{L}_{\text{basic}} = &-\tfrac{1}{4}F_{\mu\nu}^2 + (e^2f^2/2\lambda^2)A^2 + (ef/\lambda)A_\mu\partial^\mu\phi_2 + \tfrac{1}{2}\partial\phi_1^2 + \tfrac{1}{2}\partial\phi_2^2 - f^2\phi_1^2/4 \\ &- \lambda^2(\phi_1^2 + \phi_2^2)^2/16 - \lambda f\phi_1(\phi_1^2 + \phi_2^2)/4 \\ &+ eA_\mu\phi_1\partial^\mu\phi_2 + e^2A^2(\phi_1^2 + \phi_2^2 + 2f\phi_1/\lambda)/2. \end{aligned} \tag{12.3.6}$$

If we quantize with the simple gauge fixing term

$$\mathscr{L}_{\text{gf}} = -\frac{1}{2\xi}(\partial\cdot A)^2, \tag{12.3.7}$$

then renormalization is covered by the discussion of Section 9.2, where we treated spontaneous symmetry breaking in a non-gauge theory. We first renormalize in the unbroken theory (i.e., with $f^2 \to -f^2$). We need wave-function renormalizations Z_2 and Z_3, coupling renormalization $\lambda^2 \to \lambda_R^2 Y$, and mass renormalization (which is effectively $f^2 \to f^2 Z_m$). Since the Faddeev–Popov ghost is a free field, the gauge-coupling renormalization is $X = 1$. After spontaneous symmetry breaking the same renormalizations

make the Green's functions finite. Unfortunately the $A_\mu \partial^\mu \phi_2$ term in the free Lagrangian makes the Feynman rules rather messy, so it is convenient to use another gauge condition. We will discuss this in Section 12.5.

12.4 Gauge dependence of counterterms

To quantize and renormalize a gauge theory, we choose to fix the gauge in a particular way. It is important to show that physical quantities are independent of this choice. On the other hand, Green's functions of the elementary fields can certainly be gauge dependent. Since it is the Green's functions with which we work when renormalizing the theory, we must understand their gauge dependence and its effect on the renormalizations.

One important class of physical quantities which we will treat is the set of Green's functions of gauge-invariant operators. Also important is the S-matrix for physical states. In a spontaneously broken theory there are particle states that couple to the elementary fields of the theory. Some such states are physical, and we must prove that their S-matrix is gauge independent. But other states, like the Faddeev–Popov ghost, are unphysical. We will not attempt a complete treatment here.

When a gauge symmetry is unbroken there may even be no states coupled to the elementary fields. Indeed, it is commonly expected that in QCD colored states are confined. Certainly, within perturbation theory there are severe infra-red problems in obtaining the S-matrix for quarks and gluons. However, S-matrix elements of hadrons can be obtained from gauge-invariant Green's functions. Consider, for example:

$$\langle 0 | T j^a_{5\kappa}(w) j^b_{5\lambda}(x) j^c_{5\mu}(y) j^d_{5\nu}(z) | 0 \rangle, \tag{12.4.1}$$

where $j^a_{5\mu}$ is the axial isospin current. Application of the LSZ formalism will give the S-matrix for $\pi\pi \to \pi\pi$ scattering. Similarly the gauge-invariant field $\psi_i \psi_j \psi_k \varepsilon_{ijk}$ is an interpolating field for baryons. Here ψ_i is a quark field and i its color label.

There are two sorts of gauge-dependence that we will consider. The first is where we change the gauge-fixing function F_a to a different function of the fields. A specific application of our general results for this case will be given in Section 12.5 for the R_ξ-gauge. The second type of gauge-dependence is a variation of the parameter ξ. Although variation of ξ by a factor λ is equivalent to variation of F_a by a factor $\lambda^{-1/2}$, there are a number of special simplifications that will lead us to treat this second case first.

The use of BRS invariance will make our computation of gauge dependence very simple.

12.4.1 *Change of* ξ

The Green's functions and the renormalization factors in general depend on ξ. We can use the action principle to compute the dependence on ξ of a Green's function $\langle 0|TX|0\rangle$. Here X denotes any product of local fields with no explicit dependence on ξ. Then

$$\begin{aligned}\frac{\partial}{\partial\xi}\langle 0|TX|0\rangle &= \mathrm{i}\int \mathrm{d}^4y\langle 0|T{:}\frac{\partial\mathscr{L}}{\partial\xi}(y){:}X|0\rangle \\ &= \mathrm{i}\sum_Y \frac{\partial Y}{\partial\xi}\int \mathrm{d}^4y\langle 0|T{:}\frac{\partial\mathscr{L}}{\partial Y}(y){:}X|0\rangle \\ &\quad + \frac{\mathrm{i}}{2\xi^2}\int \mathrm{d}^4y\langle 0|T{:}F_a^2(y){:}X|0\rangle. \end{aligned} \tag{12.4.2}$$

In the first term we use Y to denote any of the independent renormalization factors. As usual, we use the symbol: : to indicate subtraction of the vacuum expectation value. Next we use the form of the BRS transformations to write

$$(1/\xi)F_a^2 = \delta(\bar{c}_aF_a) - \bar{c}_a\delta F_a, \tag{12.4.3}$$

and substitute for F_a^2 in (12.4.2). We can use the Ward identity

$$\delta\langle 0|T\bar{c}FX|0\rangle = 0$$

and the equation of motion

$$\langle 0|T\bar{c}_a\delta F_a(y)X|0\rangle = -\mathrm{i}\sum_a\langle 0|T\bar{c}_a\delta X/\delta\bar{c}_a(y)|0\rangle$$

to give

$$\begin{aligned}\frac{\partial}{\partial\xi}\langle 0|TX|0\rangle &= \mathrm{i}\sum_Y \frac{\partial Y}{\partial\xi}\int \mathrm{d}^4y\langle 0|Y{:}\frac{\partial\mathscr{L}}{\partial Y}{:}X|0\rangle \\ &\quad - \frac{\mathrm{i}}{2\xi}\int \mathrm{d}^4y\langle 0|T{:}\bar{c}_aF_a{:}\delta X|0\rangle + (N_c/2\xi)\langle 0|TX|0\rangle. \end{aligned} \tag{12.4.4}$$

Here N_c is the number of factors of the ghost field in X.

Suppose first that we choose to use the same counterterms for all values of ξ as at some particular value, say $\xi = 0$. Then Green's functions of elementary fields will not be finite except at this special value. But if we take a Green's function of gauge-invariant operators then (12.4.4) indicates that the Green's function is gauge independent, for such operators contain no ghosts.

Next consider a Green's function of elementary fields:

$$\langle 0|TX|0\rangle = \langle 0|T\prod_{i=1}^{N}\phi_i(x_i^\mu)|0\rangle. \tag{12.4.5}$$

The renormalizations in $\mathscr{L}$ are adjusted to keep Green's functions finite for any value of ξ. Then in (12.4.4) we find the ξ-dependence of the counterterms by requiring the

$$\sum_Y \frac{\partial Y}{\partial \xi}:\frac{\partial \mathscr{L}}{\partial Y}:$$

term to be the sum of the counterterms needed to cancel the divergences of $-:\bar{c}_a F_a:/(2\xi)$.

We consider the ultra-violet divergences of

$$\langle 0|T:\bar{c}_a F_a(y):\delta X|0\rangle. \tag{12.4.6}$$

Divergences in (12.4.6) occur either when y coincides with some set of interaction vertices or when it coincides with one of the x_i's, which are the positions of fields in X. We can renormalize Green's functions of $:\bar{c}_a F_a:$ with elementary fields by adding to it an operator $:D(y):$. Since we choose X to be a product of elementary fields, the only remaining divergences in (12.4.6) are when y coincides with the position x of a BRS varied field $\delta\phi(x)$ in δX. To renormalize this divergence we need counterterms to (12.4.6) of the form

$$\langle 0|T:D(y):\delta X|0\rangle + \sum_{i=1}^{N}\delta^{(4)}(y-x_i)\langle 0|TX|0\rangle|_{\phi_i(x_i)\to -iE_{\phi_i}(x_i)} \tag{12.4.7}$$

The counterterm when y coincides with $\delta\phi(x)$ has been written as $-iE_\phi(x)\delta(y-x)$, there are possibly derivatives of the δ-function, and the normalization factor $-i$ will be convenient later.

By use of Ward identities and equations of motion, we can write (12.4.7) as

$$-\langle 0|T:\delta D(y):X|0\rangle - \sum_\phi \langle 0|T:E_\phi(y)\delta S/\delta\phi(y):X|0\rangle. \tag{12.4.8}$$

This is the most general form of counterterm needed to keep the Green's function finite as ξ varies. It is even the correct form for the change in counterterms caused by a change in the form of F_a, as we will see in Section 12.4.4. To proceed any further we must do power-counting to determine which counterterms actually occur.

We now restrict attention to the gauge condition $F_a = \partial\cdot A_a$. The only divergent elementary Green's functions of $:\bar{c}_a F_a:$ are illustrated by their



Fig. 12.4.1. Lowest-order graph for $\langle 0|Tc:\bar{c}_aF_a:|0\rangle$.

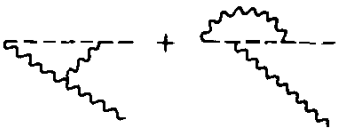


Fig. 12.4.2. Lowest-order graph for $\langle 0|Tc\,A:\bar{c}_aE_a:|0\rangle$

lowest-order cases in Figs. 12.4.1 and 12.4.2. The divergence of the first graph has a factor of the derivative at the ghost interaction. The counterterm for Fig. 12.4.2 is proportional to $(\partial_\mu \bar{c}_a)A_a^\mu$ = derivative $-\bar{c}_aF_a$. Since we integrate over all y, the only contribution to the counterterm operator D is proportional to $\bar{c}_aF_a$.

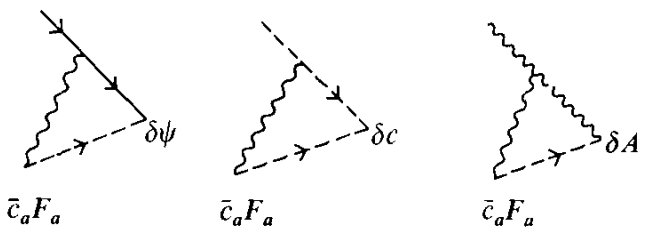


Fig. 12.4.3. Lowest-order graphs for $:\bar{c}_aF_a:$ with BRS variation of a field.

The divergence with a BRS-varied field $\delta\phi$ is logarithmic and proportional to ϕ, as in the graphs of Fig. 12.4.3. So we have $E_\phi \propto \phi$. This corresponds to a variation with ξ of the wave-function renormalizations with g_0 and M_0 fixed. For example,

$$\frac{\partial Z_2}{\partial \xi}\frac{\partial \mathscr{L}}{\partial Z_2} = \frac{1}{2}\frac{\partial \ln Z_2}{\partial \xi}(\psi \delta S/\delta\psi + \bar{\psi}\delta S/\delta\bar{\psi}) \tag{12.4.9}$$

A complication arises since we know ξ is not renormalized:

$$\begin{aligned}\frac{\partial Z_3}{\partial \xi}\frac{\partial \mathscr{L}}{\partial Z_3} &= \frac{1}{2}\frac{\partial \ln Z_3}{\partial \xi} A_\mu^a \frac{\delta}{\delta A_\mu^a}(S + F_a^2/2\xi) \\ &= \frac{1}{2}\frac{\partial \ln Z_3}{\partial \xi}(A \cdot g\delta S/\delta A + F_a^2/2\xi).\end{aligned} \tag{12.4.10}$$

Putting all this work together, we find

$$\begin{aligned}\frac{\partial}{\partial \xi}\langle 0|TX|0\rangle = &-\frac{i}{2\xi}\left(1 + \frac{\partial \ln Z_3}{\partial \ln \xi}\right)\int d^4y \langle 0|T:\bar{c}_aF_a(y):\delta X|0\rangle \\ &- \langle 0|TX|0\rangle\left\{N_\psi \frac{\partial \ln Z_2}{\partial \xi} + \frac{N_A}{2}\frac{\partial \ln Z_3}{\partial \xi}\right. \\ &\left. + N_c\left[\frac{\partial \ln(\tilde{Z}Z_3^{1/2})}{\partial \xi} + \frac{\xi}{2}\right]\right\}.\end{aligned} \tag{12.4.11}$$

This is just the counterterm structure we need. The operator $\bar{c}_aF_a$ is

multiplicatively renormalized by a factor $1+\xi\partial \ln Z_3/\partial\xi$. (In an abelian theory no renormalization is needed since $\bar{c}_a$ is a free field; so Z_3 is independent of ξ.) The other terms (where N_ϕ is the number of external ϕ fields of X) come from renormalizations of 1PI graphs including both $\bar{c}_a F_a$ and a $\delta\phi$.

Note that $\partial \ln Z_3/\partial\xi$ occurs in two places. This comes from assuming that the gauge-fixing term is $-F_a^2/(2\xi) = -\partial\cdot A^2/(2\xi)$ with no extra renormalization factor. We proved this from the Ward identity:

$$\begin{aligned} 0 &= \delta\langle 0|T\bar{c}_a(x)F_b(y)|0\rangle \\ &= (1/\xi)\langle 0|TF_a(x)F_b(y)|0\rangle + \langle 0|T\bar{c}_a\delta F_b|0\rangle \\ &= (1/\xi)\langle 0|TF_a(x)F_b(y)|0\rangle - i\delta(x-y)\delta_{ab}, \end{aligned} \tag{12.4.12}$$

where the last line follows from the ghost equation of motion. Since $F_a = \partial\cdot A_a$, its Green's functions are finite, and so ξ is finite.

We can use (12.4.11) to prove gauge invariance of the S-matrix by picking out the residue when we go to the particle pole for each external line.

If we use the correct, ξ-dependent counterterms in $\mathscr{L}$, then a gauge-invariant operator like

$$(G^a_{\mu\nu})^2 = (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_0 Z_3^{1/2} c_{abc} A^b_\mu A^c_\nu)^2 \tag{12.4.13}$$

is actually ξ-dependent; from (12.4.11) we see that is proportional to Z_3^{-1}. But the bare operator

$$(G^a_{(0)\mu\nu})^2 = (\partial_\mu A^a_{(0)\nu} - \partial_\nu A^a_{(0)\mu} + g_0 c_{abc} A^b_{(0)\mu} A^c_{(0)\nu})^2$$

is gauge independent. Both operators have UV divergences, so that they must be renormalized.

12.4.2 *Change of F_a*

Let us assume that F_a depends on a parameter κ. This allows us to interpolate between two different gauge-fixing conditions. We follow the same method as for the ξ-dependence. The change in the Lagrangian is

$$\begin{aligned} \frac{\partial\mathscr{L}}{\partial\kappa} &= \sum_Y \frac{\partial Y}{\partial\kappa}\frac{\partial\mathscr{L}}{\partial Y} - \frac{1}{\xi}F_a\frac{\partial F_a}{\partial\kappa} - \bar{c}_a\delta\frac{\partial F_a}{\partial\kappa} \\ &= \sum_Y \frac{\partial Y}{\partial\kappa}\frac{\partial\mathscr{L}}{\partial Y} - \delta\left(\bar{c}_a\frac{\partial F_a}{\partial\kappa}\right). \end{aligned} \tag{12.4.14}$$

So

$$\begin{aligned} \frac{\partial}{\partial\kappa}\langle 0|TX|0\rangle &= i\sum_Y \frac{\partial Y}{\partial\kappa}\int d^4y\langle 0|T:\frac{\partial\mathscr{L}(y)}{\partial\kappa}:X|0\rangle \\ &\quad + i\int d^4y\langle 0|T:\bar{c}_a\frac{\partial F_a}{\partial\kappa}:\delta X|0\rangle. \end{aligned} \tag{12.4.15}$$

The counterterms for the second term have exactly the same form as we met in the ξ-dependence of a Green's function, i.e., the form (12.4.8). Until the form of F_a is specified we can make no further statement about the necessary counterterms. We will examine an example of this in Section 12.5.

In the most general case we can take a Green's function of gauge-invariant operators and omit any κ-dependence of the counterterms in $\mathscr{L}$. Then the Green's functions are gauge-independent: $\partial\langle 0|TX|0\rangle/\partial\kappa = 0$.

12.5 R_ξ-gauge

To eliminate the mixing between the gauge field and the scalar field in spontaneously broken gauge theories one can use the R_ξ-gauge devised by 't Hooft (1971b), Fujikawa, Lee & Sanda (1972), and Yao (1973). Let us use the Higgs model of (12.3.5) as an example. The R_ξ-gauge is defined by using the gauge-fixing Lagrangian

$$\mathscr{L}_{\text{gf}} = -(1/2\xi)(\partial\cdot A - \kappa\xi m\phi_2)^2. \tag{12.5.1}$$

We use the parameter κ to interpolate between the R_ξ-gauge, where $\kappa = 1$, and the gauges we used in Section 12.2, where $\kappa = 0$. Also, $m = ef/\lambda$.

An interacting Faddeev–Popov ghost is needed, even in an abelian theory (which is the only case we will explicitly treat):

$$\mathscr{L}_{\text{gc}} = \partial_\mu\bar{c}\partial^\mu c - \kappa\xi m^2\bar{c}c - e\xi m\kappa\phi_1\bar{c}c. \tag{12.5.2}$$

The BRS variations of the fields are:

$$\left.\begin{aligned} \delta A_\mu &= \partial_\mu c, \\ \delta\phi_2 &= -(m + e\phi_1)c, \\ \delta\phi_1 &= e\phi_2 c, \\ \delta\bar{c} &= (\partial\cdot A - \xi\kappa m\phi_2)/\xi = F/\xi, \\ \delta c &= 0. \end{aligned}\right\} \tag{12.5.3}$$

Observe first of all that the extra couplings relative to the gauge $\kappa = 0$ are all super-renormalizable, so that the renormalizations of the dimensionless couplings can remain unchanged. These are X (which equals unity in an abelian theory), Y, Z_2, and Z_3. To determine what further renormalizations are needed, we use the method of Section 12.4.2, with $\partial F/\partial\kappa = -\xi m\phi_2$.

Thus we need the counterterms to

$$-\mathrm{i}\xi m\int \mathrm{d}^4y\langle 0|T{:}\bar{c}\phi_2(y){:}\delta X|0\rangle. \tag{12.5.4}$$

The only divergent graphs contain $\bar{c}\phi_2$, the BRS variation $\delta\phi_1$, and no other external lines. They have the form of Fig. 12.5.1, which is logarithmi-

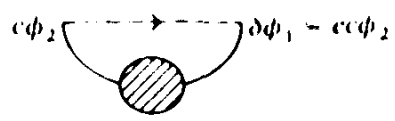


Fig. 12.5.1. Divergent graphs for (12.5.4).

cally divergent. Notice that interactions on the ghost line would make the graph convergent. The counterterm to (12.5.4) therefore has the form

$$C\int d^4y\langle 0|T\delta X/\delta\phi_1(y)|0\rangle$$
$$= iC\int d^4y\langle 0|T:\delta S/\delta\phi_1(y):X|0\rangle. \quad (12.5.5)$$

This counterterm is generated by a shift of ϕ_1 in the Lagrangian:

$$\mathscr{L}(\phi_1 + m, e, \phi_2, A, c, \bar{c}) \to \mathscr{L}(\phi_1 + Bm/e, \phi_2, A, c, \bar{c}) \quad (12.5.6)$$

with

$$\frac{m}{e}\frac{\partial B}{\partial \kappa} = C. \quad (12.5.7)$$

As far as renormalization is concerned, the effect of using the new gauge condition is to generate an ultra-violet divergence in the vacuum expectation value of the scalar field. Although this might appear strange, it is permitted to happen since the field is not gauge invariant.

Gauge-invariant Green's functions are unchanged, of course, as is the S-matrix. The ghost field, ϕ_2, and the longitudinal part of A_μ all couple to unphysical states. In this abelian theory, $F_{\mu\nu}$ is a gauge-invariant field which couples to the transverse part of the gluon, while

$$Z_m Z_2 \phi^\dagger\phi = Z_m Z_2[(\phi_1 + Bm/e)^2 + \phi_2^2]$$
$$= Z_m Z_2 (Bm/e)^2 + \phi_1[2(Bm/e)Z_m Z_2] + \cdots$$

is gauge invariant and renormalized, and couples to the ϕ_1-particle.

The Lagrangian can be written out in terms of the fields, to exhibit the interactions. It is rather fearsome-looking. We set $\kappa = 1$ and find

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4}Z_3 F_{\mu\nu}^2 - (1/2\xi)\partial\cdot A^2 + \tfrac{1}{2}m^2 B^2 Z_2 A_\mu^2 + \partial\bar{c}\partial c\\
&- m^2 B\xi\bar{c}c + A_\mu\partial^\mu\phi_2 m(Z_2 B - 1)\\
&+ \tfrac{1}{2}Z_2(\partial\phi_1^2 + \partial\phi_2^2) - \tfrac{1}{8}\phi_1^2 m^2\lambda^2 e^{-2}(3B^2 Y - Z_m)\\
&- \tfrac{1}{2}\phi_2^2 m^2[\xi + \tfrac{1}{4}\lambda^2 e^{-2}(B^2 Y - Z_m)]\\
&- \tfrac{1}{16}\lambda^2 Y(\phi_1^2 + \phi_2^2)^2 - \tfrac{1}{4}\lambda^2 m e^{-1} B Y\phi_1(\phi_1^2 + \phi_2^2)\\
&- \tfrac{1}{4}\lambda^2 m^3 e^{-3} B(B^2 Y - Z_m)\phi_1\\
&+ eZ_2 A_\mu\phi_1\overleftrightarrow{\partial^\mu}\phi_2 + \tfrac{1}{2}e^2 Z_2 A_\mu^2(\phi_1^2 + \phi_2^2)\\
&+ emBZ_2\phi_1 A_\mu^2 - \xi m\phi_1\bar{c}c. \quad (12.5.8)\end{aligned}$$

The same methods can be applied to a non-abelian theory with extra complications in the structure of the vertices – see Fujikawa, Lee & Sanda (1972).

12.6 Renormalization of gauge-invariant operators

In applications such as to deep-inelastic scattering we need the operator product expansion of gauge-invariant operators. It is natural to assume that only gauge-invariant operators appear in the expansion. Moreover, in using the operator product expansion, we need the anomalous dimensions of the operators. To compute these, we need the expression for the renormalized operators in terms of the bare operators. Again, it is natural to assume that only gauge-invariant operators are needed.

Both assumptions, taken literally, are false (Dixon & Taylor (1974), and Kluberg-Stern & Zuber (1975)). We will first treat the renormalization problem, where certain gauge-variant operators mix with gauge-invariant operators. As we will see, the gauge-variant operators vanish in physical matrix elements (Joglekar & Lee (1976), Joglekar (1977a and b)). We will simplify many parts of the proof by working in coordinate space. Then we will apply the same methods to find the operators that appear in the operator product expansion.

In the case of an ordinary global symmetry (like Lorentz invariance or isospin), an operator that is invariant under the symmetry mixes only with invariant operators. In the case of a gauge theory, we break the invariance of the action by gauge-fixing, leaving only a BRS invariance. But a gauge-invariant operator even mixes with operators that are not BRS invariant. How can this be?

To answer this question, we consider an operator $\mathcal{O}$ that is invariant under some given transformation. Let C be the sum of its counterterms, and assume the action is invariant. Then the Green's functions are invariant under the transformation, so

$$\begin{aligned} 0 &= \delta\langle 0|TX(\mathcal{O}+C)|0\rangle \\ &= \langle 0|T\delta X(\mathcal{O}+C)|0\rangle + \langle 0|TX\,\delta(\mathcal{O}+C)|0\rangle \\ &= \langle 0|T\delta X(\mathcal{O}+C)|0\rangle + \langle 0|TX\,\delta C|0\rangle. \end{aligned} \tag{12.6.1}$$

Now C is defined so that Green's functions of $\mathcal{O}+C$ with elementary fields are finite. For an ordinary symmetry, the variation of an elementary field is again an elementary field, so the first term on the right of (12.6.1) is finite. Hence the other term, an arbitrary Green's function of δC, is finite. If we use minimal subtraction, the counterterms are powers of $1/\varepsilon$, so δC is finite only if it is zero. Thus counterterms to an invariant operator are invariant.

But when δ represents a BRS transformation the variation of an elementary field is composite. So finiteness of $\langle 0|T\delta X(\mathcal{O}+C)|0\rangle$ does not follow from finiteness of the elementary Green's functions of $O+C$. Hence the counterterms to a gauge-invariant operator need not be BRS invariant, let alone gauge invariant.

This argument also shows us how to handle the problem. Let X be a product $\phi_1(x_1)\ldots\phi_N(x_N)$ of local fields. Then the only divergences of $\langle 0|T\delta X(\mathcal{O}(y)+C(y))|0\rangle$ that lack a counterterm are when y coincides with the position of a BRS-varied operator $\delta\phi_i(x_i)$ in δX. Hence, provided y is not equal to any x_i, $\langle 0|T\delta X(\mathcal{O}(y)+C(y))|0\rangle$ is finite. From (12.6.1) we then see that $\langle 0|TX\delta C|0\rangle$ is then finite. But if we use minimal subtraction this means that it is zero. Hence an arbitrary Green's function $\langle 0|TX\delta C|0\rangle$ of δC with elementary fields is zero. Thus, as an operator, $\delta_{\mathrm{BRS}}C=0$: the counterterms to a gauge-invariant operator are BRS invariant.

One can try verifying this theorem by explicit calculations (e.g. Kluberg-Stern & Zuber (1975)). These apparently contradict the theorem. However, the non-invariant counterterms must correspond to operators that vanish by the equations of motion. As usual, the treatment of derivatives of fields in covariant perturbation theory implicitly generates commutator terms. After Fourier transformation into momentum space the non-invariant operators are not manifestly zero.

What we are actually interested in are gauge-invariant operators rather than merely BRS-invariant operators, for it is only the gauge-invariant operators that have physical significance independently of the method of gauge-fixing. So we must show how it is that the gauge-variant operators that mix with $\mathcal{O}$ do not enter into physical quantities.

Let $\mathcal{O}_1,\ldots,\mathcal{O}_n$ be the set of gauge-invariant operators that mix with $\mathcal{O}$. These are the operators which have the same transformations under global symmetries (e.g. Lorentz, isospin) as $\mathcal{O}$, and whose dimension is at most that of $\mathcal{O}$. Choose this set so that it is linearly independent (after use of the equations of motion). There are three other classes of operators that mix with $\mathcal{O}$:

Class A These operators, A_i, are the BRS variation of some operator: $A_i=\delta\hat{A}_i$

Class B These operators, B_i, vanish by the equations of motion.

Class C Any other operators that mix with $\mathcal{O}$, and that are not linear combinations of A_i's, B_i's, and $\mathcal{O}_i$'s.

The nilpotence of BRS transformations ensures that $\delta A_i=0$ up to terms vanishing by the equations of motion.

The most obvious classes of BRS-invariant operators are the gauge-invariant operators $\mathcal{O}_i$ and the BRS transformation of operators $\delta\hat{A}_i$. The first result to prove is that there are no others; i.e., there are no operators of class C. The second result is that the renormalization matrix expressing the renormalized operators $[\mathcal{O}]$, $[A]$, and $[B]$ in terms of unrenormalized operators is triangular:

$$\begin{pmatrix}[\mathcal{O}]\\ [\mathbf{A}]\\ [\mathbf{B}]\end{pmatrix} = \begin{pmatrix}Z_{\mathcal{O}\mathcal{O}} & Z_{\mathcal{O}A} & Z_{\mathcal{O}B}\\ 0 & Z_{AA} & Z_{AB}\\ 0 & 0 & Z_{BB}\end{pmatrix}\begin{pmatrix}\mathcal{O}\\ \mathbf{A}\\ \mathbf{B}\end{pmatrix}. \tag{12.6.2}$$

It is easy to prove this second result. The only operators that can be counterterms to a class B operator must themselves vanish by the equations of motion, i.e., they are of class B. A Green's function of a class A operator $\delta\hat{A}_i$ can be written as

$$\langle 0|T\delta\hat{A}_i(y)X|0\rangle = -\langle 0|T\hat{A}_i\delta X|0\rangle. \tag{12.6.3}$$

Thus if the counterterm to $\hat{A}_i$ is $C(\hat{A}_i)$ then

$$\langle 0|T[\hat{A}_i + C(\hat{A}_i)]\delta X|0\rangle$$

is finite if y equals none of the x_i's. Hence the counterterm to $\delta\hat{A}_i$ is $\delta C(\hat{A}_i)$, modulo terms vanishing by the equations of motion. That is, the counterterms to $A_i = \delta\hat{A}_i$ are class A and class B only.

The proof that there are no class C operators is somewhat complicated. It is essentially a mathematical exercise in homology theory. We refer the reader to Joglekar (1977a, b) and Joglekar & Lee (1976) for proofs.

The importance of these results is as follows: both class A and class B operators vanish in physical matrix elements. For class B this is because of the equations of motion. For class A we obtain the matrix element by the LSZ reduction formula from a Green's function $\langle 0|TA_iX|0\rangle$. But there is a Ward identity

$$\langle 0|T\delta\hat{A}_iX|0\rangle = -\langle 0|T\hat{A}_i\delta X|0\rangle. \tag{12.6.4}$$

Whenever X is gauge invariant we get zero. But we may also use elementary field operators in X. A term in δX with a varied field operator $\delta\phi(x)$ has no physical particle pole for this line. This can be seen by examining the possible Feynman graphs, and observing that $\delta\phi$ contains a ghost field.

If the renormalization matrix in (12.6.2) were not triangular then renormalized operators of classes A and B would be non-zero on-shell, even though the unrenormalized operators are zero. The triangularity ensures that for physical matrix elements we have

$$[\mathcal{O}] = Z_{\mathcal{O}\mathcal{O}}\mathcal{O}. \tag{12.6.5}$$

Thus we can disregard the non-invariant operators.

12.6.1 Caveat

In practical calculations with Feynman graphs one must beware of taking (12.6.5) too glibly. Consider the calculation of $[G^{a2}_{\mu\nu}]$ in QCD. The only gauge-invariant operator to mix with it is $\bar{\psi}\psi$, so we can write (12.6.5) as

$$\begin{pmatrix}[G^{a2}_{\mu\nu}]\\ [\bar{\psi}\psi]\end{pmatrix}=\begin{pmatrix}Z_{11} & Z_{12}\\ 0 & Z_{22}\end{pmatrix}\begin{pmatrix}G^{a2}_{(0)\mu\nu}\\ \bar{\psi}\psi\end{pmatrix}. \tag{12.6.6}$$

The Z_{21}-coefficient is zero because $\bar{\psi}\psi$ has lower dimension than G^2.

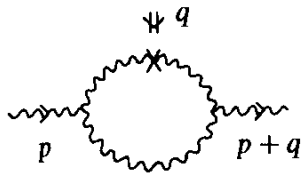


Fig. 12.6.1. One-loop graph for $\langle 0|TAA[G^2_{\mu\nu}]|0\rangle$.

We compute the one-loop term in Z_{11} from the graph of Fig. 12.6.1. There are various tensor structures for the counterterms. The $G^{a2}_{\mu\nu}$ counterterm is proportional to

$$g_{\mu\nu}p\cdot(p+q)-(p+q)_\mu p_\nu.$$

Tensors that vanish on-shell are, for example,

$$g_{\mu\nu}p^2-p_\mu p_\nu,\quad g_{\mu\nu}(p+q)^2-(p+q)_\mu(p+q)_\nu.$$

To simplify the calculations it is tempting to set $q=0$. But then these three structures are equal and it is not possible to separate the three coefficients. If one sets $p^2=0=(p+q)^2$ so that p and q are on-shell, and then multiplies the graph by polarization vectors $\varepsilon^\mu\varepsilon'^\nu$ satisfying $\varepsilon\cdot p=0=\varepsilon'\cdot(p+q)$, then one has taken a physical matrix element and only the $G^{a2}_{\mu\nu}$ counterterm survives.

The calculations of Kluberg-Stern & Zuber (1975) are at zero momentum, and they have to go to some effort to overcome the above problems.

12.7 Renormalization-group equation

Renormalization-group equations are derived in gauge theories just as in any other theory. One feature that is easy to overlook when making calculations is the variation of the gauge-fixing parameter under a renormalization group transformation. We consider the theory (12.1.1) and let G_{N_c,N_f,N_g} be the renormalized Green's function with N_g external gluons, N_f fermions, N_f antifermions, N_c ghosts, and N_c antighosts. The corresponding bare Green's function is

$$G^{(0)}_{N_c,N_f,N_g}=\tilde{Z}^{N_c}Z_2^{N_f}Z_3^{N_g/2}G_{N_c,N_f,N_g},$$

and it is RG invariant:

$$0 = \mu \frac{\mathrm{d}}{\mathrm{d}\mu} G^{(0)}$$
$$= \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial g} - \gamma_M M \frac{\partial}{\partial M} - \delta \xi \frac{\partial}{\partial \xi} \right) G^{(0)}. \tag{12.7.1}$$

Here β, γ_M and δ are obtained by requiring g_0, M_0, and $\xi_0 = Z_3 \xi$ to be RG invariant.

In the minimal subtraction scheme we have

$$\left.\begin{aligned} g_0 &= \mu^{2-d/2} g \left[1 + \sum_{n=1}^{\infty} a_n(g)/(4-d)^n \right], \\ Z_M &= 1 + \sum_{n=1}^{\infty} b_n(g)/(4-d)^n, \\ Z_i &= 1 + \sum_{n=1}^{\infty} c_{i,n}(g, \xi)/(4-d)^n. \end{aligned}\right\} \tag{12.7.2}$$

Here Z_i stands for Z_2, Z_3, or $\tilde{Z}$. The renormalization-group coefficients are

$$\left.\begin{aligned} \beta &= (d/2-2) g_0/(\partial g_0/\partial g) \\ &= (d/2-2) g + \tfrac{1}{2} g \partial a_1/\partial g, \\ \gamma_M &= \beta \partial \ln Z_M/\partial g = -\tfrac{1}{2} g \partial b_1/\partial g, \\ \gamma_i &= \beta \partial \ln Z_i/\partial g = -\tfrac{1}{2} g \partial c_{i,1}/\partial g, \\ \delta &= \gamma_3 = -\tfrac{1}{2} g \partial c_{3,1}/\partial g. \end{aligned}\right\} \tag{12.7.3}$$

Similar formulae hold in any other subtraction scheme.

Observe that, by the results of Section 12.4, g_0 and Z_M are independent of ξ so that β and γ_M are independent of ξ. But Z_2, Z_3 and $\tilde{Z}$ depend on ξ, so that γ_2, γ_3, $\tilde{\gamma}$ and δ also depend on ξ. As a special case, in an abelian theory, $g_0 = \mu^{2-d/2} g Z_3^{-1/2}$, so that Z_3, γ_3 and δ are independent of ξ.

The renormalization-group equations for renormalized Green's functions are then

$$0 = \left(\mu \frac{\mathrm{d}}{\mathrm{d}\mu} + N_c \tilde{\gamma} + N_f \gamma_2 + \tfrac{1}{2} N_g \gamma_3 \right) G$$
$$= \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial g} - \gamma_M M \frac{\partial}{\partial M} - \delta \xi \frac{\partial}{\partial \xi} + N_c \tilde{\gamma} + N_f \gamma_2 + \tfrac{1}{2} N_g \gamma_3 \right) G. \tag{12.7.4}$$

Since the anomalous dimensions are ξ-dependent, the solutions of the RG equation are a little complicated. The most convenient gauge to use is the Landau gauge $\xi = 0$, for then ξ does not vary when a renormalization group transformation is made. Tarasov Vladimirov & Zharkov (1980) have computed β to three-loop order in this theory.

12.8 Operator-product expansion

Consider the Green's function

$$M(y;x_1,\ldots,x_N)=\langle 0|Tj(y)j(0)X|0\rangle. \tag{12.8.1}$$

where j is a renormalized gauge-invariant operator, and X is a product of local fields at points $x_1,\ldots,x_N$. We have the Wilson expansion

$$M(y;x_1,\ldots,x_N)\sim\sum_i C_i(y)\langle 0|T[\mathcal{O}_i(0)]X|0\rangle. \tag{12.8.2}$$

The sum is over all renormalized operators of the appropriate dimension. This expansion is proved in a gauge theory the same way as in any other theory.

In the application to deep-inelastic scattering in Chapter 13, we will take X to be a product of two gauge-invariant operators and use the LSZ reduction formula to obtain a matrix element:

$$\begin{aligned} W&\equiv\langle P|Tj(y)j(0)|P\rangle \\ &=\sum_i C_i\langle P|[\mathcal{O}_i(0)]|P\rangle. \end{aligned} \tag{12.8.3}$$

We would like to show that only gauge-invariant operators need be included in (12.8.3). Now the proof of the operator-product expansion (in Chapter 10) treats the right-hand side of the expansion in the same way as renormalization counterterms. So the method we applied to the renormalization of gauge-invariant operators also applies to the Wilson expansion. The operators $\mathcal{O}_i$ are either gauge invariant, are the BRS variation of something, or vanish by the equations of motion. Then if we keep $y\neq 0$ in (12.8.3) (as is the case in applications) we only need gauge-invariant operators.

The renormalized operator j has, according to Section 12.6, the form

$$j=j_{\mathrm{GI}}+\delta_{\mathrm{R}}\hat{A}+B, \tag{12.8.4}$$

where j_{GI} is gauge invariant and B vanishes by the equation of motion. Hence if we take a matrix element of $j(y)j(0)$, like (12.8.3), (or if we take a gauge-invariant Green's function of $j(y)j(0)$), then we can drop the A and B terms, so that

$$\langle P|Tj(y)j(0)|P\rangle=\langle P|Tj_{\mathrm{GI}}(y)j_{\mathrm{GI}}(0)|P\rangle. \tag{12.8.5}$$

We now follow our proof in Section 12.6, starting with the Ward identity

$$\begin{aligned} 0&=\delta_{\mathrm{BRS}}\langle 0|Tj(y)j(0)X|0\rangle \\ &=\langle 0|Tj(y)j(0)\delta X|0\rangle. \end{aligned} \tag{12.8.6}$$

Here we assumed that the positions y_μ, $x_{i\mu}$, and 0 are all distinct, so that

$$\delta j=\delta B=0$$

by the equations of motion. But if $y \to 0$, then none of the $x_{i\mu}$'s are at the origin, so that:

$$\langle 0|Tj(y)j(0)\delta X|0\rangle \sim \sum_i C_i(y)\langle 0|T[\mathcal{O}_i(0)]\delta X|0\rangle$$
$$= -\sum_i C_i(y)\langle 0|T[\delta\mathcal{O}_i(0)]X|0\rangle. \qquad (12.8.7)$$

But the left-hand side is zero, by (12.8.6), while the right-hand side is its leading behavior as $y \to 0$ and is therefore also zero. If an arbitrary Green's function of an operator is zero, then the operator itself is zero, i.e.,

$$\sum_i C_i[\delta\mathcal{O}_i] = 0. \qquad (12.8.8)$$

Hence the operators needed in the expansion are those that are BRS invariant. As in Section 12.6, this means they are either gauge invariant or of classes A and B. In physical matrix elements like (12.8.3) we can therefore restrict the operators to be gauge invariant.

The practical use of this result is to compute the Wilson coefficients by taking the state $|P\rangle$ to be a state of one on-shell quark or gluon. The infrared divergences are regulated by going to $d > 4$. The Wilson coefficients and the gauge-invariant renormalization counterterms of the $\mathcal{O}_i$'s can be unambiguously obtained provided care is taken to separate the IR from the UV divergences (both of which appear as poles at $d = 4$).

12.9 Abelian theories: with and without photon mass

Consider QED with a possible mass term for the photon. The Lagrangian can be expressed in terms of unrenormalized or renormalized fields:

$$\mathscr{L}_{\text{inv}} = -\tfrac{1}{4}(F_{\mu\nu}^{(0)})^2 + \tfrac{1}{2}m_0^2(A_\mu^{(0)})^2 + \bar{\psi}_0(i\not{D} - M_0)\psi_0$$
$$= -\tfrac{1}{4}Z_3F_{\mu\nu}^2 + \tfrac{1}{2}m_0^2Z_3A_\mu^2 + Z_2\bar{\psi}(i\not{D} - M_0)\psi. \qquad (12.9.1)$$

Here we have a single vector field A_μ, and we let

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad (12.9.2)$$

$$D_\mu\psi = (\partial_\mu - ie_0A_\mu^{(0)})\psi$$
$$= (\partial_\mu - ie_0Z_3^{1/2}A_\mu)\psi. \qquad (12.9.3)$$

We will call A_μ the photon. Without a mass term for the photon, the Lagrangian (12.9.1) is invariant under the gauge transformation

$$\left.\begin{aligned} &A_\mu \to A_\mu + \partial_\mu\omega, \\ &\psi \to \psi\exp(ie_0Z_3^{1/2}\omega), \\ &\bar{\psi} \to \bar{\psi}\exp(-ie_0Z_3^{1/2}\omega). \end{aligned}\right\} \qquad (12.9.4)$$

With a mass term the Lagrangian is not invariant, but the theory can be consistently treated, as we will see.

In contrast, a non-abelian theory with a gluon mass term is not consistent. For example, there are unphysical states – the Faddeev–Popov ghosts – which violate the spin-statistics theorem, and the negative metric gluon states. In the massless theory these cancel in sums over intermediate states, but in a massive non-abelian theory they do not cancel. These problems do not occur in the abelian theory with a massive photon.

12.9.1 *BRS treatment of massive photon*

Even with a mass term for the photon, the Lagrangian is BRS invariant, if we use the standard gauge fixing. We write

$$\mathscr{L} = \mathscr{L}_{\text{inv}} - (1/2\xi)\partial\cdot A^2 + \partial_\mu\bar{c}\partial^\mu c - \xi m_0^2 Z_3 \bar{c}c. \tag{12.9.5}$$

and define BRS variations

$$\delta A_\mu = \partial_\mu c, \quad \delta\psi = \mathrm{i}e_0 Z_3^{1/2} c\psi, \quad \delta\bar{\psi} = -\mathrm{i}e_0 Z_3^{1/2}\bar{\psi}c,$$
$$\delta c = 0, \quad \delta\bar{c} = -\partial\cdot A/\xi. \tag{12.9.6}$$

Then $\mathscr{L}$ is BRS invariant, aside from an irrelevant total divergence. Since the Faddeev–Popov ghost is a free field, it can be omitted without affecting any physics (except gravity).

We may use the general methods of Section 12.2, with the result that the theory is renormalizable. Since the ghost is a free field, the renormalization factors X and $\tilde{Z}$ are both unity. Hence

$$e_0 = e_{\text{R}} Z_3^{-1/2} \tag{12.9.7}$$

and the renormalization of e is gauge independent. (If we use minimal subtraction we write $e_{\text{R}} = \mu^{2-d/2}e$.) We now find that $e_0 A_\mu^{(0)} = e_{\text{R}} A_\mu$, and that the covariant derivative and gauge transformations simplify:

$$D_\mu\psi = (\partial_\mu - \mathrm{i}e_{\text{R}} A_\mu)\psi, \tag{12.9.8}$$

$$\psi \to \mathrm{e}^{\mathrm{i}e_{\text{R}}\omega}\psi. \tag{12.9.9}$$

Using this method we also find that $m_0^2 = Z_3^{-1} m^2$, so that the ghost field has finite mass. Then the Lagrangian is

$$\mathscr{L} = -\tfrac{1}{4}Z_3 F_{\mu\nu}^2 + \tfrac{1}{2}m^2 A_\mu^2 - (1/2\xi)\partial\cdot A^2$$
$$+ Z_2\bar{\psi}(\mathrm{i}\partial\!\!\!/ + eA\!\!\!/ - M_0)\psi, \tag{12.9.10}$$

with the ghost ignored, since it is a free field. Note that there is no counterterm for the photon mass.

Similar methods apply in theories where the gauge group is a product of a

non-abelian group and one or more $U(1)$ factors. The Weinberg–Salam theory of weak and electromagnetic interactions is a simple case – with gauge group $SU(2)\otimes(1)$.

12.9.2 *Elementary treatment of abelian theory with photon mass*

The BRS methods are much more sophisticated than necessary for the abelian theory, so we will now treat the theory by elementary methods. First we write the free photon propagator for the perturbation theory of the Lagrangian (12.9.10):

$$D_{\mu\nu} = \frac{\mathrm{i}}{k^2 - m^2 + \mathrm{i}\varepsilon}(-g_{\mu\nu} + k_\mu k_\nu/m^2) - \frac{\mathrm{i}}{m^2}\frac{k_\mu k_\nu}{k^2 - \xi m^2 + \mathrm{i}\varepsilon}$$
$$= \frac{\mathrm{i}}{k^2 - m^2 + \mathrm{i}\varepsilon}\left[-g_{\mu\nu} + \frac{(1-\xi)k_\mu k_\nu}{k^2 - \xi m^2 + \mathrm{i}\varepsilon}\right]. \tag{12.9.11}$$

If $m = 0$, then we cannot remove the gauge-fixing term, for then the propagator does not exist. (This is the same as taking the limit $\xi \to \infty$.) But if m is non-zero, then the propagator exists when the gauge-fixing term is removed. However it behaves like $ik_\mu k_\nu/k^2m^2$, rather than $1/k^2$. So the theory with $m = 0$ and $\xi = \infty$ has worse divergences than usual and is not manifestly renormalizable. Even so, physical quantities are independent of ξ and the theory is renormalizable if ξ is finite, as we will see. Hence enough cancellations are present as $\xi \to \infty$ that the theory remains renormalizable if we only compute physical quantities.

Our treatment makes extensive use of the equations of motion:

$$\mathscr{L}_{A_\mu} \equiv \frac{\Delta S}{\Delta A_\mu}$$
$$\equiv Z_3 m_0^2 A^\mu + Z_3(\Box A^\mu - \partial^\mu \partial\cdot A) + \partial^\mu \partial\cdot A/\xi + e_0 Z_3^{1/2}\bar{\psi}\gamma^\mu\psi Z_2$$
$$= 0,$$
$$\mathscr{L}_{\bar{\psi}} \equiv \frac{\Delta S}{\Delta \bar{\psi}} \equiv Z_2(\mathrm{i}\not{\partial} - M_0)\psi = 0,$$
$$\mathscr{L}_{\psi} \equiv \frac{\Delta S}{\Delta \psi} \equiv Z_2\bar{\psi}(-\mathrm{i}\overleftarrow{\not{\partial}} - M_0) = 0. \tag{12.9.12}$$

We have not assumed the relation (12.9.7). Taking the divergence of the gauge-field equation of motion gives

$$0 = (\Box/\xi + m^2)\partial\cdot A. \tag{12.9.13}$$

Here, we have used the invariance of the theory under the global transformations, which are (12.9.4) with constant ω. The Noether current

for this invariance is the electromagnetic current:

$$j^\mu = -e_0 Z_3^{1/2} Z_2 \bar{\psi}\gamma^\mu\psi. \tag{12.9.14}$$

This is conserved, by the electron's equation of motion.

The result (12.9.13) that $\partial\cdot A$ is a free field is important for three reasons:

(1) It would otherwise be difficult to interpret the longitudinal part of A_μ, which has unphysical properties.
(2) The ξ-dependent part of the propagator is confined to the $k_\mu k_\nu$ term, which only contributes to $\partial\cdot A$. Then the ξ-dependence decouples from physics if $\partial\cdot A$ is free.
(3) Similarly the bad ultra-violet behavior when $\xi\to\infty$ is decoupled from physics.

To make these statements precise we will examine the Green's functions. We will prove directly:

(1) Ward identities,
(2) $m_0^2 = Z_3^{-1}m^2$, so that no mass counterterm is needed for the photon,
(3) no counterterms are needed for the $\partial\cdot A^2$ term,
(4) $e_0 = e_R Z_3^{-1/2}$, with e_R finite,
(5) Z_2 but not Z_3 depends on ξ,
(6) the S-matrix is independent of ξ.

We will also compute the exact ξ-dependence of Z_2 when minimal subtraction is used, and we will compute the exact ξ-dependence of the residue of the electron's propagator pole.

12.9.3 *Ward identities*

Green's functions of $\mathscr{L}_A$, $\mathscr{L}_\psi$ and $\mathscr{L}_{\bar\psi}$ are non-zero, since the derivatives in these operators are implicitly taken outside the time-ordering. To obtain the Ward identity corresponding to (12.9.13), we use identities like

$$\langle 0|T\mathscr{L}_{A_\mu}(x)X|0\rangle = i\langle 0|T\frac{\Delta X}{\Delta A_\mu(x)}|0\rangle. \tag{12.9.15}$$

Now

$$(\Box/\xi + m^2)\partial\cdot A = \partial_\mu\mathscr{L}_{A_\mu} - ie\bar\psi\mathscr{L}_{\bar\psi} + ie\mathscr{L}_\psi\psi. \tag{12.9.16}$$

We choose $e_0 = e_R Z_3^{-1/2}, m_0^2 = Z_3^{-1}m^2$. This gives a Lagrangian of the form (12.9.10). There are no counterterms for A^2 and $\partial\cdot A^2$, and the counterterm for $\bar\psi A\psi$ is proportional to the counterterm $Z_2 - 1$ for $\bar\psi\partial\psi$. We will prove later that this is correct. But for the moment we will choose to have our Lagrangian in this form. Our gambit is that if extra counterterms should be

needed, then they will not be available and the Green's functions will diverge.

From (12.9.16) it follows that

$$\begin{aligned}(\Box_x/\xi + m^2)\langle 0|T\partial\cdot A(x)X|0\rangle &= i\frac{\partial}{\partial x^\mu}\langle 0|T\Delta X/\Delta A_\mu(x)|0\rangle \\ &\quad + e_R\langle 0|T\bar\psi(x)\Delta X/\Delta\bar\psi(x)|0\rangle - e_R\langle 0|T\psi(x)\Delta X/\Delta\psi(x)|0\rangle \\ &= i\frac{\delta_{\text{gauge}}}{\delta\omega(x)}\langle 0|TX|0\rangle, \end{aligned} \tag{12.9.17}$$

where the gauge variation is computed according to (12.9.4).

A convenient way to write these results is as follows:

(1) Let ϕ be a free scalar boson field of mass $\xi^{1/2}m$.

(2) Let $\hat\delta$(field) = gauge variation of 'field' with parameter ϕ, i.e.,

$$\begin{aligned}\hat\delta A_\mu &= \partial_\mu\phi, \\ \hat\delta\psi &= ie_R\phi\psi, \\ \hat\delta\bar\psi &= -ie_R\phi\bar\psi.\end{aligned}$$

(3) Then (12.9.17) is

$$(1/\xi)\langle 0|T\partial\cdot AX|0\rangle = \langle 0|T\phi(x)\hat\delta X|0\rangle. \tag{12.9.18}$$

The field ϕ is similar to the Faddeev–Popov ghost except for being a boson (this will be important later). In a non-abelian theory the nilpotence of the BRS transformation is crucial to proving Ward identities and is proved using the anticommutation of c with itself. A field introduced in the same way as ϕ has to be the ghost field and must be a fermion. In (12.9.18) we can choose ϕ to be a boson.

12.9.4 *Counterterms proportional to A^2 and $\partial\cdot A^2$*

Let us apply (12.9.17) to the case $X = A_\nu(y)$:

$$(\Box_x/\xi + m^2)\frac{\partial}{\partial x^\mu}\langle 0|TA^\mu(x)A_\nu(y)|0\rangle = i\frac{\partial}{\partial x^\nu}\delta^{(d)}(x-y). \tag{12.9.19}$$

This equation implies that no counterterms proportional to A^2 or to $\partial\cdot A^2$ are needed. For suppose otherwise. Then consider the lowest order in which there is a divergence in the photon propagator. The divergence comes from the insertion of a divergent self-energy graph in the free photon propagator. The divergence has the form of Fig. 12.9.1, where the cross denotes the counterterm to this divergence. Power-counting indicates that

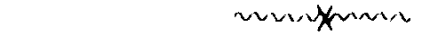


Fig. 12.9.1. Counterterm to photon's self-energy.

the only divergences are proportional to $k_\mu k_\nu$ and $g_{\mu\nu}$. A divergence $g_{\mu\nu}k^2 - k_\mu k_\nu$ is cancelled by the wave-function renormalization Z_3 for the photon. Insertion of the extra counterterms in the propagator gives a divergent contribution to the left-hand side of (12.9.19). But the left-hand side is finite. So there are in fact no divergences. Hence $m_0^2 = Z_3^{-1}m^2$. It also follows that the photon self-energy is transverse.

Applying (12.9.17) to $X = A_\mu(y)A_\nu(z)$ and to $A_\kappa(w)A_\lambda(y)A_\nu(z)$ shows that no counterterms cubic or quartic in A are needed.

12.9.5 *Relation between e_0 and Z_3*

Apply (12.9.17) to $X = \psi(y)\bar{\psi}(z)$:

$$(\Box_x/\xi + m^2)\frac{\partial}{\partial x^\mu}\langle 0|TA^\mu(x)\psi\bar{\psi}|0\rangle$$
$$= e\langle 0|T\psi(y)\bar{\psi}(z)|0\rangle[\delta(x-y) - \delta(x-z)]. \tag{12.9.20}$$

We assumed $e_0 = Z_3^{-1/2}e_R$. If this is not the correct counterterm, then let the ψ–$\bar{\psi}$–A 1PI vertex first diverge at N-loops. Then the left-hand side of (12.9.20) diverges at order e_R^{2N+1}, while there is a counterterm $Z_2 - 1$ available to make the right-hand side finite. Hence the left-hand side does not diverge.

12.9.6 *Gauge dependence*

The gauge variation of a Green's function $\langle 0|TX|0\rangle$ is

$$\frac{\partial}{\partial\xi}\langle 0|TX|0\rangle = \frac{\mathrm{i}}{2\xi^2}\int \mathrm{d}^4x\langle 0|T[\partial\cdot A^2(x) - \langle 0|\partial\cdot A^2|0\rangle]X|0\rangle$$
$$+ \mathrm{i}\frac{\partial \ln Z_2}{\partial\xi}\int \mathrm{d}^4x\langle 0|T\bar{\psi}\mathscr{L}_{\bar{\psi}}X|0\rangle$$
$$= \frac{\mathrm{i}}{2\xi}\int \mathrm{d}^4x\langle 0|T\partial\cdot A(x)\phi(x)\hat{\delta}X|0\rangle$$
$$- N_\psi\langle 0|TX|0\rangle\partial \ln Z_2/\partial\xi$$
$$= \frac{\mathrm{i}}{2}\int \mathrm{d}^4x\langle 0|T[\phi^2(x) - \langle 0|\phi^2|0\rangle]\hat{\delta}^2X|0\rangle$$
$$- N_\psi\langle 0|TX|0\rangle\partial \ln Z_2/\partial\xi. \tag{12.9.21}$$

Here N_ψ is the number of ψ fields in X (which equals the number of $\bar{\psi}$ fields).

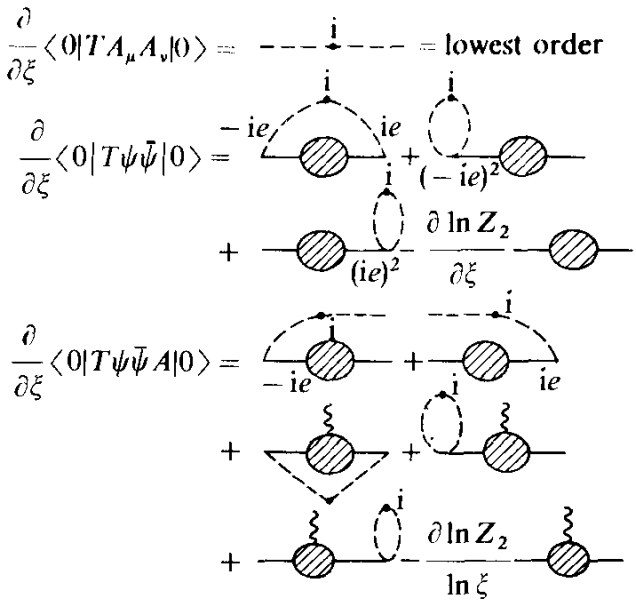


Fig. 12.9.2. Examples of (12.9.21).

The disappearance of the factor $\frac{1}{2}$ in the last equality is due to the identical nature of the two ϕ fields. Also, $\phi^2 - \langle 0|\phi^2|0\rangle$ appears, rather than ϕ^2, because the pure vacuum graphs never occur on the left-hand side. (A better derivation uses two different fields ϕ_1 and ϕ_2 for each application of the Ward identities.) Notice that we cannot derive this equation if ϕ is a fermion, for then $\phi^2 = 0$. So, in a non-abelian theory, the derivation must be replaced by the more complicated one which leads to the weaker (12.4.4). The graphical structure of (12.9.21) for a few simple cases is shown in Fig. 12.9.2.

From (12.9.21) it is immediate that all Green's functions of gauge-invariant fields are ξ-independent.

The only graphs with UV divergences are those with the simple loop (Fig. 12.9.3) attached to a $\bar{\psi}$ or a ψ vertex together with those with the $\partial \ln Z_2/\partial\xi$ factor. We therefore find that

$$-\frac{\partial \ln Z_2}{\partial \xi} + ie_R^2 \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 - \xi m^2)^2} \tag{12.9.22}$$

is UV finite. Hence

$$\frac{\partial \ln Z_2}{\partial \xi} = -\frac{e^2}{16\pi^2}\Gamma(2 - d/2)\left(\frac{\xi m^2}{4\pi\mu^2}\right)^{d/2-2} + \text{finite}.$$



Fig. 12.9.3. All UV divergent graphs contributing to (12.9.21) contain this loop.

If we use minimal subtraction then

$$Z_2 = Z_2(\xi = 0)\exp\left[\frac{e^2\xi}{8\pi^2(d-4)}\right]. \tag{12.9.23}$$

To obtain the ξ-dependence of the S-matrix, we recall the LSZ formula. It tells us to consider the corresponding Green's function, and pick out the poles in its external momenta. Since we use transverse polarization ($\varepsilon\cdot k = 0$) for photons this picks out graphs with the loop Fig. 12.9.3 and the $\partial \ln Z_2/\partial\xi$ terms. Then

$$S = \prod_{\substack{\text{ext}\\ \text{Fermions}}}\left(\frac{\not{p} - M_{\text{ph}}}{iz_2^{1/2}}\right)\prod_{\substack{\text{ext}\\ \text{photons}}}\left(\frac{p^2 - m_{\text{ph}}^2}{iz_3^{1/2}}\right)\langle 0|T\tilde{X}|0\rangle. \tag{12.9.24}$$

From Fig. 12.9.2 for $\langle 0|T\psi\bar{\psi}|0\rangle$ and for $\langle 0|TA_\mu A_\nu|0\rangle$, we see that the physical masses and the residue of the photon pole are ξ-independent, while the gauge dependence of z_2 exactly cancels the gauge dependence of the particle pole coefficient of $\langle 0|T\tilde{X}|0\rangle$.

We can compute explicitly the ξ-dependence of the residue, z_2, of the pole of the fermion propagator, if we use minimal subtraction:

$$\frac{\partial \ln z_2}{\partial\xi} = \lim_{d\to 4}(12.9.22)$$

$$= \frac{e^2}{16\pi^2}\left[\gamma + \ln\left(\frac{\xi m^2}{4\pi\mu^2}\right)\right],$$

i.e.,

$$z_2 = z_2(\xi = 0)\exp\left\{\frac{e^2\xi}{16\pi^2}\left[\gamma - 1 + \ln\left(\frac{\xi m^2}{4\pi\mu^2}\right)\right]\right\}. \tag{12.9.25}$$

Since $\xi_0 = \xi Z_3$, $e_0^2 = Z_3^{-1}e^2$, and $m_0^2 = Z_3^{-1}m^2$, the combinations $e^2\xi$ and ξm^2 in (12.9.25) are RG invariant.

12.9.7 *Renormalization-group equation*

Using our knowledge of the renormalization of e_0, ξ_0, and m_0^2 we find the RG equation for a Green's function of $N_f\psi$'s, $N_f\bar{\psi}$'s, and N_A A's to be

$$0 = (\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial e} - \gamma_M M\frac{\partial}{\partial M} + \gamma_3 m^2\frac{\partial}{\partial m^2} - \gamma_3\xi\frac{\partial}{\partial\xi} + N_f\gamma_2 + \tfrac{1}{2}N_A\gamma_3)G, \tag{12.9.26}$$

with

$$\beta = \tfrac{1}{2}e\gamma_3 + (d/2 - 2)e.$$

If we use minimal subtraction then the only ξ-dependent coefficient is γ_2, and

from (12.9.23) we find that

$$\gamma_2 = \gamma_2(\xi = 0) + \xi e^2/8\pi^2. \tag{12.9.27}$$

The results (12.9.23), (12.9.25), and (12.9.27) were derived by Lautrup (1976) and by Collins (1975a).

12.10 Unitary gauge for massive photon

The unitary gauge is the limit $\xi \to \infty$. Since the S-matrix and gauge-invariant operators are ξ-independent, this limit exists for them. But for gauge-variant operators there are severe ultra-violet divergences. Thus the limits $\xi \to \infty$, $d \to 4$ do not commute.

We may take $\xi \to \infty$ first, in the regulated theory. The resulting UV divergences at $d = 4$ may be cancelled by extra counterterms beyond those that we have already considered. Since the S-matrix is gauge-independent, all these counterterms must vanish by the equations of motion.

13

Anomalies

A characteristic feature of relativistic quantum field theories is that symmetries of the classical theory are not always present after quantization. We do not mean here the spontaneous breaking that is characterized by a non-invariant vacuum and by the presence of the Goldstone bosons. Rather we mean a situation where there is no conserved current for the symmetry despite the absence of any terms in the action that appear to break the symmetry. Such breaking of a symmetry is called anomalous.

If the classical action is invariant, then a naive application of Noether's theorem gives us a conserved current. That is, there is no anomalous symmetry breaking. What prevents the argument from being correct is the presence of UV divergences. The current is a composite operator, i.e., a product of elementary fields at the same point, and to define it, some kind of regularization and renormalization is needed. The renormalization may invalidate the equations used to prove Noether's theorem.

For simplicity, we will consider only global symmetries, as opposed to local, or gauge, symmetries. The simplest cases of global symmetries were considered in Chapter 9. These could be treated by using an ultra-violet regulator that preserved the symmetry. The proof of Noether's theorem can then be made in the cut-off theory. We showed that only symmetric counterterms are needed. Consequently the symmetry remains good after the cut-off is removed.

However, not all symmetries can be preserved after regularization. The case which we will treat in this chapter is that of chiral symmetries. These are transformations that act independently on the left- and right-handed components of Dirac fields. These are particularly interesting because sometimes the anomalous breaking of chiral symmetries cancels. Indeed there is a theorem, first proved by Adler & Bardeen (1969), that if anomalous breaking of a chiral symmetry is zero to one-loop order then it is zero to all orders.

Our treatment will use dimensional regularization. Chiral symmetry is valid in the physical space-time dimension $d = 4$, but not when $d \neq 4$. The anomaly in, say, an equation of current conservation will be an operator with

in effect a coefficient proportional to $d-4$, where d is the space-time dimension. This would vanish at $d=4$, were it not for the existence of ultra-violet divergences which allow the anomaly operator to have a pole at $d=4$ so that a non-zero anomaly results at $d=4$. We will see explicitly how this works.

There are two key issues. The first is to derive simple forms for the anomaly. Of these the most dramatic is the Adler–Bardeen theorem that tells us that in some cases there is complete cancellation of the anomaly. The second issue is to derive the results in a form that is applicable to the physical theory, i.e., after renormalization and removal of the cut-off. Even though a particular derivation depends on the choice of a particular regularization scheme, the final results must be independent of this choice.

Note that to prove existence of an anomaly, it is not sufficient to say that the symmetry in question is broken in the regulated theory. The breaking may go away after removal of the cut-off. For example, if one uses a lattice cut-off, then Poincaré invariance is lost. However, one must prove that Poincaré invariance is restored in the renormalized continuum limit, if the theory is to agree with real world phenomena.

Aside from the case of chiral transformations, there are a number of other important situations where there are anomalies. One of the simplest is that of dilatations. These are scale transformations on space-time: $x^\mu \to \lambda x^\mu$. A classical Lagrangian is scale-invariant if it contains no dimensional parameters, like a mass scale. But to cut-off ultra-violet divergences we necessarily introduce a mass scale. The symmetry is necessarily broken in the regulated theory and the question arises of whether the symmetry remains broken after the theory is renormalized and the cut-off is removed. This answer is, in general, yes, if the theory has interactions. The reason is that there is, in fact, a mass scale hidden in the renormalized theory, as we saw when we discussed dimensional transmutation and the renormalization group, in Chapter 7. Detailed treatments can be found in the literature (see Collins (1976), Brown (1980) and references therein). The simple form for the Ward identity is known as the Callan–Symanzik equation (Callan (1970) and Symanzik (1970b)). The information contained in the Callan–Symanzik equation is in fact also contained in the renormalization-group equation that we studied in Chapter 7.

Other situations which we will not treat include the following: chiral gauge theories (see Costa *et al.* (1977), Bandelloni *et al.* (1980), Piguet & Rouet (1981)), conformal transformations (Sarkar (1974)) and supersymmetries (Piguet & Rouet (1981), Clark, Piguet & Sibold (1979, 1980), and Piguet & Sibold (1982a, b, c)).

13.1 Chiral transformations

We will consider QCD with two flavors of quark: up and down. (We could have more flavors, but no essentially new ideas would be needed.) The Lagrangian is (2.11.7). If the quarks were all massless then the classical theory is invariant under the following transformations of the quark fields:

$$\psi \to \exp\{i[\tfrac{1}{2}(1-\gamma_5)(\omega_L^0 + \omega_L^a t^a) + \tfrac{1}{2}(1+\gamma_5)(\omega_R^0 + \omega_R^a t^a)]\}\psi. \quad (13.1.1)$$

Here the matrices t^a are the generators of the isospin group acting on the flavor indices. The transformations (13.1.1) form a group that we will call $U(1)_L \otimes U(1)_R \otimes SU(2)_L \otimes SU(2)_R$. The symmetry of the QCD Lagrangian under these transformations is broken by mass terms for the quarks. Since the masses of the u and d quarks are small, the chiral symmetries are only weakly broken.

These symmetries and their breaking were understood well before the advent of QCD – see, for example, Treiman, Gross & Jackiw (1972). Treatments of chiral symmetries in the light of QCD can be found in Marciano & Pagels (1978) and in Llewellyn-Smith (1980). Thus it is unnecessary to go into details here. What we will emphasize is how the potential for anomalies arises.

Since these transformations involve γ_5, they are in some sense coupled to the spin structure of the theory. Since spin is related to the symmetries of space-time, we can expect trouble when the theory is regulated, for imposition of an ultra-violet cut-off must alter the space-time structure.

Notice that there is a $U(1) \otimes SU(2)$ subgroup not involving γ_5; these transformations have $\omega_L = \omega_R$. For them the treatment of Chapter 9 is correct. The corresponding Noether currents are

$$\begin{aligned} j^\mu &= Z\bar{\psi}\gamma^\mu\psi, \\ j_a^\mu &= Z\bar{\psi}\gamma^\mu t^a\psi. \end{aligned} \quad (13.1.2)$$

The conserved charge derived from j^μ is the conserved quark number, while the transformations generated by j_a^μ are just ordinary isospin transformations.

For the other generators of chiral transformations, let us define axial currents

$$\begin{aligned} j_5^\mu &= Z\bar{\psi}\tfrac{1}{2}[\gamma^\mu, \gamma_5]\psi, \\ j_{a5}^\mu &= Z\bar{\psi}\tfrac{1}{2}[\gamma^\mu, \gamma_5]t^a\psi. \end{aligned} \quad (13.1.3)$$

In four dimensions $\gamma^\mu\gamma_5 = -\gamma_5\gamma^\mu$, so that we could have written $\gamma^\mu\gamma_5$ in place of the commutator in (13.1.3). However, because we will use dimensional regularization, we must use the form (13.1.3) in order to ensure that the currents are hermitian.

To define the theory we must regulate its ultra-violet divergences, and for this we will use dimensional regularization. We will see that the regulated theory is not invariant under the transformations generated by the axial currents j_5^μ and j_{5a}^μ. This allows the possibility of an anomaly. When the cut-off is removed we will find that we can arrange for the non-singlet currents j_{a5}^μ to be without anomaly. Although we will not demonstrate it, it is true that the singlet current necessarily has an anomalous divergence.

Our remarks above were addressed to the case that all the quarks are massless. But in the real world, there are quark mass terms in the Lagrangian, so we now generalize our discussion. The vector singlet current j^μ is the current for quark number – i.e., 1/3 of baryon number – so this remains an exact symmetry. The $SU(2)$ symmetry given by the vector currents is broken by quark mass differences:

$$\partial_\mu j_a^\mu = -\mathrm{i}\bar{\psi}[t^a, M]\psi, \tag{13.1.4}$$

where M is the quark mass matrix. Since the masses of the u and d quarks are small we have an approximate isospin symmetry of strong interactions. By the theory given in Chapter 9, the currents j_a^μ and j^μ, as defined by (13.1.2), are finite, since the breaking is from mass terms (Symanzik (1970a)).

The axial symmetries are broken by the anomaly as well as by the quark mass terms:

$$\begin{aligned} \partial_\mu j_5^\mu &= -2\mathrm{i}\bar{\psi}\gamma_5 M\psi + \text{anomaly}, \\ \partial_\mu j_{a5}^\mu &= -\mathrm{i}\bar{\psi}\gamma_5\{t^a, M\}\psi. \end{aligned} \tag{13.1.5}$$

The u- and d-quark masses are light enough to give us an approximate $SU(2)\otimes SU(2)$ symmetry. The axial part appears to be spontaneously broken as well as explicitly broken. The abnormally light mass of the pion is usually taken to mean that it would be a Goldstone boson if $m_u = m_d = 0$. See Marciano & Pagels (1978) and Gasser & Leutwyler (1982) for more details.

13.2 Definition of γ_5

In order to formulate consistently the dimensional regularization of theories with fermions we had to define an infinite set of matrices γ^μ, $(\mu = 0, 1, 2, \ldots)$. As we saw in Chapter 4, they satisfied the algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}1, \quad \mathrm{tr}\, 1 = 4. \tag{13.2.1}$$

Our task is now to find a generalization of the matrix which at $d = 4$ is called γ_5. Its four dimensional definition is

$$\gamma_5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3; \tag{13.2.2}$$

and it satisfies the anticommutation relation

$$\{\gamma^\mu, \gamma_5\} = 0 \quad (d = 4). \tag{13.2.3}$$

It would be natural to assume that this relation (13.2.3) can be maintained for arbitrary values of d. Unfortunately an inconsistency arises, as we will now show, when we wish to compute the trace of γ_5 with a product of the ordinary γ^μ's. The ultimate result will be that stated in Section 4.6, where we used the definition (13.2.2) of γ_5 for all values of d. Then γ_5 has mixed commutation and anticommutation relations, (4.6.3).

We now demonstrate the inconsistency in trace calculations, starting with $\operatorname{tr}\gamma_5$:

$$\begin{aligned} d\operatorname{tr}(\gamma_5) &= \operatorname{tr}(\gamma_5\gamma^\mu\gamma_\mu) \\ &= \operatorname{tr}(\gamma_\mu\gamma^5\gamma_\mu) \\ &= -\operatorname{tr}(\gamma_5\gamma_\mu\gamma^\mu) \\ &= -d\operatorname{tr}(\gamma_5). \end{aligned} \tag{13.2.4}$$

In the first and last lines we used

$$\gamma_\mu\gamma^\mu = \tfrac{1}{2}\{\gamma_\mu, \gamma^\mu\} = g^\mu_\mu 1 = d.$$

In the second line we used cyclicity of a trace, and in the third line we assumed (13.2.3). From (13.2.4) we see that $\operatorname{tr}\gamma_5 = 0$ except at $d = 0$. Now when we apply dimensional regularization we wish to obtain a result that is a meromorphic function of d. Hence we must have $\operatorname{tr}\gamma_5 = 0$ for all d.

Similarly

$$\begin{aligned} d\operatorname{tr}\gamma_5\gamma_\mu\gamma_\nu &= \operatorname{tr}\gamma_5\gamma_\mu\gamma_\nu\gamma_\lambda\gamma^\lambda \\ &= \operatorname{tr}\gamma^\lambda\gamma_5\gamma_\mu\gamma_\nu\gamma_\lambda \\ &= -\operatorname{tr}\gamma_5\gamma^\lambda\gamma_\mu\gamma_\nu\gamma_\lambda \\ &= -2g^\lambda_\mu\operatorname{tr}\gamma_5\gamma_\nu\gamma_\lambda + 2g^\lambda_\nu\operatorname{tr}\gamma_5\gamma_\mu\gamma_\lambda - d\operatorname{tr}\gamma_5\gamma_\mu\gamma_\nu \\ &= -2\operatorname{tr}\gamma_5\{\gamma_\mu, \gamma_\nu\} + (4-d)\operatorname{tr}\gamma_5\gamma_\mu\gamma_\nu \\ &= (4-d)\operatorname{tr}\gamma_5\gamma_\mu\gamma_\nu. \end{aligned} \tag{13.2.5}$$

Here we used the result $\operatorname{tr}\gamma_5 = 0$. Hence we find that $(d-2)\operatorname{tr}\gamma_5\gamma_\mu\gamma_\nu$ is zero.

Now, the same technique can be used to prove that

$$(4-d)\operatorname{tr}(\gamma_5\gamma_\kappa\gamma_\lambda\gamma_\mu\gamma_\nu) = 0. \tag{13.2.6}$$

At $d = 4$ this equation permits the usual non-zero trace of γ_5 with four other Dirac matrices. However, if the trace is to be meromorphic in d, equation (13.2.6) shows that it must be zero at $d = 4$, and we can therefore not obtain normal physics at $d = 4$. This is the inconsistency referred to at the start of this section.

We are therefore forced to drop one of the hypotheses that led to (13.2.6). Candidate hypotheses for removal include:

(1) the anticommutation relation (13.2.3) ('t Hooft & Veltman (1972a)),
(2) the use of dimensional regularization for fermion loops (Bardeen (1972), and Chanowitz, Furman & Hinchliffe (1979)).

These last authors have shown how to calculate with a totally anticommuting γ_5. Bardeen chooses to use a regulator other than dimensional regularization for all fermion loops. On the other hand, Chanowitz *et al.* regulate fermion loops with an even number of γ_5's dimensionally. Their procedure is useful for low-order graphs, since the Ward identities are preserved for graphs without fermion loops. However, we then lose the use of dimensional regularization as a complete regulator; the theorems that we derived in Section 6.6 and in Chapter 12 no longer apply. The details of higher-order calculations by this method have not been spelled out.

Therefore let us follow 't Hooft & Veltman (1972a) and Breitenlohner & Maison (1977a) and change the anticommutation relation. In fact, we may use the definition (13.2.2) for all values of d, just as stated in Section 4.6. Our definition is, of course, not completely Lorentz covariant, since the first four dimensions are picked out as special. But this is not an overwhelming objection, for our actual physics is confined to these dimensions. An important advantage of the definition is that it gives a concrete construction of γ_5. We are therefore guaranteed consistency.

One notational inconvenience arises. We have a set of matrices γ^μ for $\mu = 0, 1, 2, \ldots$. Usually we only refer explicitly to the first four; the rest are referred to collectively. But there is γ^μ with $\mu = 5$. It is not the same as γ_5 defined by (13.2.2). However, we bow to standard usage and use γ_5 to denote the matrix defined in (13.2.2). Confusion should be rare.

The γ_5 so defined has mixed commutation and anticommutation relations that follow from (13.2.2) and the properties of γ^μ. These were stated in (4.6.3). The derivation of (13.2.4)–(13.2.6) now fails, because γ_5 does not anticommute with all of the γ^μ's. We can derive the correct relations easily. Since the trace of an odd number of γ^μ's is zero, we have

$$\operatorname{tr}(\gamma^\mu \gamma_5) = 0, \quad \operatorname{tr}(\gamma_5 \gamma^\kappa \gamma^\lambda \gamma^\mu) = 0, \text{ etc.} \tag{13.2.7}$$

We may read off the trace of γ_5 with an even number of γ^μ's from (4.5.13) and its relatives with more than four γ-matrices. Thus we have

$$\left.\begin{aligned} \operatorname{tr}(\gamma_5) &= 0, \\ \operatorname{tr}(\gamma_5 \gamma^\mu \gamma^\nu) &= 0. \end{aligned}\right\} \tag{13.2.8}$$

Finally

$$\mathrm{tr}\,(\gamma_5\gamma_\kappa\gamma_\lambda\gamma_\mu\gamma_\nu) = 4\mathrm{i}\varepsilon_{\kappa\lambda\mu\nu}, \tag{13.2.9}$$

where $\varepsilon_{\kappa\lambda\mu\nu}$ is given by (4.6.2).

The relations (4.6.3) mean that the chiral transformations (13.1.1) no longer generate a symmetry if d is not equal to 4. To discuss the resulting problems, the following notation is useful (Breitenlohner & Maison (1977a)):

$$\left.\begin{aligned} \hat{g}_{\mu\nu} &= \begin{cases} g_{\mu\nu}, & \text{if } \mu \text{ and } \nu \text{ are 4 or larger,} \\ 0, & \text{otherwise;} \end{cases} \\ \hat{V}^\mu &= \hat{g}^{\mu\nu}V_\nu. \end{aligned}\right\} \tag{13.2.10}$$

Here V^μ is any vector. Then $\hat{g}_{\mu\nu}$ is a projector onto the unphysical dimensions. Thus, for example,

$$\left.\begin{aligned} \hat{g}^\mu_\mu &= g^{\mu\nu}\hat{g}_{\mu\nu} = \hat{g}^{\mu\nu}\hat{g}_{\mu\nu} = d-4, \\ \{\hat{V}, \gamma_5\} &= \sum_{\mu>3} V_\mu\{\gamma^\mu, \gamma_5\} = 2\hat{V}\gamma_5 = 2\gamma_5\hat{V}. \end{aligned}\right\} \tag{13.2.11}$$

We may also define projections onto the four physical dimensions:

$$\left.\begin{aligned} \bar{g}^{\mu\nu} &= \begin{cases} g^{\mu\nu} & \text{if } \mu \text{ and } \nu \text{ are less than 4,} \\ 0 & \text{otherwise;} \end{cases} \\ \bar{V}^\mu &= \bar{g}^{\mu\nu}V_\nu. \end{aligned}\right\} \tag{13.2.12}$$

The following results are elementary, but will prove useful:

$$\left.\begin{aligned} \hat{\gamma}^\mu\hat{\gamma}^\nu\hat{\gamma}_\mu &= (6-d)\hat{\gamma}^\nu, \\ \hat{\gamma}^\mu\bar{\gamma}^\nu\hat{\gamma}_\mu &= (4-d)\bar{\gamma}^\nu, \\ \bar{\gamma}^\mu\bar{\gamma}^\nu\bar{\gamma}_\mu &= -2\bar{\gamma}^\nu, \\ \bar{\gamma}^\mu\hat{\gamma}^\nu\bar{\gamma}_\mu &= -4\hat{\gamma}^\nu. \end{aligned}\right\} \tag{13.2.13}$$

Let us define $\omega^a = (\omega^a_{\mathrm{L}} + \omega^a_{\mathrm{R}})/2$ and $\omega^a_5 = (-\omega^a_{\mathrm{L}} + \omega^a_{\mathrm{R}})/2$. Then the variation of $\mathscr{L}$ under the chiral transformations (13.1.1) is

$$\delta\mathscr{L} = \mathrm{i}\bar{\psi}(2\omega^0_5\gamma_5 + \omega^a[M, t^a] + \omega^a_5\gamma_5\{M, t^a\})\psi - 2\bar{\psi}\gamma_5\hat{\not{D}}(\omega^0_5 + \omega^a_5 t^a)\psi. \tag{13.2.14}$$

Hence the divergences of the axial currents are

$$\begin{aligned} \partial_\mu j^\mu_{5a} &= -\mathrm{i}\bar{\psi}\gamma_5\{t^a, M\}\psi + \bar{\psi}\gamma_5\hat{\not{D}}t^a\psi \\ &= -\mathrm{i}\bar{\psi}\gamma_5\{t^a, M\}\psi + \bar{\psi}\gamma_5\{t^a, \hat{\not{D}}\}\psi/2, \end{aligned} \tag{13.2.15}$$

$$\partial_\mu j^\mu_5 = -2\mathrm{i}\bar{\psi}\gamma_5 M\psi + \bar{\psi}\gamma_5\hat{\not{D}}\psi. \tag{13.2.16}$$

The second term in each equation can potentially give an anomaly when we let $d \to 4$.

13.3 Properties of axial currents

There are a number of somewhat different situations in which axial currents appear. The original papers on chiral anomalies primarily addressed anomalies in Ward identities of the chiral currents of strong interactions. More general cases have since been worked out, with corresponding generalizations of the Adler–Bardeen theorem. In this section we will list the various cases and state what is known.

13.3.1 *Non-anomalous currents*

The following properties apply, for example, to the non-singlet axial currents in QCD:

(1) If there is no anomaly to one-loop order (as for the current j^{μ}_{a5}), then there is no anomaly to all orders. The Ward identities of the current with elementary fields have no anomalies.
(2) Under the same condition as in (1) there is no anomaly in the two-current Ward identities.
(3) Under the same condition there is an anomaly in a three-current Ward identity (like the one for $\partial_{\kappa}\langle 0|Tj^{\kappa}_{a5}(x)j^{\mu}_{b}(y)j^{\nu}_{c}(z)|0\rangle$). However, the only non-zero term in the anomaly is the one-loop contribution.

The theorem that the complete anomaly in these cases is determined by the one-loop value is due to Adler & Bardeen (1969).

The lack of anomalies in the Ward identities of one current with elementary fields is essential if the currents are to generate the correct transformation law for the fields. These transformations imply commutation relations for the currents. Since these commutators are also given by the Ward identities with two currents, the two-current identities must be anomaly-free.

No such consistency requirement applies to the three-current Ward identity. The value of its anomaly is related to the decay rate for $\pi^0 \to 2\gamma$, and the lack of higher-order corrections enables a successful prediction to be made easily. (See Adler (1970); for reviews from the point of view of QCD, see Marciano & Pagels (1978) and Llewellyn-Smith (1980).)

13.3.2 *Anomalous currents*

The singlet current j_5 has an anomalous divergence. It has the form

$$\partial_{\mu} j^{\mu}_{5} = C(g) G^{a}_{\mu\nu} \tilde{G}^{a\mu\nu} + \text{mass terms}, \tag{13.3.1}$$

where $G^{a}_{\mu\nu}$ is the gluon field strength tensor and $\tilde{G}^{a}_{\mu\nu}$ is its dual:

$$\tilde{G}^{a\mu\nu} = \varepsilon^{\mu\nu\kappa\lambda} G^{a}_{\kappa\lambda}. \tag{13.3.2}$$

The Adler Bardeen theorem asserts that the coefficient of $C'(g)$ is equal to its one-loop value

$$C(g) = N_{fl} g^2/(32\pi^2).$$

However, the operators j_5 and $G\tilde{G}$ need renormalization, and there is not an obvious natural renormalization condition. So the value of C is susceptible to change by redefinition of the renormalization prescription. We will not treat this case here.

13.3.3 *Chiral gauge theories*

Theories like the Weinberg–Salam theory of weak interactions have a gauged chiral symmetry. It is essential that there be no anomaly, for otherwise the theory is not renormalizable and loses other important properties (Gross & Jackiw (1972) and Korthals Altes & Perrottet (1972)). A generalization of the Adler–Bardeen theorem is that there is no anomaly to any order of perturbation theory if there is none to one-loop order. Proofs have been given by Becchi, Rouet & Stora (1976), and by Costa *et al.* (1977).

13.3.4 *Supersymmetric theories*

Supersymmetric theories have potential anomalies similar to the chiral anomalies. A completely general treatment has not yet been given, but many particular cases have been treated – see Piguet & Rouet (1981), Piguet & Sibold (1982a, b, c), Clark, Piguet & Sibold (1979, 1980), and Jones & Leveille (1982).

13.4 Ward identity for bare axial current

Without use of the equations of motion the divergence of the non-singlet current j^μ_{a5} is

$$\begin{aligned}\partial_\mu j^\mu_{a5} &= \mathrm{i} Z_2 \bar{\psi}\gamma_5 t^a(\mathrm{i}\not{D} - M_0)\psi + hc \\ &\quad + \mathrm{i} Z_2 \bar{\psi}\{M_0, t^a\}\gamma_5\psi + \tfrac{1}{4} Z_2 \bar{\psi} t^a \{\overrightarrow{\not{D}}, \gamma_5\}\psi \\ &= D^a_{\mathrm{em}} + D^a_M + D^a_{\mathrm{anom}}. \end{aligned} \tag{13.4.1}$$

The first term we will call the equation of motion term. When inserted in a Green's function with elementary fields it gives

$$\begin{aligned}&\langle 0|T D^a_{\mathrm{em}}(x)\textstyle\prod A \prod_i \psi(y_i) \prod_j \bar{\psi}(z_j)|0\rangle \\ &= \sum_i \delta(x - y_i)\langle 0|T\textstyle\prod A \prod \psi \prod \bar{\psi}|0\rangle_{\psi(y_i)\to -\gamma_5 t^a \psi} \\ &\quad + \sum_i \delta(x - z_j)\langle 0|T\textstyle\prod A \prod \psi \prod \bar{\psi}|0\rangle_{\bar{\psi}(z_j)\to \bar{\psi}\gamma_5 t^a}. \end{aligned} \tag{13.4.2}$$

The variations of the fields are just their chiral transformations multiplied by i. (We set $\omega_R^a = -\omega_L^a = \omega_5^a$ in (13.1.1) to obtain these transformations.) This equation (13.4.2) has the expected right-hand side for the Ward identity in the absence of anomalies.

The mass term on the right of (13.4.1) is expected; it is the non-anomalous breaking.

The anomaly term is D^a_{anom}. If we insert it in a Green's function with no divergences of any kind, then we may set $d=4$. The result must be zero because D^a_{anom} vanishes if only the first four γ-matrices are used (for then $\{\not{D}, \gamma_5\} = 0$). But if the Green's function has an overall divergence, or if it has a divergent subgraph that contains the vertex for D^a_{anom}, then the result may be non-zero at $d=4$, as we will see. The full Ward identity reads

$$\frac{\partial}{\partial x^\mu}\langle 0|Tj^\mu_{a5}(x)\prod A\prod\psi\prod\bar{\psi}|0\rangle = \text{right-hand side of (13.4.2)} + \text{mass term} + \langle 0|TD^a_{\text{anom}}(x)\prod A\prod\psi\prod\bar{\psi}|0\rangle. \quad (13.4.3)$$

Recall that in the case of a symmetry such as the isospin $SU(2)$ of QCD that has no anomaly, we used its Ward identity to prove the current finite. The only possible counterterm for the current is proportional to itself, so finiteness of the divergence of the current, $\partial\cdot j$, implies finiteness of the current itself. The Ward identities imply that the divergence of the current is finite. However, for the axial currents the extra term in (13.4.3) prevents this argument from being made.

13.4.1 *Renormalization of operators in Ward identities*

Our aim will be to construct a finite current j^μ_{Ra5} that at $d=4$ satisfies a non-anomalous Ward identity:

$$\frac{\partial}{\partial x^\mu}\langle 0|Tj^\mu_{Ra5}(x)\prod A\prod\psi\prod\bar{\psi}|0\rangle = \text{right-hand side of (13.4.2).} \quad (13.4.4)$$

The first step is to observe that the only counterterm to j^μ_{a5} is itself. No other operators have the correct dimension and quantum numbers. So we can define a minimally subtracted operator

$$[j^\mu_{a5}] = Z_5 j^\mu_{a5} = \tfrac{1}{2}Z_5 Z_2 \bar{\psi}[\gamma^\mu, \gamma_5]t^a\psi. \quad (13.4.5)$$

Throughout this section we will use square brackets to indicate minimal subtraction. So the renormalization factor Z_5 has the form

$$Z_5 = 1 + \text{poles at } d=4.$$

To obtain the operator j_{Ra5}, we will later show that we have to invoke a further finite renormalization.

We will show that the Ward identity (13.4.4) is not correct if j_{Ra5} is replaced by the minimally subtracted $[j_{a5}]$. Rather we must make a further finite renormalization to obtain a Ward identity without anomalies. Thus we have

$$\begin{aligned} j^{\mu}_{Ra5} &= z_5[j^{\mu}_{a5}] \\ &= \tfrac{1}{2} z_5 Z_5 Z_2 \bar{\psi}[\gamma^{\mu}, \gamma_5] t^a \psi, \end{aligned} \tag{13.4.6}$$

where $z_5(g)$ is a finite factor.

The anomaly operator D^a_{anom} is a dimension four scalar quantity, so there are several operators with which it can mix. It is proportional to the ε-tensor times a fourth rank tensor, which we will call $\Delta^{\kappa\lambda\mu\nu}$. Since $\varepsilon_{\kappa\lambda\mu\nu}$ appears nowhere in the Lagrangian, the tensor $\Delta^{\kappa\lambda\mu\nu}$ is invariant under all Lorentz transformations. Given these restrictions, a complete list of operators that mix with D^a_{anom} is

$$\partial_{\mu} j^{\mu}_{a5}, D^a_{\text{anom}}, \bar{\psi}\{M_0, t^a\}\gamma_5\psi.$$

No operator involving only ghost and gluon fields can be constructed such that it mixes with D^a_{anom}. The linearity in $\varepsilon_{\kappa\lambda\mu\nu}$ implies the presence of four factors of vector objects (derivatives or A-fields). Therefore the coefficient of the operator is independent of mass, by our usual results. Gauge invariance of D^a_{anom} allows a restricted set of gauge variant counterterms (see Chapter 12) none of which have low enough dimension to appear. Since j_{a5} is even under charge conjugation, so is D^a_{anom}. Therefore the only allowed counterterm proportional to quark mass has a flavor factor $\{M, t^a\}$ rather than the commutator $[M, t^a]$; this gives us the operator D^a_M that appears in (13.4.1).

We can therefore write the minimally subtracted operator corresponding to $[D^a_{\text{anom}}]$ as

$$[D^a_{\text{anom}}] = Z_a D^a_{\text{anom}} + Z_{a5}\partial\cdot j_{a5} + Z_{am} D^a_M. \tag{13.4.7}$$

The equation of motion operator is finite by itself – see (13.4.2) – so we have

$$[D^a_{\text{em}}] = D^a_{\text{em}}. \tag{13.4.8}$$

Note that the definition of D^a_{em} includes some counterterms, but these are manufactured from the wave-function, mass, and coupling renormalizations in the Lagrangian.

We also need the renormalization of the mass operator

$$[D^a_M] = Z_{5M} D^a_M. \tag{13.4.9}$$

It is somewhat unobvious that the renormalization is multiplicative. The easiest method is to examine the γ-matrix structure of self-energy graphs with

an insertion of $\bar{\psi} t^a \gamma_5 \psi$ or $\bar{\psi} \gamma_5 \psi$. These operators can then be shown to be multiplicatively renormalizable with a common factor.

By use of (13.4.6)–(13.4.9) we can express the equation (13.4.1), for the divergence of the bare axial current, in terms of renormalized operators to find

$$[D^a_{\text{em}}] = Z_5^{-1}(1 - Z_{a5}Z_a^{-1})\partial_\mu[j^\mu_{a5}] - [D^a_M]Z_{5M}^{-1}(1 + Z_{aM}Z_a^{-1}) - Z_a^{-1}[D^a_{\text{anom}}]. \tag{13.4.10}$$

The renormalized operators on the right are all linearly independent, so the only way a linear combination of them can equal the finite left-hand side is for the coefficients to be finite. Since we use minimal subtraction this implies:

$$\left.\begin{aligned} Z_a &= 1, \\ Z_{aM} &= Z_{5M} - 1, \\ Z_{a5} &= 1 - Z_5. \end{aligned}\right\} \tag{13.4.11}$$

We therefore have the renormalized Ward identity:

$$\frac{\partial}{\partial x^\mu}\langle 0|T[j^\mu_{a5}]\prod A\prod\psi\prod\bar{\psi}|0\rangle = \text{r.h.s. of (13.4.2)} + \langle 0|T([D^a_M] + [D^a_{\text{anom}}])\prod A\prod\psi\prod\bar{\psi}|0\rangle, \tag{13.4.12}$$

which apparently still has an anomaly. Before showing how the anomaly in fact disappears, let us examine some low-order graphs.

13.5 One-loop calculations

The tree approximation for the two-point Green's function of $[D^a_{\text{anom}}]$ is given by Fig. 13.5.1. To save algebra we will set quark masses to zero. The graph's value is

$$\frac{i\not{p}'}{p'^2}(-i)(\hat{p}' + \hat{p})\gamma_5 t^a \frac{i\not{p}}{p^2}. \tag{13.5.1}$$

If we let the external momenta be physical, i.e., in the first four dimensions, then this vanishes. The vertex for D^a_{anom} has the property we define as evanescence: it vanishes when the cut-off is removed and we go to the physical renormalized theory. We will formulate a precise definition of evanescence later, when we have understood the subtleties associated with inserting the vertex inside loops.

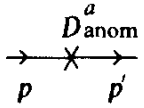


Fig. 13.5.1. Tree approximation for two-point Green's function of $[D^a_{\text{anom}}]$.

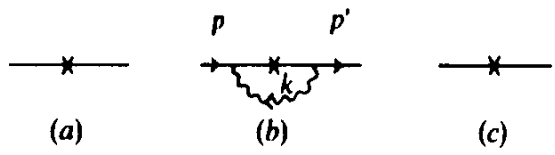


Fig. 13.5.2. Graphs up to order g^2 for two-point Green's function of j^{μ}_{a5}.

Next let us examine the graphs for the two-point function of the current j^{μ}_{a5}, as given in Fig. 13.5.2. Graph (*a*) has the value

$$\frac{\mathrm{i}\not{p}'}{p'^2}\bar{\gamma}^{\mu}\gamma_5 t^a \frac{\mathrm{i}\not{p}}{p^2}. \tag{13.5.2}$$

Taking the divergence is the same as multiplying by $\mathrm{i}(p'-p)^{\mu}$. The result is

$$\begin{aligned}
&\frac{\mathrm{i}\not{p}'}{p'^2}(\not{\bar{p}}' - \not{\bar{p}})\mathrm{i}\gamma_5 t^a \frac{\mathrm{i}\not{p}}{p^2} \\
&\quad = \frac{\mathrm{i}\not{p}'}{p'^2}(\not{\bar{p}}'\gamma_5 + \gamma_5 \not{\bar{p}})\mathrm{i}t^a \frac{\mathrm{i}\not{p}}{p^2} \\
&\quad = \frac{\mathrm{i}\not{p}'}{p'^2}(\not{p}'\gamma_5 + \gamma_5\not{p} - \not{\hat{p}}'\gamma_5 - \gamma_5\not{\hat{p}})\mathrm{i}t^a \frac{\mathrm{i}\not{p}}{p^2} \\
&\quad = -\gamma_5 t^a \mathrm{i}\not{p}/p^2 - \mathrm{i}(\not{p}'/p'^2)\gamma_5 t^a + (13.5.1).
\end{aligned} \tag{13.5.3}$$

When we set $\hat{p} = \hat{p}' = 0$, to get the four-dimensional result, we obtain the lowest-order case of the chiral Ward identity.

The next order graph is Fig. 13.5.2(*b*). Its value, with the external propagators amputated, is

$$\Gamma_{2b} = \frac{\mathrm{i}g^2 C_{\mathrm{F}}}{16\pi^4}(2\pi\mu)^{4-d}\int \mathrm{d}^d k \frac{\gamma_\nu(\not{p}' + \not{k})\bar{\gamma}^{\mu}\gamma_5 t^a(\not{p} + \not{k})\gamma^\nu}{k^2(p'+k)^2(p+k)^2}. \tag{13.5.4}$$

This is evidently divergent. It is easy to calculate the pole at $d = 4$:

$$\begin{aligned}
\mathrm{pole}\,(\Gamma_{2b}) &= \frac{g^2 t^a C_{\mathrm{F}}}{32\pi^2}\,\mathrm{pole}\left\{\frac{1}{4-d}\gamma_\nu\gamma_\kappa\bar{\gamma}^{\mu}\gamma_5\gamma^\kappa\gamma^\nu\right\} \\
&= \frac{g^2 t^a C_{\mathrm{F}}}{32\pi^2}\,\mathrm{pole}\left\{\frac{1}{4-d}\bar{\gamma}^{\mu}\gamma_5(d-6)^2\right\} \\
&= \frac{g^2}{8\pi^2}\frac{C_{\mathrm{F}} t^a}{4-d}\bar{\gamma}^{\mu}\gamma_5.
\end{aligned} \tag{13.5.5}$$

Here we have twice used the result that

$$\begin{aligned}
\gamma_\kappa\bar{\gamma}^{\mu}\gamma_5\gamma^\kappa &= \bar{\gamma}_\kappa\bar{\gamma}^{\mu}\gamma_5\bar{\gamma}^\kappa + \hat{\gamma}_\kappa\bar{\gamma}^{\mu}\gamma_5\hat{\gamma}^\kappa \\
&= 2\bar{\gamma}^{\mu}\gamma_5 - \hat{\gamma}_\kappa\hat{\gamma}^\kappa\bar{\gamma}^{\mu}\gamma_5 \\
&= (6-d)\bar{\gamma}^{\mu}\gamma_5.
\end{aligned} \tag{13.5.6}$$

There is a counterterm graph implicit in the definition $j^{\mu}_{a5} = Z_2\bar{\psi}\bar{\gamma}^{\mu}\gamma_5 t^a\psi$, with the quark wave-function renormalization given by

$$Z_2 = 1 - \frac{g^2 C_F}{8\pi^2(4-d)} + O(g^4). \tag{13.5.7}$$

The resulting counterterm graph Fig. 13.5.2(*c*) therefore cancels the UV divergence of graph (*b*), leaving a finite result. No additional renormalization is needed:

$$Z_5 = 1 + O(g^4). \tag{13.5.8}$$

Let us next take the divergence of (13.5.5) plus its counterterm, by multiplying by $i(p'-p)_{\mu}$. It is left as an exercise for the reader to verify that the Ward identity (13.4.3) holds at this order. What we will do is examine the graphs of order g^2 for the Green's function of D^a_{anom}. These are listed in Fig. 13.5.3. Note that the definition includes a covariant derivative:

$$\begin{aligned}\tfrac{1}{2}Z_2\bar{\psi}\{\overleftrightarrow{\not{D}}, \gamma_5\}t^a\psi &= \tfrac{1}{2}\bar{\psi}Z_2\{\overleftrightarrow{\not{\partial}}, \gamma_5\}t^a\psi - ig_0 Z_2\bar{\psi}\{\not{A}_0, \gamma_5\}t^a\psi \\ &= Z_2\bar{\psi}\overleftrightarrow{\not{\partial}}\gamma_5\psi - 2ig_0 Z_2\bar{\psi}\hat{\not{A}}_0\gamma_5 t^a\psi. \end{aligned}\tag{13.5.9}$$

The $\bar{\psi}\hat{\not{A}}\gamma_5\psi$ term gives rise to the graphs (*b*) and (*c*).

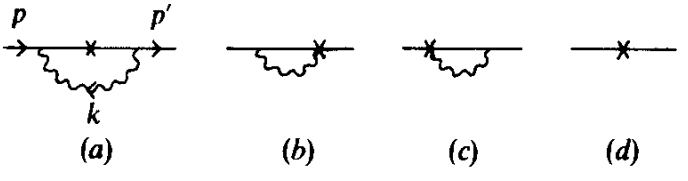


Fig. 13.5.3. Graphs of order g^2 for two-point Green's function of $[D^a_{\text{anom}}]$.

Graph (*a*) equals

$$\begin{aligned}\Gamma_{3a} &= \frac{C_F g^2 t^a}{16\pi^4}(2\pi\mu)^{4-d}\int d^d k \frac{\gamma_\nu(\not{p}' + \not{k})(\hat{\not{p}}' + \hat{\not{p}} + 2\hat{\not{k}})\gamma_5(\not{p} + \not{k})\gamma^\nu}{k^2(p'+k)^2(p+k)^2} \\ &= \frac{ig^2 C_F}{16\pi^2}t^a(2\pi\mu)^{4-d}\int_0^1 dx\int_0^{1-x} dy \times \\ &\quad \times \{\Gamma(3-d/2)D^{d/2-3}\gamma_\nu[\not{p}'(1-y) - \not{p}x][\hat{\not{p}}'(1-2y) + \hat{\not{p}}(1-2x)] \\ &\quad \times \gamma_5[\not{p}(1-x) - \not{p}'y]\gamma^\nu \\ &\quad - \tfrac{1}{2}\Gamma(2-d/2)D^{d/2-2}\gamma_\nu[2(\not{p}'(1-y) - \not{p}x)\hat{\gamma}_\kappa\gamma_5\gamma^\kappa \\ &\qquad\qquad + 2\gamma_\kappa\hat{\gamma}^\kappa\gamma_5(\not{p}(1-x) - \not{p}'y) \\ &\qquad\qquad + \gamma_\kappa(\hat{\not{p}}'(1-2y) + \hat{\not{p}}(1-2x))\gamma_5\gamma^\kappa]\gamma^\nu\}, \end{aligned}\tag{13.5.10}$$

where

$$D = -p^2x(1-x) - p'^2y(1-y) + 2p\cdot p'xy. \tag{13.5.11}$$

If it were not that there is an ultra-violet divergence given by the

$\Gamma(2-d/2)$, we could let p and p' be four-dimensional, and then set $d=4$ to obtain zero. (Note that $\hat{\gamma}_\kappa\gamma^\kappa=\hat{\gamma}_\kappa\hat{\gamma}^\kappa=d-4$.) The pole prevents this argument from being made. First of all, notice that the pole is

$$\begin{aligned}\text{pole}(\Gamma_{3a}) &= -\frac{ig^2C_F}{96\pi^2}t^a\,\text{pole}\left\{\frac{1}{4-d}\gamma_\nu\gamma_\kappa(\hat{\not{p}}'+\hat{\not{p}})\gamma_5\gamma^\kappa\gamma^\nu\right\}\\ &= \frac{-3ig^2C_F}{8\pi^2}t^a\frac{(\hat{\not{p}}'+\hat{\not{p}})}{4-d}.\end{aligned}\tag{13.5.12}$$

The manipulations on the Dirac matrices are easy to do incorrectly, so let us be careful. We need the following result

$$\begin{aligned}\gamma_\kappa\hat{\gamma}^\mu\gamma_5\gamma^\kappa &= \bar{\gamma}_\kappa\hat{\gamma}^\mu\gamma_5\bar{\gamma}^\kappa+\hat{\gamma}_\kappa\hat{\gamma}^\mu\gamma_5\hat{\gamma}^\kappa\\ &= \bar{\gamma}_\kappa\bar{\gamma}^\kappa\hat{\gamma}^\mu\gamma_5-\hat{\gamma}_\kappa\hat{\gamma}^\kappa\hat{\gamma}^\mu\gamma_5+2\hat{\gamma}^\mu\gamma_5\\ &= (10-d)\hat{\gamma}^\mu\gamma_5.\end{aligned}\tag{13.5.13}$$

In the first line we split γ^κ into a four-dimensional piece $\bar{\gamma}^\kappa$ and a $(d-4)$-dimensional piece $\hat{\gamma}^\kappa$. Then in the second line we used the commutators or anticommutators of $\bar{\gamma}^\kappa$ and $\hat{\gamma}^\kappa$ with $\hat{\gamma}^\mu$ and γ_5.

The graphs of Fig. 13.5.3(*b*) and (*c*) may be evaluated similarly. The sum of the pole terms for all three graphs is

$$\text{pole}\,(\Gamma_{3a}+\Gamma_{3b}+\Gamma_{3c})=\frac{-ig^2C_F}{8\pi^2}t^a\frac{(\hat{\not{p}}'+\hat{\not{p}})\gamma_5}{4-d},\tag{13.5.14}$$

which is cancelled by the counterterm Fig. 13.5.3(*d*). This is in agreement with our general result (13.4.11).

We are now ready to compute the value at $d=4$ of the sum of the graphs of Fig. 13.5.3. Considerable simplification occurs. Since $\hat{p}$ and $\hat{p}'$ are now zero, the term in (13.5.10) that multiplies $\Gamma(3-d/2)$ vanishes. Similarly the last term multiplying $\Gamma(2-d/2)$ gives zero. The remaining two terms have a factor $\hat{\gamma}^\kappa\hat{\gamma}_\kappa=d-4$, which cancels the pole to leave a finite result:

$$\Gamma_{3a}\big|_{d=4,\hat{p}=\hat{p}'=0}=\frac{ig^2C_F}{8\pi^2}t^a(\bar{\not{p}}'-\bar{\not{p}})\gamma_5.\tag{13.5.15}$$

Similarly Γ_{3b} and Γ_{3c} give

$$(\Gamma_{3b}+\Gamma_{3c})\big|_{d=4,\hat{p}=\hat{p}'=0}=\frac{ig^2C_F}{8\pi^2}t^a(\bar{\not{p}}'-\bar{\not{p}})\gamma_5.\tag{13.5.16}$$

Effectively Fig. 13.5.3 sums to $g^2C_F/(4\pi^2)$ times the vertex for $\partial\cdot j_{a5}$. It is easy to understand why the result should be of this form. Without the loop integration, the vertex for D^a_{anom} vanishes when $\hat{k}=\hat{p}=\hat{p}'=0$. When we include the integration over the components of $\hat{k}$, we can get a non-zero value for the graphs even if $\hat{p}=\hat{p}'=0$. However, the

evanescence property of the basic vertex implies that it has effectively a factor $d-4$. We only get a finite result by multiplying by an ultra-violet divergence – so the effect at $d=4$ is of a local operator.

A general theory of evanescent operators can be worked out. The results simply generalize what we have learnt from examples:

(1) We define an evanescent vertex as one that is finite and that vanishes in a tree graph when we set $d=4$ and when all momenta and polarizations are four-dimensional.
(2) A Green's function or a graph or an operator is evanescent if it is finite at $d=4$ and if it vanishes when its external momenta and polarizations are four-dimensional.
(3) Consider a graph containing an evanescent vertex. If the graph is completely finite then it is evanescent. ('Completely finite' means that the graph and all its subgraphs have negative degree of divergence.)
(4) A renormalized operator $[E]$ whose basic vertex is evanescent has the following expansion:

$$[E]=\sum_V C_{EV}[V]+\text{evanescent operators.} \tag{13.5.17}$$

The sum is over operators V whose basic vertices are non-evanescent. The only operators that are needed are the ones that according to the usual power-counting and symmetry requirements will mix with E under renormalization. The general proof is left as an exercise to the reader.

13.6 Non-singlet axial current has no anomaly

13.6.1 *Reduction of anomaly*

The only operator that can appear on the right-hand side of (13.5.17) for the case $E=D^a_{\text{anom}}$ is $\partial\cdot[j_{a5}]$. The restrictions are that it be pseudoscalar, isovector, gauge invariant and of dimension at most four. (If we had non-zero quark masses, then the operator $[D^a_M]$ could also appear.) So we have

$$[D^a_{\text{anom}}]=C(g)\partial\cdot[j^\mu_{a5}]+\text{evanescent}, \tag{13.6.1}$$

which, when substituted into the renormalized identity (13.4.12), gives

$$\frac{\partial}{\partial x^\mu}\langle 0|T(1-C)[j^\mu_{a5}]\prod A\prod\psi\prod\bar{\psi}|0\rangle$$
$$=\text{r.h.s. of (13.4.2)}+\text{evanescent}. \tag{13.6.2}$$

So we should define the renormalized current

$$\begin{aligned} j^{\mu}_{Ra5} &= (1 - C)[j^{\mu}_{a5}] \\ &= (1 - C)Z_5^{-1} j^{\mu}_{a5}, \end{aligned} \tag{13.6.3}$$

which is (13.4.6) with $z_5 = 1 - C$. For the physical four-dimensional theory this implies that j^{μ}_{Ra5} has Ward identities with no anomaly, viz. (13.4.4), as we wished to prove.

From our calculations of Fig. 13.5.3 we see that

$$C = g^2 C_F/(4\pi^2) + O(g^4). \tag{13.6.4}$$

Our proof has been long and involves a general theory of evanescent operators summarized in Section 13.5. The basic idea, however, is simple. The only way an anomaly can appear in the physical theory is when a divergence cancels an effective factor of $d - 4$ for the evanescence of an anomaly. The anomaly in the four-dimensional theory is a local operator, and the only possible operators are those which power-counting would allow as counterterms to $\partial \cdot j$.

In the case of our isovector current, the only such operator is $\partial \cdot j$ itself. So a finite renormalization (13.6.3) serves to eliminate the anomaly at $d = 4$.

13.6.2 *Renormalized current has no anomalous dimension*

Let us apply the renormalization-group operator $\mu \mathrm{d}/\mathrm{d}\mu$ to the Ward identity (13.4.4). For the right-hand side we get

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} \text{right-hand side} = \text{right-hand side} \{\Sigma \text{ anomalous dimensions of fields}\}.$$

To get the same result for the left-hand side, the current j^{μ}_{Ra5} must have zero anomalous dimension (when $d = 4$):

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} j_{Ra5} = 0. \tag{13.6.5}$$

This is a sensible result: the current j^{μ}_{Ra5} is a physical object, and it should not depend on how we parametrize the theory by a renormalized coupling.

Useful consequences follow, for the minimally subtracted current does have an anomalous dimension:

$$\mu \mathrm{d}[j_{a5}]/\mathrm{d}\mu = -\gamma_5[j_{a5}], \tag{13.6.6}$$

where

$$\begin{aligned} \gamma_5(g) &= \mu \mathrm{d} \ln(Z_5)/\mathrm{d}\mu \\ &= [(2 - d/2)g + \beta(g)]\partial \ln(Z_5)/\partial g. \end{aligned} \tag{13.6.7}$$

Now, the coefficient C is a function of g; it is dimensionless and even in the presence of masses cannot depend on them, just like the renormalization factor Z_5. We also define it to have no cut-off dependence, since it is a factor between renormalized operators at $d=4$.

We therefore have

$$\begin{aligned} 0 &= \mu \mathrm{d} j_{Ra5}/\mathrm{d}\mu \\ &= -\beta(\partial C/\partial g)[j_{a5}] - (1-C)\gamma_5[j_{a5}], \end{aligned}$$

so that

$$\beta \partial(1-C)/\partial g = \gamma_5(1-C). \tag{13.6.8}$$

It follows that

$$1 - C = \exp\left\{ \int_0^g \mathrm{d}g' \gamma_5(g')/\beta(g') \right\}, \tag{13.6.9}$$

where we used as a boundary condition the fact that C has a perturbation expansion starting at order g^2. In order that the integral in (13.6.9) be convergent, the order g^2 term in γ_5 must vanish.

So from the definition (13.6.7) it must be that Z_5 has no order g^2 term; this we know by explicit calculation. Moreover, we know from (13.6.4) the one-loop value of C, so that

$$\gamma_5 = -A_1 g^4 C_F/(2\pi^2) + O(g^6), \tag{13.6.10}$$

where the one-loop term in β is $-A_1 g^3$. Hence (13.6.10) gives us a prediction of the leading divergence in Z_5:

$$Z_5 = 1 - \frac{A_1 C_F g^4}{4\pi^2(4-d)} + O(g^6). \tag{13.6.11}$$

The reader is invited to check this by Feynman graph calculations.

We may use the techniques of Chapter 7 to sum the divergences. We find the full renormalization factor of j^{ν}_{Ra5} to be

$$\begin{aligned} \frac{j^{\mu}_{Ra5}}{j^{\mu}_{a5}} &= (1-C)Z_5^{-1} \\ &= \exp\left\{ \int_0^g \mathrm{d}g' \gamma_5(g') \left[\frac{1}{\beta(g')} - \frac{1}{(d/2-2)g' + \beta(g')} \right] \right\} \\ &= \exp\left\{ \int_0^g \mathrm{d}g' \frac{\gamma_5(g')(d/2-2)g'}{\beta(g')[(d/2-2)g' + \beta(g')]} \right\} \\ &= 1 + (d-4)O[\ln(g^2/(4-d))] \\ &\to 1 \quad \text{as } d \to 4. \end{aligned} \tag{13.6.12}$$

Evidently the Noether current is finite in the complete theory.

13.7 Three-current Ward identity; the triangle anomaly

13.7.1 *General form of anomaly*

We consider the Green's function

$$D^{\lambda\mu\nu}_{abc}(p,p') \equiv \int \mathrm{d}^4x\,\mathrm{d}^4y\,\mathrm{e}^{\mathrm{i}p\cdot x+\mathrm{i}p'\cdot y}\langle 0|\,Tj^{\lambda}_{Ra5}(0)j^{\mu}_{b}(x)j^{\nu}_{c}(y)|0\rangle. \quad (13.7.1)$$

Only connected graphs contribute; Lorentz invariance forces the vacuum expectation value of a current to be zero. The currents are all renormalized currents, so all subdivergences are cancelled by counterterms, and the only possible infinity in (13.7.1) is an overall divergence. In fact there is no overall divergence, as we will now show.

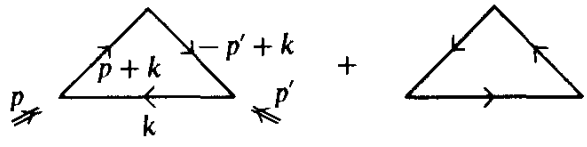


Fig. 13.7.1. Lowest-order graph for (13.7.1).

Individual graphs for (13.7.1) have a linear divergence, as can be seen from, say, Fig. 13.7.1. Any divergence must be linear in external momenta and proportional to the ε-tensor. The only possibility is

$$\varepsilon^{\kappa\lambda\mu\nu}[a(d)p_{\kappa}+b(d)p'_{\kappa}]. \quad (13.7.2)$$

There is also the constraint of conservation of the vector current. This is expressed by constructing a Ward identity in the dimensionally regularized theory.

Consider

$$p_{\mu}D^{\lambda\mu\nu} = -\mathrm{i}\int \mathrm{d}^dx\,\mathrm{d}^dy\,\mathrm{e}^{\mathrm{i}p\cdot x+\mathrm{i}p'\cdot y}\frac{\partial}{\partial x^{\mu}}\langle 0|\,Tj^{\lambda}_{Ra5}(0)j^{\mu}_{b}(x)j^{\nu}_{c}(y)|0\rangle. \quad (13.7.3)$$

By use of the result

$$\mathrm{i}\partial_{\mu}Z_2\bar{\psi}\gamma^{\mu}t^a\psi = Z_2\bar{\psi}t^a(\mathrm{i}\not{\partial}-M_0)\psi - Z\bar{\psi}(-\mathrm{i}\overleftarrow{\not{\partial}}-M_0)t^a\psi$$

and the equations of motion, we find

$$p_{\mu}D^{\lambda\mu\nu} = \int \mathrm{d}^dx\mathrm{i}e^{\mathrm{i}(p+p')\cdot x}\langle 0|\,Tj^{\lambda}_{Ra5}(0)Z_2\bar{\psi}[t^c,t^b]\gamma^{\nu}\psi(x)|0\rangle$$
$$+\int \mathrm{d}^dy\mathrm{i}e^{\mathrm{i}p'\cdot y}\langle 0|\,TZ_2\bar{\psi}[t^a,t^b]\bar{\gamma}^{\lambda}\gamma_5\psi\,j^{\nu}_{c}(y)|0\rangle. \quad (13.7.4)$$

In these equations we assumed that the currents j^{μ}_{b} and j^{ν}_{c} are conserved. Each of the terms in (13.7.4) is a Green's function of a vector and a pseudovector

current. Parity invariance forces them to be zero, so

$$p_\mu D^{\lambda\mu\nu} = 0. \tag{13.7.5a}$$

Similarly

$$p'_\nu D^{\lambda\mu\nu} = 0. \tag{13.7.5b}$$

The counterterm (13.7.2) must therefore give zero when multiplied by p_μ or p'_ν. This forces the whole counterterm to be zero; the Green's function (13.7.1) is finite as it stands.

In the regulated theory the axial current is not conserved, so we cannot prove the Ward identity $(p+p')_\lambda D^{\lambda\mu\nu} = 0$ by the same manipulations. Indeed we have

$$(p+p')_\lambda D^{\lambda\mu\nu} = \text{commutator terms} + \int \mathrm{d}^4x \mathrm{d}^4y \mathrm{e}^{\mathrm{i}(p\cdot x + p'\cdot y)} \times \\ \times \langle 0 | T E_a(0) j^\mu_b(x) j^\nu_c(y) | 0 \rangle. \tag{13.7.6}$$

Here E is the evanescent operator in (13.6.2):

$$\begin{aligned} E &= [D^a_{\text{anom}}] - C\partial\cdot[j_{a5}] \\ &= [D^a_{\text{anom}}] - C(1-C)^{-1}\partial\cdot j_{Ra5}. \end{aligned} \tag{13.7.7}$$

The commutator terms in (13.7.6) vanish, as in the Ward identity for the vector currents. Hence, finiteness of the left-hand side implies that the Green's function of E with j^μ_b and j^ν_c is finite. Even though graphs for it are quadratically divergent, the divergences cancel.

Now, E is an evanescent operator. This means that its Green's functions with elementary fields vanish in the four-dimensional theory. The general theory of evanescent operators, which we summarized at the end of Section 13.5, then tells us that the only way that the right-hand side of (13.7.6) will fail to vanish is for E to be part of a graph or subgraph with overall degree of divergence at least zero. Now, the definition of E has ensured that these subgraphs are all evanescent. Hence we are left with the complete graphs. So we have (at $d = 4$)

$$\begin{aligned} (p+p')_\lambda D^{\lambda\mu\nu} &= A(g)\varepsilon_{\mu\nu\alpha\beta}p^\alpha p'^\beta \varepsilon_{abc} \\ &= \tfrac{1}{2}A\varepsilon_{\mu\nu\alpha\beta}(p+p')^\alpha(p-p')^\beta\varepsilon_{abc}. \end{aligned} \tag{13.7.8}$$

The tensor structure is the only one possible. The coefficient A is dimensionless at $d = 4$, so it can only be a function of g. Note that the right-hand side of (13.7.8) obeys vector current conservation, so that

$$\begin{aligned} p_\mu(p+p')_\lambda D^{\lambda\mu\nu} &= 0, \\ p'_\nu(p+p')_\lambda D^{\lambda\mu\nu} &= 0, \end{aligned}$$

as should be.

13.7.2 *One-loop value*

The lowest-order value of A is easily computed from the graphs of Fig. 13.7.1.

$$A\varepsilon_{\mu\nu\alpha\beta}p^{\alpha}p'^{\beta} = \tfrac{3}{4}\int\frac{\mathrm{d}^d k}{(2\pi)^d}\frac{\mathrm{tr}\,[(\hat{\not{p}}-\hat{\not{p}}'+2\hat{\not{k}})\gamma_5(\not{p}+\not{k})\gamma^{\mu}\not{k}\gamma^{\nu}(\not{k}-\not{p}')]}{k^2(p+k)^2(p'-k)^2} + \text{charge conjugate}. \tag{13.7.9}$$

The factor 3 is the number of quark colors. To evaluate this, notice that

$$\left.\begin{aligned}\mathrm{tr}\,(\gamma_5\gamma^{\kappa}\gamma^{\lambda}) &= 0,\\ \mathrm{tr}\,(\gamma_5\gamma^{\kappa}\gamma^{\lambda}\gamma^{\mu}\gamma^{\nu}) &= 4\mathrm{i}\varepsilon^{\kappa\lambda\mu\nu}.\end{aligned}\right\} \tag{13.7.10}$$

Since $\varepsilon_{\kappa\lambda\mu\nu}$ is restricted to the first four dimensions, it follows that the trace of γ_5 with four Dirac matrices is zero if one of the matrices is a $\hat{\gamma}$:

$$\mathrm{tr}\,(\gamma_5\hat{\gamma}^{\kappa}\gamma^{\lambda}\gamma^{\mu}\gamma^{\nu}) = 0. \tag{13.7.11}$$

Let us commute the $-\not{p}'+\not{k}$ in (13.7.9) to the left, and use (13.7.11) whenever possible. We will also set $\hat{p}=\hat{p}'=0$. The result is

$$\begin{aligned}&\mathrm{tr}\,[(\hat{\not{p}}-\hat{\not{p}}'+2\hat{\not{k}})\gamma_5(\not{p}+\not{k})\gamma^{\mu}\not{k}\gamma^{\nu}(-\not{p}'+\not{k})]\\ &\quad=\mathrm{tr}\,[2\hat{\not{k}}\gamma_5(\not{p}+\not{k})\gamma^{\mu}\not{k}(\not{p}'-\not{k})\gamma^{\nu}]+0\\ &\quad=\mathrm{tr}\,[2\hat{\not{k}}\gamma_5(\not{p}+\not{k})\gamma^{\mu}\not{k}\not{p}'\gamma^{\nu}]+0\\ &\quad=-\mathrm{tr}\,[2\hat{\not{k}}\gamma_5(\not{p}+\not{k})\not{k}\gamma^{\mu}\not{p}'\gamma^{\nu}]+0\\ &\quad=-\mathrm{tr}\,[2\hat{\not{k}}\gamma_5\not{p}\not{k}\gamma^{\mu}\not{p}'\gamma^{\nu}]+0\\ &\quad=-\mathrm{tr}\,[2\hat{\not{k}}\not{k}\gamma_5\not{p}\gamma^{\mu}\not{p}'\gamma^{\nu}]+0.\end{aligned} \tag{13.7.12}$$

The terms indicated by '0' vanish by use of (13.7.11). The charge conjugate term gives an equal contribution.

We now have

$$A\varepsilon_{\mu\nu\alpha\beta}p^{\alpha}p'^{\beta} = 3\int\frac{\mathrm{d}^d k}{(2\pi)^d}\frac{\mathrm{tr}\,[\hat{\not{k}}\not{k}\gamma_5\not{p}\gamma^{\mu}\not{p}'\gamma^{\nu}]}{k^2(p'-k)^2(p+k)^2}. \tag{13.7.13}$$

This is now only a logarithmically divergent integral. After use of Feynman parameters the standard result (4.4.14) gives

$$\begin{aligned}A\varepsilon_{\mu\nu\alpha\beta}p^{\alpha}p'^{\beta} &= \frac{3\mathrm{i}}{16\pi^2}\int_0^1\mathrm{d}x\int_0^{1-x}\mathrm{d}y\,\tfrac{1}{2}\Gamma(2-d/2)\times\\ &\qquad\times\frac{\mathrm{tr}\,[\hat{\gamma}_{\kappa}\gamma^{\kappa}\gamma_5\not{p}\gamma^{\mu}\not{p}'\gamma^{\nu}]}{[-p^2x(1-x)-p'^2y(1-y)+2p\cdot p'xy]}+O(d-4)\\ &= -\frac{3}{8\pi^2}\varepsilon_{\mu\nu\alpha\beta}p^{\alpha}p'^{\beta}.\end{aligned} \tag{13.7.14}$$

The evanescence of the vertex has effectively given a factor of $d-4$ which

cancels the UV pole to leave a finite result

$$A = -3/(8\pi^2) + O(g^2). \tag{13.7.15}$$

This result for the anomaly in the axial Ward identity for (13.7.1) was first found by Adler (1969), and Bell & Jackiw (1969).

13.7.3 *Higher orders*

There are, in fact, (Adler & Bardeen (1969)) no higher-order corrections to the anomaly for (13.7.1). We will follow the proof due to Zee (1972). The basic idea is simple. Each of the currents in (13.7.1) is RG invariant, and there is no overall counterterm. Therefore this Green's function is invariant when we make an RG transformation. The anomaly must therefore be invariant also. But the anomaly coefficient $A(g)$ depends on the coupling g and on no other parameter of the theory. We can change g arbitrarily by changing the renormalization mass μ. Hence A is independent of g.

This proof may easily be written out. Renormalization-group invariance of $D^{\lambda\mu\nu}_{abc}$ is the equation

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} D^{\lambda\mu\nu}_{abc} = 0.$$

Hence

$$\mu \frac{\mathrm{d}}{\mathrm{d}\mu} A(g) \varepsilon_{abc} \varepsilon^{\mu\nu\alpha\beta} p_\alpha p'_\beta = (p + p')_\lambda \mu \frac{\mathrm{d}}{\mathrm{d}\mu} D^{\lambda\mu\nu}_{abc} = 0.$$

Since $\mu \mathrm{d}/\mathrm{d}\mu = \mu\partial/\partial\mu + \beta\partial/\partial g$, this gives

$$\beta \partial A/\partial g = 0.$$

Then A is independent of g, so it equals its lowest-order value:

$$A = -3/(8\pi^2) \text{ exactly.} \tag{13.7.16}$$

This is a very striking result. The proof we have given is very simple, but the reader should not suppose it is not a deep result. The whole power of the renormalization apparatus is needed for its derivation. We first had to show that there is no anomaly in the divergence of the axial current. Then we had to show that there was no counterterm needed to make $D^{\lambda\mu\nu}_{abc}$ finite. These results involved considerable cancelation of UV infinities. Since the anomaly in $\partial \cdot j_{a5}$ disappears when the UV cut-off is removed, it can affect a Ward identity only by being enhanced by a UV infinity which has not made its appearance earlier.

Thus the anomaly is associated with a UV pole implicit in the Feynman graphs. It is precisely for this reason that it must have the dependence on the

parameters of the theory and on the external momenta that is characteristic of a renormalization counterterm. In particular, it is polynomial in the momenta and masses of the degree determined by UV power-counting. Once this is clear, the most general possible form of the anomaly is (13.7.8). The final step to show that A is independent of g is trivial.

An important phenomenological consequence of the anomaly is a calculation of the decay rate for $\pi^0 \to 2\gamma$ (see Marciano & Pagels (1978) and Llewellyn-Smith (1980)). The amplitude is proportional to the number of quark colors, so the decay rate is proportional to the square of this number. The measured rate in fact agrees with the standard theory that there are three colors.

14

Deep-inelastic scattering

In this chapter we will show how the operator-product expansion can be used to compute the cross-section for deep-inelastic scattering. Since the calculation is fairly straightforward, this process is one of the classic tests of quantum chromodynamics (QCD).

The process is the scattering of a lepton of momentum l^μ on a hadron of momentum p^μ, with the only observed particle in the final state being the lepton:

$$l + N \to l' + \text{anything}. \tag{14.0.1}$$

In practice one uses a beam of electrons, muons, or neutrinos, and the hadrons in the target are nucleons. There are a number of cases for which there are experimental data, the cases being distinguished by the types of lepton involved:

$$\left.\begin{aligned} &e + N \to e + \text{anything}, \\ &\mu + N \to \mu + \text{anything}, \\ &\nu + N \to \nu + \text{anything}, \\ &\nu + N \to (e \text{ or } \mu) + \text{anything}. \end{aligned}\right\} \tag{14.0.2}$$

The basic reason for measuring the cross-section for these processes is to study the fundamental constituents (quarks and gluons) of the hadron. Suppose we have a scattering of a lepton on a small-sized constituent, and that the momentum transfer is large, so that the scattering happens over a small time-scale. The weak and electromagnetic interactions have a small coupling, so to a good approximation the lepton does not interact again. Moreover the interactions that turn the final state, involving the struck quark, into hadrons happen on a much longer time-scale. So these interactions do not interfere with the basic Born graph. Hence, we should be able to calculate the cross-section for the process (14.0.1) rather simply in terms of the cross-section for lepton–quark scattering (Fig. 14.0.1). The approximation in which final-state interactions of the hadrons are ignored is called the impulse approximation. We can use it because we choose to sum the cross-section over all hadronic final states.

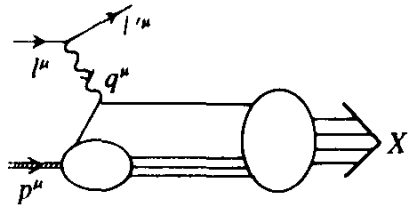


Fig. 14.0.1. Parton model for deep-inelastic scattering.

A more mathematical formulation of the same idea is the parton model, explained in Section 14.2. There are, however, many weak points in the intuitive discussion just given, and we must remedy them. In the remainder of this chapter we will compute the cross-section from the theory of strong interactions (QCD). The parton model will in fact give a qualitatively correct approximation to the real cross-section. Our intuitive argument shows why deep-inelastic scattering is simple enough to permit calculations.

Our treatment in this chapter is based on material to be found, among other places, in Gross (1976) and Treiman, Jackiw & Gross (1972).

14.1 Kinematics, etc.

We will compute the cross-section to lowest order in weak and electromagnetic interactions. Then the amplitudes contributing to the process (14.0.1) have the form of Fig. 14.1.1 where a boson is exchanged between the lepton and the hadron. The boson can be a photon, a W or a Z. At high enough energy, it is also necessary to include Higgs boson exchange. Higher-order weak and electromagnetic corrections do not need to be included, except for soft photon effects. We will ignore the soft photon corrections here, and will concentrate on understanding the strong-interaction corrections.

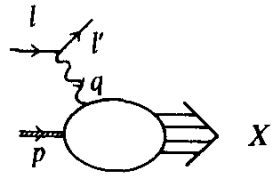


Fig. 14.1.1. Amplitude for process contributing to deep-inelastic scattering.

We first review the kinematics of the process. The two independent Lorentz invariants of the hadron system are chosen to be

$$Q^2 = -q^\mu q_\mu,$$
$$\nu = p\cdot q, \tag{14.1.1}$$

where q^μ is the momentum transfer from the leptons. The mass of the final hadron state X is then

$$m_X^2 = (p+q)^2 = m_N^2 + 2\nu - Q^2. \tag{14.1.2}$$

In the laboratory frame, where the initial hadron is at rest, we can express Q^2 and ν in terms of the initial and final lepton energies E and E', and of the lepton scattering angle θ:

$$Q^2 = 2EE'(1 - \cos\theta).$$
$$\nu = m_N(E - E'). \tag{14.1.3}$$

We have neglected lepton masses with respect to E and E'. The following inequalities hold

$$\nu \geq Q^2/2 \geq 0. \tag{14.1.4}$$

The region we will investigate is where both Q^2 and m_X^2 get large in a fixed ratio. We define the Bjorken scaling variable $x = Q^2/(2\nu)$. Then we let Q^2 get large with x fixed. This is called the Bjorken limit or the deep-inelastic region. In this region the missing mass m_X is large:

$$m_X^2 = Q^2(1/x - 1), \tag{14.1.5}$$

(where we neglect m_N^2 compared with Q^2). In order that m_X^2 be positive we must have $0 \leq x \leq 1$. The Bjorken limit applies only if x is not at its endpoints. An equivalent variable to x that is sometimes used is $\omega = 1/x$.

The cross-section is given by

$$E'\frac{d\sigma}{d^3p'} = \frac{1}{2\pi E'}\frac{d\sigma}{dE'd\cos\theta}$$
$$= \frac{\pi}{4ME}g_W^4 \sum_X |\langle l'|j_\lambda^{lept}|l\rangle\langle X|j_\nu^{had}|N\rangle D^{\lambda\nu}(q)|^2 \delta^{(4)}(p_X^\mu - p^\mu - q^\mu)$$
$$= \left(\frac{g_W^2}{4\pi}\right)^2 \frac{1}{ME} D^{\kappa\mu}(q)^* D^{\lambda\nu}(q) L_{\kappa\lambda}(l,l') W_{\mu\nu}(p,q). \tag{14.1.6}$$

Here j_λ^{lept} and j_ν^{had} are the currents to which the exchanged vector boson couples, g_W is its coupling, and $D^{\lambda\nu}(q)$ is its propagator. The lepton tensor $L_{\kappa\lambda}$ is easily computed in the tree approximation for

$$L_{\kappa\lambda} = \langle l|j_\kappa^{lept}(0)|l'\rangle\langle l'|j_\lambda^{lept}(0)|l\rangle. \tag{14.1.7}$$

The hadron tensor is equal to

$$W_{\mu\nu}(p,q) = \frac{1}{4\pi}\sum_X \langle p|j_\mu(0)^\dagger|X\rangle\langle X|j_\nu(0)|p\rangle(2\pi)^4\delta^{(4)}(p_X^\mu - p^\mu - q^\mu)$$
$$= \frac{1}{4\pi}\int d^4y e^{iq\cdot y}\langle p|j_\mu(y)^\dagger j_\nu(0)|p\rangle, \tag{14.1.8}$$

where the normalization is the standard convention.

Deep-inelastic scattering is the region where Q gets large with x fixed. This is not the short-distance limit we treated in Chapter 10. There we took all components of q^μ to infinity in a fixed ratio, so that $Q \to \infty$ with $Q/p\cdot q$ fixed,

i.e., x is proportional to Q. This is not in the physical region $0 \leq x \leq 1$ for deep-inelastic scattering. Luckily we can use a dispersion relation (Christ, Hasslacher & Mueller (1972)) to relate the deep-inelastic limit of $W_{\mu\nu}$ to the short-distance limit of the time-ordered matrix element

$$T_{\mu\nu} = \int d^4y e^{iq\cdot y} \langle p | T j_\mu(y)^\dagger j_\nu(0) | 0 \rangle. \tag{14.1.9}$$

We will perform this analysis in Section 14.3.

Let us now decompose $W_{\mu\nu}$ into tensors with scalar coefficients. We will assume that the hadron is unpolarized, i.e., that we average over its spin states. Then the most general form of $W_{\mu\nu}$ is

$$\begin{aligned} W_{\mu\nu} = {} & W_1(-g_{\mu\nu} + q_\mu q_\nu/q^2) + W_2(p_\mu - q_\mu \nu/q^2)(p_\nu - q_\nu \nu/q^2) \\ & - iW_3 \varepsilon_{\mu\nu\alpha\beta} p^\alpha q^\beta/(2M^2) + W_4 q_\mu q_\nu/M^2 \\ & + W_5(p_\mu q_\nu + q_\mu p_\nu)/(2M^2) - iW_6(p_\mu q_\nu - q_\mu p_\nu)/(2M^2). \end{aligned} \tag{14.1.10}$$

The scalar coefficients $W_i(Q^2, \nu)$ are called the structure functions. The normalizations are such that they are dimensionless. A number of properties follow from symmetries and basic quantum mechanics (Treiman, Jackiw, & Gross (1972)):

(1) Each W_i is real.
(2) Time reversal invariance of strong interactions implies that $W_6 = 0$.

In the case of a purely electromagnetic process, the current j_μ^{had} is conserved, so that $q^\mu W_{\mu\nu} = 0$. Hence $W_4 = W_5 = 0$. Moreover the electromagnetic current is a pure vector, and strong interactions are parity invariant. So $W_3 = 0$.

In the case of neutrino scattering, the currents are only conserved if quark masses are zero. In that case the only non-zero structure functions are W_1, W_2 and W_3. It turns out that when the masses are non-zero the contributions of W_4 and W_5 to the cross-section are suppressed by a factor of order $m_l m_q/Q^2$, where m_l and m_q are lepton and quark masses. We will discuss this further in Section 14.8.

When we compute the structure functions in the Bjorken limit, $Q^2 \to \infty, x$ fixed, we will find that they behave as a certain power of Q^2 times logarithms. The power is $(Q^2)^0$ for W_1, $1/Q^2$ for W_2 and W_3, and $1/Q^4$ for W_4 and W_5. So it is convenient to anticipate these results and define scaling structure functions $F_i(x, Q^2)$ which depend only logarithmically on Q^2. The standard definitions are:

$$\left.\begin{aligned} F_1(x, Q^2) &= W_1(Q^2, \nu = Q^2/(2x)), \\ F_i(x, Q^2) &= \nu W_i/M^2, \quad \text{for } i = 2 \text{ and } 3, \\ F_i(x, Q^2) &= \nu^2 W_i/M^4, \quad \text{for } i = 4 \text{ and } 5. \end{aligned}\right\} \tag{14.1.11}$$

14.2 Parton model

Before we discuss the true theoretical predictions for deep-inelastic scattering, let us discuss the parton model (Bjorken & Paschos (1969)). This is the simplest model for the process, and contains the essence of the correct physics. One considers the initial state hadron in the overall center-of-mass for the whole scattering. Now time dilation slows down processes in the hadron so that they typically occur on a rather long time-scale T of order Q/m_N^2. Then the scattering happens on a much shorter time-scale $T \sim 1/Q$. This implies that the hadron may be regarded as an assembly of non-interacting point-like constituents. These are what were originally (Feynman (1972)) called partons. (We now identify them as quarks and gluons.) Since the hadron in this frame is ultra-relativistic, we must regard the partons as massless and as each moving parallel to the hadron with a certain fraction z of its momentum.

The structure functions of the hadron are obtained by computing the partons' structure functions in tree approximation and by then summing over all partons weighted by their number density.

Let $f_{a/N}(z)\mathrm{d}z$ be the number of partons of type a in hadron N with fraction z to $z + \mathrm{d}z$ of its momentum. Then, for example, the electromagnetic structure functions are:

$$\begin{aligned} 2F_1^{\mathrm{em}}(x) &= x^{-1}F_2^{\mathrm{em}}(x) \\ &= \tfrac{4}{9}[f_{u/N}(x) + f_{\bar{u}/N}(x)] + \tfrac{1}{9}\sum_{q=d,s}[f_{q/N}(x) + f_{\bar{q}/N}(x)] \\ &\quad + \text{heavy quark terms.} \end{aligned} \tag{14.2.1}$$

The relation $F_2 = 2xF_1$ is characteristic of the spin-$\frac{1}{2}$ of the quarks (Callan & Gross (1969)). Notice that F_1 and F_2 are independent of Q^2. This is the property known as scaling. Experimentally, F_1 and F_2 obey this property approximately.

It is common in the literature to use the terms 'structure function' and 'quark distribution' interchangeably, because of the parton model relation between them (14.2.1), which is also approximately true in QCD – see Section 14.7. However, it is important to distinguish the two terms. Structure functions are coefficients of certain tensors in a current correlation function. They can be defined for other processes, e.g., the Drell–Yan process (Lam & Tung (1978)). On the other hand the quark distribution functions are exactly what their name implies: probability distributions of quarks in a hadron.

The parton model is correct in a super-renormalizable theory (Drell, Levy and Yan (1969, 1970a, b, c)). However, in a renormalizable theory like QCD, there are processes inside a hadron that happen significantly on all time scales

down to zero. Then the basic assumption of the parton model does not hold. However, the fact that it is short times that cause the problems allows the operator-product expansion to come into play to solve the problem. There are two approaches, essentially equivalent for deep-inelastic scattering:

(1) Use a dispersion relation to show that moments of the structure functions (i.e.,

$$F_{i,N}(Q^2) = \int_0^1 dx\, x^{N-1} F_i(x, Q^2)) \tag{14.2.2}$$

can be directly computed by the Wilson expansion. This approach was initiated by Christ, Hasslacher & Mueller (1972), and it is the one we will use.

(2) One can generalize the derivation of the operator-product expansion (Amati, Petronzio & Veneziano (1978), Ellis *et al.* (1979), Libby & Sterman (1978), Stirling (1978)). For deep-inelastic scattering, the result is equivalent to the first method without the taking of moments.

14.3 Dispersion relations and moments

Consider the time-ordered Green's function $T_{\mu\nu}$ defined by (14.1.9). It can be expanded in scalar structure functions $T_1, T_2, T_3, T_4, T_5, T_6$, just like $W_{\mu\nu}$. We will only need T_1, T_2, and T_3. The operator-product expansion derived in Chapter 10 can be applied to $T_{\mu\nu}$ when Q^2 and ν get large with Q^2/ν^2 fixed. As we have seen, this is not the scaling region, for we have $x \to \infty$ instead of x fixed. However, we will relate $T_{\mu\nu}$ to $W_{\mu\nu}$ by a dispersion relation. Then we will see that information on $W_{\mu\nu}$ in the physical region can be obtained from the operator-product expansion for $T_{\mu\nu}$.

If Q^2 is fixed and positive then each T_i is analytic in the ν-plane. There are cuts going out to infinity from the thresholds $\nu = \pm Q^2/2$. See Fig. 14.3.1. (This is a standard property. It can be proved by expressing

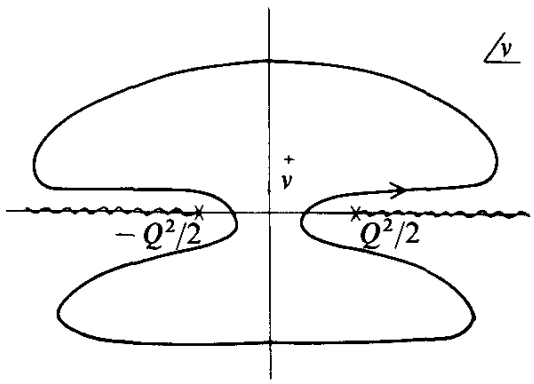


Fig. 14.3.1. Analyticity of $W_{\mu\nu}$ and contour to derive (14.3.1).

$\langle N|Tj_\mu(y)j_\nu(0)|N\rangle$ in terms of $W_{\mu\nu}$ and noting that $W_{\mu\nu}$ is zero if $|\nu|<Q^2/2$.) The discontinuities across the cut are:

$$\left.\begin{aligned} T_{\mu\nu}(p,q)\big|_{\nu-i\varepsilon}^{\nu+i\varepsilon} &= 4\pi W_{\mu\nu}(p,q), \quad (\text{if } \nu>0).\\ T_{\mu\nu}(p,q)\big|_{\nu+i\varepsilon}^{\nu-i\varepsilon} &= 4\pi \bar{W}_{\nu\mu}(p,q), \quad (\text{if } \nu<0), \end{aligned}\right\} \tag{14.3.1}$$

where $\bar{W}_{\mu\nu}$ is $W_{\mu\nu}$ with j replaced by its hermitian conjugate $j^\dagger$:

$$\bar{W}_{\mu\nu}=\frac{1}{4\pi}\int d^4y e^{iq\cdot y}\langle p|j_\mu(y)j_\nu(0)^\dagger|p\rangle. \tag{14.3.2}$$

For the electromagnetic or neutral-current processes the current is hermitian, so that $\bar{W}_{\mu\nu}=W_{\mu\nu}$. But if $W_{\mu\nu}$ is for charged-current neutrino scattering then $\bar{W}_{\mu\nu}$ gives the structure functions for antineutrino scattering.

By Cauchy's theorem we have

$$T_i(Q^2,\nu)=\frac{1}{2\pi i}\int_C \frac{d\nu'}{\nu'-\nu}T_i(Q^2,\nu'), \tag{14.3.3}$$

where C is any contour enclosing ν, as shown in Fig. 14.3.1. We will be able to compute the T_i's in the short-distance limit $\nu/Q^2\to 0$. So suppose we expand $T_i(Q^2,\nu)$ in a power series in $1/x=2\nu/Q^2$:

$$\begin{aligned} T_1 &= \sum_{n=0}^{\infty} T_{1,n}(Q^2)x^{-n},\\ \nu T_i/M^2 &= \sum_{n=0}^{\infty} T_{i,n}(Q^2)x^{-n} \quad (i=2 \text{ or } 3). \end{aligned} \tag{14.3.4}$$

(We expand $\nu T_i/M^2$ (if i is 2 or 3) in analogy with (14.1.11).) Then from (14.3.3)

$$T_{i,n}(Q^2)=\frac{1}{2\pi i}\int_C \frac{d\nu'}{\nu'}\left(\frac{Q^2}{2\nu'}\right)^n T_i(Q^2,\nu')(\nu'/M^2)^{a_i}, \tag{14.3.5}$$

where $a_i=0$ if $i=1$ and $a_i=1$ if $i=2$ or 3. If n is large enough to give convergence at $|\nu'|=\infty$, then we can deform the contour and pick up only the contribution from the discontinuity of T_i across the cuts:

$$\begin{aligned} T_{i,n}(Q^2)=&\frac{2}{i}\int_{Q^2/2}^{\infty}\frac{d\nu'}{\nu'}\left(\frac{Q^2}{2\nu'}\right)^n W_i(Q^2,\nu')\left(\frac{\nu'}{M^2}\right)^{a_i}\\ &+\frac{2}{i}\int_{-\infty}^{-Q^2/2}\frac{d\nu'}{\nu'}\left(\frac{Q^2}{2\nu'}\right)^n \bar{W}_i(Q^2,-\nu')\left(\frac{\nu'}{M^2}\right)^{a_i}. \end{aligned} \tag{14.3.6}$$

Finally we write the right-hand side in terms of the scaling functions $F_i(x,Q^2)$:

$$\begin{aligned} T_{i,n}(Q^2) &= -2i\int_0^1 dx' x'^{n-1}[F_i(x',Q^2)+(-1)^{n+a_i}\bar{F}_i(x',Q^2)]\\ &= -2i[F_{i,n}(Q^2)+(-1)^{n+a_i}\bar{F}_i(Q^2)]. \end{aligned} \tag{14.3.7}$$

This is the dispersion relation referred to earlier. It relates the power series expansion of $T_{\mu\nu}$ about $\nu = 0$ to the moments of the F's, which are defined by

$$F_{i,N}(Q^2) = \int_0^1 \mathrm{d}x\, x^{N-1} F_i(x, Q^2). \tag{14.3.8}$$

The relation (14.3.7) only applies if the integral is convergent. For small enough values of n it diverges. If $\nu^{a_i} W_i$ behaves like ν^p as $\nu \to \infty$, then we have convergence only if n is greater than p. Now the limit $\nu \to \infty$ at fixed Q^2 is a Regge limit (elastic scattering of a virtual boson off a hadron at energy m_X^2). So there is a standard expectation (Treiman, Jackiw & Gross (1972)) that $p = 1$ for $i = 1$ or 3 and $p = 0$ for $i = 2$. This is equivalent to $F_1, F_2/x, F_3$ all behaving like $1/x$ as $x \to 0$. In the parton model this would correspond to a $1/x$ behavior for quark distributions, and is roughly what is measured experimentally.

Our theoretical predictions will give all the terms in the series expansions of the T_i's, e.g. (14.3.4). Those coefficients for which (14.3.7) does not apply will not have any direct implications for deep-inelastic cross-sections.

14.4 Expansion for scalar current

To explain without a slew of indices the method for computing moments of structure functions, let us first work out the case where j_μ is replaced by a hermitian *scalar* operator j. For example, j might be $ZZ_m \bar{q}_i q_i$, appropriate to the coupling of a scalar boson to a particular flavor i of quark. We have a single scalar structure function:

$$F(x, Q^2) = W(\nu, Q^2) = (1/4\pi) \int \mathrm{d}^4 y \mathrm{e}^{\mathrm{i}q\cdot y} \langle p | j(y) j(0) | p \rangle, \tag{14.4.1}$$

while the time-ordered function is:

$$T(\nu, Q^2) = \int \mathrm{d}^4 y \mathrm{e}^{\mathrm{i}q\cdot y} \langle p | \, T j(y) j(0) | p \rangle. \tag{14.4.2}$$

(Notice that we choose j to be a renormalized operator. As is the case for the ϕ^2 operator for a scalar field, the renormalization factor for $\bar{q}q$ is the same as the mass renormalization factor, so $[\bar{q}q] = ZZ_m \bar{q}q = Z_m \bar{q}_0 q_0$, where q_0 is the bare quark field.)

The dispersion-relation argument applied to the series

$$T(Q^2/(2x), Q^2) = \sum_{n=0}^{\infty} T_n(Q) x^{-n} \tag{14.4.3}$$

gives

$$T_n(Q) = -2\mathrm{i}[F_n(Q) + (-1)^n \bar{F}_n(Q)]. \tag{14.4.4}$$

Since the current is hermitian, we have $F(x, Q) = \bar{F}(x, Q)$, so that

$$T_n(Q) = \begin{cases} -4iF_n(Q), & \text{if } n \text{ is even,} \\ 0, & \text{if } n \text{ is odd.} \end{cases} \tag{14.4.5}$$

Regge theory suggests $T \sim 1/x$ as $x \to 0$, so this equation is valid only if n is bigger than 1.

We now apply the operator-product expansion. The results of Chapter 10 apply in the limit $Q^2 \to \infty$ with ν^2/Q^2 fixed, i.e., with $x =$ constant times Q. Now

$$T_n x^{-n} = Q^{-n}[T_n(2\nu/Q)^n], \tag{14.4.6}$$

so that it would appear that all but the $n = 0$ term are non-leading by a power of Q^2 and that we only have a reliable prediction for $n = 0$. But the relation (14.4.5) is not expected to hold unless $n > 1$.

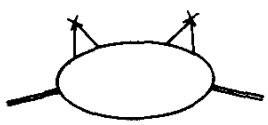


Fig. 14.4.1. $T(\nu, Q)$.

To remedy this problem, we must find an object for which $x^{-n}T_n$ contains the leading-power behavior as $Q \to \infty$. This is done by making a partial wave decomposition in the t-channel. That is, we treat $T(\nu, Q)$ (Fig. 14.4.1) as we would treat a scattering amplitude, and decompose it in terms of angular momenta:

$$\langle p|Tj(y)j(0)|0\rangle = \sum_{J=0}^{\infty} \langle p|V_J|p\rangle M_J(y\cdot q/q^2). \tag{14.4.7}$$

Here $\langle p|V_J|p\rangle$ is the reduced matrix element of some operator $V_{J,m}$ of spin J, and the M_J are appropriate polynomials in $y\cdot q/q^2$. (The operator V_J is not necessarily local.)

Now we perform an operator-product expansion of $Tj(y)j(0)$, keeping the leading-power behavior for each spin:

$$\langle p|Tj(y)j(0)|p\rangle \sim -2i\sum_{J,a}(-i)^J \langle p|\mathcal{O}^{(J,a)}_{\mu_1\ldots\mu_J}|p\rangle C^{Ja}(y^2)y^{\mu_1}\ldots y^{\mu_J}. \tag{14.4.8}$$

Here the operator $\mathcal{O}^{Ja}_{\mu_1\ldots\mu_J}$ is a local operator of spin J. The label 'a' denotes different operators of the same spin. Only the symmetric part of the operator is relevant, and in order that it be of definite spin, it must be traceless. Since

the hadron is unpolarized the matrix element has the form

$$\langle p|\mathcal{O}^{Ja}_{\mu_1\ldots\mu_J}|p\rangle = \langle p|\mathcal{O}^{Ja}|p\rangle(p_{\mu_1}\ldots p_{\mu_J} - \text{traces}), \tag{14.4.9}$$

where $\langle p|\mathcal{O}^{Ja}|p\rangle$ is the reduced matrix element – a scalar quantity. The normalizations in (14.4.8) have been adjusted for later convenience.

The leading power of y as $y\to 0$ is obtained by using operators of the minimum possible dimension. In QCD these are

$$\left.\begin{aligned}\mathcal{O}^{Jq}_{\mu_1\ldots\mu_J} &= 2^{1-J}\bar{q}\gamma_{\mu_1}\mathrm{i}D_{\mu_2}\ldots\mathrm{i}D_{\mu_J}q, \text{ symmetrized minus traces,}\\ \mathcal{O}^{Jg}_{\mu_1\ldots\mu_J} &= -2^{3-J}G_{\mu_1\nu}\mathrm{i}D_{\mu_2}\ldots\mathrm{i}D_{\mu_{J-1}}G^{\nu}_{\mu_J}, \text{ symmetrized minus traces.}\end{aligned}\right\} \tag{14.4.10}$$

In accordance with the results of Section 12.8 we have kept only gauge-invariant operators. (We use D_μ to denote the covariant derivative, and q to denote the field of a quark of flavor q. Sums over color indices are implicit in (14.4.10). Only hermitian operators are needed.)

Our usual power-counting arguments imply that the behavior of the scalar coefficient $C^{Ja}(y^2)$ in (14.4.8) is

$$(y^2)^{-\dim(C)} \text{ times logarithms,}$$

with

$$\begin{aligned}\dim(C) &= 2\dim(j) - \dim(\mathcal{O}^{Ja}_{\mu_1\ldots\mu_J}) + J\\ &= 2\dim(j) - \dim[\langle p|\mathcal{O}^{Ja}|p\rangle/\langle p|p\rangle].\end{aligned} \tag{14.4.11}$$

The dimension minus the spin of an operator is evidently the important quantity here; it is called the twist of the operator. The leading twist is two, for the operators of (14.4.10), and for them $C(y)\sim y^{-4}$ modulo logarithms.

Fourier transformation of (14.4.8) now gives

$$T(\nu, Q^2) = -2\mathrm{i}\sum_{J,a}\langle p|\mathcal{O}^{Ja}|p\rangle\tilde{C}^{Ja}(Q)x^{-J} + \text{non-leading powers of } Q, \tag{14.4.12}$$

where we define

$$\tilde{C}^{Ja}(Q) = Q^{2J}(\partial/\partial Q^2)^J\int \mathrm{d}^4 y\, \mathrm{e}^{\mathrm{i}q\cdot y}C^{Ja}(y^2). \tag{14.4.13}$$

Perturbative calculations of $T(\nu, Q^2)$ will give us $\tilde{C}^{Ja}$ in (14.4.13). We can then obtain moments of $F(x, Q)$ from the dispersion relation:

$$2F_J(Q) = \sum_a \langle p|\mathcal{O}^{Ja}|p\rangle\tilde{C}^{Ja}(Q)[1 + O(1/Q)], \tag{14.4.14}$$

if J is greater than one.

14.5 Calculation of Wilson coefficients

There are two parts to the calculations. The first part is to do low-order calculations of the Wilson coefficients

$$\tilde{C}^{Ja}(g, Q/\mu),$$

where we have now explicitly indicated the dependence on all parameters. Higher-order corrections have logarithms of Q/μ, so the second part of the calculation is to compute the anomalous dimensions and then to do a renormalization-group transformation to set μ to be of order Q. Thus we write

$$2F_n(Q) = \sum_{a,a'} \langle p|\mathcal{O}^{Ja}|p\rangle_{(\mu)} \tilde{C}^{Ja'}(g(Q),1) M_{a,a'}(g(Q), Q/\mu). \quad (14.5.1)$$

The subscript (μ) on the matrix element denotes renormalization with unit of mass μ. The matrix M is obtained by solving the renormalization group equation for C, and the $\tilde{C}^{Ja'}(g(Q),1)$ is well approximated by its lowest-order term in perturbation theory. Measurements of deep-inelastic scattering at one value of Q are enough to give $\langle p|\mathcal{O}^{Ja}|p\rangle$, and then (14.5.1) predicts the moments of the structure functions at other values of Q.

14.5.1 *Lowest-order Wilson coefficients*

The Wilson coefficients are independent of the target, so we may calculate them with the hadron state replaced by a quark state. In tree approximation we have the expansion sketched in Fig. 14.5.1. The scalar 'current' j is the renormalized operator $ZZ_m\bar{q}_i q_i = [\bar{q}_i q_i]$, for a particular quark flavor i. Since the Wilson coefficients are independent of mass, we set quark masses to zero. Then from the graphs of Fig. 14.5.1 we find

$$\begin{aligned} T &= \frac{\mathrm{i}\,\mathrm{tr}[\not{p}(\not{p}+\not{q})]}{2\quad (p+q)^2} + \frac{\mathrm{i}\,\mathrm{tr}[\not{p}(\not{p}-\not{q})]}{2\quad (p-q)^2} \\ &= -2\mathrm{i}\frac{1/x^2}{1-1/x^2}. \end{aligned} \quad (14.5.2)$$

The factor $\frac{1}{2}$ comes from averaging over the spin of the quark. For a quark of any other flavor than i, or for a gluon, we have $T = 0$ to this order.

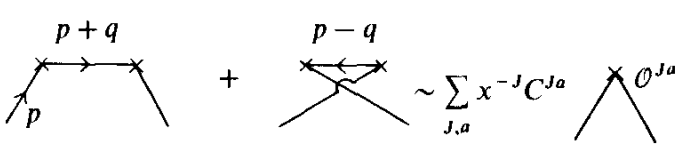


Fig. 14.5.1. Wilson expansion of $\langle 0|Tjj|0\rangle$ to lowest order.

To compute the Wilson coefficients we also need the matrix elements of the operators defined by (14.4.10). In a quark state with flavor i we have

$$\begin{aligned}\langle pi|\mathcal{O}^{Ji'}_{\mu_1\ldots\mu_J}|pi\rangle &= \tfrac{1}{2}\delta_{ii'}\,\mathrm{tr}(\not{p}\gamma_{\mu_1}p_{\mu_2}\cdots p_{\mu_J}) \\ &= 2\delta_{ii'}p_{\mu_1}\cdots p_{\mu_J},\end{aligned} \tag{14.5.3}$$

in tree approximation, with the label i' denoting a quark flavor or a gluon. For a gluon state and gluon operator

$$\langle pg|O^{Jg}_{\mu_1\ldots\mu_J}|pg\rangle = 2p_{\mu_1}\cdots p_{\mu_J}. \tag{14.5.4}$$

All other matrix elements are zero. Hence the non-zero reduced matrix elements are all equal to 2.

Comparison of (14.5.2)–(14.5.4) with the operator-product expansion (14.4.12) shows that

$$\left.\begin{aligned}&\tilde{C}^{Ji} = 1 + O(g^2), \quad \text{if} \quad J \geq 1 \text{ and is even,} \\ &\tilde{C}^{Ji'} = O(g^2), \quad \text{if} \quad i' \neq i, \text{ or if } i' \text{ is a gluon.}\end{aligned}\right\} \tag{14.5.5}$$

14.5.2 *Anomalous dimensions*

The anomalous dimension of the operator $ZZ_m\bar{q}_iq_i$ is γ_m:

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}(ZZ_m\bar{q}q) = \gamma_m(g)ZZ_m\bar{q}q, \tag{14.5.6}$$

and we let $\gamma^J_{aa'}(g)$ be the anomalous dimension matrix of the $\mathcal{O}^{Ja}$'s:

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}\mathcal{O}^{Ja} = \sum_{a'}\gamma_{aa'}(g)\mathcal{O}^{Ja'}. \tag{14.5.7}$$

(Operators of different spin do not mix.) As shown in Section 10.5 in a simpler case, the Wilson expansion then implies a renormalization-group equation for the Wilson coefficients:

$$\left[\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g}\right]C^{Ja}(g, Q/\mu) = -\sum_{a'}\gamma_{a'a}C^{Ja'} + \gamma_m C^{Ja} \tag{14.5.8}$$

This equation would be trivial to solve were it not that it is a matrix equation.

The lowest-order counterterms for the operators are generated by the graphs of Fig. 14.5.2. It is evident that the different operators mix. To solve the RG equation we must diagonalize the anomalous dimension matrix at order g^2. The first step is to recall that the counterterms are independent of quark masses. So the renormalizations respect the $SU(n_{\mathrm{fl}})$ symmetry of the flavor space. Therefore let us now choose a new basis for the twist-2 operators.

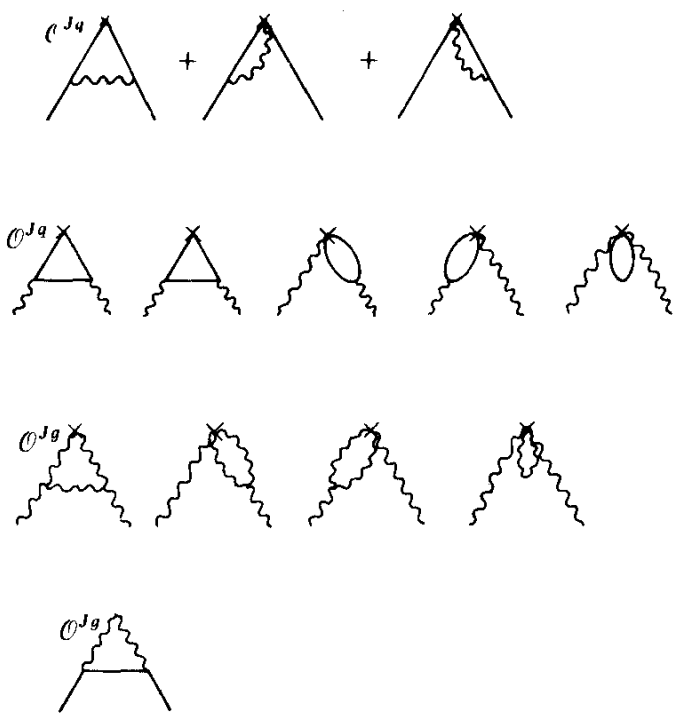


Fig. 14.5.2. Lowest-order divergences of the operators in the Wilson expansion of $\langle 0|Tjj|0\rangle$.

There is a multiplet of non-singlet operators:

$$\mathcal{O}^{J,NS,a} = 2^{1-J}\bar{\psi}\gamma_{\mu_1}\mathrm{i}\overleftrightarrow{D}_{\mu_2}\ldots\mathrm{i}\overleftrightarrow{D}_{\mu_J}\lambda^a\psi, \tag{14.5.9}$$

where the λ^a's are the $n^{\mathrm{fl}} \times n^{\mathrm{fl}}$ matrices that generate the flavor symmetry. There are two singlet operators:

$$\left.\begin{aligned} &\mathcal{O}^{Jg}, \text{ defined by (14.4.10)},\\ &\mathcal{O}^{JS} = \sum_{\mathrm{flav},i} \mathcal{O}^{Ji}. \end{aligned}\right\} \tag{14.5.10}$$

The renormalizations are

$$\begin{pmatrix} [\mathcal{O}^{Jg}] \\ [\mathcal{O}^{JS}] \\ [\mathcal{O}^{JNS}] \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} & 0 \\ Z_{21} & Z_{22} & 0 \\ 0 & 0 & Z_{NS} \end{pmatrix} \begin{pmatrix} \mathcal{O}_0^{Jg} \\ \mathcal{O}_0^{JS} \\ \mathcal{O}_0^{JNS} \end{pmatrix}, \tag{14.5.11}$$

$$[\mathcal{O}_\alpha^J] = \sum_\beta Z_{\alpha\beta}\mathcal{O}_{\beta,0}^J,$$

from which the anomalous dimension matrix

$$\gamma_{\alpha\beta} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 \\ \gamma_{21} & \gamma_{22} & 0 \\ 0 & 0 & \gamma_{NS} \end{pmatrix} \tag{14.5.12}$$

is obtained by

$$\sum_{\beta} \gamma_{\alpha\beta} Z_{\beta\gamma} = \beta \frac{\partial}{\partial g} Z_{\alpha\gamma}. \tag{14.5.13}$$

A calculation of the divergences of the graphs of Fig. 14.5.2 gives (Gross (1976))

$$\begin{aligned}
\gamma^J_{22} &= \gamma^J_{NS} = \frac{g^2}{6\pi^2}\left[1 - \frac{2}{J(J+1)} + 4\sum_{k=2}^{J}\frac{1}{k}\right], \\
\gamma^J_{11} &= \frac{g^2}{8\pi^2}\left[1 - \frac{12}{J(J-1)} - \frac{12}{(J+1)(J+2)} + 12\sum_{2}^{J}\frac{1}{k} + \frac{2}{3}n_{f1}\right], \\
\gamma^J_{21} &= -\frac{g^2 n_{f1}}{4\pi^2}\frac{(J^2+J+2)}{J(J+1)(J+2)}, \\
\gamma^J_{12} &= -\frac{g^2}{3\pi^2}\frac{(J^2+J+2)}{J(J^2-1)}.
\end{aligned} \tag{14.5.14}$$

14.5.3 *Solution of RG equation – non-singlet*

The Wilson coefficients of the non-singlet operators evolve very simply:

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu} C^J_{NS}(g(\mu), Q/\mu) = (-\gamma^J_{NS} + \gamma_m) C^J_{NS}. \tag{14.5.15}$$

An approximate solution can be found by taking the one-loop approximation for the anomalous dimensions. This gives

$$C^J_{NS}(g(\mu), Q/\mu) = C^J_{NS}(g(Q), 1)\left[\frac{\ln(Q/\Lambda)}{\ln(\mu/\Lambda)}\right]^{[\gamma^{(1)}_{NS,J} - \gamma^{(1)}_m]/(2A_1)} \tag{14.5.16}$$

where $\gamma^{(1)}_{NS,J}$ and $\gamma^{(1)}_m$ denote the coefficients of $g^2/4\pi^2$ in $\gamma^J_{NS}(g)$ and $\gamma_m(g)$, and $-A_1$ is the coefficient of $g^3/(4\pi^2)$ in $\beta(g)$. We may replace $C^{J,NS}(g(Q), 1)$ by its value in tree approximation. The accuracy of (14.5.16) may be systematically improved by taking more terms in the perturbation expansions of β, γ_{NS}, γ_m, and C.

The singlet coefficients may be obtained by diagonalizing the 2×2 matrix of anomalous dimensions. Then there are two linear combinations of singlet coefficients that have simple behavior like (14.5.16).

14.6 OPE for vector and axial currents

We will now apply the operator-product expansion to the structure functions of T for a weak or electromagnetic current. The argument is a simple

generalization of the treatment in Sections 14.4 and 14.5 of deep-inelastic scattering with a scalar current. We have

$$\begin{aligned} T_{\mu\nu}(p,q) \sim & -2\mathrm{i}\int \mathrm{d}^4 y \mathrm{e}^{\mathrm{i}q\cdot y} \sum_{J,a} \langle p|\mathcal{O}^{Ja}_{\mu_1\ldots\mu_J}|p\rangle (-\mathrm{i})^J \times \\ & \times \left\{ y^{\mu_1}\ldots y^{\mu_J}\left(-g_{\mu\nu}\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial y^\mu \partial y^\nu}\right) C_1^{Ja}(y^2) \right. \\ & - y^{\mu_3}\ldots y^{\mu_J}\left(g^{\mu_1}_{\mu}\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial y^\mu \partial y_{\mu_1}}\right)\left(g^{\mu_2}_{\nu}\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial y^\nu \partial y_{\mu_2}}\right) C_2^{Ja}(y^2) \\ & \left. - \mathrm{i}y^{\mu_2}\ldots y^{\mu_J}\varepsilon_{\mu\nu}{}^{\mu_1\beta}\frac{\partial}{\partial y^\beta} C_3^{Ja}(y^2)\right\}. \end{aligned} \tag{14.6.1}$$

The derivatives acting on the C_i^{Ja}'s are arranged to be in combinations with zero divergence, to correspond to the condition $q^\mu T_{\mu\nu} = 0$. Fourier transformation as at (14.4.13), with suitable factors of Q^2, gives

$$\begin{aligned} T_{\mu\nu} \sim & -2\mathrm{i}\sum_{J,a} \langle p|\mathcal{O}^{Ja}|p\rangle \times \\ & \times \{(-g_{\mu\nu} + q_\mu q_\nu/q^2)x^{-J}\tilde{C}_1^{Ja}(Q) \\ & + (1/\nu)(p_\mu - q_\mu \nu/q^2)(p_\nu - q_\nu \nu/q^2)x^{1-J}\tilde{C}_2^{Ja}(Q) \\ & - \mathrm{i}\varepsilon_{\mu\nu\alpha\beta}p^\alpha q^\beta (1/2\nu)x^{-J}\tilde{C}_3^{Ja}(Q)\}. \end{aligned} \tag{14.6.2}$$

The set of leading twist operators is the same as in Section 14.4, and we have arranged normalizations so that the $\tilde{C}_i$'s are dimensionless in leading twist.

Hence the moments of the structure functions satisfy

$$\begin{aligned} F_{1,J}(Q) + (-1)^J \bar{F}_{1,J}(Q) &= \sum_a \tilde{C}_1^{Ja}(Q)\langle p|\mathcal{O}^{Ja}|p\rangle + \text{correction}, \\ F_{2,J-1}(Q) + (-1)^J \bar{F}_{2,J-1}(Q) &= \sum_a \tilde{C}_2^{Ja}(Q)\langle p|\mathcal{O}^{Ja}|p\rangle + \text{correction}, \\ F_{3,J}(Q) + (-1)^{J+1} \bar{F}_{3,J}(Q) &= \sum_a \tilde{C}_3^{Ja}(Q)\langle p|\mathcal{O}^{Ja}|p\rangle + \text{correction}. \end{aligned} \tag{14.6.3}$$

These equations are valid for $J > 1$, and the corrections are of order $1/Q^2$ times logarithms.

14.6.1 *Wilson coefficients – electromagnetic case*

To compute the lowest-order Wilson coefficient we consider deep-inelastic scattering on a quark target. The graphs are the same as in the case of a scalar current, Fig. 14.5.1, except that the current operator is now the elec-

tromagnetic current. We find

$$T_{\mu\nu}^{(\text{lowest order})} = \tfrac{1}{2}e_q^2\left\{\operatorname{tr}\left[\not{p}\gamma_\mu \frac{\mathrm{i}(\not{p}+\not{q})}{(p+q)^2}\gamma_\nu\right] + \operatorname{tr}\left[\not{p}\gamma_\nu \frac{\mathrm{i}(\not{p}-\not{q})}{(p-q)^2}\gamma_\mu\right]\right\}$$

$$= 2\mathrm{i}e_q^2\left\{(-g_{\mu\nu}+q_\mu q_\nu/q^2)\frac{1/x^2}{1/x^2-1}\right.$$

$$\left. + (p_\mu - q_\mu \nu/q^2)(p_\nu - q_\nu \nu/q^2)\frac{1}{\nu}\frac{2/x}{1/x^2-1}\right\}. \quad (14.6.4)$$

Expanding about $1/x = 0$ gives

$$T_1 = -2\mathrm{i}e_q^2 \sum_{n=0}^{\infty} (1/x)^{2n+2} + O(g^2),$$

$$\nu T_2/M^2 = -4\mathrm{i}e_q^2 \sum_{n=0}^{\infty} (1/x)^{2n+1} + O(g^2). \quad (14.6.5)$$

The Wilson coefficients are therefore

$$\left.\begin{aligned} \tilde{C}_1^{Jq} &= e_q^2/2 + O(g^2) \\ \tilde{C}_2^{Jq} &= e_q^2 + O(g^2) \end{aligned}\right\} \text{ if } J \text{ is even and } \geq 2,$$

$$\begin{aligned} C_3 &= 0, \\ \tilde{C}_i^{Jg} &= 0 + O(g^2). \end{aligned} \quad (14.6.6)$$

The relation $\tilde{C}_2 = 2\tilde{C}_1$ corresponds to the Callan–Gross relation $F_2 = 2xF_1$ in the parton model. Since the renormalization-group equation is the same for both $\tilde{C}_1^{Ja}$ and $\tilde{C}_2^{Ja}$, the Callan–Gross relation is true in QCD with corrections of $O(g(Q)^2)$. These corrections are from the Wilson coefficient, and have been calculated. See Buras (1981) for an up-to-date list of references.

The renormalization-group equations are the same as in the scalar case, except that γ_m is replaced by zero since the anomalous dimension of a conserved current is zero.

14.7 Parton interpretation of Wilson expansion

The use of moments in comparing theory and experiment is not very convenient, since cross-sections are needed outside the range in which they are measured. A more convenient form can be derived in which an expansion is obtained for the structure functions themselves. We will just summarize the results. More details can again be found in Buras (1981).

It is sufficient to examine the case of a scalar current for which the expansion is

$$F(x, Q) = \sum_a \int_x^1 \frac{\mathrm{d}z}{z} f_{a/N}(z, \mu) C_a(z/x; Q/\mu, g(\mu)). \quad (14.7.1)$$

The sum is over all species of parton, i.e., flavors of quark and antiquark, and gluon. The generalized Wilson coefficient C_a is effectively a structure function for deep-inelastic scattering of a parton state, while $f_{a/N}(z,\mu)$ is a parton distribution. Lowest-order calculations reproduce the parton-model result, so that

$$\begin{aligned} C_q &= C_{\bar{q}} = e_q{}^2\delta(z/x-1)/2, \\ C_g &= 0. \end{aligned} \tag{14.7.2}$$

The renormalization group equation for f has the form

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}f_{a/N}(x,\mu) = \sum_b \int_x^1 \frac{\mathrm{d}z}{z}\gamma_{ba}(z/x,g)f_{b/N}(z,\mu). \tag{14.7.3}$$

This is called the Altarelli–Parisi (1977) equation – it was first derived by these authors in leading logarithmic approximation from an heuristic argument. Later derivations (Collins & Soper (1982a), and Curci, Furmanski & Petronzio (1980)) are more complete.

Integro-differential equations like (14.7.3) are not particularly easy to work with. One mathematical simplification that can be made is to take moments, with the result that the operator-product expansion of Sections 14.5 and 14.6 is recovered. This will enable us to see that the two methods are essentially equivalent. However, the mathematically more complicated method using convolutions, as in (14.7.3), gives more physical insight, can be extended to other processes (Buras (1981)), and can be used without knowing structure functions at small x.

To see the equivalence of the two methods, let us define the following moments:

$$\begin{aligned} f_{a/N}^{(n)}(\mu) &= \int_0^1 \mathrm{d}z\, z^{n-1} f_{a/N}(z,\mu), \\ C_a^{(n)} &= \int_0^1 \mathrm{d}(x/z)(x/z)^{n-1} C_a(z,x;Q/\mu,g(\mu)), \\ \gamma_{ba}^{(n)} &= \int_0^1 \mathrm{d}(x/z)(x/z)^{n-1}\gamma_{ba}(z,x;g(\mu)). \end{aligned} \tag{14.7.4}$$

Then (14.7.1) implies

$$F_n(Q) = \sum_a f_{a/N}^{(n)}(\mu)C^{(n)}(g,Q/\mu). \tag{14.7.5}$$

This expansion has the same form as (14.4.14). In fact, moments of the f's are the same as the matrix elements of the twist-2 operators:

$$\langle N|\mathcal{O}^{Ja}|N\rangle = \int_0^1 \mathrm{d}x\, x^{J-1}[f_{a/N}(x)+f_{\bar{a}/N}(x)], \tag{14.7.6}$$

where there is a contribution from each type of parton and antiparton. This equation can be proved (Collins & Soper (1982a), and Curci, Furmanski & Petronzio (1980)) provided only that the same renormalization prescription (e.g., minimal subtraction) is used for both the operators and for the parton distributions.

14.8 W_4 and W_5

So far we have ignored F_4 and F_5. They are zero if the current j^μ in $W_{\mu\nu}$ is conserved. But the weak-interaction currents are not conserved if quark masses are non-zero:

$$\partial_\mu(\bar{\psi}\gamma^\mu(1-\gamma_5)\lambda^a\psi) = \frac{\mathrm{i}}{2}\bar{\psi}([m,\lambda^a]-\gamma_5\{m,\lambda^a\})\psi$$
$$\equiv \mathrm{i}D^a/2. \tag{14.8.1}$$

Here m is the quark mass matrix. It follows that W_4 and W_5 are non-zero in neutrino scattering. However, we should regard the ratio m/Q as setting the scale for their effect on the cross-section (Jaffe & Llewellyn-Smith (1973) and Llewellyn-Smith (1972)). They therefore give a small contribution to the deep-inelastic cross-section. Since W_4 and W_5 are therefore non-leading in the Bjorken limit, they are difficult to compute directly.

A convenient technique to compute W_4 and W_5 is to consider

$$q^\mu W_{\mu\nu} = (q^2 W_4 + p\cdot q W_5/2)q_\nu/M^2 + q^2 p_\nu W_5/2M^2. \tag{14.8.2}$$

We have the operator formula

$$q^\mu W_{\mu\nu} = \int \mathrm{d}^4 y \mathrm{e}^{\mathrm{i}q\cdot y}\langle p|j_\nu(y)D(0)|p\rangle, \tag{14.8.3}$$

so we can compute W_4 and W_5 by making an operator-product expansion for $j_\mu(y)D(0)$. Since there is an explicit factor of quark mass m in the expression for D, this operator behaves as a dimension 3 rather than a dimension 4 operator. This suppresses the Wilson coefficients for W_4 and W_5 by a power of Q^2. Since, in (14.1.10), the tensors multiplying W_4 and W_5 have a q_μ or q_ν in them, a similar suppression occurs because the lepton masses are much less than Q, as is seen by contracting q_μ or q_ν with the lepton tensor (14.1.7). The result is that W_4 and W_5 make a negligible contribution to the cross-section at large Q.

A detailed treatment of W_4 and W_5 within QCD can be made but has not yet been published. It generalizes the results of Jaffe & Llewellyn-Smith (1973), who worked within the parton model.

References

Adler, S. (1969). *Phys. Rev.* **177**, 2426.
Adler, S. (1970). In *Lectures on Elementary Particles and Quantum Field Theory*, eds. S. Deser *et al.* (M.I.T., Cambridge, Mass.).
Adler, S. & Bardeen, W. (1969). *Phys. Rev.* **182**, 1517.
Aizenman, M. (1981). *Phys. Rev. Lett.* **47**, 1.
Altarelli, G. & Parisi, G. (1977). *Nucl. Phys.* **B126**, 298.
Amati, D., Petronzio, R. & Veneziano, G. (1978). *Nucl. Phys.* **B146**, 29.
Appelquist, T. & Carazzone, J. (1975). *Phys. Rev.* **D11**, 2856.
Ashmore, J. F. (1972). *Nuovo Cimento Lett.* **4**, 289.
Bandelloni, G., Becchi, C., Blasi, A. & Collina, R. (1980). *Comm. Math. Phys.* **72**, 239.
Bardeen, W. A. (1972). In *Proceedings XVI International Conference on High Energy Physics* (National Accelerator Laboratory, Batavia, IL).
Bardeen, W. A., Buras, A. J., Duke, D. W. & Muta, T. (1978). *Phys. Rev.* **D18**, 3998.
Baym, G. (1960). *Phys. Rev.* **117**, 886.
Becchi, C., Rouet, A. & Stora, R. (1975). *Comm. Math. Phys.* **42**, 127.
Becchi, C., Rouet, A. & Stora, R. (1976). *Ann. Phys. (N.Y.)* **98**, 287.
Beg, M. A. B. & Sirlin, A. (1982). *Phys. Rep.* **88**, 1.
Bell, J. S. & Jackiw, R. (1969). *Nuovo Cimento* **51**, 47.
Bernard, C. (1974). *Phys. Rev.* **D9**, 3312.
Bjorken, J. D. (1982). *Physica Scripta* **25**, 69.
Bjorken, J. D. & Drell, S. D. (1966). *Relativistic Quantum Fields* (McGraw-Hill. New York).
Bjorken, J. D. & Paschos, E. A. (1969). *Phys. Rev.* **185**, 1975.
Bogoliubov, N. N. & Parasiuk, O. (1957). *Acta Math.* **97**, 227.
Bogoliubov, N. N. & Shirkov, D. V. (1955/6). *Usp. Fiz. Nauk* **55**, 149 (1955). (German trans.: *Fortschr. Phys.* **4**, 438 (1956).)
Bogoliubov, N. N. & Shirkov, D. V. (1980). *Introduction to the Theory of Quantized Fields* (3rd edn) (Wiley, New York).
Bollini, C. G. & Giambiagi, J. J. (1972). *Phys. Lett.* **40B**, 566.
Brandt, R. A. (1967). *Ann. Phys. (N.Y.)* **44**, 221.
Brandt, R. A. (1976). *Nucl. Phys.* **B116**, 413.
Breitenlohner, P. & Maison, D. (1977a). *Comm. Math. Phys.* **52**, 11.
Breitenlohner, P. & Maison, D. (1977b). *Comm. Math. Phys.* **52**, 39.
Breitenlohner, P. & Maison, D. (1977c). *Comm. Math. Phys.* **52**, 55.
Brown, L. S. (1980). *Ann. Phys. (N.Y.)* **126**, 135.
Brown, L. S. & Collins, J. C. (1980). *Ann. Phys. (N.Y.)* **130**, 215.
Buras, A. J. (1981). In *Proceedings of Lepton–Photon Conference* (Bonn).
Buras, A. J., Floratos, E. G., Ross, D. A. & Sachrajda, C. T. (1977). *Nucl. Phys.* **B131**, 308.

Callan, C. G. (1970). *Phys. Rev.* D**2**, 1541.
Callan, C. G., Coleman, S. & Jackiw, R. (1970). *Ann. Phys. (N.Y.)* **59**, 42.
Callan, C. G. & Gross, D. (1969). *Phys. Rev. Lett.* **22**, 156.
Caswell, W. E. & Kennedy, A. D. (1982). *Phys. Rev.* D**25**, 392.
Caswell, W. E. & Zanon, D. (1981). *Nucl. Phys.* B**182**, 125.
Celmaster, W. & Gonsalves, R. J. (1979). *Phys. Rev.* D**20**, 1420.
Chanowitz, M., Furman, M. & Hinchliffe, I. (1979). *Nucl. Phys.* B**159**, 225.
Cheng, T. P., Eichten, E. & Li, L.F. (1974). *Phys. Rev.* D**9**, 2259.
Chetyrkin, K. G., Kataev, A. L. & Tkachov, F. V. (1981). *Phys. Lett.* **99B**, 147, and erratum.
Chetyrkin, K. G. & Tkachov, F. V. (1981). *Nucl. Phys.* B**192**, 159.
Christ, N., Hasslacher, B. & Mueller, A. (1972). *Phys. Rev.* D**6**, 3543.
Cicuta, G. M. & Montaldi, E. (1972). *Nuovo Cimento Lett.* **4**, 329.
Clark, T. E., Piguet, O. & Sibold, K. (1979). *Nucl. Phys.* B**159**, 1.
Clark, T. E., Piguet, O. & Sibold, K. (1980). *Nucl. Phys.* B**172**, 201.
Coleman, S. (1973). *Comm. Math. Phys.* **31**, 259.
Coleman, S. & Gross, D. J. (1973). *Phys. Rev. Lett.* **31**, 851.
Coleman, S. & Jackiw, R. (1971). *Ann. Phys. (N.Y.)* **67**, 552.
Coleman, S. & Weinberg, E. (1973). *Phys. Rev.* D**7**, 1888.
Collins, J. C. (1974). *Nucl. Phys.* B**80**, 341.
Collins, J. C. (1975a). PhD. Thesis: Cambridge University.
Collins, J. C. (1975b). *Nucl. Phys.* B**92**, 477.
Collins, J. C. (1976). *Phys. Rev.* D**14**, 1965.
Collins, J. C. (1980). *Phys. Rev.* D**22**, 1478.
Collins, J. C. & Soper, D. E. (1981). *Nucl. Phys.* B**193**, 381.
Collins, J. C. & Soper, D. E. (1982a). *Nucl. Phys.* B**194**, 445.
Collins, J. C. & Soper, D. E. (1982b). *Nucl. Phys.* B**197**, 446.
Collins, J. C., Wilczek, F. & Zee, A. (1978). *Phys. Rev.* D**18**, 242.
Cornwall, J. M. & Norton, R. E. (1973). *Phys. Rev.* D**8**, 3338.
Costa, G., Marinucci, T., Tonin, M. & Julve, J. (1977). *Nuovo Cimento* **38A**, 373.
Creutz, M. (1980). *Phys. Rev.* D**21**, 2308.
Creutz, M. (1983). *Quarks, Gluons and Lattices* (Cambridge University Press).
Creutz, M. & Moriarty, K. J. M. (1982). *Phys. Rev.* D**26**, 2166.
Curci, G., Furmanski, W. & Petronzio, R. (1980). *Nucl. Phys.* B**175**, 27.
Cvitanovic, P. (1977). *Nucl. Phys.* B**130**, 114.
Dirac, P. A. M. (1981). In *Perturbative Quantum Chromodynamics* (Tallahassee, 1981), ed. Duke, D. W. & Owens, J. F. (AIP, New York).
Dixon, J. A. & Taylor, J. C. (1974). *Nucl. Phys.* B**78**, 552.
Drell, S. D., Levy, D. & Yan, T. M. (1969). *Phys. Rev.* **187**, 2159.
Drell, S. D., Levy, D. & Yan, T. M. (1970a). *Phys. Rev.* D**1**, 1035.
Drell, S. D., Levy, D. & Yan, T. M. (1970b). *Phys. Rev.* D**1**, 1617.
Drell, S. D., Levy, D. & Yan, T. M. (1970c). *Phys. Rev.* D**1**, 2404.
Dyson, F. J. (1952). *Phys. Rev.* **85**, 631.
Ellis, R. K., Georgi, H., Machacek, M., Politzer, H. D. & Ross, G. G. (1979). *Nucl. Phys.* B**152**, 285.
Faddeev, L. D. & Popov, V. M. (1967). *Phys. Lett.* **25B**, 29.
Fayet, P. & Ferrara, S. (1977). *Phys. Rep.* **32**, 249.
Fetter, A. L. & Walecka, J. D. (1971). *Quantum Theory of Many-Body Systems* (McGraw-Hill, New York).
Feynman, R. P. (1972). *Photon–Hadron Interactions* (Benjamin, New York).
Fisher, M. E. & Gaunt, D. S. (1964). *Phys. Rev.* **133**, A224.
Freedman, D. Z., Muzinich, I. J. & Weinberg, E. J. (1974) *Ann. Phys. (N.Y.)* **87**, 95.

Fujikawa, K., Lee, B. W. & Sanda, A. I. (1972). *Phys. Rev.* **D6**, 2923.
Gasser, J. & Leutwyler, H. (1982). *Phys. Rep.* **87**, 77.
Gastmans, R. & Meuldermans, R. (1973). *Nucl. Phys.* **B63**, 277.
Gastmans, R., Verwaest, J. & Meuldermans, R. (1976). *Nucl. Phys.* **B105**, 454.
Gell-Mann, M. & Low, F. E. (1951). *Phys. Rev.* **84**, 350.
Gell-Mann, M. & Low, F. E. (1954). *Phys. Rev.* **95**, 1300.
Georgi, H. & Politzer, H. D. (1976). *Phys. Rev.* **D14**, 1829.
Gildener, E. & Weinberg, S. (1976). *Phys. Rev.* **D13**, 3333.
Glimm, J. & Jaffe, A. (1981). *Quantum Physics: A Functional Integral Point of View* (Springer-Verlag, New York).
Goldstone, J., Salam, A. & Weinberg, S. (1962). *Phys. Rev.* **127**, 965.
Gross, D. (1976). In *Methods in Field Theory*, ed. Balian, R. & Zinn-Justin, J. (North-Holland, Amsterdam).
Gross, D. & Jackiw, R. (1972). *Phys. Rev.* **D6**, 477.
Gross, D. & Neveu, A. (1974). *Phys. Rev.* **D10**, 3235.
Hahn, Y. & Zimmermann, W. (1968). *Comm. Math. Phys.* **10**, 330.
Hasenfratz, A. & Hasenfratz, P. (1980). *Phys. Lett.* **93B**, 165.
Hawking, S. W. & Israel, W. (1979). *General Relativity* (Cambridge University Press).
Hepp, K. (1966). *Comm. Math. Phys.* **2**, 301.
Hohenberg, P. (1967). *Phys. Rev.* **158**, 383.
Itzykson, C. & Zuber, J.-B. (1980). *Quantum Field Theory* (McGraw-Hill, New York).
Jackiw, R. (1968). *Ann. Phys. (N.Y.)* **48**, 292.
Jackiw, R. & Johnson, K. (1973). *Phys. Rev.* **D8**, 2386.
Jaffe, R. & Llewellyn-Smith, C. H. (1973). *Phys. Rev.* **D7**, 2506.
Joglekar, S. (1976). *Ann. Phys. (N.Y.)* **100**, 395.
Joglekar, S. (1977a). *Ann. Phys. (N.Y.)* **108**, 233.
Joglekar, S. (1977b). *Ann. Phys. (N.Y.)* **109**, 210.
Joglekar, S. & Lee, B. W. (1976). *Ann. Phys. (N.Y.)* **97**, 160.
Jones, D. R. T. & Leveille, J. P. (1982). *Nucl. Phys.* **B206**, 473.
Kazama, Y. & Yao, Y.-P. (1982). *Phys. Rev.* **D25**, 1605.
Kluberg-Stern, H. & Zuber, J. B. (1975). *Phys. Rev.* **D12**, 467.
Kogut, J. (1983). *Rev. Mod. Phys.* **55**, 775.
Korthals Altes, C. P. & Perrottet, M. (1972). *Phys. Lett.* **39B**, 546.
Lam, C. S. & Tung, W.-K. (1978). *Phys. Rev.* **D18**, 2447.
Lam, Y.-M. P. (1972). *Phys. Rev.* **D6**, 2145.
Landau, L. D. & Pomeranchuk, I. (1955). *Dokl. Akad. Nauk USSR* **102**, 489.
Langacker, P. (1981). *Phys. Rep.* **72**, 185.
LaRue, G. S., Phillips, J. D. & Fairbank, W. M. (1981). *Phys. Rev. Lett.* **46**, 967
Lautrup, B. (1976). *Nucl. Phys.* **B105**, 23.
Lee, B. W. (1976), In *Methods in Field Theory*, ed. Balian, R. & Zinn-Justin, J. (North-Holland, Amsterdam).
Lee, B. W. & Zinn-Justin, J. (1972). *Phys. Rev.* **D5**, 3121.
Libby, S. & Sterman, G. (1978). *Phys. Rev.* **D18**, 3252.
Llewellyn-Smith, C. H. (1972). *Phys. Rep.* **3**, 261.
Llewellyn-Smith, C. H. (1980). In *Quantum Flavordynamics, Quantum Chromodynamics, and Unified Theories*, ed. Mahanthappa, K. T. & Randa, J. (Plenum, New York).
Lowenstein, J. H. (1971). *Phys. Rev.* **D4**, 2281.
Lowenstein, J. H., Weinstein, M. & Zimmermann, W. (1974a). *Phys. Rev.* **D10**, 1854.

Lowenstein, J. H., Weinstein, M. & Zimmermann, W. (1974b). *Phys. Rev.* **D10**, 2500.
Lurié, D. (1968). *Particles and Fields* (Interscience, New York).
Macfarlane, A. J. & Woo, G. (1974). *Nucl. Phys.* **B77**, 91; **B86**, 548(E).
Mack, G. (1972). Kaiserslautern Lectures. *Springer Lecture Notes in Physics*, vol. 17, ed. Ruhl, W. & Vancura, A.
Marciano, W. & Pagels, H. (1978). *Phys. Rep.* **36**, 138.
Marciano, W. & Sirlin, A. (1975). *Nucl. Phys.* **B88**, 86.
Mermin, N. D. & Wagner, H. (1966). *Phys. Rev. Lett.* **17**, 1133.
Mueller, A. H. (1978). *Phys. Rev.* **D18**, 3705.
Mueller, A. H. (1979). *Phys. Rev.* **D20**, 2037.
Mueller, A. H. (1981). *Phys. Rep.* **73**, 237.
Oppenheimer, J. R. (1930). *Phys. Rev.* **35**, 461.
Ovrut, B. A. & Schnitzer, H. J. (1980). *Phys. Rev.* **D21**, 3369.
Pauli, W. & Villars, F. (1949). *Rev. Mod. Phys.* **21**, 434.
Piguet, O. & Rouet, A. (1981). *Phys. Rep.* **76**, 1.
Piguet, O. & Sibold, K. (1982a). *Nucl. Phys.* **B197**, 247.
Piguet, O. & Sibold, K. (1982b). *Nucl. Phys.* **B197**, 257.
Piguet, O. & Sibold, K. (1982c). *Nucl. Phys.* **B197**, 272.
Ramond, P. (1981). *Field Theory: A Modern Primer* (Benjamin, New York).
Ross, G. G. (1981). *Rep. Prog. Phys.* **44**, 655.
Sarkar, S. (1974). *Nucl. Phys.* **B83**, 108.
Schwinger, J. (1951). *Phys. Rev.* **82**, 664.
Shaw, R. (1955). Cambridge University Thesis.
Slavnov, A. (1972). *Theor. Math. Phys.* **10**, 153 (Russian); **10**, 99 (English translation).
Speer, E. R. (1974). *J. Math. Phys.* **15**, 1.
Speer, E. R. & Westwater, M. J. (1971). *Ann. Inst. Henri Poincaré* **A14**, 1.
Stirling, W. J. (1978). *Nucl. Phys.* **B145**, 477.
Streater, R. F. & Wightman, A. S. (1978). *PCT, Spin and Statistics, and All That* (Benjamin, New York).
Stueckelberg, E. C. G. & Petermann, A. (1953). *Helv. Phys. Acta* **26**, 499.
Sudakov, V. (1956). *Zh. Eksp. Teor. Fiz.* **30**, 87 (English trans.: *Sov. Phys. JETP* **3**, 65).
Symanzik, K. (1970a). *Comm. Math. Phys.* **16**, 48.
Symanzik, K. (1970b). *Comm. Math. Phys.* **18**, 227.
Symanzik, K. (1973). *Comm. Math. Phys.* **34**, 7.
Symanzik, K. (1982). In *Proceedings* XXII *International Conference on High Energy Physics*, ed. Petian, P. & Porneuf, M. (Editions de Physique, Les Ulis, France).
Tarasov, O. V., Vladimirov, A. A. & Zharkov, A. Yu. (1980). *Phys. Lett.* **93B**, 429.
Taylor, J. C. (1971). *Nucl. Phys.* **B33**, 436.
't Hooft, G. (1971a). *Nucl. Phys.* **B33**, 173.
't Hooft, G. (1971b). *Nucl. Phys.* **B35**, 167.
't Hooft, G. (1973). *Nucl. Phys.* **B61**, 455.
't Hooft, G. & Veltman M. (1972a). *Nucl. Phys.* **B44**, 189.
't Hooft, G. & Veltman M. (1972b). *Nucl. Phys.* **B50**, 318.
Tkachov, F. V. (1981). *Phys. Lett.* **100B**, 65.
Toussaint, D. (1978). *Phys. Rev.* **D18**, 1628.
Treiman, S. B., Jackiw, R. & Gross, D. J. (1972). *Lectures on Current Algebra and Its Applications* (Princeton University Press).

Veltman, M. (1977). *Nucl. Phys.* **B123**, 89.
Vladimirov, A. A., Kazakov, D. I. & Tarasov, O. V. (1979). *Zh. Eksp. Teor. Fiz.* **77**, 1035 (English trans.: *Sov. Phys. JETP* **50**, 521).
Ward, J. C. (1950). *Phys. Rev.* **78**, 182.
Weinberg, S. (1960). *Phys. Rev.* **118**, 838.
Weinberg, S. (1973). *Phys. Rev.* **D8**, 3497.
Weinberg, S. (1974). In *Proceedings XVII International Conference on High Energy Physics* (London, 1974), ed. Smith, J. R. (Rutherford Lab., Chilton, Didcot, Berkshire, England).
Weinberg, S. (1976). In *Gauge Theories and Modern Field Theory*, eds. Arnowitt, R. & Pran Nath (M.I.T., Cambridge, Mass.).
Weinberg, S. (1980). *Phys. Lett.* **91B**, 51.
Wilson, K. G. (1969). *Phys. Rev.* **179**, 1699.
Wilson, K. G. (1973). *Phys. Rev.* **D7**, 2911.
Wilson, K. G. (1974). *Phys. Rev.* **D10**, 2445.
Wilson, K. G. & Fisher, M. E. (1972). *Phys. Rev. Lett.* **28**, 240.
Wilson, K. G. & Kogut, J. (1974). *Phys. Rep.* **12**, 75.
Witten, E. (1976). *Nucl. Phys.* **B104**, 445.
Yang, C. N. & Mills, R. (1954). *Phys. Rev.* **96**, 191.
Yao, Y. P. (1973). *Phys. Rev.* **D7**, 1647.
Zee, A. (1972). *Phys. Rev. Lett.* **29**, 1198.
Zimmermann, W. (1968). *Comm. Math. Phys.* **11**, 1.
Zimmermann, W. (1969). *Comm. Math. Phys.* **15**, 208.
Zimmermann, W. (1970). In *Lectures on Elementary Particles and Quantum Field Theory*, eds. S. Deser *et al.* (M.I.T., Cambridge, Mass).
Zimmermann, W. (1973a). *Ann. Phys. (N.Y.)* **77**, 536.
Zimmermann, W. (1973b). *Ann. Phys. (N.Y.)* **77**, 570.

Index

ε-expansion 62
ϕ^3 theory 36
  one-loop calculations 38
Λ parameter of QCD 188

Adler–Bardeen theorem 338
Altarelli–Parisi equation 370
anomalies 331
  Adler–Bardeen theorem 338
  not in non-singlet axial current 346
  one-loop calculation 342
  triangle 349
anomalously broken symmetries 18, 331
asymptotic freedom 187

bare parameters
  definition 10
  renormalization group calculation 209
    dimensional regularization 217
basic graph 90
basic Lagrangian 89
  definition 11
BPH 53
  see also renormalization prescriptions
BPHZ 53, 133
  see also renormalization prescriptions
broken symmetries 18
  anomalous 18
  dynamically 19
  explicitly 18
  spontaneously 18
    not at $d = 2$ 26
    perturbation theory 26, 247
    renormalization 247, 314
BRS, *see* gauge theories

canonical momentum field 5
causality 6
charge conjugation symmetry, for gauge theories 35
chiral symmetry 17, 333
  and anomalies 338
commutation relations
  canonical equal-time 5
  from Green's functions 6
composite operators 138
  definition 146
  equations of motion 155, 158
  mixing under renormalization 149
  properties 152
  renormalization, example 142
  tensor operators 150
  twist 363
  Ward identities 160
  see also operator product expansion
continuum limit 8
  see also renormalization
coordinate space, renormalization in 59, 277
cosmological term, and renormalization 118
counterterm Lagrangian 11, 89
cut-off
  dimensional regularization 13, 53, 62
    for infra-red 56
  lattice 8
  Pauli–Villars 12
  see also individual entries

decoupling theorem 222
  examples 226
  exceptions 223
  manifest decoupling 230
  proof 233
  renormalization group 240
deep inelastic scattering 354
  Altarelli–Parisi equation 370
  calculation of Wilson coefficients 364
  operator product expansion 361
  partons and operator product expansion 369
  W_4 and W_5 357, 371
degree of divergence 45
differentiation with respect to parameters in Lagrangian 163
dimensional regularization 13, 53, 62
  axioms 64
  definition 67
    convergence 68
  Dirac matrices in 83, 334
  list of integrals 81
  properties 73

reduced to ordinary integration 67
uniqueness 65
dimensional transmutation 203
Dirac matrices
γ_5 at $d \neq 4$ 86, 334
definition in d dimensions 83
dynamically broken symmetries 19

equations of motion
Green's function 6,13
momentum space 15
operator 5
Euclidean space-time 8.
evanescent operators 346
explicit symmetry breaking 18

factorization theorems 1, 187
Faddeev–Popov ghost, *see* gauge theories
Feynman rules
for gauge theories 34
see also perturbation theory
fixed points 190
forest formula 109
3-loop example 106
functional integral
for bosons 6
for fermions 27
Monte-Carlo calculation 13

gauge dependence of counterterms 309, 327
gauge invariant operators 316
gauge theories 28
BRS invariance 32
charge conjugation symmetry 35
Faddeev–Popov ghost 30
Feynman rules 34
gauge invariance v. gauge independence 31
gauge-fixing 30
operator product expansion 321
proof of renormalizability 298
quantization 29
renormalization 293
renormalization group 319
Slavnov–Taylor identities 32
Gell-Mann–Low formula 22
General Relativity
as example of gauge theory 17
cosmological term and renormalization 118
generating functional
for Green's functions 7
global symmetries
renormalization 244
spontaneously broken, invariance of counterterms 247
unbroken, invariance of counterterms 245
Goldstone bosons, *see* Nambu–Goldstone bosons
Goldstone model 247
Grassmann algebra 27
Green's functions
definition 6
equations of motion 13
functional integral 7
generating functional 7
momentum space 15

Hamiltonian, field theory 5

infra-red behavior, and renormalization group 187, 191

Landau ghost 196
large mass expansion 222
large momentum behavior of Green's functions 185
lattice 8
leading logarithms 193
non-renormalization group 196
list of integrals 81
low-energy behavior 191

mass-shell renormalization prescription 48
massive photon, renormalization 324
minimal subtraction 56, 135
see also renormalization prescriptions
Monte-Carlo 13

Nambu–Goldstone bosons 21, 252
Noether's theorem 18
non-renormalizability 49
in real physics 122
non-renormalization of currents 162
normal products, *see* composite operators

operator product expansion 257
and parton model 369
and renormalization group 274
deep inelastic scattering 361
examples 139, 258
gauge theories 321
in coordinate- and momentum-space 257
proof 266
to define composite operators 258
overlapping divergence 99
oversubtraction 130
on 1PR graphs 131

partition function, functional integral as 8
parton model 355, 358, 369
Pauli–Villars cut-off 12

perturbation theory 21
as saddle point expansion 26
Feynman rules for gauge theories 34
maximum accuracy with asymptotic freedom 189
not convergent 24, 189
perturbative QCD 1
Poincaré symmetry 17
propagator 22
short-distance behavior 278
thermal 280

QCD
Λ parameter 188
asymptotic freedom 188
Feynman rules 34
perturbative 1
see also gauge theories
QED
gauge dependence of counterterms 327
in external field 291
renormalization 293, 322
renormalization group 329
Ward identities 326
see also gauge theories
quantum chromodynamics, *see* QCD
quantum electrodynamics, *see* QED
quantum field theory
analog of statistical mechanics 8
axioms 6
review 4
standard properties 6
standard theories 36
see also,
ϕ^3 theory; QCD; QED
quantum mechanics
as field theory at $d = 1$ 6
R_ξ-gauge 314
reconstruction theorem 6
regulator, *see* cut-off
renormalizability 49, 116
renormalization 9, 88
and continuum limit 9
axial current 340
coordinate space 59, 277
examples 281
examples 38
for massless theories 59
forest formula 109
gauge dependence of counterterms 309, 327
gauge invariant operators 316
gauge theories 293
with symmetry breaking 308, 314
in external field 291
locality of counterterms 97, 125
non-zero temperature 286
oversubtraction 130
prescriptions 52
QED 293, 322
with photon mass 323
renormalization group 50, 168
Λ parameter 188
asymptotic freedom 187
change of renormalization prescription 200
coefficients 180
composite operators 219
computation of bare parameters 209
dimensional regularization 217
cut-off procedure, dependence on 206
decoupling theorem 240
dimensional transmutation 203
equation 183
examples 50, 172
fixed points 190
gauge theories 319
infra-red behavior 187, 191
Landau ghost 196
large momentum behavior 185
leading logarithms 193
minimal subtraction 56, 181
several couplings 197
solution of RG equation 184
renormalization prescriptions 52
BPH 53
BPHZ 53, 133
mass-shell 48
minimal subtraction 135
at one loop 56
spontaneously broken symmetries 253
renormalized field 10
renormalized mass, and physical mass 51

Slavnov–Taylor identities 32
spontaneously broken symmetries 18
absence of IR divergences 252
choice of renormalization prescription 253
renormalization 247
renormalization of current 251
super-renormalizability, definition 118
supersymmetries 17, 339
symmetries 15, 244
BRS, *see* gauge theories
charge conjugation, *see* gauge theories
gauge 28
Ward identities 19, 160
see also broken symmetries; global symmetries

thermodynamic limit of functional integral 8
triangle anomaly 349

twist 363

ultra-violet regulators 12
  see also cut-off; dimensional regularization
unitary gauge 324, 330

Ward identities 16, 19, 160
  axial current 339, 349
  QED 326
  renormalization 246–256
Weinberg's theorem 125
Wick rotation 8, 39
Wightman axioms 6
Wilson expansion, *see* operator product expansion

Yang–Mills theory, *see* gauge theories