Where is the minimum of this potential? If we take this expression at face value, we find that $V_{\text{eff}}(\phi_{\text{cl}})$ passes through zero when ϕ_{cl} reaches the very small value

$$\phi_{\text{cl}}^2 = e^{3/2}\frac{M^2}{\lambda}\cdot\exp\left[-\frac{(4\pi)^2}{(N+8)\lambda}\right],$$

and, near this point, attains a minimum with a nonzero value of ϕ_{cl}. But the zero occurs by the cancellation of the leading term against the quantum correction. In other words, perturbation theory breaks down completely before we can address the question of whether $V_{\text{eff}}(\phi_{\text{cl}})$, for $\mu^2 = 0$, has a symmetry-breaking minimum. It seems that our present tools are quite inadequate to resolve this case.

Although it is far from obvious, these two problems turn out to be related to each other. One of our major results in Chapter 12 will be an explanation of the interrelation of M^2, λ, and μ^2 displayed in Eq. (11.80). Then, in Chapter 13, we will use the insight we have gained from this analysis to solve completely the second problem of the appearance of large logarithms. Before beginning that study, however, there are a few issues we have yet to discuss in the more formal aspects of the renormalization of theories with spontaneously broken symmetry.

11.5 The Effective Action as a Generating Functional

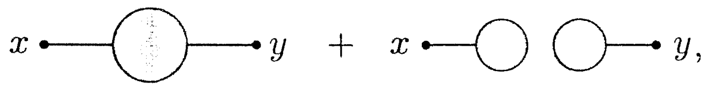
Now that we have defined the effective action and computed it for one particular theory, let us return to our goal of understanding the renormalization of theories with hidden symmetry. In Section 11.6 we will use the effective action as a tool in achieving this goal. First, however, we must investigate in more detail the relation between the effective action and Feynman diagrams.

We saw in Section 9.2 that the functional derivatives of $Z[J]$ with respect to $J(x)$ produce the correlation functions of the scalar field (see, for example, Eq. (9.35)). In other words, $Z[J]$ is the *generating functional* of correlation functions. Our goal now is to show that $\Gamma[\phi_{\text{cl}}]$ is also such a generating functional; specifically, it is the generating functional of one-particle-irreducible (1PI) correlation functions. Since the 1PI correlation functions figure prominently in the theory of renormalization, this result will be central in the discussion of renormalization in the following section.

To begin, let us consider the functional derivatives not of $\Gamma[\phi_{\text{cl}}]$, but of $E[J] = i\log Z[J]$. The first derivative, given in Eq. (11.44), is precisely $-\langle\phi(x)\rangle$. The second derivative is

$$\begin{aligned}\frac{\delta^2 E[J]}{\delta J(x)\delta J(y)} &= -\frac{i}{Z}\int\mathcal{D}\phi\, e^{i\int(\mathcal{L}+J\phi)}\phi(x)\phi(y)\\ &\qquad + \frac{i}{Z^2}\int\mathcal{D}\phi\, e^{i\int(\mathcal{L}+J\phi)}\phi(x)\cdot\int\mathcal{D}\phi\, e^{i\int(\mathcal{L}+J\phi)}\phi(y)\\ &= -i\Big[\langle\phi(x)\phi(y)\rangle - \langle\phi(x)\rangle\langle\phi(y)\rangle\Big].\end{aligned}\tag{11.82}$$

If we were to compute the term $\langle\phi(x)\phi(y)\rangle$ from Feynman diagrams, there would be two types of contributions:



(11.83)

where each circle corresponds to a sum of *connected* diagrams. The second term in the last line of Eq. (11.82) cancels the second, disconnected, term of (11.83). Thus the second derivative of $E[J]$ contains only those contributions to $\langle\phi(x)\phi(y)\rangle$ that come from connected Feynman diagrams. Let us call this object the *connected correlator*:

$$\frac{\delta^2 E[J]}{\delta J(x)\delta J(y)} = -i\,\langle\phi(x)\phi(y)\rangle_{\text{conn}}\,. \tag{11.84}$$

Similarly, the third functional derivative of $E[J]$ is

$$\begin{aligned}\frac{\delta^3 E[J]}{\delta J(x)\delta J(y)\delta J(z)} &= \Big[\langle\phi(x)\phi(y)\phi(z)\rangle - \langle\phi(x)\phi(y)\rangle\langle\phi(z)\rangle - \langle\phi(x)\phi(z)\rangle\langle\phi(y)\rangle \\ &\qquad - \langle\phi(y)\phi(z)\rangle\langle\phi(x)\rangle + 2\langle\phi(x)\rangle\langle\phi(y)\rangle\langle\phi(z)\rangle\Big] \\ &= \langle\phi(x)\phi(y)\phi(z)\rangle_{\text{conn}}\,.\end{aligned} \tag{11.85}$$

In each successive derivative of $E[J]$ all contributions cancel except for those from fully connected diagrams. The general formula for n derivatives is

$$\frac{\delta^n E[J]}{\delta J(x_1)\cdots\delta J(x_n)} = (i)^{n+1}\,\langle\phi(x_1)\cdots\phi(x_n)\rangle_{\text{conn}}\,. \tag{11.86}$$

We therefore refer to $E[J]$ as the *generating functional of connected correlation functions.*

So much for $E[J]$. Now what about the functional derivatives of the effective action? Consider first the derivative of Eq. (11.48) with respect to $J(y)$:

$$\frac{\delta}{\delta J(y)}\frac{\delta\Gamma}{\delta\phi_{\text{cl}}(x)} = -\delta(x-y).$$

We can rewrite the left-hand side of this equation using the chain rule, to obtain

$$\begin{aligned}\delta(x-y) &= -\int d^4z\,\frac{\delta\phi_{\text{cl}}(z)}{\delta J(y)}\,\frac{\delta^2\Gamma}{\delta\phi_{\text{cl}}(z)\delta\phi_{\text{cl}}(x)} \\ &= \int d^4z\,\frac{\delta^2 E}{\delta J(y)\delta J(z)}\,\frac{\delta^2\Gamma}{\delta\phi_{\text{cl}}(z)\delta\phi_{\text{cl}}(x)} \\ &= \left(\frac{\delta^2 E}{\delta J\delta J}\right)_{yz}\left(\frac{\delta^2\Gamma}{\delta\phi_{\text{cl}}\delta\phi_{\text{cl}}}\right)_{zx}.\end{aligned} \tag{11.87}$$

In the second line we have used Eq. (11.44). The last line is an abstract representation of the second line, where we think of each of the second derivatives as an infinite-dimensional matrix, with the integral over z represented by matrix multiplication. What we have shown is that these two matrices are inverses of each other:

$$\left(\frac{\delta^2 E}{\delta J \delta J}\right) = \left(\frac{\delta^2 \Gamma}{\delta \phi_{\text{cl}} \delta \phi_{\text{cl}}}\right)^{-1}. \tag{11.88}$$

Now according to Eq. (11.84), the first of these matrices is $-i$ times the conected two-point function, that is, the exact propagator of the field ϕ. Let us call this propagator $D(x,y)$:

$$\left(\frac{\delta^2 E}{\delta J(x) \delta J(y)}\right) = -i \left\langle \phi(x) \phi(y) \right\rangle_{\text{conn}} \equiv -iD(x,y). \tag{11.89}$$

We will therefore refer to the other matrix (times $-i$) as the *inverse propagator*:

$$\left(\frac{\delta^2 \Gamma}{\delta \phi_{\text{cl}}(x) \delta \phi_{\text{cl}}(y)}\right) = iD^{-1}(x,y). \tag{11.90}$$

This provides an interpretation, of sorts, for the second functional derivative of the effective action. This interpretation becomes more concrete if we go to momentum space. On a translation-invariant vacuum state (one with ϕ_{cl} constant), the matrix $D(x,y)$ must be diagonal in momentum:

$$D(x,y) = \int \frac{d^4p}{(2\pi)^4} e^{-ip\cdot(x-y)} \widetilde{D}(p). \tag{11.91}$$

We showed in Eq. (7.43) that the momentum-space propagator $\widetilde{D}(p)$ is a geometric series in one-particle-irreducible Feynman diagrams. The Fourier transform of $D^{-1}(x,y)$ then gives the inverse propagator:

$$\widetilde{D}^{-1}(p) = -i(p^2 - m^2 - M^2(p^2)), \tag{11.92}$$

where $M^2(p)$ is the sum of one-particle-irreducible two-point diagrams.

To evaluate higher derivatives of the effective action we again use the chain rule,

$$\frac{\delta}{\delta J(z)} = \int d^4w \, \frac{\delta \phi_{\text{cl}}(w)}{\delta J(z)} \frac{\delta}{\delta \phi_{\text{cl}}(w)} = i \int d^4w \, D(z,w) \frac{\delta}{\delta \phi_{\text{cl}}(w)}, \tag{11.93}$$

together with the standard rule for differentiating matrix inverses:

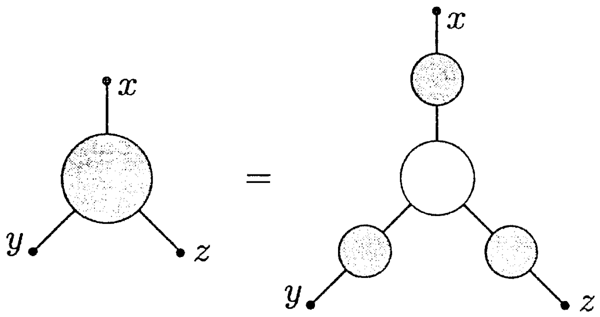
$$\frac{\partial}{\partial \alpha} M^{-1}(\alpha) = -M^{-1} \frac{\partial M}{\partial \alpha} M^{-1}. \tag{11.94}$$

Applying these identities to Eq. (11.88), we find (with some abbreviated notation)

$$\begin{aligned}\frac{\delta^3 E[J]}{\delta J_x \delta J_y \delta J_z} &= i \int d^4w \, D(z,w) \, \frac{\delta}{\delta \phi^{\text{cl}}_w} \left(\frac{\delta^2 \Gamma}{\delta \phi^{\text{cl}}_x \delta \phi^{\text{cl}}_y}\right)^{-1} \\ &= i \int d^4w \, D_{zw} \, (-1) \int d^4u \int d^4v (-iD_{xu}) \frac{\delta^3 \Gamma}{\delta \phi^{\text{cl}}_u \delta \phi^{\text{cl}}_v \delta \phi^{\text{cl}}_w} (-iD_{vy})\end{aligned}$$

$$= i \int d^4u\, d^4v\, d^4w\, D_{xu} D_{yv} D_{zw} \frac{\delta^3 \Gamma}{\delta\phi^{\text{cl}}_u \delta\phi^{\text{cl}}_v \delta\phi^{\text{cl}}_w}. \tag{11.95}$$

This relation is more clearly expressed diagrammatically. The left-hand side is the connected three-point function. If we extract exact propagators as indicated in (11.95), this decomposes as follows:



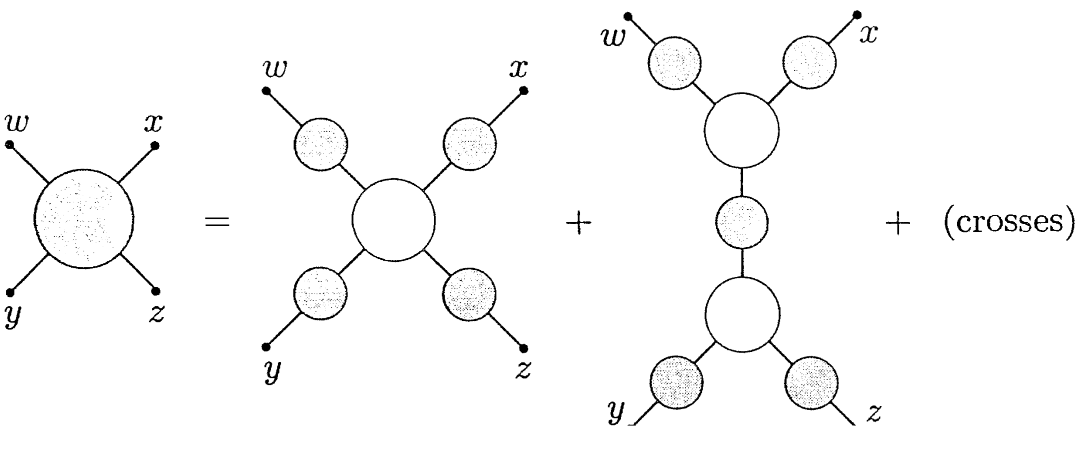
In this picture, each dark gray circle represents the sum of connected diagrams, while the light gray circle on the right-hand side represents the third derivative of $i\Gamma[\phi_{\text{cl}}]$. We see that the third derivative of $i\Gamma[\phi_{\text{cl}}]$ is just the connected correlation function with all three full propagators removed, that is, the *one-particle-irreducible* three-point function:

$$\frac{\delta^3 \Gamma}{\delta\phi_{\text{cl}}(x)\phi_{\text{cl}}(y)\phi_{\text{cl}}(z)} = i\, \langle \phi(x)\phi(y)\phi(z)\rangle_{\text{1PI}}\,.$$

By similar, if increasingly complicated, manipulations, one can derive the same relation for each successive derivative of Γ. For example, differentiating Eq. (11.95), we eventually find (using matrix notation with repeated indices implicitly integrated over)

$$\begin{aligned}\frac{-i\delta^4 E}{\delta J_w \delta J_x \delta J_y \delta J_z} = D_{sw} D_{xt} D_{yu} D_{zv} \Bigg[& \frac{i\delta^4 \Gamma}{\delta\phi^{\text{cl}}_s \delta\phi^{\text{cl}}_t \delta\phi^{\text{cl}}_u \delta\phi^{\text{cl}}_v} \\ & + \frac{i\delta^3 \Gamma}{\delta\phi^{\text{cl}}_s \delta\phi^{\text{cl}}_t \delta\phi^{\text{cl}}_r} D_{qr} \frac{i\delta^3 \Gamma}{\delta\phi^{\text{cl}}_q \delta\phi^{\text{cl}}_u \delta\phi^{\text{cl}}_v} + (t \leftrightarrow u) + (t \leftrightarrow v) \Bigg].\end{aligned}$$

Since the left-hand side of this equation is the connected four-point function, we can rewrite it diagrammatically as



As above, the dark gray circles represent the sum of connected diagrams, while the light gray circles represent i times various derivatives of Γ. Subtracting the last three terms from each side removes all one-particle reducible pieces from the connected four-point function and so identifies the fourth derivative of $i\Gamma$ as the one-particle-irreducible four-point function. The general relation (for $n \geq 3$) is

$$\frac{\delta^n \Gamma[\phi_{\rm cl}]}{\delta\phi_{\rm cl}(x_1)\cdots\delta\phi_{\rm cl}(x_n)} = -i\,\langle\phi(x_1)\cdots\phi(x_n)\rangle_{\rm 1PI}\,. \tag{11.96}$$

In other words, the effective action is the generating functional of one-particle-irreducible correlation functions.

This conclusion implies that Γ contains the complete set of physical predictions of the quantum field theory. Let us review how this information is encoded. The vacuum state of the field theory is identified as the minimum of the effective potential. The location of the minimum determines whether the symmetries of the Lagrangian are preserved or spontaneously broken. The second derivative of Γ is the inverse propagator. The poles of the propagator, or the zeros of the inverse propagator, give the values of the particle masses. Thus the particle masses m^2 are determined as the values of p^2 that solve the equation

$$\widetilde{D}^{-1}(p^2) = \int d^4x\, e^{ip\cdot(x-y)} \frac{\delta\Gamma}{\delta\phi\delta\phi}(x,y) = 0. \tag{11.97}$$

The higher derivatives of Γ are the one-particle-irreducible amplitudes. These can be connected by full propagators and joined together to construct four- and higher-point connected amplitudes, which give the S-matrix elements. Thus, from the knowledge of Γ, we can reconstruct the qualitative behavior of the quantum field theory, its pattern of symmetry-breaking, and then the quantitative details of its particles and their interactions.

11.6 Renormalization and Symmetry: General Analysis

In our analysis of the divergences of quantum field theories (especially in the paragraph below Eq. (10.4)), we noted that the basic divergences of Feynman integrals are associated with one-particle-irreducible diagrams. Thus we might expect that the effective action will be a useful object in discussing the renormalizability of quantum field theories, especially those with spontaneously broken symmetry. In this section we will make use of the effective action in precisely this way.

In Section 11.4, we saw in a particular example that the formalism for calculating the effective action provides the counterterms needed to remove the ultraviolet divergences, at least at the one-loop level. These counterterms were exactly those of the original Lagrangian. We will now argue that this set of counterterms is always sufficient—to all orders and for any renormalizable field theory—by applying the power-counting arguments of Section 10.1

directly to the computation of the effective action. We will use the language of scalar field theories, but the arguments can be generalized to theories of spinor and vector fields.

Consider first the computation of the effective potential for constant (x-independent) classical fields, in a field theory with an arbitrary number of fields ϕ^i. The effective potential has mass dimension 4, so we expect that $V_{\text{eff}}(\phi_{\text{cl}})$ will have divergent terms up to Λ^4. To understand these divergences, expand $V_{\text{eff}}(\phi_{\text{cl}})$ in a Taylor series:

$$V_{\text{eff}}(\phi_{\text{cl}}) = A_0 + A_2^{ij}\phi_{\text{cl}}^i\phi_{\text{cl}}^j + A_4^{ijk\ell}\phi_{\text{cl}}^i\phi_{\text{cl}}^j\phi_{\text{cl}}^k\phi_{\text{cl}}^\ell + \cdots.$$

In theories without a symmetry $\phi^i \to -\phi^i$, there might also be terms linear and cubic in ϕ^i; we omit these for simplicity. The coefficients A_0, A_2, A_4 have mass dimension, respectively, 4, 2, and 0; thus we expect them to contain Λ^4, Λ^2, and $\log \Lambda$ divergences, respectively. The power-counting analysis predicts that all higher terms in the Taylor series expansion should be finite. The constant term A_0 is independent of ϕ_{cl}; it has no physical significance. However, the divergences in A_2 and A_4 appear in physical quantities, since these coefficients enter the inverse propagator (11.90) and the irreducible four-point function (11.96) and therefore appear in the computation of S-matrix elements. There is one further coefficient in the effective action that has non-negative mass dimension by power counting; this is the coefficient of the term quadratic in $\partial_\mu\phi_{\text{cl}}^i$, which appears when the effective action is evaluated for a nonconstant background field:

$$\Delta\Gamma[\phi_{\text{cl}}] = \int d^4x\, B_2^{ij}\,\partial_\mu\phi_{\text{cl}}^i\partial^\mu\phi_{\text{cl}}^j. \tag{11.98}$$

All other coefficients in the Taylor expansion of the effective action in powers of ϕ_{cl}^i are finite by power counting.

We can now argue that the counterterms of the original Lagrangian suffice to remove the divergences that might appear in the computation of $\Gamma[\phi_{\text{cl}}]$. The argument proceeds in two steps. We first use the BPHZ theorem to argue that the divergences of Green's functions can be removed by adjusting a set of counterterms corresponding to the possible operators that can be added to the Lagrangian with coefficients of mass dimension greater than or equal to zero. The coefficients of these counterterms are in 1-to-1 correspondence with the coefficients A_2, A_4, and B_2 of the effective action. Next, we use the fact that the effective action is manifestly invariant to the original symmetry group of the model. This is true even if the vacuum state of the model has spontaneous symmetry breaking. This symmetry of the effective action follows from the analysis of Section 11.4, since the method we presented there for computing the effective action is manifestly invariant to the original symmetry of the Lagrangian. Combining these two results, we conclude that the effective action can always be made finite by adjusting the set of counterterms that are invariant to the original symmetry of the theory, even if this symmetry is spontaneously broken. By using the results of Section 11.5, which explain how to construct the Green's functions of the theory from the functional derivatives of the effective action, this conclusion of renormalizability extends to all the Green's functions of the theory.

To make this abstract argument more concrete, we will demonstrate in a simple example how the functional derivatives of the effective action yield a set of Feynman diagrams whose divergences correspond to symmetric counterterms. Let us, then, return once again to the $O(N)$-invariant linear sigma model and compute the second functional derivative of $\Gamma[\phi_{\rm cl}]$. If the whole formalism we have constructed hangs together, we should be able to recognize the result as the Feynman diagram expansion of the inverse propagator, with divergences corresponding to the counterterms of $O(N)$-symmetric scalar field theory.

To begin, we write out expression (11.63) explicitly for this model:

$$\Gamma[\phi_{\rm cl}] = \int d^4x \Big(\frac{1}{2}(\partial_\mu^2 \phi_{\rm cl}^i)^2 + \frac{1}{2}\mu^2 (\phi_{\rm cl}^i)^2 - \frac{\lambda}{4}((\phi_{\rm cl}^i)^2)^2 + \frac{1}{2}\log\det[-i\mathcal{D}^{ij}] + \cdots \Big), \tag{11.99}$$

where

$$-i\mathcal{D}^{ij} = -\frac{\delta^2 \mathcal{L}}{\delta\phi^i \delta\phi^j} = \partial^2 \delta^{ij} + \big(\lambda(\phi_{\rm cl}^k(x))^2 - \mu^2\big)\delta^{ij} + 2\lambda \phi_{\rm cl}^i(x)\phi_{\rm cl}^j(x). \tag{11.100}$$

For constant $\phi_{\rm cl}^i$, $\mathcal{D}^{ij}$ is the operator that, acting on a given component of the scalar field, equals the Klein-Gordon operator with mass squared given by Eq. (11.69). This is the leading-order approximation to the inverse propagator of the linear sigma model.

To find the higher-order corrrections to the inverse propagator, we must compute the second functional derivative of the quantum correction terms in $\Gamma[\phi_{\rm cl}]$. From (11.99), we find

$$\frac{\delta^2 \Gamma}{\delta\phi_{\rm cl}^i(x)\delta\phi_{\rm cl}^j(y)} = \frac{\delta^2 \mathcal{L}}{\delta\phi_{\rm cl}^i(x)\delta\phi_{\rm cl}^j(y)} + \frac{i}{2}\frac{\delta^2}{\delta\phi_{\rm cl}^i(x)\delta\phi_{\rm cl}^j(y)} \log\det[-i\mathcal{D}] + \cdots .$$

The first term is just the Klein-Gordon operator $i\mathcal{D}^{ij}\delta(x-y)$. To compute the second term, use identity (9.77) for determinants of matrices:

$$\frac{\partial}{\partial\alpha}\log\det M(\alpha) = \frac{\partial}{\partial\alpha}\operatorname{tr}\log M(\alpha) = \operatorname{tr} M^{-1}\frac{\partial M}{\partial\alpha}. \tag{11.101}$$

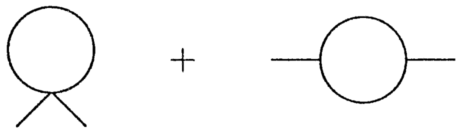
Using this identity, we find

$$\begin{aligned} &\frac{1}{2}\frac{\delta}{\delta\phi_{\rm cl}^k(z)}\log\det[-i\mathcal{D}] \\ &\qquad = \operatorname{Tr}\Big[\lambda\Big(\phi_{\rm cl}^k(z)\delta^{ij} + \phi_{\rm cl}^i(z)\delta^{jk} + \phi_{\rm cl}^j(z)\delta^{ik}\Big)(-i\mathcal{D}^{-1})^{ij}(z,z)\Big] \\ &\qquad = i\lambda\Big(\phi_{\rm cl}^k(z)\delta^{ij} + \phi_{\rm cl}^i(z)\delta^{jk} + \phi_{\rm cl}^j(z)\delta^{ik}\Big)(\mathcal{D}^{-1})^{ij}(z,z). \end{aligned} \tag{11.102}$$

The quantity $(\mathcal{D}^{-1})^{ij}(x,y)$ is the Klein-Gordon propagator. To differentiate a second time, we can use the identity (11.94); this yields

$$\begin{aligned}
&\frac{1}{2}\frac{\delta^2}{\delta\phi^k_{\text{cl}}(z)\delta\phi^\ell_{\text{cl}}(w)}\log\det[-i\mathcal{D}] \\
&\qquad = i\lambda(\delta^{k\ell}\delta^{ij}+\delta^{ik}\delta^{j\ell}+\delta^{i\ell}\delta^{jk})(\mathcal{D}^{-1})^{ij}(z,z)\delta(z-w) \\
&\qquad\quad - 2\lambda^2(\phi^k_{\text{cl}}(z)\delta^{ij}+\phi^i_{\text{cl}}(z)\delta^{jk}+\phi^j_{\text{cl}}(z)\delta^{ik})(\mathcal{D}^{-1})^{im}(z,w) \\
&\qquad\quad \cdot(\phi^\ell_{\text{cl}}(z)\delta^{mn}+\phi^m_{\text{cl}}(z)\delta^{n\ell}+\phi^n_{\text{cl}}(z)\delta^{m\ell})(\mathcal{D}^{-1})^{nj}(w,z). \qquad (11.103)
\end{aligned}$$

This is expected to be the formal correction to the inverse propagator at one-loop order, and indeed we can recognize in (11.103) the values of the one-loop diagrams



Notice how, in this derivation, every functional derivative on $\mathcal{D}^{-1}$ adds another propagator to the diagram and thus lowers the degree of divergence, in conformity with our general arguments in Section 10.1.

This example illustrates that the successive functional derivatives of $\Gamma[\phi_{\text{cl}}]$ are computed by a Feynman diagram expansion, with propagators and vertices that depend on the classical field. When the classical field is a constant, the propagators reduce to ordinary Klein-Gordon propagators and so the BPHZ theorem applies. All ultraviolet divergences can be removed from all of the amplitudes obtained by differentiating $\Gamma[\phi_{\text{cl}}]$ by the use of the most general set of mass, vertex, and field-strength renormalizations. At the same time, the perturbation theory is manifestly invariant to the symmetry of the original Lagrangian, and so the only divergences that appear—and thus the only counterterms required—are those that respect this symmetry. In general, then, all amplitudes of a renormalizable theory of scalar fields invariant under a symmetry group can be made finite using only the set of counterterms invariant to the symmetry. This gives a complete and quite satisfactory answer to the question posed at the beginning of Section 11.2.

The computation of the effective action in spatially varying background fields has not been analyzed at the level of rigor involved in the proof of the BPHZ theorem. However, it is expected that in this situation also, the standard set of counterterms for the symmetric theory should suffice. We can argue this intuitively by using the fact that the ultraviolet divergences of Feynman diagrams are local in spacetime. Thus, to understand the divergences of a computation in a background $\phi_{\text{cl}}(x)$ that is smoothly varying, we can divide spacetime into small boxes, in each of which $\phi_{\text{cl}}(x)$ is approximately constant, and expand in the derivatives $\partial_\mu\phi_{\text{cl}}(x)$. In this expansion in powers of $\partial_\mu\phi_{\text{cl}}(x)$, the Taylor series coefficients are functional derivatives of Γ in a constant background, which we know can be renormalized. The conclusion of this intuitive argument has been checked at the two-loop level for several nontrivial background field configurations.

Our general result on the renormalization of theories with spontaneously broken symmetry has an important implication for the physical predictions of these theories. In a renormalizable field theory, the most basic quantities of the theory cannot be predicted, because they are the quantities that must be specified as part of the definition of the theory. For example, in QED, the mass and charge of the electron must be adjusted from outside in order to define the theory. The predictions of QED are quantities that do not appear in the basic Lagrangian, for example, the anomalous magnetic moment of the electron. In renormalizable theories with spontaneously broken symmetry, however, the symmetry-breaking produces a large number of distinct masses and couplings, which depend on the relatively small number of parameters of the original symmetric theory. After the original parameters of the theory are fixed, any additional observable of the theory can be predicted unambiguously. For example, in the linear sigma model studied in this chapter, we took the values of the four-point coupling λ and the vacuum expectation value $\langle\phi\rangle$ as input parameters; we then calculated the mass of the σ particle in terms of these parameters in an unambiguous way.

There is a general argument that implies that, once we fix the parameters of the Lagrangian, we must find an unambiguous, finite formula for the σ mass in ϕ^4 theory, or, more generally, for any additional parameter of a renormalizable quantum field theory. In general, this parameter will be determined at the classical level in terms of the couplings in the Lagrangian. For the example of the σ mass in the linear sigma model, this classical relation is

$$m - \sqrt{2}\lambda\,\langle\phi\rangle = 0, \tag{11.104}$$

where m is the mass of the σ and λ gives the four-ϕ scattering amplitude at threshold. In general, loop corrections will modify this relation, contributing some nonzero expression to the right-hand side of this equation. However, since Eq. (11.104) is valid at the classical level however the parameters of the Lagrangian are modified, it holds equally well when we add counterterms to the Lagrangian and then adjust these counterterms order by order. Thus, the counterterms must give zero contributions to the right-hand side of Eq. (11.104). Therefore, the perturbative corrections to Eq. (11.104) must be automatically ultraviolet-finite. A relation of this type, true at the classical level for all values of the couplings in the Lagrangian, but corrected by loop effects, is called a *zeroth-order natural relation.* The argument we have given implies that, for any such relation, the loop corrections are finite and constitute predictions of the quantum field theory. We will see another example of such a relation in Problem 11.2.

Goldstone's Theorem Revisited

As a final application of the effective action formalism, let us return to the question of whether Goldstone's theorem is valid in the presence of quantum corrections. Recall that we proved this theorem at the classical level at the end of Section 11.1: We showed in (11.13) that, if the Lagrangian has a continuous symmetry that is spontaneously broken, the matrix of second derivatives of the classical potential $V(\phi)$ has a corresponding zero eigenvalue. According to Eq. (11.11), this implies that the classical theory contains a massless scalar particle, associated with the spontaneously broken symmetry.

Using the effective action formalism, this argument can be repeated almost verbatim in the full quantum field theory. The effective potential $V_{\rm eff}(\phi_{\rm cl})$ encapsulates the full solution to the theory, including all orders of quantum corrections. At the same time, it satisfies the general properties of the classical potential: It is invariant to the symmetries of the theory, and its minimum gives the vacuum expectation value of ϕ. This means that the argument we gave in (11.13) works in exactly the same way for $V_{\rm eff}$ as it does for V: If a continuous symmetry of the original Lagrangian is spontaneously broken by $\langle\phi\rangle$, the matrix of second derivatives of $V_{\rm eff}(\phi_{\rm cl})$ has a zero eigenvalue along the symmetry direction.

We now argue that, just as at the classical level, the presence of such a zero eigenvalue implies the existence of a massless scalar particle. In our discussion of the general properties of the effective action, we showed that its second functional derivative is the inverse propagator, and that, through Eq. (11.97), this derivative yields the spectrum of masses in the quantum theory. Let us rewrite Eq. (11.97) for a theory that contains several scalar fields:

$$\int d^4x\, e^{-ip\cdot(x-y)} \frac{\delta\Gamma}{\delta\phi^i\delta\phi^j}(x,y) = 0. \tag{11.105}$$

A particle of mass m corresponds to a zero eigenvalue of this matrix equation at $p^2 = m^2$. Now set $p = 0$. This implies that we differentiate $\Gamma[\phi_{\rm cl}]$ with respect to constant fields. Thus, we can replace $\Gamma[\phi_{\rm cl}]$ by its value with constant classical fields, which is just the effective potential. We find that the quantum field theory contains a scalar particle of zero mass when the matrix of second derivatives,

$$\frac{\partial^2 V_{\rm eff}}{\partial\phi^i_{\rm cl}\partial\phi^j_{\rm cl}},$$

has a zero eigenvalue. This completes the proof of Goldstone's theorem.

This argument for Goldstone's theorem illustrates the power of the effective action formalism. The formalism gives a geometrical picture of spontaneous symmetry breaking that is valid to any order in quantum corrections. As a bonus, it is built up from objects that are renormalized in a simple way. This formalism will prove useful in understanding the applications of spontaneously broken symmtry that occur, in several different contexts, throughout the rest of this book.

Problems

11.1 Spin-wave theory.

(a) Prove the following wonderful formula: Let $\phi(x)$ be a free scalar field with propagator $\langle T\phi(x)\phi(0)\rangle = D(x)$. Then

$$\left\langle Te^{i\phi(x)}e^{-i\phi(0)}\right\rangle = e^{[D(x)-D(0)]}.$$

(The factor $D(0)$ gives a formally divergent adjustment of the overall normalization.)

(b) We can use this formula in Euclidean field theory to discuss correlation functions in a theory with spontaneously broken symmetry for $T < T_C$. Let us consider only the simplest case of a broken $O(2)$ or $U(1)$ symmetry. We can write the local spin density as a complex variable

$$s(x) = s^1(x) + is^2(x).$$

The global symmetry is the transformation

$$s(x) \to e^{-i\alpha}s(x).$$

If we assume that the physics freezes the modulus of $s(x)$, we can parametrize

$$s(x) = Ae^{i\phi(x)}$$

and write an effective Lagrangian for the field $\phi(x)$. The symmetry of the theory becomes the translation symmetry

$$\phi(x) \to \phi(x) - \alpha.$$

Show that (for $d > 0$) the most general renormalizable Lagrangian consistent with this symmetry is the free field theory

$$\mathcal{L} = \tfrac{1}{2}\rho(\vec{\nabla}\phi)^2.$$

In statistical mechanics, the constant ρ is called the *spin wave modulus*. A reasonable hypothesis for ρ is that it is finite for $T < T_C$ and tends to 0 as $T \to T_C$ from below.

(c) Compute the correlation function $\langle s(x)s^*(0)\rangle$. Adjust A to give a physically sensible normalization (assuming that the system has a physical cutoff at the scale of one atomic spacing) and display the dependence of this correlation function on x for $d = 1, 2, 3, 4$. Explain the significance of your results.

11.2 A zeroth-order natural relation. This problem studies an $N = 2$ linear sigma model coupled to fermions:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi^i)^2 + \frac{1}{2}\mu^2(\phi^i)^2 - \frac{\lambda}{4}((\phi^i)^2)^2 + \bar{\psi}(i\partial\!\!\!/)\psi - g\bar{\psi}(\phi^1 + i\gamma^5\phi^2)\psi \tag{1}$$

where ϕ^i is a two-component field, $i = 1, 2$.

(a) Show that this theory has the following global symmetry:

$$\begin{aligned}\phi^1 &\to \cos\alpha\,\phi^1 - \sin\alpha\,\phi^2,\\ \phi^2 &\to \sin\alpha\,\phi^1 + \cos\alpha\,\phi^2,\\ \psi &\to e^{-i\alpha\gamma^5/2}\,\psi.\end{aligned} \tag{2}$$

Show also that the solution to the classical equations of motion with the minimum energy breaks this symmetry spontaneously.

(b) Denote the vacuum expectation value of the field ϕ^i by v and make the change of variables

$$\phi^i(x) = (v + \sigma(x), \pi(x)). \tag{3}$$

Write out the Lagrangian in these new variables, and show that the fermion acquires a mass given by

$$m_f = g \cdot v. \tag{4}$$

(c) Compute the one-loop radiative correction to m_f, choosing renormalization conditions so that v and g (defined as the $\psi\psi\pi$ vertex at zero momentum transfer) receive no radiative corrections. Show that relation (4) receives nonzero corrections but that these corrections are *finite*. This is in accord with our general discussion in Section 11.6.

11.3 The Gross-Neveu model. The Gross-Neveu model is a model in two spacetime dimensions of fermions with a discrete chiral symmetry:

$$\mathcal{L} = \bar{\psi}_i i\partial\!\!\!/\psi_i + \tfrac{1}{2}g^2(\bar{\psi}_i\psi_i)^2$$

with $i = 1, \ldots, N$. The kinetic term of two-dimensional fermions is built from matrices γ^μ that satisfy the two-dimensional Dirac algebra. These matrices can be 2×2:

$$\gamma^0 = \sigma^2, \qquad \gamma^1 = i\sigma^1,$$

where σ^i are Pauli sigma matrices. Define

$$\gamma^5 = \gamma^0\gamma^1 = \sigma^3;$$

this matrix anticommutes with the γ^μ.

(a) Show that this theory is invariant with respect to

$$\psi_i \to \gamma^5\psi_i,$$

and that this symmetry forbids the appearance of a fermion mass.

(b) Show that this theory is renormalizable in 2 dimensions (at the level of dimensional analysis).

(c) Show that the functional integral for this theory can be represented in the following form:

$$\int \mathcal{D}\psi\; e^{i\int d^2x\,\mathcal{L}} = \int \mathcal{D}\psi\mathcal{D}\sigma\; \exp\left[i\int d^2x\left\{\bar{\psi}_i i\partial\!\!\!/\psi_i - \sigma\bar{\psi}_i\psi_i - \frac{1}{2g^2}\sigma^2\right\}\right],$$

where $\sigma(x)$ (not to be confused with a Pauli matrix) is a new scalar field with no kinetic energy terms.

(d) Compute the leading correction to the effective potential for σ by integrating over the fermion fields ψ_i. You will encounter the determinant of a Dirac operator; to evaluate this determinant, diagonalize the operator by first going to Fourier components and then diagonalizing the 2×2 Pauli matrix associated with each Fourier mode. (Alternatively, you might just take the determinant of this 2×2 matrix.) This 1-loop contribution requires a renormalization proportional to σ^2 (that is, a renormalization of g^2). Renormalize by minimal substraction.

(e) Ignoring two-loop and higher-order contributions, minimize this potential. Show that the σ field acquires a vacuum expectation value which breaks the symmetry of part (a). Convince yourself that this result does not depend on the particular renormalization condition chosen.

(f) Note that the effective potential derived in part (e) depends on g and N according to the form

$$V_{\text{eff}}(\sigma_{\text{cl}}) = N \cdot f(g^2 N).$$

(The overall factor of N is expected in a theory with N fields.) Construct a few of the higher-order contributions to the effective potential and show that they contain additional factors of N^{-1} which suppress them if we take the limit $N \to \infty$, $(g^2 N)$ fixed. In this limit, the result of part (e) is unambiguous.

Chapter 12

The Renormalization Group

In the past two chapters, our main goal has been to determine when, and how, the cancellation of ultraviolet divergences in quantum field theory takes place. We have seen that, in a large class of field theories, the divergences appear only in the values of a few parameters: the bare masses and coupling constants, or, in renormalized perturbation theory, the counterterms. Aside from the shift in these parameters, virtual particles with very large momenta have no effect on computations in these theories.

The cancellation of ultraviolet divergences is essential if a theory is to yield quantitative physical predictions. But, at a deep level, the fact that high-momentum virtual quanta can have so little effect on a theory is quite surprising. One of the essential features of quantum field theory is locality, that is, the fact that fields at different spacetime points are independent degrees of freedom with independent quantum fluctuations. The quantum fluctuations at arbitrarily short distances appear in Feynman diagram computations as virtual quanta with arbitrarily high momenta. In a renormalizable theory, the loop integrals over virtual-particle momenta are always dominated by values comparable to the finite external particle momenta. But why? It is not easy to understand how the quantum fluctuations associated with extremely short distances can be so innocuous as to affect a theory only through the values of a few of its parameters.

This chapter begins with a physical picture, due to Kenneth Wilson, that explains this unusual and counterintuitive simplification. This picture generalizes the idea of the distance- or scale-dependent electric charge, introduced at the end of Chapter 7, and suggests that all of the parameters of a renormalizable field theory can usefully be thought of as scale-dependent entities. We will see that this scale dependence is described by simple differential equations, called *renormalization group* equations. The solutions of these equations will lead to physical predictions of a completely new type: predictions that, under certain circumstances, the correlation functions of a quantum field exhibit unusual but computable scaling laws as a function of their coordinates.

12.1 Wilson's Approach to Renormalization Theory

Wilson's method is based on the functional integral approach to field theory, in which the degrees of freedom of a quantum field are variables of integration. In this approach, one can study the origin of ultraviolet divergences by isolating the dependence of the functional integral on the short-distance degrees of freedom of the field.* In this section, we will illustrate this idea in the simplest example of ϕ^4 theory.

To make our analysis more concrete, we will drop the elegant but somewhat mysterious method of dimensional regularization in this section and instead use a sharp momentum cutoff. Since we will be working here only in ϕ^4 theory, we will not be concerned that this cutoff makes it difficult to satisfy Ward identities. Wilson's analysis can be adapted to QED and other situations where this subtlety is important, but the case of ϕ^4 theory is sufficient to give us the basic qualitative results of this approach.

In Section 9.2, we constructed the Green's functions of ϕ^4 theory in terms of a functional integral representation of the generating functional $Z[J]$. The basic integration variables are the Fourier components of the field $\phi(k)$, so $Z[J]$ is given concretely by the expression

$$Z[J] = \int \mathcal{D}\phi \, e^{i\int[\mathcal{L}+J\phi]} = \left(\prod_k \int d\phi(k)\right) e^{i\int[\mathcal{L}+J\phi]}. \tag{12.1}$$

To impose a sharp ultraviolet cutoff Λ, we restrict the number of the integration variables displayed in (12.1). That is, we integrate only over $\phi(k)$ with $|k| \leq \Lambda$, and set $\phi(k) = 0$ for $|k| > \Lambda$.

This modification of the functional integral suggests a method for assessing the influence of the quantum fluctuations at very short distances or very large momenta. In the functional integral representation, these fluctuations are represented by the integrals over the Fourier components of ϕ with momenta near the cutoff. Why not explicitly perform the integrals over these variables? Then we can compare the result to the original functional integral, and determine precisely the influence of these high-momentum modes on the physical predictions of the theory.

Before beginning this analysis, though, we must introduce one modification. At first sight, it seems most natural to define the ultraviolet cutoff in Minkowski space. However, a cutoff $k^2 \leq \Lambda^2$ is not completely effective in controlling large momenta, since in lightlike directions the components of k can be very large while k^2 remains small. We will therefore consider the cutoff to be imposed on the Euclidean momenta obtained after Wick rotation. Equivalently, we consider the Euclidean form of the functional integral, presented in Section 9.3, and restrict its variables $\phi(k)$, with k Euclidean, to $|k| \leq \Lambda$.

*Wilson's ideas are reviewed in K. G. Wilson and J. Kogut, *Phys. Repts.* **12C**, 75 (1974).

The transition to Euclidean space also brings us closer to the connection between renormalization theory and statistical mechanics advertised in Chapter 8. As we saw in Section 9.3, the Euclidean functional integral for ϕ^4 theory has precisely the same form as the continuum description of the statistical mechanics of a magnet. The field $\phi(x)$ is interpreted as the fluctuating spin field $s(x)$. A real magnet is built of atoms, and the atomic spacing provides a physical cutoff, a shortest distance over which fluctuations can take place. The cut-off functional integral models the effects of this atomic size in a crude way.

By pursuing this analogy, we can derive some physical intuition about the effects of the ultraviolet cutoff in a field theory. In a magnet, it is quite easy to visualize statistical fluctuations of the spins at the atomic scale. In fact, for values of the temperature away from any critical points, the statistical fluctuations are restricted to this scale; over distances of tens of atomic spacings, the magnet already shows its homogeneous macroscopic behavior. We have seen in Chapter 8 that we can approximate the correlation function of the spin field by the propagator of a Euclidean ϕ^4 theory. In this approximation,

$$\langle s(x)s(0)\rangle = \int \frac{d^4k}{(2\pi)^4}\,\frac{e^{ik\cdot x}}{k^2+m^2} \underset{|x|\to\infty}{\longrightarrow} \frac{1}{4\pi^2|x|^2}e^{-m|x|}. \tag{12.2}$$

As long as the temperature is far from the critical temperature, the size of the "mass" m is determined by the one natural scale in the problem, the atomic spacing. Thus, we expect $m \approx \Lambda$. In our field theory calculations, we were specifically interested in the situation where $m \ll \Lambda$, and we adjusted the parameters of the theory to satisfy this condition. In describing a magnet, it appears that no such adjustment is called for.

However, we saw in Chapter 8 that there is one circumstance in which the correlations of the spin field are much longer than the atomic spacing, so that, indeed, $m \ll \Lambda$. When the spin system begins to magnetize, just in the vicinity of the critical point, the spins become correlated over arbitrarily long distances as the fluctuating spins attempt to choose their eventual direction of magnetization. To study these long-range correlations in a magnet, one must carefully adjust the temperature to bring the system into the vicinity of the phase transition. In the same way, we can imagine making a fine adjustment of the parameter m of ϕ^4 theory to bring the quantum field theory into a region of parameters where we do find correlations of the field $\phi(x)$ over distances much larger than $1/\Lambda$.

Integrating Over a Single Momentum Shell

With this introduction, we will now carry out the integration over the high-momentum degrees of freedom of ϕ. We begin by writing the functional integral (12.1) more explicitly for the case of ϕ^4 theory. We apply the cutoff prescription described earlier, and set $J = 0$ for simplicity. Then

$$Z = \int [\mathcal{D}\phi]_\Lambda \exp\Big(-\int d^dx \Big[\frac{1}{2}(\partial_\mu\phi)^2 + \frac{1}{2}m^2\phi^2 + \frac{\lambda}{4!}\phi^4\Big]\Big), \tag{12.3}$$

where

$$[\mathcal{D}\phi]_\Lambda = \prod_{|k|<\Lambda} d\phi(k). \tag{12.4}$$

In the Lagrangian of Eq. (12.3), m and λ are the bare parameters, and so there are no counterterms. As in our study of the superficial degree of divergence, it will be useful to carry out this analysis in an arbitrary spacetime dimension d.

We now divide the integration variables $\phi(k)$ into two groups. Choose a fraction $b < 1$. The variables $\phi(k)$ with $b\Lambda \le |k| < \Lambda$ are the high-momentum degrees of freedom that we will integrate over. To label these degrees of freedom, let us define

$$\hat{\phi}(k) = \begin{cases} \phi(k) & \text{for } b\Lambda \le |k| < \Lambda; \\ 0 & \text{otherwise.} \end{cases}$$

Next, let us define a new $\phi(k)$, which is identical to the old for $|k| < b\Lambda$ and zero for $|k| > b\Lambda$. Then we can replace the old ϕ in the Lagrangian with $\phi + \hat{\phi}$, and rewrite Eq. (12.3) as

$$\begin{aligned} Z &= \int \mathcal{D}\phi \int \mathcal{D}\hat{\phi} \exp\Big(-\int d^dx \Big[\frac{1}{2}(\partial_\mu\phi + \partial_\mu\hat{\phi})^2 + \frac{1}{2}m^2(\phi+\hat{\phi})^2 + \frac{\lambda}{4!}(\phi+\hat{\phi})^4\Big]\Big) \\ &= \int \mathcal{D}\phi\, e^{-\int \mathcal{L}(\phi)} \int \mathcal{D}\hat{\phi} \exp\Big(-\int d^dx \Big[\frac{1}{2}(\partial_\mu\hat{\phi})^2 + \frac{1}{2}m^2\hat{\phi}^2 \\ &\qquad + \lambda\Big(\frac{1}{6}\phi^3\hat{\phi} + \frac{1}{4}\phi^2\hat{\phi}^2 + \frac{1}{6}\phi\hat{\phi}^3 + \frac{1}{4!}\hat{\phi}^4\Big)\Big]\Big). \end{aligned} \tag{12.5}$$

In the final expression we have gathered all terms independent of $\hat{\phi}$ into $\mathcal{L}(\phi)$. Note that quadratic terms of the form $\phi\hat{\phi}$ vanish, since Fourier components of different wavelengths are orthogonal.

The next few paragraphs will explain how to perform the integral over $\hat{\phi}$. This integration will transform (12.5) into an expression of the form

$$Z = \int [\mathcal{D}\phi]_{b\Lambda} \exp\Big(-\int d^dx\, \mathcal{L}_{\text{eff}}\Big), \tag{12.6}$$

where $\mathcal{L}_{\text{eff}}(\phi)$ involves only the Fourier components $\phi(k)$ with $|k| < b\Lambda$. We will see that $\mathcal{L}_{\text{eff}}(\phi) = \mathcal{L}(\phi)$ plus corrections proportional to powers of λ. These correction terms compensate for the removal of the large-k Fourier components $\hat{\phi}$, by supplying the interactions among the remaining $\phi(k)$ that were previously mediated by fluctuations of the $\hat{\phi}$.

To carry out the integrals over the $\hat{\phi}(k)$, we use the same method that we applied in Section 9.2 to derive Feynman rules. In fact, we will see below that the new terms in $\mathcal{L}_{\text{eff}}$ can be written in a diagrammatic form. In this analysis, we treat the quartic terms in (12.5), all proportional to λ, as perturbations.

Since we are mainly interested in the situation $m^2 \ll \Lambda^2$, we will also treat the mass term $\frac{1}{2}m^2\hat{\phi}^2$ as a perturbation. Then the leading-order term in the portion of the Lagrangian involving $\hat{\phi}$ is

$$\int \mathcal{L}_0 = \int\limits_{b\Lambda\leq|k|<\Lambda} \frac{d^dk}{(2\pi)^d}\,\hat{\phi}^*(k)k^2\hat{\phi}(k). \tag{12.7}$$

This term leads to a propagator

$$\overbrace{\hat{\phi}(k)\hat{\phi}(p)} = \frac{\int \mathcal{D}\hat{\phi}\, e^{-\int\mathcal{L}_0}\hat{\phi}(k)\hat{\phi}(p)}{\int \mathcal{D}\hat{\phi}\, e^{-\int\mathcal{L}_0}} = \frac{1}{k^2}(2\pi)^d\delta^{(d)}(k+p)\Theta(k), \tag{12.8}$$

where

$$\Theta(k) = \begin{cases} 1 & \text{if } b\Lambda \leq |k| < \Lambda; \\ 0 & \text{otherwise.} \end{cases} \tag{12.9}$$

We will regard the remaining $\hat{\phi}$ terms in Eq. (12.5) as perturbations, and expand the exponential. The various contributions from these perturbations can be evaluated by using Wick's theorem with (12.8) as the propagator.

First consider the term that results from expanding to one power of the $\phi^2\hat{\phi}^2$ term in the exponent of (12.5). We find

$$-\int d^dx\, \frac{\lambda}{4}\phi^2\overbrace{\hat{\phi}\hat{\phi}} = -\frac{1}{2}\int \frac{d^dk_1}{(2\pi)^d}\,\mu\,\phi(k_1)\phi(-k_1), \tag{12.10}$$

where the coefficient μ is the result of contracting the two $\hat{\phi}$ fields:

$$\mu = \frac{\lambda}{2}\int\limits_{b\Lambda\leq|k|<\Lambda} \frac{d^dk}{(2\pi)^d}\frac{1}{k^2} = \frac{\lambda}{(4\pi)^{d/2}\Gamma(\frac{d}{2})}\frac{1-b^{d-2}}{d-2}\Lambda^{d-2}. \tag{12.11}$$

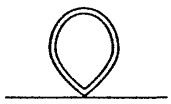
The term (12.10) could just as well have arisen from an expansion of the exponential

$$\exp\Bigl(-\int d^dx\,\frac{1}{2}\mu\phi^2 + \cdots\Bigr). \tag{12.12}$$

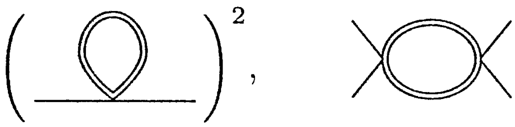
We will soon see that the rest of the perturbation series also organizes itself into this form. The coefficient μ therefore gives a positive correction to the m^2 term in $\mathcal{L}$.

The higher orders of the perturbation theory in the correction terms can be worked out in a similar way. As in our derivation of the standard perturbation theory for ϕ^4 theory, it is useful to adopt a diagrammatic notation. Represent the propagator (12.8) by a double line. This propagator will connect pairs of fields $\hat{\phi}$ from the various quartic interactions. Represent the fields ϕ in these interactions, which are not integrated over, as single external lines.

Then, for example, the contribution of (12.10) corresponds to the following diagram:



At order λ^2, we will have, among other contributions, terms involving the contractions of two interaction terms $\lambda\phi^2\hat{\phi}^2$. Each term corresponds to a vertex connecting two single lines and two double lines. There are two possible contractions:



(12.13)

Of these, the first, which is a disconnected diagram, supplies the order-λ^2 term in the exponential (12.12). The second is a new contribution, which will become a correction to the ϕ^4 interaction in $\mathcal{L}(\phi)$.

Let us now evaluate this second contribution. For simplicity, we consider the limit in which the external momenta carried by the factors ϕ are very small compared to $b\Lambda$, so we can ignore them. Then this diagram has the value

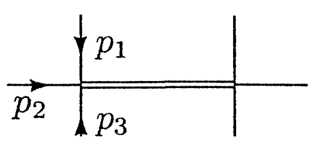
$$-\frac{1}{4!}\int d^dx\,\zeta\,\phi^4, \tag{12.14}$$

where

$$\zeta = -4!\frac{2}{2!}\left(\frac{\lambda}{4}\right)^2 \int\limits_{b\Lambda\leq|k|<\Lambda} \frac{d^dk}{(2\pi)^d}\left(\frac{1}{k^2}\right)^2 = \frac{-3\lambda^2}{(4\pi)^{d/2}\Gamma(\frac{d}{2})}\frac{(1-b^{d-4})}{d-4}\Lambda^{d-4}$$
$$\underset{d\to 4}{\longrightarrow} -\frac{3\lambda^2}{16\pi^2}\log\frac{1}{b}. \tag{12.15}$$

The 2 in the numerator counts the two possible contractions; there are no additional combinatoric factors from counting external legs or vertices. In the analysis of ϕ^4 theory in Section 10.2, we encountered a similar diagram, integrated over a range of momenta from 0 to Λ, producing a logarithmic ultraviolet divergence. In Wilson's treatment this divergence is not a pathology but simply a sign that the diagram is receiving contributions from all momentum scales. Indeed, it receives an equal contribution from each logarithmic interval between the momentum scales m and Λ. We will see below that the (finite) contribution to this diagram from each momentum interval has a natural physical importance.

The diagrammatic perturbation theory we have described not only generates contributions proportional to ϕ^2 and ϕ^4 but also to higher powers of ϕ. For example, the following diagram generates a contribution to a ϕ^6 interaction:



$$\propto \frac{\lambda^2}{(p_1+p_2+p_3)^2}\Theta(p_1+p_2+p_3). \tag{12.16}$$

There are also derivative interactions, which arise when we no longer neglect the external momenta of the diagrams. A more exact treatment would Taylor-expand in these momenta; for instance, in addition to expression (12.14), we would obtain terms with two powers of external momenta, which we could rewrite as

$$-\frac{1}{4}\int d^d x\,\eta\,\phi^2(\partial_\mu\phi)^2. \tag{12.17}$$

We would also find terms with four, six, and more powers of the momenta carried by the ϕ. In general, the procedure of integrating out the $\hat{\phi}$ generates all possible interactions of the fields ϕ and their derivatives.

The diagrammatic corrections can be simplified slightly by resumming them as an exponential. We have seen already in (12.13) that our diagrammatic expansion generates disconnected diagrams. By the same combinatoric argument that we used in Eq. (4.52), we can rewrite the sum of the series as the exponential of the sum of the connected diagrams. This leads precisely to expression (12.6), with

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}(\partial_\mu\phi)^2 + \frac{1}{2}m^2\phi^2 + \frac{1}{4!}\lambda\phi^4 + (\text{sum of connected diagrams}). \tag{12.18}$$

The diagrammatic contributions include corrections to m^2 and λ, as well as all possible higher-dimension operators. We can now use the new Lagrangian $\mathcal{L}_{\text{eff}}(\phi)$ to compute correlation functions of the $\phi(k)$, or to compute S-matrix elements. Since the $\phi(k)$ include only momenta up to $b\Lambda$, the loop diagrams in such a calculation would be integrated only up to that lowered cutoff. The correction terms in (12.18) precisely compensate for this change.

One might well be puzzled by the appearance of higher-dimension operators in Eq. (12.18). We chose the original Lagrangian of ϕ^4 theory to contain only renormalizable interactions. At first sight, it is disturbing that all possible nonrenormalizable interactions appear when we integrate out the variables $\hat{\phi}$. However, we will see below that our procedure actually keeps the contributions of these nonrenormalizable interactions under control. In fact, our analysis will imply that the presence of nonrenormalizable interactions in the original Lagrangian, defined to be used with very large cutoff Λ, has negligible effect on physics at scales much less than Λ.

Renormalization Group Flows

Let us now make a more careful comparison of the new functional integral (12.6) and the one we started with (12.3). The most convenient way to do this is to rescale distances and momenta in (12.6) according to

$$k' = k/b, \qquad x' = xb, \tag{12.19}$$

so that the variable k' is integrated over $|k'| < \Lambda$. Let us express the explicit form of (12.18) schematically as

$$\int d^dx\,\mathcal{L}_{\text{eff}} = \int d^dx\Big[\frac{1}{2}(1+\Delta Z)(\partial_\mu\phi)^2 + \frac{1}{2}(m^2+\Delta m^2)\phi^2 + \frac{1}{4}(\lambda+\Delta\lambda)\phi^4 + \Delta C(\partial_\mu\phi)^4 + \Delta D\phi^6 + \cdots\Big]. \tag{12.20}$$

In terms of the rescaled variable x', this becomes

$$\int d^dx\,\mathcal{L}_{\text{eff}} = \int d^dx'\,b^{-d}\Big[\frac{1}{2}(1+\Delta Z)b^2(\partial'_\mu\phi)^2 + \frac{1}{2}(m^2+\Delta m^2)\phi^2 + \frac{1}{4}(\lambda+\Delta\lambda)\phi^4 + \Delta C b^4(\partial'_\mu\phi)^4 + \Delta D\phi^6 + \cdots\Big]. \tag{12.21}$$

Throughout this analysis, we have treated all terms beyond the first as small perturbations. As long as the original couplings are small, this is still a valid approximation in treating (12.21).

The original functional integral led to the propagator (12.8). The new action (12.21) will give rise to exactly the same propagator, if we rescale the field ϕ according to

$$\phi' = \big[b^{2-d}(1+\Delta Z)\big]^{1/2}\phi. \tag{12.22}$$

After this rescaling, the unperturbed action returns to its initial form, while the various perturbations undergo a transformation:

$$\int d^dx\,\mathcal{L}_{\text{eff}} = \int d^dx'\Big[\frac{1}{2}(\partial'_\mu\phi')^2 + \frac{1}{2}m'^2\phi'^2 + \frac{1}{4}(\lambda'\phi'^4 + C'(\partial'_\mu\phi')^4 + D'\phi'^6 + \cdots\Big]. \tag{12.23}$$

The new parameters of the Lagrangian are

$$\begin{aligned} m'^2 &= (m^2+\Delta m^2)(1+\Delta Z)^{-1}b^{-2},\\ \lambda' &= (\lambda+\Delta\lambda)(1+\Delta Z)^{-2}b^{d-4},\\ C' &= (C+\Delta C)(1+\Delta Z)^{-2}b^{d},\\ D' &= (D+\Delta D)(1+\Delta Z)^{-3}b^{2d-6}, \end{aligned} \tag{12.24}$$

and so on. (The original Lagrangian had $C = D = 0$, but the same equations would apply if the initial values of C and D were nonzero.) All of the corrections, Δm^2, $\Delta\lambda$, and so on, arise from diagrams and thus are small compared to the leading terms if perturbation theory is justified.

By combining the operation of integrating out high-momentum degrees of freedom with the rescaling (12.19), we have rewritten this operation as a transformation of the Lagrangian. Continuing this procedure, we could integrate over another shell of momentum space and transform the Lagrangian further. Successive integrations produce further iterations of the transformation (12.24). If we take the parameter b to be close to 1, so that the shells of momentum space are infinitesimally thin, the transformation becomes a continuous one. We can then describe the result of integrating over the high-momentum degrees of freedom of a field theory as a trajectory or a flow in the space of all possible Lagrangians.

For historical reasons, these continuously generated transformations of Lagrangians are referred to as the *renormalization group*. They do not form a group in the formal sense, because the operation of integrating out degrees of freedom is not invertible. On the other hand, they are most certainly connected to renormalization, as we will now see.

Imagine that we wish to compute a correlation function of fields whose momenta p_i are all much less than Λ. We could compute this correlation function perturbatively using either the original Lagrangian $\mathcal{L}$, or the effective Lagrangian $\mathcal{L}_{\text{eff}}$ obtained after integrating over all momentum shells down to the scale of the external momenta p_i. Both procedures must ultimately yield the same result. But in the first case, the effects of high-momentum fluctuations of the field do not show up until we compute loop diagrams. In the second case, these effects have already been absorbed into the new coupling constants (m', λ', etc.), so their influence can be seen directly from the Lagrangian. In the first procedure, the large shifts from the original (bare) parameters to the values appropriate to low-momentum processes appear suddenly in one-loop diagrams, and seem to invalidate the use of perturbation theory. In the second approach, these corrections are introduced slowly and systematically. A perturbative treatment is valid at every step as long as the effective coupling constants such as λ' remain small.

However, the parameters of the effective Lagrangian may be very different from those of the original Lagrangian, since we must iterate the transformation (12.24) many times to get from the large momentum Λ down to the momentum scale of typical experiments. Let us therefore look more closely at how the Lagrangian tends to vary under the renormalization group transformations.

The simplest case to consider is a Lagrangian in the vicinity of the point $m^2 = \lambda = C = D = \cdots = 0$, where all the perturbations vanish. We have defined our transformation so that this point is left unchanged; we say that the free-field Lagrangian

$$\mathcal{L}_0 = \tfrac{1}{2}(\partial_\mu \phi)^2 \tag{12.25}$$

is a *fixed point* of the renormalization group transformation.

In the vicinity of $\mathcal{L}_0$, we can ignore the terms Δm^2, $\Delta\lambda$, etc., in the iteration equations (12.24) and keep only those terms that are linear in the perturbations. This gives an especially simple transformation law:

$$m'^2 = m^2 b^{-2}, \quad \lambda' = \lambda b^{d-4}, \quad C' = C b^d, \quad D' = D b^{2d-6}, \quad \text{etc.} \tag{12.26}$$

Since $b < 1$, those parameters that are multiplied by negative powers of b

grow, while those that are multiplied by positive powers of b decay. If the Lagrangian contains growing coefficients, these will eventually carry it away from $\mathcal{L}_0$.

It is conventional to speak of the various terms in the effective Lagrangian as a set of local operators that can be added as perturbations to $\mathcal{L}_0$. We call the operators whose coefficients grow during the recursion procedure *relevant* operators. The coefficients that die away are associated with *irrelevant* operators. For example, the scalar field mass operator ϕ^2 is always relevant, while the ϕ^4 operator is relevant if $d < 4$. If the coefficient of some operator is multiplied by b^0 (for example, the operator ϕ^4 in $d = 4$), we call this operator *marginal*; to find out whether its coefficient grows or decays, we must include the effect of higher-order corrections.

In general, an operator with N powers of ϕ and M derivatives has a coefficient that transforms as

$$C'_{N,M} = b^{N(d/2-1)+M-d} C_{N,M}. \tag{12.27}$$

Notice that the coefficient is just $(d_{N,M} - d)$, where $d_{N,M}$ is the mass dimension of the operator as computed at the end of Section 10.1. In other words, relevant and marginal operators about the free theory $\mathcal{L}_0$ correspond precisely to super-renormalizable and renormalizable interaction terms in the power-counting analysis of Section 10.1.

We can also understand the evolution of coefficients near the free-field fixed point using straightforward dimensional analysis. An operator with mass dimension d_i has a coefficient with dimension $(\text{mass})^{d-d_i}$. The natural order of magnitude for this mass is the cutoff Λ. Thus, if $d_i < d$, the perturbation is increasingly important at low momenta. On the other hand, if $d_i > d$, the relative size of this term decreases as $(p/\Lambda)^{d_i - d}$ as the momentum $p \to 0$; thus the term is truly irrelevant.

We have now shown that, at least in the vicinity of the zero-coupling fixed point, an arbitrarily complicated Lagrangian at the scale of the cutoff degenerates to a Lagrangian containing only a finite number of renormalizable interactions. It is instructive to compare this result with the conclusions of Chapter 10. There we took the philosophy that the cutoff Λ should be disposed of by taking the limit $\Lambda \to \infty$ as quickly as possible. We found that this limit gives well-defined predictions only if the Lagrangian contains no parameters with negative mass dimension. From this viewpoint, it seemed exceedingly fortunate that QED, for example, contained no such parameters, since otherwise this theory would not yield well-defined predictions.

Wilson's analysis takes just the opposite point of view, that any quantum field theory is defined fundamentally with a cutoff Λ that has some physical signficance. In statistical mechanical applications, this momentum scale is the inverse atomic spacing. In QED and other quantum field theories appropriate to elementary particle physics, the cutoff would have to be associated with some fundamental graininess of spacetime, perhaps a result of quantum fluctuations in gravity. We discuss some speculations on the nature of this

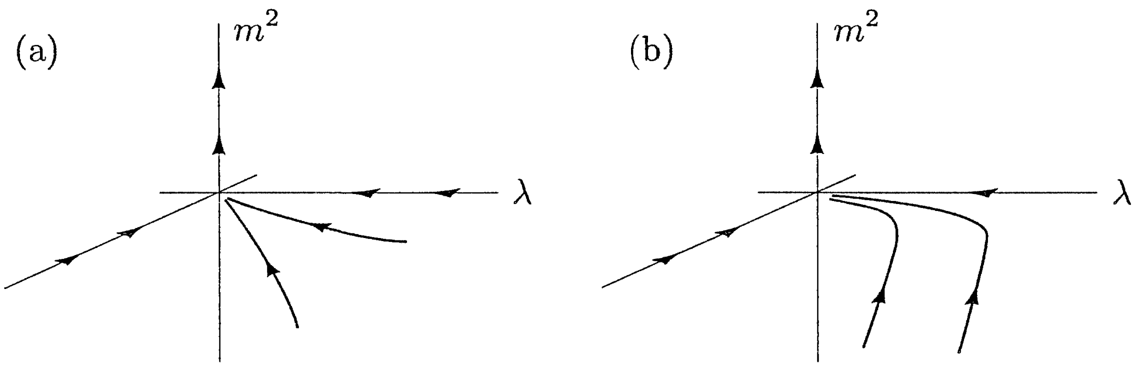


Figure 12.1. Renormalization group flows near the free-field fixed point in scalar field theory: (a) $d > 4$; (b) $d = 4$.

cutoff in the Epilogue. But whatever this scale is, it lies far beyond the reach of present-day experiments. The argument we have just given shows that this circumstance *explains* the renormalizability of QED and other quantum field theories of particle interactions. Whatever the Lagrangian of QED was at its fundamental scale, as long as its couplings are sufficiently weak, it must be described at the energies of our experiments by a renormalizable effective Lagrangian.

On the other hand, we should emphasize that these simple conclusions can be altered by sufficiently strong field theory interactions. Away from the free-field fixed point, the simple transformation laws (12.26) receive corrections proportional to higher powers of the coupling constants. If these corrections are large enough, they can halt or reverse the renormalization group flow. They could even create new fixed points, which would give new types of $\Lambda \to \infty$ limits.

To illustrate the possible influences of interactions in a relatively simple context, let us discuss the renormalization group flows near $\mathcal{L}_0$ for the specific case of ϕ^4 theory. It is instructive to consider the three cases $d > 4$, $d = 4$, and $d < 4$ in turn. When $d > 4$, the only relevant operator is the scalar field mass term. Then the renormalization group flows near $\mathcal{L}_0$ have the form shown in Fig. 12.1(a). The ϕ^4 interaction and possible higher-order interactions die away, while the mass term increases in importance.

In previous chapters, we have always discussed ϕ^4 theory in the limit in which the mass is small compared to the cutoff. Let us take a moment to rewrite this condition in the language of renormalization group flows. In the course of the flow, the effective mass term m'^2 becomes large and eventually comes to equal the current cutoff. For example, near the free-field fixed point, after n iterations, $m'^2 = m^2 b^{-2n}$, and eventually there is an n such that $m'^2 \sim \Lambda^2$. At this point, we have integrated out the entire momentum region between the original Λ and the effective mass of the scalar field. The mass term then suppresses the remaining quantum fluctuations. In general, the criterion that the scalar field mass is small compared to the cutoff is equivalent to the statement that $m'^2 \sim \Lambda^2$ only after a large number of iterations of the renormalization group transformation.

This criterion is met whenever the initial conditions for the renormalization group flow are adjusted so that the trajectory passes very close to a fixed point. In principle, the flow could begin far away, along the direction of an irrelevant operator. The original value of m^2 need not be particularly small, as long as this original value is canceled by corrections arising from the diagrammatic contributions to $\mathcal{L}_{\text{eff}}$. Thus we could imagine constructing a scalar field theory in $d > 4$ by writing a complicated nonlinear Lagrangian, but adjusting the original m^2 so the trajectory that begins at this Lagrangian eventually passes close to the free-field fixed point $\mathcal{L}_0$. In this case, the effective theory at momenta small compared to the cutoff should be extremely simple: It will be a free field theory with negligible nonlinear interaction. As will be discussed in the next chapter, this remarkable prediction has been verified in mathematical models of magnetic systems in more than four dimensions: Even though the original model is highly nonlinear, the correlation function of spins near the phase transition has the free-field form given by the higher-dimensional analogue of Eq. (12.2).

Next consider the case $d = 4$. For this case, Eq. (12.26) does not give enough information to tell us whether the ϕ^4 interaction is important or unimportant at large distances. So we must go back to the complete transformation law (12.24). The leading contribution to $\Delta\lambda$ is given by Eq. (12.15). The leading contribution to ΔZ is of order λ^2 and can be neglected. (This is just what happened with the first correction to δ_Z in Section 10.2.) Thus we find the transformation

$$\lambda' = \lambda - \frac{3\lambda^2}{16\pi^2}\log(1/b). \tag{12.28}$$

This says that λ slowly decreases as we integrate out high-momentum degrees of freedom.

The diagram contributing to the correction $\Delta\lambda$ has the same structure as the one-loop diagrams computed in Section 10.2. In fact, these are essentially the same diagrams, and differ only in whether the integrals are carried out iteratively or all at once. However, whereas the diagrams in Section 10.2 had ultraviolet divergences, the corresponding diagram in Wilson's approach is well defined and gives the coefficient of a simple evolution equation of the coupling constant. This transformation gives a first example of the reinterpretation of ultraviolet divergences that we will make in this chapter.

The transformation law (12.28) implies that the renormalization group flows near $\mathcal{L}_0$ have the form shown in Fig. 12.1(b), with one slowly decaying direction. If we follow the flows far enough, the behavior should again be that of a free field. This picture has the puzzling implication that four-dimensional interacting ϕ^4 theory does not exist in the limit in which the cutoff goes to infinity. We will discuss this result further—and explain why it nevertheless makes sense to use ϕ^4 theory as a model field theory—in Section 12.3.

Finally consider the case $d < 4$. Now λ becomes a relevant parameter.

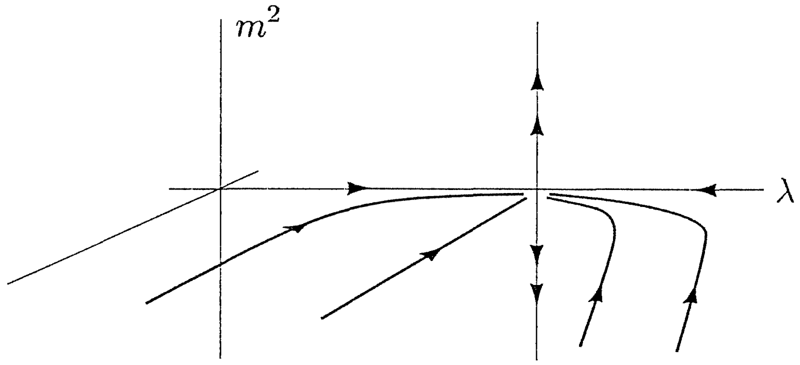


Figure 12.2. Renormalization group flows near the free-field fixed point in scalar field theory: $d < 4$.

The theory thus flows away from the free theory $\mathcal{L}_0$ as we integrate out degrees of freedom; at large distances, the ϕ^4 interaction becomes increasingly important. However, when λ becomes large, the nonlinear corrections such as that displayed in Eq. (12.28) must also be considered. If we include this specific effect in $d < 4$, we find the recursion formula

$$\lambda' = \left[\lambda - \frac{3\lambda^2}{(4\pi)^{d/2}\Gamma(\frac{d}{2})}\frac{b^{d-4}-1}{4-d}\Lambda^{d-4}\right]b^{d-4}. \tag{12.29}$$

This equation implies that there is a value of λ at which the increase due to rescaling is compensated by the decrease caused by the nonlinear effect. At this value, λ is unchanged when we integrate out degrees of freedom. The corresponding Lagrangian is a second fixed point of the renormalization group flow. In the limit $d \to 4$, the flow (12.29) tends to (12.28) and so the new fixed point merges with the free field fixed point. For d sufficiently close to 4, the new fixed point will share with $\mathcal{L}_0$ the property that the mass parameter m^2 is increased by the iteration. Then the mass operator will be a relevant operator near the new fixed point, so that the renormalization group flows will have the form shown in Fig. 12.2.

In this example, the new fixed point of the renormalization group had a Lagrangian with couplings weak enough that the transformation equations could be computed in perturbation theory. In principle, one could also find fixed points whose Lagrangians are strongly coupled, so that the renormalization group transformations cannot be understood by Feynman diagram analysis. Many examples of such fixed points are known in exactly solvable model field theories in two dimensions.† However, up to the present, all of the examples of quantum field theories that are important for physical applications have been found to be controlled either by the free field fixed point or by fixed points, like the one described in the previous paragraph, that approach the free-field fixed point in a specific limit. No one understands why this should be. This observation implies that Feynman diagram analysis has

†We mention some of these examples, and discuss other nonperturbative approaches to quantum field theory, in the Epilogue.

unexpected power in evaluating the physical consequences of quantum field theories.

One more aspect of ϕ^4 theory deserves comment. Since the mass term, $m^2\phi^2$, is a relevant operator, its coefficient diverges rapidly under the renormalization group flow. We have seen above that, in order to end up at the desired value of m^2 at low momentum, we must imagine that the value of m^2 in the original Lagrangian has been adjusted very delicately. This adjustment has a natural interpretation in a magnetic system as the need to sensitively adjust the temperature to be very close to the critical point. However, it seems quite artificial when applied to the quantum field theory of elementary particles, which purports to be a fundamental theory of Nature. This problem appears only for scalar fields, since for fermions the renormalization of the mass is proportional to the bare mass rather than being an arbitrary additive constant. Perhaps this is the reason why there seem to be no elementary scalar fields in Nature. We will return to this question in the Epilogue.

12.2 The Callan-Symanzik Equation

Wilson's picture of renormalization, as a flow in the space of possible Lagrangians, is beautifully intuitive, and gives us a deep understanding of why Nature should be describable in terms of renormalizable quantum field theories. In addition, however, this idea can be applied to extract further quantitative predictions from these theories. In the remainder of this chapter we will develop a formalism for extracting these predictions. Specifically, we will see that Wilson's picture leads to predictions for the form of the high- and low-momentum behavior of correlation functions. In the simplest cases, the correlation functions turn out to scale as powers of their external momenta, with power laws that do not appear at any fixed order of perturbation theory.

It is possible to derive these predictions directly from Wilson's procedure of integrating out slices in momentum space, as Wilson originally did. However, now that we understand the basic idea of renormalization group flows, it will be technically easier to work in the more familiar context of ordinary renormalized perturbation theory. The discussion of the previous section was physically motivated but technically complex. It involved awkward integrals over finite domains, and used the artificial parameter b, which must cancel out in any final results. Furthermore, we know from Section 7.5 that a cutoff regulator leads to even more trouble in QED, since it conflicts with the Ward identity. The discussion of the present section will be much more abstract and formal, but it will remove these technical problems. In this section and the next we will derive a flow equation for the coupling constant, similar to the one we derived in Section 12.1. To obtain the flows of the most general Lagrangians, we will need some additional tools, to be developed in Sections 12.4 and 12.5.

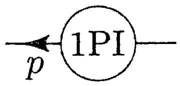
How can we hope to obtain information on renormalization group flows from the expressions for renormalized Green's functions, in which the cutoff has already been taken to infinity? We must first realize that renormalized quantum field theories correspond to a restricted class of the full set of possible Lagrangians that we considered in the previous section. In Wilson's language, a renormalized field theory with the cutoff taken arbitrarily large corresponds to a trajectory that takes an arbitrarily long time to evolve to a large value of the mass parameter. Such a trajectory must, then, pass arbitrarily close to a fixed point, which we will assume to be the weak-coupling fixed point. In the slow evolution past this fixed point, the irrelevant operators in the original Lagrangian die away, and we are left only with the relevant and marginal operators. The coefficients of these operators are in one-to-one correspondence with the parameters of the renormalizable field theory. Thus, in working with a renormalized field theory, we are throwing away information on the evolution of irrelevant perturbations, but keeping information on the flows of relevant and marginal perturbations.

The flows of these parameters cannot be determined from the cutoff dependence, because, in this framework, the cutoff has already been sent to infinity. However, we have an alternative, though more abstract, tool at our disposal. The parameters of a renormalized field theory are determined by a set of renormalization conditions, which are applied at a certain momentum scale (called the *renormalization scale*). By looking at how the parameters of the theory depend on the renormalization scale, we can recover the information contained in the renormalization group flows of the previous section.

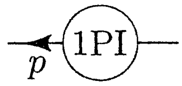
We consider first the specific case of ϕ^4 theory in four dimensions, where the coupling constant λ is dimensionless and the corresponding operator is marginal. For simplicity, we will also assume that the mass term m^2 has been adjusted to zero, so that the theory sits just at its critical point. We will perform this analysis in Minkowski space, using spacelike reference momenta. However, the analysis would be essentially identical if carried out in Euclidean space. If we wish to consider renormalization group predictions at timelike momenta, we must consider the possibilities of new singularities which make the analysis more complicated. These include both physical thresholds and the Sudakov double logarithms discussed in Section 6.4. We postpone discussion of these complications until Chapters 17 and 18.

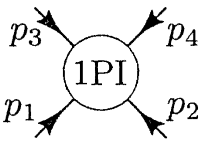
Renormalization Conditions

To define the theory properly, we must specify the renormalization conditions. In Chapter 10 we used a natural set of renormalization conditions (10.19) for ϕ^4 theory, defined in terms of the physical mass m. However, in a theory where $m = 0$, these conditions cannot be used because they lead to singularities in the counterterms. (Consider, for example, the limit $m^2 \to 0$ of Eq. (10.24).) To avoid such singularities, we choose an arbitrary momentum scale M and impose the renormalization conditions at a spacelike momentum p with $p^2 = -M^2$:



$$= 0 \quad \text{at } p^2 = -M^2;$$

$$\frac{d}{dp^2}\Bigg($$  $$\Bigg) = 0 \quad \text{at } p^2 = -M^2;$$



$$= -i\lambda \qquad \text{at } (p_1+p_2)^2 = (p_1+p_3)^2 = (p_1+p_4)^2 = -M^2. \tag{12.30}$$

The parameter M is called the *renormalization scale*. These conditions define the values of the two- and four-point Green's functions at a certain point and, in the process, remove all ultraviolet divergences. Speaking loosely, we say that we are "defining the theory at the scale M".

These new renormalization conditions take some getting used to. The second condition, in particular, implies that the two-point Green's function has a coefficient of 1 at the unphysical momentum $p^2 = -M^2$, rather than on shell (at $p^2 = 0$):

$$\langle\Omega|\,\phi(p)\phi(-p)\,|\Omega\rangle = \frac{i}{p^2} \qquad \text{at } p^2 = -M^2.$$

Here ϕ is the renormalized field, related to the bare field ϕ_0 by a scale factor that we again call Z:

$$\phi = Z^{-1/2}\phi_0. \tag{12.31}$$

This Z, however, is not the residue of the physical pole in the two-point Green's function of bare fields, as it was in Chapters 7 and 10. Instead, we now have

$$\langle\Omega|\,\phi_0(p)\phi_0(-p)\,|\Omega\rangle = \frac{iZ}{p^2} \qquad \text{at } p^2 = -M^2.$$

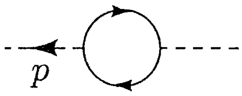
The Feynman rules for renormalized perturbation theory are the same as in Chapter 10, with the same relation between Z and the counterterm δ_Z,

$$\delta_Z = Z - 1.$$

Now, however, the counterterms δ_Z and δ_λ must be adjusted to maintain the new conditions (12.30).

The first renormalization condition in (12.30) holds the physical mass of the scalar field fixed at zero. We saw in Chapter 10 that, in ϕ^4 theory, the one-loop propagator correction is momentum-independent and is completely canceled by the mass renormalization counterterm. At two-loop order, however, the situation becomes more complicated, and the propagator corrections require both mass and field strength renormalizations. In more general scalar field theories, such as the Yukawa theory example considered at the end of Section 10.2, this complication arises already at one-loop order. Since the field strength renormalization counterterm will play an important role in the discussion below, it will be helpful to discuss briefly how we will treat this double subtraction.

The evaluation of propagator corrections has some special simplifications for the case of a massless scalar field, which we consider here, and specifically with the use of dimensional regularization. Consider, for example, the one-loop propagator correction in Yukawa theory. In Section 10.2 we found an expression of the form



$$\sim \frac{\Gamma(1-\frac{d}{2})}{\Delta^{1-d/2}}, \tag{12.32}$$

where Δ is a linear combination of the fermion mass m_f and p^2. If we compute the diagram using massless propagators only, Δ is proportional to p^2. Expression (12.32) has a pole at $d = 2$, corresponding to the quadratically divergent mass renormalization. However, the residue of this pole is independent of p^2, so we can completely cancel the pole with the mass counterterm δ_m. This allows us to analytically continue (12.32) to $d = 4$. Then this expression takes the form

$$-p^2\Big(\frac{1}{2-d/2} + \log\frac{1}{-p^2} + C\Big), \tag{12.33}$$

and gives no additional mass shift but only a field strength renormalization. The remaining divergence is canceled by the counterterm δ_Z. If we adopt the rule that we should simply continue expressions of the form (12.32) to $d = 4$, we can forget about the counterterm δ_m altogether.

In a regularization scheme with a momentum cutoff, the contributions to δ_m and δ_Z become tangled up with one another. Then it is more awkward to define the massless limit. In the following discussion, we will assume the use of dimensional regularization. However, to emphasize the physical role of the cutoff, we will write expressions of the form (12.33) as

$$-p^2\Big(\log\frac{\Lambda^2}{-p^2} + C\Big). \tag{12.34}$$

The logarithmically divergent terms proportional to p^2 will agree with the divergences obtained with a momentum cutoff; the constant terms will not agree, but these will drop out of our final results.

In ϕ^4 theory, where the one-loop propagator correction is momentum-independent, the one-loop diagram is simply set to zero by this prescription. Then the preceding analysis applies to the two-loop and higher correction terms.

The generalization of the analysis of this section to massive scalar field theory requires some additional formalism, which we postpone to Section 12.5.

The Callan-Symanzik Equation

In the renormalization conditions (12.30), the renormalization scale M is arbitrary. We could just as well have defined the same theory at a different scale M'. By "the same theory", we mean a theory whose bare Green's functions,

$$\langle\Omega| T\phi_0(x_1)\phi_0(x_2)\cdots\phi_0(x_n)|\Omega\rangle,$$

are given by the same functions of the bare coupling constant λ_0 and the cutoff Λ. These functions make no reference to M. The dependence on M enters only when we remove the cutoff dependence by rescaling the fields and eliminating λ_0 in favor of the renormalized coupling λ. The renormalized Green's functions are numerically equal to the bare Green's functions, up to a rescaling by powers of the field strength renormalization Z:

$$\langle\Omega| T\phi(x_1)\phi(x_2)\cdots\phi(x_n)|\Omega\rangle = Z^{-n/2}\langle\Omega| T\phi_0(x_1)\phi_0(x_2)\cdots\phi_0(x_n)|\Omega\rangle. \tag{12.35}$$

The renormalized Green's functions could be defined equally well at another scale M', using a new renormalized coupling λ' and a new rescaling factor Z'.

Let us write more explicitly the effect of an infinitesimal shift of M. Let $G^{(n)}(x_1,\cdots,x_n)$ be the connected n-point function, computed in renormalized perturbation theory:

$$G^{(n)}(x_1,\cdots,x_n) = \langle\Omega| T\phi(x_1)\cdots\phi(x_n)|\Omega\rangle_{\text{connected}}. \tag{12.36}$$

Now suppose that we shift M by δM. There is a corresponding shift in the coupling constant and the field strength such that the bare Green's functions remain fixed:

$$\begin{aligned} M &\to M + \delta M,\\ \lambda &\to \lambda + \delta\lambda,\\ \phi &\to (1+\delta\eta)\phi. \end{aligned} \tag{12.37}$$

Then the shift in any renormalized Green's function is simply that induced by the field rescaling,

$$G^{(n)} \to (1+n\delta\eta)G^{(n)}.$$

If we think of $G^{(n)}$ as a function of M and λ, we can write this transformation as

$$dG^{(n)} = \frac{\partial G^{(n)}}{\partial M}\delta M + \frac{\partial G^{(n)}}{\partial\lambda}\delta\lambda = n\delta\eta G^{(n)}. \tag{12.38}$$

Rather than writing this relation in terms of $\delta\lambda$ and $\delta\eta$, it is conventional to define the dimensionless parameters

$$\beta \equiv \frac{M}{\delta M}\delta\lambda; \qquad \gamma \equiv -\frac{M}{\delta M}\delta\eta. \tag{12.39}$$

Making these substitutions in Eq. (12.38) and multiplying through by $M/\delta M$, we obtain

$$\left[M\frac{\partial}{\partial M} + \beta\frac{\partial}{\partial\lambda} + n\gamma\right]G^{(n)}(x_1,\cdots,x_n;M,\lambda) = 0. \tag{12.40}$$

The parameters β and γ are the same for every n, and must be independent of the x_i. Since the Green's function $G^{(n)}$ is renormalized, β and γ cannot depend on the cutoff, and hence, by dimensional analysis, these functions cannot depend on M. Therefore they are functions only of the dimensionless variable λ. We conclude that any Green's function of massless ϕ^4 theory must satisfy

$$\left[M\frac{\partial}{\partial M}+\beta(\lambda)\frac{\partial}{\partial\lambda}+n\gamma(\lambda)\right]G^{(n)}(\{x_i\};M,\lambda)=0. \tag{12.41}$$

This relation is called the Callan-Symanzik equation.‡ It asserts that there exist two universal functions $\beta(\lambda)$ and $\gamma(\lambda)$, related to the shifts in the coupling constant and field strength, that compensate for the shift in the renormalization scale M.

The preceding argument generalizes without difficulty to other massless theories with dimensionless couplings. In theories with multiple fields and couplings, there is a γ term for each field and a β term for each coupling. For example, we can define QED at zero electron mass by introducing a renormalization scale as in Eqs. (12.30). The renormalization conditions for the propagators are applied at $p^2=-M^2$, and those for the vertex at a point where all three invariants are of order $-M^2$. Then the renormalized Green's functions of this theory satisfy the Callan-Symanzik equation

$$\left[M\frac{\partial}{\partial M}+\beta(e)\frac{\partial}{\partial e}+n\gamma_2(e)+m\gamma_3(e)\right]G^{(n,m)}(\{x_i\};M,e)=0, \tag{12.42}$$

where n and m are, respectively, the number of electron and photon fields in the Green's function $G^{(n,m)}$ and γ_2 and γ_3 are the rescaling functions of the electron and photon fields.

Computation of β and γ

Before we work out the implications of the Callan-Symanzik equation, let us look more closely at the functions β and γ that appear in it. From their definitions (12.39), we see that they are proportional to the shift in the coupling constant and the shift in the field normalization, respectively, when the renormalization scale M is increased. The behavior of the coupling constant as a function of M is of particular interest, since it determines the strength of the interaction and the conditions under which perturbation theory is valid. We will see in the next section that the shift in the field strength is also reflected directly in the values of Green's functions.

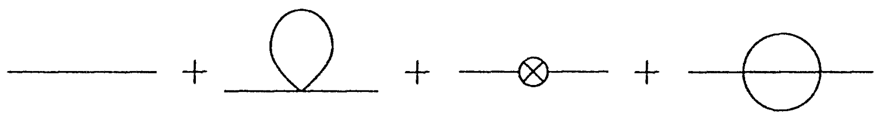
The easiest way to compute the Callan-Symanzik functions is to begin with explicit perturbative expressions for some conveniently chosen Green's functions. If we insist that these expressions satisfy the Callan-Symanzik equation, we will obtain equations that can be solved for β and γ. Because the

‡C. G. Callan, *Phys. Rev.* **D2**, 1541 (1970), K. Symanzik, *Comm. Math. Phys.* **18**, 227 (1970).

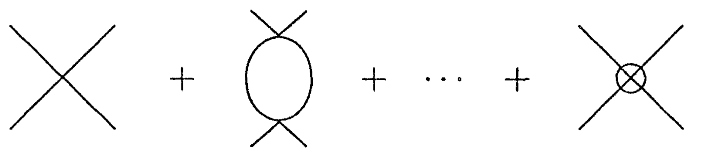
M dependence of a renormalized Green's function originates in the counterterms that cancel its logarithmic divergences, we will find that the β and γ functions are simply related to these counterterms, or equivalently, to the coefficients of the divergent logarithms. The precise formulae that relate β and γ to the counterterms will depend on the specific renormalization prescription and other details of the calculational scheme. At one-loop order, however, the expressions for β and γ are simple and unambiguous.

As a first example, let us calculate the one-loop contributions to $\beta(\lambda)$ and $\gamma(\lambda)$ in massless ϕ^4 theory. We can simplify the analysis by working in momentum space rather than coordinate space. Our strategy will be to apply the Callan-Symanzik equation to the diagrammatic expressions for the two- and four-point Green's functions.

The two-point function is given by

$G^{(2)}(p) =$  $+\cdots .$

In massless ϕ^4 theory, the one-loop propagator correction is completely canceled by the mass counterterm. Then the first nontrivial correction to the propagator comes from the two-loop diagram and its counterterm, and is of order λ^2. Meanwhile, the four-point function is given by

$G^{(4)} =$  $+\ \mathcal{O}(\lambda^3)\,,$

where we have omitted the canceled one-loop propagator corrections to the external legs. The diagrams of order λ^3 include nonvanishing two-loop propagator corrections to the external legs.

To calculate β, we apply the Callan-Symanzik equation to the four-point function:

$$\Big[M\frac{\partial}{\partial M} + \beta(\lambda)\frac{\partial}{\partial\lambda} + 4\gamma(\lambda)\Big]G^{(4)}(p_1,\ldots,p_4) = 0. \tag{12.43}$$

Borrowing our result (10.21) from Section 10.2, we can write $G^{(4)}$ as

$$G^{(4)} = \big[-i\lambda + (-i\lambda)^2\big[iV(s) + iV(t) + iV(u)\big] - i\delta_\lambda\big]\cdot\prod_{i=1,\ldots,4}\frac{i}{p_i^2},$$

where $V(s)$ represents the loop integral in (10.20). Our renormalization condition (12.30) requires that the correction terms cancel at $s = t = u = -M^2$. The order-λ^2 vertex counterterm is therefore

$$\delta_\lambda = (-i\lambda)^2\cdot 3V(-M^2) = \frac{3\lambda^2}{2(4\pi)^{d/2}}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{(x(1-x)M^2)^{2-d/2}}. \tag{12.44}$$

The last expression follows from setting $m = 0$ and $p^2 = -M^2$ in Eq. (10.23) for $V(p^2)$. In the limit as $d \to 4$, Eq. (12.44) becomes

$$\delta_\lambda = \frac{3\lambda^2}{2(4\pi)^2}\left[\frac{1}{2-d/2} - \log M^2 + \text{finite}\right], \tag{12.45}$$

where the finite terms are independent of M. This counterterm gives $G^{(4)}$ its M dependence:

$$M\frac{\partial}{\partial M}G^{(4)} = \frac{3i\lambda^2}{(4\pi)^2}.$$

Let us assume for the moment that $\gamma(\lambda)$ has no term of order λ; we will justify this in the next paragraph. Then the Callan-Symanzik equation (12.43) can be satisfied to order λ^2 only if the β function of ϕ^4 theory is given by

$$\beta(\lambda) = \frac{3\lambda^2}{16\pi^2} + \mathcal{O}(\lambda^3). \tag{12.46}$$

Next, consider the Callan-Symanzik equation for the two-point function:

$$\left[M\frac{\partial}{\partial M} + \beta(\lambda)\frac{\partial}{\partial\lambda} + 2\gamma(\lambda)\right]G^{(2)}(p) = 0. \tag{12.47}$$

Since, to one-loop order, there are no propagator corrections to $G^{(2)}$, no dependence on M or λ is introduced to order λ. Thus the γ function is zero to this order:

$$\gamma = 0 + \mathcal{O}(\lambda^2). \tag{12.48}$$

This justifies the assumption made in the previous paragraph. The two-loop propagator correction is divergent, and its counterterm contains a term of order λ^2 which depends on M. This contributes to the first term in Eq. (12.47). Since β is of order λ^2 and the corrections to $G^{(2)}$ are of order λ^2, the leading contributions to the second term in (12.47) are of order λ^3. Thus γ acquires a nonzero contribution in order λ^2. This leading contribution to γ is computed in Problem 13.2.

The preceding example illustrates how β and γ can be calculated in more general theories with dimensionless couplings. In such theories, the M dependence of Green's functions enters through the field-strength and vertex counterterms, which are used to subtract the divergent logarithms. The lowest-order expressions for β and γ can be computed directly from these counterterms, or from the coefficients of the divergent logarithms.

In any renormalizable massless scalar field theory, the two-point Green's function has the generic form

$$\begin{aligned} G^{(2)}(p) &= \text{———} \quad + \quad \text{(loop diagrams)} \quad + \quad \text{—}\otimes\text{—} \quad + \cdots \\ &= \frac{i}{p^2} + \frac{i}{p^2}\left(A\log\frac{\Lambda^2}{-p^2} + \text{finite}\right) + \frac{i}{p^2}(ip^2\delta_Z)\frac{i}{p^2} + \cdots. \end{aligned} \tag{12.49}$$

The M dependence of this expression, to lowest order, comes entirely from the counterterm δ_Z. Applying the Callan-Symanzik equation to $G^{(2)}(p)$, and neglecting the β term (which is always smaller by at least one power of the coupling constant), we find

$$-\frac{i}{p^2}M\frac{\partial}{\partial M}\delta_Z + 2\gamma\frac{i}{p^2} = 0,$$

or

$$\gamma = \frac{1}{2}M\frac{\partial}{\partial M}\delta_Z \qquad \text{(to lowest order).} \tag{12.50}$$

To make this result more explicit, note that the counterterm must be

$$\delta_Z = A\log\frac{\Lambda^2}{M^2} + \text{finite}$$

in order to cancel the divergent logarithm in $G^{(2)}$. Thus γ is simply the coefficient of the logarithm:

$$\gamma = -A \qquad \text{(to lowest order).} \tag{12.51}$$

In most theories (e.g., Yukawa theory or QED), the first logarithmic divergence in δ_Z occurs at the one-loop level. However, even in ϕ^4 theory, formulae (12.50) and (12.51) are true for the first nonvanishing term in δ_Z, in this case the two-loop contribution.* By replacing the scalar field propagator (i/p^2) with a fermion propagator ($i/\not{p}$), we could repeat this argument line for line to compute the γ function for a fermion field in terms of its field strength counterterm δ_Z.

We can derive similar expressions for the β function of a generic dimensionless coupling constant g, associated with an n-point vertex. Taking propagator corrections into account, the full connected Green's function, to one-loop order, has the general form

$$\begin{aligned} G^{(n)} &= \begin{pmatrix}\text{tree-level}\\ \text{diagram}\end{pmatrix} + \begin{pmatrix}\text{1PI loop}\\ \text{diagrams}\end{pmatrix} + \begin{pmatrix}\text{vertex}\\ \text{counterterm}\end{pmatrix} + \begin{pmatrix}\text{external leg}\\ \text{corrections}\end{pmatrix} \\ &= \Big(\prod_i \frac{i}{p_i^2}\Big)\Big[-ig - iB\log\frac{\Lambda^2}{-p^2} - i\delta_g + (-ig)\sum_i\Big(A_i\log\frac{\Lambda^2}{-p_i^2} - \delta_{Zi}\Big)\Big] \\ &\qquad + \text{finite terms.} \end{aligned} \tag{12.52}$$

In this expression, p_i are the momenta on the external legs, and p^2 represents a typical invariant built from these momenta. We assume that renormalization conditions are applied at a point where all such invariants are spacelike and of order $-M^2$. The M dependence of this expression comes from the counterterms δ_g and δ_{Zi}. Applying the Callan-Symanzik equation, we obtain

$$M\frac{\partial}{\partial M}\Big(\delta_g - g\sum_i \delta_{Zi}\Big) + \beta(g) + g\sum_i \frac{1}{2}M\frac{\partial}{\partial M}\delta_{Zi} = 0,$$

*At one loop, formula (12.33) implies that we can also identify A as the coefficient of $2/(4-d)$ in the 1PI self-energy, in the limit $d \to 4$. This relation changes in higher loops. However, Eq. (12.50) remains correct.

or

$$\beta(g) = M\frac{\partial}{\partial M}\Big(-\delta_g + \frac{1}{2}g\sum_i \delta_{Z i}\Big) \qquad \text{(to lowest order).} \tag{12.53}$$

To be more explicit, we note that

$$\delta_g = -B\log\frac{\Lambda^2}{M^2} + \text{finite.}$$

Thus the β function is just a combination of the coefficients of the divergent logarithms:

$$\beta(g) = -2B - g\sum_i A_i \qquad \text{(to lowest order).} \tag{12.54}$$

Notice that the finite parts of counterterms are independent of M and therefore never contribute to β or γ. This means that, to compute the leading terms in the Callan-Symanzik functions, we needn't be too precise in specifying renormalization conditions: Any momentum scale of order M^2 will yield the same results. The divergent parts of the counterterms can be estimated simply by setting all invariants inside of logarithms equal to M^2, as we did above in our expression for the n-point Green's function.

As in the computation of γ, this argument can be applied almost without change to coupling constants for fields with spin. In Yukawa theory, for example, we consider the three-point function with one incoming fermion, one outgoing fermion, and one scalar, with momenta p_1, p_2, and p_3, respectively. Then the tree-level expression for the three-point function is

$$\frac{i}{\not{p}_1}\frac{i}{\not{p}_2}\frac{1}{p_3^2}(-ig). \tag{12.55}$$

The one-loop corrections replace the quantity $(-ig)$ by the expression in brackets in Eq. (12.52). Then formulae (12.53) and (12.54) hold also for the β function of this theory.

Similar expressions also apply in QED, though there are a number of small complications. The first comes in computing the γ function for the photon propagator. In Eq. (7.74), we saw that the general form of the photon propagator in Feynman gauge is

$$D^{\mu\nu}(q) = D(q)\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\Big) + \frac{-i}{q^2}\frac{q^\mu q^\nu}{q^2}. \tag{12.56}$$

The coefficient of the last term in (12.56) depends on the gauge. Fortunately, this term drops out of all gauge-invariant observables. Thus it makes sense to concentrate on the first term, projecting all external photons onto their transverse components. Projecting the photon propagator, we see that $D(q)$ satisfies the Callan-Symanzik equation. Since the corrections to this function have the form (12.49), the arguments following that formula are valid for photons as well as for electrons and scalars. Thus, to leading order,

$$\gamma_2 = \frac{1}{2}M\frac{\partial}{\partial M}\delta_2, \qquad \gamma_3 = \frac{1}{2}M\frac{\partial}{\partial M}\delta_3, \tag{12.57}$$

where δ_2 and δ_3 are the counterterms defined in Section 10.3.

Similarly, we may consider the three-point connected Green's function $\langle\overline{\psi}(p_1)\psi(p_2)A_\mu(q)\rangle$, projected onto transverse components of the photon. At leading order, this function equals

$$\frac{i}{\not{p}_1}(-ie\gamma^\mu)\frac{i}{\not{p}_2}\frac{-i}{q^2}\Big(g^{\mu\nu}-\frac{q^\mu q^\nu}{q^2}\Big).$$

The divergent one-loop corrections have the same form, with $(-ie)$ replaced by logarithmically divergent terms. Thus, Eq. (12.53) gives the lowest-order expression for the β function:

$$\beta(e)=M\frac{\partial}{\partial M}\big(-\delta_1+e\delta_2+\frac{e}{2}\delta_3\big). \tag{12.58}$$

To find explicit expressions for the Callan-Symanzik functions of QED, we must write expressions for the counterterms δ_1, δ_2, δ_3. In Section 10.3, we evaluated these counterterms using on-shell renormalization conditions with massive fermions. We must now re-evaluate these terms for massless fermions and renormalization at $-M^2$. Fortunately, we need only evaluate the logarithmically divergent pieces of these counterterms, which are identical in the two cases. Reading from Eqs. (10.43) and (10.44), we find

$$\begin{aligned} e^{-1}\delta_1=\delta_2 &= -\frac{e^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}} + \text{ finite},\\ \delta_3 &= -\frac{e^2}{(4\pi)^2}\frac{4}{3}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}} + \text{ finite}. \end{aligned} \tag{12.59}$$

Using formulae (12.57) and (12.59), we obtain at leading order

$$\gamma_2(e)=\frac{e^2}{16\pi^2},\qquad \gamma_3(e)=\frac{e^2}{12\pi^2}. \tag{12.60}$$

And from Eq. (12.58), we find

$$\beta(e)=\frac{e^3}{12\pi^2}. \tag{12.61}$$

It is important to remember that the expression we have used for δ_2 explicitly assumes the use of Feynman gauge. In fact, γ_2 depends on the gauge parameter, and this makes sense, because Green's functions of individual ψ and $\overline{\psi}$ fields are not gauge invariant. On the other hand, the QED vacuum polarization, and therefore γ_3 and β, are gauge invariant.

The Meaning of β and γ

We can obtain a deeper insight into the nature of β and γ by expressing them in terms of the parameters of bare perturbation theory: Z, λ_0, and Λ for the case of ϕ^4 theory.

First recall that the bare and renormalized field are related by

$$\phi(p) = Z(M)^{-1/2}\phi_0(p). \tag{12.62}$$

This equation expresses the dependence of the field rescaling on M. If M is increased by δM, the renormalized field is shifted by

$$\delta\eta = \frac{Z(M+\delta M)^{-1/2}}{Z(M)^{-1/2}} - 1.$$

Hence our original definition (12.39) of γ gives us immediately

$$\gamma(\lambda) = \frac{1}{2}\frac{M}{Z}\frac{\partial}{\partial M}Z. \tag{12.63}$$

Since $\delta_Z = Z - 1$ (Eq. (10.17)), this formula is in agreement with (12.50) to leading order. Formula (12.63), however, is an exact relation. This expression clarifies the relation of γ to the field strength rescaling. However, it obscures the fact that γ is independent of the cutoff Λ. To understand this aspect of γ, we have to go back to the original definition of this function in terms of renormalized Green's functions, whose cutoff independence follows from the renormalizability of the theory.

Similarly, we can find an instructive expression for β in terms of the parameters of bare perturbation theory. Our original definition of β in Eq. (12.39) made use of a quantity $\delta\lambda$, defined to be the shift of the renormalized coupling λ needed to preserve the values of the bare Green's functions when the renormalization point is shifted infinitesimally. Since the bare Green's functions depend on the bare coupling λ_0 and the cutoff, this definition can be rewritten as

$$\beta(\lambda) = M\frac{\partial}{\partial M}\lambda\bigg|_{\lambda_0,\Lambda}. \tag{12.64}$$

Thus the β function is the rate of change of the renormalized coupling at the scale M corresponding to a fixed bare coupling. Recalling our analysis in Section 12.1, it is tempting to associate $\lambda(M)$ with the coupling constant λ' obtained by integrating out degrees of freedom down to the scale M. With this correspondence, the β function is just the rate of the renormalization group flow of the coupling constant λ. A positive sign for the β function indicates a renormalized coupling that increases at large momenta and decreases at small momenta. We can see explicitly that this relation works for ϕ^4 theory, to leading order in λ, by comparing Eqs. (12.28) and (12.46). We will justify this correspondence further in the following section.

The equality of the exact formula (12.64) with the first-order formula (12.53) again follows from the counterterm definitions (10.17). As with (12.63), it is not obvious that this formula for $\beta(\lambda)$ is independent of Λ, but that fact again follows from renormalizability. Conversely, it is possible to prove the renormalizability of ϕ^4 theory by demonstrating, order by order in perturbation theory, that expressions (12.63) and (12.64) are independent of Λ.[†]

12.3 Evolution of Coupling Constants

Now that we have discussed all of the ingredients of the Callan-Symanzik equation, let us investigate its implications. We begin by finding the explicit solution to the Callan-Symanzik equation for the simplest situation, the two-point Green's function of a scalar field theory. This solution will clarify the physical implications of the equation. In particular, it will cement the relation suggested at the end of the previous section, which identifies the β function with the rate of the renormalization group flow of the coupling constant. We will then use this relation to discuss the qualitative features of the renormalization group flow in renormalizable field theories.

Solution of the Callan-Symanzik Equation

We would like to solve the Callan-Symanzik equation for the two-point Green's function, $G^{(2)}(p)$, in a theory with a single scalar field. Since $G^{(2)}(p)$ has dimensions of $(\text{mass})^{-2}$, we can express its dependence on p and M as

$$G^{(2)}(p) = \frac{i}{p^2} g(-p^2/M^2). \tag{12.65}$$

This equation allows us to trade the derivative with respect to M for a derivative with respect to p^2. For the remainder of this chapter, we will use the variable p to represent the magnitude of the spacelike momentum: $p = (-p^2)^{1/2}$. Then we can rewrite the Callan-Symanzik equation as

$$\left[p\frac{\partial}{\partial p} - \beta(\lambda)\frac{\partial}{\partial \lambda} + 2 - 2\gamma(\lambda)\right] G^{(2)}(p) = 0. \tag{12.66}$$

In free field theory, β and γ vanish and we recover the trivial result

$$G^{(2)}(p) = \frac{i}{p^2}. \tag{12.67}$$

In an interacting theory, β and γ are nonzero functions of λ. However, it is still possible to write the explicit solution to the Callan-Symanzik equation, using the method of characteristics. Equivalently (for those not well versed in the theory of partial differential equations), we will apply a lovely hydrodynamic-bacteriological analogy due to Sidney Coleman.[‡] Imagine a narrow pipe running in the x direction, containing a fluid whose velocity

[†]Callan has given a beautiful proof of the renormalizability of ϕ^4 theory, based on proving that the Callan-Symanzik equation holds order by order in λ, in his article in *Methods in Field Theory*, R. Balian and J. Zinn-Justin, eds. (North Holland, Amsterdam, 1976).

[‡]Coleman (1985), chap. 3.

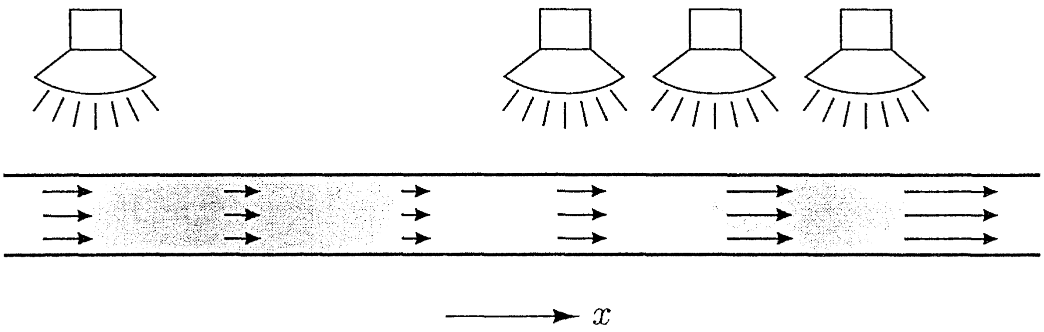


Figure 12.3. Coleman's bacteriological analogy to the Callan-Symanzik equation. The pipe is inhabited by bacteria with a given initial density $D_i(x)$. The growth rate (determined by the illumination) and flow velocity are given functions of x. The problem is to determine the density $D(t, x)$ at all subsequent times.

is $v(x)$, as shown in Fig. 12.3. The pipe is inhabited by bacteria, whose density is $D(t, x)$ and whose rate of growth is $\rho(x)$. Then the future behavior of the function $D(t, x)$ is governed by the differential equation

$$\left[\frac{\partial}{\partial t} + v(x)\frac{\partial}{\partial x} - \rho(x)\right] D(t, x) = 0. \tag{12.68}$$

The second term allows for the fact that the bacteria are swept along with the fluid, so their present density here determines their future density not here, but some distance ahead. This equation is identical to Eq. (12.66), with the replacements

$$\begin{aligned} \log(p/M) &\leftrightarrow t, \\ \lambda &\leftrightarrow x, \\ -\beta(\lambda) &\leftrightarrow v(x), \\ 2\gamma(\lambda){-}2 &\leftrightarrow \rho(x), \\ G^{(2)}(p, \lambda) &\leftrightarrow D(t, x). \end{aligned} \tag{12.69}$$

Now suppose we know the initial concentration of the bacteria: $D(t, x) = D_i(x)$ at time $t = 0$. Then we can determine the concentration of bacteria in a fluid element at the point x at any later time by computing the history of that fluid element and then integrating the rate of growth along that path. Consider the fluid element that is at x at the time t. We can find out where it was at time zero by integrating its motion backward in time. The position of this element at time $t = 0$ is given by $\bar{x}(t; x)$, which satisfies the differential equation

$$\frac{d}{dt'}\bar{x}(t'; x) = -v(\bar{x}), \qquad \text{with} \qquad \bar{x}(0, x) = x. \tag{12.70}$$

Then, immediately,

$$\begin{aligned} D(t,x) &= D_i\big(\bar{x}(t;x)\big)\cdot \exp\Big(\int_0^t dt'\,\rho\big(\bar{x}(t';x)\big)\Big) \\ &= D_i\big(\bar{x}(t;x)\big)\cdot \exp\Big(\int_{\bar{x}(t)}^{x} dx'\,\frac{\rho(x')}{v(x')}\Big). \end{aligned} \tag{12.71}$$

Now bring this solution back to our field theory problem by replacing each bacteriological parameter with its corresponding field theory parameter. The time $t=0$ corresponds to $-p^2 = M^2$, and the initial concentration $D_i(x)$ becomes an unknown function $\hat{\mathcal{G}}(\lambda)$. Then

$$G^{(2)}(p,\lambda) = \hat{\mathcal{G}}\big(\bar{\lambda}(p;\lambda)\big)\cdot \exp\Big(-\int_{p'=M}^{p'=p} d\log(p'/M)\cdot 2\big[1-\gamma\big(\bar{\lambda}(p';\lambda)\big)\big]\Big), \tag{12.72}$$

where $\bar{\lambda}(p;\lambda)$ solves

$$\frac{d}{d\log(p/M)}\bar{\lambda}(p;\lambda) = \beta(\bar{\lambda}), \qquad \bar{\lambda}(M;\lambda) = \lambda. \tag{12.73}$$

This differential equation describes the flow of a modified coupling constant $\bar{\lambda}(p;\lambda)$ as a function of momentum. The rate of this flow is just the β function. Thus, this flow is strongly reminiscent of the dependence of the renormalized coupling on the renormalization scale given by Eq. (12.64). We will refer to $\bar{\lambda}(p)$ as the *running coupling constant.* Its equation (12.73) is often called the *renormalization group equation.*

One can check directly that (12.72) solves the Callan-Symanzik equation by using the identity

$$\int_{\lambda}^{\bar{\lambda}} \frac{d\lambda'}{\beta(\lambda')} = \int_{p'=M}^{p'=p} d\log(p'/M), \tag{12.74}$$

from which it follows that

$$\Big(p\frac{\partial}{\partial p} - \beta(\lambda)\frac{\partial}{\partial\lambda}\Big)\bar{\lambda} = 0. \tag{12.75}$$

A convenient way of writing the solution (12.72) is

$$G^{(2)}(p,\lambda) = \frac{i}{p^2}\mathcal{G}\big(\bar{\lambda}(p;\lambda)\big)\cdot \exp\Big(2\int_M^p d\log(p'/M)\,\gamma\big(\bar{\lambda}(p';\lambda)\big)\Big), \tag{12.76}$$

in which $\mathcal{G}(\bar{\lambda})$ is a function that must be determined. This function cannot be determined from the general principles of renormalization theory. Instead, we must compute $G^{(2)}(p)$ as a perturbation series in λ and match terms to the expansion of (12.76) as a series in the same parameter. For the two-point function in ϕ^4 theory, this matching is rather trivial: $\mathcal{G}(\bar{\lambda}) = 1 + \mathcal{O}(\bar{\lambda}^2)$.

The preceding analysis can be applied to any family of Green's functions that are related by uniform rescaling of the momenta. Consider, for example, the connected four-point function of ϕ^4 theory evaluated at spacelike momenta p_i such that $p_i^2 = -P^2$, $p_i \cdot p_j = 0$, so that s, t, and u are of order $-P^2$. To leading order in perturbation theory, this function is given by

$$G^{(4)}(P) = \left(\frac{i}{P^2}\right)^4 (-i\lambda). \tag{12.77}$$

Using the fact that $G^{(4)}$ has dimensions of $(\text{mass})^{-8}$, we can exchange M for P in the Callan-Symanzik equation and write this equation as

$$\left[P\frac{\partial}{\partial P} - \beta(\lambda)\frac{\partial}{\partial \lambda} + 8 - 4\gamma(\lambda)\right] G^{(4)}(P;\lambda) = 0. \tag{12.78}$$

The solution to this equation is

$$G^{(4)}(P;\lambda) = \frac{1}{P^8}\mathcal{G}^{(4)}(\bar{\lambda}(p;\lambda)) \cdot \exp\left(4\int_M^p d\log(p'/M)\, \gamma(\bar{\lambda}(p';\lambda))\right). \tag{12.79}$$

This formula must agree with (12.77) to leading order in λ; this matching requires that

$$\mathcal{G}^{(4)}(\bar{\lambda}(p;\lambda)) = -i\bar{\lambda} + \mathcal{O}(\bar{\lambda}^2). \tag{12.80}$$

We can now see the physical implication of the Callan-Symanzik equation. The ordinary Feynman perturbation series for a Green's function depends both on the coupling constant λ and on the dimensionless parameter $\log(-p^2/M^2)$. The perturbation theory can be badly behaved even when λ is small if the ratio p^2/M^2 is large. The solutions (12.76) and (12.79) reorganize this dependence into a function of the running coupling constant and an exponential scale factor. We consider these two pieces in turn.

The first factor in Eqs. (12.76) and (12.79) is a function of the running coupling constant, evaluated at the momentum scale p. If p were of order M, the renormalization scale, this function would essentially be the ordinary perturbative evaluation of the Green's function. The results (12.76) and (12.79) instruct us to make use of this same expression at the scale p, but to replace λ with a new coupling constant $\bar{\lambda}$ appropriate to that scale. Thus, the running coupling constant $\bar{\lambda}(p)$ is precisely the effective coupling constant of the renormalization group flow. This interpretation is particularly clear in the solution (12.79) for $G^{(4)}(P)$, since this function directly measures the strength of the ϕ^4 coupling constant.

The exponential factor in Eqs. (12.76) and (12.79) has an equally simple interpretation: It is the accumulated field strength rescaling of the correlation function from the reference point M to the actual momentum p at which the

Green's function is evaluated. This factor receives a multiplicative contribution from each intermediate scale between M and p. Each of these contributions is, appropriately, computed using the running coupling constant at that particular scale.

As a check on these formal arguments, we can use the explicit form of the β function of ϕ^4 theory found in Eq. (12.46) and the renormalization group equation (12.73) to evaluate the running coupling constant of ϕ^4 theory. This running coupling constant satisfies the differential equation

$$\frac{d}{d\log(p/M)}\bar{\lambda} = \frac{3\bar{\lambda}^2}{16\pi^2}, \qquad \text{with} \qquad \bar{\lambda}(M;\lambda) = \lambda. \tag{12.81}$$

Integrating, we find

$$\left(\frac{3}{16\pi^2}\right)^{-1}\left[\frac{1}{\lambda} - \frac{1}{\bar{\lambda}}\right] = \log\frac{p}{M},$$

and thus,

$$\bar{\lambda}(p) = \frac{\lambda}{1-(3\lambda/16\pi^2)\log(p/M)}. \tag{12.82}$$

Many properties of the solution to the Callan-Symanzik equation are visible in this relation. First, the expansion of this formula for $\bar{\lambda}$ to order λ^2 agrees precisely with Eq. (12.28), the rate of the renormalization group flow from Wilson's method. Second, this expression for the running coupling constant goes to zero at a logarithmic rate as $p \to 0$. This coincides with our expectation that a positive value for the β function should imply an effective coupling that becomes stronger at large momenta and weaker at small momenta.

If we expand the running coupling constant $\bar{\lambda}(p)$ in powers of λ, we find that the successive powers of the coupling constant are multiplied by powers of logarithms,

$$\lambda^{n+1}(\log p/M)^n,$$

which become large and invalidate a simple perturbation expansion for p much greater or much less than M. We have seen this problem of large logarithms arising several times in our diagram calculations, and we have remarked on it specifically as a problem in the discussion following Eq. (11.81). We now see that the renormalization group gives a partial solution to this problem. In this example, and in many others that we will study, the Callan-Symanzik equation tells us how to sum these large logarithms into the running coupling constant and multiplicative rescalings. If the running coupling constant becomes large, as happens in ϕ^4 theory for $p \to \infty$, the perturbation expansion will break down anyway, and we will need more advanced methods. However, if the running coupling constant becomes small, as for ϕ^4 theory as $p \to 0$, we will have successfully organized the powers of logarithms into a meaningful and controlled expression. The specific problem posed at the end of Section 11.4 will be solved explicitly by this method in Section 13.2.

An Application to QED

For a more concrete application of the Callan-Symanzik equation, we can look again at the electromagnetic potential between static charges, $V(\mathbf{x})$, which we studied in Section 7.5. At very short distances or at large momenta, we can ignore the electron mass in the computation of QED corrections to this potential. In this approximation, the potential should obey the Callan-Symanzik equation of massless QED. We could write this equation either for $V(\mathbf{x})$ itself or for its Fourier transform; we choose to work in Fourier space in order to make contact more easily with the results of Section 7.5.

We define the massless limit of QED by specifying a renormalization scale M at which the renormalized coupling e_r is defined. If M is taken close to the electron mass m, at the point where the massless approximation is just becoming valid, then the value of e_r will be close to the physical electron charge e. The potential between static charges is a measurable energy, so its normalization is unambiguous and is not shifted from one renormalization point to another. Thus the Callan-Symanzik equation for the Fourier transform of the potential has no γ term, being simply

$$\left[M\frac{\partial}{\partial M}+\beta(e_r)\frac{\partial}{\partial e_r}\right]V(q;M,e_r)=0. \tag{12.83}$$

The Fourier transform of the potential has dimensions of $(\text{mass})^{-2}$, so we can trade dependence on M for dependence on q as in the scalar field theory discussion above. This gives

$$\left[q\frac{\partial}{\partial q}-\beta(e_r)\frac{\partial}{\partial e_r}+2\right]V(q;M,e_r)=0. \tag{12.84}$$

Equation (12.84) is almost the same as Eq. (12.66), so we can immediately write down the solution as a special case of (12.76):

$$V(q,e_r)=\frac{1}{q^2}\mathcal{V}\big(\bar{e}(q;e_r)\big), \tag{12.85}$$

where $\bar{e}(q)$ is the solution of the renormalization group equation

$$\frac{d}{d\log(q/M)}\bar{e}(q;e_r)=\beta(\bar{e}),\qquad \bar{e}(M;e_r)=e_r. \tag{12.86}$$

By comparing this formula for $V(q)$ to the leading-order result

$$V(q)\approx\frac{e^2}{q^2},$$

we can identify $\mathcal{V}(\bar{e})=\bar{e}^2+\mathcal{O}(\bar{e}^4)$. Then

$$V(q,e_r)=\frac{\bar{e}^2(q;e_r)}{q^2}, \tag{12.87}$$

up to corrections that are suppressed by powers of e_r^2 and contain no compensatory large logarithms of q/M.

To turn Eq. (12.87) into a completely explicit formula, we need only solve the renormalization group equation (12.86). Using the QED β function (12.61), we can integrate (12.86) to find

$$\frac{12\pi^2}{2}\left(\frac{1}{e_r^2}-\frac{1}{\bar{e}^2}\right)=\log\frac{q}{M}.$$

This simplifies to

$$\bar{e}^2(q)=\frac{e_r^2}{1-(e_r^2/6\pi^2)\log(q/M)}. \tag{12.88}$$

This result is almost identical to the formula for the effective electric charge that we found in Eq. (7.96). To cement the identification, set M to be of order the electron mass, $M^2 = Am^2$, and approximate e_r at this point by e, with $\alpha = e^2/4\pi$. Then Eq. (12.88) takes the form

$$\bar{\alpha}(q)=\frac{\alpha}{1-(\alpha/3\pi)\log(-q^2/Am^2)}. \tag{12.89}$$

The particular choice $A = \exp(5/3)$ reproduces Eq. (7.96). Of course, we could not find this exact correspondence without the detailed one-loop calculation of Section 7.5. Nevertheless, our present analysis produces the correct asymptotic formula for the effective charge. Furthermore, our present formalism can be applied to any renormalizable quantum field theory; it does not rely on the special symmetries of QED that we exploited in Section 7.5.

Alternatives for the Running of Coupling Constants

Now that we have computed the behavior of the running coupling constant in two specific quantum field theories, let us consider more generally what behaviors of the running coupling constant are possible in principle. We continue to restrict our discussion to renormalizable theories in the massless limit, with a single dimensionless coupling constant λ.

By the arguments of the previous section, the Green's functions in any such theory obey a Callan-Symanzik equation. The solution of this equation depends on a running coupling constant, $\bar{\lambda}(p)$, which satisfies a differential equation

$$\frac{\partial}{\partial\log(p/M)}\bar{\lambda}=\beta(\bar{\lambda}), \tag{12.90}$$

in which the function $\beta(\lambda)$ is computable as a power series in the coupling constant. In the examples we have just discussed, the leading coefficient in this power series was positive. However, as a matter of principle, three behaviors are possible in the region of small λ:

$$\begin{aligned}
&(1)\quad \beta(\lambda)>0;\\
&(2)\quad \beta(\lambda)=0;\\
&(3)\quad \beta(\lambda)<0.
\end{aligned}$$

Examples of quantum fields are known that exhibit each of these behaviors.

We have already seen how, in theories of the first class, the running coupling constant goes to zero in the infrared, leading to definite predictions about the small-momentum behavior of the theory. However, the running coupling constant becomes large in the region of high momenta. Thus the short-distance behavior of the theory cannot be computed using Feynman diagram perturbation theory. In fact, in the examples studied above, the coupling constant formally goes to infinity at a large but finite value of the momentum; thus it is not even clear that these theories possess a nontrivial limit $\Lambda \to \infty$. A Feynman diagram analysis is useful in such theories if one is mainly interested in large-distance or macroscopic behavior. In Chapter 13 we will use this observation to solve problems in the statistical mechanics of systems with critical points.

In theories of the second class, the coupling constant does not flow. In these theories, the running coupling constant is independent of the momentum scale, and thus equal to the bare coupling. This means that there can be no ultraviolet divergences in the relation of coupling constants. The only possible ultraviolet divergences in such theories are those associated with field rescaling, which automatically cancel in the computation of S-matrix elements. Such theories are called *finite* quantum field theories. Before the emergence of our modern understanding of renormalization, these theories would have been embraced as the solution to the problem of ultraviolet infinities. But in fact the known finite field theories in four dimensions are very special constructions—the so-called gauge theories with extended supersymmetry—with no known physical application.

In theories of the third class, the running coupling constant becomes large in the large-distance regime and becomes small at large momenta or short distances. Imagine, for instance, that the sign of the QED β function were reversed:

$$\beta(e) = -\tfrac{1}{2}Ce^3. \tag{12.91}$$

Then, following our earlier analysis, we would have

$$\bar{e}^2(p) = \frac{e^2}{1 + Ce^2 \log(p/M)}. \tag{12.92}$$

This coupling constant tends to zero at a logarithmic rate as the momentum scale increases. Such theories are called *asymptotically free*. In theories of this class, the short-distance behavior is completely solvable by Feynman diagram methods. Though ultraviolet divergences appear in every order of perturbation theory, the renormalization group tells us that the sum of these divergences is completely harmless. If we interpret these theories in terms of a bare coupling e_b and a finite cutoff Λ, the result (12.92) indicates that there is a smooth limit in which e_b tends to zero as Λ tends to infinity. Thus, asymptotically free theories give another, more sophisticated, resolution of the problem of ultraviolet divergences. In Chapter 17, we will see that asymptotic freedom plays an essential role in the formulation of a field theory that describes the strong interactions of elementary particle physics.

Now that we have enumerated the possibilities for the renormalization group flow in the region of weak coupling, let us turn our attention to the region of strong coupling. Here we will not be able to compute the β function quantitatively, but we can at least use the renormalization group equation to discuss qualitatively the possibilities for the coupling constant flow. All of our explicit solutions for running coupling constants—Eqs. (12.82), (12.88), and (12.92)—predict that the running coupling becomes infinite at a finite value of the momentum p. For example, according to Eq. (12.82), the running coupling constant of ϕ^4 theory should diverge at

$$p \sim M \exp\left(\frac{16\pi^2}{3\lambda}\right). \tag{12.93}$$

It is possible that this is the true behavior of the quantum field theory, but we have not proved this, because when the running coupling constant becomes large, the approximation we have made, ignoring the higher-order terms in the β function, is no longer valid. It is a logical possibility that the higher terms of the β function are negative, so that the β function has the form shown in Fig. 12.4(a). In this case the β function has a zero at a nonzero value λ_*. When $\bar{\lambda}$ approaches this value, the renormalization group flow slows to a halt; thus $\lambda = \lambda_*$ would be a nontrivial fixed point of the renormalization group. In this model, the running coupling constant $\bar{\lambda}$ tends to λ_* in the limit of large momentum.

For the specific case of ϕ^4 theory in four dimensions, we have strong evidence from numerical studies that there is no such nontrivial fixed point. However, we will soon demonstrate that there is a nontrivial fixed point in ϕ^4 theory in $d < 4$, and many more examples are known. It is thus worthwhile to explore the implications of a fixed point in the renormalization group flow.

For a β function of the form of Fig. 12.4(a), the β function behaves in the vicinity of the fixed point as

$$\beta \approx -B(\lambda - \lambda_*), \tag{12.94}$$

where B is a positive constant. For $\bar{\lambda}$ near λ_*,

$$\frac{d}{d\log p}\bar{\lambda} \approx -B(\bar{\lambda} - \lambda_*). \tag{12.95}$$

The solution of this equation is

$$\bar{\lambda}(p) = \lambda_* + C\left(\frac{M}{p}\right)^B. \tag{12.96}$$

Thus, $\bar{\lambda}$ indeed tends to λ_* as $p \to \infty$, and the rate of approach is governed by the slope of the β function at the fixed point.

This behavior has a dramatic consequence for the exact solution (12.72) of the Callan-Symanzik equation for $G(p)$. For p sufficiently large, the integral

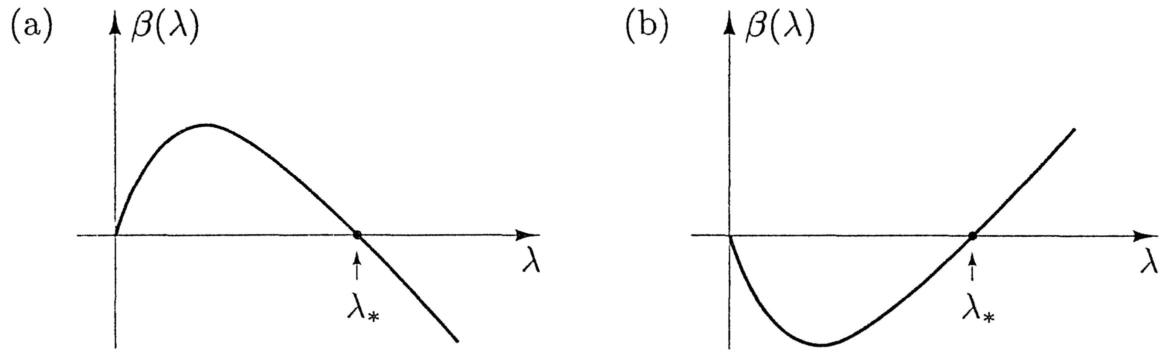


Figure 12.4. Possible forms of the β function with nontrivial zeros: (a) ultraviolet-stable fixed point; (b) infrared-stable fixed point.

in the exponential factor in this equation will be dominated by values of p for which $\bar{\lambda}(p)$ is close to λ_*. Then

$$\begin{aligned} G(p) &\approx \mathcal{G}(\lambda_*) \exp\Big[-\Big(\log\frac{p}{M}\Big)\cdot 2\big(1-\gamma(\lambda_*)\big)\Big] \\ &\approx C\cdot\Big(\frac{1}{p^2}\Big)^{1-\gamma(\lambda_*)}. \end{aligned} \tag{12.97}$$

Thus the two-point correlation function returns to the form of a simple scaling law, but with a power law different from that expected by dimensional analysis. At the fixed point we have a scale-invariant quantum field theory in which the interactions of the theory affect the law of rescaling. The shift of the exponent $\gamma(\lambda_*)$ is called the *anomalous dimension* of the scalar field. By convention, the function $\gamma(\lambda)$ is often called the anomalous dimension even if there is no fixed point in the theory.

A similar behavior is possible in an asymptotically free theory. If the β function has the form shown in Fig. 12.4(b), the running coupling constant will tend to a fixed point λ_* as $p \to 0$. The two-point correlation function of fields $G(p)$ will tend to a power law as in (12.97) for asymptotically small momenta. The two cases shown in Figs. 12.4(a) and (b) are called, respectively, *ultraviolet-stable* and *infrared-stable* fixed points.

In the previous section, we saw that the leading-order expressions for the Callan-Symanzik functions β and γ are related in a simple way to the ultraviolet divergent parts of the one-loop counterterms. However, we noted that, in higher orders of perturbation theory, β and γ depend on the specific renormalization conventions used to define the Green's functions. Still, there are some properties of these functions that are independent of any convention. The coefficient of the logarithm in the denominator of such expressions as (12.82) or (12.89) can be determined unambiguously from experiments that measure this coupling constant. This confirms the convention independence of the first β-function coefficient. Experiments sensitive to the coupling constant can also determine the existence of a zero of the β function at strong coupling, and the rate of approach to this asymptote. Thus the existence of a zero of the β function (but not necessarily the value of λ_*), the slope B at the zero, and the value of the anomalous dimension at the fixed point should all be independent of the conventions used to compute β and γ.

12.4 Renormalization of Local Operators

The analysis of the previous two sections has been restricted to quantum field theories with only dimensionless coefficients, that is, strictly renormalizable field theories in the massless limit. It is not difficult to generalize this formalism to theories with mass terms and other operators whose coefficients have mass dimension. However, it is worthwhile to first devote some attention to an intermediate step, by analyzing the renormalization group properties of matrix elements of local operators. This is an interesting problem in its own right, and we will devote considerable space to the applications of this formalism in Chapter 18.

Matrix elements of local operators appear often in quantum field theory calculations. Typically one considers a set of interacting particles that couple weakly to an additional particle, which mediates new forces. Consider, for example, the theory of strongly interacting quarks perturbed by the effects of weak decay processes. The weak interaction is mediated by a massive vector boson, the W. Let us write the interaction of the quarks with the W very schematically as

$$\delta\mathcal{L} = \frac{g}{\sqrt{2}} W_\mu \bar{\psi}\gamma^\mu(1-\gamma^5)\psi, \tag{12.98}$$

and assign the W boson the propagator

$$\frac{-ig^{\mu\nu}}{q^2 - m_W^2 + i\epsilon}. \tag{12.99}$$

(We will discuss this interaction more correctly in Section 18.2 and in Chapter 20.) Exchange of a W boson leads to the interaction shown in Fig. 12.5. For momentum transfers small compared to m_W, we can ignore the q^2 in the W propagator and write this interaction as the matrix element of the operator

$$\frac{g^2}{2m_W^2}\mathcal{O}(x), \quad \text{where} \quad \mathcal{O}(x) = \bar{\psi}\gamma^\mu(1-\gamma^5)\psi\,\bar{\psi}\gamma_\mu(1-\gamma^5)\psi. \tag{12.100}$$

In the spirit of Wilson's renormalization group procedure, we can say that, on distance scales larger than m_W^{-1}, the W boson can be integrated out, leaving over the interaction (12.100).

How would we analyze the effects of the operator (12.100) on strongly interacting particles composed of quarks and antiquarks? A useful way to begin is to compute the Green's function of the operator $\mathcal{O}$ together with fields that create and destroy quarks. If we approximate the theory of quarks by a theory of free fermions, it is easy to compute these Green's functions; for

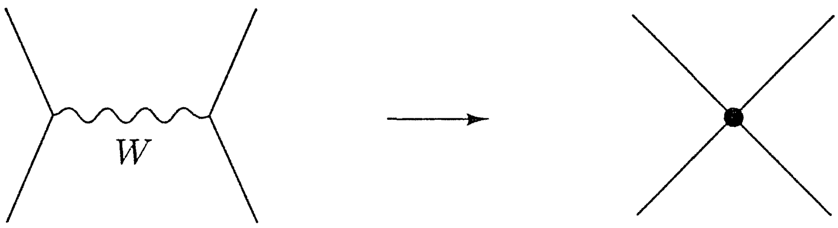


Figure 12.5. Interaction of quarks generated by the exchange of a W boson.

example:

$$\begin{aligned} &\langle \psi(p_1)\overline{\psi}(-p_2)\psi(p_3)\overline{\psi}(-p_4)\mathcal{O}(0)\rangle \\ &\quad = S_F(p_1)\gamma^\mu(1-\gamma^5)S_F(p_2)\, S_F(p_3)\gamma_\mu(1-\gamma^5)S_F(p_4). \end{aligned} \tag{12.101}$$

However, in an interacting field theory, the answer will be much more complicated. Some of these complications will involve the low-energy interactions of quarks, and we will leave them outside of the present discussion. However, in a renormalizable theory of quark interactions, one will also find that Green's functions containing $\mathcal{O}$ have new ultraviolet divergences. The one-loop corrections to (12.101) will contain diagrams that evaluate to the right-hand side of (12.101) times a divergent integral. These diagrams can be interpreted as field strength renormalizations of the operator $\mathcal{O}$. As with correlation functions of elementary fields, we can obtain finite and well-defined matrix elements of local operators only if we establish conventions for the normalization of local operators and introduce operator rescalings in the form of counterterms, order by order in perturbation theory, to preserve these conventions. More specifically, in a massless, renormalizable field theory of the fermions ψ, we should make the convention that Eq. (12.101) is exact at some spacelike normalization point for which $p_1^2 = p_2^2 = p_3^2 = p_4^2 = -M^2$. Then we should add a counterterm of the form $\delta_{\mathcal{O}}\mathcal{O}(x)$, and adjust this counterterm at each order of perturbation theory to insure that these relations are preserved. We refer to the operator satisfying the normalization condition (12.101) at M^2 as $\mathcal{O}_M$.

The renormalized operator $\mathcal{O}_M$ is a rescaled version of the operator $\mathcal{O}_0$ built of bare fields,

$$\mathcal{O}_0(x) = \overline{\psi}_0\gamma^\mu(1-\gamma^5)\psi_0\overline{\psi}_0\gamma_\mu(1-\gamma^5)\psi_0. \tag{12.102}$$

As we did for the elementary fields, we can write this relation as

$$\mathcal{O}_0 = Z_{\mathcal{O}}(M)\mathcal{O}_M. \tag{12.103}$$

This allows us to write the generalization of the relation (12.35) between Green's functions of bare and renormalized fields. Let us return to the language of scalar field theories and consider $\mathcal{O}(x)$ to be a local operator in a scalar field theory. Define

$$G^{(n;1)}(p_1,\cdots,p_n;k) = \langle \phi(p_1)\cdots\phi(p_n)\mathcal{O}_M(k)\rangle\,. \tag{12.104}$$

Then $G^{(n;1)}$ is related to a Green's function of bare fields by

$$G^{(n)}(p_1,\cdots,p_n;k) = Z(M)^{-n/2}Z_{\mathcal{O}}(M)^{-1}\langle\phi_0(p_1)\cdots\phi_0(p_n)\mathcal{O}_0(k)\rangle . \tag{12.105}$$

Repeating the derivation of Eqs. (12.63) and (12.64), we find that the Green's functions containing a local operator obey the Callan-Symanzik equation

$$\left[M\frac{\partial}{\partial M} + \beta(\lambda)\frac{\partial}{\partial\lambda} + n\gamma(\lambda) + \gamma_{\mathcal{O}}(\lambda)\right]G^{(n)} = 0, \tag{12.106}$$

where

$$\gamma_{\mathcal{O}} = M\frac{\partial}{\partial M}\log Z_{\mathcal{O}}(M). \tag{12.107}$$

It often happens that a quantum field theory contains several operators with the same quantum numbers. For example, in quantum electrodynamics, the operators $\bar\psi[\gamma^\mu D^\nu + \gamma^\nu D^\mu]\psi$ and $F^{\mu\lambda}F^\nu{}_\lambda$ are both symmetric tensors with zero electric charge; in addition, both operators have mass dimension 4. Such operators, with the same quantum numbers and the same mass dimension, can be mixed by quantum corrections.* For such a set of operators $\{\mathcal{O}^i\}$, the relation of renormalized and bare operators must be generalized to

$$\mathcal{O}_0^i = Z_{\mathcal{O}}^{ij}(M)\mathcal{O}_M^j. \tag{12.108}$$

This relation in turn implies that the anomalous dimension function $\gamma_{\mathcal{O}}$ in the Callan-Symanzik equation must be generalized to a matrix,

$$\gamma_{\mathcal{O}}^{ij} = [Z_{\mathcal{O}}^{-1}(M)]^{ik}M\frac{\partial}{\partial M}[Z_{\mathcal{O}}(M)]^{kj}. \tag{12.109}$$

Most of our applications of (12.106) in Chapter 18 will require this generalization.

On the other hand, there are some operators for which the rescaling and anomalous dimensions are especially simple. If $\mathcal{O}$ is the quark number current $\bar\psi\gamma^\mu\psi$, its normalization is fixed once and for all because the associated charge

$$Q = \int d^3x\,\bar\psi\gamma^0\psi$$

is just the conserved integer number of quarks minus antiquarks in a given state. More generally, for any conserved current J^μ, $Z_J(M) = 1$ and $\gamma_J = 0$. The same argument applies to the energy-momentum tensor. Thus, in the QED example above, the specific linear combination

$$T^{\mu\nu} = \frac{1}{2}\bar\psi[\gamma^\mu D^\nu + \gamma^\nu D^\mu]\psi + \frac{1}{4}F^{\mu\lambda}F^\nu{}_\lambda \tag{12.110}$$

receives no rescaling and no anomalous dimension. This linear combination of operators must be an eigenvector of the matrix γ^{ij} with eigenvalue zero.

---

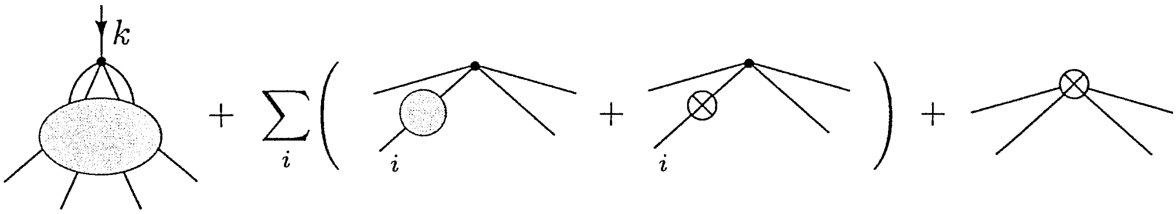
*Our assumption that we are working in a massless field theory constrains the possibilities for operator mixing. In a massive field theory, operators of a given dimension can also mix with operators of lower dimension.

So far, our discussion of operator matrix elements has been rather abstract. To make it more concrete, we will construct a formula for computing γ_O to leading order from one-loop counterterms, and then apply this formula to a simple example in ϕ^4 theory.

To find a simple formula for γ_O, we follow the same path that took us from Eq. (12.52) to the formula (12.53) for the β function. Consider an operator whose normalization condition is based on a Green's function with m scalar fields:

$$G^{(m;1)} = \langle \phi(p_1) \cdots \phi(p_m) O_M(k) \rangle . \tag{12.111}$$

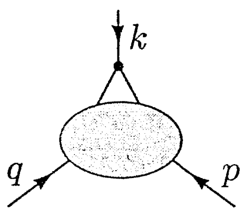
To compute this Green's function to one-loop order, we find the set of diagrams:



The last diagram is the counterterm δ_O needed to maintain the renormalization condition. Notice that the counterterm δ_Z also appears. If we insist that this sum of diagrams satisfies the Callan-Symanzik equation (12.106) to leading order in λ, we find, analogously to (12.53), the relation

$$\gamma_O(\lambda) = M \frac{\partial}{\partial M} \Big(-\delta_O + \frac{m}{2} \delta_Z \Big). \tag{12.112}$$

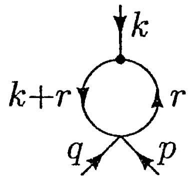
As a specific example of the use of this formula, let us compute the anomalous dimension γ_O of the mass operator ϕ^2 in ϕ^4 theory. There is a small subtlety involved in this computation. The Feynman diagrams of ϕ^4 theory generate an additive mass renormalization, which must be removed by the mass counterterm at each order in perturbation theory. We would like to define the mass operator as a perturbation which we can add to the massless theory defined in this way. To clarify the distinction between the underlying mass, which is renormalized to zero, and the explicit mass perturbation, we will analyze a Green's function of ϕ^2 in which this operator carries a specific nonzero momentum. We thus choose to define the normalization of ϕ^2 by the convention



$$= \big\langle \phi(p)\phi(q)\phi^2(k) \big\rangle = \frac{i}{p^2}\frac{i}{q^2} \cdot 2 \tag{12.113}$$

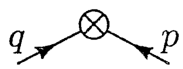
at $p^2 = q^2 = k^2 = -M^2$.

The one-particle-irreducible one-loop correction to (12.113) is



$$= \frac{i}{p^2}\frac{i}{q^2}\int\frac{d^4r}{(2\pi)^4}\delta^{k\ell}\cdot(-i\lambda)\frac{i}{r^2}\frac{i}{(k+r)^2}$$
$$= \frac{i}{p^2}\frac{i}{q^2}\Big[-\frac{\lambda}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}\Big], \tag{12.114}$$

where Δ is a function of the external momenta. At $-M^2$, this contribution must be canceled by a counterterm diagram,



$$= \frac{i}{p^2}\frac{i}{q^2}2\delta^{ij}\delta_{\phi^2}. \tag{12.115}$$

Thus, the counterterm must be

$$\delta_{\phi^2} = \frac{\lambda}{2(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \tag{12.116}$$

Since δ_Z is finite to order λ, this is the only contribution to (12.112), and we find

$$\gamma_{\phi^2} = \frac{\lambda}{16\pi^2}. \tag{12.117}$$

This function can be used together with the γ and β functions of pure massless ϕ^4 theory to discuss the scaling of Green's functions that include the mass operator.

12.5 Evolution of Mass Parameters

Finally, we discuss the renormalization group for theories with masses. We note, though, that although we treat these masses as arbitrary parameters, we will continue to use renormalization conventions that are independent of mass, and we will often treat the masses as small parameters. This approach breaks down at momentum scales much less than the scale of masses, but it is sufficient, and simpler than alternative approaches, for most practical applications of the renormalization group.

In the previous section, we worked out the scaling of Green's functions containing one power of the mass operator. It is a small step to generalize this discussion to include an arbitrary number ℓ of mass operators; one simply finds the equation (12.106) with the coefficient ℓ in front of the term $\gamma_{\mathcal{O}}$. Now consider what would happen if we add the mass operator directly to the Lagrangian of the massless ϕ^4 theory, treating this operator as a perturbation. If $\mathcal{L}_M$ is the massless Lagrangian renormalized at the scale M, the new Lagrangian will be

$$\mathcal{L}_M + \tfrac{1}{2}m^2\phi_M^2. \tag{12.118}$$

The Green's function of n scalar fields in the theory (12.118) could be expressed as a perturbation series in the mass parameter m^2. The coefficient of $(m^2)^\ell$ would be a joint correlation function of the n scalar fields with ℓ powers of ϕ_M^2, and would therefore satisfy the Callan-Symanzik equation (12.106) with the extra factor ℓ as noted above. In general, we can use the operator $m^2(\partial/\partial m^2)$ to count the number of insertions ℓ of ϕ^2. Then the Green's functions of the massive ϕ^4 theory, renormalized according to the mass-independent scheme, satisfy the equation

$$\left[M\frac{\partial}{\partial M}+\beta(\lambda)\frac{\partial}{\partial\lambda}+n\gamma(\lambda)+\gamma_{\phi^2}m^2\frac{\partial}{\partial m^2}\right]G^{(n)}(\{p_i\};M,\lambda,m^2)=0. \quad (12.119)$$

This argument extends to any perturbation of massless ϕ^4 theory. In the general case,

$$\mathcal{L}(C_i)=\mathcal{L}_M+\int d^4x\, C_i\mathcal{O}^i_M(x), \quad (12.120)$$

and the Green's functions of this perturbed theory satisfy

$$\left[M\frac{\partial}{\partial M}+\beta(\lambda)\frac{\partial}{\partial\lambda}+n\gamma(\lambda)+\sum_i\gamma_i(\lambda)C_i\frac{\partial}{\partial C_i}\right]G^{(n)}(\{p_i\};M,\lambda,\{C_i\})=0. \quad (12.121)$$

To interpret this equation, it will help to make a slight change to bring the notation in line with our new viewpoint. Let d_i be the mass dimension of the operator $\mathcal{O}^i$. Then rewrite (12.120) by representing each coefficient C_i as a power of M and a dimensionless coefficient ρ_i:

$$\mathcal{L}(\rho_i)=\mathcal{L}_M+\int d^4x\,\rho_i M^{4-d_i}\mathcal{O}^i_M(x). \quad (12.122)$$

The size of each ρ_i indicates the importance of the corresponding operator at the scale M. This new convention introduces further explicit M dependence into the Green's functions, which is compensated by a rescaling of the ρ_i. Thus (12.121) must be modified to

$$\left[M\frac{\partial}{\partial M}+\beta\frac{\partial}{\partial\lambda}+n\gamma+\sum_i\left[\gamma_i(\lambda)+d_i-4\right]\rho_i\frac{\partial}{\partial\rho_i}\right]G^{(n)}(\{p_i\};M,\lambda,\{\rho_i\})=0. \quad (12.123)$$

The meaning of this equation becomes clearer if we define

$$\beta_i=(d_i-4+\gamma_i)\rho_i. \quad (12.124)$$

Then

$$\left[M\frac{\partial}{\partial M}+\beta\frac{\partial}{\partial\lambda}+\sum_i\beta_i\frac{\partial}{\partial\rho_i}+n\gamma\right]G^{(n)}(\{p_i\};M,\lambda,\{\rho_i\})=0. \quad (12.125)$$

Now all of the coupling constants ρ_i appear on the same footing as λ. We can solve this generalized Callan-Symanzik equation using the same method as in Section 12.3, by introducing bacteria, which now live in a multidimensional velocity field (β, β_i). The solution will depend on a set of running coupling constants which obey the equations

$$\frac{d}{d\log(p/M)}\bar{\rho}_i = \beta_i(\bar{\rho}, \bar{\lambda}). \tag{12.126}$$

It is interesting to examine this flow of coupling constants for the case where all the dimensionless parameters λ, ρ_i are small, so that we are close to the free scalar field Lagrangian. In this situation, we can ignore the contribution of γ_i to β_i; then

$$\frac{d}{d\log(p/M)}\bar{\rho}_i = [d_i - 4 + \cdots]\bar{\rho}_i. \tag{12.127}$$

The solution to this equation is

$$\bar{\rho}_i = \rho_i\left(\frac{p}{M}\right)^{d_i-4}. \tag{12.128}$$

Operators with mass dimension greater than 4, corresponding to nonrenormalizable interactions, become less important as a power of p as $p \to 0$. This is exactly the behavior that we found in Eq. (12.27) using Wilson's method. Since we have now generalized the Callan-Symanzik equation to incorporate the most general perturbation of the free-field Lagrangian, it is pleasing that we recover the full structure of the Wilson flow of coupling constants. In addition, this more formal method gives us a way to compute the corrections to the Wilson flow due to $\lambda\phi^4$ interactions, order by order in λ, using Feynman diagrams.

We can move one step closer to the generality of Section 12.1 by moving from four dimensions to an arbitrary dimension d. We require only two small changes in the formalism. First, the operator ϕ^4 acquires a dimensionful coefficient when $d \neq 4$, and we must take account of this. We have seen in the discussion below Eq. (10.13) that a scalar field has mass dimension $(d-2)/2$. Thus, the operator ϕ^4 has mass dimension $(2d-4)$, and so its coefficient has dimension $4-d$. To implement the renormalization group, we redefine λ so that this coefficient remains dimensionless in d dimensions. We treat the mass term similarly, replacing $m^2 \to \rho_m M^2$. Thus the expansion of the Lagrangian about the free scalar field theory $\mathcal{L}_0$ reads:

$$\mathcal{L} = \mathcal{L}_0 - \frac{1}{2}\rho_m M^2\phi_M^2 - \frac{1}{4}\lambda M^{4-d}\phi_M^4 + \cdots. \tag{12.129}$$

The second required change in the formalism is that of recomputing the β and γ functions in the new dimension. To order λ, the result is surprisingly innocuous. Consider, for example, the computation of γ_{ϕ^2}, Eq. (12.114). This computation, which was performed in dimensional regularization, is essentially unchanged. For general values of d, the derivative of the counterterm δ_{ϕ^2} with respect to $\log M$ still involves the factor

$$M\frac{\partial}{\partial M}\left(\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}\right) = -2 + \mathcal{O}(4-d). \tag{12.130}$$

This observation holds for all of the γ_i, and the β function is shifted only by the contribution of the mass dimension of λ. Thus, for d near 4,

$$\begin{aligned}\beta &= (d-4)\lambda + \beta^{(4)}(\lambda) + \cdots,\\ \beta_m &= [-2 + \gamma_{\phi^2}^{(4)}]\rho_m + \cdots,\\ \beta_i &= [d_i - d + \gamma_i^{(4)}]\rho_i + \cdots,\end{aligned} \tag{12.131}$$

where the functions with a superscript (4) are the four-dimensional results obtained earlier in this section, and the omitted correction terms are of order $\lambda\cdot(d-4)$. The precise form of these corrections depends on the renormalization scheme.†

Using the explicit four-dimensional result (12.46) for β, we now find

$$\beta = -(4-d)\lambda + \frac{3\lambda^2}{16\pi^2}. \tag{12.132}$$

For $d \geq 4$, this function is positive and predicts that the coupling constant flows smoothly to zero at large distances. However, when $d < 4$, this $\beta(\lambda)$ has the form shown in Fig. 12.4(b). Thus it generates just the coupling constant flow that we discussed from Wilson's viewpoint below Eq. (12.29). At small values of λ, the coupling constant increases in importance with increasing distance, as dimensional analysis predicts. However, at larger λ, the coupling constant decreases as a result of its own nonlinear effects. These two tendencies come into balance at the zero of the beta function,

$$\lambda_* = \frac{16\pi^2}{3}(4-d), \tag{12.133}$$

which gives a nontrivial fixed point of the renormalization group flows in scalar field theory for $d < 4$. If we formally consider values of d close to 4, this fixed point occurs in a region where the coupling constant is small and we can use Feynman diagrams to investigate its properties. This fixed point, which was discovered by Wilson and Fisher,‡ has important consequences for statistical mechanics, which we will discuss in Chapter 13.

Critical Exponents: A First Look

As an application of the formalism of this section, let us calculate the renormalization group flow of the coefficient of the mass operator in ϕ^4 theory. This is found by integrating Eq. (12.126), using the value of β_m from (12.131):

$$\frac{d}{d\log p}\bar{\rho}_m = [-2 + \gamma_{\phi^2}(\bar{\lambda})]\bar{\rho}_m. \tag{12.134}$$

---

†This expansion is displayed to rather high order in E. Brezin, J. C. Le Gillou, and J. Zinn-Justin, *Phys. Rev.* **D9**, 1121 (1974).

‡K. G. Wilson and M. E. Fisher, *Phys. Rev. Lett.* **28**, 240 (1972).

For $\lambda = 0$, this equation gives the trivial relation

$$\bar{\rho}_m = \rho_m \left(\frac{M}{p}\right)^2. \tag{12.135}$$

If we recall that we originally defined $\rho_m = m^2/M^2$, this is just a complicated way of saying that, when p becomes of order m, the mass term becomes an important term in the Lagrangian. At this point, the correlations in the ϕ field begin to die away exponentially. The characteristic range of correlations, which in statistical mechanics would be called the correlation length ξ, is given by

$$\xi \sim p_0^{-1}, \qquad \text{where } \bar{\rho}_m(p_0) = 1. \tag{12.136}$$

If we evaluate this criterion, we find $\xi \sim (M^2 \rho_m)^{-1/2}$, that is, $\rho \sim m^{-1}$, as we would have expected.

However, the application of this criterion at the fixed point λ_* gives a much more interesting result. If we set $\bar{\lambda} = \lambda_*$, then Eq. (12.134) has the solution

$$\bar{\rho}_m = \rho_m \left(\frac{M}{p}\right)^{2-\gamma_{\phi^2}(\lambda_*)}. \tag{12.137}$$

This gives a nontrivial relation

$$\xi \sim \rho_m^{-\nu}, \tag{12.138}$$

where the exponent ν is given formally by the expression

$$\nu = \frac{1}{2 - \gamma_{\phi^2}(\lambda_*)}. \tag{12.139}$$

Using the results (12.133) and (12.117), we can evaluate this explicitly for d near 4:

$$\nu^{-1} = 2 - \frac{1}{3}(4-d). \tag{12.140}$$

Wilson and Fisher showed that this expression can be extended to a systematic expansion of ν in powers of $\epsilon = (4-d)$.

Because the exponent ν has an interpretation in statistical mechanics, it is directly measurable in the realistic case of three dimensions. In the statistical mechanical interpretation of scalar field theory, ρ_m is just the parameter that one must adjust finely to bring the system to the critical temperature. Thus ρ_m is proportional to the deviation from the critical temperature, $(T - T_C)$. Our field theoretic analysis thus implies that the correlation length in a magnet grows as $T \to T_C$ according to the scaling relation

$$\xi \sim (T - T_C)^{-\nu}. \tag{12.141}$$

It also gives a definite, and somewhat unusual, prediction for the value of ν. It predicts that ν is close to the value 1/2 suggested by the Landau approximation studied in Chapter 8 (Eq. (8.16)), but that ν differs from this value by some systematic corrections.

A scaling behavior of the type (12.141) is observed in magnets, and it is known that several definite scaling laws occur, depending on the symmetry of the spin ordering. Magnets can be characterized by the number of fluctuating spin components: N=1 for magnets with a preferred axis, N=2 for magnets with a preferred plane, and N=3 for magnets that are isotropic in three-dimensional space. The experimental value of ν depends on this parameter. The ϕ^4 field theory discussed in this chapter contained only one fluctuating field; this is the analogue of a magnet with one spin component. In Chapter 11, we considered a generalization of ϕ^4 theory to a theory of N fields with $O(N)$ symmetry. We might guess that this system models magnets of general N.

If this correspondence is correct, Eq. (12.140) gives a prediction for the value of ν in magnets with a preferred axis. In Section 13.1, we will repeat the analysis leading to this equation in the $O(N)$-symmetric ϕ^4 theory and derive the formula

$$\nu^{-1} = 2 - \frac{N+2}{N+8}(4-d), \tag{12.142}$$

valid for general N to first order in $(4-d)$. For the cases $N = 1, 2, 3$ and $d = 3$, this formula predicts

$$\nu = 0.60,\ 0.63,\ 0.65. \tag{12.143}$$

For comparison, the best current experimental determinations of ν in magnetic systems give*

$$\nu = 0.64,\ 0.67,\ 0.71 \tag{12.144}$$

for $N = 1, 2, 3$. The prediction (12.143) gives a reasonable first approximation to the experimental results.

The ability of quantum field theory to predict the critical exponents gives a concrete application both of the formal connection between quantum field theory and statistical mechanics and of the flows of coupling constants predicted by the renormalization group. However, there is another experimental aspect of critical behavior that is even more remarkable, and more persuasive. Critical behavior can be studied not only in magnets but also in fluids, binary alloys, superfluid helium, and a host of other systems. It has long been known that, for systems with this disparity of microscopic dynamics, the scaling exponents at the critical point depend only on the dimension N of the fluctuating variables and not on any other detail of the atomic structure. Fluids, binary alloys, and uniaxial magnets, for example, have the same critical exponents. To the untutored eye, this seems to be a miracle. But for a quantum field theorist, this conclusion is the natural outcome of the renormalization group idea, in which most details of the field theoretic interaction are described by operators that become irrelevant as the field theory finds its proper, simple, large-distance behavior.

*For further details, see Table 13.1 and the accompanying discussion.

Problems

12.1 Beta functions in Yukawa theory. In the pseudoscalar Yukawa theory studied in Problem 10.2, with masses set to zero,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{\lambda}{4!}\phi^4 + \bar{\psi}(i\partial\!\!\!/)\psi - ig\bar{\psi}\gamma^5\psi\phi,$$

compute the Callan-Symanzik β functions for λ and g:

$$\beta_\lambda(\lambda, g), \qquad \beta_g(\lambda, g),$$

to leading order in coupling constants, assuming that λ and g^2 are of the same order. Sketch the coupling constant flows in the λ-g plane.

12.2 Beta function of the Gross-Neveu model. Compute $\beta(g)$ in the two-dimensional Gross-Neveu model studied in Problem 11.3,

$$\mathcal{L} = \bar{\psi}_i i\partial\!\!\!/\psi_i + \tfrac{1}{2}g^2(\bar{\psi}_i\psi_i)^2,$$

with $i = 1, \ldots, N$. You should find that this model is asymptotically free. How was that fact reflected in the solution to Problem 11.3?

12.3 Asymptotic symmetry. Consider the following Lagrangian, with two scalar fields ϕ_1 and ϕ_2:

$$\mathcal{L} = \frac{1}{2}\left((\partial_\mu\phi_1)^2 + (\partial_\mu\phi_2)^2\right) - \frac{\lambda}{4!}(\phi_1^4 + \phi_2^4) - \frac{2\rho}{4!}(\phi_1^2\phi_2^2).$$

Notice that, for the special value $\rho = \lambda$, this Lagrangian has an $O(2)$ invariance rotating the two fields into one another.

(a) Working in four dimensions, find the β functions for the two coupling constants λ and ρ.

(b) Write the renormalization group equation for the ratio of couplings ρ/λ. Show that, if $\rho/\lambda < 3$ at a renormalization point M, this ratio flows toward the condition $\rho = \lambda$ at large distances. Thus the $O(2)$ internal symmetry appears asymptotically.

(c) Write the β functions for λ and ρ in $4 - \epsilon$ dimensions. Show that there are nontrivial fixed points of the renormalization group flow at $\rho/\lambda = 0, 1, 3$. Which is the most stable? Sketch the pattern of coupling constant flows. This flow implies that the critical exponents are those of a symmetric two-component magnet.

Chapter 13

Critical Exponents and Scalar Field Theory

The idea of running coupling constants and renormalization-group flows gives us a new language with which to discuss the qualitative behavior of scalar field theory. In our first discussion of ϕ^4 theory, each value of the coupling constant—and, more generally, each form of the potential and each spacetime dimension—gave a separate problem to be explored. But in Chapter 12, we saw that ϕ^4 theories with different values of the coupling are connected by renormalization-group flows, and that the pattern of these flows changes continuously with the spacetime dimension. In this context, it makes sense to ask the very general question: How does ϕ^4 theory behave as a function of the dimension? This chapter will give a detailed answer to this question.

The central ingredient in our analysis will be the Wilson-Fisher fixed point discussed in Section 12.5. This fixed point exists in spacetime dimensions d with $d < 4$; in those dimensions it controls the renormalization group flows of massless ϕ^4 theory. The scalar field theory has manifest or spontaneously broken symmetry according to the sign of the mass parameter m^2. Near $m^2 = 0$, the theory exhibits scaling behavior with anomalous dimensions whose values are determined by the renormalization group equations. For $d > 4$, the Wilson-Fisher fixed point disappears, and only the free-field fixed point remains. Again, the theory exhibits two distinct phases, but now the behavior at the transition is determined by the renormalization group flows near the free-field fixed point, so the scaling laws are those that follow from simple dimensional analysis.

The continuation of these results to Euclidean space has important implications for the theory of phase transitions in magnets and fluids. As we discussed in the previous chapter, the ideas of the renormalization group imply that the power-law behaviors of thermodynamic quantities near a phase transition point are determined by the behavior of correlation functions in a Euclidean ϕ^4 theory. The results stated in the previous paragraph then imply the following conclusions for critical scaling laws: For statistical systems in a space of dimension $d > 4$, the scaling laws are just those following from simple dimensional analysis. These predictions are precisely those of Landau theory, which we discussed in Chapter 8. On the other hand, for $d < 4$, the critical scaling laws are modified, in a way that we can compute using the renormalization group.

In $d = 4$, we are on the boundary between the two types of scaling behavior. This corresponds to the situation in which ϕ^4 theory is precisely renormalizable. In this case, the dimensional analysis predictions are corrected, but only by logarithms. We will analyze this case specifically in Section 13.2.

Though it is not obvious, the case $d = 2$ provides another boundary. Here the transition to spontaneous symmetry breaking is described by a different quantum field theory, which becomes renormalizable in two dimensions. In Section 13.3, we will introduce that theory, called the *nonlinear sigma model*, and show how its renormalization group behavior merges smoothly with that of ϕ^4 theory. By combining all of the results of this chapter, we will obtain a quantitative understanding of the behavior of ϕ^4 theory, and of critical phenomena, over the whole range of spacetime dimensions.

13.1 Theory of Critical Exponents

At the end of Chapter 12, we used properties of the renormalization group for scalar field theory to make a prediction about the behavior of correlations near the critical point of a thermodynamic system. We argued that the range of correlations, the correlation length ξ, should increase to infinity as one approaches the critical point, according to the scaling law (12.141). The exponent in this equation, called ν, should depend only on the symmetry of the order parameter. We argued, further, that this exponent is related to the anomalous dimension of a local operator in ϕ^4 theory, and that it can be computed from Feynman diagrams. In this section, we will show that similar conclusions apply more generally to a large number of scaling laws associated with a critical point.

To begin, we will define systematically a set of *critical exponents*, exponents of scaling laws that describe the thermodynamic behavior in the vicinity of the critical point. We will then show, using the Callan-Symanzik equation, that all these exponents can be reduced to two basic anomalous dimensions. Finally, we will compare this remarkable prediction of quantum field theory to experiment.

In suggesting a set of critical scaling laws, we begin with the behavior of the correlation function of fluctuations of the ordering field. For definiteness, we will use the language appropriate to a magnet, as in Chapter 8. We will compute classical thermal expectation values as correlation functions in a Euclidean quantum field theory, as explained in Section 9.3. The fluctuating field will be called the spin field $s(x)$, its integral will be the magnetization M, the the external field that couples to $s(x)$ will be called the magnetic field H. (In deference to the magnetization, we will denote the renormalization scale in the Callan-Symanzik equation by μ in this section.)

Define the two-point correlation function by

$$G(x) = \langle s(x)s(0)\rangle \,, \tag{13.1}$$

or by the connected expectation value, if we are in the magnetized phase where

$\langle s(x) \rangle \neq 0$. Away from the critical point, $G(x)$ should decay exponentially, according to

$$G(x) \sim \exp[-|x|/\xi]. \tag{13.2}$$

The approach to the critical point is characterized by the parameter

$$t = \frac{T - T_C}{T_C}. \tag{13.3}$$

Then we expect that, as $t \to 0$, the correlation length should increase to infinity. Define the exponent ν, (12.141), by the formula

$$\xi \sim |t|^{-\nu}. \tag{13.4}$$

Just at $t = 0$, the correlation function should decay only as a power law. Define the exponent η by the formula

$$G(x) \sim \frac{1}{|x|^{d-2+\eta}}, \tag{13.5}$$

where d is the Euclidean space dimension.

The behaviors of thermodynamic quantities near the critical point define a number of additional exponents. Typically, the specific heat of the thermodynamic system diverges as $t \to 0$; define the exponent α by the formula for the specific heat at fixed external field $H = 0$:

$$C_H \sim |t|^{-\alpha}. \tag{13.6}$$

Since the ordering sets in at $t = 0$, the magnetization at zero field tends to zero as $t \to 0$ from below. Define the exponent β (not to be confused with the Callan-Symanzik function) by

$$M \sim |t|^{\beta}. \tag{13.7}$$

Even at $t = 0$ one has a nonzero magnetization at nonzero magnetic field. Write the law by which this magnetization tends to zero as $H \to 0$ as the relation

$$M \sim H^{1/\delta}. \tag{13.8}$$

Finally, the magnetic susceptibility diverges at the critical point; we write this divergence as the relation

$$\chi \sim |t|^{-\gamma}. \tag{13.9}$$

Equations (13.4)–(13.9) define a set of critical exponents α, β, γ, δ, ν, η, which can be measured experimentally for a variety of thermodynamic systems.*

In Chapter 12 we argued, following Wilson, that a thermodynamic system near its critical point can be described by a Euclidean quantum field theory. At the level of the atomic scale, the Lagrangian of this quantum field theory may be complicated; however, when we have integrated out the small-scale

*A variety of further critical exponents and relations are presented in M. E. Fisher, *Repts. Prog. Phys.* **30**, 615 (1967).

degrees of freedom, this Lagrangian simplifies. If we adjust a parameter of the theory to insure the presence of long-range correlations, the Lagrangian must closely approach a fixed point of the renormalization group. Generically, the Lagrangian will approach the fixed point with a single unstable or relevant direction, corresponding to the mass parameter of ϕ^4 theory. In $d < 4$, this is the Wilson-Fisher fixed point. In $d \geq 4$, it is the free-field fixed point. For definiteness, we will assume $d < 4$ in the following discussion.

Exponents of the Spin Correlation Function

In this setting, we can study the behavior of the spin-spin correlation function $G(x)$. By the argument just reviewed, $G(x)$ is proportional to the two-point correlation function of a Euclidean scalar field theory. The technology introduced in the previous chapter can be applied directly. The correlation function obeys the Callan-Symanzik equation (12.125),

$$\Big[\mu\frac{\partial}{\partial\mu} + \sum_i \beta_i \frac{\partial}{\partial\rho_i} + 2\gamma\Big] G(x; \mu, \{\rho_i\}) = 0. \tag{13.10}$$

Here we include the ϕ^4 coupling λ among the generalized couplings ρ_i.

By dimensional analysis, in d dimensions,

$$G(x) = \frac{1}{|x|^{d-2}} g(\mu|x|, \{\rho_i\}), \tag{13.11}$$

where g is an arbitrary function of the dimensionless parameters. (This is the Fourier transform of the statement that $G(p) \sim p^{-2}$ times a dimensionless function.) From this starting point, we can solve the Callan-Symanzik equation (13.10) by the method of Section 12.3, and find

$$G(x) = \frac{1}{|x|^{d-2}} h(\{\bar{\rho}_i(x)\}) \cdot \exp\Bigg[-2\int_{1/\mu}^{|x|} d\log(|x'|)\, \gamma(\{\bar{\rho}(x')\})\Bigg], \tag{13.12}$$

where h is a dimensionless initial condition. The running coupling constants $\bar{\rho}_i$ obey the differential equation

$$\frac{d}{d\log(1/\mu|x|)} \bar{\rho}_i = \beta_i(\{\rho_j\}). \tag{13.13}$$

We studied the solution to this equation in Section 12.5. We saw there that, for flows that come to the vicinity of the Wilson-Fisher fixed point, the dimensionless coefficient of the mass operator grows as one moves toward large distances, while the other dimensionless parameters become small. Let λ_* be the location of the fixed point. Then we can write more explicitly

$$\begin{aligned} \bar{\rho}_m &= \rho_m(\mu|x|)^{2-\gamma_{\phi^2}(\lambda_*)}, \\ \bar{\rho}_i &= \rho_i(\mu|x|)^{-A_i}, \end{aligned} \tag{13.14}$$

where $A_i > 0$ for $i \neq m$. If the deviation of λ from the fixed point is treated as one of the ρ_i, by defining

$$\rho_\lambda = \lambda - \lambda_*, \tag{13.15}$$

this parameter also decreases in importance as a power of $|x|$, as we demonstrated in Eq. (12.96). In the language of Section 12.1, all of the parameters ρ_i multiply irrelevant operators, except for ρ_m, which multiplies a relevant operator.

To approach the critical point, we adjust the parameters of the underlying theory so that, at some scale $(1/\mu)$ near the atomic scale, $\rho_m \ll 1$. If ρ_m is adjusted by tuning the temperature of the thermodynamic system, then $\rho_m \sim t$. The critical scaling laws will be valid if there is a region of distance scales where $\bar{\rho}_m$ remains small while the other $\bar{\rho}_i$ can be neglected. The scaling laws can then be computed by evaluating the solution to the Callan-Symanzik equation with $\bar{\rho}_m$ given by (13.14) and the other $\bar{\rho}_i$ set equal to zero. The corrections to this approximation can be shown to be proportional to positive powers of t.

In this approximation, we should evaluate the function $\gamma(\lambda)$ in (13.12) at $\bar{\rho}_\lambda = 0$, that is, at the fixed point. Using this value and the solution for $\bar{\rho}_m$, Eq. (13.12) becomes

$$G(x) = \frac{1}{|x|^{d-2}} \cdot \frac{1}{(\mu|x|)^{2\gamma(\lambda_*)}} \cdot h\big(t(\mu|x|)^{2-\gamma_{\phi^2}(\lambda_*)}\big). \tag{13.16}$$

This equation implies the scaling laws (13.5) and (13.4): For the argument of h sufficiently small, $G(x)$ obeys Eq. (13.5), with

$$\eta = 2\gamma(\lambda_*). \tag{13.17}$$

At large distances, h must fall off exponentially, since this function is derived from a scalar field propagator. From the argument of h, we deduce that this exponential must be of the form

$$\exp[-|x|(\mu t^\nu)], \tag{13.18}$$

where, as in (12.139),

$$\nu = \frac{1}{2 - \gamma_{\phi^2}(\lambda_*)}. \tag{13.19}$$

This is precisely the scaling law (13.2), (13.4), with the identification of ν in terms of the anomalous dimension of the operator ϕ^2.

Exponents of Thermodynamic Functions

The thermodynamic critical scaling laws can be derived in a similar way, by studying the scaling behavior of macroscopic thermodynamic variables. These are derived from the Gibbs free energy, or, in the language of quantum fields, from the effective potential of the scalar field theory. Since the effective potential, and, more generally, the effective action, are constructed from correlation functions, these quantities should satisfy Callan-Symanzik equations. We will now construct those equations and then use them to identify the thermodynamic critical exponents.

In Eq. (11.96), we showed that the effective action Γ depends on the classical field $\phi_{\rm cl}$ in such a way that the nth derivative of Γ with respect to $\phi_{\rm cl}$ gives the one-particle-irreducible n-point function of the field theory. Thus we can reconstruct Γ from the 1PI functions by writing the Taylor series

$$\Gamma[\phi_{\rm cl}] = i\sum_2^\infty \frac{1}{n!}\int dx_1\cdots dx_n\,\phi_{\rm cl}(x_1)\cdots\phi_{\rm cl}(x_n)\,\Gamma^{(n)}(x_1,\ldots,x_n), \qquad (13.20)$$

where the $\Gamma^{(n)}$ are the 1PI amplitudes.

To find the Callan-Symanzik equation satisfied by $\Gamma[\phi_{\rm cl}]$, it is easiest to first work out the equation satisfied by $\Gamma^{(n)}$. We begin by considering the irreducible three-point function $\Gamma^{(3)}$. This function is defined as

$$\Gamma^{(3)}(p_1,p_2,p_3) = \frac{1}{G^{(2)}(p_1)G^{(2)}(p_2)G^{(2)}(p_3)}G^{(3)}(p_1,p_2,p_3). \qquad (13.21)$$

Rescaling with factors $Z(\mu)$, we see that $\Gamma^{(3)}$ is related to the irreducible three-point function of bare fields by

$$\Gamma^{(3)}(p_1,p_2,p_3) = Z(\mu)^{+3/2}\Gamma_0^{(3)}(p_1,p_2,p_3).$$

Similarly, the irreducible n-point function is related to the corresponding function of bare fields by

$$\Gamma^{(n)} = Z(\mu)^{n/2}\Gamma_0^{(n)}. \qquad (13.22)$$

This relation is identical in form to the corresponding relation for the full Green's functions, Eq. (12.35), except for the change of sign in the exponent. From this point, we can follow the logic used to derive the Callan-Symanzik equation for Green's functions, Eq. (12.41); the only difference is that the $n\gamma$ term enters with the opposite sign. Thus we find

$$\Big[\mu\frac{\partial}{\partial\mu} + \beta(\lambda)\frac{\partial}{\partial\lambda} - n\gamma(\lambda)\Big]\Gamma^{(n)}(\{p_i\};\mu,\lambda) = 0. \qquad (13.23)$$

To convert this to an equation for the effective action, note that, on the right-hand side of Eq. (13.20), the function $\Gamma^{(n)}$ is accompanied by n powers of the classical field. Then Eq. (13.23), integrated with n powers of $\phi_{\rm cl}$ and summed over n, is equivalent to the equation

$$\Big[\mu\frac{\partial}{\partial\mu} + \beta(\lambda)\frac{\partial}{\partial\lambda} - \gamma(\lambda)\int dx\,\phi_{\rm cl}(x)\frac{\delta}{\delta\phi_{\rm cl}(x)}\Big]\Gamma([\phi_{\rm cl}];\mu,\lambda) = 0. \qquad (13.24)$$

The operator multiplying $\gamma(\lambda)$ counts the number of powers of $\phi_{\rm cl}$ in each term of the Taylor expansion. By specializing Eq. (13.24) to the case of constant $\phi_{\rm cl}$, we find the Callan-Symanzik equation for $V_{\rm eff}$:

$$\Big[\mu\frac{\partial}{\partial\mu} + \beta(\lambda)\frac{\partial}{\partial\lambda} - \gamma\phi_{\rm cl}\frac{\partial}{\partial\phi_{\rm cl}}\Big]V_{\rm eff}(\phi_{\rm cl},\mu,\lambda) = 0. \qquad (13.25)$$

To apply Eq. (13.25) to the problem of critical exponents, we first convert this equation to the notation of statistical mechanics by replacing $\phi_{\rm cl}$ with the magnetization M, the conjugate source J by H, and the effective potential $V_{\rm eff}$ by the Gibbs free energy $\mathbf{G}(M)$. At the same time, we will generalize λ to the full set of couplings ρ_i. Then (13.25) takes the form

$$\left[\mu\frac{\partial}{\partial\mu}+\sum_i\beta_i\frac{\partial}{\partial\rho_i}-\gamma M\frac{\partial}{\partial M}\right]\mathbf{G}(M,\mu,\{\rho_i\})=0. \tag{13.26}$$

Now let us find the solution to this equation. As before, we begin from a statement of dimensional analysis. In d dimensions, the effective potential has mass dimension d, and a scalar field has mass dimension $(d-2)/2$. Thus

$$\mathbf{G}(M,\mu,\{\rho_i\})=M^{2d/(d-2)}\hat{g}\big((M\mu^{-(d-2)/2}),\{\rho_i\}\big), \tag{13.27}$$

where $\hat{g}$ is a new dimensionless function. Inserting (13.27) into (13.26), we see that $\hat{g}$ satisfies

$$\left[\sum_i\beta_i\frac{\partial}{\partial\rho_i}-\Big(\frac{d-2}{2}+\gamma\Big)M\frac{\partial}{\partial M}-d\frac{2}{d-2}\gamma\right]\hat{g}\big((M\mu^{(d-2)/2}),\{\rho_i\}\big)=0, \tag{13.28}$$

that is,

$$\left[M\frac{\partial}{\partial M}-\sum_i\frac{2\beta_i}{(d-2+2\gamma)}\frac{\partial}{\partial\rho_i}+\frac{4d\gamma}{(d-2)(d-2+2\gamma)}\right]\hat{g}=0. \tag{13.29}$$

Solving this equation, we find

$$\begin{aligned}\mathbf{G}(M)&=M^d\hat{h}(\{\bar{\rho}_i(M)\})\\&\quad\times\exp\left[-\int_{\mu^{(d-2)/2}}^{M}d\log(M')\frac{4d\gamma}{(d-2)(d-2+2\gamma)}(\{\bar{\rho}(M')\})\right],\end{aligned} \tag{13.30}$$

where the running coupling constants $\bar{\rho}_i$ obey

$$\frac{d}{d\log M}\bar{\rho}_i=\frac{2\beta_i(\{\bar{\rho}_i\})}{d-2+2\gamma(\{\bar{\rho}_i\})}. \tag{13.31}$$

As in our discussion of the spin correlation function, we specialize to the critical region by assuming that we are on a renormalization group flow that passes close to the Wilson-Fisher fixed point. We again ignore the effects of irrelevant operators. Then we should set

$$\begin{aligned}\bar{\rho}_m&=\rho_m\big(M\mu^{-(d-2)/2}\big)^{-2(2-\gamma_{\phi^2}(\lambda_*))/(d-2+2\gamma(\lambda_*))},\\\bar{\rho}_i&=0\qquad\text{for } i\neq m,\end{aligned} \tag{13.32}$$

with $\rho_m\sim t$. In this approximation, the Gibbs free energy takes the form

$$\begin{aligned}\mathbf{G}(M,t)=M^{2/d-2}&\cdot\big(M\mu^{(d-2)/2}\big)^{4d\gamma(\lambda_*)/(d-2+2\gamma(\lambda_*))}\\&\cdot\hat{h}\big(t(M\mu^{(d-2)/2})^{-2(2-\gamma_{\phi^2}(\lambda_*))/(d-2+2\gamma(\lambda_*))}\big),\end{aligned} \tag{13.33}$$

where $\hat{h}$ is a smooth initial condition.

To simplify the form of the exponents in this expression, we anticipate some of the results below and replace

$$\begin{aligned} \beta &= \frac{d-2+2\gamma(\lambda_*)}{2(2-\gamma_{\phi^2}(\lambda_*))}, \\ \delta &= \frac{2d}{d-2+2\gamma(\lambda_*)} - 1 = \frac{d+2-2\gamma(\lambda_*)}{d-2+2\gamma(\lambda_*)}. \end{aligned} \tag{13.34}$$

We must demonstrate that these new exponents indeed correspond to the ones we have defined in Eqs. (13.7) and (13.8). With these replacements (and ignoring the dependence on μ from here on), we find for $\mathbf{G}$ the scaling formula

$$\mathbf{G}(M,t) = M^{1+\delta}\hat{h}(tM^{-1/\beta}), \tag{13.35}$$

where $\hat{h}$ has a smooth limit as $t \to 0$. An equivalent way to represent this formula is

$$\mathbf{G}(M,t) = t^{\beta(1+\delta)}\hat{f}(Mt^{-\beta}). \tag{13.36}$$

The scaling laws for thermodynamic quantities follow immediately from these relations. Along the line $t=0$, we find from (13.35) that

$$H = \frac{\partial \mathbf{G}}{\partial M} = \hat{h}(0)M^{\delta}, \tag{13.37}$$

which is precisely (13.8). Below the criticial temperature, we find the nonzero value of the magnetization by minimizing $\mathbf{G}$ with respect to M. In the scaling region, this minimum occurs at the minimum m_0 of the function $\hat{f}(m)$ in (13.36). This leads to relation (13.7), in the form

$$Mt^{-\beta} = m_0. \tag{13.38}$$

If we work above T_C and in zero field, the minimum of $\hat{f}$ must occur at $M=0$. Then

$$\mathbf{G}(t) \sim t^{\beta(1+\delta)}. \tag{13.39}$$

To compute the specific heat, we differentiate twice with respect to temperature; this gives the scaling law (13.6), with

$$2-\alpha = \beta(1+\delta) = \frac{2d}{2-\gamma_{\phi^2}(\lambda_*)}. \tag{13.40}$$

Finally, we must construct the scaling law for the magnetic susceptibility. From (13.36), the scaling law for H at nonzero t is

$$H = \frac{\partial \mathbf{G}}{\partial M} = t^{\beta\delta}\hat{f}'(Mt^{-\beta}). \tag{13.41}$$

The inverse of this relation is the scaling law

$$M = t^{\beta}\hat{c}(Ht^{-\beta\delta}). \tag{13.42}$$

The magnetic susceptibility at zero field is then

$$\chi = \left(\frac{\partial M}{\partial H}\right)_t = \hat{c}'(0)t^{-(\delta-1)\beta}. \tag{13.43}$$

Thus, we confirm Eq. (13.9), with the identification

$$\gamma = (\delta - 1)\beta = \frac{2(1-\gamma(\lambda_*))}{2-\gamma_{\phi^2}(\lambda_*)}. \tag{13.44}$$

We have now found explicit expressions for all of the various critical exponents in terms of the Callan-Symanzik functions. As the dimensionality d approaches 4 from below, the fixed point λ_* tends to zero. Then the six critical exponents approach the values that they would attain in simple dimensional analysis:

$$\begin{aligned} &\eta = 0; \qquad \nu = \tfrac{1}{2}; \qquad \alpha = 0; \qquad \beta = \tfrac{1}{2}; \\ &\qquad\quad \gamma = 1; \qquad \delta = 3. \end{aligned} \tag{13.45}$$

It is no surprise that the values of η, ν and β given in (13.45) are those that we derived in Chapter 8 from the the Landau theory of critical phenomena. The other values can similarly be shown to follow from Landau theory. The renormalization group analysis tells us how to systematically correct the predictions of Landau theory to take proper account of the large-scale fluctuations of the spin field.

Notice that all of the exponents associated with thermodynamic quantities are constructed from the same ingredients as the exponents associated with the correlation function. From the field theory viewpoint, this is obvious, since all of the scaling laws in the field theory must ultimately follow from the anomalous dimensions of the operators $\phi(x)$ and $\phi^2(x)$, which are precisely $\gamma(\lambda_*)$ and $\gamma_{\phi^2}(\lambda_*)$. This result, however, has an interesting experimental consequence: It implies model-independent relations among critical exponents. For example, in any system with a critical point, this theory predicts

$$\alpha = 2 - d\nu, \qquad \beta = \tfrac{1}{2}(d-2+\eta)\nu. \tag{13.46}$$

These relations test the general framework of identifying a critical point with the fixed point of a renormalization group flow.

In addition, the field theoretic approach to critical phenomena predicts that critical exponents are *universal*, in the sense that they take the same values in condensed matter systems that approach the same scalar field fixed point in the limit $T \to T_C$.

Values of the Critical Exponents

Finally, scalar field theory actually predicts the values of $\gamma(\lambda_*)$ and $\gamma_{\phi^2}(\lambda_*)$, either from the expansion in powers of $\epsilon = 4 - d$ described in Section 12.5 or by direct expansion of the β and γ functions in powers of λ. We can use these expressions to generate quantitative predictions for the critical exponents. We gave an example of such a prediction at the end of Section 12.5, when we presented in Eq. (12.143) the first two terms of an expansion for ν. We now return to this question to give field-theoretic predictions for all of the critical exponents.

In our discussion at the end of Section 12.5, we remarked that magnets with different numbers of fluctuating spin components are observed to have different values for the critical exponents. An optimistic hypothesis would be that any thermodynamic system with N fluctuating spin components, or, more generally, N fluctuating thermodynamic variables at the critical point, would be described by the same fixed point field theory with N scalar fields. A natural candidate for this fixed point would be the Wilson-Fisher fixed point of the $O(N)$-symmetric ϕ^4 theory discussed in Chapter 11. We will now describe the computation of critical exponents for general values of N in this theory.

As a first step, we should compute the values of the functions $\beta(\lambda)$, $\gamma(\lambda)$, and $\gamma_{\phi^2}(\lambda)$ in four dimensions. This computation parallels the analysis done in Chapter 12 for ordinary ϕ^4 theory, so we will only indicate the changes that need to be made for this case. Just as in ordinary ϕ^4 theory, the propagator of the massless $O(N)$-symmetric theory receives no field strength corrections in one-loop order, and so the one-loop term in $\gamma(\lambda)$ again vanishes. In Problem 13.2, we compute the leading, two-loop, contribution to $\gamma(\lambda)$ in $O(N)$-symmetric ϕ^4 theory:

$$\gamma = (N+2)\frac{\lambda^2}{4(8\pi^2)^2} + \mathcal{O}(\lambda^3). \tag{13.47}$$

The one-loop contribution to the β function in ϕ^4 theory is derived from the one-loop vertex counterterm δ_λ, given in Eq. (12.44). For the $O(N)$-symmetric case, we computed the divergent part of the corresponding vertex counterterm in Section 11.2; from Eq. (11.22),

$$\delta_\lambda = \frac{\lambda^2}{(4\pi)^{d/2}}(N+8)\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}} + \text{finite}. \tag{13.48}$$

Following the logic to Eq. (12.46), or using Eq. (12.54), we find

$$\beta = (N+8)\frac{\lambda^2}{8\pi^2} + \mathcal{O}(\lambda^3). \tag{13.49}$$

This reduces to the β function of ϕ^4 theory if we set $N = 1$ and replace $\lambda \to \lambda/6$, as indicated below Eq. (11.5). Finally, to compute γ_{ϕ^2}, we must repeat the computation done at the end of Section 12.4. If we consider, instead of (12.113), the Green's function $\langle \phi^i(p)\phi^j(q)\phi^2(k)\rangle$, and replace the vertex of ϕ^4 theory by the four-point vertex following from the Lagrangian (11.5), the factor $(-i\lambda)$ in the first line of (12.114) is replaced by

$$(-2i\lambda)[\delta^{ij}\delta^{k\ell} + \delta^{ik}\delta^{j\ell} + \delta^{i\ell}\delta^{jk}]\cdot\delta^{k\ell} = -2i\lambda(N+2)\delta^{ij}.$$

Then

$$\gamma_{\phi^2} = (N+2)\frac{\lambda}{8\pi^2} + \mathcal{O}(\lambda^2). \tag{13.50}$$

Next, we consider the same theory in $(4-\epsilon)$ dimensions. The β function now becomes

$$\beta = -\epsilon\lambda + (N+8)\frac{\lambda^2}{8\pi^2}, \tag{13.51}$$

so there is a Wilson-Fisher fixed point at

$$\lambda_* = \frac{8\pi^2}{(N+8)\epsilon}. \tag{13.52}$$

At this fixed point, we find

$$\gamma(\lambda_*) = \frac{N+2}{4(N+8)^2}\epsilon^2 + \cdots, \qquad \gamma_{\phi^2}(\lambda_*) = \frac{N+2}{N+8}\epsilon + \cdots. \tag{13.53}$$

From these two results, we can work out predictions for the whole set of critical exponents to order ϵ. As an example, inserting (13.53) into (13.19), we find

$$\nu^{-1} = 2 - \frac{N+2}{N+8}\epsilon + \mathcal{O}(\epsilon^2), \tag{13.54}$$

as claimed at the end of Section 12.5.

In our discussion in Chapter 12, we claimed that the predictions of critical exponents are in rough agreement with experimental data. However, by computing to higher order, one can obtain a much more precise comparison of theory and experiment. The ϵ expansion of critical exponents has now been worked out through order ϵ^5. More impressively, the λ expansion for critical exponents in $d=3$ has been worked out through order λ^9. By summing this perturbation series, it is possible to obtain very precise estimates of the anomalous dimensions $\gamma(\lambda_*)$ and $\gamma_{\phi^2}(\lambda_*)$ and, through them, precise predictions for the critical exponents.

A comparison of these values to direct determinations of the critical exponents is given in Table 13.1. The column labeled 'QFT' gives values of critical exponents obtained by anomalous dimension calculations using ϕ^4 perturbation theory in three dimensions. The column labeled 'Experiment' lists a selection of experimental determinations of the critical exponents in a variety of systems. These include the liquid-gas critical point in Xe, CO_2, and other fluids, the critical point in binary fluid mixtures with liquid-liquid phase separation, the order-disorder transition in the atomic arrangement of the Cu–Zn alloy β-brass, the superfluid transition in ^{4}He, and the order-disorder transitions in ferromagnets (EuO, EuS, Ni) and antiferromagnets ($RbMnF_3$). The agreement between experimental determinations of the exponents in different systems is a direct test of universality. For the case of systems with a single order parameter ($N=1$), there is a remarkable diversity of physical systems that are characterized by the same critical exponents.

The column labeled 'Lattice' contains estimates of critical exponents in abstract lattice statistical mechanical models. For these simplified models, the statistical mechanical partition function can be calculated in an expansion for large temperature. With some effort, these expansions can be carried out to

Table 13.1. Values of Critical Exponents for Three-Dimensional Statistical Systems

Exponent	Landau	QFT	Lattice	Experiment	
$N = 1$ Systems:					
γ	1.0	1.241 (2)	1.239 (3)	1.240 (7)	binary liquid
				1.22 (3)	liquid-gas
				1.24 (2)	β-brass
ν	0.5	0.630 (2)	0.631 (3)	0.625 (5)	binary liquid
				0.65 (2)	β-brass
α	0.0	0.110 (5)	0.103 (6)	0.113 (5)	binary liquid
				0.12 (2)	liquid-gas
β	0.5	0.325 (2)	0.329 (9)	0.325 (5)	binary liquid
				0.34 (1)	liquid-gas
η	0.0	0.032 (3)	0.027(5)	0.016 (7)	binary liquid
				0.04 (2)	β-brass
$N = 2$ Systems:					
γ	1.0	1.316 (3)	1.32 (1)		
ν	0.5	0.670 (3)	0.674 (6)	0.672 (1)	superfluid ^{4}He
α	0.0	−0.007 (6)	0.01 (3)	−0.013 (3)	superfluid ^{4}He
$N = 3$ Systems:					
γ	1.0	1.386 (4)	1.40 (3)	1.40 (3)	EuO, EuS
				1.33 (3)	Ni
				1.40 (3)	$RbMnF_3$
ν	0.5	0.705 (3)	0.711 (8)	0.70 (2)	EuO, EuS
				0.724 (8)	$RbMnF_3$
α	0.0	−0.115 (9)	−0.09 (6)	−0.011 (2)	Ni
β	0.5	0.365 (3)	0.37 (5)	0.37 (2)	EuO, EuS
				0.348 (5)	Ni
				0.316 (8)	$RbMnF_3$
η	0.0	0.033 (4)	0.041 (14)		

The values of critical exponents in the column 'QFT' are obtained by resumming the perturbation series for anomalous dimensions at the Wilson-Fisher fixed point in $O(N)$-symmetric ϕ^4 theory in three dimensions. The values in the column 'Lattice' are based on analysis of high-temperature series expansions for lattice statistical mechanical models. The values in the column 'Experiment' are taken from experiments on critical points in the systems described. In all cases, the numbers in parentheses are the standard errors in the last displayed digits. This table is based on J. C. Le Guillou and J. Zinn-Justin, *Phys. Rev.* **B21**, 3976 (1980), with some values updated from J. Zinn-Justin (1993), Chapter 27. A full set of references for the last two columns can be found in these sources.

15 terms or more. By resumming these series, one can obtain direct theoretical estimates of the critical exponents, with an accuracy comparable to that of the best experiments. The comparison between these values and experiment tests the identification of experimental systems with the simple Hamiltonians that were the starting point for our renormalization group analysis.

The agreement of all three types of determinations of the critical exponents presents an impressive picture. The picture is certainly not perfect, and a careful inspection of Table 13.1 reveals some significant discrepancies. But, in general, the evidence is compelling that quantum field theory provides the basic explanation for the thermodynamic critical behavior of a broad range of physical systems.

13.2 Critical Behavior in Four Dimensions

Now that we have discussed the general theory of critical exponents for $d < 4$, let us concentrate some attention on the case $d = 4$. This case obviously has special interest for the applications of quantum field theory to elementary particle physics. In addition, we now know that $d = 4$ lies on a boundary at which the Wilson-Fisher fixed point disappears. We would like to understand the special behavior of quantum field theory predictions at this boundary.

The most obvious difference between $d < 4$ and $d = 4$ is that, while in the former case the deviation of λ from the fixed point multiplies an irrelevant operator, in the case $d = 4$, λ multiplies a marginal operator. We have seen in Eq. (12.82) that, at small momenta or large distances, the running value of λ still approaches its fixed point, now located at $\lambda = 0$. However, this approach is described by a much slower function, not a power but only a logarithm of the distance scale. Thus it is normally not correct to ignore the deviation of λ from the fixed point. Including this effect, we find additional logarithmic terms, analogous to the dependence of correlation functions on $\log p$ that we already know characterizes a renormalizable field theory.

To give a nontrivial illustration of this logarithmic dependence, we return to a problem that we postponed at the end of Chapter 11. In Eq. (11.81), we obtained the expression for the effective potential of ϕ^4 theory to second order in λ, in the limit of vanishing mass parameter:

$$V_{\text{eff}} = \frac{1}{4}\phi_{\text{cl}}^4\Big[\lambda + \frac{\lambda^2}{(4\pi)^2}(N+8)\Big(\log(\lambda\phi_{\text{cl}}^2/M^2) - \frac{3}{2}\Big)\Big]. \tag{13.55}$$

(Note that we now return to our standard notation, in which M is the renormalization scale and μ is a mass parameter.) This expression seemed to have a minimum for very small values of ϕ_{cl}, but only at values so small that

$$\big|\lambda \log(\lambda\phi_{\text{cl}}^2/M^2)\big| \sim 1. \tag{13.56}$$

Since, at the nth order of perturbation theory, one finds n powers of this logarithm, Eq. (13.56) implies that the higher-order terms in λ are not necessarily negligible. What we need is a technique that sums these terms.

This summation is provided by the Callan-Symanzik equation. From (13.24) or (13.25), the Callan-Symanzik equation for the effective potential in the massless limit of four-dimensional ϕ^4 theory is

$$\left[M\frac{\partial}{\partial M}+\beta(\lambda)\frac{\partial}{\partial\lambda}-\gamma\phi_{\rm cl}\frac{\partial}{\partial\phi_{\rm cl}}\right]V_{\rm eff}(\phi_{\rm cl},M,\lambda)=0. \tag{13.57}$$

As before, we can solve for $V_{\rm eff}$ by combining this equation with the predictions of dimensional analysis. In $d=4$,

$$V_{\rm eff}(\phi_{\rm cl},M,\lambda)=\phi_{\rm cl}^4 v(\phi_{\rm cl}/M,\lambda). \tag{13.58}$$

Then v satisfies

$$\left[\phi_{\rm cl}\frac{\partial}{\partial\phi_{\rm cl}}-\frac{\beta}{1+\gamma}\frac{\partial}{\partial\lambda}+\frac{4\gamma}{1+\gamma}\right]v=0. \tag{13.59}$$

This equation for v can be solved by our standard methods, to give

$$v(\phi/M,\lambda)=v_0(\bar\lambda)\exp\left[-\int_M^{\phi_{\rm cl}} d\log\phi_{\rm cl}\,\frac{4\gamma}{1+\gamma}(\bar\lambda(\phi_{\rm cl}))\right], \tag{13.60}$$

where $\bar\lambda$ satisfies

$$\frac{d}{d\log(\phi_{\rm cl}/M)}\bar\lambda=\frac{\beta(\bar\lambda)}{1+\gamma(\bar\lambda)}. \tag{13.61}$$

However, since we are working only to the order of the leading loop corrections, and since $\gamma(\lambda)$ is zero to this order, we can ignore the exponential in (13.60). In addition, we can ignore the denominator on the right-hand side of (13.61), so that this equation reduces to the more standard form of the equation (12.73) for the running coupling constant. Thus, using the leading-order Callan-Symanzik function, we find

$$V_{\rm eff}(\phi_{\rm cl})=v_0(\bar\lambda(\phi_{\rm cl}))\phi_{\rm cl}^4. \tag{13.62}$$

The function v_0 in (13.62) is not determined by the Callan-Symanzik equation. To find this function, we compare (13.62) to the result (13.55) that we obtained from our explicit one-loop evaluation of the effective potential. The precise constraint is the following: After choosing the function $v_0(\bar\lambda)$, substitute for $\bar\lambda$ the solution (12.82) to the renormalization group equation,

$$\bar\lambda(\phi_{\rm cl})=\frac{\lambda}{1-(\lambda/8\pi^2)\log(\phi_{\rm cl}/M)}. \tag{13.63}$$

Then expand the result in powers of λ and drop terms of order λ^3 and higher. If v_0 is chosen correctly, the result should agree with (13.55). Applying this criterion, we find the following result for the effective potential:

$$V_{\rm eff}(\phi_{\rm cl})=\frac{1}{4}\phi_{\rm cl}^4\left[\bar\lambda+\frac{\bar\lambda^2}{(4\pi)^2}(N+8)\left(\log\bar\lambda-\frac{3}{2}\right)\right], \tag{13.64}$$

where $\bar\lambda$ is given by (13.63).

The error in Eq. (13.64) comes in the determination of v_0 as a power series in $\bar{\lambda}$. Thus this error is of order $\bar{\lambda}^3$. As $\phi_{\rm cl} \to 0$, $\bar{\lambda} \to 0$, and so the representation (13.64) becomes more and more accurate. Thus this formula successfully sums the powers of the dangerous logarithm (13.56). Viewed as a function of $\phi_{\rm cl}$, (13.64) has its minimum at $\phi_{\rm cl} = 0$. Thus the apparent symmetry-breaking minimum of (13.55) is indeed an artifact of the incomplete perturbation expansion and disappears in a more complete treatment. This resolves the question that we raised in Section 11.4. We should note that, in more complicated examples, an apparent symmetry-breaking minimum of the effective potential found in the one-loop order of perturbation theory can survive a renormalization-group analysis. An example is given in the Final Project for Part II.

The procedure we have followed in this argument is called the *renormalization group improvement* of perturbation theory. The technique can be applied equally well to the computation of correlation functions and other predictions of Feynman diagram perturbation theory: One compares the solution of the Callan-Symanzik equation to the result of a straightforward perturbation theory computation to the same order in the coupling constant, choosing the undetermined function in the renormalization group solution in such a way as to reproduce the perturbation theory result. In this way, one finds a more compact formula in which large logarithms such as those in (13.56) are resummed into running coupling constants. This resummation produces the dependence of correlation functions on the logarithm of the mass scale that characterizes a field theory with a marginal or renormalizable perturbation.

In the case of ϕ^4 theory, the running coupling constant goes to zero at small momenta and becomes large at large momenta. Since the error term in improved perturbation theory is a power of $\bar{\lambda}$, the improved perturbation theory becomes accurate at small momenta but goes out of control at large momenta. This accords with our physical intuition: We would expect perturbation theory to be accurate only when the running coupling constant stays small.

In an asymptotically free theory, where the running coupling constant becomes small at large momenta, we can find accurate expressions for correlation functions at large momenta using renormalization-group-improved perturbation theory. In Chapters 17 and 18 we will use this idea as our major tool in analyzing the short-distance behavior of the strong interactions.

13.3 The Nonlinear Sigma Model

To complete our study of scalar field theory, we will discuss a nonlinear theory of scalar fields, whose structure is very different from that of ϕ^4 theory. This theory, called the *nonlinear sigma model*, was first proposed as an alternative description of spontaneous symmetry breaking. It will be interesting to us for three reasons. First, it provides a simple explicit example of an asymptotically free theory. Second, it will give us a second dimensional expansion with which we can study the Wilson-Fisher fixed point. Then we can see where the Wilson-Fisher fixed point goes in the space of Lagrangians for dimensions d well below 4. Finally, we will show that the nonlinear sigma model is exactly solvable in a limit that is different from the standard weak-coupling limit. This solution will give us further insight into the dependence of symmetry breaking on spacetime dimensionality.

The $d = 2$ Nonlinear Sigma Model

We begin our study in two dimensions. In $d = 2$, a scalar field is dimensionless; thus, any theory of scalar fields ϕ^i with a Lagrangian of the form

$$\mathcal{L} = f_{ij}(\{\phi^i\})\partial_\mu\phi^i\partial^\mu\phi^j \tag{13.65}$$

has dimensionless couplings and so is renormalizable. Since any function $f(\{\phi^i\})$ leads to a renormalizable theory, this class of scalar field theories contains an infinite number of marginal parameters. To restrict these possible parameters, we must impose some symmetries on the theory.

A simple choice is to take the scalar fields ϕ^i to form an N-component unit vector field $n^i(x)$, constrained to satisfy

$$\sum_{i=1}^{N}\left|n^i(x)\right|^2 = 1. \tag{13.66}$$

If we insist that the field theory has $O(N)$ symmetry, the function f in (13.65) can depend only on the invariant length of $\vec{n}(x)$, which is constrained by (13.66). Thus, the most general possible choice for f is a constant. Similarly, the only possible nonderivative interaction $g(\{n^i\})$ that one might add to (13.65) is a constant, and this would have no effect on the Green's functions of $\vec{n}$. With these restrictions, the most general Lagrangian one can build from $\vec{n}(x)$ with two derivatives and $O(N)$ symmetry is

$$\mathcal{L} = \frac{1}{2g^2}\left|\partial_\mu\vec{n}\right|^2. \tag{13.67}$$

This theory has a straightforward physical interpretation. It is a phenomenological description of a system with $O(N)$ symmetry spontaneously broken by the vacuum expectation value of a field that transforms as a vector of $O(N)$. Consider, for example, the situation in N-component ϕ^4 theory in its spontaneously broken phase. The field ϕ^i acquires a vacuum expectation value, which we can write in terms of a magnitude and a direction parameterized by a unit vector

$$\langle\phi^i\rangle = \phi_0 n^i(x). \tag{13.68}$$

The fluctuations of ϕ_0 correspond to a massive field, the field called σ in Chapter 11. The fluctuations of the direction of the unit vector $\vec{n}(x)$ correspond to the $N-1$ Goldstone bosons. Notice that $\vec{n}$ has N components subject to the one constraint (13.66), and so contains $N-1$ degrees of freedom. Formally, the nonlinear sigma model is the limit of ϕ^4 theory as the mass of the σ field is taken to infinity while ϕ_0 is held constant.

Despite this suggestive connection, we will first analyze the nonlinear sigma model on its own footing as an independent quantum field theory. It is convenient to solve the constraint and parametrize $\vec{n}$ by $N-1$ Goldstone boson fields π^k:

$$n^i = \left(\pi^1, \cdots, \pi^{N-1}, \sigma\right), \tag{13.69}$$

where, by definition,

$$\sigma = (1-\pi^2)^{1/2}. \tag{13.70}$$

The configuration $\pi^k = 0$ corresponds to a uniform state of spontaneous symmetry breaking, oriented in the N direction. The representation (13.69) implies that

$$\left|\partial_\mu n^i\right|^2 = \left|\partial_\mu \vec{\pi}\right|^2 + \frac{(\vec{\pi}\cdot\partial_\mu\vec{\pi})^2}{1-\pi^2}. \tag{13.71}$$

Then the Lagrangian (13.67) takes the form

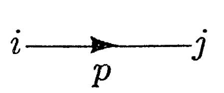
$$\mathcal{L} = \frac{1}{2g^2}\left[\left|\partial_\mu\vec{\pi}\right|^2 + \frac{(\vec{\pi}\cdot\partial_\mu\vec{\pi})^2}{1-\pi^2}\right]. \tag{13.72}$$

Notice that there is no mass term for the field $\vec{\pi}$, as required by Goldstone's theorem.

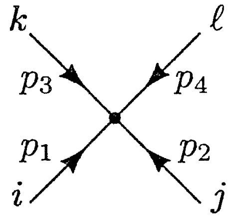
The perturbation theory for the π^k field can be read off straightforwardly by expanding the Lagrangian in powers of π^k:

$$\mathcal{L} = \frac{1}{2g^2}\left|\partial_\mu\vec{\pi}\right| + \frac{1}{2g^2}(\vec{\pi}\cdot\partial_\mu\vec{\pi})^2 + \cdots. \tag{13.73}$$

This leads to the Feynman rules shown in Fig. 13.1, plus additional vertices with all even numbers of π^k fields. Since the Lagrangian (13.67) is the most general $O(N)$-symmetric Lagrangian with dimensionless coefficients that can be built out of these fields, the theory can be made finite by renormalization of the coupling constant g and $O(N)$-symmetric rescaling of the fields π^k and σ. In renormalized perturbation theory, there are divergences and counterterms



$$= \frac{ig^2}{p^2}\delta^{ij}$$



$$\begin{aligned} = -\frac{i}{g^2}\big[&(p_1+p_2)\cdot(p_3+p_4)\delta^{ij}\delta^{k\ell} \\ &+(p_1+p_3)\cdot(p_2+p_4)\delta^{ik}\delta^{j\ell} \\ &+(p_1+p_4)\cdot(p_2+p_3)\delta^{i\ell}\delta^{jk}\big] \end{aligned}$$

Figure 13.1. Feynman rules for the nonlinear sigma model.

for each possible $2n$-π vertex; however, these counterterms are all related by the basic requirement that the bare Lagrangian preserve the $O(N)$ symmetry.

We now compute the Callan-Symanzik functions for this theory. Since the theory is renormalizable, its Green's functions obey the Callan-Symanzik equation for some functions β, γ. Explicitly,

$$\Big[M\frac{\partial}{\partial M}+\beta(g)\frac{\partial}{\partial g}+n\gamma(g)\Big]G^{(n)}=0, \tag{13.74}$$

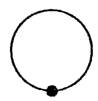
where $G^{(n)}$ is a Green's function of n fields π^k or σ. To identify the β and γ functions, to the leading order in perturbation theory, we compute two simple Green's functions to one-loop order and then see what forms are necessary if the Callan-Symanzik equation is to be satisfied.

The first Green's function we consider is

$$G^{(1)}=\langle\sigma(x)\rangle\,. \tag{13.75}$$

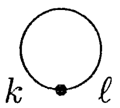
Expanding the definition (13.70), we find

$$\langle\sigma(0)\rangle = 1-\tfrac{1}{2}\left\langle\pi^2(0)\right\rangle+\cdots \quad = \quad 1 \quad + \tag{13.76}$$



To evaluate this formula, we use the propagator of Fig. 13.1 to compute

$$\left\langle\pi^k(0)\pi^\ell(0)\right\rangle = \tag{13.77}$$



$$= \int\frac{d^dk}{(2\pi)^d}\frac{ig^2}{k^2-\mu^2}\delta^{k\ell}.$$

We have added a small mass μ as an infrared cutoff. Then

$$\left\langle\pi^k(0)\pi^\ell(0)\right\rangle = \frac{g^2}{(4\pi)^{d/2}}\frac{\Gamma(1-\frac{d}{2})}{(\mu^2)^{1-d/2}}\delta^{k\ell}. \tag{13.78}$$

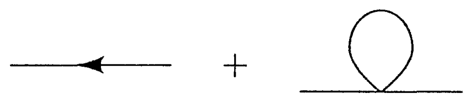
Using this result in our expression for $\langle\sigma\rangle$ and then subtracting at the momentum scale M, we find

$$\begin{aligned}\langle\sigma\rangle &= 1-\frac{1}{2}(N-1)\frac{g^2}{(4\pi)^{d/2}}\Gamma(1-\tfrac{d}{2})\left(\frac{1}{(\mu^2)^{1-d/2}}-\frac{1}{(M^2)^{1-d/2}}\right)+\mathcal{O}(g^4) \\ &\underset{d\to 2}{\longrightarrow} 1-\frac{g^2(N-1)}{8\pi}\log\frac{M^2}{\mu^2}+\mathcal{O}(g^4). \end{aligned} \tag{13.79}$$

This expression satisfies the Callan-Symanzik equation to order g^2 only if

$$\gamma(g) = \frac{g^2(N-1)}{4\pi} + \mathcal{O}(g^4). \tag{13.80}$$

Next, consider the π^k two-point function,



$$\begin{aligned}\langle \pi^k(p)\pi^\ell(-p)\rangle &= \quad + \quad + \cdots \\ &= \frac{ig^2}{p^2}\delta^{k\ell} + \frac{ig^2}{p^2}(-i\Pi^{k\ell})\frac{ig^2}{p^2} + \cdots .\end{aligned} \tag{13.81}$$

In evaluating $\Pi^{k\ell}$ from the Feynman rules in Fig. 13.1, we again encounter the integral (13.77), and also the integral

$$\begin{aligned}\langle \partial_\mu \pi^k(0)\partial^\mu \pi^\ell(0)\rangle &= \int \frac{d^dk}{(2\pi)^d}\frac{ig^2k^2}{k^2-\mu^2}\delta^{k\ell} \\ &= \frac{g^2}{(4\pi)^{d/2}}\frac{\frac{d}{2}\Gamma(-\frac{d}{2})}{(\mu^2)^{-d/2}}.\end{aligned} \tag{13.82}$$

This formula has no pole at $d=0$, and for $d>0$ it is proportional to a positive power of μ^2; hence, we can set this contraction to zero. Then

$$\Pi^{k\ell}(p) = -\delta^{k\ell}p^2\frac{g^2}{(4\pi)^{d/2}}\frac{\Gamma(1-\frac{d}{2})}{(\mu^2)^{1-d/2}}. \tag{13.83}$$

Subtracting at M as above and taking the limit $d\to 2$, we find

$$\begin{aligned}\langle \pi^k(p)\pi^\ell(-p)\rangle &= \frac{ig^2}{p^2}\delta^{k\ell} + \frac{ig^2}{p^2}\Big(+ip^2\frac{g^2}{4\pi}\log\frac{M^2}{\mu^2}\Big)\frac{ig^2}{p^2} + \cdots . \\ &= \frac{i}{p^2}\delta^{k\ell}\Big(g^2 - \frac{g^4}{4\pi}\log\frac{M^2}{\mu^2} + \mathcal{O}(g^6)\Big).\end{aligned} \tag{13.84}$$

Applying the Callan-Symanzik equation to this result gives

$$\begin{aligned}\Big[M\frac{\partial}{\partial M} + \beta(g)\frac{\partial}{\partial g} + 2\gamma(g)\Big]\langle \pi^k(p)\pi^\ell(-p)\rangle &= 0, \\ = \frac{i\delta^{k\ell}}{p^2}\big[-\frac{g^4}{2\pi} + \beta(g)\cdot 2g + 2\gamma(g)\big].\end{aligned} \tag{13.85}$$

Inserting the result (13.80) for $\gamma(g)$, we find

$$\beta(g) = -(N-2)\frac{g^3}{4\pi} + \mathcal{O}(g^5). \tag{13.86}$$

At $N=2$ precisely, the beta function vanishes. This is not an accident but rather is a nontrivial check of our calculation. For $N=2$, we can make the change of variables $\pi^1 = \sin\theta$; then $\sigma = \cos\theta$, and the Lagrangian takes the form

$$\mathcal{L} = \frac{1}{2g^2}(\partial_\mu\theta)^2. \tag{13.87}$$

This is a free field theory for the field $\theta(x)$, so it can have no renormalization group flow.

For $N > 2$, the β function is *negative*: This theory is asymptotically free. The running coupling constant $\bar{g}$ becomes small at small distances and grows large at large distances.

In quantum electrodynamics, we found an appealing physical picture for the sign of the coupling constant evolution. As we discussed in Section 7.5, the process of virtual pair creation makes the vacuum a dielectric medium, which screens electric charge. One would therefore expect the effective Coulomb interaction of charge to decrease at large distances and increase at small distances. It is easy to imagine that a similar screening phenomenon might occur in any quantum field theory. Thus, it is surprising that, in this theory, we have found by explicit calculation that the coupling constant evolution has the opposite sign. What is the physical explanation for this?

In fact, the original derivation of the asymptotic freedom of the nonlinear sigma model, due to Polyakov,† gave a clear physical argument for the sign of the evolution. Now that we have derived the β function by the automatic method of the Callan-Symanzik equation, let us review Polyakov's more physical derivation.

Polyakov analyzed the nonlinear sigma model using Wilson's momentum-slicing technique, which we discussed in Section 12.1. Consider, then, the nonlinear sigma model defined with a momentum cutoff in place of the dimensional regulator. As in Section 12.1, we work in Euclidean space with initial cutoff Λ.

The original integration variables are the Fourier components of the unit vector field $n^i(x)$. We wish to integrate out of the functional integral those Fourier components corresponding to momenta k in the range $b\Lambda \le |k| < \Lambda$. If the remaining components are Fourier-transformed back to coordinate space, they describe a coarse-grained average of the original unit vector field. This averaged field can be rescaled so that it is again a unit vector at each point. Call this averaged and rescaled field $\tilde{n}^i$. Then we can write the relation of n^i and $\tilde{n}^i$ as follows:

$$n^i(x) = \tilde{n}^i(x)\bigl(1-\phi^2\bigr)^{1/2} + \sum_{a=1}^{N-1} \phi_a(x) e_a^i(x). \tag{13.88}$$

In this equation, the vectors $\vec{e}_a(x)$ form a basis of unit vectors orthogonal to $\tilde{n}(x)$. In Polyakov's picture, $\tilde{n}(x)$ and the $\vec{e}_a(x)$ are slowly varying. On the other hand, the coefficients $\phi_a(x)$ contain only Fourier components in the range $b\Lambda \le |k| < \Lambda$. These are the variables we integrate over to achieve the renormalization group transformation.

---

†A. M. Polyakov, *Phys. Lett.* **59B**, 79 (1975).

To set up the integral over ϕ_a, we first work out

$$\partial_\mu n^i = \partial_\mu \tilde{n}^i (1-\phi^2)^{1/2} - \tilde{n}^i \left(\frac{\phi \cdot \partial_\mu \phi}{(1-\phi^2)^{1/2}} \right) + \partial_\mu \phi_a e^i_a + \phi_a \partial_\mu e^i_a. \tag{13.89}$$

By the definition of $\tilde{n}$, $\vec{e}_a$, these vectors satisfy

$$|\tilde{n}|^2 = 1; \qquad \tilde{n} \cdot \vec{e}_a = 0. \tag{13.90}$$

Taking the derivative of these identities, we find

$$\tilde{n} \cdot \partial_\mu \tilde{n} = 0; \qquad \tilde{n} \cdot \partial_\mu \vec{e}_a + \partial_\mu \tilde{n} \cdot \vec{e}_a = 0. \tag{13.91}$$

Using the identities in (13.90) and (13.91), we can compute the Lagrangian of the nonlinear sigma model through terms quadratic in the ϕ_a:

$$\begin{aligned}\mathcal{L} = \frac{1}{2g^2}|\partial_\mu n^i|^2 = \frac{1}{2g^2}\Big[&|\partial_\mu \tilde{n}^i|^2 (1-\phi^2) + (\partial_\mu \phi_a)^2 + 2(\phi_a \partial^\mu \phi_b)(\vec{e}_a \cdot \partial_\mu \vec{e}_b) \\ &+ \partial^\mu \phi_a \partial_\mu \tilde{n} \cdot e_a + \phi_a \phi_b \partial^\mu \vec{e}_a \cdot \vec{\partial}_\mu e_b + \cdots \Big].\end{aligned} \tag{13.92}$$

We will consider the second term of (13.92) to be the zeroth-order Lagrangian for ϕ_a. Thus,

$$\mathcal{L}_0 = \frac{1}{2g^2}(\partial_\mu \phi_a)^2, \tag{13.93}$$

which gives the propagator

$$\langle \phi_a(p) \phi_b(-p) \rangle = \frac{g^2}{p^2}\delta_{ab}, \tag{13.94}$$

restricted to the momentum region $b\Lambda \leq |p| < \Lambda$. This propagator can be used to integrate the remaining terms of the Lagrangian over the ϕ_a. Borrowing the integrals from the derivation of (13.84), we can set

$$\langle \phi_a(0) \partial_\mu \phi_b(0) \rangle = \langle \partial^\mu \phi_a(0) \partial_\mu \phi_b(0) \rangle = 0 \tag{13.95}$$

and

$$\langle \phi_a(0) \phi_b(0) \rangle = \delta_{ab} \frac{g^2}{4\pi} \log \frac{\Lambda^2}{(b\Lambda)^2}. \tag{13.96}$$

Then, after the integral over ϕ, the new Lagrangian is given approximately by

$$\mathcal{L}_{\text{eff}} = \frac{1}{2g^2}\Big[|\partial_\mu \tilde{n}|^2 (1 - \langle \phi^2 \rangle) + \langle \phi_a \phi_b \rangle \, \partial_\mu \vec{e}_a \cdot \partial^\mu \vec{e}_b + \mathcal{O}(g^4) \Big], \tag{13.97}$$

where the expectation values of ϕ_a are given by (13.96).

To simplify this further, we must simplify the structure $(\partial_\mu \vec{e}_a)^2$ that appears in the second term of (13.97). Introduce a complete basis of vectors:

$$(\partial_\mu \vec{e}_a)^2 = (\tilde{n} \cdot \partial_\mu \vec{e}_a)^2 + (\vec{e}_c \cdot \partial_\mu \vec{e}_a)^2. \tag{13.98}$$

The second term on the right is a new structure, associated with the torsion of the coordinate system for e_a; it turns out to correspond to an irrelevant operator induced by the renormalization procedure. The first term, however, can be put into a familiar form by using the two identities (13.91):

$$(\tilde{n}\cdot\partial_\mu\vec{e}_a)^2 = (\vec{e}_a\cdot\partial_\mu\tilde{n})^2 = (\partial_\mu\tilde{n})^2. \tag{13.99}$$

Then

$$\begin{aligned}\mathcal{L}_{\text{eff}} &= \frac{1}{2g^2}\left[|\partial_\mu\tilde{n}|^2\left(1-(N-1)\frac{g^2}{4\pi}\log\frac{1}{b^2}+\frac{g^2}{4\pi}\log\frac{1}{b^2}\right)+\cdots\right]\\ &= \frac{1}{2}\left(g^2+\frac{g^4}{2\pi}(N-2)\log\frac{1}{b}+\cdots\right)^{-1}|\partial_\mu\tilde{n}|^2.\end{aligned} \tag{13.100}$$

The quantity in parentheses is the square of a running coupling constant. To the order of our calculation, this quantity satisfies

$$\frac{d}{d\log b}\bar{g} = -(N-2)\frac{\bar{g}^3}{4\pi}, \tag{13.101}$$

in agreement with (13.86).

In this calculation, the sign of the coupling constant renormalization comes from the fact that the effective length of the unit vector $\vec{n}$ is reduced by averaging over short-wavelength fluctuations. This lowers the effective action associated with a configuration in which the direction of $\vec{n}$ changes over a displacement Δx (see Fig. 13.2). Looking back at (13.67), we see that a decrease of the magnitude of $\mathcal{L}$ for the same configuration of $\vec{n}$ can be interpreted as an *increase* of the effective coupling. Thus the nonlinear sigma model is more strongly coupled, or, in terms of the physical configuration of the $\vec{n}$ field, more disordered, at large distances.

Our calculation implies that, if any two-dimensional statistical system apparently has spontaneously broken symmetry and Goldstone bosons, then, at large distances, the ordering disappears. This is an unexpected conclusion. However, this conclusion is in accord with a theorem proved by Mermin and Wagner‡ that a two-dimensional system with a continuous symmetry cannot support an ordered state in which a symmetry-breaking field has a nonzero vacuum expectation value. This theorem applies to the case $N=2$ as well as to $N>2$. We have motivated this theorem in Problem 11.1.

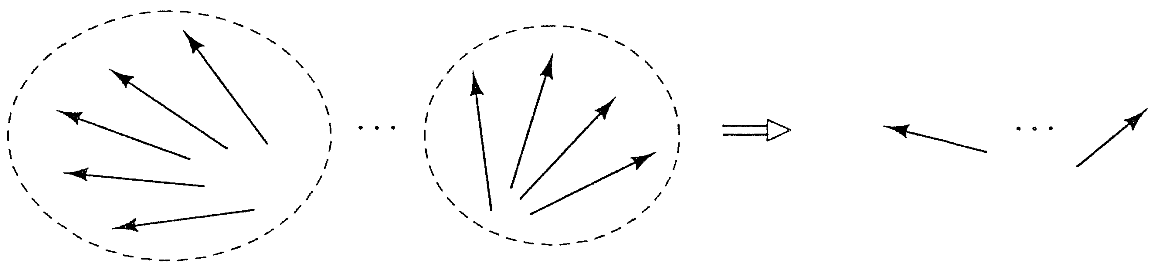
The Nonlinear Sigma Model for $2<d<4$

We now extend the results of this analysis to dimensions $d>2$. In general d, we will continue to define the action of the nonlinear sigma model by

$$\int d^dx\,\mathcal{L} = \int d^dx\,\frac{1}{2g^2}(\partial_\mu\vec{n})^2, \tag{13.102}$$

where $\vec{n}$ is still dimensionless, since it obeys the constraint $|\vec{n}|^2=1$. Thus g has the dimensions $(\text{mass})^{(d-2)/2}$. We define a dimensionless coupling by

‡N. D. Mermin and H. Wagner, *Phys. Rev. Lett.* **17**, 1133 (1966).



$$\int d^2x \, \frac{1}{2g^2} (\partial_\mu n)^2 \quad \Longrightarrow \quad \int d^2x \, \frac{1}{2\bar{g}^2} (\partial_\mu \tilde{n})^2$$

Figure 13.2. Averaging of the direction of $\vec{n}$, and its interpretation as an increase of the running coupling constant.

writing

$$T = g^2 M^{d-2}, \tag{13.103}$$

just as we did in (12.122). If (13.102) is viewed as the Boltzmann weight of a partition function, then T is a dimensionless variable proportional to the temperature.

From (13.103), we can find the β function for T in d dimensions, in analogy to Eq. (12.131):

$$\beta(T) = (d-2) + 2g\beta^{(2)}(g), \tag{13.104}$$

where the factor of $2g$ in the second term comes from the definition $T \sim g^2$. Since $\vec{n}$ is dimensionless, the γ function is unchanged from the two-dimensional result when expressed in terms of dimensionless couplings. Thus, in $d = 2 + \epsilon$,

$$\begin{aligned} \beta(T) &= +\epsilon T - (N-2)\frac{T^2}{2\pi}; \\ \gamma(T) &= (N-1)\frac{T}{4\pi}. \end{aligned} \tag{13.105}$$

Notice that the β function for T has a nontrivial zero, which approaches $T = 0$ as $\epsilon \to 0$. This zero is located at

$$T_* = \frac{2\pi\epsilon}{N-2}. \tag{13.106}$$

The form of the β function is sketched in Fig. 13.3. In contrast to the Wilson-Fisher zero in $d = 4 - \epsilon$, discussed in Section 12.5, this is an ultraviolet-stable fixed point. The flows to the infrared go out from this fixed point. Since T is proportional to the temperature of the corresponding statistical system, $t \to 0$ is the state of complete order, while $t \gg 1$ is the state of complete disorder. This agrees with the intuition that accompanied Polyakov's derivation of the β function. The fixed point T_* corresponds to the critical temperature. Thus, the critical temperature tends to zero as $d \to 2$, in accord with the Mermin-Wagner theorem.

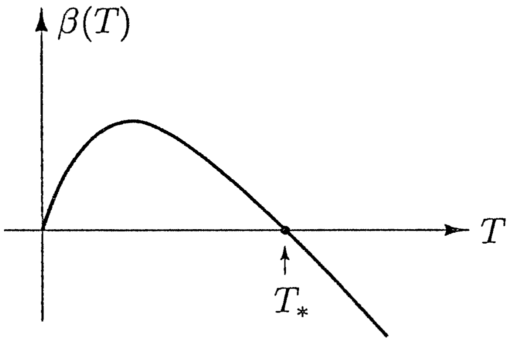


Figure 13.3. The form of the β function in the nonlinear sigma model for $d > 2$.

We can now compute the critical exponents of the nonlinear sigma model in an expansion in $\epsilon = d - 2$. The exponent η is given straightforwardly by

$$\eta = 2\gamma(T_*) = \frac{\epsilon}{N-2}. \tag{13.107}$$

To find the second exponent ν, we need to identify the relevant perturbation that corresponds to the renormalization group flow away from the fixed point for $T \neq T_C$. This is just the deviation of T from T_*:

$$\rho_T = T - T_*. \tag{13.108}$$

From the renormalization group equation for the running coupling constant, we find that the running ρ_T obeys

$$\frac{d}{d\log p}\bar{\rho}_T = \left[\frac{d}{dT}\beta(T)\Big|_{T=T_*}\right]\cdot\bar{\rho}_T. \tag{13.109}$$

The quantity in brackets is negative. As in Eqs. (12.134) and (12.137), we can identify this quantity with $(-1/\nu)$: At a momentum $p \ll M$,

$$\bar{\rho}_T(p) = \rho_T\left(\frac{p}{M}\right)^{-1/\nu}; \tag{13.110}$$

thus $\bar{\rho}(p)$ becomes of order 1 at a momentum that is the inverse of $\xi \sim (T-T_*)^{-\nu}$, as required. Using the explicit form of the β function from (13.105), we find

$$\nu = \frac{1}{\epsilon}, \tag{13.111}$$

independent of N to this order in ϵ. (Of course, these results apply only for $N \geq 3$.) The thermodynamic critical exponents can be found from (13.107) and (13.111) using the model-independent relations derived in Section 13.1. When the values found here for ν and η are extrapolated to $d = 3$ (that is, $\epsilon = 1$), the agreement with experiment is not spectacular, but the results at least suggest that the fixed point we have found here may be the continuation of the Wilson-Fisher fixed point to the vicinity of two dimensions.

Exact Solution for Large N

It is possible to obtain further insight into the nature of this fixed point by attacking the nonlinear sigma model using another approach. Since the nonlinear sigma model depends on a parameter N, the number of components of the unit vector, it is reasonable to ask how this model behaves as $N \to \infty$. We now show that if we take this limit holding $g^2 N$ fixed, we can obtain an exact solution to the model with nontrivial behavior.

The manipulations that lead to this solution are most clear if we work in Euclidean space, regarding the Lagrangian as the Boltzmann weight of a spin system. Then we must compute the functional integral

$$Z = \int \mathcal{D}n \, \exp\left[-\int d^dx \, \frac{1}{2g_0^2}(\partial_\mu n)^2\right] \cdot \prod_x \delta(n^2(x) - 1). \tag{13.112}$$

Here g_0 is the bare value of the coupling constant, while the product of delta functions, one at each point, enforces the constraint. Introduce an integral representation of the delta functions; this requires a second functional integral over a Lagrange multiplier variable $\alpha(x)$:

$$Z = \int \mathcal{D}\alpha \mathcal{D}n \, \exp\left[-\int d^dx \, \frac{1}{2g_0^2}(\partial_\mu n)^2 - \frac{i}{2g_0^2}\int d^dx \, \alpha(n^2 - 1)\right]. \tag{13.113}$$

Now the variable n is unconstrained and appears in the exponent only quadratically. Thus, we can integrate over n, to obtain

$$\begin{aligned} Z &= \int \mathcal{D}\alpha \, \left(\det[-\partial^2 + i\alpha(x)]\right)^{-N/2} \exp\left[\frac{i}{2g_0^2}\int d^dx \, \alpha\right]. \\ &= \int \mathcal{D}\alpha \, \exp\left[-\frac{N}{2}\operatorname{tr}\log(-\partial^2 + i\alpha) + \frac{i}{2g_0^2}\int d^dx \, \alpha\right]. \end{aligned} \tag{13.114}$$

Since we are taking the limit $N \to \infty$ with $g_0^2 N$ held fixed, both terms in the exponent are of order N. Thus it makes sense to evaluate the integral by steepest descents. This entails dominating the integral by the value of the function $\alpha(x)$ that minimizes the exponent. To determine this configuration, we compute the functional derivative of the exponent with respect to $\alpha(x)$. This gives the variational equation

$$\frac{N}{2}\langle x| \frac{1}{-\partial^2 + i\alpha} |x\rangle = \frac{1}{2g_0^2}. \tag{13.115}$$

The left-hand side of this equation must be constant and real; thus, we should look for a solution in which $\alpha(x)$ is constant and pure imaginary. Write

$$\alpha(x) = -im^2; \tag{13.116}$$

then m^2 obeys

$$N\int \frac{d^dk}{(2\pi)^d}\frac{1}{k^2 + m^2} = \frac{1}{g_0^2}. \tag{13.117}$$

We will study this equation first in $d = 2$. If we define the integral in (13.117) with a momentum cutoff, we can evaluate this integral and find the equation for m:

$$\frac{N}{2\pi}\log\frac{\Lambda}{m} = \frac{1}{g_0^2}. \tag{13.118}$$

We can make this equation finite by the renormalization

$$\frac{1}{g_0^2} = \frac{1}{g^2} + \frac{N}{2\pi}\log\frac{\Lambda}{M}, \tag{13.119}$$

which introduces an arbitrary renormalization scale M. Then we can solve for m, to find

$$m = M\exp\left[-\frac{2\pi}{g^2 N}\right]. \tag{13.120}$$

This is a nonzero, $O(N)$-invariant mass term for the N unconstrained components of $\vec{n}$. In this solution, $\langle\vec{n}\rangle = 0$ and the symmetry is unbroken, for any value of g^2 or T.

The solution of the theory does depend on the arbitrary renormalization scale M; this dependence simply reflects the arbitrariness of the definition of the renormalized coupling constant. The statement that m follows unambiguously from an underlying theory with fixed bare coupling and cutoff is precisely the statement that m obeys the Callan-Symanzik equation with no overall rescaling:

$$\left[M\frac{\partial}{\partial M} + \beta(g^2)\frac{\partial}{\partial g}\right]m(g^2, M) = 0. \tag{13.121}$$

Using the large-N limit of (13.86),

$$\beta(g) = -\frac{g^3 N}{4\pi}, \tag{13.122}$$

it is easy to check that (13.121) is satisfied. Conversely, the validity of (13.121) with (13.122) tells us that Eq. (13.122) is an *exact* representation of the β function to all orders in $g^2 N$ in the limit of large N. The corrections to (13.122) are of order $(1/N)$ or, equivalently, of order g^2 with no compensatory factor of N. Equation (13.122) agrees with our earlier calculation (13.86) to this order.

Now let us redo this exercise in $d > 2$. In this case, the integral in (13.117) diverges as a power of the cutoff. Even when the dependence on Λ is removed by renormalization, this change in behavior leads to a change in the dependence of the integral on m, which has important physical implications.

It is not difficult to work out the integral in (13.117) as an expansion in (Λ/m). One finds:

$$\int\frac{d^d k}{(2\pi)^d}\frac{1}{k^2+m^2} = \begin{cases} C_1\Lambda^{d-2} - C_2 m^{d-2} + \cdots & \text{for } d < 4, \\ C_1\Lambda^{d-2} - \tilde{C}_2 m^2\Lambda^{d-4} + \cdots & \text{for } d > 4, \end{cases} \tag{13.123}$$

where C_1, C_2, $\tilde{C}_2$ are some functions of d. In particular,

$$C_1 = \left[2^{d-1}\pi^{(d+1)/2}\Gamma\left(\frac{d-1}{2}\right)(d-2)\right]^{-1}. \tag{13.124}$$

In $d > 4$, the first derivative of the integral with respect to m^2 is smooth as $m^2 \to 0$; this is the reason for the change in behavior.

In the case $d = 2$, the left-hand side of (13.117) covered the whole range from 0 to ∞ as m was varied; thus, we could always find a solution for any value of g_0^2. In $d > 2$, this is no longer true. Equation (13.117) can be solved for m only if Ng_0^2 is greater than the critical value

$$Ng_C^2 = \left(C_1\Lambda^{d-2}\right)^{-1}. \tag{13.125}$$

Just at the boundary, $m = 0$. For bare couplings weaker than (13.125), it is possible to lower the value of the effective action by giving one component of $\vec{n}$ a vacuum expectation value while keeping the other components massless. Thus (13.125) is the criterion for the second-order phase transition in this model. Equation (13.124) implies that the critical value of g_0^2, which is proportional to the critical temperature, goes to zero as $d \to 2$, in accord with our renormalization-group analysis.

In the symmetric phase of the nonlinear sigma model, the mass m determines the exponential fall-off of correlations, so $\xi = m^{-1}$. Thus we can determine the exponent ν by solving for the dependence of m on the deviation of g_0^2 from the critical temperature. Write

$$t = \frac{g_0^2 - g_C^2}{g_C^2}. \tag{13.126}$$

Then, in $2 < d < 4$, we can use (13.123) to solve (13.117) for m for small values of t. This gives

$$\frac{1}{Ng_C^4} \cdot t = C_2 m^{d-2}, \tag{13.127}$$

which implies $m \sim t^\nu$ with

$$\nu = \frac{1}{d-2}, \qquad 2 < d < 4. \tag{13.128}$$

Similarly,

$$\nu = \frac{1}{2}, \qquad d \geq 4. \tag{13.129}$$

The discontinuity in the dependence of ν on d is exactly what we predicted from renormalization group analysis. For $d > 4$, the value of ν goes over to the prediction of naive dimensional analysis. The value of ν given by (13.128) is in precise agreement with (13.111), in the expansion $\epsilon = d - 2$, and with the $N \to \infty$ limit of (12.142), in the expansion $\epsilon = 4 - d$. Apparently, all of our results for critical exponents mesh in a very satisfying way.

By combining all of our results, we arrive at a pleasing picture of the behavior of scalar field theory as a function of spacetime dimensionality. Above four dimensions, any scalar field interaction is irrelevant and the expected behavior is trivial. Just at four dimensions, the coupling constant tends to zero only logarithmically at large scale, giving rise to a renormalizable theory with predictions such as those in Section 13.2. Below four dimensions, the theory is intrinsically a theory of interacting scalar fields, dominated by the Wilson-Fisher fixed point. The coupling at this fixed point is small near four dimensions but grows large as the dimensionality decreases. Finally (for $N > 2$), as $d \to 2$, the fixed-point theory approaches the weak-coupling limit of a completely different Lagrangian with the same symmetries, the nonlinear sigma model.

This evolution of the behavior of the model as a function of d illustrates the main point of the previous two chapters: The qualitative behavior of a quantum field theory is determined not by the fundamental Lagrangian, but rather by the nature of the renormalization group flow and its fixed points. These, in turn, depend only on the basic symmetries that are imposed on the family of Lagrangians that flow into one another. This conclusion signals, at the deepest level, the importance of symmetry principles in determining the fundamental laws of physics.

Problems

13.1 Correction-to-scaling exponent. For critical phenomena in $4-\epsilon$ dimensions, the irrelevant contributions that disappear most slowly are those associated with the deviation of the coupling constant λ from its fixed-point value. This gives the most important nonuniversal correction to the scaling laws derived in Section 13.1. By studying the solution of the Callan-Symanzik equation, show that if the bare value of λ differs slightly from λ_*, the Gibbs free energy receives a correction

$$G(M,t) \to G(M,t)\cdot\left(1+(\lambda-\lambda_*)t^{\omega\nu}\hat{k}(tM^{-1/\beta})\right).$$

This formula defines a new critical exponent ω, called the *correction-to-scaling exponent.* Show that

$$\omega = \frac{d}{d\lambda}\beta\Big|_{\lambda_*} = \epsilon + \mathcal{O}(\epsilon^2).$$

13.2 The exponent η. By combining the result of Problem 10.3 with an appropriate renormalization prescription, show that the leading term in $\gamma(\lambda)$ in ϕ^4 theory is

$$\gamma = \frac{\lambda^2}{12(4\pi)^2}.$$

Generalize this result to the $O(N)$-symmetric ϕ^4 theory to derive Eq. (13.47). Compute the leading-order (ϵ^2) contribution to η.

13.3 The CP^N model. The nonlinear sigma model discussed in the text can be thought of as a quantum theory of fields that are coordinates on the unit sphere. A slightly more complicated space of high symmetry is complex projective space,

CP^N. This space can be defined as the space of $(N+1)$-dimensional complex vectors $(z_1, \ldots, z_{N+1})$ subject to the condition

$$\sum_j |z_j|^2 = 1,$$

with points related by an overall phase rotation identified, that is,

$$(e^{i\alpha} z_1, \ldots, e^{i\alpha} z_{N+1}) \quad \text{identified with} \quad (z_1, \ldots, z_{N+1}).$$

In this problem, we study the two-dimensional quantum field theory whose fields are coordinates on this space.

(a) One way to represent a theory of coordinates on CP^N is to write a Lagrangian depending on fields $z_j(x)$, subject to the constraint, which also has the local symmetry

$$z_j(x) \to e^{i\alpha(x)} z_j(x),$$

independently at each point x. Show that the following Lagrangian has this symmetry:

$$\mathcal{L} = \frac{1}{g^2}\left[|\partial_\mu z_j|^2 + |z_j^* \partial_\mu z_j|^2\right].$$

To prove the invariance, you will need to use the constraint on the z_j, and its consequence

$$z_j^* \partial_\mu z_j = -(\partial_\mu z_j^*) z_j.$$

Show that the nonlinear sigma model for the case $N = 3$ can be converted to the CP^N model for the case $N = 1$ by the substitution

$$n^i = z^* \sigma^i z,$$

where σ^i are the Pauli sigma matrices.

(b) To write the Lagrangian in a simpler form, introduce a scalar Lagrange multiplier λ which implements the constraint and also a vector Lagrange multiplier A_μ to express the local symmetry. More specifically, show that the Lagrangian of the CP^N model is obtained from the Lagrangian

$$\mathcal{L} = \frac{1}{g^2}\left[|D_\mu z_j|^2 - \lambda(|z_j|^2 - 1)\right],$$

where $D_\mu = (\partial_\mu + iA_\mu)$, by functionally integrating over the fields λ and A_μ.

(c) We can solve the CP^N model in the limit $N \to \infty$ by integrating over the fields z_j. Show that this integral leads to the expression

$$Z = \int \mathcal{D}A\mathcal{D}\lambda \exp\left[-N \operatorname{tr}\log(-D^2 - \lambda) + \frac{i}{g^2}\int d^2x \lambda\right],$$

where we have kept only the leading terms for $N \to \infty$, $g^2 N$ fixed. Using methods similar to those we used for the nonlinear sigma model, examine the conditions for minimizing the exponent with respect to λ and A_μ. Show that these conditions have a solution at $A_\mu = 0$ and $\lambda = m^2 > 0$. Show that, if g^2 is renormalized at the scale M, m can be written as

$$m = M \exp\left[-\frac{2\pi}{g^2 N}\right].$$

(d) Now expand the exponent about $A_\mu = 0$. Show that the first nontrivial term in this expansion is proportional to the vacuum polarization of massive scalar fields. Evaluate this expression using dimensional regularization, and show that it yields a standard kinetic energy term for A_μ. Thus the strange nonlinear field theory that we started with is finally transformed into a theory of $(N+1)$ massive scalar fields interacting with a massless photon.

The Coleman-Weinberg Potential

In Chapter 11 and Section 13.2 we discussed the effective potential for an $O(N)$-symmetric ϕ^4 theory in four dimensions. We computed the perturbative corrections to this effective potential, and used the renormalization group to clarify the behavior of the potential for small values of the scalar field mass. After all this work, however, we found that the qualitative dependence of the theory on the mass parameter was unchanged by perturbative corrections. The theory still possessed a second-order phase transition as a function of the mass. The loop corrections affected this picture only in providing some logarithmic corrections to the scaling behavior near the phase transition.

However, loop corrections are not always so innocuous. For some systems, they can change the structure of the phase transition qualitatively. This Final Project treats the simplest example of such a system, the *Coleman-Weinberg model.* The analysis of this model draws on a broad variety of topics discussed in Part II; it provides a quite nontrivial application of the effective potential formalism and the use of the renormalization group equation. The phenomenon displayed in this exercise reappears in many contexts, from displacive phase transitions in solids to the thermodynamics of the early universe.

This problem makes use of material in starred sections of the book, in particular, Sections 11.3, 11.4, and 13.2. Parts (a) and (e), however, depend only on the unstarred material of Part II. We recommend part (e) as excellent practice in the computation of renormalization group functions.

The Coleman-Weinberg model is the quantum electrodynamics of a scalar field in four dimensions, considered for small values of the scalar field mass. The Lagrangian is

$$\mathcal{L} = -\tfrac{1}{4}(F_{\mu\nu})^2 + (D_\mu\phi)^\dagger D^\mu\phi - m^2\phi^\dagger\phi - \tfrac{\lambda}{6}(\phi^\dagger\phi)^2,$$

where $\phi(x)$ is a complex-valued scalar field and $D_\mu\phi = (\partial_\mu + ieA_\mu)\phi$.

(a) Assume that $m^2 = -\mu^2 < 0$, so that the symmetry $\phi(x) \to e^{-i\alpha}\phi(x)$ is spontaneously broken. Write out the expression for $\mathcal{L}$, expanded around the broken-symmetry state by introducing

$$\phi(x) = \phi_0 + \frac{1}{\sqrt{2}}\big[\sigma(x) + i\,\pi(x)\big],$$

where ϕ_0, $\sigma(x)$, and π are real-valued. Show that the A_μ field acquires a mass. This mechanism of mass generation for vector fields is called the *Higgs mechanism.* We will study it in great detail in Chapter 20.

(b) Working in Landau gauge ($\partial^\mu A_\mu = 0$), compute the one-loop correction to the effective potential $V(\phi_{\rm cl})$. Show that it is renormalized by counterterms for m^2 and λ. Renormalize by minimal subtraction, introducing a renormalization scale M.

(c) In the result of part (b), take the limit $\mu^2 \to 0$. The result should be an effective potential that is scale-invariant up to logarithms containing M. Analyze this expression for λ very small, of order $(e^2)^2$. Show that, with this choice of coupling constants, $V(\phi_{\rm cl})$ has a symmetry-breaking minimum at a value of $\phi_{\rm cl}$ for which no logarithm is large, so that a straightforward perturbation theory analysis should be valid. Thus the $\mu^2 = 0$ theory, for this choice of coupling constants, still has spontaneously broken symmetry, due to the influence of quantum corrections.

(d) Sketch the behavior of $V(\phi_{\rm cl})$ as a function of m^2, on both sides of $m^2 = 0$, for the choice of coupling constants made in part (c).

(e) Compute the Callan-Symanzik β functions for e and λ. You should find

$$\beta_e = \frac{e^3}{48\pi^2}, \qquad \beta_\lambda = \frac{1}{24\pi^2}\left(5\lambda^2 - 18e^2\lambda + 54e^4\right).$$

Sketch the renormalization group flows in the (λ, e^2) plane. Show that every renormalization group trajectory passes through the region of coupling constants considered in part (c).

(f) Construct the renormalization-group-improved effective potential at $\mu^2 = 0$ by applying the results of part (e) to the calculation of part (c). Compute $\langle\phi\rangle$ and the mass of the σ particle as a function of λ, e^2, M. Compute the ratio m_σ/m_A to leading order in e^2, for $\lambda \ll e^2$.

(g) Include the effects of a nonzero m^2 in the analysis of part (f). Show that m_σ/m_A takes a minimum nonzero value as m^2 increases from zero, before the broken-symmetry state disappears entirely. Compute this value as a function of e^2, for $\lambda \ll e^2$.

(h) The Lagrangian of this problem (in its Euclidean form) is equivalent to the Landau free energy for a superconductor in d dimensions, coupled to an electromagnetic field. This expression is known as the Landau-Ginzburg free energy. Compute the β functions for this system and sketch the renormalization group flows for $d = 4 - \epsilon$. Describe the qualitative behavior you would expect for the superconducting phase transition in three dimensions. (For realistic superconductors, the value of e^2—after it is made dimensionless in the appropriate way—is very small. The effect you will find is expected to be important only for $|T - T_C|/T_C < 10^{-5}$.)

Part III

Non-Abelian Gauge Theories

Chapter 14

Invitation: The Parton Model of Hadron Structure

In Part II of this book, we explored the structure of quantum field theories in a formal way. We developed sophisticated calculational algorithms (Chapter 10), derived a formalism for the extraction of scaling laws and asymptotic behavior (Chapter 12), and worked out some of the consequences of spontaneously broken symmetry (Chapter 11). Much of this formalism turned out to have unexpected applications in statistical mechanics. However, we have not yet investigated its implications for elementary particle physics. To do so, we must first ask which particular quantum field theories describe the interactions of elementary particles.

Since the mid-1970s, most high-energy physicists have agreed that the elementary particles that make up matter are a set of fermions, interacting primarily through the exchange of vector bosons. The elementary fermions include the *leptons* (the electron, its heavy counterparts μ and τ, and a neutral, almost massless neutrino corresponding to each of these species), and the *quarks*, whose bound states form the particles with nuclear interactions, mesons and baryons (collectively called hadrons). These fermions interact through three forces: the strong, weak, and electromagnetic interactions. Of these, the *strong interaction* is responsible for nuclear binding and the interactions of the constituents of nuclei, while the *weak interaction* is responsible for radioactive beta decay processes. The *electromagnetic interaction* is the familiar Quantum Electrodynamics, coupled minimally to all charged quarks and leptons. It is not clear that these three forces suffice to explain the most subtle properties of the elementary fermions—we will discuss this question in Chapter 20—but these three forces are certainly the most prominent. All three are now understood to be mediated by the exchange of vector bosons. The equations describing the electromagnetic interaction were discovered by Maxwell, and their quantum mechanical implications have been treated in detail in Part I. The correct theories of the weak and strong interactions were discovered much later.

By the late 1950s, studies of the helicity dependence of weak interaction cross sections and decay rates had shown that the weak interaction involves a coupling of vector currents built of quark and lepton fields.* It was thus natural to assume that the weak interaction is due to the exchange of very heavy vector bosons, and indeed, such bosons, the W and Z particles, were discovered in experiments at CERN in 1982. But a complete theory of the weak interaction must include not only the correct couplings of the bosons to fermions, but also the equations of motion of the boson fields themselves, the analogue, for the W and Z, of Maxwell's equations. Finding the correct form of these equations was not straightforward, because Maxwell's equations prohibit the generation of a mass for the vector particle. The proper reconciliation of the generalized Maxwell equations with the nonzero W and Z masses turned out to require incorporating into the theory a spontaneously broken symmetry. Chapters 20 and 21 treat this subject in some detail, describing the interplay of vector field theories with spontaneously broken symmetry. This interplay leads to new twists and new phenomena, beyond those discussed in our treatment of spontaneous symmetry breaking in Chapter 11. A complete theory of the weak interaction also requires the simultaneous incorporation of the electromagnetic interaction, forming a unified structure as first hypothesized by Glashow, Weinberg, and Salam.

On the other hand, it was for a long time completely obscure that a theory of exchanged vector bosons could correctly describe the strong interaction. Part of the mystery was that quarks do not exist as isolated species. Their existence, and eventually their quantum numbers, had to be deduced from the spectrum of observable strongly interacting particles. But, in addition, there were complications due to the fact that the strong interactions are strong. The Feynman diagram expansion assumes that the coupling constant is small; when the coupling becomes strong, a large number of diagrams are important (if the series converges at all) and it becomes impossible to pick out the contributions of the elementary interaction vertices. The crucial clue that the strong interactions have a vector character arose from what at first seemed to be just another mystery, the observation that the strong interactions turn themselves off when the momentum transfer is large, in a sense that we will now describe.

Almost Free Partons

In Section 5.1 we computed the cross section for the QED process $e^+e^- \to \mu^+\mu^-$. We then remarked that the corresponding cross section for e^+e^- annihilation into hadrons could be computed in the same way, using a simplistic model in which the quarks are treated as noninterating fermions. This method gives a surprisingly accurate formula for the cross section, capturing its most important qualitative features. But we deferred the explanation of this puzzle: How can a model of noninteracting quarks represent the behavior of a force that, under other circumstances, is extremely *strong*?

*For an overview of weak interaction phenomenology, see Perkins (1987), Chapter 7, or any other modern particle physics text.

In fact, there are many circumstances in the study of the strong interaction at high energy in which this force has an unexpectedly weak effect. Historically, the first of these appeared in proton-proton collisions. At high energy, above 10 GeV or so in the center of mass, collisions of protons (or any other hadrons) produce large numbers of pions. One might have imagined that these pions would fill all of the allowed phase space, but, in fact, they are mainly produced with momenta almost collinear with the collision axis. The probability of producing a pion with a large component of momentum transverse to the collision axis falls off exponentially in the value of this transverse momentum, suppressing the production substantially for transverse momenta greater than a few hundred MeV.

This phenomenon of limited transverse momentum led to a picture of a hadron as a loosely bound assemblage of many components. In this picture, a proton struck by another proton would be torn into a cloud of pieces. These pieces would have momenta roughly collinear with the original momentum of the proton and would eventually reform into hadrons moving along the collision axis. By hypothesis, these pieces could not absorb a large momentum transfer. We can characterize this hypothesis mathematically as follows: In a high-energy collision, the momenta of the two initial hadrons are almost lightlike. The shattered pieces of the hadrons, arrayed along the collision axis, also have lightlike momenta parallel to the original momentum vectors. This final state can be produced by exchanging momenta q among the pieces in such a way that, though the components of q might be large, the invariant q^2 is always small. The ejection of a hadron at large transverse momentum would require large (spacelike) q^2, but such a process was very rare. Thus it was hypothesized that hadrons were loose clouds of constituents, like jelly, which could not absorb a large q^2.

This picture of hadronic structure was put to a crucial test in the late 1960s, in the SLAC-MIT deep inelastic scattering experiments.† In these experiments, a 20 GeV electron beam was scattered from a hydrogen target, and the scattering rate was measured for large deflection angles, corresponding to large invariant momentum transfers from the electron to a proton in the target. The large momentum transfer was delivered through the electromagnetic rather than the strong interaction, so that the amount of momentum delivered could be computed from the momentum of the scattered electron. In models in which hadrons were complex and softly bound, very low scattering rates were expected.

Instead, the SLAC-MIT experiments saw a substantial rate for hard scattering of electrons from protons. The total reaction rate was comparable to what would have been expected if the proton were an elementary particle scattering according to the simplest expectations from QED. However, only in rare cases did a single proton emerge from the scattering process. The largest part

†For a description of these experiments and their ramifications, see J. I. Friedman, H. W. Kendall, and R. E. Taylor, *Rev. Mod. Phys.* **63**, 573 (1991).

of the rate came from the *deep inelastic* region of phase space, in which the electromagnetic impulse shattered the proton and produced a system with a large number of hadrons.

How could one reconcile the presence of electromagnetic hard scattering processes with the virtual absence of hard scattering in strong interaction processes? To answer this question, Bjorken and Feynman advanced the following simple model, called the *parton model*: Assume that the proton is a loosely bound assemblage of a small number of constituents, called *partons*. These include quarks (and antiquarks), which are fermions carrying electric charge, and possibly other neutral species responsible for their binding. By assumption, these constituents are incapable of exchanging large momenta q^2 through the strong interactions. However, the quarks have the electromagnetic interactions of elementary fermions, so that an electron scattering from a quark can knock it out of the proton. The struck quark then exchanges momentum softly with the remainder of the proton, so that the pieces of the proton materialize as a jet of hadrons. The produced hadrons should be collinear with the direction of the original struck parton.

The parton model, incomplete though it is, imposes a strong constraint on the cross section for deep inelastic electron scattering. To derive this constraint, consider first the cross section for electron scattering from a single constituent quark. We discussed the related process of electron-muon scattering in Section 5.4, and we can borrow that result. Since we imagine the reaction to occur at very high energy, we will ignore all masses. The square of the invariant matrix element in the massless limit is written in a simple form in Eq. (5.71):

$$\frac{1}{4}\sum_{\text{spins}} |\mathcal{M}|^2 = \frac{8e^4 Q_i^2}{\hat{t}^2}\left(\frac{\hat{s}^2+\hat{u}^2}{4}\right), \tag{14.1}$$

where $\hat{s}$, $\hat{t}$, $\hat{u}$ are the Mandelstam variables for the electron-quark collision and Q_i is the electric charge of the quark in units of $|e|$. Recall from Eq. (5.73) that, for a collision involving massless particles, $\hat{s}+\hat{t}+\hat{u}=0$. Then the differential cross section in the center of mass system is

$$\begin{aligned}\frac{d\sigma}{d\cos\theta_{\text{CM}}} &= \frac{1}{2\hat{s}}\frac{1}{16\pi}\frac{8e^4Q_i^2}{\hat{t}^2}\left(\frac{\hat{s}^2+\hat{u}^2}{4}\right)\\ &= \frac{\pi\alpha^2 Q_i^2}{\hat{s}}\left(\frac{\hat{s}^2+(\hat{s}+\hat{t})^2}{\hat{t}^2}\right).\end{aligned} \tag{14.2}$$

Or, since $\hat{t} = -\hat{s}(1-\cos\theta_{\text{CM}})/2$,

$$\frac{d\sigma}{d\hat{t}} = \frac{2\pi\alpha^2 Q_i^2}{\hat{s}^2}\left(\frac{\hat{s}^2+(\hat{s}+\hat{t})^2}{\hat{t}^2}\right). \tag{14.3}$$

To make use of this result, we must relate the invariants $\hat{s}$ and $\hat{t}$ to experimental observables of electron-proton inelastic scattering. The kinematic variables are shown in Fig. 14.1. The momentum transfer q from the electron

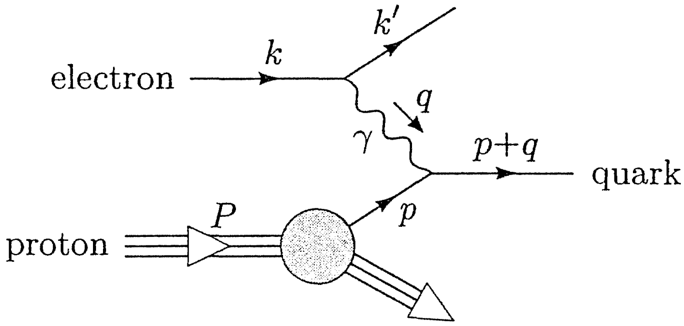


Figure 14.1. Kinematics of deep inelastic electron scattering in the parton model.

can be measured by measuring the final momentum and energy of the electron, without using any information from the hadronic products. Since q^μ is a spacelike vector, one conventionally expresses its invariant square in terms of a positive quantity Q, with

$$Q^2 \equiv -q^2. \tag{14.4}$$

Then the invariant $\hat{t}$ is simply $-Q^2$.

Expressing $\hat{s}$ in terms of measurable quantities is more difficult. If the collision is viewed from the electron-proton center of mass frame, and we visualize the proton as a loosely bound collection of partons (and continue to ignore masses), we can characterize a given parton by the fraction of the proton's total momentum that it carries. We denote this *longitudinal fraction* by the parameter ξ, with $0 < \xi < 1$. For each species i of parton, for example, up-type quarks with electric charge $Q_i = +2/3$, there will be a function $f_i(\xi)$ that expresses the probability that the proton contains a parton of type i and longitudinal fraction ξ. The expression for the total cross section for electron-proton inelastic scattering will contain an integral over the value of ξ for the struck parton. The momentum vector of the parton is then $p = \xi P$, where P is the total momentum of the proton. Thus, if k is the initial electron momentum,

$$\hat{s} = (p+k)^2 = 2p\cdot k = 2\xi P\cdot k = \xi s, \tag{14.5}$$

where s is the square of the electron-proton center of mass energy.

Remarkably, ξ can also be determined from measurements of only the electron momentum, if one makes the assumption that the electron-parton scattering is elastic. Since the scattered parton has a mass small compared to s and Q^2,

$$0 \approx (p+q)^2 = 2p\cdot q + q^2 = 2\xi P\cdot q - Q^2. \tag{14.6}$$

Thus

$$\xi = x, \qquad \text{where} \quad x \equiv \frac{Q^2}{2P\cdot q}. \tag{14.7}$$

From each scattered electron, one can determine the values of Q^2 and x for the scattering process. The parton model then predicts the event distribution

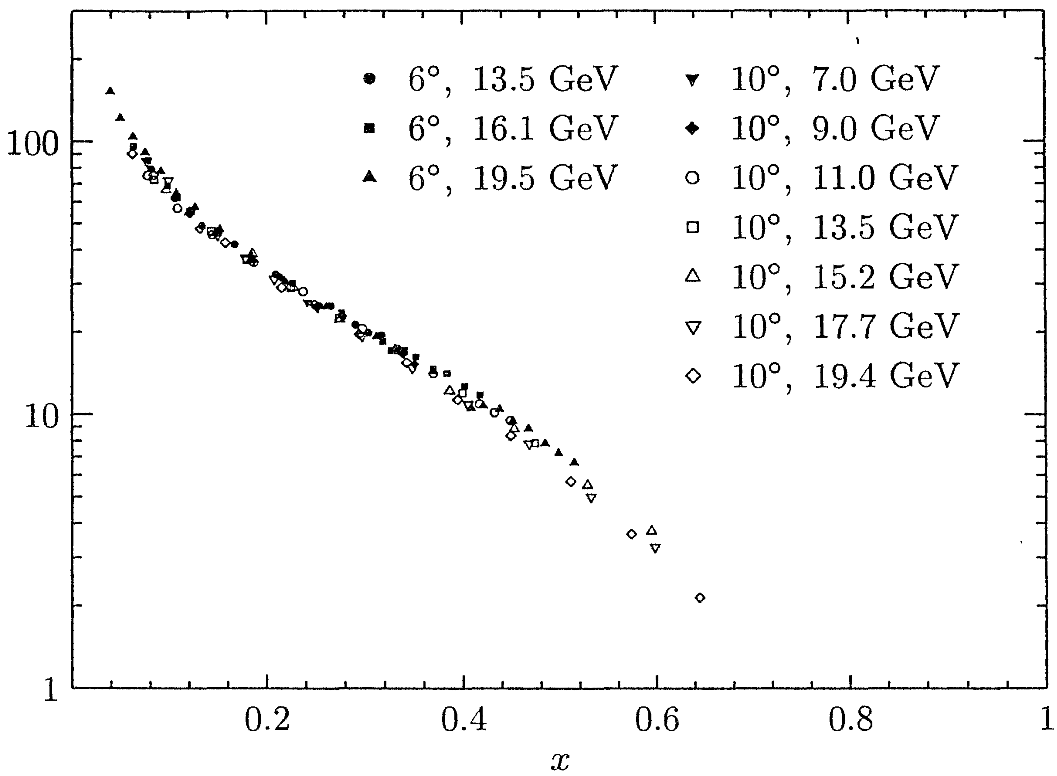


Figure 14.2. Test of Bjorken scaling using the e^-p deep inelastic scattering cross sections measured by the SLAC-MIT experiment, J. S. Poucher, et. al., *Phys. Rev. Lett.* **32**, 118 (1974). We plot $d^2\sigma/dxdQ^2$ divided by the factor (14.9) against x, for the various initial electron energies and scattering angles indicated. The data span the range 1 GeV^2 $< Q^2 <$ 8 GeV^2.

in the x-Q^2 plane. Using the parton distribution functions $f_i(\xi)$, evaluated at $\xi = x$, and the cross-section formula (14.3), we find the distribution

$$\frac{d^2\sigma}{dxdQ^2} = \sum_i f_i(x)Q_i^2 \cdot \frac{2\pi\alpha^2}{Q^4}\left[1 + \left(1 - \frac{Q^2}{xs}\right)^2\right]. \tag{14.8}$$

The distribution functions $f_i(x)$ depend on the details of the structure of the proton and it is not known how to compute them from first principles. But formula (14.8) still makes a striking prediction, that the deep inelastic scattering cross section, when divided by the factor

$$\frac{1 + (1 - Q^2/xs)^2}{Q^4} \tag{14.9}$$

to remove the kinematic dependence of the QED cross section, gives a quantity that depends only on x and is independent of Q^2. This behavior is known as *Bjorken scaling.* Indeed, the data from the SLAC-MIT experiment exhibited Bjorken scaling to about 10% accuracy for values of Q above 1 GeV, as shown in Fig. 14.2.

Bjorken scaling is, essentially, the statement that the structure of the proton looks the same to an electromagnetic probe no matter how hard the proton is struck. In the frame of the proton, the energy of the exchanged virtual photon is

$$q^0 = \frac{P \cdot q}{m} = \frac{Q^2}{2xm}, \tag{14.10}$$

where m is the proton mass. The reciprocal of this energy transfer is, roughly, the duration of the scattering process as seen by the components of the proton. This time should be compared to the reciprocal of the proton mass, which is the characteristic time over which the partons interact. The deep inelastic regime occurs when $q^0 \gg m$, that is, when the scattering is very rapid compared to the normal time scales of the proton. Bjorken scaling implies that, during such a rapid scattering process, interactions among the constituents of the proton can be ignored. We might imagine that the partons are approximately free particles over the very short times scales corresponding to energy transfers of a GeV or more, though they have strong interactions on longer time scales.

Asymptotically Free Partons

The picture of the proton structure implied by Bjorken scaling was beautifully simple, but it raised new, fundamental questions. In quantum field theory, fermions interact by exchanging virtual particles. These virtual particles can have arbitrarily high momenta, hence the fluctuations associated with them can occur on arbitrarily short time scales. Quantum field theory processes do not turn themselves off at short times to reveal free-particle equations. Thus the discovery of Bjorken scaling suggested a conflict between the observation of almost free partons and the basic principles of quantum field theory.

The resolution of this paradox came from the renormalization group. In Chapter 12 we saw that coupling constants vary with distance scale. In QED and ϕ^4 theory, we found that the couplings become strong at large momenta and weak at small momenta. However, we noted the possibility that, in some theories, the coupling constant could have the opposite behavior, becoming strong at small momenta or large times but weak at large momenta or short times. We referred to such behavior as *asymptotic freedom.* Section 13.3 discussed an example of an asymptotically free quantum field theory, the nonlinear sigma model in two dimensions. The problem posed in the previous paragraph would be resolved if there existed a suitable asymptotically free quantum field theory in four dimensions that could describe the interaction and binding of quarks. Then, at least to some level of approximation, the strong interaction described by this theory would turn off in large-momentum-transfer or short-time processes.

At the time of the discovery of Bjorken scaling, no asymptotically free field theories in four dimensions were known. Then, in the early 1970s, ‘t Hooft, Politzer, Gross, and Wilczek discovered a class of such theories. These are the *non-Abelian gauge theories*: theories of interacting vector bosons that can be constructed as generalizations of quantum electrodynamics. It was subsequently shown that these are the only asymptotically free field theories in four dimensions. This discovery gave the crucial clue for the construction of the fundamental theory of the strong interactions. Apparently, the quarks are bound together by interacting vector bosons (called *gluons*) of precisely this type.

However, these gauge theories cannot precisely reproduce the expectations of strict Bjorken scaling. The differences between the free parton model and the quantum field theory model with asymptotic freedom appear when one moves to a higher level of accuracy in measurements of deep inelastic scattering and other strong interaction processes involving large momentum transfer. In an asymptotically free quantum field theory, the coupling constant is still nonzero at any finite momentum transfer. In fact, the final evolution of the coupling to zero is very slow, logarithmic in momentum. Thus, at some level, one must find small corrections to Bjorken scaling, associated with the exchange or emission of high-momentum gluons. Similarly, the other qualitative simplifications of hadron physics at high momentum transfer—for example, the phenomenon of limited transverse momentum in hadron-hadron collisions—should be only approximate, receiving corrections due to gluon exchange and emission. Thus the predictions of an asymptotically free theory of the strong interaction are twofold. On one hand, such a theory predicts qualitative simplifications of behavior at high momentum. But, on the other hand, such a theory predicts a specific pattern of corrections to this behavior.

In fact, particle physics experiments of the 1970s revealed precisely this picture. Bjorken scaling was found to be only an approximate relation, showing violations that correspond to a slow evolution of the parton distributions $f_i(x)$ over a logarithmic scale in Q^2. The rate of particle production in hadron-hadron collisions was found to decrease only as a power rather than exponentially at very large values of the transverse momentum, and the particles produced at large transverse momentum were shown to be associated with jets of hadrons created by the soft evolution of a hard-scattered quark or gluon. Most remarkably, the forms of the cross sections found for these and other deviations from scaling did, finally, give direct evidence for the vector character of the elementary field that mediates the strong interaction.

We will review all of these phenomena in Chapter 17, as we study the particular gauge theory that describes the strong interactions. First, however, we must learn how to construct non-Abelian gauge theories and how to work out their predictions using Feynman diagrams. Throughout our analysis of these theories, the renormalization group will play an essential role. One of the very beautiful aspects of the study of non-Abelian gauge theories is the way in which the most powerful general ideas of quantum field theory acquire even more strength as they intertwine with the specific features of these particular, intricately built models. This interplay between general principles and the specific features of gauge theories will be the major theme of Part III of this book.

Chapter 15

Non-Abelian Gauge Invariance

So far in this book we have worked with a rather limited class of quantum fields and interactions, restricting our attention to scalar field theories, Yukawa theory, and Quantum Electrodynamics. It is hardly surprising that these theories are not sufficient to describe all of the known interactions of elementary particles. But what other theories are possible, given that the Lagrangian of a renormalizable theory can contain no terms of mass dimension higher than 4?

The most natural theories to try next would be ones with interactions among vector fields, of the form $A^\mu A^\nu \partial_\mu A_\nu$ or A^4. Sensible theories of this type are difficult to construct, however, because of the negative-norm states produced by the time component A^0 of the vector field operator. In Section 5.5 we saw that these negative-norm states cause no difficulty in QED: They are effectively canceled out by the longitudinal polarization states, by virtue of the Ward identity. The Ward identity, in turn, follows from the invariance of the QED Lagrangian under local gauge transformations. Perhaps, then, if we can generalize the principle of local gauge invariance, it will lead us to the construction of other sensible theories of vector particles.

The goal of this chapter is to do just that. First we will return briefly to the study of QED, this time taking the gauge symmetry to be fundamental and deriving the rest of the theory from this principle. Then, in Section 15.2, we will see that the gauge invariance of electrodynamics is only the most trivial example of an infinite-parameter symmetry, and that the more general examples lead to other interesting Lagrangians. These field theories, the first of which was constructed by Yang and Mills,* generalize electrodynamics in a profound way. They are theories of multiple vector particles, whose interactions are strongly constrained by the symmetry principle. In subsequent chapters we will study the quantization of these theories and their application to the real world of elementary particle physics.

*C. N. Yang and R. Mills, *Phys. Rev.* **96**, 191 (1954).

15.1 The Geometry of Gauge Invariance

In Section 4.1 we wrote down the Lagrangian of Quantum Electrodynamics and noted the curious fact that it is invariant under a very large group of transformations (4.6), allowing an independent symmetry transformation at every point in spacetime. This invariance is the famous *gauge symmetry* of QED. From the modern viewpoint, however, gauge symmetry is not an incidental curiosity, but rather the fundamental principle that determines the form of the Lagrangian. Let us now review the elements of the theory, taking the modern viewpoint.

We begin with the complex-valued Dirac field $\psi(x)$, and stipulate that our theory should be invariant under the transformation

$$\psi(x) \to e^{i\alpha(x)}\psi(x). \tag{15.1}$$

This is a phase rotation through an angle $\alpha(x)$ that varies arbitrarily from point to point. How can we write a Lagrangian that is invariant under this transformation? As long as we consider terms in the Lagrangian that have no derivatives, this is easy: We simply write the same terms that are invariant to global phase rotations. For example, the fermion mass term

$$m\bar{\psi}\psi(x)$$

is permitted by global phase invariance, and the local invariance gives no further restriction.

The difficulty arises when we try to write terms including derivatives. The derivative of $\psi(x)$ in the direction of the vector n^μ is defined by the limiting procedure

$$n^\mu \partial_\mu \psi = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\big[\psi(x+\epsilon n) - \psi(x)\big]. \tag{15.2}$$

However, in a theory with local phase invariance, this definition is not very sensible, since the two fields that are subtracted, $\psi(x+\epsilon n)$ and $\psi(x)$, have completely different transformations under the symmetry (15.1). The quantity $\partial_\mu\psi$, in other words, has no simple tranformation law and no useful geometric interpretation.

In order to subtract the values of $\psi(x)$ at neighboring points in a meaningful way, we must introduce a factor that compensates for the difference in phase transformations from one point to the next. The simplest way to do this is to define a scalar quantity $U(y,x)$ that depends on the two points and has the transformation law

$$U(y,x) \to e^{i\alpha(y)}U(y,x)e^{-i\alpha(x)} \tag{15.3}$$

simultaneously with (15.1). At zero separation, we set $U(y,y)=1$; in general, we can require $U(y,x)$ to be a pure phase: $U(y,z) = \exp[i\phi(y,x)]$. With this definition, the objects $\psi(y)$ and $U(y,x)\psi(x)$ have the same transformation law, and we can subtract them in a manner that is meaningful despite the local symmetry. Thus we can define a sensible derivative, called the *covariant derivative*, as follows:

$$n^\mu D_\mu \psi = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\left[\psi(x+\epsilon n) - U(x+\epsilon n, x)\psi(x)\right]. \tag{15.4}$$

To make this definition explicit, we need an expression for the comparator $U(y,x)$ at infinitesimally separated points. If the phase of $U(y,x)$ is a continuous function of the positions y and x, then $U(y,x)$ can be expanded in the separation of the two points:

$$U(x+\epsilon n, x) = 1 - ie\,\epsilon n^\mu A_\mu(x) + \mathcal{O}(\epsilon^2). \tag{15.5}$$

Here we have arbitrarily extracted a constant e. The coefficient of the displacement ϵn^μ is a new vector field $A_\mu(x)$. Such a field, which appears as the infinitesimal limit of a comparator of local symmetry transformations, is called a *connection*. The covariant derivative then takes the form

$$D_\mu \psi(x) = \partial_\mu \psi(x) + ieA_\mu \psi(x). \tag{15.6}$$

By inserting (15.5) into (15.3), one finds that A_μ transforms under this local gauge transformation as

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu \alpha(x). \tag{15.7}$$

To check that all of these expressions are consistent, we can transform $D_\mu \psi(x)$ according to Eqs. (15.1) and (15.7):

$$\begin{aligned} D_\mu \psi(x) \to &\left[\partial_\mu + ie\big(A_\mu - \frac{1}{e}\partial_\mu \alpha\big)\right] e^{i\alpha(x)}\psi(x) \\ &= e^{i\alpha(x)}\big(\partial_\mu + ieA_\mu\big)\psi(x) = e^{i\alpha(x)} D_\mu \psi(x). \end{aligned} \tag{15.8}$$

Thus the covariant derivative transforms in the same way as the field ψ, exactly as we constructed it to in the original definition (15.4).

We have now recovered most of the familiar ingredients of the QED Lagrangian. From our current viewpoint, however, the definition of the covariant derivative and the transformation law for the connection A_μ follow from the postulate of local phase rotation symmetry. Even the very existence of the vector field A_μ is a consequence of local symmetry: Without it we could not write an invariant Lagrangian involving derivatives of ψ.

More generally, our present analysis gives us a way to construct all possible Lagrangians that are invariant under the local symmetry. In any term with derivatives of ψ, replace these with covariant derivatives. According to Eq. (15.8), these transform in exactly the same manner as ψ itself. Therefore any combination of ψ and its covariant derivatives that is invariant under a global phase rotation (and only these combinations) will also be locally invariant.

To complete the construction of a locally invariant Lagrangian, we must find a kinetic energy term for the field A_μ: a locally invariant term that depends on A_μ and its derivatives, but not on ψ. This term can be constructed

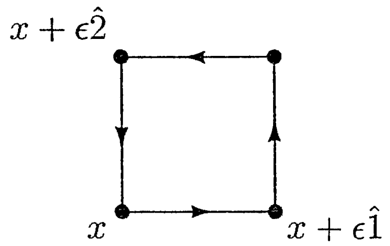


Figure 15.1. Construction of the field strength by comparisons around a small square in the $(1, 2)$ plane.

either integrally, from the comparator $U(y, x)$, or infinitesimally, from the covariant derivative.

Working from $U(y, x)$, we will need to extend our explicit formula (15.5) to the next term in the expansion in ϵ. Using the assumption that $U(y, x)$ is a pure phase and the restriction $(U(x, y))^\dagger = U(y, x)$, it follows that

$$U(x + \epsilon n, x) = \exp\bigl[-ie\epsilon n^\mu A_\mu(x + \tfrac{\epsilon}{2}n) + \mathcal{O}(\epsilon^3)\bigr]. \tag{15.9}$$

(Relaxing these restrictions introduces additional vector fields into the theory; this is an unnecessary complication.) Using this expansion for $U(y, x)$, we link together comparisons of the phase direction around a small square in spacetime. For definiteness, we take this square to lie in the $(1, 2)$-plane, as defined by the unit vectors $\hat{1}$, $\hat{2}$ (see Fig. 15.1). Define $\mathbf{U}(x)$ to be the product of the four comparisons around the corners of the loop:

$$\begin{aligned}\mathbf{U}(x) \equiv\ & U(x, x + \epsilon\hat{2})U(x + \epsilon\hat{2}, x + \epsilon\hat{1} + \epsilon\hat{2}) \\ & \times U(x + \epsilon\hat{1} + \epsilon\hat{2}, x + \epsilon\hat{1})U(x + \epsilon\hat{1}, x).\end{aligned} \tag{15.10}$$

The transformation law (15.3) for U implies that $\mathbf{U}(x)$ is locally invariant. In the limit $\epsilon \to 0$, it will therefore give us a locally invariant function of A_μ. To find the form of this function, insert the expansion (15.9) to obtain

$$\begin{aligned}\mathbf{U}(x) = \exp\Bigl\{ - i\epsilon e\bigl[&-A_2(x + \tfrac{\epsilon}{2}\hat{2}) - A_1(x + \tfrac{\epsilon}{2}\hat{1} + \epsilon\hat{2}) \\ &+ A_2(x + \epsilon\hat{1} + \tfrac{\epsilon}{2}\hat{2}) + A_1(x + \tfrac{\epsilon}{2}\hat{1})\bigr] + \mathcal{O}(\epsilon^3)\Bigr\}.\end{aligned} \tag{15.11}$$

When we expand the exponent in powers of ϵ, this reduces to

$$\mathbf{U}(x) = 1 - i\epsilon^2 e\bigl[\partial_1 A_2(x) - \partial_2 A_1(x)\bigr] + \mathcal{O}(\epsilon^3). \tag{15.12}$$

Therefore the structure

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{15.13}$$

is locally invariant. Of course, $F_{\mu\nu}$ is the familiar electromagnetic field tensor, and its invariance under (15.7) can be checked directly. The preceding construction, however, shows us the geometrical origin of the structure of $F_{\mu\nu}$. Any function that depends on A_μ only through $F_{\mu\nu}$ and its derivatives is locally invariant. More general functions, such as the vector field mass term

$A_\mu A^\mu$, transform under (15.7) in ways that cannot be compensated and thus cannot appear in an invariant Lagrangian.

A related argument for the invariance of $F_{\mu\nu}$ can be made using the covariant derivative. We have seen above that, if a field has the local transfomation law (15.1), then its covariant derivative has the same transformation law. Thus the second covariant derivative of ψ also transforms according to (15.1). The same conclusion holds for the commutator of covariant derivatives:

$$[D_\mu, D_\nu]\psi(x) \to e^{i\alpha(x)}[D_\mu, D_\nu]\psi(x). \tag{15.14}$$

However, the commutator is not itself a derivative at all:

$$\begin{aligned}[D_\mu, D_\nu]\psi &= [\partial_\mu, \partial_\nu]\psi + ie\big([\partial_\mu, A_\nu] - [\partial_\nu, A_\mu]\big)\psi - e^2[A_\mu, A_\nu]\psi \\ &= ie\big(\partial_\mu A_\nu - \partial_\nu A_\mu\big)\cdot\psi.\end{aligned} \tag{15.15}$$

That is,

$$[D_\mu, D_\nu] = ieF_{\mu\nu}. \tag{15.16}$$

On the right-hand side of (15.14), the factor $\psi(x)$ accounts for the entire transformation law, so the multiplicative factor $F_{\mu\nu}$ must be invariant. One can visualize the commutator of covariant derivatives as the comparison of comparisons across a small square; fundamentally, therefore, this argument is equivalent to that of the previous paragraph.

Whatever the method of proving the invariance of $F_{\mu\nu}$, we have now assembled all of the ingredients we need to write the most general locally invariant Lagrangian for the electron field ψ and its associated connection A_μ. This Lagrangian must be a function of ψ and its covariant derivatives, and of $F_{\mu\nu}$ and its derivatives, and must be invariant to global phase transformations. Up to operators of dimension 4, there are only four possible terms:

$$\mathcal{L}_4 = \bar\psi(i\not{D})\psi - \frac{1}{4}(F_{\mu\nu})^2 - ic\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu} - m\bar\psi\psi. \tag{15.17}$$

By adjusting the normalization of the fields ψ and A_μ, we have set the coefficients of the first two terms to their standard values. This normalization of A_μ requires the arbitrary scale factor e in our original definition (15.5) of A_μ. The third term violates the discrete symmetries P and T, so we may exclude it if we postulate these symmetries.† Then $\mathcal{L}_4$ contains only two free parameters, the scale factor e and the coefficient m.

By using operators of dimension 5 and 6, we can form many additional gauge-invariant combinations:

$$\mathcal{L}_6 = ic_1\bar\psi\sigma_{\mu\nu}F^{\mu\nu}\psi + c_2(\bar\psi\psi)^2 + c_3(\bar\psi\gamma^5\psi)^2 + \cdots. \tag{15.18}$$

More allowed terms appear at each higher order in mass dimension. But all of these terms are nonrenormalizable interactions. In the language of Section 12.1, they are irrelevant to physics in four dimensions in the limit where the cutoff is taken to infinity.

†The general systematics of P, C, and T violation are discussed in Section 20.3.

We have now reached a remarkable conclusion. We began by postulating that the electron field obeys the local symmetry (15.1). From this postulate, we showed that there must be an electromagnetic vector potential. Further, the symmetry principle implies that the most general Lagrangian in four dimensions that is renormalizable (or relevant, in Wilson's sense) is the general form $\mathcal{L}_4$. If we insist that this Lagrangian also be invariant under time reversal or parity, we are led uniquely to the Maxwell-Dirac Lagrangian that is the basis of quantum electrodynamics.

15.2 The Yang-Mills Lagrangian

If the simple geometrical constructions of the previous section yield Maxwell's theory of electrodynamics, then surely it must be possible to construct other interesting theories by starting with more general geometrical principles. Yang and Mills proposed that the argument of the previous section could be generalized from local phase rotation invariance to invariance under any continuous symmetry group. In this section, we will introduce this generalization of local symmetry. For most of the discussion, we will consider our local symmetry to be the three-dimensional rotation group, $O(3)$ or $SU(2)$, since in this case the necessary group theory should be familiar. At the end of the section, we will generalize further to the case of an arbitrary local symmetry.

Consider, then, the following generalization of the phase rotation (15.1): Instead of a single fermion field, we start with a doublet of Dirac fields,

$$\psi = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}, \tag{15.19}$$

which transform into one another under abstract three-dimensional rotations as a two-component spinor:

$$\psi \to \exp\Big(i\alpha^i \frac{\sigma^i}{2}\Big)\psi. \tag{15.20}$$

Here σ^i are the Pauli sigma matrices, and, as usual, a sum over repeated indices is implied. It is important to distinguish this abstract transformation from a rotation in physical three-dimensional space; in their original paper, Yang and Mills considered (ψ_1, ψ_2) to be the proton-neutron doublet as it is transformed under isotopic spin. As in the case of a phase rotation, it is not hard to construct Lagrangians for ψ that are invariant to (15.20) as a global symmetry.

We now promote (15.20) to a local symmetry, by insisting that the Lagrangian be invariant to this transformation for α^i an arbitrary function of x. Write this transformation as

$$\psi(x) \to V(x)\psi(x), \qquad \text{where } V(x) = \exp\Big(i\alpha^i(x)\frac{\sigma^i}{2}\Big). \tag{15.21}$$

We can construct a suitable Lagrangian by applying the methods of the previous section. However, we will encounter a number of additional complications, due to the fact that there are now three orthogonal symmetry motions, which do not commute with one another. This feature is sufficiently important to earn a special name for theories that have it: We refer to the *Abelian* symmetry group of electrodynamics, and the *non-Abelian* symmetry group of the more general theories. The field theory associated with a noncommuting local symmetry is termed a *non-Abelian gauge theory.*

To construct a Lagrangian that is invariant under this new group of transformations, we must again define a covariant derivative that transforms in a simple way. Again we use the definition (15.4), but since ψ is now a two-component object, the comparator $U(y,x)$ must be a 2×2 matrix. The transformation law for $U(y,x)$ is now

$$U(y,x) \to V(y)\,U(y,x)\,V^{\dagger}(x), \tag{15.22}$$

where $V(x)$ is as in (15.21), and again we set $U(y,y) = 1$. At points $x \neq y$ we can consistently restrict $U(y,x)$ to be a unitary matrix. Near $U = 1$, any such matrix can be expanded in terms of the Hermitian generators of $SU(2)$; thus for infinitesimal separation we can write

$$U(x+\epsilon n, x) = 1 + ig\epsilon n^{\mu} A^{i}_{\mu}\frac{\sigma^{i}}{2} + \mathcal{O}(\epsilon^{2}). \tag{15.23}$$

Here g is a constant, extracted for later convenience. Inserting this expansion into the definition (15.4) of the covariant derivative, we find the following expression for the covariant derivative associated with local $SU(2)$ symmetry:

$$D_{\mu} = \partial_{\mu} - igA^{i}_{\mu}\frac{\sigma^{i}}{2}. \tag{15.24}$$

This covariant derivative requires three vector fields, one for each generator of the transformation group.

We can find the gauge transformation law of the connection A^{i}_{μ} by inserting the expansion (15.23) into the transformation law (15.22):

$$1 + ig\epsilon n^{\mu} A^{i}_{\mu}\frac{\sigma^{i}}{2} \to V(x+\epsilon n)\Big(1 + ig\epsilon n^{\mu} A^{i}_{\mu}\frac{\sigma^{i}}{2}\Big)V^{\dagger}(x). \tag{15.25}$$

We must expand the right-hand side to order ϵ, taking care that the various Pauli matrices do not commute with one another. The expansion of $V(x+\epsilon n)$ is conveniently done using the identity

$$\begin{aligned} V(x+\epsilon n)V^{\dagger}(x) &= \Big[\Big(1 + \epsilon n^{\mu}\frac{\partial}{\partial x^{\mu}} + \mathcal{O}(\epsilon^{2})\Big)V(x)\Big]V^{\dagger}(x) \\ &= 1 + \epsilon n^{\mu}\Big(\frac{\partial}{\partial x^{\mu}}V(x)\Big)V^{\dagger}(x) + \mathcal{O}(\epsilon^{2}) \\ &= 1 + \epsilon n^{\mu} V(x)\Big(-\frac{\partial}{\partial x^{\mu}}V^{\dagger}(x)\Big) + \mathcal{O}(\epsilon^{2}). \end{aligned} \tag{15.26}$$

Then the terms proportional to ϵn^μ in (15.25) give the transformation

$$A^i_\mu(x)\frac{\sigma^i}{2} \to V(x)\Big(A^i_\mu(x)\frac{\sigma^i}{2} + \frac{i}{g}\partial_\mu\Big)V^\dagger(x). \tag{15.27}$$

The derivative acts on $V^\dagger(x) = \exp(-i\alpha^i\sigma^1/2)$; it is not so easy to compute this derivative explicitly because the exponent does not necessarily commute with its derivative. For infinitesimal transformations we can expand $V(x)$ to first order in α. In this case we obtain

$$A^i_\mu\frac{\sigma^i}{2} \to A^i_\mu\frac{\sigma^i}{2} + \frac{1}{g}(\partial_\mu\alpha^i)\frac{\sigma^i}{2} + i\big[\alpha^i\frac{\sigma^i}{2}, A^j_\mu\frac{\sigma^j}{2}\big] + \cdots. \tag{15.28}$$

The last term in this transformation law is new, and arises from the noncommutativity of the local transformations. By combining this relation with the infinitesimal form of the fermion transformation,

$$\psi \to \Big(1 + i\alpha^i\frac{\sigma^i}{2}\Big)\psi + \cdots, \tag{15.29}$$

we can check the infinitesimal transformation of the covariant derivative:

$$\begin{aligned} D_\mu\psi \to \Big(\partial_\mu - igA^i_\mu\frac{\sigma^i}{2} - i(\partial_\mu\alpha^i)\frac{\sigma^i}{2} + g\big[\alpha^i\frac{\sigma^i}{2}, A^j_\mu\frac{\sigma^j}{2}\big]\Big)\Big(1 + i\alpha^k\frac{\sigma^k}{2}\Big)\psi \\ = \Big(1 + i\alpha^i\frac{\sigma^i}{2}\Big)D_\mu\psi, \end{aligned} \tag{15.30}$$

up to terms of order α^2. It is not difficult to check using (15.27) and (15.21) that, even for finite transformations, the covariant derivative has the same transformation law as the field on which it acts.

Using the covariant derivative, we can build the most general gauge-invariant Lagrangians involving ψ. But to write a complete Lagrangian, we must also find gauge-invariant terms that depend only on A^i_μ. To do this, we construct the analogue of the electromagnetic field tensor. We will use the second method of the previous section, working from the commutator of covariant derivatives. The transformation law of the covariant derivative implies that

$$[D_\mu, D_\nu]\psi(x) \to V(x)[D_\mu, D_\nu]\psi(x). \tag{15.31}$$

At the same time, by writing out the commutator using formula (15.24), we can show, as in the Abelian case, that $[D_\mu, D_\nu]$ is not a differential operator but merely a multiplicative factor (now a matrix) acting on ψ. This time, however, there is a new feature: The last term in the expansion of the commutator no longer vanishes. Instead, we find

$$[D_\mu, D_\nu] = -igF^i_{\mu\nu}\frac{\sigma^i}{2}, \tag{15.32}$$

with

$$F^i_{\mu\nu}\frac{\sigma^i}{2} = \partial_\mu A^i_\nu\frac{\sigma^i}{2} - \partial_\nu A^i_\mu\frac{\sigma^i}{2} - ig\big[A^i_\mu\frac{\sigma^i}{2}, A^j_\nu\frac{\sigma^j}{2}\big]. \tag{15.33}$$

We can simplify this relation by applying the standard commutation relations of Pauli matrices:

$$\left[\frac{\sigma^i}{2}, \frac{\sigma^j}{2}\right] = i\epsilon^{ijk}\frac{\sigma^k}{2}. \tag{15.34}$$

Then

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\epsilon^{ijk} A^j_\mu A^k_\nu. \tag{15.35}$$

The transformation law for the field strength follows from Eqs. (15.21) and (15.31):

$$F^i_{\mu\nu}\frac{\sigma^i}{2} \to V(x) F^j_{\mu\nu}\frac{\sigma^j}{2} V^\dagger(x). \tag{15.36}$$

The infinitesimal form is

$$F^i_{\mu\nu}\frac{\sigma^i}{2} \to F^i_{\mu\nu}\frac{\sigma^i}{2} + \left[i\alpha^i\frac{\sigma^i}{2}, F^j_{\mu\nu}\frac{\sigma^j}{2}\right]. \tag{15.37}$$

Notice that the field strength is no longer a gauge-invariant quantity. It cannot be, since there are now three field strengths, each associated with a given direction of rotation in the abstract space. However, it is easy to form gauge-invariant combinations of the field strengths. For example,

$$\mathcal{L} = -\frac{1}{2}\,\mathrm{tr}\left[\left(F^i_{\mu\nu}\frac{\sigma^i}{2}\right)^2\right] = -\frac{1}{4}\left(F^i_{\mu\nu}\right)^2 \tag{15.38}$$

is a gauge-invariant kinetic energy term for A^i_μ. Notice that, in contrast to the case of electrodynamics, this Lagrangian contains cubic and quartic terms in A^i_μ. Thus, this Lagrangian describes a nontrivial, interacting field theory, called *Yang-Mills theory.* This is the simplest example of a non-Abelian gauge theory.

To construct a theory of Yang-Mills vector fields interacting with fermions, we simply add the gauge-field Lagrangian (15.38) to the familiar Dirac Lagrangian, with the ordinary derivative of ψ replaced by the covariant derivative. The result looks almost identical to the QED Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\not{D})\psi - \frac{1}{4}(F^i_{\mu\nu})^2 - m\bar{\psi}\psi. \tag{15.39}$$

This is the famous Yang-Mills Lagrangian. Like that of QED, it depends on two parameters: the scale factor g (which is analogous to the electron charge) and the fermion mass m. By varying this Lagrangian, we find the classical equations of motion of the gauge theory. These are the Dirac equation for the fermion field and the equation

$$\partial^\mu F^i_{\mu\nu} + g\epsilon^{ijk} A^{j\mu} F^k_{\mu\nu} = -g\bar{\psi}\gamma_\nu\frac{\sigma^i}{2}\psi \tag{15.40}$$

for the vector field.

Everything that we have done for the $SU(2)$ symmetry transformation (15.20) generalizes easily to any other continuous group of symmetries. The full range of possible symmetry groups is enumerated and classified in Section 15.4. For any such group, however, the general expressions for elements of the Lagrangian are quite similar. Consider any continuous group of transformations, represented by a set of $n \times n$ unitary matrices V. Then the basic fields $\psi(x)$ will form an n-plet, and transform according to

$$\psi(x) \to V(x)\psi(x), \tag{15.41}$$

where the x dependence of V makes the transformation local. In infinitesimal form, $V(x)$ can be expanded in terms of a set of basic generators of the symmetry group, which can be represented as Hermitian matrices t^a:

$$V(x) = 1 + i\alpha^a(x)t^a + \mathcal{O}(\alpha^2). \tag{15.42}$$

Now one can carry through the whole analysis from Eq. (15.22) to Eq. (15.33) for a general local symmetry group simply by replacing

$$\frac{\sigma^i}{2} \longrightarrow t^a \tag{15.43}$$

at each step of the analysis.

To generalize the explicit expression (15.35) for the field tensor, we need to know the commutation relations of the matrices t^a. It is conventional to write these in the standard form

$$[t^a, t^b] = if^{abc}t^c, \tag{15.44}$$

where f^{abc} is a set of numbers called *structure constants.* This object replaces ϵ^{ijk} in Eq. (15.34). It is conventional to choose a basis for the matrices t^a such that f^{abc} is completely antisymmetric; we will prove that this is always possible in Section 15.4.

We can now recapitulate all of our results as follows. The covariant derivative associated with the general transformation (15.41) is

$$D_\mu = \partial_\mu - igA^a_\mu t^a; \tag{15.45}$$

it contains one vector field for each independent generator of the local symmetry. The infinitesimal tranformation laws for ψ and A^a_μ are

$$\begin{aligned} \psi &\to (1 + i\alpha^a t^a)\psi; \\ A^a_\mu &\to A^a_\mu + \frac{1}{g}\partial_\mu\alpha^a + f^{abc}A^b_\mu\alpha^c. \end{aligned} \tag{15.46}$$

The finite transformation of A^a_μ has exactly the form of (15.27):

$$A^a_\mu(x)t^a \to V(x)\Big(A^a_\mu(x)t^a + \frac{i}{g}\partial_\mu\Big)V^\dagger(x). \tag{15.47}$$

These transformation laws imply that the covariant derivative of ψ has the same transformation law as ψ itself. The field tensor is defined by

$$[D_\mu, D_\nu] = -igF^a_{\mu\nu}t^a, \tag{15.48}$$

or more explicitly,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu. \tag{15.49}$$

This quantity has the infinitesimal transformation

$$F^a_{\mu\nu} \to F^a_{\mu\nu} - f^{abc} \alpha^b F^c_{\mu\nu}. \tag{15.50}$$

From Eqs. (15.46) and (15.50), one can show that any globally symmetric function of ψ, $F^a_{\mu\nu}$, and their covariant derivatives is also locally symmetric, and is therefore a candidate for a term in a gauge-invariant Lagrangian. However, there are very few permissible terms up to dimension 4. The most general gauge-invariant Lagrangian that is renormalizable and conserves P and T is again given by Eq. (15.39). The corresponding classical equation of motion is

$$\partial^\mu F^a_{\mu\nu} + g f^{abc} A^{b\mu} F^c_{\mu\nu} = -g j^a_\nu, \tag{15.51}$$

where

$$j^a_\nu = \bar{\psi} \gamma_\nu t^a \psi \tag{15.52}$$

is the global symmetry current of the fermion field.

Notice that the nonlinear terms in the Yang-Mills Lagrangian (15.39) appear in the covariant derivative, where they are proportional to t^a, and in the field tensor, where they are proportional to f^{abc}. Thus the form of the interactions in a non-Abelian gauge theory is dictated by the local symmetry. The nonlinear interactions of the vector field with itself are proportional to the commutators of symmetry generators and thus explicitly require the non-Abelian nature of the symmetry group.

15.3 The Gauge-Invariant Wilson Loop

In both of the previous sections we made use of the *comparator*, $U(y, x)$, which converts the fermion gauge transformation law at point x to that at point y. So far, in writing expressions for this object, it has sufficed to assume that x and y are infinitesimally separated. However, it is worthwhile to think further about the comparator in the case where x and y are far apart. This discussion will give us further insights into the geometry of gauge invariance, and will reveal some additional useful functions of the gauge field which we will put to work in Chapter 19.

We first return to the Abelian theory and expand upon our discussion of $U(y, x)$ in that context. In Eq. (15.10) we constructed a product of comparators on a path that wound around a small square. We showed that this product $\mathrm{U}(x)$ is not trivial, even though we eventually return to the starting point; rather, we found that $\mathrm{U}(x)$ differs from 1 by a term proportional to the electromagnetic field strength and to the area of the square. This is a particular case of a general conclusion: The comparator between two points x and y at finite separation depends on the path taken from x to y.

To explain this statement, it is useful to reverse some of the logic of Section 15.1. We begin from the connection A_μ, which we assume to have the transformation law (15.7), and construct $U(z, y)$ as a function of A_μ that transforms according to (15.3). It is not difficult to verify that the expression

$$U_P(z,y) = \exp\Big[-ie\int_P dx^\mu A_\mu(x)\Big] \tag{15.53}$$

meets this criterion if the integral is taken along any path P that runs from y to z. This object $U_P(z, y)$ is called the *Wilson line.*‡ Expression (15.53) gives an explicit realization of the abstract comparator $U(z, y)$ for points at finite separation.

A crucial property of the Wilson line is that it depends on the path P. If P is a closed path that returns to y, we obtain the *Wilson loop*,

$$U_P(y,y) = \exp\Big[-ie\oint_P dx^\mu A_\mu(x)\Big]. \tag{15.54}$$

This quantity is a nontrivial function of A_μ that is, by construction, locally gauge invariant. In fact, all gauge-invariant functions of A_μ can be thought of as combinations of Wilson loops for various choices of the path P. To motivate this claim, we use Stokes's theorem to rewrite the Wilson loop as

$$U_P(y,y) = \exp\Big[-i\frac{e}{2}\int_\Sigma d\sigma^{\mu\nu}F_{\mu\nu}\Big], \tag{15.55}$$

where Σ is a surface that spans the closed loop P, $d\sigma^{\mu\nu}$ is an area element on this surface, and $F_{\mu\nu}$ is the field tensor (15.13). This relation between the Wilson loop and the field strength is illustrated in Fig. 15.2. Since the Wilson loop is gauge invariant, this argument gives one more way to visualize the gauge invariance of the field strength. Conversely, since (almost) all gauge-invariant functions of A_μ can be built up from $F_{\mu\nu}$, this expression gives weight to the statement that $U_P(y, y)$ is the most general gauge invariant.

Both the Wilson line and the Wilson loop can be generalized to the non-Abelian case. Here, however, additional subtleties arise when we consider exponentials of noncommuting matrices. Let us first construct the Wilson line, which now transforms according to Eq. (15.22). It is not correct to make a straightforward rewriting of (15.53) with the integral of $A^a_\mu t^a$ in the exponent, since these matrices do not necessarily commute at different points. Instead, we must order these matrices in a particular way. We will now give the correct ordering prescription and then prove its transformation law.

Let s be a parameter of the path P, running from 0 at $x = y$ to 1 at $x = z$. Then define the Wilson line as the power-series expansion of the exponential, with the matrices in each term ordered so that higher values of s stand to the

‡This path-dependent phase was used long before Wilson's work, in Schwinger's early papers on QED, and in Y. Aharonov and D. Böhm, *Phys. Rev.* **115**, 485 (1959).

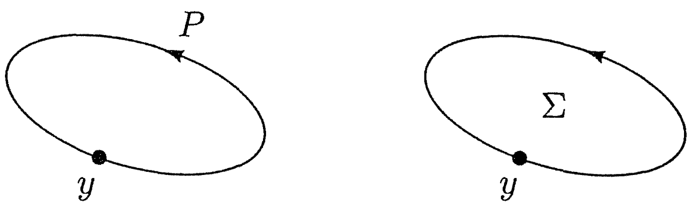


Figure 15.2. The Wilson loop integral is taken around an arbitrary loop. It can also be expressed as a flux integral of the field strength over a surface spanning the loop.

left. This prescription is called *path-ordering* and is denoted by the symbol $P\{\}$. Thus the Wilson line is written

$$U_P(z,y) = P\left\{\exp\left[ig\int_0^1 ds\frac{dx^\mu}{ds}A_\mu^a(x(s))t^a\right]\right\}. \tag{15.56}$$

This expression is similar to the time-ordered exponential that we wrote for the interaction-picture propagator in Eq. (4.23). Pursuing this analogy, one can show that this expression for U_P is the solution of a differential equation similar to (4.24):

$$\frac{d}{ds}U_P(x(s),y) = \left(ig\frac{dx^\mu}{ds}A_\mu^a(x(s))t^a\right)U_P(x(s),y). \tag{15.57}$$

(Here we consider U_P to be a continuous function of the parameter s, rather than fixing $s=1$ at the endpoint.)

To show that expression (15.56) is the correct generalization of the Wilson line, we must show that it satisfies the correct gauge transformation law (15.22). This follows from the differential equation (15.57), which can be rewritten as

$$\frac{dx^\mu}{ds}\,D_\mu U_P(x,y) = 0. \tag{15.58}$$

Now let A^V represent the gauge transform of a field configuration A, and use these arguments to denote explicitly the dependence of gauge functions on the gauge field. We would like to show that

$$U_P(z,y,A^V) = V(z)U_P(z,y,A)V^\dagger(y), \tag{15.59}$$

which is equivalent to (15.22). In (15.30) we proved, in its infinitesimal version, the relation

$$D_\mu(A^V)\,V(x) = V(x)\,D_\mu(A). \tag{15.60}$$

This relation implies that the right-hand side of (15.59) satisfies (15.58) for the gauge field A^V if $U_P(z,y,A)$ satisfies this equation for the gauge field A. But the solution of a first-order differential equation with a fixed boundary condition is unique. Thus, if $U_P(z,y)$ is defined to be the solution of (15.57) or (15.58), it indeed has the transformation law (15.59).

The Wilson line associated with a closed path returning to y transforms only with the gauge parameter at y; however, it is not a gauge invariant:

$$U_P(y,y) \to V(y)U_P(y,y)V^\dagger(y). \tag{15.61}$$

To understand this transformation better, one can work out the expression for $U_P(x,x)$, where the path is the small square in the $(1,2)$ plane shown in Fig. 15.1. In addition to the terms in Eq. (15.11), there are additional corrections of order ϵ^2 coming from products of $(A^a_\mu t^a)$ factors from pairs of sides, which sum up to a commutator of these factors. One finds

$$U_P(x,x) = 1 + ig\epsilon^2 F^a_{12}(x)t^a + \mathcal{O}(\epsilon^3), \tag{15.62}$$

where $F^a_{\mu\nu}$ is given by the full expression in (15.49). If we then expand the transformation law (15.61) in powers of ϵ, the term of order ϵ^2 is the transformation law of $F^a_{\mu\nu}$ given in Eq. (15.36).

To convert the Wilson line for a closed path into a true gauge invariant, take the trace. By cyclic invariance, (15.61) implies

$$\operatorname{tr} U_P(x,x) \to \operatorname{tr} U_P(x,x). \tag{15.63}$$

Thus for a non-Abelian gauge theory, we define the Wilson loop to be the trace of the Wilson line around a closed path.

Let us evaluate $\operatorname{tr} U_P(x,x)$ more explicitly for the case of an $SU(2)$ gauge group. If $U(\epsilon)$ is any 2×2 unitary matrix that tends to 1 as $\epsilon \to 0$, we can expand it in ϵ as follows:

$$\begin{aligned} U(\epsilon) &= \exp\left[i(\epsilon\beta^i + \epsilon^2\gamma^i + \cdots)\frac{\sigma^i}{2}\right] \\ &= 1 + i(\epsilon\beta^i + \epsilon^2\gamma^i + \cdots)\frac{\sigma^i}{2} - \frac{1}{2}(\epsilon\beta^i\cdot\epsilon\beta^j + \cdots)\frac{\sigma^i}{2}\frac{\sigma^j}{2} + \cdots. \end{aligned} \tag{15.64}$$

Then, since the Pauli matrices are traceless and satisfy $\operatorname{tr}[\sigma^i\sigma^j] = 2\delta^{ij}$,

$$\operatorname{tr} U(\epsilon) = 2 - \frac{1}{4}\epsilon^2(\beta^i)^2 + \mathcal{O}(\epsilon^3). \tag{15.65}$$

Applying this formula to Eq. (15.62), we find

$$\operatorname{tr} U_P(x,x) = 2 - \frac{1}{4}g^2\epsilon^4(F^i_{12})^2 + \mathcal{O}(\epsilon^5). \tag{15.66}$$

Thus the gauge invariance of $(F^i_{\mu\nu})^2$ can be derived from a geometrical argument, just as in the Abelian case. Using the notation that will be introduced in the next section, one can show that the same argument goes through for any gauge group.

15.4 Basic Facts about Lie Algebras

At the end of Section 15.2 we saw that the class of non-Abelian gauge theories is very large. To work with these theories most efficiently, it is worthwhile to pause and consider the general properties of the continuous groups on which they are based. In this section we will enumerate all the possible groups that can be used to construct non-Abelian gauge theories. We will then compute some numerical factors, built out of group transformation matrices, that are needed in performing explicit calculations in quantized gauge theories.*

To a mathematician, a group is made up of abstract entities that obey certain algebraic rules. In quantum mechanics, however, we are interested specifically in groups of unitary operators that act on the vector space of quantum states. We focus our attention on continuously generated groups, that is, groups that contain elements arbitrarily close to the identity, such that the general element can be reached by the repeated action of these infinitesimal elements. Then any infinitesimal group element g can be written

$$g(\alpha) = 1 + i\alpha^a T^a + \mathcal{O}(\alpha^2). \tag{15.67}$$

The coefficients of the infinitesimal group parameters α^a are Hermitian operators T^a, called the *generators* of the symmetry group. A continuous group with this structure is called a *Lie group.*

The set of generators T^a must span the space of infinitesimal group transformations, so the commutator of generators T^a must be a linear combination of generators. Thus the commutation relations of the operators T^a can be written

$$[T^a, T^b] = if^{abc}T^c; \tag{15.68}$$

the numbers f^{abc} are called *structure constants.* The vector space spanned by the generators, with the additional operation of commutation, is called a *Lie Algebra.*

The commutation relations (15.68) and the identity

$$[T^a, [T^b, T^c]] + [T^b, [T^c, T^a]] + [T^c, [T^a, T^b]] = 0 \tag{15.69}$$

imply that the structure constants obey

$$f^{ade}f^{bcd} + f^{bde}f^{cad} + f^{cde}f^{abd} = 0, \tag{15.70}$$

called the *Jacobi identity.* From the mathematician's viewpoint (considering the generators to be abstract entities rather than Hermitian operators), the

*In this section we will state, without proof, some general results from the theory of continuous groups. There are many excellent books that review these mathematical results systematically. Among these, we recommend especially Cahn (1984), for a brief but incisive discussion, and S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic Press, 1978), which gives for an elegant and rigorous account. R. Slansky, *Phys. Repts.* **79**, 1 (1981), has compiled an especially useful set of tables of group-theoretic identities relevant to the construction of non-Abelian gauge theories.

Jacobi identity is an axiom that must be satisfied in order for a given set of commutation rules to define a Lie algebra.

The commutation relations of the Lie algebra completely determine the group multiplication law of an associated Lie group sufficiently close to the identity. For large enough transformations, additional global questions come into play; to give a familiar example, $SU(2)$ and $O(3)$ have the same commutation relations but different global structure. However, the Lagrangian of a non-Abelian gauge theory depends only on the Lie algebra of the local symmetry group, so we will ignore these global questions from here on.

Classification of Lie Algebras

For the application to gauge theories, the local symmetry is normally a unitary transformation of a set of fields. Thus we are primarily interested in Lie algebras that have finite-dimensional Hermitian representations, leading to finite-dimensional unitary representations of the corresponding Lie group. We will also assume that the number of generators is finite. Such Lie algebras are called *compact*, because these conditions imply that the Lie group is a finite-dimensional compact manifold.

If one of the generators T^a commutes with all of the others, it generates an independent continuous Abelian group. Such a group, which has the structure of the group of phase rotations

$$\psi \to e^{i\alpha}\psi, \tag{15.71}$$

we call $U(1)$. If the algebra contains no such commuting elements, so that the group contains no $U(1)$ factors, then we call the algebra *semi-simple.* If, in addition, the Lie algebra cannot be divided into two mutually commuting sets of generators, the algebra is *simple.* A general Lie algebra is the direct sum of non-Abelian simple components and additional Abelian generators.

Surprisingly, the basic conditions that a Lie algebra be compact and simple turn out to be extremely restrictive. In one of the triumphs of nineteenth-century mathematics, Killing and Cartan classified all possible compact simple Lie algebras. Almost all of these algebras belong to one of three infinite families, with only five exceptions. The three infinite families are the algebras corresponding to the so-called *classical groups*, whose structures are conveniently defined in terms of particular matrix representations. The definitions of the three families of classical groups are as follows:

1. *Unitary transformations of N-dimensional vectors.* Let ξ and η be complex N-vectors. A general linear transformation then has the form

$$\eta_a \to U_{ab}\eta_b, \qquad \xi_a \to U_{ab}\xi_b. \tag{15.72}$$

We say that this transformation is *unitary* if it preserves the inner product $\eta_a^*\xi_a$. The pure phase transformations

$$\xi_a \to e^{i\alpha}\xi_a \tag{15.73}$$

form a $U(1)$ subgroup which commutes with all other unitary transformations; we remove this subgroup to form a simple Lie group, called $SU(N)$; it consists of all $N \times N$ unitary transformations satisfying $\det(U) = 1$. The generators of $SU(N)$ are represented by $N \times N$ Hermitian matrices t^a, subject to the condition that they be orthogonal to the generator of (15.73):

$$\mathrm{tr}[t^a] = 0. \tag{15.74}$$

There are $N^2 - 1$ independent matrices satisfying these conditions.

2. *Orthogonal transformations of N-dimensional vectors.* This is the subgroup of unitary $N \times N$ transformations that preserves the symmetric inner product

$$\eta_a E_{ab} \xi_b, \qquad \text{with } E_{ab} = \delta_{ab}. \tag{15.75}$$

This is the usual vector product, and so this group is the rotation group in N dimensions, $SO(N)$. (Adding the reflection gives the group $O(N)$.) There is an independent rotation corresponding to each plane in N dimensions, so $SO(N)$ has $N(N-1)/2$ generators.

3. *Symplectic transformations of N-dimensional vectors.* This is the subgroup of unitary $N \times N$ transformations, for N even, that preserves the antisymmetric inner product

$$\eta_a E_{ab} \xi_b, \qquad \text{with } E_{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{15.76}$$

where the elements of the matrix are $N/2 \times N/2$ blocks. This group is called $Sp(N)$; it has $N(N+1)/2$ generators.

Beyond these three families, there are five more *exceptional* Lie algebras, denoted in Cartan's classification system as G_2, F_4, E_6, E_7, and E_8. Of these, E_6 and E_8 have been applied as local symmetry groups in interesting unified models of the fundamental interactions. However, we will not consider these exceptional groups further in this book. In fact, most of our examples will involve only $SU(N)$ groups.

Representations

Once we have specified the local symmetry group, the fields that appear in the Lagrangian most naturally transform according to a finite-dimensional unitary representation of this group. Thus we might next ask how to systematically find all such representations of any given Lie group. Recall that for the group $SU(2)$, the representations can be constructed directly from the commutation relations, using the raising and lowering operators J_+ and J_-. This construction can be generalized to find the finite-dimensional representations of any compact Lie algebra. In this book, however, we will work with relatively simple representations whose structure we can work out by less formal methods.

Before discussing representations of Lie algebras, we should review some general aspects of group representations. Given a symmetry group G, a finite-dimensional unitary representation of the group's Lie algebra is a set of $d \times d$ Hermitian matrices t^a that satisfy the commutation relations (15.68). The size d is the *dimension* of the representation. An arbitrary representation can generally be decomposed by finding a basis in which all representation matrices are simultaneously block-diagonal. Through this change of basis, we can write the representation as the direct sum of *irreducible* representations. We denote the representation matrices in the irreducible representation r by t^a_r.

It is standard practice to adopt a normalization convention for the matrices t^a_r, based on traces of their products. If the Lie algebra is semi-simple, the matrices t^a_r themselves are traceless. Consider, however, the trace of the product of two generator matrices:

$$\operatorname{tr}[t^a_r t^b_r] \equiv D^{ab}. \tag{15.77}$$

As long as the generator matrices are Hermitian, the matrix D^{ab} is positive definite. Let us choose a basis for the generators T^a so that this matrix is proportional to the identity. It can be shown that, once this is done for one irreducible representation, it is true for all irreducible representations. Thus, in this basis,

$$\operatorname{tr}[t^a_r t^b_r] = C(r)\delta^{ab}, \tag{15.78}$$

where $C(r)$ is a constant for each representation r. Equation (15.78) and the commutation relations (15.68) yield the following representation of the structure constants:

$$f^{abc} = -\frac{i}{C(r)} \operatorname{tr}\left\{[t^a_r, t^b_r]t^c_r\right\}. \tag{15.79}$$

This equation implies that f^{abc} is totally antisymmetric.

For each irreducible representation r of G, there is an associated *conjugate* representation $\bar{r}$. The representation r yields the infinitesimal transformation

$$\phi \to (1 + i\alpha^a t^a_r)\phi. \tag{15.80}$$

The complex conjugate of this transformation,

$$\phi^* \to (1 - i\alpha^a (t^a_r)^*)\phi^*, \tag{15.81}$$

must also be the infinitesimal element of a representation of G. Thus the conjugate representation to r has representation matrices

$$t^a_{\bar{r}} = -(t^a_r)^* = -(t^a_r)^T. \tag{15.82}$$

Since $\phi^*\phi$ is invariant to unitary transformations, it is possible to combine fields transforming in the representations r and $\bar{r}$ to form a group invariant.

It is possible that the representation $\bar{r}$ may be equivalent to r, if there is a unitary transformation U such that $t^a_{\bar{r}} = U t^a_r U^\dagger$. If so, the representation r is *real*. In this case, there is a matrix G_{ab} such that, if η and ξ belong to the representation r, the combination $G_{ab}\eta_a\xi_b$ is an invariant. It is sometimes useful to distinguish the case in which G_{ab} is symmetric from that in which G_{ab} is antisymmetric. In the former case the representation is *strictly real*; in the latter case it is *pseudoreal.* Both cases occur already in $SU(2)$: The invariant combination of two vectors is $v_a w_a$, so the vector is a real representation; the invariant combination of two spinors is $\epsilon^{\alpha\beta}\eta_\alpha\xi_\beta$, so the spinor is a pseudoreal representation.

With this language we can discuss the simplest representations of the classical groups. In $SU(N)$, the basic irreducible representation (often called the *fundamental* representation) is the N-dimensional complex vector. For $N > 2$ this representation is complex, so that there is a second, inequivalent, representation $\overline{N}$. (In $SU(2)$ this representation is the pseudoreal spinor representation.) In $SO(N)$, the basic N-dimensional vector is a (strictly) real representation. In $Sp(N)$, the N-dimensional vector is a pseudoreal representation.

Another irreducible representation, present for any simple Lie algebra, is the one to which the generators of the algebra belong. This representation is called the *adjoint representation* and denoted by $r = G$. The representation matrices are given by the structure constants:

$$(t^b_G)_{ac} = if^{abc}. \tag{15.83}$$

With this definition, the statement that t^a_G satisfies the Lie algebra

$$\left([t^b_G, t^c_G]\right)_{ae} = if^{bcd}(t^d_G)_{ae} \tag{15.84}$$

is just a rewriting of the Jacobi identity (15.70). Since the structure constants are real and antisymmetric, $t^a_G = -(t^a_G)^*$; thus the adjoint representation is always a real representation. From the descriptions of the Lie groups given above, the dimension of the adjoint representation $d(G)$ is given, for the classical groups, by

$$d(G) = \begin{cases} N^2 - 1 & \text{for } SU(N), \\ N(N-1)/2 & \text{for } SO(N), \\ N(N+1)/2 & \text{for } Sp(N). \end{cases} \tag{15.85}$$

The identification of f^{abc} as a representation matrix allows us to gain further insight into some of the quantities introduced in Section 15.2. The covariant derivative acting on a field in the adjoint representation is

$$\begin{aligned} (D_\mu\phi)_a &= \partial_\mu\phi_a - igA^b_\mu(t^b_G)_{ac}\phi_c \\ &= \partial_\mu\phi_a + gf^{abc}A^b_\mu\phi_c. \end{aligned} \tag{15.86}$$

Thus we can recognize the infinitesimal form of the gauge transformation of the vector field in (15.46) as the motion

$$A^a_\mu \to A^a_\mu + \frac{1}{g}(D_\mu\alpha)^a. \tag{15.87}$$

The gauge field equation of motion (15.51) can be rewritten as

$$(D^\mu F_{\mu\nu})^a = -gj^a_\nu. \tag{15.88}$$

In both of these expressions, the arbitrary-looking terms involving f^{abc} arise naturally as part of a covariant derivative. An additional identity follows from considering the antisymmetric double commutator of covariant derivatives,

$$\epsilon^{\mu\nu\lambda\sigma}[D_\nu,[D_\lambda,D_\sigma]].$$

This quantity vanishes by its total antisymmetry, in the same way as (15.69). This result can be reduced to the identity

$$\epsilon^{\mu\nu\lambda\sigma}(D_\nu F_{\lambda\sigma})^a = 0. \tag{15.89}$$

This equation, called the *Bianchi identity* of a non-Abelian gauge theory, is the analogue of the homogeneous Maxwell equations in electrodynamics.

The Casimir Operator

In $SU(2)$, we characterize representations by the eigenvalue of the total spin J^2. In fact, for any simple Lie algebra, the operator

$$T^2 = T^a T^a \tag{15.90}$$

(with the repeated index summed, as always) commutes with all group generators:

$$\begin{aligned}[][T^b, T^aT^a] &= (if^{bac}T^c)T^a + T^a(if^{bac}T^c)\\ &= if^{bac}\{T^c, T^a\},\end{aligned} \tag{15.91}$$

which vanishes by the antisymmetry of f^{abc}. In other words, T^2 is an invariant of the algebra; this implies that T^2 takes a constant value on each irreducible representation. Thus, the matrix representation of T^2 is proportional to the unit matrix:

$$t^a_r t^a_r = C_2(r)\cdot \mathbf{1}, \tag{15.92}$$

where $\mathbf{1}$ is the $d(r)\times d(r)$ unit matrix and $C_2(r)$ is a constant, called the *quadratic Casimir operator*, for each representation. For the adjoint representation, Eq. (15.92) is more conveniently written as

$$f^{acd}f^{bcd} = C_2(G)\delta^{ab}. \tag{15.93}$$

Casimir operators appear very often in computations in non-Abelian gauge theories. Furthermore, the related invariant $C(r)$ given by (15.78) is simply related to the Casimir operator: If we contract (15.78) with δ^{ab} and evaluate the left-hand side using (15.92), we find

$$d(r)C_2(r) = d(G)C(r). \tag{15.94}$$

Thus it will be useful for us to compute $C_2(r)$ for the simplest $SU(N)$ representations.

For $SU(2)$, the fundamental two-dimensional representation is the spinor representation, which is given in terms of Pauli matrices by

$$t_2^a = \frac{\sigma^a}{2}. \tag{15.95}$$

These satisfy $\mathrm{tr}[t_2^a t_2^b] = \frac{1}{2}\delta^{ab}$. We will choose the generators of $SU(N)$ so that three of these are the generators (15.95), acting on the first two components of the N-vector ξ. Then, for any matrices of the fundamental representation,

$$\mathrm{tr}[t_N^a t_N^b] = \tfrac{1}{2}\delta^{ab}. \tag{15.96}$$

This convention fixes the values of $C(r)$ and $C_2(r)$ for all of the irreducible representations of $SU(N)$. For the fundamental representations N and $\overline{N}$, $C(N)$ is given directly by (15.96), and $C_2(N)$ follows from (15.94). We find

$$C(N) = \frac{1}{2}, \qquad C_2(N) = \frac{N^2-1}{2N}. \tag{15.97}$$

To compute the Casimir operator for the adjoint representation, we build up this representation from the product of the N and $\overline{N}$. Let us first discuss the product of irreducible representations more generally. The direct product of two representations r_1, r_2 is a representation of dimension $d(r_1)\cdot d(r_2)$. An object that transforms according to this representation can be written as a tensor Ξ_{pq}, in which the first index transforms according to r_1, the second according to r_2. In general, such a product can be decomposed into a direct sum of irreducible representations; symbolically, we write

$$r_1 \times r_2 = \sum r_i. \tag{15.98}$$

The representation matrices in the representation $r_1 \times r_2$ are

$$t^a_{r_1\times r_2} = t^a_{r_1} \otimes 1 + 1 \otimes t^a_{r_2}, \tag{15.99}$$

where the first matrix of each product acts on the first index of Ξ_{pq} and the second matrix acts on the second index.

The Casimir operator in the product representation is

$$(t^a_{r_1\times r_2})^2 = (t^a_{r_1})^2 \otimes 1 + 2t^a_{r_1} \otimes t^a_{r_2} + 1 \otimes (t^a_{r_2})^2.$$

Take the trace; since the matrices t^a_r are traceless, the trace of the second term on the right is zero. Then

$$\mathrm{tr}(t^a_{r_1\times r_2})^2 = \big(C_2(r_1) + C_2(r_2)\big)d(r_1)d(r_2). \tag{15.100}$$

On the other hand, the decomposition (15.98) implies

$$\mathrm{tr}(t^a_{r_1\times r_2})^2 = \sum C_2(r_i)d(r_i). \tag{15.101}$$

Equating (15.100) and (15.101), we find a useful identity for $C_2(r)$.

Now apply this identity to the product of the N and $\overline{N}$ representations of $SU(N)$. In this case, the tensor Ξ_{pq} can contain a term proportional to the invariant δ_{pq}. The remaining (N^2-1) independent components of Ξ_{pq}

transform as a general traceless $N \times N$ tensor; the matrices that effect these transformations make up the adjoint representation of $SU(N)$. In this case Eq. (15.98) becomes explicitly

$$N \times \overline{N} = 1 + (N^2 - 1). \tag{15.102}$$

For this decomposition, Eqs. (15.100) and (15.101) imply the identity

$$\left(2 \cdot \frac{N^2 - 1}{2N}\right) N^2 = 0 + C_2(G) \cdot (N^2 - 1). \tag{15.103}$$

Thus, for $SU(N)$,

$$C_2(G) = C(G) = N. \tag{15.104}$$

Some additional examples of the computation of quadratic Casimir operators are given in Problem 15.5. However, the examples we have discussed in this section, combined with the basic group-theoretic concepts that we have reviewed, already provide enough material to carry out the most important computations of physical interest in non-Abelian gauge theories.

Problems

15.1 Brute-force computations in $SU(3)$. The standard basis for the fundamental representation of $SU(3)$ is

$$t^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad t^2 = \frac{1}{2}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad t^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$t^4 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad t^5 = \frac{1}{2}\begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$

$$t^6 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad t^7 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad t^8 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

(a) Explain why there are exactly eight matrices in the basis.

(b) Evaluate all the commutators of these matrices, to determine the structure constants of $SU(3)$. Show that, with the normalizations used here, f^{abc} is totally antisymmetric. (This exercise is tedious; you may wish to check only a representative sample of the commutators.)

(c) Check the orthogonality condition (15.78), and evaluate the constant $C(r)$ for this representation.

(d) Compute the quadratic Casimir operator $C_2(r)$ directly from its definition (15.92), and verify the relation (15.94) between $C_2(r)$ and $C(r)$.

15.2 Write down the basis matrices of the adjoint representation of $SU(2)$. Compute $C(G)$ and $C_2(G)$ directly from their definitions (15.78) and (15.92).

15.3 Coulomb potential.

(a) Using functional integration, compute the expectation value of the Wilson loop in pure quantum electrodynamics without fermions. Show that

$$\langle U_P(z,z)\rangle = \exp\left[-ie^2 \oint_P dx^\mu \oint_P dy^\nu g_{\mu\nu} \frac{1}{2\pi^2(x-y)^2}\right],$$

with x and y integrated around the closed curve P.

(b) Consider the Wilson loop of a rectangular path of (spacelike) width R and (timelike) length T, $T \gg R$. Compute the expectation value of the Wilson loop in this limit and compare to the general expression for time evolution,

$$\langle U_P\rangle = \exp[-iE(R)T],$$

where $E(R)$ is the energy of the electromagnetic sources corresponding to the Wilson loop. Show that the potential energy of these sources is just the Coulomb potential, $V(R) = -e^2/4\pi R$.

(c) Assuming that the propagator of the non-Abelian gauge field is given by the Feynman gauge expression

$$\left\langle A^a_\mu(x) A^b_\mu(y)\right\rangle = \int \frac{d^4p}{(2\pi)^4} \frac{-ig_{\mu\nu}\delta^{ab}}{p^2} e^{-ip\cdot(x-y)},$$

compute the expectation value of a non-Abelian Wilson loop to order g^2. The result will depend on the representation r of the gauge group in which one chooses the matrices that appear in the exponential. Show that, to this order, the Coulomb potential of the non-Abelian gauge theory is $V(R) = -g^2 C_2(r)/4\pi R$.

15.4 Scalar propagator in a gauge theory. Consider the equation for the Green's function of the Klein-Gordon equation:

$$(\partial^2 + m^2) D_F(x,y) = -i\delta^{(4)}(x-y).$$

We can find an interesting representation for this Green's function by writing

$$D_F(x,y) = \int_0^\infty dT\, D(x,y,T),$$

where $D(x,y,T)$ satisfies the Schrödinger equation

$$\left[i\frac{\partial}{\partial T} - (\partial^2 + m^2)\right] D(x,y,T) = i\delta(T)\delta^{(4)}(x-y).$$

Now, represent $D(x,y,T)$ using the functional integral solution of the Schrödinger equation presented in Section 9.1.

(a) Using the explicit formula of the propagator of the Schrödinger equation, show that this integral formula gives the standard expression for the Feynman propagator.

(b) Using the method just described, show that the expression

$$D_F(x,y) = \int_0^\infty dT \int \mathcal{D}x \,\exp\left[i\int dt\,\frac{1}{2}\left(\left(\frac{dx^\mu}{dt}\right)^2 - m^2\right) - ie\int dt\,\frac{dx^\mu}{dt}A_\mu(x)\right]$$

is a functional integral representation for the scalar field propagator in an arbitrary background electromagnetic field. Show, in particular, that the functional integral satisfies the relevant Schrödinger equation. Notice that this integral depends on A_μ through the Wilson line.

(c) Generalize this expression to a non-Abelian gauge theory. Show that the functional integral solves the relevant Schrödinger equation only if the group matrices in the exponential for the Wilson line are path-ordered.

15.5 Casimir operator computations. An alternative strategy for computing the quadratic Casimir operator is to compute $C(r)$ in the formula

$$\mathrm{tr}[t_r^a t_r^b] = C(r)\delta^{ab}$$

by choosing t^a and t^b to lie in an $SU(2)$ subgroup of the gauge group.

(a) Under an $SU(2)$ subgroup of a general group G, an irreducible representation r of G will decompose into a sum of representations of $SU(2)$:

$$r \to \sum j_i,$$

where the j_i are the spins of $SU(2)$ representations. Show that

$$3C(r) = \sum_i j_i(j_i+1)(2j_i+1).$$

(b) Under an $SU(2)$ subgroup of $SU(N)$, the fundamental representation N transforms as a 2-component spinor ($j=\frac{1}{2}$) and $(N-2)$ singlets. Use this relation to check the formula $C(N)=\frac{1}{2}$. Show that the adjoint representation of $SU(N)$ decomposes into one spin 1, $2(N-2)$ spin-$\frac{1}{2}$'s, plus singlets, and use this decomposition to check that $C(G)=N$.

(c) Symmetric and antisymmetric 2-index tensors form irreducible representations of $SU(N)$. Compute $C_2(r)$ for each of these representations. The direct sum of these representations is the product representation $N\times N$. Verify that your results for $C_2(r)$ satisfy the identity for product representations that follows from Eqs. (15.100) and (15.101).

Chapter 16

Quantization of Non-Abelian Gauge Theories

The previous chapter showed how to construct Lagrangians with non-Abelian gauge symmetry. However, this is only the first step in the process of relating the idea of non-Abelian gauge invariance to the real interactions of particle physics. We must next work out the rules for computing Feynman diagrams containing the non-Abelian gauge vector particles, then use these rules to compute scattering amplitudes and cross sections. This chapter will develop the technology needed for such calculations.

Alongside this technical discussion, we will study how the gauge symmetry affects the Feynman amplitudes. In any theory with a local symmetry, some degrees of freedom of the fields that appear in the Lagrangian are *unphysical*, in the sense that they can be adjusted arbitrarily by gauge transformations. In electrodynamics, the components of the field $A_\mu(k)$ proportional to k^μ lie along the symmetry directions. We saw in Section 9.4 that this fact has two important consequences. First, the propagator of the field A_μ is ambiguous; there are multiple expressions for the propagator, which follow equally well from the QED Lagrangian. Second, the vertices of electrodynamics are such that this ambiguity makes no difference in the calculation of cross sections. For example, Eq. (9.58) displays a continuous family of photon propagators, one for each value of the continuous parameter ξ; but we saw immediately that all dependence of S-matrix elements on ξ is eliminated by the Ward identity. Non-Abelian gauge theories contain similar ambiguities and cancellations, but, as we will see in this chapter, the structure of the cancellations is more intricate.

An additional goal of this chapter is to compute the Callan-Symanzik β function, and hence determine the behavior of the running coupling constant, for non-Abelian gauge theories. As discussed in Chapter 14, these theories are in fact *asymptotically free*: The coupling constant becomes weak at large momenta. This result indicates the applicability of non-Abelian gauge theory to model the strong interactions. We will be able to derive this result once we have determined the correct Feynman rules for non-Abelian gauge theories.

16.1 Interactions of Non-Abelian Gauge Bosons

Most of the Feynman rules for non-Abelian gauge theory can be read directly from the Yang-Mills Lagrangian, following the method of Section 9.2. However, when we quantized the electromagnetic field in Section 9.4, we saw that the functional integral over a gauge field must be defined carefully, and that the subtle aspects of this construction can introduce new ingredients into the quantum theory. In this section we will see how far we can go in the non-Abelian theory by ignoring these subtleties. In Section 16.2 we will carry out a more proper derivation of the Feynman rules, through a careful analysis of the functional integral.

Feynman Rules for Fermions and Gauge Bosons

The Yang-Mills Lagrangian, as derived in the previous chapter, is

$$\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2 + \bar{\psi}(i\not{D} - m)\psi, \tag{16.1}$$

where the index a is summed over the generators of the gauge group G, and the fermion multiplet ψ belongs to an irreducible representation r of G. The field strength is

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu, \tag{16.2}$$

where f^{abc} are the structure constants of G. The covariant derivative is defined in terms of the representation matrices t^a_r by

$$D_\mu = \partial_\mu - igA^a_\mu t^a_r. \tag{16.3}$$

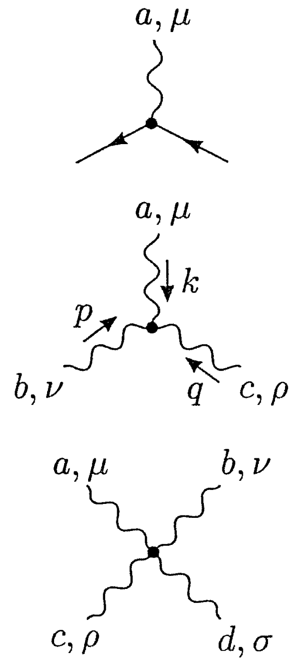
From now on we will drop the subscript r except where it is needed for clarity.

The Feynman rules for this Lagrangian can be derived from a functional integral over the fields ψ, $\bar{\psi}$, and A^a_μ. Imagine expanding the functional integral in perturbation theory, starting with the free Lagrangian, at $g = 0$. The free theory contains of a number of free fermions equal to the dimension $d(r)$ of the representation r, and a number of free vector bosons equal to the number $d(G)$ of generators of G. Using the methods of Section 9.5, it is straightforward to derive the fermion propagator

$$\langle \psi_{i\alpha}(x)\bar{\psi}_{j\beta}(y)\rangle = \int \frac{d^4k}{(2\pi)^4}\left(\frac{i}{\not{k} - m}\right)_{\alpha\beta} \delta_{ij}\, e^{-ik\cdot(x-y)}, \tag{16.4}$$

where α, β are Dirac indices and i, j are indices of the symmetry group: $i, j = 1, \ldots, d(r)$. In analogy with electrodynamics, we would guess that the propagator of the vector fields is

$$\langle A^a_\mu(x) A^b_\nu(y)\rangle = \int \frac{d^4k}{(2\pi)^4}\left(\frac{-ig_{\mu\nu}}{k^2}\right) \delta^{ab}\, e^{-ik\cdot(x-y)}, \tag{16.5}$$



$$= \quad ig\gamma^\mu t^a$$

$$= \quad gf^{abc}\big[g^{\mu\nu}(k-p)^\rho + g^{\nu\rho}(p-q)^\mu + g^{\rho\mu}(q-k)^\nu\big]$$

$$= \quad -ig^2\big[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma}-g^{\mu\sigma}g^{\nu\rho}) + f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma}-g^{\mu\sigma}g^{\nu\rho}) + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma}-g^{\mu\rho}g^{\nu\sigma})\big]$$

Figure 16.1. Feynman rules for fermion and gauge boson vertices of a non-Abelian gauge theory.

with $a, b = 1, \dots, d(G)$. We will derive this formula in the next section.

To find the vertices, we write out the nonlinear terms in (16.1). If $\mathcal{L}_0$ is the free field Lagrangian, then

$$\begin{aligned}\mathcal{L} = \mathcal{L}_0 + gA^a_\lambda\bar{\psi}\gamma^\lambda t^a\psi - gf^{abc}(\partial_\kappa A^a_\lambda)A^{\kappa b}A^{\lambda c} \\ - g^2(f^{eab}A^a_\kappa A^b_\lambda)(f^{ecd}A^{\kappa c}A^{\lambda d}).\end{aligned} \tag{16.6}$$

The first of the three nonlinear terms gives the fermion-gauge boson vertex

$$ig\gamma^\mu t^a; \tag{16.7}$$

this is a matrix that acts on the Dirac and gauge indices of the fermions. The second nonlinear term leads to a three gauge boson vertex. To work out this vertex, we first choose a definite convention for the external momenta and Lorentz and gauge indices. A suitable convention is shown in Fig. 16.1, with all momenta pointing inward. Consider first contracting the external gauge particle with momentum k to the first factor of A^a_μ, the gauge particle with momentum p to the second, and the gauge particle of momentum q to the third. The derivative contributes a factor $(-ik_\kappa)$ if the momentum points into the diagram. Then this contribution is

$$-igf^{abc}(-ik^\nu)g^{\mu\rho}. \tag{16.8}$$

In all, there are 3! possible contractions, which alternate in sign according to the total antisymmetry of f^{abc}. The sum of these is exhibited in Fig. 16.1. Finally, the last term of (16.6) leads to a four gauge boson vertex. Following the conventions of Fig. 16.1, one possible contraction gives the contribution

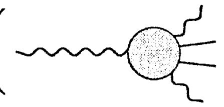
$$-ig^2 f^{eab}f^{ecd}g^{\mu\rho}g^{\nu\sigma}. \tag{16.9}$$

There are 4! possible contractions, of which sets of 4 are equal to one another. The sum of these contributions is shown in Fig. 16.1.

Notice that all of these vertices involve the same coupling constant g. We derived the vertices, and thus the equality of the coupling constants, as a part of our construction of the Lagrangian from the principle of non-Abelian gauge invariance. However, it is also possible to see the need for this equality *a posteriori*, from the properties of Feynman amplitudes.

Equality of Coupling Constants

One property that we expect from Feynman amplitudes in non-Abelian gauge theories is that they should satisfy Ward identities similar to those of QED. These Ward identities express the conservation of the symmetry currents, which follows already from the global symmetry of the theory. In QED, the simplest form of the Ward identity was obtained by putting external electrons and positrons on shell. In non-Abelian gauge theories, the gauge bosons also carry charge and so these must also be put on shell to remove contact terms. With all external particles on shell, the amplitude for production of a gauge boson should obey

$k^\mu \Big($  $\Big) = 0.$ (16.10)

This identity is not only an indication of the local gauge symmetry, but is physically important in its own right. Like the photon, the non-Abelian gauge boson has only two physical polarization states. In QED, the on-shell Ward identity expressed the fact that the orthogonal, unphysical polarization states are not produced in scattering processes. The on-shell Ward identity will play a similar role in the non-Abelian case.

Let us check the Ward identity in a simple case, the lowest-order diagrams contributing to fermion-antifermion annihilation into a pair of gauge bosons. In order g^2, there are three diagrams, shown in Fig. 16.2. The first two diagrams are similar to the QED diagrams that we studied in Section 5.5; they sum to

$$\begin{aligned} i\mathcal{M}^{\mu\nu}_{1,2}\epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2) = (ig)^2\bar{v}(p_+)\Big\{\gamma^\mu t^a \frac{i}{\not{p} - \not{k}_2 - m}\gamma^\nu t^b \\ + \gamma^\nu t^b \frac{i}{\not{k}_2 - \not{p}_+ - m}\gamma^\mu t^a\Big\} u(p)\, \epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2). \end{aligned} \tag{16.11}$$

The vectors $\epsilon(k_i)$ are the gauge boson polarization vectors; for physical polarizations, these satisfy $k_i^\mu \epsilon_\mu(k_i) = 0$. To check the Ward identity (16.10), we

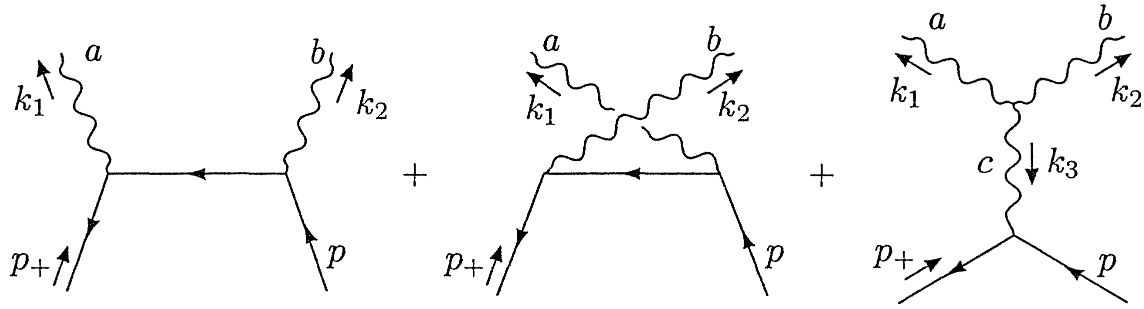


Figure 16.2. Diagrams contributing to fermion-antifermion annihilation to two gauge bosons.

replace $\epsilon^*_\nu(k_2)$ in (16.11) by $k_{2\nu}$. This gives

$$\begin{aligned} i\mathcal{M}^{\mu\nu}_{1,2}\epsilon^*_{1\mu}k_{2\nu} = (ig)^2\bar{v}(p_+)\Big\{\gamma^\mu t^a \frac{i}{\not{p}-\not{k}_2-m}\not{k}_2 t^b \\ + \not{k}_2 t^b \frac{i}{\not{k}_2-\not{p}_+-m}\gamma^\mu t^a\Big\}u(p)\,\epsilon^*_{1\mu}. \end{aligned} \tag{16.12}$$

Since

$$(\not{p}-m)u(p)=0 \qquad \text{and} \qquad \bar{v}(p_+)(-\not{p}_+-m)=0, \tag{16.13}$$

we can add these quantities to $\not{k}_2$ in the first and second terms of (16.12), to cancel the denominators. This gives

$$i\mathcal{M}^{\mu\nu}_{1,2}\epsilon^*_{1\mu}k_{2\nu} = (ig)^2\bar{v}(p_+)\Big\{-i\gamma^\mu[t^a,t^b]\Big\}u(p)\,\epsilon^*_{1\mu}. \tag{16.14}$$

In the Abelian case, this expression would vanish. In the non-Abelian case, however, the residual term is nonzero and depends on the commutator of gauge group generators:

$$i\mathcal{M}^{\mu\nu}_{1,2}\epsilon^*_{1\mu}k_{2\nu} = -g^2\bar{v}(p_+)\gamma^\mu u(p)\,\epsilon^*_{1\mu}\cdot f^{abc}t^c. \tag{16.15}$$

We need to find another contribution to cancel this term. Notice, however, that this term has the group index structure of a fermion-gauge boson vertex ($g\gamma^\mu t^c$) multiplied by a three gauge boson vertex (gf^{abc}). This is just the structure of the third diagram in Fig. 16.2.

To check that the cancellation works, let us evaluate the third diagram:

$$\begin{aligned} i\mathcal{M}^{\mu\nu}_3\epsilon^*_{1\mu}\epsilon^*_{2\nu} = ig\bar{v}(p_+)\gamma_\rho t^c u(p)\,\frac{-i}{k_3^2}\,\epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2) \\ \times\, gf^{abc}\big[g^{\mu\nu}(k_2-k_1)^\rho + g^{\nu\rho}(k_3-k_2)^\mu + g^{\rho\mu}(k_1-k_3)^\nu\big], \end{aligned}$$

with $k_3 = -k_1-k_2$. If we replace $\epsilon_\nu(k_2)$ with $k_{2\nu}$, then eliminate k_2 using momentum conservation, the expression in brackets simplifies as follows:

$$\begin{aligned} &\epsilon^*_\nu(k_2)\big[g^{\mu\nu}(k_2-k_1)^\rho + g^{\nu\rho}(k_3-k_2)^\mu + g^{\rho\mu}(k_1-k_3)^\nu\big] \\ &\quad\to k_2^\mu(k_2-k_1)^\rho + k_2^\rho(k_3-k_2)^\mu + g^{\rho\mu}(k_1-k_3)\cdot k_2 \\ &\quad= g^{\rho\mu}k_3^2 - k_3^\rho k_3^\mu - g^{\rho\mu}k_1^2 + k_1^\rho k_1^\mu. \end{aligned} \tag{16.16}$$

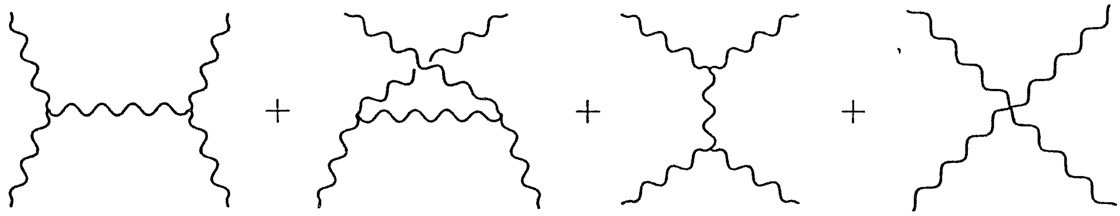


Figure 16.3. Diagrams contributing to gauge boson-gauge boson scattering.

Let us assume that the other gauge boson, with momentum k_1, is on shell ($k_1^2 = 0$), and that it has transverse polarization ($k_1^\mu \epsilon_\mu(k_1) = 0$). Then the third and fourth terms in the last line vanish. Furthermore, the term $k_3^\rho k_3^\mu$ vanishes when it is contracted with the fermion current. In the remaining term, the factor k_3^2 cancels the gauge boson propagator, and we are left with

$$i\mathcal{M}_3^{\mu\nu}\epsilon_{1\mu}^* k_{2\nu} = +g^2\bar{v}(p_+)\gamma^\mu u(p)\,\epsilon_{1\mu}^* \cdot f^{abc}t^c, \tag{16.17}$$

which precisely cancels (16.15).

Notice that this cancellation takes place only if the value of the coupling constant in the three-boson vertex is identical to that in the fermion-boson vertex. In a similar way, the Ward identity cannot be satisfied among the diagrams for boson-boson scattering, shown in Fig. 16.3, unless the coupling constant g in the four-boson vertex is identical to that in the three-boson vertex. Thus, the coupling constants of all three nonlinear terms in the Yang-Mills Lagrangian *must* be equal in order to preserve the Ward identity and avoid the production of bosons with unphysical polarization states. Conversely, the non-Abelian gauge symmetry guarantees that these couplings *are* equal. The symmetry thus accomplishes exactly what we hoped it would in our discussion at the beginning of Chapter 15, giving us a consistent theory of physical vector particle interactions.

A Flaw in the Argument

The preceding argument has one serious deficiency. At the final stage, we needed to assume that the second gauge boson was transverse. However, one might have expected that this information would come out of the argument rather than having to be put in. In QED, the Feynman diagrams predict that, when an electron and a positron annihilate to form two photons, only the physical transverse polarization states of the photons are produced. Amplitudes to produce other photon polarizations cancel each other to yield zero, as we saw in Eq. (5.80). This statement is not true for the non-Abelian gauge theory Feynman rules that we have worked with so far.

To state the discrepancy more concretely, we introduce some notation. Let $k^\mu = (k^0, \mathbf{k})$ be a lightlike vector: $k^2 = 0$. Then there are two purely spatial vectors orthogonal to $\mathbf{k}$. If k is the momentum of a vector boson, these are the two transverse polarizations. To construct an orthogonal basis, we must include also the longitudinal polarization state, with polarization vector parallel to $\mathbf{k}$, and the timelike polarization state. It is most convenient to work with the two lightlike linear combinations of these states, with polarization vectors parallel to the vectors k^μ and $\tilde{k}^\mu = (k^0, -\mathbf{k})$. These two unphysical polarization states of a massless vector particle can be written as follows:

$$\epsilon^+_\mu(k) = \left(\frac{k^0}{\sqrt{2}|\mathbf{k}|}, \frac{\mathbf{k}}{\sqrt{2}|\mathbf{k}|}\right); \qquad \epsilon^-_\mu(k) = \left(\frac{k^0}{\sqrt{2}|\mathbf{k}|}, -\frac{\mathbf{k}}{\sqrt{2}|\mathbf{k}|}\right). \tag{16.18}$$

We will refer to $\epsilon^+(k)$ and $\epsilon^-(k)$ as the *forward* and *backward* lightlike polarization vectors. Denote the two transverse polarization states $\epsilon^T_{i\mu}(k)$, for $i = 1, 2$. These four polarization vectors obey the orthogonality relations

$$\begin{gathered} \epsilon^T_i \cdot \epsilon^{*T}_j = -\delta_{ij}, \qquad \epsilon^+ \cdot \epsilon^T_i = \epsilon^- \cdot \epsilon^T_i = 0, \\ (\epsilon^+)^2 = (\epsilon^-)^2 = 0, \qquad \epsilon^+ \cdot \epsilon^- = 1. \end{gathered} \tag{16.19}$$

They also satisfy the completeness relation

$$g_{\mu\nu} = \epsilon^-_\mu \epsilon^{+*}_\nu + \epsilon^+_\mu \epsilon^{-*}_\nu - \sum_i \epsilon^T_{i\mu} \epsilon^{T*}_{i\nu}. \tag{16.20}$$

Using this notation, we can express concretely the gap in the argument for the Ward identity. The Feynman diagrams of Fig. 16.2 apparently predict that there is a nonzero amplitude to produce a forward-polarized gauge boson together with a backward-polarized gauge boson. For this case, we substitute $\epsilon^{-*}_\mu(k_1)$ and $\epsilon^{+*}_\nu(k_2)$ for the two polarization vectors. Then the term proportional to $k_1^\rho k_1^\mu$ in Eq. (16.16) no longer vanishes; it now yields

$$\begin{aligned} i\mathcal{M} &= ig\bar{v}(p_+)\gamma_\rho t^c u(p)\, \frac{-i}{k_3^2}\, \epsilon^{-*}_\mu(k_1) \cdot \frac{1}{\sqrt{2}|\mathbf{k}_2|} \cdot g f^{abc}\big[-k_1^\rho k_1^\mu\big] \\ &= ig\bar{v}(p_+)\gamma_\rho t^c u(p)\, \frac{-i}{k_3^2} \cdot (-g) f^{abc} k_1^\rho \cdot \frac{|\mathbf{k}_1|}{|\mathbf{k}_2|}. \end{aligned} \tag{16.21}$$

Can we simply ignore this totally unphysical process? We are free to do so in calculations of leading-order amplitudes, but the process will come back to haunt us in loop diagrams. Recall from Section 7.3 how the optical theorem (7.49) links the imaginary part of a loop diagram to the square of a corresponding scattering amplitude, obtained by cutting the diagram across the loop. If we apply the optical theorem to the diagram shown in Fig. 16.4, we obtain a paradox. In the gauge boson loop on the left-hand side we can replace the $g_{\mu\nu}$ factors in the propagators with sums over all four polarization vectors (16.20). The theorem thus implies that all four polarizations, even the unphysical ones, should be included for the final-state gauge bosons on the right-hand side. We are faced with a choice of allowing the production of unphysical states or violating the optical theorem. A third alternative, equally unattractive, would be to discard our expression (16.5) for the gauge boson propagator. Clearly, we are missing some crucial element of the quantum-mechanical structure of non-Abelian gauge theories.

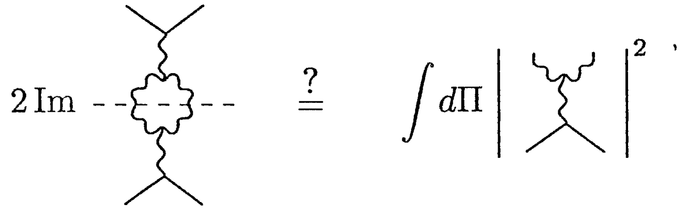


Figure 16.4. A paradox for the optical theorem in gauge theories.

16.2 The Faddeev-Popov Lagrangian

It is not surprising that we have found a problem with our Feynman rules for non-Abelian gauge theories, since we were not very careful in deriving them. In particular, we did not actually derive expression (16.5) for the gauge field propagator. In this section we will remedy this by going through a formal derivation of this expression. We will find that, although expression (16.5) is indeed correct, it is incomplete: It must be supplemented by additional rules of a completely new type.

To define the functional integral for a theory with non-Abelian gauge invariance, we will use the Faddeev-Popov method, as introduced in Section 9.4 to quantize the electromagnetic field. Our present discussion will follow Section 9.4 closely. However, as we have by now come to expect, the case of non-Abelian local symmetry brings with it new tricks and surprises.

First consider the quantization of the pure gauge theory, without fermions. To derive the Feynman rules, we must define the functional integral

$$\int \mathcal{D}A \, \exp\Big[i\int d^4x\Big(-\tfrac{1}{4}(F^a_{\mu\nu})^2\Big)\Big]. \tag{16.22}$$

As in the Abelian case, the Lagrangian is unchanged along the infinite number of directions in the space of field configurations corresponding to local gauge transformations. To compute the functional integral we must factor out the integrations along these directions, constraining the remaining integral to a much smaller space.

As in electrodynamics, we will constrain the gauge directions by applying a gauge-fixing condition $G(A) = 0$ at each point x. Following Faddeev and Popov, we can introduce this constraint by inserting into the functional integral the identity (9.53):

$$1 = \int \mathcal{D}\alpha(x)\, \delta\big(G(A^\alpha)\big)\, \det\Big(\frac{\delta G(A^\alpha)}{\delta\alpha}\Big). \tag{16.23}$$

Here A^α is the gauge field A transformed through a finite gauge transformation as in (15.47):

$$(A^\alpha)^a_\mu t^a = e^{i\alpha^a t^a}\big[A^b_\mu t^b + \frac{i}{g}\partial_\mu\big]e^{-i\alpha^c t^c}. \tag{16.24}$$

In evaluating the determinant, the infinitesimal form of this transformation will be more useful:

$$(A^\alpha)^a_\mu = A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c = A^a_\mu + \frac{1}{g} D_\mu \alpha^a, \tag{16.25}$$

where D_μ is the covariant derivative (15.86) acting on a field in the adjoint representation. Note that, as long as the gauge-fixing function $G(A)$ is linear, the functional derivative $\delta G(A^\alpha)/\delta\alpha$ is independent of α.

Since the Lagrangian is gauge invariant, we can replace A by A^α in the exponential of (16.22). Then, as in the Abelian case, we can interchange the order of the functional integrals over A and α, and then change variables in the inner integral from A to $A' = A^\alpha$. The transformation (16.24) looks more complicated than in the Abelian case, but it is nothing more than a linear shift of the A^a_μ, followed by a unitary rotation of the various components of the symmetry multiplet $A^a_\mu(x)$ at each point. Both of these operations preserve the measure

$$\mathcal{D}A = \prod_x \prod_{a,\mu} dA^a_\mu. \tag{16.26}$$

Thus $\mathcal{D}A = \mathcal{D}A'$, under the integral over α. Just as in the Abelian case, the integral over gauge motions α can be factored out of the functional integral into an overall normalization, leaving us with

$$\int \mathcal{D}A\, e^{iS[A]} = \left(\int \mathcal{D}\alpha\right) \int \mathcal{D}A\, e^{iS[A]}\, \delta\big(G(A)\big)\, \det\Big(\frac{\delta G(A^\alpha)}{\delta\alpha}\Big). \tag{16.27}$$

This normalization factor cancels in the computation of correlation functions of gauge-invariant operators.

From this point, the derivation of the gauge boson propagator proceeds as for the photon propagator. We choose the generalized Lorentz gauge condition

$$G(A) = \partial^\mu A^a_\mu(x) - \omega^a(x), \tag{16.28}$$

with a Gaussian weight for ω^a as in Eq. (9.56). The manipulations of Section 9.4 then lead to the class of gauge field propagators

$$\big\langle A^a_\mu(x) A^b_\nu(y)\big\rangle = \int \frac{d^4k}{(2\pi)^4} \frac{-i}{k^2+i\epsilon}\left(g_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k^2}\right)\delta^{ab} e^{-ik\cdot(x-y)}, \tag{16.29}$$

with a freely adjustable gauge parameter ξ. Our guess (16.5) corresponds to the choice $\xi = 1$, called the *Feynman-'t Hooft gauge*.

So far, this whole derivation parallels the case of electrodynamics. Here, however, there is one more nontrivial ingredient. In QED, the determinant in Eq. (16.23) was independent of A, so this quantity could be treated as just another contribution to the normalization factor. In the non-Abelian case this is no longer true: Using the infinitesimal form (16.25) of the gauge transformation, we can evaluate

$$\frac{\delta G(A^\alpha)}{\delta\alpha} = \frac{1}{g}\partial^\mu D_\mu, \tag{16.30}$$

acting on a field in the adjoint representation; this operator depends on A. The functional determinant of (16.30) thus contributes new terms to the Lagrangian.

Faddeev and Popov chose to represent this determinant as a functional integral over a new set of anticommuting fields belonging to the adjoint representation:

$$\det\left(\frac{1}{g}\partial^\mu D_\mu\right) = \int \mathcal{D}c\mathcal{D}\bar{c}\, \exp\left[i\int d^4x\, \bar{c}(-\partial^\mu D_\mu)c\right]. \tag{16.31}$$

We derived this formal identity in Eq. (9.69), using our rules for fermionic functional integrals. (The factor of $1/g$ is absorbed into the normalization of the fields c and $\bar{c}$.) But to give the correct identity, c and $\bar{c}$ must be anticommuting fields that are scalars under Lorentz transformations. The quantum excitations of these fields have the wrong relation between spin and statistics to be physical particles. However, we can nevertheless treat these excitations as additional particles in the computation of Feynman diagrams. These new fields and their particle excitations are called *Faddeev-Popov ghosts.*

If we temporarily suppress our curiosity about the physical interpretation of the ghosts, we can work out their Feynman rules. We write the ghost Lagrangian more explicitly as

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a\left(-\partial^2\delta^{ac} - g\partial^\mu f^{abc}A^b_\mu\right)c^c. \tag{16.32}$$

The first term gives a ghost propagator,

$$\left\langle c^a(x)\bar{c}^b(y)\right\rangle = \int \frac{d^4k}{(2\pi)^4}\frac{i}{k^2}\delta^{ab}e^{-ik\cdot(x-y)}. \tag{16.33}$$

In a diagram, this propagator carries an arrow that shows the flow of ghost number, as in Fig. 16.5. In the interaction term of (16.32), the derivative stands to the left of the gauge field; this implies that this derivative is evaluated with the momentum coming out of the vertex along the ghost line. The explicit Feynman rule is shown in Fig. 16.5. As with the other vertices we have encountered, the coupling constant g that appears in this vertex must be equal to the coupling constant g in the three-boson vertex in order to avoid upsetting the Ward identities.

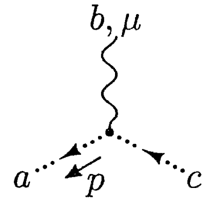
There are no further subtleties in the construction of the perturbation theory for non-Abelian gauge theories. In particular, it is straightforward to include fermions. The final Lagrangian, including all of the effects of Faddeev-Popov gauge fixing, is

$$\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2 + \frac{1}{2}\xi(\partial^\mu A^a_\mu)^2 + \bar{\psi}(i\not{D} - m)\psi + \bar{c}^a(-\partial^\mu D^{ac}_\mu)c^c. \tag{16.34}$$

This Lagrangian leads to the propagator (16.29), and to the set of Feynman rules for vertices shown in Figs. 16.1 and 16.5.

The argument we have just completed suffices to derive the Feynman diagram expansion of any correlation function of gauge-invariant operators in a non-Abelian gauge theory. At the end of Section 9.4, we explained that the

$$= \frac{i\delta^{ab}}{p^2}$$



$$= g f^{abc} p^{\mu}$$

Figure 16.5. Feynman rules for Faddeev-Popov ghosts.

Faddeev-Popov gauge-fixing technique also gives the correct gauge-invariant expressions for S-matrix elements. This remains true in the non-Abelian case. However, the argument given in Section 9.4 relied upon the cancellation in QED of the emission probabilities for timelike and longitudinal photons, and we have already found that this cancellation does not go through in the non-Abelian case. In Section 16.4 we will construct a more sophisticated argument, in which the Faddeev-Popov ghosts play an essential role, that will correctly generalize our previous argument to non-Abelian gauge theories.

16.3 Ghosts and Unitarity

We might now ask whether the new ingredients that we found in the previous section, the Faddeev-Popov ghosts, can resolve the paradox that we encountered at the end of Section 16.1. There we saw that the first diagram in Fig. 16.6 contains a nonzero contribution to its imaginary part that does not correspond to a possible final state with physical gauge boson polarizations. We will now compute this contribution more carefully. We must then add a new potential contribution from the ghosts, shown as the second diagram in Fig. 16.6.

Let us call the amplitude for fermion-fermion annihilation into gauge bosons, which we studied in Section 16.1,

$$i\mathcal{M}^{\mu\nu}\epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2); \tag{16.35}$$

the amplitude for two gauge bosons to convert to a fermion-antifermion pair will be, correspondingly, $\mathcal{M}'$. Then, following the Cutkosky rules of Section 7.3, we find the imaginary part of the first digram in Fig. 16.6 by replacing the cut gauge boson propagator with momentum k_i by

$$-ig_{\mu\nu}\cdot(-2\pi i)\delta(k_i^2). \tag{16.36}$$

Replacing both propagators gives two delta functions, turning the four-dimensional integrals over the gauge boson momenta into three-dimensional phase space integrals, as in the example in Section 7.3. We are thus left with the expression

$$\tfrac{1}{2}(i\mathcal{M}^{\mu\nu})g_{\mu\rho}g_{\nu\sigma}(i\mathcal{M}'^{\rho\sigma}), \tag{16.37}$$

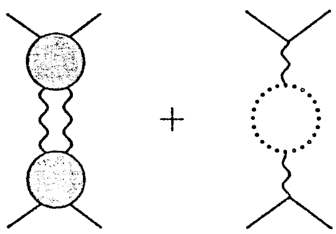


Figure 16.6. The diagram on the left, in which each circle represents the sum of the three contributions of Fig. 16.2, gives a possible problem for the optical theorem. The ghost diagram on the right cancels the anomalous terms.

integrated over the phase space of two massless particles. The factor 1/2 is a symmetry factor for the Feynman diagram or, equivalently, a correction to the phase space integral for identical particles.

Now introduce the representation (16.20) for $g_{\mu\rho}$ and $g_{\nu\sigma}$. The pieces that involve only transverse polarizations correspond to the expected imaginary parts necessary to satisfy the optical theorem. We need not consider these terms further. The cross terms between physical and unphysical polarizations vanish: We showed in Section 16.1 that

$$i\mathcal{M}^{\mu\nu}\epsilon_\mu^{T*}(k_1)\epsilon_\nu^{+*}(k_2) = 0. \tag{16.38}$$

The same identity holds if $\mathcal{M}$ is replaced by $\mathcal{M}'$, and if ϵ^+ is replaced by ϵ^-. Furthermore, the amplitude vanishes if both polarization vectors are forward or both are backward. The only surviving terms are the cross terms between forward and backward polarization, which yield the expression

$$\tfrac{1}{2}\left[(i\mathcal{M}^{\mu\nu}\epsilon_\mu^{-*}\epsilon_\nu^{+*})(i\mathcal{M}'^{\rho\sigma}\epsilon_\rho^{+}\epsilon_\sigma^{-}) + (i\mathcal{M}^{\mu\nu}\epsilon_\mu^{+*}\epsilon_\nu^{-*})(i\mathcal{M}'^{\rho\sigma}\epsilon_\rho^{-}\epsilon_\sigma^{+})\right], \tag{16.39}$$

integrated over phase space. We worked out the value of the first factor in Eq. (16.21), and the contraction with $\mathcal{M}'$ is very similar. Substituting these results, expression (16.39) becomes

$$\begin{aligned}\frac{1}{2}&\Big(ig\bar{v}(p_+)\gamma_\mu t^c u(p)\cdot\frac{-i}{(k_1+k_2)^2}\cdot(-gf^{abc}k_1^\mu)\Big)\\ &\times\Big(ig\bar{u}(p')\gamma_\rho t^d v(p'_+)\cdot\frac{-i}{(k_1+k_2)^2}\cdot(-gf^{abd}(-k_2)^\rho)\Big) + (k_1\leftrightarrow k_2).\end{aligned} \tag{16.40}$$

Using the identity

$$\bar{v}(p_+)\gamma_\mu(k_1+k_2)^\mu u(p) = \bar{v}(p_+)\gamma_\mu(p+p_+)^\mu u(p) = 0, \tag{16.41}$$

we see that the two terms added in (16.40) are equal.

Now add the contribution from the Faddeev-Popov ghosts. Using the Feynman rules in Fig. 16.5, we can assemble the amplitude for fermion-antifermion annihilation into a pair of ghosts:

$$i\mathcal{M}_{\text{ghost}} = ig\bar{v}(p_+)\gamma_\mu t^c u(p)\cdot\frac{-i}{(k_1+k_2)^2}\cdot(-gf^{abc}k_1^\mu). \tag{16.42}$$

This is precisely the first half of expression (16.40). Similarly, the amplitude for the ghost-antighost pair to annihilate into fermions is equal to the second half of (16.40). Finally, since Faddeev-Popov ghost fields anticommute, we must supply a factor of -1 for each ghost loop. Thus the ghost contribution exactly cancels the contribution of unphysical gauge boson polarizations to the Cutkosky cut of the diagrams in Fig. 16.6.

This example illustrates a general physical interpretation of Faddeev-Popov ghosts. These "particles" serve as negative degrees of freedom to cancel the effects of the unphysical timelike and longitudinal polarization states of the gauge bosons. The simplest effect of the ghosts can already be seen from the determinants that appear when one integrates over the gauge and ghost fields in the Faddeev-Popov Lagrangian (16.34). In a general dimension d, working in Feynman gauge and at zero coupling for simplicity, the functional integral over the gauge and ghost fields in (16.34) yields

$$\left(\det[-\partial^2]\right)^{-d/2} \cdot \left(\det[-\partial^2]\right)^{+1}. \tag{16.43}$$

The second determinant, which appears with a positive exponent because the ghost fields anticommute, cancels the contribution to the first determinant of two components of the field A_μ. This physical effect was illustrated, using the language of Section 9.4, in Problem 9.2.

16.4 BRST Symmetry

To show how this cancellation extends to the complete interacting theory, Becchi, Rouet, Stora, and Tyutin introduced as a beautiful formal tool a new symmetry of the gauge-fixed Lagrangian (16.34), which involves the ghost in an essential way.* This *BRST symmetry* has a continuous parameter that is an anticommuting number. To write the symmetry in its simplest form, let us rewrite the Faddeev-Popov Lagrangian by introducing a new (commuting) scalar field B^a:

$$\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2 + \bar{\psi}(i\not{D} - m)\psi - \frac{1}{2\xi}(B^a)^2 + B^a\partial^\mu A^a_\mu + \bar{c}^a(-\partial^\mu D^{ac}_\mu)c^c. \tag{16.44}$$

The new field B^a has a quadratic term without derivatives, so it is not a normal propagating field. The functional integral over B^a can be done by completing the square in (16.44); this procedure brings us back precisely to Eq. (16.34). A field of this type, which appears in the functional integral but has no independent dynamics, is called an *auxilliary field.*

*C. Becchi, A. Rouet, and R. Stora, *Ann. Phys.* **98**, 287 (1976); I. V. Tyutin, Lebedev Institute preprint (1975, unpublished); M. Z. Iofa and I. V. Tyutin, *Theor. Math. Phys.* **27**, 316 (1976).

Now let ϵ be an infinitesimal anticommuting parameter, and consider the following infinitesimal transformation of the fields in (16.44):

$$\begin{aligned}\delta A^a_\mu &= \epsilon D^{ac}_\mu c^c \\ \delta\psi &= ig\epsilon c^a t^a \psi \\ \delta c^a &= -\tfrac{1}{2} g\epsilon f^{abc} c^b c^c \\ \delta\bar{c}^a &= \epsilon B^a \\ \delta B^a &= 0.\end{aligned} \tag{16.45}$$

The transformation of the fields A^a_μ and ψ is a local gauge transformation whose parameter is proportional to the ghost field: $\alpha^a(x) = g\epsilon c^a(x)$. Thus, the first two terms of (16.44) are invariant to (16.45). The third term is trivially invariant. The transformation of A^a_μ in the fourth term cancels the transformation of $\bar{c}^a$ in the last term. Finally, we must examine the transformation of the last ingredient in (16.44):

$$\begin{aligned}\delta(D^{ac}_\mu c^c) &= D^{ac}_\mu \delta c^c + g f^{abc} \delta A^b_\mu c^c \\ &= -\tfrac{1}{2} g\epsilon\partial_\mu (f^{abc} c^b c^c) - \tfrac{1}{2} g^2 \epsilon f^{abc} f^{cde} A^b_\mu c^d c^e \\ &\qquad + g\epsilon f^{abc} (\partial_\mu c^b) c^c + g^2 \epsilon f^{abc} f^{bde} A^d_\mu c^e c^c.\end{aligned} \tag{16.46}$$

The two terms of order g manifestly cancel. By using the anticommuting nature of the ghost fields and exchanging the names of indices, we can write the remaining two terms as

$$-\tfrac{1}{2} g^2 f^{abc} f^{cde} \big(A^b_\mu c^d c^e + A^d c^e c^b + A^e c^b c^d\big), \tag{16.47}$$

which vanishes by the Jacobi identity (15.70). Apparently, the BRST transformation (16.45) is a global symmetry of the gauge-fixed Lagrangian (16.44), for any value of the gauge parameter ξ.

The BRST transformation has one more remarkable feature, which is a natural consequence of its anticommuting nature. Let $Q\phi$ be the BRST transformation of the field ϕ: $\delta\phi = \epsilon Q\phi$. For example, $QA^a_\mu = D^{ac}_\mu c^c$. Then, for any field, the BRST variation of $Q\phi$ vanishes:

$$Q^2\phi = 0. \tag{16.48}$$

The vanishing of (16.46) proves this identity for the second BRST variation of the gauge field. For the ghost field,

$$Q^2 c^a = \tfrac{1}{2} g^2 f^{abc} f^{bde} c^c c^d c^e, \tag{16.49}$$

which vanishes by the Jacobi identity. It is straightforward to check that the second BRST variations of the other fields in (16.44) also vanish.

To describe the implications of identity (16.48), we now consider studying the effective theory (16.44) in the Hamiltonian picture after canonical quantization. Because the Lagrangian has the continuous symmetry (16.45), the theory will have a conserved current, and the integral of the time component of this current will be a conserved charge Q that commutes with H. The action of Q on field configurations will be just that described in the previous paragraph. The relation (16.48) is equivalent to the operator identity

$$Q^2 = 0. \tag{16.50}$$

We say that the BRST operator Q is *nilpotent.*

A nilpotent operator that commutes with H divides the eigenstates of H into three subspaces. Many eigenstates of H must be annihilated by Q so that (16.50) can be satisfied. Let $\mathcal{H}_1$ be the subspace of states that are *not* annihilated by Q. Let $\mathcal{H}_2$ be the subspace of states of the form

$$|\psi_2\rangle = Q\,|\psi_1\rangle\,, \tag{16.51}$$

where $|\psi_1\rangle$ is in $\mathcal{H}_1$. According to (16.50), acting Q again on these states gives zero. Finally, let $\mathcal{H}_0$ be the subspace of states $|\psi_0\rangle$ that satisfy $Q\,|\psi_0\rangle = 0$ but that cannot be written in the form (16.51). The subspace $\mathcal{H}_2$ is quite peculiar, because any two states in this subspace have zero inner product:

$$\langle\psi_{2a}|\psi_{2b}\rangle = \langle\psi_{1a}|\,Q\,|\psi_{2b}\rangle = 0 \tag{16.52}$$

by (16.50). By the same argument, the states of $\mathcal{H}_2$ have zero inner product with the states of $\mathcal{H}_0$.

These considerations seem extremely abstract, but they have a direct physical correspondence.† To see this, consider single-particle states of the non-Abelian gauge theory in the limit of zero coupling. According to the transformation (16.45), Q converts the forward component of A^a_μ to a ghost field; equivalently, Q converts a single forward-polarized gauge boson to a ghost. At $g = 0$, Q annihilates the one-ghost state. At the same time, Q converts the antighost state to a quantum of B^a. To identify this state, note that the Lagrangian (16.44) implies the classical field equation

$$\frac{1}{\xi}B^a = \partial^\mu A^a_\mu. \tag{16.53}$$

Thus the quanta of the field B^a are those quanta of A^a_μ with polarization vectors such that $k^\mu\epsilon_\mu(k) \neq 0$; these are the backward-polarized gauge bosons.

We have now seen that, among the single-particle states of the gauge theory, forward gauge bosons and antighosts belong to $\mathcal{H}_1$, ghosts and backward gauge bosons belong to $\mathcal{H}_2$, and transverse gauge bosons belong to $\mathcal{H}_0$. More generally, it can be shown that asymptotic states containing ghosts,

†The following argument is presented only at an intuitive level. For a rigorous discussion, see T. Kugo and I. Ojima, *Prog. Theor. Phys.* **66**, 1 (1979).

antighosts, or gauge bosons of unphysical polarization always belong to $\mathcal{H}_1$ or $\mathcal{H}_2$, while the asymptotic states in $\mathcal{H}_0$ are those with only transversely polarized gauge bosons. The BRST operator thus gives a precise relation between the unphysical gauge boson polarization states and the ghosts and antighosts as positive and negative degrees of freedom.

In Section 9.4, we argued that the Faddeev-Popov prescription gave the correct, gauge-invariant result for a certain subclass of S-matrix elements, from which we could compute the physical scattering cross sections of transversely polarized gauge bosons. These S-matrix elements were constructed by putting operators in the far past to create transversely polarized gauge bosons, adiabatically turning on the gauge coupling, adiabatically turning off the gauge coupling, and then placing operators in the far future to annihilate gauge bosons with transverse polarization. However, this argument had a possible problem: If the states created as collections of transversely polarized bosons in the far past could evolve into states that contained gauge bosons of other polarizations in the far future, the S-matrix projected between transverse gauge boson states would not be unitary. This problem would also lead to the technical problem discussed in the previous section: The Cutkosky cuts of diagrams contributing to S-matrix elements would have nonzero contributions from unphysical polarizations. In Section 9.4, we used an argument special to the Abelian case to show that these problems do not arise in QED. In the non-Abelian case, the removal of unphysical gauge boson polarizations is more subtle, and we have seen that it involves the ghosts in an essential way. To resolve this subtle problem, we apply the principle of BRST symmetry.

Let $|A;\mathrm{tr}\rangle$ be an external state that contains no ghosts or antighosts and only gauge bosons with transverse polarization. We wish to show that the S-matrix projected onto such states is unitary:

$$\sum_C \langle A;\mathrm{tr}|\, S^\dagger \,|C;\mathrm{tr}\rangle \,\langle C;\mathrm{tr}|\, S \,|B;\mathrm{tr}\rangle = \langle A;\mathrm{tr}|\, \mathbf{1} \,|B;\mathrm{tr}\rangle . \tag{16.54}$$

As we explained above, the physical states $|A;\mathrm{tr}\rangle$ belong to—and, in fact, span—the subspace $\mathcal{H}_0$ defined by the BRST operator. In particular, all of these states are annihilated by Q. Since Q commutes with the Hamiltonian, the time evolution of any such state must also produce a state annihilated by Q. Thus,

$$Q \cdot S\,|A;\mathrm{tr}\rangle = 0. \tag{16.55}$$

This implies that the states $S\,|A;\mathrm{tr}\rangle$ must be linear combinations of states in $\mathcal{H}_0$ and $\mathcal{H}_2$. However, states in $\mathcal{H}_2$ have zero inner product with one another and with states in $\mathcal{H}_0$. Thus the inner product of any two states of the form $S\,|A;\mathrm{tr}\rangle$ comes only from the overlap of the components in $\mathcal{H}_0$, so we can write

$$\langle A;\mathrm{tr}|\, S^\dagger \cdot S \,|B;\mathrm{tr}\rangle = \sum_C \langle A;\mathrm{tr}|\, S^\dagger \,|C;\mathrm{tr}\rangle \,\langle C;\mathrm{tr}|\, S \,|B;\mathrm{tr}\rangle . \tag{16.56}$$

Since the full S-matrix is unitary, this relation implies that the restricted S-matrix is also unitary, Eq. (16.54). In addition, (16.56) implies that the sum of the Cutkosky cuts of diagrams contributing to the S-matrix in a given order is equal to the sum of the cuts involving transverse gauge bosons only. Thus, the cancellation between diagrams that produce pairs of gauge bosons with unphysical polarizations and those that produce ghosts is a general property that persists to all orders in perturbation theory.

Since the BRST transformation generates a continuous symmetry, it generates a set of Ward identities. These identities are similar in structure to the Ward identities of the non-Abelian gauge symmetry, since the BRST symmetry contains a gauge transformation whose parameter is the ghost field. However, the identities that follow from BRST symmetry are simpler. We will not study the Ward identities of non-Abelian gauge theory further in this book. However, when one discusses the renormalization of gauge theories at a higher level, the central identities among renormalization constants that follow from the Ward identities are most easily derived using the BRST symmetry.‡

16.5 One-Loop Divergences of Non-Abelian Gauge Theory

Now that we have discussed the general properties of tree-level diagrams in non-Abelian gauge theories, we turn our attention to diagrams with loops. As always in quantum field theory, some of these loop diagrams will diverge, and we must take care to treat the divergent integrals correctly.

The Lagrangian of a non-Abelian gauge theory (15.39) contains no interactions of dimension higher than 4. Therefore, by the general arguments of Chapter 10, this Lagrangian is renormalizable, in the sense that the divergences can be removed by a finite number of counterterms. However, in non-Abelian gauge theories, as in QED, the gauge symmetries of the theory imply stronger restrictions on the structure of the divergences. In QED, provided that we use a gauge-invariant regulator, there are only four possible divergent coefficients, which are subtracted by the counterterms for the electromagnetic vertex (δ_1), for the electron and photon field strength (δ_2 and δ_3), and for the electron mass (δ_m). In particular, the possibility of a photon mass renormalization is excluded by gauge invariance. Furthermore, the two counterterms δ_1 and δ_2 are equal to one another, and cancel in the evaluation of the electron-photon vertex function, as a consequence of the Ward identity. Non-Abelian gauge symmetries imply similar restrictions on the divergences of Feynman diagrams. In this section, we will illustrate some of these restrictions through examples of one-loop digrams.

‡An introduction to the Ward identities of the BRST symmetry is given by Taylor (1976).

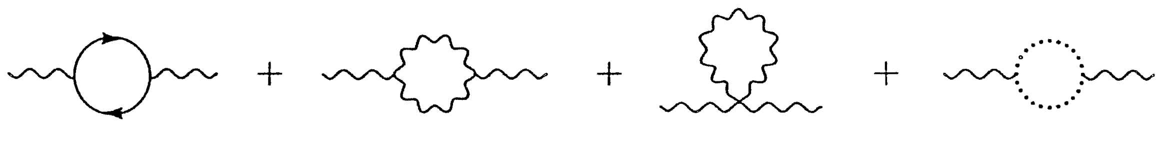
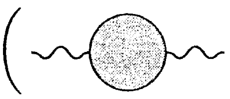


Figure 16.7. Contributions to the gauge boson self-energy in order g^2.

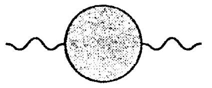
The Gauge Boson Self-Energy

In QED, the strongest constraints of gauge invariance come in the evaluation of the photon self-energy. The Ward identity implies the relation



$$q^\mu \left(\quad \right) = 0, \tag{16.57}$$

which in turn implies that the photon self-energy diagrams have the structure

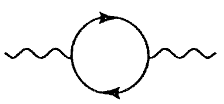


$$= i(q^2 g^{\mu\nu} - q^\mu q^\nu)\Pi(q^2). \tag{16.58}$$

The only divergence possible is a logarithmically divergent contribution to $\Pi(q^2)$. In non-Abelian gauge theories, (16.57) still holds, so the self-energy again has the Lorentz structure (16.58). However, the cancellations that lead to this structure are more complex. Here we will exhibit these cancellations by computing the gauge boson self-energy in detail at the one-loop level. In order to preserve gauge invariance, we will use dimensional regularization.

The contributions of order g^2 to the gauge boson self-energy are shown in Fig. 16.7. (In addition to these 1PI diagrams, there are three "tadpole" diagrams; but these automatically vanish, as in QED, by the argument given below Eq. (10.5).) The fermion loop diagram can be considered separately from the other diagrams, since in principle we could include any number of fermions in the theory. We will see below that the contributions of the three remaining diagrams interlock in an essential way.

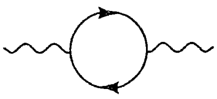
Let us first calculate the fermion loop diagram. The Feynman rule for the vertices in this diagram is identical to the QED Feynman rule, except for the addition of a group matrix t^a that acts on the fermion gauge group indices. The value of this diagram is therefore the same as in QED, Eq. (7.90), multiplied by a trace over group matrices:



$$\begin{aligned} &= \mathrm{tr}[t^a t^b]\, i(q^2 g^{\mu\nu} - q^\mu q^\nu) \\ &\quad \times \frac{-g^2}{(4\pi)^{d/2}} \int_0^1 dx\, 8x(1-x) \frac{\Gamma(2-\frac{d}{2})}{(m^2 - x(1-x)q^2)^{2-d/2}}. \end{aligned}$$

The value of the trace is given by Eq. (15.78): $\mathrm{tr}[t^a t^b] = C(r)\delta^{ab}$. In a theory with several species of fermions, there would be a diagram of this type for each species. We will be mainly interested in the divergent part of this diagram, which is independent of the fermion mass. If there are n_f species of fermions, all in the same representation r, then the total contribution of fermion loop diagrams takes the form

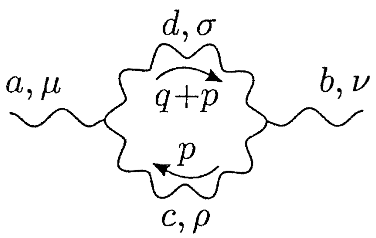
$$\sum_{\text{fermions}} \Big(\quad \Big)$$



$$= i(q^2 g^{\mu\nu} - q^\mu q^\nu)\delta^{ab}\Big(\frac{-g^2}{(4\pi)^2}\cdot\frac{4}{3}n_f C(r)\Gamma(2-\tfrac{d}{2}) + \cdots\Big). \tag{16.59}$$

Now consider the three diagrams from the pure gauge sector. The contribution of these diagrams depends on the gauge; we will use Feynman-'t Hooft gauge, $\xi = 1$.

Using the three-gauge-boson vertex from Fig. 16.1, we can write the first of the three diagrams as



$$= \frac{1}{2}\int \frac{d^4p}{(2\pi)^4}\frac{-i}{p^2}\frac{-i}{(p+q)^2}\, g^2 f^{acd} f^{bcd} N^{\mu\nu}, \tag{16.60}$$

where the numerator structure is

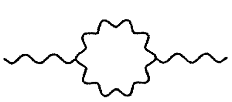
$$\begin{aligned} N^{\mu\nu} = &\big[g^{\mu\rho}(q-p)^\sigma + g^{\rho\sigma}(2p+q)^\mu + g^{\sigma\mu}(-p-2q)^\rho\big] \\ &\times \big[\delta^\nu{}_\rho (p-q)_\sigma + g_{\rho\sigma}(-2p-q)^\nu + \delta^\nu{}_\sigma (p+2q)_\rho\big]. \end{aligned}$$

The overall factor of 1/2 is a symmetry factor. The contraction of structure constants can be evaluated using Eq. (15.93): $f^{acd}f^{bcd} = C_2(G)\delta^{ab}$.

To simplify the expression further, combine denominators in the standard way:

$$\frac{1}{p^2}\frac{1}{(p+q)^2} = \int_0^1 dx\, \frac{1}{((1-x)p^2 + x(p+q)^2)^2} = \int_0^1 dx\, \frac{1}{(P^2-\Delta)^2}, \tag{16.61}$$

where $P = p + xq$ and $\Delta = -x(1-x)q^2$. Then (16.60) can be rewritten

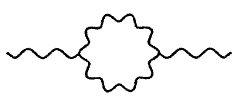


$$= -\frac{g^2}{2}C_2(G)\delta^{ab}\int_0^1 dx \int \frac{d^4P}{(2\pi)^4}\frac{1}{(P^2-\Delta)^2}N^{\mu\nu}.$$

The numerator structure can be simplified by eliminating p in favor of P, discarding terms linear in P^μ (which integrate symmetrically to zero), and replacing $P^\mu P^\nu$ with $g^{\mu\nu}P^2/d$ (also by symmetry):

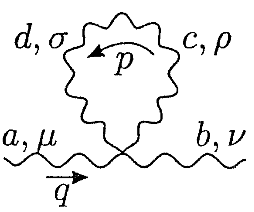
$$
\begin{aligned}
N^{\mu\nu} = & -g^{\mu\nu}\left[(2q+p)^2 + (q-p)^2\right] - d(q+2p)^\mu(q+2p)^\nu \\
& + \left[(2q+p)^\mu(q+2p)^\nu + (q-p)^\mu(2q+p)^\nu - (q+2p)^\mu(q-p)^\nu \right.\\
& \left. + (\mu \leftrightarrow \nu)\right] \\
\to & -g^{\mu\nu}P^2 \cdot 6(1-\tfrac{1}{d}) - g^{\mu\nu}q^2\left[(2-x)^2 + (1+x)^2\right] \\
& + q^\mu q^\nu\left[(2-d)(1-2x)^2 + 2(1+x)(2-x)\right].
\end{aligned}
$$

The final step in the evaluation is to Wick-rotate and apply the integration formulae (7.85) and (7.86). This brings the diagram into the following form:



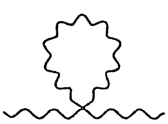
$$
\begin{aligned}
= & \frac{ig^2}{(4\pi)^{d/2}} C_2(G)\delta^{ab} \int_0^1 dx \, \frac{1}{\Delta^{2-d/2}} \\
& \times \Big(\Gamma(1-\tfrac{d}{2})\, g^{\mu\nu}q^2 \left[\tfrac{3}{2}(d-1)x(1-x)\right] \\
& \quad + \Gamma(2-\tfrac{d}{2})\, g^{\mu\nu}q^2 \left[\tfrac{1}{2}(2-x)^2 + \tfrac{1}{2}(1+x)^2\right] \\
& \quad - \Gamma(2-\tfrac{d}{2})\, q^\mu q^\nu \left[(1-\tfrac{d}{2})(1-2x)^2 + (1+x)(2-x)\right]\Big).
\end{aligned}
\tag{16.62}
$$

Next consider the diagram with a four-gauge-boson vertex. Using the vertex Feynman rule in Fig. 16.1, we find



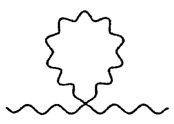
$$
\begin{aligned}
= & \frac{1}{2}\int \frac{d^4p}{(2\pi)^4} \frac{-ig_{\rho\sigma}}{p^2}\delta^{cd}\,(-ig^2) \\
& \times \left[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho})\right. \\
& \quad + f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) \\
& \quad \left. + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})\right].
\end{aligned}
\tag{16.63}
$$

The factor 1/2 in the first line is a symmetry factor. The first combination of structure constants in the vertex factor vanishes by antisymmetry; the second and third can be reduced by the use of Eq. (15.93). We then find simply



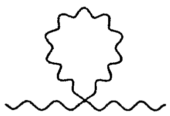
$$
= -g^2 C_2(G)\delta^{ab}\int \frac{d^4p}{(2\pi)^4}\frac{1}{p^2}\cdot g^{\mu\nu}(d-1). \tag{16.64}
$$

In dimensional regularization, the integral over p gives a pole at $d=2$ but yields zero as $d \to 4$. We could simply discard this diagram and trust that the pole at $d=2$ is canceled by the other two diagrams. It is instructive, however, and no more difficult, to demonstrate the cancellation explicitly. To do so, we can force the integral to look like that of the previous diagram, multiplying the integrand by 1 in the form $(q+p)^2/(q+p)^2$. We then combine denominators as before, and eliminate p in favor of the shifted variable $P = p + xq$. After dropping the term linear in P, we obtain



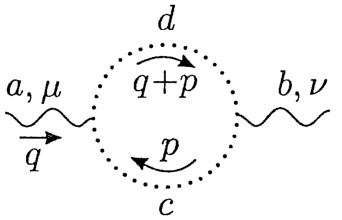
$$= -g^2 C_2(G)\delta^{ab}\int_0^1 dx \int \frac{d^4P}{(2\pi)^4}\frac{1}{(P^2-\Delta)^2} g^{\mu\nu}(d-1)[P^2+(1-x)^2q^2].$$

We can now Wick-rotate and integrate over P to obtain



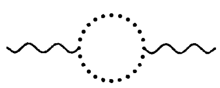
$$\begin{aligned} &= \frac{ig^2}{(4\pi)^{d/2}}C_2(G)\delta^{ab}\int_0^1 dx\,\frac{1}{\Delta^{2-d/2}} \\ &\quad\times\Big(-\Gamma(1-\tfrac{d}{2})\,g^{\mu\nu}q^2\left[\tfrac{1}{2}d(d-1)x(1-x)\right] \\ &\qquad -\Gamma(2-\tfrac{d}{2})\,g^{\mu\nu}q^2\left[(d-1)(1-x)^2\right]\Big). \end{aligned} \tag{16.65}$$

Expressions (16.62) and (16.65), by themselves, do not add to any reasonable value: The pole at $d=2$ does not cancel, and the sum does not have a transverse Lorentz structure. To bring the gauge boson self-energy into its desired form, we must include the diagram with a ghost loop. According to the rules shown in Fig. 16.5, this diagram is



$$= (-1)\int \frac{d^4p}{(2\pi)^4}\frac{i}{p^2}\frac{i}{(p+q)^2}\, g^2 f^{dac}(p+q)^\mu f^{cbd}p^\nu. \tag{16.66}$$

There is no symmetry factor in this case, but there is a factor of -1 because the ghost fields anticommute. The ghost diagram can be simplified using the same set of tricks that we applied to the previous two: combine denominators, shift the integral to P, Wick-rotate, and integrate over P using dimensional regularization. The result is



$$\begin{aligned} &= \frac{ig^2}{(4\pi)^{d/2}}C_2(G)\delta^{ab}\int_0^1 dx\,\frac{1}{\Delta^{2-d/2}} \\ &\quad\times\Big(-\Gamma(1-\tfrac{d}{2})\,g^{\mu\nu}q^2\left[\tfrac{1}{2}x(1-x)\right] \\ &\qquad +\Gamma(2-\tfrac{d}{2})\,q^\mu q^\nu\left[x(1-x)\right]\Big). \end{aligned} \tag{16.67}$$

Now we are ready to put these results together. In the sum of the three diagrams, the coefficient of $\Gamma(1-\frac{d}{2})g^{\mu\nu}q^2x(1-x)$ is

$$\tfrac{1}{2}(3d-3-d^2+d-1) = (1-\tfrac{d}{2})(2-d). \tag{16.68}$$

The first factor cancels the pole of the gamma function at $d=2$. Thus, the sum of the three diagrams has no quadratic divergence and no gauge boson mass renormalization. Notice that the ghost diagram plays an essential role in this cancellation.

After the pole at $d = 2$ is canceled, $\Gamma(1-\frac{d}{2})$ becomes $\Gamma(2-\frac{d}{2})$. This term therefore combines with the others that are proportional to $\Gamma(2-\frac{d}{2})g^{\mu\nu}q^2$, to give a total coefficient of

$$(d-2)x(1-x) + \tfrac{1}{2}(2-x)^2 + \tfrac{1}{2}(1+x)^2 - (d-1)(1-x)^2. \tag{16.69}$$

Since the best way to simplify this expression is not obvious, let us put it aside and work first with the coefficient of $\Gamma(2-\frac{d}{2})q^\mu q^\nu$:

$$-(1-\tfrac{d}{2})(1-2x)^2 - (1+x)(2-x) + x(1-x) = -(1-\tfrac{d}{2})(1-2x)^2 - 2.$$

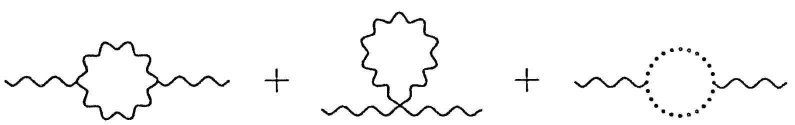
If the total self-energy is to be proportional to $(g^{\mu\nu}q^2 - q^\mu q^\nu)$, it must be possible to reduce expression (16.69) to this same form (times -1). To do so, note that Δ is symmetric with respect to $x \leftrightarrow (1-x)$, and therefore we can substitute $(1-x)$ for x in any term of the numerator. In particular, terms that are linear in x can be transformed as follows:

$$x \to \tfrac{1}{2}x + \tfrac{1}{2}(1-x) = \tfrac{1}{2}.$$

In the end, the sum of the three pure-gauge diagrams simplifies to

$$\frac{ig^2}{(4\pi)^{d/2}}C_2(G)\delta^{ab}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}(g^{\mu\nu}q^2 - q^\mu q^\nu)\big[(1-\tfrac{d}{2})(1-2x)^2 + 2\big]. \tag{16.70}$$

This expression is manifestly transverse, as required by the Ward identity of the non-Abelian gauge theory. For future reference, we record the ultraviolet divergent part of (16.70):



$$= i(q^2g^{\mu\nu} - q^\mu q^\nu)\delta^{ab}\Big(\frac{-g^2}{(4\pi)^2}\cdot\Big(-\frac{5}{3}\Big)C_2(G)\Gamma(2-\tfrac{d}{2}) + \cdots\Big). \tag{16.71}$$

As we noted above, the result (16.70) depends on the gauge used in the calculation. In any gauge, the boson self-energy is transverse and free of quadratic divergences. However, the coefficient of the transverse Lorentz structure may depend on ξ. It turns out that, for a general value of ξ, the coefficient of the ultraviolet divergence in (16.71) is modified according to

$$-\frac{5}{3} \to -\Big(\frac{13}{6} - \frac{\xi}{2}\Big). \tag{16.72}$$

The fact that the boson self-energy depends on the gauge does not contradict the general theorem that S-matrix elements are independent of ξ. The full set of one-loop corrections to a gauge theory S-matrix element always involves a number of different radiative corrections to vertices and propagators; the gauge dependence cancels in an intricate fashion among these various terms.

The β Function

The simplest calculation that involves a gauge-invariant combination of radiative corrections is the computation of the leading term of the Callan-Symanzik β function of a non-Abelian gauge theory. The invariance of the leading term of β could be argued intuitively, by saying that the coupling constant of the gauge theory should not evolve to large values in one scheme of calculation while it stays small in another scheme. In Section 17.2 we will demonstrate this result more cleanly by showing that the leading coefficient of the β function can be extracted from a physical cross section and so must be gauge independent. (Surprisingly, this conclusion actually applies to the first *two* coefficients of the β function, written as a power series in g.)

Recall from Section 12.2 that the β function gives the rate at which the renormalized coupling constant changes as the renormalization scale M is increased. Since Green's functions depend on M through the counterterms that subtract ultraviolet divergences, β can be computed from the counterterms that enter an appropriately chosen Green's function. For example, in Eq. (12.58), we saw that the β function of QED can be computed from the counterterms for the electron-photon vertex, the electron self-energy, and the photon self-energy. The same derivation goes through in the case of a non-Abelian gauge theory. Thus, to lowest order,

$$\beta(g) = gM\frac{\partial}{\partial M}\left(-\delta_1 + \delta_2 + \tfrac{1}{2}\delta_3\right), \tag{16.73}$$

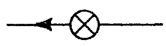
with the conventions for the counterterm vertices shown in Fig. 16.8. In QED, the first two terms cancel by the Ward identity, so β depends only on δ_3. In the non-Abelian case, all three terms contribute. The most difficult to compute is δ_3, but we have nearly done so already by computing the gauge-boson self-energy diagrams. Let us now complete this calculation of the β function of non-Abelian gauge theory.

In order for the counterterm δ_3 to cancel the divergence of Eqs. (16.59) and (16.71), it must be of the form

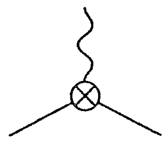
$$\delta_3 = \frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}\left[\frac{5}{3}C_2(G) - \frac{4}{3}n_f C(r)\right], \tag{16.74}$$

where M is the renormalization scale. Depending on the precise renormalization conditions used, there may be additional finite contributions to δ_3, but these do not contribute to the β function (to one-loop order). Similarly, the finite parts of δ_2 and δ_1 will depend on the details of the renormalization scheme. However, as we saw in Section 12.2, the one-loop contribution to the β function is the same in any scheme in which amplitudes are renormalized at a point where all momentum invariants are of the same order M^2. In dimensional regularization, a logarithmic divergence always takes the form $\Gamma(2-\frac{d}{2})/\Delta^{2-d/2}$, where Δ is some combination of momentum invariants. Thus, to compute the β function, we can simply set $\Delta = M^2$ in such expressions.

$$= -i\big(k^2 g^{\mu\nu} - k^\mu k^\nu\big)\delta^{ab}\delta_3$$



$$= i\not{p}\delta_2$$



$$= igt^a\gamma^\mu\delta_1$$

Figure 16.8. Counterterms needed for computing fermion interactions in a non-Abelian gauge theory.

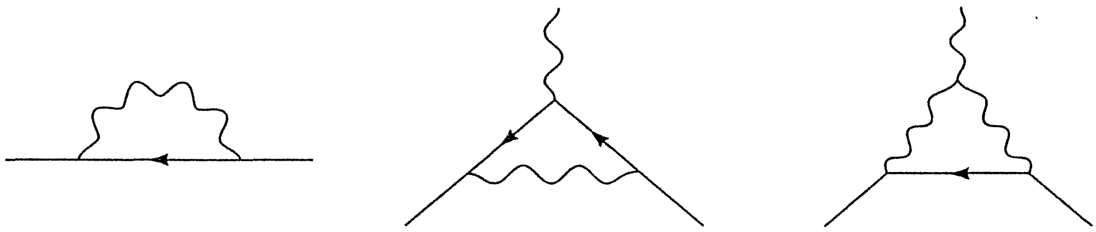
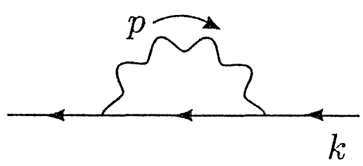


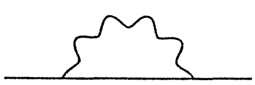
Figure 16.9. Diagrams whose divergences are subtracted by the counterterms δ_2 and δ_1.

To complete the computation of the β function, we must compute δ_2 and δ_1 to the same level of approximation. The fermion self-energy counterterm δ_2 cancels the divergence proportional to $\not{k}$ in the first diagram of Fig. 16.9. In Feynman-'t Hooft gauge, the value of this diagram is



$$= \int \frac{d^4p}{(2\pi)^4}(ig)^2\gamma^\mu t^a \frac{i(\not{p}+\not{k})}{(p+k)^2}\gamma_\mu t^a \frac{-i}{p^2}. \tag{16.75}$$

Since the divergence in the field strength renormalization is independent of the fermion mass, we have simplified (16.75) by setting the mass to zero. The product of group matrices equals the quadratic Casimir operator, by definition (15.92). The Dirac matrix structure can be reduced using a contraction identity (7.89). The rest of the calculation follows the same steps as for the boson self-energy diagrams:



$$\begin{aligned} &= g^2 C_2(r)(d-2)\int \frac{d^4p}{(2\pi)^4}\frac{(\not{p}+\not{k})}{(p+k)^2p^2} \\ &= g^2 C_2(r)(d-2)\int_0^1 dx \int \frac{d^4P}{(2\pi)^4}\frac{(1-x)\not{k}}{(P^2-\Delta)^2} \\ &= \frac{ig^2}{(4\pi)^{d/2}}C_2(r)\not{k}\int_0^1 dx\,(1-x)(d-2)\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}} \end{aligned}$$

$$= \frac{ig^2}{(4\pi)^2} \not{k} C_2(r)\Gamma(2-\tfrac{d}{2}) + \cdots . \tag{16.76}$$

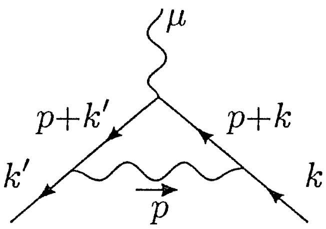
(Here $P = p + xk$ and $\Delta = -x(1-x)k^2$.)

The divergent part of this expression must be canceled by the second counterterm diagram of Fig. 16.8. Thus, if the renormalization scale is M, the counterterm must be

$$\delta_2 = -\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}} \cdot C_2(r), \tag{16.77}$$

plus finite terms. We note that, like δ_3, δ_2 depends on the gauge; for example, δ_2 has no one-loop divergence in Landau gauge ($\xi = 0$).

To determine δ_1, we must compute the second and third diagrams of Fig. 16.9. The second diagram, computed in Feynman-'t Hooft gauge and for massless fermions, is



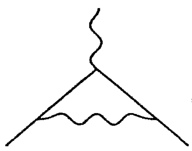
$$= \int \frac{d^4p}{(2\pi)^4} g^3\, t^b t^a t^b\, \frac{\gamma^\nu(\not{p} + \not{k}')\gamma^\mu(\not{p} + \not{k})\gamma_\nu}{(p+k')^2(p+k)^2p^2}. \tag{16.78}$$

The gauge group matrices can be simplified according to

$$\begin{aligned} t^b t^a t^b &= t^b t^b t^a + t^b[t^a, t^b] \\ &= C_2(r)t^a + it^b f^{abc} t^c \\ &= C_2(r)t^a + \tfrac{1}{2} i f^{abc} \cdot i f^{bcd} t^d \\ &= [C_2(r) - \tfrac{1}{2}C_2(G)]t^a. \end{aligned} \tag{16.79}$$

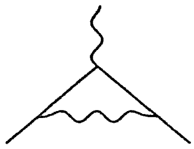
In the third line we have used the antisymmetry of f^{abc} to rewrite the matrix product as a commutator; in the last line we have used Eq. (15.93).

The diagrams computed earlier in this section had positive superficial degrees of divergence, so we needed to extract their logarithmic divergences carefully. The integral in (16.78), however, is superficially logarithmically divergent, and so the coefficient of this divergence can be extracted easily by considering the limit in which the integration variable p is much greater than any external momentum. In this limit, the diagram is estimated as follows:



$$\sim g^3[C_2(r) - \tfrac{1}{2}C_2(G)]t^a \int \frac{d^4p}{(2\pi)^4}\frac{\gamma^\nu \not{p}\gamma^\mu\not{p}\gamma_\nu}{p^2\cdot p^2\cdot p^2}. \tag{16.80}$$

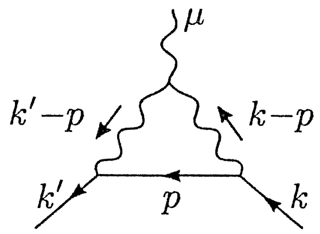
If we replace $p^\rho p^\sigma$ by $g^{\rho\sigma}p^2/d$ in the numerator of (16.80), this expression simplifies easily:



$$\sim g^3[C_2(r)-\tfrac{1}{2}C_2(G)]t^a(2-d)^2\frac{1}{d}\gamma^\mu\int\frac{d^4p}{(2\pi)^4}\frac{1}{(p^2)^2}$$
$$\sim \frac{ig^3}{(4\pi)^2}[C_2(r)-\tfrac{1}{2}C_2(G)]t^a\gamma^\mu\big(\Gamma(2-\tfrac{d}{2})+\cdots\big). \tag{16.81}$$

This estimate gives the correct coefficient of the divergent term. It drops completely the finite terms in the vertex function, but we do not need these to compute the β function.

The third diagram of Fig. 16.9 can be analyzed in the same way. Its value, in Feynman-'t Hooft gauge and for massless fermions, is

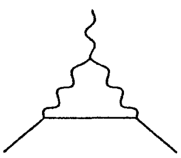


$$= \int\frac{d^4p}{(2\pi)^4}(ig\gamma_\nu t^b)\frac{i\not{p}}{p^2}(ig\gamma_\rho t^c)\frac{-i}{(k'-p)^2}\frac{-i}{(k-p)^2}$$
$$\times\, gf^{abc}\big[g^{\mu\nu}(q+k'-p)^\rho+g^{\nu\rho}(-k'-k+2p)^\mu$$
$$+\,g^{\rho\mu}(k-p-q)^\nu\big]. \tag{16.82}$$

The gauge matrix product can be reduced as follows:

$$f^{abc}t^bt^c=\frac{1}{2}f^{abc}\cdot if^{bcd}t^d=\frac{i}{2}C_2(G)t^a.$$

Again we can determine the logarithmic divergence of this diagram by neglecting all external momenta in comparison with p. A straightforward calculation then yields



$$\sim\frac{g^3}{2}C_2(G)t^a\int\frac{d^4p}{(2\pi)^4}\gamma_\nu\not{p}\gamma_\rho\frac{g^{\mu\nu}p^\rho-2g^{\nu\rho}p^\mu+g^{\rho\mu}p^\nu}{(p^2)^3}$$
$$\sim\frac{g^3}{2}C_2(G)t^a\frac{1}{d}\int\frac{d^4p}{(2\pi)^4}\frac{1}{(p^2)^2}\big[\gamma^\mu\gamma^\rho\gamma_\rho-2\gamma^\rho\gamma^\mu\gamma_\rho+\gamma^\sigma\gamma_\sigma\gamma^\mu\big]$$
$$\sim\frac{ig^3}{(4\pi)^2}\frac{3}{2}C_2(G)\,t^a\gamma^\mu\big(\Gamma(2-\tfrac{d}{2})+\cdots\big). \tag{16.83}$$

In the second line we have again replaced $p^\rho p^\sigma$ with $g^{\rho\sigma}p^2/d$.

The sum of the divergences in results (16.81) and (16.83) must be canceled by the third counterterm diagram in Fig 16.8. With a renormalization scale of M, we find

$$\delta_1=-\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}\big[C_2(r)+C_2(G)\big]. \tag{16.84}$$

Notice that δ_1 is not equal to δ_2, as would have been true in the Abelian case; here δ_1 has an extra term, proportional to $C_2(G)$.

We are now ready to compute the β function. Plugging the three counterterms (16.74), (16.77), and (16.84) into our formula (16.73), we find

$$\beta(g) = (-2)\frac{g^3}{(4\pi)^2}\Big[\big(C_2(r)+C_2(G)\big) - C_2(r) + \frac{1}{2}\Big(\frac{5}{3}C_2(G) - \frac{4}{3}n_f C(r)\Big)\Big];$$

that is,

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\Big[\frac{11}{3}C_2(G) - \frac{4}{3}n_f C(r)\Big]. \tag{16.85}$$

Notice that, at least for small values of n_f, the β function is *negative* and so non-Abelian gauge theories are asymptotically free. This is a result of exceptional physical importance, first discovered by 't Hooft, Politzer, and Gross and Wilczek.* We will discuss the physical interpretation of this result further in Section 16.7, and in the next several chapters. However, for the rest of this section, we will resist the temptation to pursue the physics and instead complete our technical analysis of the divergences of non-Abelian gauge theories.

Relations among Counterterms

In the analysis just completed, we computed the β function of a non-Abelian gauge theory from the divergences of the fermion vertex and field strength renormalizations. One might visualize that we were computing the running of the coupling constant at the fermion-gauge boson vertex. Alternatively, we could have studied the divergences of the three-gauge-boson vertex or the four-gauge-boson vertex, and thus computed the running of these coupling constants. However, we saw already in Section 16.1 that non-Abelian gauge invariance knits together these separate coupling constants and requires their equality. Thus we might expect that these different calculations should produce the same value of the β function.

To clarify this issue, let us carefully enumerate all the counterterms that appear in a non-Abelian gauge theory. We start from the Lagrangian (16.34), regarded as a combination of bare fields and a bare coupling constant. In the following discussion, we denote bare quantities by the subscript 0. Then,

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}(\partial_\mu A^a_{0\nu} - \partial_\nu A^a_{0\mu})^2 + \bar\psi_0(i\partial\!\!\!/ - m_0)\psi_b - \bar c^a_0 \partial^2 c^a_0 \\ &+ g_0 A^a_{0\mu}\bar\psi_0\gamma^\mu\psi_0 - g_0 f^{abc}(\partial_\mu A^a_{0\nu})A^b_{0\mu}A^c_{0\nu} \\ &- \; g^2(f^{eab}A^a_{0\mu}A^b_{0\nu})(f^{ecd}A^c_{0\mu}A^d_{0\nu}) - g\bar c^a_0 f^{abc}\partial^\mu A^0_{b\mu}c^c_0.\end{aligned} \tag{16.86}$$

We choose $\xi = 0$ for simplicity. We now rescale the fields to the renormalized field strengths by extracting the factors Z_2, Z_3, Z^c_2 for the gauge bosons,

*G. 't Hooft, unpublished; H. D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973); D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1323 (1973).

fermions, and ghosts, and shift the coupling to the renormalized coupling g. The Lagrangian then takes the form

$$\mathcal{L} = \mathcal{L}_{\text{ren}} + \mathcal{L}_{\text{c.t.}},$$

where $\mathcal{L}_{\text{ren}}$ is the Lagrangian (16.34) and $\mathcal{L}_{\text{c.t.}}$ takes the form

$$\begin{aligned}\mathcal{L}_{\text{c.t.}} = &-\frac{1}{4}\delta_3(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)^2 + \bar{\psi}(i\delta_2 \not{\partial} - \delta_m)\psi - \delta_2^c \bar{c}^a \partial^2 c^a \\ &+ g\delta_1 A_\mu^a \bar{\psi}\gamma^\mu \psi - g\delta_1^{3g} f^{abc}(\partial_\mu A_\nu^a) A_\mu^b A_\nu^c \\ &- \ g^2\delta_1^{4g}(f^{eab}A_\mu^a A_\nu^b)(f^{ecd}A_\mu^c A_\nu^d) - g\delta_1^c \bar{c}^a f^{abc}\partial^\mu A_\mu^b c^c,\end{aligned} \tag{16.87}$$

with the counterterms defined by

$$\begin{aligned}&\delta_2 = Z_2 - 1, \qquad \delta_3 = Z_3 - 1, \qquad \delta_2^c = Z_2^c - 1, \qquad \delta_m = Z_2 m_b - m, \\ &\delta_1 = \frac{g_0}{g} Z_2 (Z_3)^{1/2} - 1, \qquad \delta_1^{3g} = \frac{g_0}{g}(Z_3)^{3/2} - 1, \\ &\delta_1^{4g} = \frac{g_0^2}{g^2}(Z_3)^2 - 1, \qquad \delta_1^c = \frac{g_0}{g} Z_2^c (Z_3)^{1/2} - 1.\end{aligned} \tag{16.88}$$

Notice that these eight counterterms depend on five underlying parameters; thus, there are three relations among them. The situation is very similar to that for the scalar theories with spontaneously broken symmetry that we studied in Chapter 11. The underlying symmetry of the theory—here, local gauge invariance—implies relations among the divergent amplitudes of the theory and among the counterterms required to cancel them. In the present case, a set of five renormalization conditions uniquely specifies all of the counterterms in a way that removes all divergences from the theory.

This program is especially simple at one-loop order. In this case we can expand g_b/g and the various Z factors about 1, keeping only the leading-order contribution to each counterterm. Then the three relations among the counterterms can be written

$$\delta_1 - \delta_2 = \delta_1^{3g} - \delta_3 = \tfrac{1}{2}(\delta_1^{4g} - \delta_3) = \delta_1^c - \delta_2^c. \tag{16.89}$$

It is instructive to check explicitly that the values of δ_1^{3g}, δ_1^{4g}, and δ_1^c determined from (16.89) indeed remove the divergences of the corresponding vertex diagrams; this is the subject of Problem 16.1. Using relations (16.89), it is easy to show that the one-loop calculation of the β function will yield the same value, whichever gauge boson vertex is used in the computation. More generally, consider a non-Abelian gauge theory with many different species of particles, bosons and fermions, which couple to the gauge field. Then, to one-loop order, the quantity

$$\delta_1^i - \delta_2^i,$$

where δ_1^i is the vertex counterterm for species i and δ_2^i is the corresponding field strength counterterm, takes a universal value. This value is gauge dependent, so that the gauge dependence of its divergent part cancels the gauge dependence of δ_3 in the computation of the β function.

In our discussion of the counterterms of QED at the end of Section 10.3, we remarked that the relation between δ_1 and δ_2 insured that all electrically charged species see a common universal value of the coupling constant e. In non-Abelian gauge theories, the relations (16.89) and their higher-loop generalizations preserve the universality of the non-Abelian couplings. In QED, we were able to obtain an even stronger relation, $\delta_1 = \delta_2$ or $Z_1 = Z_2$, from the absolute normalization of the matrix elements of the vector current. However, in non-Abelian gauge theories, the corresponding vector current $j^{\mu a} = \bar{\psi}\gamma^\mu t^a \psi$ transforms under local gauge transformations in the adjoint representation. Thus the Faddeev-Popov prescription cannot be used to compute matrix elements of this current unambiguously, and thus the normalization of these matrix elements is not preserved by the perturbation theory.

16.6 Asymptotic Freedom: The Background Field Method

In the previous section, we saw that the β function of a non-Abelian gauge theory with a sufficiently small number of fermions is negative. This result is important enough that it is worthwhile to derive it twice. The preceding derivation was straightforward but not very illuminating. In this section we give a second derivation of the same result, which is more abstract but much cleaner and more transparent.

The method of this section reflects the spirit of Wilson's idea of integrating out the high-momentum degrees of freedom, while taking proper care to preserve gauge invariance. We will compute the effective action of a non-Abelian gauge theory for a fixed, slowly varying, classical background gauge field $A_\mu^a(x)$. By adopting a canonical normalization of this field, we can interpret the coefficient of the effective action as a running coupling constant. This method is analogous to Polyakov's method for computing the β function of the nonlinear sigma model, presented in Section 13.3.

Background Field Perturbation Theory

To set up the computation, rescale the gauge field $gA_\mu^a \rightarrow A_\mu^a$. In this normalization, the gauge coupling is removed from the covariant derivative and moved to the coefficient of the gauge field kinetic energy term. We thus start from the Lagrangian

$$\mathcal{L} = \frac{1}{4g^2}(F_{\mu\nu}^a)^2 + \bar{\psi}(i\not{D})\psi, \tag{16.90}$$

with

$$\begin{aligned} D_\mu &= \partial_\mu - iA^a_\mu t^a, \\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A^b_\mu A^c_\nu, \end{aligned} \tag{16.91}$$

and the fermion mass set to zero for simplicity. The transformation laws of A^a_μ and ψ are also independent of the coupling constant:

$$\delta A^a_\mu = \partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c, \qquad \delta\psi = i\alpha^a t^a \psi. \tag{16.92}$$

On the other hand, the coupling constant g will appear in the gauge field propagator.

Next, split the gauge field into a classical background field and a fluctuating quantum field:

$$A^a_\mu \to A^a_\mu + \mathcal{A}^a_\mu. \tag{16.93}$$

We will treat the classical part A^a_μ as a fixed field configuration and the fluctuating part $\mathcal{A}^a_\mu$ as the integration variable of the functional integral. From here on, we will use the symbol D_μ to denote the covariant derivative with respect to the background field: $D_\mu = \partial_\mu - iA^a_\mu t^a$. Then

$$\bar\psi(i\not{D})\psi \to \bar\psi(i\not{D})\psi + \mathcal{A}^a_\mu \bar\psi \gamma^\mu t^a \psi. \tag{16.94}$$

The Yang-Mills field strength decomposes as follows:

$$\begin{aligned} F^a_{\mu\nu} \to\ & \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A^b_\mu A^c_\nu \\ & + \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + f^{abc}(A^b_\mu \mathcal{A}^c_\nu - A^b_\nu \mathcal{A}^c_\mu) + f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu \\ =\ & F^a_{\mu\nu} + D_\mu \mathcal{A}^a_\nu - D_\nu \mathcal{A}^a_\mu + f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu, \end{aligned} \tag{16.95}$$

where, in the last line, $F^a_{\mu\nu}$ is the field strength of the classical field, and D_μ is the covariant derivative in the adjoint representation, Eq. (15.86). Notice that, both in (16.94) and in (16.95), the derivative ∂_μ appears only as a part of the covariant derivative with respect to the background field.

If the background field A^a_μ is regarded as fixed, the Lagrangian has a local gauge symmetry implemented by transformations on $\mathcal{A}^a_\mu$:

$$\mathcal{A}^a_\mu \to \mathcal{A}^a_\mu + D_\mu \alpha^a + f^{abc} \mathcal{A}^b_\mu \alpha^c. \tag{16.96}$$

To define the functional integral, we must gauge-fix using the Faddeev-Popov procedure. We choose a gauge-fixing condition that is covariant with respect to the background gauge field:

$$G(A) = D^\mu \mathcal{A}^a_\mu - \omega^a, \tag{16.97}$$

instead of (16.28). The Faddeev-Popov determinant involves the variation of this operator with respect to the gauge transformation (16.96). As in Section 16.2, we can promote the gauge-fixing term to the exponent, to quantize the theory in the background field analogue of Feynman-'t Hooft gauge. Then the gauge-fixed Lagrangian is

$$\begin{aligned}\mathcal{L}_{\mathrm{FP}} = &-\frac{1}{4g^2}\left(F^a_{\mu\nu} + D_\mu\mathcal{A}^a_\nu - D_\nu\mathcal{A}^a_\mu + f^{abc}\mathcal{A}^b_\mu\mathcal{A}^c_\mu\right)^2 - \frac{1}{2g^2}(D^\mu\mathcal{A}^a_\mu)^2 \\ &+ \bar{\psi}(i\not{D} + \mathcal{A}^a_\mu\gamma^\mu t^a)\psi + \bar{c}^a(-D^2 - D^\mu f^{abc}\mathcal{A}^b_\mu)c^c.\end{aligned} \tag{16.98}$$

The Lagrangian (16.98) is gauge-fixed, but it is invariant under a local symmetry that transforms both $\mathcal{A}^a_\mu$ and the background field A^a_μ:

$$\begin{aligned} A^a_\mu &\to A^a_\mu + D_\mu\beta^a \\ \mathcal{A}^a_\mu &\to \mathcal{A}^a_\mu - f^{abc}\beta^b\mathcal{A}^c_\mu \\ \psi &\to \psi + i\beta^a t^a\psi \\ c^a &\to c^a - f^{abc}\beta^b c^c. \end{aligned} \tag{16.99}$$

Under this transformation, $\mathcal{A}^a_\mu$ transforms as a matter field in the adjoint representation, while A^a_μ carries the part of the local gauge transformation proportional to $\partial_\mu\beta^a$. To prove that (16.99) is a symmetry of (16.98), we need only note that (16.98) is globally invariant, and that A^a_μ appears in (16.98) only as a part of the covariant derivative and the field strength. The transformation (16.98) is also a symmetry of the functional measure. Thus, if we functionally integrate over $\mathcal{A}^a_\mu$, ψ, and c^a to compute the effective action, the result must be invariant to local gauge transformations of A^a_μ. This observation greatly simplifies the analysis of the effective action.

One-Loop Correction to the Effective Action

Let us now compute the effective action, using the method of Section 11.4. To compute $\Gamma[A^a_\mu]$ to one-loop order, we drop terms linear in the fluctuating field $\mathcal{A}^a_\mu$ and then integrate over the terms quadratic in $\mathcal{A}^a_\mu$ and the fermion and ghost fields. This produces functional determinants, which we can evaluate into an appropriate form for an effective action.

To carry out this program, we must work out the terms in (16.98) quadratic in each of the various fields. The terms quadratic in $\mathcal{A}^a_\mu$ are:

$$\mathcal{L}_{\mathcal{A}} = -\frac{1}{2g^2}\left\{\tfrac{1}{2}(D_\mu\mathcal{A}^a_\nu - D_\nu\mathcal{A}^a_\mu)^2 + F^{a\mu\nu}f^{abc}\mathcal{A}^b_\mu\mathcal{A}^b_\nu + (D^\mu\mathcal{A}^a_\mu)^2\right\}. \tag{16.100}$$

After integrating by parts, we can rewrite this as

$$\mathcal{L}_{\mathcal{A}} = -\frac{1}{2g^2}\left\{\mathcal{A}^a_\mu[-(D^2)^{ab}g^{\mu\nu} + (D^\nu D^\mu)^{ab} - (D^\mu D^\nu)^{ab}]\mathcal{A}^b - \mathcal{A}^a f^{abc}F^{b\mu\nu}\mathcal{A}^c\right\}. \tag{16.101}$$

The term in brackets contains the commutator of covariant derivatives. This can be simplified using (15.48); the result combines with the last term to give

$$\mathcal{L}_{\mathcal{A}} = -\frac{1}{2g^2}\left\{\mathcal{A}^a_\mu[-(D^2)^{ac}g^{\mu\nu} - 2f^{abc}F^{b\mu\nu}]\mathcal{A}^c\right\}. \tag{16.102}$$

The first term is part of a covariant d'Alembertian operator. The second term seems quite special, but we can put it into a form that will be convenient later as follows: First, we recognize that $F^b_{\mu\nu}$ is contracted with a group generator in the adjoint representation. Next, we introduce the matrix (3.18) that is the generator of Lorentz transformations on 4-vectors:

$$(\mathcal{J}^{\rho\sigma})_{\alpha\beta} = i(\delta^\rho_\alpha \delta^\sigma_\beta - \delta^\sigma_\alpha \delta^\rho_\beta). \tag{16.103}$$

With these replacements, we can write (16.102) in the form

$$\mathcal{L}_\mathcal{A} = -\frac{1}{2g^2}\{\mathcal{A}^a_\mu[-(D^2)^{ac}g^{\mu\nu} + 2(\tfrac{1}{2}F^b_{\rho\sigma}\mathcal{J}^{\rho\sigma})^{\mu\nu}(t^b_G)^{ac}]\mathcal{A}^c\}. \tag{16.104}$$

The object in brackets can be considered as a generalized d'Alembertian for fluctuations on the background field.

Next, we reduce the quadratic terms in fermion fields in a similar way. The quadratic Lagrangian for the fermion field is

$$\mathcal{L}_\psi = \bar{\psi}(i\not{D})\psi. \tag{16.105}$$

Integrating over the fermion fields, we find the determinant of the operator $(i\not{D})$. This is conveniently expressed as the square root of the determinant of the operator

$$\begin{aligned}(i\not{D})^2 &= -\gamma^\mu\gamma^\nu D_\mu D_\nu \\ &= \left(-\tfrac{1}{2}\{\gamma^\mu,\gamma^\nu\} - \tfrac{1}{2}[\gamma^\mu,\gamma^\nu]\right)D_\mu D_\nu \\ &= -D^2 + 2i\left(\frac{i}{4}[\gamma^\mu,\gamma^\nu]\right)D_\mu D_\nu.\end{aligned} \tag{16.106}$$

In the last line, the commutator of Dirac matrices forms the generator of Lorentz transformations in the spinor representation, $S^{\mu\nu}$ (3.23). Since this object is antisymmetric in its indices, the product $D_\mu D_\nu$ that is contracted with it can be replaced by half of their commutator. Then (16.106) takes the form

$$(i\not{D})^2 = -D^2 + 2(\tfrac{1}{2}F^b_{\rho\sigma}S^{\rho\sigma})t^b, \tag{16.107}$$

where t^a is now given in the representation of the fermions. This is just the d'Alembertian in (16.104), rewritten for the new set of spin and gauge quantum numbers. If the theory contains n_f species of fermions, the fermionic functional integral gives the determinant of (16.107) raised to the power $n_f/2$.

The quadratic term in ghosts is simply

$$\mathcal{L}_c = \bar{c}^a[-(D^2)^{ab}]c^b; \tag{16.108}$$

This contains the same d'Alembertian operator written for the case of spin zero.

To summarize all of these results, we define the general covariant background-field d'Alembertian as

$$\Delta_{r,j} = -D^2 + 2(\tfrac{1}{2}F^b_{\rho\sigma}\mathcal{J}^{\rho\sigma}), \tag{16.109}$$

acting on a field of representation r and spin j. The square of the covariant derivative gives the normal, convective, minimal coupling of the particle described by $\Delta_{r,j}$ to the gauge field. The additional term is a magnetic moment interaction with the gauge field, whose strength corresponds to a g-factor $g = 2$. Using this general expression, we can write the effective action for the classical fields A^a_μ, to one-loop order, as

$$\begin{aligned} e^{i\Gamma[A]} &= \int \mathcal{D}A\mathcal{D}\psi\mathcal{D}c \exp\Big[i\int d^4x\,(\mathcal{L}_{\text{FP}} + \mathcal{L}_{\text{c.t.}})\Big] \\ &= \exp\Big[i\int d^4x\big(-\frac{1}{4g^2}(F^a_{\mu\nu})^2 + \mathcal{L}_{\text{c.t.}}\big)\Big] \\ &\qquad \cdot \big(\det\Delta_{G,1}\big)^{-1/2}\big(\det\Delta_{r,1/2}\big)^{+n_f/2}\,\big(\det\Delta_{G,0}\big)^{+1}, \end{aligned} \tag{16.110}$$

where $\mathcal{L}_{\text{c.t.}}$ is the counterterm Lagrangian and the three determinants are the results of evaluating the gauge field, fermion, and ghost functional integrals. Additional loop corrections to the effective action are suppressed by another factor of g^2.

Since each integral contributing to (16.110) is invariant to (16.99), each determinant will be a gauge-invariant functional of A^a_μ. If we expand the determinants in powers of the background field, we should then find a series of terms that begins

$$\log\det\Delta_{r,j} = i\int d^4x\,\Big(\frac{1}{4}\mathbf{C}_{r,j}(F^a_{\mu\nu})^2 + \cdots\Big), \tag{16.111}$$

where the succeeding terms contain higher-dimension gauge-invariant operators. The coefficient $\mathbf{C}_{r,j}$ can depend on the representation r and the spin j. This first term of the expansion modifies the zeroth-order effective action according to

$$\frac{1}{4g^2}(F^a_{\mu\nu})^2 \to \frac{1}{4}\Big(\frac{1}{g^2} + \frac{1}{2}\mathbf{C}_{G,1} - \mathbf{C}_{G,0} - \frac{n_f}{2}\mathbf{C}_{r,1/2}\Big)(F^a_{\mu\nu})^2. \tag{16.112}$$

The factors $\mathbf{C}_{r,j}$ are dimensionless but, since they arise from a one-loop computation, we should expect that they are logarithmically divergent:

$$\mathbf{C}_{r,j} = c_{r,j}\log\frac{\Lambda^2}{k^2} + \cdots, \tag{16.113}$$

where k is a momentum characterizing the variation of the background field. The counterterm δ_3 removes the divergence; if we impose a renormalization condition at the scale M, then the addition of (16.113) and its counterterm gives the result (16.112) with the replacement

$$\mathbf{C}_{r,j} = c_{r,j}\log\frac{M^2}{k^2} + \cdots. \tag{16.114}$$

Then the original fixed coupling constant in the effective action is replaced by a running coupling constant

$$\frac{1}{g^2(k^2)} = \frac{1}{g^2} + \left(\frac{1}{2}c_{G,1} - c_{G,0} - \frac{n_f}{2}c_{r,1/2}\right)\log\frac{M^2}{k^2}, \tag{16.115}$$

or

$$g^2(k^2) = \frac{g^2}{1 - (\frac{1}{2}c_{G,1} - c_{G,0} - \frac{n_f}{2}c_{r,1/2})g^2\log k^2/M^2}. \tag{16.116}$$

By comparing this form to Eq. (12.88), we see that this running coupling constant is the solution to the renormalization group equation for the β function

$$\beta(g) = \left(\frac{1}{2}c_{G,1} - c_{G,0} - \frac{n_f}{2}c_{r,1/2}\right)g^3. \tag{16.117}$$

Thus, by calculating the $c_{r,j}$, we can directly obtain the leading coefficient of the β function.

Computation of the Functional Determinants

To compute $c_{r,j}$, we must work out the first term in the expansion of the determinant in powers of the external field. To expand the determinant, we proceed as in the example in Section 9.5. Write

$$\Delta_{r,j} = -\partial^2 + \Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})}, \tag{16.118}$$

where

$$\begin{aligned} \Delta^{(1)} &= = i[\partial^\mu A^a_\mu t^a + A^a_\mu t^a \partial^\mu] \\ \Delta^{(2)} &= g^2 A^{a\mu} t^a A^b_\mu t^b \\ \Delta^{(\mathcal{J})} &= 2(\tfrac{1}{2}F^b_{\rho\sigma}\mathcal{J}^{\rho\sigma}). \end{aligned} \tag{16.119}$$

The pieces $\Delta^{(1)}$ and $\Delta^{(\mathcal{J})}$ contain one power of the external field; $\Delta^{(2)}$ contains two powers of A^a_μ. Treating these terms as perturbations, we write

$$\begin{aligned} \log\det\Delta_{r,j} &= \log\det[-\partial^2 + (\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})})] \\ &= \log\det[-\partial^2] + \log\det[1 + (-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})})] \\ &= \log\det[-\partial^2] + \operatorname{tr}\log[1 + (-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})})] \\ &= \log\det[-\partial^2] + \operatorname{tr}[(-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})}) + \cdots]. \end{aligned} \tag{16.120}$$

The first term of the right in (16.120) is an irrelevant constant. The terms in this expansion that are linear in A^a_μ vanish by gauge invariance (or, more explicitly, because $\operatorname{tr}[t^a] = 0$). The quadratic terms in A^a_μ must organize themselves into the structure of (16.111), plus terms with higher derivatives.

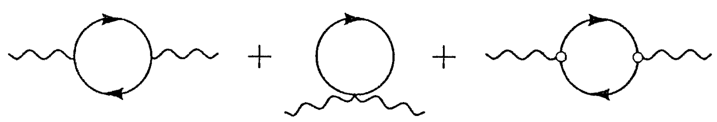


Figure 16.10. Terms quadratic in the external field in the expansion of $\log\det\Delta_{r,j}$. The special vertex arises from the $F^{\rho\sigma}\mathcal{J}_{\rho\sigma}$ coupling.

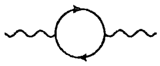
The terms in (16.111) quadratic in A^a_μ can be written in Fourier space as

$$\log\det\Delta_{r,j} = \frac{i}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu(-k)A^b_\nu(k)(k^2\gamma^{\mu\nu} - k^\mu k^\nu)\cdot\left[\mathrm{C}_{r,j} + \mathcal{O}(k^2)\right]. \tag{16.121}$$

We will now compute these terms explicitly from (16.120) and bring them into the form of (16.121). The terms with two powers of A^a_μ in the expansion (16.120) are those with one power of $\Delta^{(2)}$ or two powers of $\Delta^{(1)}$ or $\Delta^{(\mathcal{J})}$. Further, terms linear in $\Delta^{(\mathcal{J})}$ are proportional to $\mathrm{tr}[\mathcal{J}^{\rho\sigma}] = 0$, so the cross term between these two structures vanishes. The three remaining contributions correspond to the Feynman diagrams shown in Fig. 16.10.

The term involving two powers of $\Delta^{(1)}$ is

$$-\tfrac{1}{2}\,\mathrm{tr}\big[(-\partial^2)^{-1}\Delta^{(1)}(-\partial^2)^{-1}\Delta^{(1)}\big] =$$



$$= \tfrac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\,\mathrm{tr}\,\frac{1}{p^2}(2p+k)^\mu t^a\frac{1}{(p+k)^2}(2p+k)^\nu t^b, \tag{16.122}$$

where the trace is now simply a trace over gauge and spin indices. The factor 1/2 comes from the expansion of the logarithm. The term involving one power of $\Delta^{(2)}$ is

$$\mathrm{tr}\big[(-\partial^2)^{-1}\Delta^{(2)}\big] =$$

$$= \int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\,\mathrm{tr}\,\frac{1}{p^2}g^{\mu\nu}t^a t^b. \tag{16.123}$$

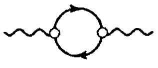
As Fig. 16.10 suggests, these two contributions are precisely proportional to the contribution of a scalar particle to the QED vacuum polarization, times the factor

$$\mathrm{tr}[t^a t^b] = C(r)d(j)\delta^{ab}, \tag{16.124}$$

where $d(j)$ is the number of spin components. The values of the diagrams can be worked out using the methods of the previous section (or simply recalled from Problem 9.1). One finds that the two diagrams together sum up to the gauge-invariant form (16.121), to give

$$\tfrac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu(-k)A^b_\nu(k)(k^2g^{\mu\nu} - k^\mu k^\nu)\cdot\left[i\frac{C(r)d(j)}{3(4\pi)^2}\Gamma(2-\tfrac{d}{2}) + \cdots\right]. \tag{16.125}$$

The term involving two powers of $\Delta^{(\mathcal{J})}$ is

$$-\tfrac{1}{2}\,\mathrm{tr}\big[(-\partial^2)^{-1}\Delta^{(\mathcal{J})}(-\partial^2)^{-1}\Delta^{(\mathcal{J})}\big] =$$ 

$$= -\tfrac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\,\mathrm{tr}\,\frac{1}{p^2}(2ik_\rho g_{\mu\sigma}\mathcal{J}^{\rho\sigma})t^a\frac{1}{(p+k)^2}(-2ik_\alpha g_{\nu\beta}\mathcal{J}^{\alpha\beta})t^b. \tag{16.126}$$

To evaluate this, define $C(j)$ as the trace over spin indices

$$\mathrm{tr}[\mathcal{J}^{\rho\sigma}\mathcal{J}^{\alpha\beta}] = (g^{\rho\alpha}g^{\sigma\beta} - g^{\rho\beta}g^{\sigma\alpha})C(j). \tag{16.127}$$

It is straightforward to work out from the explicit expressions that

$$C(j) = \begin{cases} 0 & \text{scalars;} \\ 1 & \text{Dirac spinors;} \\ 2 & \text{4-vectors.} \end{cases} \tag{16.128}$$

Then (16.126) can be evaluated as

$$-\frac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^2}\frac{1}{(p+k)^2}(k^2g^{\mu\nu} - k^\mu k^\nu)4C(r)C(j)$$
$$= \frac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu(-k)A^b_\nu(k)(k^2g^{\mu\nu} - k^\mu k^\nu)\Big(-i\frac{4C(r)C(j)}{(4\pi)^2}\Gamma(2-\tfrac{d}{2}) + \cdots\Big). \tag{16.129}$$

Adding (16.125) and (16.129), we find that the coefficient $\mathbf{C}_{r,j}$ in (16.111) is given by

$$\mathbf{C}_{r,j} = \frac{1}{(4\pi)^2}\big[\tfrac{1}{3}d(j) - 4C(j)\big]C(r)\Gamma(2-\tfrac{d}{2}). \tag{16.130}$$

Thus,

$$c_{r,j} = \frac{1}{(4\pi)^2}\big[\tfrac{1}{3}d(j) - 4C(j)\big]C(r), \tag{16.131}$$

or explicitly,

$$c_{r,j} = \frac{C(r)}{(4\pi)^2}\cdot\begin{cases} +1/3 & \text{scalars;} \\ -8/3 & \text{Dirac spinors;} \\ -20/3 & \text{4-vectors.} \end{cases} \tag{16.132}$$

Notice that, whenever the magnetic moment term is nonzero, it dominates, and that its coefficient is opposite in sign from the convective term.

Inserting the values from (16.132) into (16.117), we find

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\Big(\frac{11}{3}C_2(G) - \frac{4}{3}n_fC(r)\Big). \tag{16.133}$$

We thus confirm the conclusion of the previous section, that non-Abelian gauge theories with sufficiently few fermions are asymptotically free.

16.7 Asymptotic Freedom: A Qualitative Explanation

In the previous two sections† we twice calculated the β function in non-Abelian gauge theory:

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}C_2(G) - \frac{4}{3}n_f C(r)\right). \tag{16.134}$$

Here n_f is the number of fermion species in representation r, $C(r)$ is the constant appearing in the orthogonality relation (15.78) for the representation matrices, and $C_2(G)$ is the quadratic Casimir operator of the adjoint representation of the group, defined in Eq. (15.92). In an $SU(N)$ gauge theory with fermions in the fundamental representation, this result becomes

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}N - \frac{2}{3}n_f\right). \tag{16.135}$$

The overall minus sign implies that, for sufficiently small n_f, non-Abelian gauge theories are asymptotically free. In this case the running coupling constant tends to zero at large momenta, according to Eq. (12.92):

$$g^2(k) = \frac{g^2}{1 + \frac{g^2}{(4\pi)^2}\left(\frac{11}{3}N - \frac{2}{3}n_f\right)\log(k^2/M^2)}. \tag{16.136}$$

The asymptotic freedom of non-Abelian gauge theories is a surprising conclusion. When we first encountered the running of the electromagnetic coupling in Section 7.5, we found it easy to understand the direction of the coupling constant flow: The vacuum acquires a dielectric property due to virtual electron-positron pair creation, causing the effective electric charge to decrease at large distances. In non-Abelian gauge theories, according to Eq. (16.134), the fermions still produce such an effect. Furthermore, since the non-Abelian gauge bosons are charged, they should produce an additional screening effect. But according to Eq. (16.134), the net effect of the non-Abelian gauge bosons is *opposite* in sign. Apparently there must be other, competing, effects, which overcome the effect of screening and change the sign of the β function.

The precise form of these effects depends on the gauge. They are simplest to describe in the Coulomb gauge, for which the gauge fixing condition is

$$\partial_i A^{ai} = 0. \tag{16.137}$$

We will not work out the full quantization in this gauge; rather, we will just describe its qualitative features. As in electrodynamics, the field quanta in Coulomb gauge are described in a non-Lorentz-covariant manner as transversely polarized photons. There are no timelike or longitudinal photons and

†Section 16.7 draws on the main result of 16.5 and 16.6, but does not depend on these earlier sections. However, even if you have not read Section 16.5, you may wish to skim pages 522 through 531 to get an overview of how the β function can be calculated.

no propagating ghosts. However, there is a Coulomb potential, described by the field A^{a0}, which obeys a constraint equation analogous to Gauss's law. Not surprisingly, in the non-Abelian case, Gauss's law takes the gauge-covariant form

$$D_i E^{ai} = g\rho^a, \tag{16.138}$$

where $E^{ai} = F^{a0i}$ and ρ^a is the charge density of the global symmetry current of the fermions. Recall from Eq. (15.86) that the covariant derivative acting on a field in the adjoint representation is

$$(D_\mu \phi)^a = \partial_\mu \phi^a + g f^{abc} A^b_\mu \phi^c.$$

To analyze the consequences of Eq. (16.138), we will choose an example as simple and explicit as possible. Let the gauge group be $SU(2)$, so that $a = 1, 2, 3$ and $f^{abc} = \epsilon^{abc}$. Let us compute the Coulomb potential of a point charge of magnitude $+1$ with the orientation $a = 1$. We will solve for E^{ai} using an iteration process, putting the gauge-field term of the covariant derivative on the right-hand side of the equation:

$$\partial_i E^{ai} = g\delta^{(3)}(\mathbf{x})\delta^{a1} + g\epsilon^{abc} A^{bi} E^{ci}. \tag{16.139}$$

The second term on the right shows that, in a non-Abelian gauge theory, a region containing vector potentials and electric fields that are parallel in physical space and perpendicular in the group space is a source of electric field.

The implication of Eq. (16.139) is worked out pictorially in Fig. 16.11. The leading term on the right-hand side of (16.139) implies a $1/r^2$ electric field of type $a = 1$ radiating from $\mathbf{x} = 0$. Somewhere in space, this electric field will cross with a bit of vector potential A^{ai} arising as a fluctuation of the vacuum. For definiteness, let us assume that this fluctuation has $a = 2$ and points in some diagonal direction, as shown in Fig. 16.11(a). Then the second term on the right-hand side of Eq. (16.139) is negative for $a = 3$: There is a sink of the field E^{3i} at this location, as shown in Fig. 16.11(b). These new fields are, in two locations, parallel or antiparallel to the original A^{ai} field fluctuation. Looking again at the second term of Eq. (16.139), we see that there is a source of electric field with $a = 1$ closer to the origin, and a sink of electric field with $a = 1$ farther away. This is an induced electric dipole in the vacuum, shown in Fig. 16.11(c). But look at the signs: This dipole is oriented toward the original charge, and thus serves to amplify rather than screen it! The effect of the original charge thus becomes stronger at larger distances.

The competition between this antiscreening effect and the screening due to virtual pairs of gauge bosons must be worked out quantitatively. When this is done,‡ one finds that the antiscreening effect is 12 times larger.

In this argument, it is a set of dynamical features peculiar to the non-Abelian gauge theory that enables the coupling constant to be amplified rather

‡T. Appelquist, M. Dine, and I. Muzinich, *Phys. Lett.* **69B**, 231 (1977).

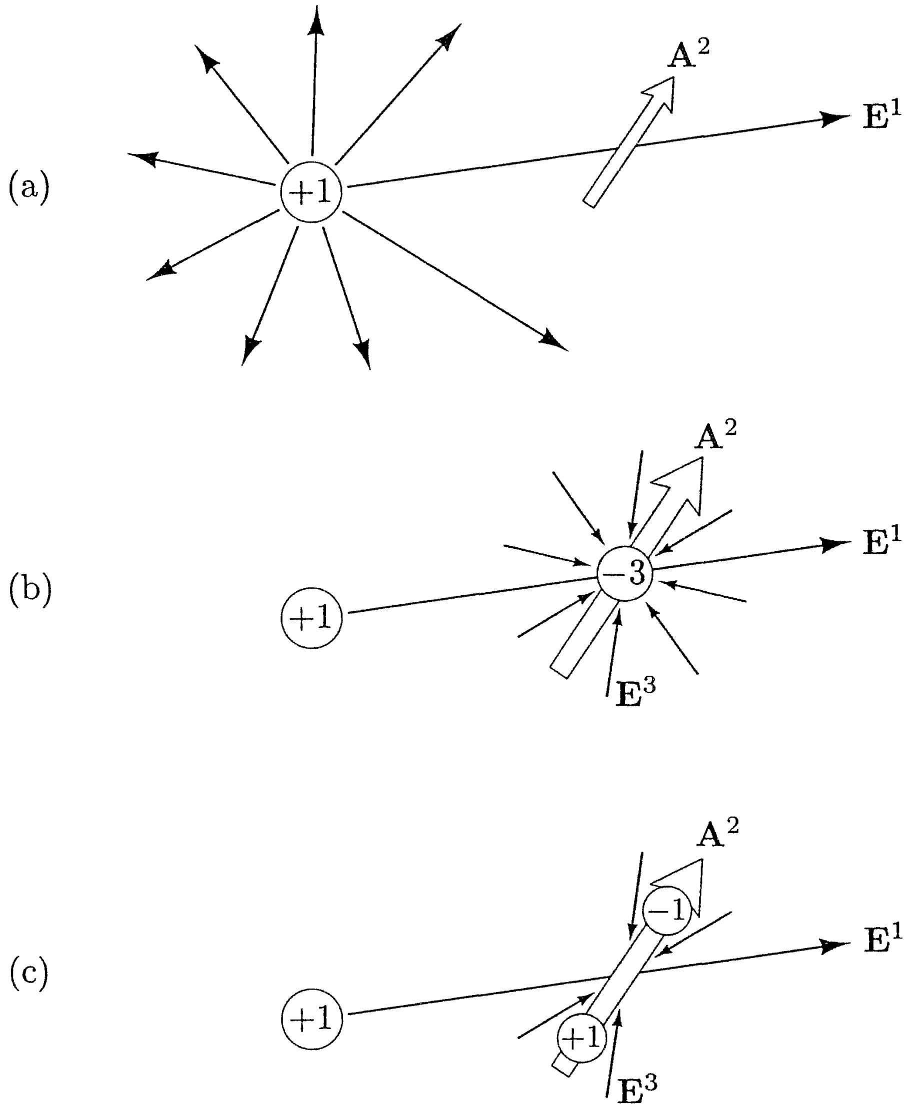


Figure 16.11. The effect of vacuum fluctuations on the Coulomb field of an $SU(2)$ gauge theory. In (a), a fluctuation $\mathbf{A}^2$ occurs on top of the $1/r^2$ field $\mathbf{E}^1$. These combined fields generate a sink of the field $\mathbf{E}^3$, as shown in (b). The $\mathbf{E}^3$ field, in turn, combines with $\mathbf{A}^2$ to create an effective $\mathbf{E}^1$ dipole, shown in (c). The dipole points toward the original charge, enhancing its field at large distances.

than screened at large distances. This suggests that asymptotic freedom might be a special property of non-Abelian gauge theories. Although the statement can be proved only by exhausting other cases, it does actually turn out to be true: Among renormalizable quantum field theories in four spacetime dimensions, only the non-Abelian gauge theories are asymptotically free.* We have already seen in Chapter 14 that asymptotic freedom was suggested experimentally as a property of the strong interactions. In the following chapter we will build a model of the strong interactions out of a non-Abelian gauge theory and explore its properties in detail.

*S. Coleman and D. J. Gross, *Phys. Rev. Lett.* **31**, 851 (1973).

Problems

16.1 Arnowitt-Fickler gauge. Perform the Faddeev-Popov quantization of Yang-Mills theory in the gauge $A^{3a} = 0$, and write the Feynman rules. Show that there are no propagating ghosts, and that the gauge field is reduced to two positive-metric degrees of freedom. (Although the gauge condition violates Lorentz invariance, this symmetry is restored in the calculation of gauge-invariant S-matrix elements.)

16.2 Scalar field with non-Abelian charge. Consider a non-Abelian gauge theory with gauge group G. Couple to this theory a complex scalar field in the representation r.

(a) Show that the Feynman rules for the scalar field are a simple modification of the Feynman rules displayed for scalar QED in Problem 9.1(a).

(b) Compute the contribution of this scalar field to the β function, and show that the full β function for this theory is

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}C_2(G) - \frac{1}{3}C(r)\right).$$

16.3 Counterterm relations. In Section 16.5, we computed the divergent parts of δ_1, δ_2, and δ_3. Compute the divergent parts of the remaining counterterms in Eq. (16.88) to one-loop order in Feynman-'t Hooft gauge, and verify explicitly that relations (16.89) are consistent with the removal of ultraviolet divergences. Notice that the ghost field strength renormalization is finite to one-loop order; thus, consideration of the ghost vertex gives a particularly simple way to complete the derivation of the β function.

Chapter 17

Quantum Chromodynamics

The key to constructing a realistic model of the strong interactions is the phenomenon of asymptotic freedom. Chapter 14 described the experimental discovery of this phenomenon, while Chapter 16 presented the theoretical proof that non-Abelian gauge theories are asymptotically free. We are now ready to explore the consequences of these discoveries.

We will begin by arguing that the most natural candidate for a model of the strong interactions is the non-Abelian gauge theory with gauge group $SU(3)$, coupled to fermions (quarks) in the fundamental representation. This theory is known as *Quantum Chromodynamics*, or QCD. After some general discussion of QCD in Section 17.1, we will investigate a number of specific QCD scattering processes in Sections 17.2 through 17.4. The most interesting application of QCD, however, is of a somewhat more sophisticated nature; it comes in the prediction of a pattern of slow violations of the Bjorken scaling relation discussed in Chapter 14. Section 17.5 develops the additional theoretical tools that are needed to understand these violations.

Although this chapter includes many references to experiments, we remind the reader that, for QCD as for QED or critical phenomena, this book is primarily a textbook of theoretical methods rather than a review and interpretation of experimental data. The details of experimental techniques and results on strong interaction physics are reviewed in a number of excellent texts (see the bibliography). We hope that this chapter will give the theoretical foundation necessary to illuminate and interpret these results.

17.1 From Quarks to QCD

Our current theoretical picture of the strong interactions began with the identification of the elementary fermions that make up the proton and other hadrons. As the properties of these fermions became better understood, the nature of their interactions became tightly constrained, in a way that led eventually to a unique candidate theory. In order to appreciate the uniqueness of this theory, we begin this chapter with a simplified history of how it arose.

In 1963, Gell-Mann and Zweig proposed a model that explained the spectrum of strongly interacting particles in terms of elementary constituents called *quarks.* Mesons were expected to be quark-antiquark bound states. Indeed, the lightest mesons have just the correct quantum numbers to justify this interpretation; they are spin-0 and spin-1 states of odd parity, just as we found for fermion-antifermion bound states of zero orbital angular momentum in Chapter 3. Baryons were interpreted as bound states of three quarks. To explain the electric charges and other quantum numbers of hadrons, Gell-Mann and Zweig needed to assume three species of quarks, up (u), down (d), and strange (s). Additional hadrons discovered since that time require the existence of three more species: charm (c), bottom (b), and top (t). To make baryons with integer charges, the quarks needed to be assigned fractional electric charge: $+2/3$ for u, c, t, and $-1/3$ for d, s, b. Then, for example, the proton would be a bound state of uud, while the neutron would be a bound state of udd. The six types of quarks are conventionally referred to as *flavors.*

The quark model had great success in predicting new hadronic states, and in explaining the strengths of electromagnetic and weak-interaction transitions among different hadrons. In particular, the quark model naturally incorporates the most important symmetry relations among strongly interacting particles. If one assumes that the u and d quarks have identical masses and interactions, the $SU(2)$ group that acts as a unitary rotation of u and d states,

$$\begin{pmatrix} u \\ d \end{pmatrix} \rightarrow U \begin{pmatrix} u \\ d \end{pmatrix}, \tag{17.1}$$

should be a symmetry of the strong interactions. Indeed, both in nuclear and in elementary particle physics, the quantum states form multiplets of this $SU(2)$ symmetry, called *isotopic spin* or *isospin.* Similarly, since the strange quark is only a little heavier than the u and d quarks, it makes sense to consider the symmetry of unitary transformations of the triplet (u, d, s). Gell-Mann and Ne'eman showed that the elementary particles naturally fill out irreducible representations of this $SU(3)$ symmetry.

Despite the phenomenological success of the original quark model, it had two serious problems. First, despite considerable effort, free particles with fractional charge could not be found. Second, the spectrum of baryons required the assumption that the wavefunction of the three quarks be totally symmetric under the interchange of the quark spin and flavor quantum numbers, contradicting the expectation that quarks, which must have spin 1/2, should obey Fermi-Dirac statistics. The need for this symmetry is most clearly illustrated in the fact that one of the lightest excited states of the nucleon is a spin-3/2 particle with charge +2, the Δ^{++}. This particle is readily interpreted as a uuu bound state with zero orbital angular momentum and all three quark spins parallel.

To reconcile the baryon spectrum with the spin-statistics theorem, Han and Nambu, Greenberg, and Gell-Mann proposed that quarks carry an additional, unobserved quantum number, called *color.* They introduced the *ad hoc* assumption that baryon wavefunctions must be totally antisymmetric in color quantum numbers. Then, if the quark wavefunctions are totally symmetric in spin and flavor, they are totally antisymmetric overall, in agreement with Fermi-Dirac statistics. The simplest model of color would be to assign quarks to the fundamental representation of a new, internal $SU(3)$ global symmetry. Suppressing for a moment the spin and flavor quantum numbers, we can represent quarks by q_i, where $i = 1, 2, 3$ is the color index. Thus quarks transform under the fundamental, or "3", representation of the color $SU(3)$ symmetry. Antiquarks, $\bar{q}^i$, transform in the $\bar{3}$ representation. The inner product of a 3 and a $\bar{3}$ is an invariant of $SU(3)$. One can also form an invariant by using the totally antisymmetric combination of three 3's, ϵ_{ijk}: This object transforms under a unitary transformation according to

$$\epsilon_{ijk} \to U_{ii'}U_{jj'}U_{kk'}\epsilon_{i'j'k'} = (\det U)\epsilon_{ijk}, \tag{17.2}$$

and so is invariant under $SU(3)$ transformations, which have $\det U = 1$. Under the postulate that all hadron wavefunctions must be invariant under $SU(3)$ symmetry transformations, these two types of combinations are the only simple ones allowed:

$$\bar{q}^i q_i, \qquad \epsilon^{ijk} q_i q_j q_k, \qquad \epsilon_{ijk}\bar{q}^i\bar{q}^j\bar{q}^k. \tag{17.3}$$

That is, the assumption that physical hadrons are singlets under color implies that the only possible light hadrons are the mesons, baryons, and antibaryons.

Like the original quark model, the color hypothesis was phenomenologically successful but raised additional questions: Why should quarks have this seemingly superfluous property, and what mechanism insures that all hadron wavefunctions are color singlets? The answers to these questions came not from hadron spectroscopy, but from the deep-inelastic scattering experiments described in Chapter 14 and the ensuing search for a theory of parton binding with the property of asymptotic freedom. When it was discovered that non-Abelian gauge theories have this property, all that remained was to identify the correct gauge group and fermion representation. Since the color symmetry had no other obvious physical role, it was natural to identify this symmetry with the gauge group, with the colors being the gauge quantum numbers of the quarks. This reasoning resulted in a model of the strong interactions as a system of quarks, of the various flavors, each assigned to the fundamental representation of the local gauge group $SU(3)$. The quanta of the $SU(3)$ gauge field are called *gluons*, and the theory is known as Quantum Chromodynamics, or QCD.

In this book, we will mainly investigate the properties of QCD in the high-energy regime, where the coupling constant has become small. However, we should point out that one can also study QCD in the regime of strong coupling, using an approximation scheme introduced by Wilson in which the continuum gauge theory is replaced by a discrete statistical mechanical system on a four-dimensional Euclidean lattice. Using this approximation, Wilson showed that, for sufficiently strong coupling, QCD exhibits *confinement of color*: The only finite-energy asymptotic states of the theory are those that are singlets of color $SU(3)$. Thus the *ad hoc* assumption that explains the

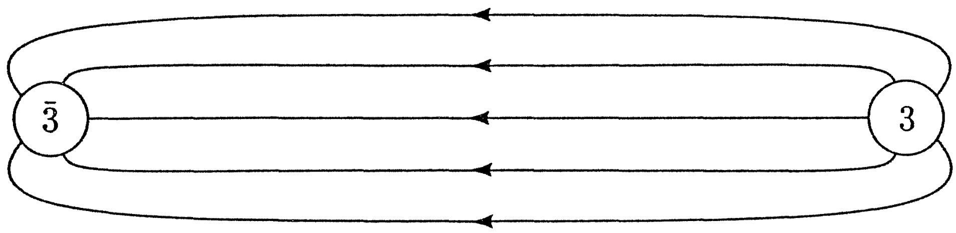


Figure 17.1. Gauge electric field configuration associated with the separation of color sources in a strong-coupling gauge theory.

spectrum of hadrons turns out to be a consequence of the non-Abelian gauge theory coupling to color. If one attempts to separate a color-singlet state into colored components—for example, to dissociate a meson into a quark and an antiquark—a tube of gauge field forms between the two sources, as shown in Fig. 17.1. In a non-Abelian gauge theory with sufficiently strong coupling, this tube has fixed radius and energy density, so the energy cost of separating color sources grows proportionally to the separation. A force law of this type can consistently be weak at short distances and strong at long distances, accounting for the fact that isolated quarks are not observed. We will discuss the large-distance, strong-coupling limit of gauge theories further in the Epilogue.

The short-distance limit of Quantum Chromodynamics can be readily studied using the Feynman diagram technology that we have developed in previous chapters. Here asymptotic freedom makes the coupling weak, and there is a sensible diagrammatic perturbation theory that begins from the model of free quarks and gluons. The following sections treat the elementary interactions among quarks and gluons that can be observed in high-energy experiments.

17.2 e^+e^- Annihilation into Hadrons

The simplest reaction involving quarks is the production of quark pairs in e^+e^- annihilation, a process that we treated already in Section 5.1. There we analyzed this process only at the most elementary level, viewing it as a pure QED reaction in which free quarks are created by a virtual photon. The diagram for this lowest-order process is shown in Fig. 17.2(a). The computation of the total cross section includes a sum over the various color states of the quark fields, and so provides a confirmation that the number of allowed colors is 3. Combining the color factor with the square of the quark electric charges, we found (Eq. (5.16))

$$\sigma(e^+e^- \to \text{hadrons}) = \sigma_0 \cdot 3 \cdot \sum_f Q_f^2, \tag{17.4}$$

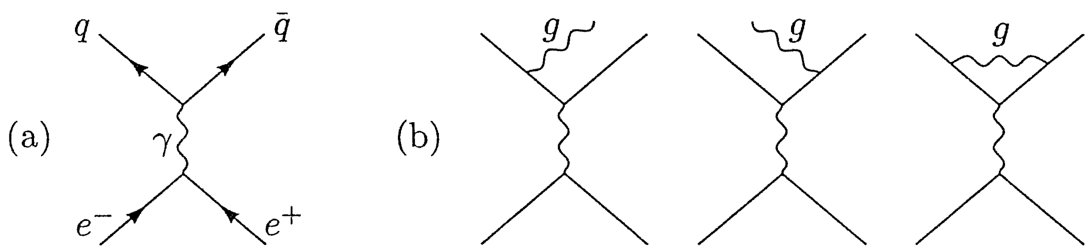


Figure 17.2. Diagrams contributing to the process $e^+e^- \to$ hadrons in QCD: (a) the leading-order diagram; (b) corrections of order α_s.

where σ_0 is the QED cross section for $e^+e^- \to \mu^+\mu^-$,

$$\sigma_0 = \frac{4\pi\alpha^2}{3s}, \tag{17.5}$$

and the sum in (17.4) is taken over quark flavors. This formula assumes that the center of mass energy is high enough that we can ignore the quark masses.

When we couple the quarks to an $SU(3)$ gauge theory, we add many important processes that affect both the value of this cross section and the final states that it includes. Some of the most important effects cannot be discussed within the context of perturbation theory. In particular, though the leading diagram contains free quarks, the particles that emerge from the reaction are color-singlet mesons and baryons. However, we will find that QCD perturbation theory with quarks and gluons does make a number of important predictions for e^+e^- annihilation to hadrons. The ideas that we develop in working out these predictions will also apply to many other strong-interaction processes.

Total Cross Section

The leading corrections to the rate of e^+e^- annihilation due to gluon exchange and emission are shown in Fig. 17.2(b). These are precisely the diagrams computed in the Final Project of Part I. The first two diagrams give a cross section of order g^2, where g is the $SU(3)$ gauge coupling, to produce a gluon in addition to the quark and antiquark. The third diagram must be summed in the amplitude with the leading diagram to produce a correction to the rate of $q\bar{q}$ production without gluon emission. In Part I, we computed these two contributions as if the strong interactions were an Abelian gauge theory. To obtain the corresponding expressions in QCD, we need only multiply the Abelian formulae by the group theory factor

$$\text{tr}[t^a t^a] = C_2(r) \cdot \text{tr}[1] = \frac{4}{3} \cdot 3, \tag{17.6}$$

where we have used Eq. (15.97) to evaluate $C_2(r)$ for the fundamental representation of $SU(3)$. The factor of 3 is the same color sum that appears in Eq. (17.4). Thus we can obtain the correct formulae for QCD from those of the Final Project by making the replacement

$$g^2 \to \frac{4}{3}g^2, \qquad \text{or} \qquad \alpha_g \to \frac{4}{3}\alpha_s, \tag{17.7}$$

where

$$\alpha_s = \frac{g^2}{4\pi} \tag{17.8}$$

is the strong-interaction analogue of the fine-structure constant.

The end result of the Final Project of Part I was a formula for the total cross section to produce hadrons in e^+e^- annihilation. If we replace α_g with $(4/3)\alpha_s$, that result becomes

$$\sigma(e^+e^- \to \text{hadrons}) = \sigma_0 \cdot \Big(3\sum_f Q_f^2\Big) \cdot \Big[1 + \frac{\alpha_s}{\pi} + \mathcal{O}(\alpha_s^2)\Big]. \tag{17.9}$$

This is actually the sum of the rates for two elementary processes, $e^+e^- \to q\bar{q}$ (including the correction from the third diagram of Fig. 17.2(b)) and $e^+e^- \to q\bar{q}g$. Although the rate for each of these processes is divergent as the gluon mass is taken to zero, that divergence cancels when they are combined. This is another example of the phenomenon of infrared divergence cancellations that we studied for the example of electron scattering in Sections 6.4 and 6.5. There we showed that the dressing of the final state by the emission of soft and collinear photons does not affect the overall scattering rate. Here, we see again that infrared divergences cancel in the total rate, although the sum over real and virtual gluon corrections leaves over a simple numerical correction.

It is not difficult to understand the cancellation of infrared logarithms intuitively. The original process $e^+e^- \to q\bar{q}$ is extremely rapid: Since the virtual photon is off-shell by an amount $q^2 = s$, the quarks are created in a time $1/\sqrt{s}$. However, the emission of collinear gluons, and the virtual corrections associated with the exchange of soft gluons, occur over a much longer time scale. In the diagrams with gluon emission, the virtual quark or antiquark is off-mass-shell by an amount $p_{\perp g}^2$, where $p_{\perp g}$ is the transverse momentum of the gluon relative to the $q\bar{q}$ system. Thus this virtual state survives for a time $1/p_{\perp g}$ before it decays. Such a slow process cannot affect the probability that a $q\bar{q}$ pair was produced; it can only affect the properties of the final state into which the $q\bar{q}$ system will evolve. By this logic, the only perturbative corrections that can affect the total cross section are those for which $p_{\perp g} \sim \sqrt{s}$. Another way to express this conclusion would be to argue that, after contributions from the infrared-sensitive regions have canceled, the contributions that remain come from the region of large real or virtual gluon momenta. By either argument, formula (17.9) should be a meaningful prediction of QCD perturbation theory, even though it involves an integral over the region of soft gluon emission.

The Running of α_s

Formula (17.9) depends on the QCD coupling constant α_s, which must be defined at some renormalization point M. This is in contrast to the QED coupling constant, which we defined in a natural way by on-shell renormalization. In QCD, we would like to avoid discussing on-shell quarks, since these are strongly interacting particles that are significantly affected by nonperturbative forces. The use of an arbitrary renormalization point M allows us to avoid this problem. We will define α_s by renormalization conditions imposed at a large momentum scale M where the coupling is small; this value of α_s can then be used to predict the results of scattering processes with any large momentum transfer.

However, the use of renormalization at a scale M in a computation involving momentum invariants of order P^2 involves some subtlety when P^2 and M^2 are very different. In our discussion of Section 12.3, we saw that, in this circumstance, Feynman diagrams with n loops typically contain correction terms proportional to $(\alpha_s \log(P^2/M^2))^n$. Fortunately, we can absorb these corrections into the lowest-order terms by using the renormalization group to replace the fixed renormalized coupling with a running coupling constant.

To illustrate how this analysis applies to QCD, let us examine the implications of the Callan-Symanzik equation for the e^+e^- annihilation total cross section σ, viewed as a function of s, a renormalization scale M, and the value of α_s at the renormalization scale. Like the QED potential (12.87), the e^+e^- total cross section is an observable quantity and so its normalization is independent of any conventions. It therefore obeys a Callan-Symanzik equation with $\gamma = 0$:

$$\left[M\frac{\partial}{\partial M} + \beta(g)\frac{\partial}{\partial g}\right]\sigma(s, M, \alpha_s) = 0. \tag{17.10}$$

By dimensional analysis, we can write

$$\sigma = \frac{c}{s} f\left(\frac{s}{M^2}, \alpha_s\right), \tag{17.11}$$

were c is a dimensionless constant. Then the Callan-Symanzik equation implies that f depends on its arguments only through the running coupling constant $\alpha_s(Q) = \bar{g}^2/4\pi$, evaluated at $Q^2 = s$. The coupling constant $\bar{g}$ is defined to satisfy the renormalization group equation

$$\frac{d}{d\log(Q/M)}\bar{g} = \beta(\bar{g}), \tag{17.12}$$

with initial condition $\alpha_s(M) = \alpha_s$. For QCD with three colors and n_f approximately massless quarks, the β function is given by Eq. (16.135):

$$\beta(g) = -\frac{b_0}{(4\pi)^2}, \qquad \text{with } b_0 = 11 - \frac{2}{3}n_f. \tag{17.13}$$

Then the solution of the renormalization group equation is

$$\alpha_s(Q) = \frac{\alpha_s}{1 + (b_0\alpha_s/2\pi)\log(Q/M)}. \tag{17.14}$$

The explicit dependence of σ on α_s can be found by matching the successive terms in the expansion of $f(\alpha_s(\sqrt{s}))$ to the terms in the perturbative expansion. To the order of the first corrections, we find simply

$$\sigma = \sigma_0 \cdot \Big(3\sum_f Q_f^2\Big) \cdot \Big[1 + \frac{\alpha_s(\sqrt{s})}{\pi} + +\mathcal{O}(\alpha_s^2(\sqrt{s}))\Big]. \tag{17.15}$$

Thus the Callan-Symanzik equation instructs us to replace the fixed renormalized coupling α_s with the running coupling constant $\alpha_s(Q)$, evaluated at $Q^2 = s$.

Because the fixed coupling α_s depends on the arbitrary renormalization point M, it is sometimes useful to remove it from our formulae completely. To do this, we define a mass scale conventionally called Λ (not to be confused with an ultraviolet cutoff!) satisfying

$$1 = g^2(b_0/8\pi^2)\log(M/\Lambda). \tag{17.16}$$

Then Eq. (17.14) can be rearranged into the form

$$\alpha_s(Q^2) = \frac{2\pi}{b_0\log(Q/\Lambda)}. \tag{17.17}$$

This formula is the clearest expression of the statement that $\alpha_s(Q)$ becomes small as $(\log(Q))^{-1}$ for large Q. The momentum scale Λ is the scale at which α_s becomes strong as Q^2 is decreased.

Experimental measurements of the rate of this reaction and others yield a value of $\Lambda \approx 200$ MeV. QCD perturbation theory is valid only when Q is somewhat larger than this, say above $Q = 1$ GeV, where $\alpha_s(Q) \approx 0.4$. The strong interactions become strong at distances larger than $\sim 1/\Lambda$, which is roughly the size of the light hadrons.

Although the example of the e^+e^- annihilation cross section is especially simple, since it depends on only one momentum invariant, similar conclusions carry over to other QCD predictions. In analyzing strong-interaction processes that are sensitive to the quark and gluon substructure, we will find leading-order formulae for the reaction cross sections that depend on the renormalized coupling α_s. To make these expressions satisfy the Callan-Symanzik equation, we must replace this fixed coupling with the running coupling constant $\alpha_s(Q)$, evaluated with Q of the order of the momentum invariants of the reaction. Since the running coupling constant depends only logarithmically on Q, we need not worry about choosing Q precisely. If we guess the proper scale incorrectly by a factor of 2, this induces an error in $\alpha_s(Q)$ that is of order $(\log(Q))^{-2} \sim \alpha_s^2(Q)$. Conversely, this ambiguity would be resolved by computing to the next order in α_s.

Before concluding this formal treatment of the e^+e^- annihilation cross section, we should add one qualification. At the beginning of Section 12.2, we remarked that renormalization group predictions can be complicated by the appearance of physical thresholds and their associated singularities, and so we stated these predictions only for when the relevant momentum invariant P^2 was large and spacelike. In the present chapter, we will be concerned with cross sections for quark and gluon reactions, evaluated on-shell. This introduces additional complications of principle. For example, in order to apply the Callan-Symanzik equation to $\sigma(s)$, we needed to know that this quantity contains no infrared divergences whose regulator might provide another dimensionful parameter. Throughout this chapter we will assume that similiar cancellations of divergences associated with soft and collinear gluons occur in the processes of interest to us. The complete proof of these cancellations in QCD can be carried through, but it is rather technical.* In some cases, an alternative point of view is possible; one can justify the use of the renormalization group to analyze an on-shell amplitude by relating it to Green's functions evaluated in the spacelike region. This method of analysis, which brings its own insights, will be the main subject of Chapter 18.

Gluon Emission and Jet Production

A second result of the Final Project of Part I was a formula for the differential cross section for $q\bar{q}$ production with gluon emission. Transcribing this formula to QCD using (17.7) gives the following result: Let x_1, x_2, x_3 be the ratios of the quark, antiquark, and gluon energies to the electron beam energy. These satisfy $0 < x_i < 1$ and $x_1+x_2+x_3 = 2$. Then the cross section for $e^+e^- \to q\bar{q}g$ is given by

$$\frac{d\sigma}{dx_1 dx_2}(e^+e^- \to q\bar{q}g) = \sigma_0 \cdot \Big(3\sum_f Q_f^2\Big) \cdot \frac{2\alpha_s}{3\pi}\frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}. \qquad (17.18)$$

This cross section is singular as x_1 or x_2 approaches 1. The limit $x_1 \to 1$ corresponds to configurations in which the quark has the maximum possible energy, while the antiquark and the gluon go off in the opposite direction, sharing the remaining energy. Then the antiquark and gluon have almost collinear lightlike momentum vectors and so form a system of very small invariant mass. Similarly, the limit $x_2 \to 1$ corresponds to configurations in which the quark and gluon are collinear. These singularities are responsible for the divergence of the integrated cross section in the limit of vanishing gluon mass.

How should we interpret these singularities? In our general treatment of bremsstrahlung in Section 6.1, we saw that the emission of a photon by

*For a review of the theorems justifying the formulae of perturbative QCD, see J. C. Collins and D. E. Soper, *Ann. Rev. Nucl. Sci.* **37**, 383 (1987).

a scattered electron is enhanced, for collinear radiation, by a factor of order $\log(q^2/m^2)$, where m is the mass of the electron. Thus the total rate for emitting a collinear photon formally diverges in the limit of zero mass. The same conclusion holds for the emission of gluons by quarks. A divergence that appears for collinear emissions in the limit of zero mass is called a *mass singularity.* In QED, we saw that the mass singularity signals a real physical effect of strong collinear radiation when q^2 is large. In QCD, we might expect strong gluon radiation in this limit, but we must think carefully about how this radiation appears experimentally. Whether a collinear gluon is radiated or not, the quark and antiquark emerging from the reaction will undergo further soft interactions with the other products. These processes must continue, producing quark-antiquark pairs and emitting and absorbing gluons, until all colored particles are collected into color-singlet hadrons. Thus the presence of one or more collinear gluons will have no noticeable effect on the final state, which consists of two back-to-back jets of hadrons. For this reason, formula (17.18) is of no use when the gluon transverse momentum is less than the typical scale of soft gluon interactions, roughly 1 GeV.

When the gluon is emitted with substantial transverse momentum with respect to the $q\bar{q}$ axis, however, it is not possible for subsequent soft exchanges to recall or reverse this transverse momentum. In this case, the $q\bar{q}g$ system evolves into a system of three distinct jets of hadrons. Thus, sufficiently far from the collinear regions, we can interpret Eq. (17.18) as the cross section for producing events with three hadronic jets having energies x_1, x_2, x_3 times the electron beam energy.

By an analysis similar to that given above for the total cross section, we can improve Eq. (17.18) by replacing the fixed coupling constant α_s with a running $\alpha_s(Q)$. A reasonable choice for Q is the transverse momentum of the gluon, $p_{\perp g}$, described below Eq. (17.9). If this transverse momentum is too small, however, $\alpha_s(Q)$ will be large, and our leading-order formula will break down. This is a second reason why we cannot use formula (17.18) when the transverse momentum of the gluon is less than about 1 GeV.

On the other hand, when the gluon transverse momentum is much larger than 1 GeV, there is no reason to distrust QCD perturbation theory. Soft processes cannot disturb the three-jet nature of the hadronic state, and asymptotic freedom insures that the coupling constant is small, so that the leading order of perturbation theory will be a good approximation.

The three-jet cross section (17.18) is a good example of the type of prediction that one obtains from the use of perturbation theory in QCD. We describe a strong-interaction process by the invariant momentum transfer Q it gives to hadronic constituents. QCD perturbation theory makes predictions about the flow of energy and momentum in such a reaction into the final system of jets of hadrons. If Q is small, perturbation theory is invalid, and we obtain no useful prediction. However, if Q is large, the asymptotic freedom of QCD implies that Feynman diagrams for quarks and gluons will correctly predict the behavior of the final system of hadronic jets.

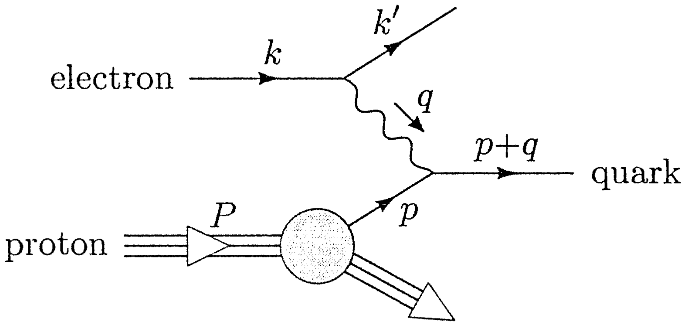


Figure 17.3. Deep inelastic scattering in QCD. The diagram shows the flow of momentum when a high-energy electron scatters from a quark taken from the wavefunction of the proton.

17.3 Deep Inelastic Scattering

After e^+e^- annihilation into hadrons, the next simplest reaction involving strongly interacting particles is electron scattering from a proton, or from some other hadron. At the most elementary level, this reaction can be viewed as the electromagnetic scattering of an electron from a quark inside the proton.† A way to visualize the process is shown in Fig. 17.3. Call the proton momentum P, and the initial quark momentum p. Call the initial and final momenta of the electron k and k'. If the final electron momentum is measured, one can deduce the momentum $q = k - k'$ transferred by the virtual photon to the hadronic system. The vector q is spacelike, and one conventionally writes $q^2 = -Q^2$.

If the invariant momentum transfer Q^2 is large, the quark is ejected from the proton in a manner that cannot be balanced by subsequent soft processes. These soft processes will, however, create gluons and quark-antiquark pairs that eventually neutralize the color and cause the struck quark to materialize as a jet of hadrons in the direction of the momentum transfer from the electron. Typically the total invariant mass of the final hadronic system is large, since the struck quark carries a large momentum with respect to the other "spectator" quarks. In this case, the process is referred to as *deep inelastic* scattering.

To derive a first approximation to the cross section for electron-proton scattering, we consider this reaction from a frame in which the electron and proton are moving rapidly toward each other, for example, the electron-proton center-of-mass frame. We assume that the center-of-mass energy is large enough that we can ignore the proton mass in working out the kinematics. Then the proton has an almost lightlike momentum along the collision axis. The constituents of the proton also have lightlike momenta, which

†The electron could just as well be a muon; all the same formulae apply in this case. Leptons can also scatter from quarks via the neutral-current weak interaction, as we will see in Chapter 20.

are almost collinear with the momentum of the proton. This is because a constituent cannot acquire a large transverse momentum except through exchange of a hard gluon, a process that is suppressed by the smallness of α_s at large momentum scales. Thus, to leading order in QCD perturbation theory, we can write

$$p = \xi P, \tag{17.19}$$

where ξ is a number between 0 and 1, called the *longitudinal fraction* of the constituent. To leading order in α_s we can also ignore gluon emission or exchanges during the collision process. The cross section for electron-proton scattering is then given by the cross section for electron-quark scattering at given longitudinal fraction ξ, multiplied by the probability that the proton contains a quark at that value of ξ, integrated over ξ.

But the probability that the proton contains a certain constituent with a certain momentum fraction cannot be computed using QCD perturbation theory, since it depends on the soft processes that determine the structure of the proton as a bound state of quarks and gluons. We will therefore consider this probability to be an unknown function, to be determined from experiment. Eventually, we will need to make use of such a probability function for each species of quark, antiquark, and gluon that can be found in the wavefunction of the proton. Collectively, these constituents are called *partons*. For each parton species f, we write the probability of finding a constituent of the proton of type f at longitudinal fraction ξ as

$$\begin{pmatrix} \text{probability of finding constituent } f \\ \text{with longitudinal fraction } \xi \end{pmatrix} = f_f(\xi) d\xi. \tag{17.20}$$

The functions $f_f(\xi)$ are called the *parton distribution functions*. Using this notation, the cross section for electron-proton inelastic scattering is given to leading order in α_s by the expression

$$\begin{aligned} \sigma\big(e^-(k)p(P) \to e^-(k') + X\big) \\ = \int_0^1 d\xi \sum_f f_f(\xi)\, \sigma\big(e^-(k)q_f(\xi P) \to e^-(k') + q_f(p')\big), \end{aligned} \tag{17.21}$$

where X stands for any hadronic final state. The sum in (17.21) contains contributions from constituent antiquarks as well as constituent quarks.

Equation (17.21) is equivalent to the formula (14.8) that we constructed for this reaction in Chapter 14. Now we see that this formula is justified by the smallness of the QCD coupling constant at large momentum scales. It is important to remember, however, that (17.21) is not the complete prediction of QCD, but only the first term of an expansion in α_s; this level of approximation is called the *parton model*. The higher-order QCD corrections to Eq. (17.21) will involve modifications both to the electron-quark scattering cross section and to the parton distribution functions. The most important of these corrections are discussed in Section 17.5.

In the same way, all other reactions of the proton that involve large momentum transfer also have parton model descriptions. In QCD, all of these reaction cross sections are computed from scattering amplitudes for quarks and gluons. The initial motion of the partons for any process is described by the same parton distribution functions $f_f(\xi)$ that appear in deep inelastic scattering.

Let us now work out the explicit leading-order formula for the deep inelastic scattering cross section, reviewing the analysis in Chapter 14. In Eq. (14.3), we wrote the leading-order differential cross section for the parton-level process,

$$\frac{d\sigma}{d\hat{t}}(e^-q \to e^-q) = \frac{2\pi\alpha^2 Q_f^2}{\hat{s}^2}\left[\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}\right]. \tag{17.22}$$

In general, we will use the symbols $\hat{s}$, $\hat{t}$, $\hat{u}$ to denote the Mandelstam variables for two-body scattering processes at the parton level. These variables must be related to observable properties of the hadronic system or the scattered electron. For massless initial and final particles,

$$\hat{s}+\hat{t}+\hat{u}=0.$$

In the case of deep inelastic scattering,

$$\hat{t} = -Q^2$$

and

$$\hat{s} = 2p\cdot k = 2\xi P\cdot k = \xi s.$$

Thus the cross section for deep inelastic scattering at fixed Q^2 is given by

$$\frac{d\sigma}{dQ^2} = \int_0^1 d\xi \sum_f f_f(\xi) Q_f^2 \frac{2\pi\alpha^2}{Q^4}\left[1+\left(1-\frac{Q^2}{\xi s}\right)^2\right]\theta(\xi s - Q^2). \tag{17.23}$$

The final factor expresses the kinematic constraint $\hat{s} \geq |\hat{t}|$. Expression (17.23) should be an accurate first approximation to the deep inelastic scattering cross section when Q^2 is large. In that case, the corrections to this formula from hard gluon emissions and exchanges will be of order $\alpha_s(Q^2)$.

We also showed in Chapter 14 that the measurement of the scattered electron momentum k' and thus the momentum transfer q uniquely determines an allowed value of ξ for elastic electron-quark scattering. This value is given by Eq. (14.7):

$$\xi = x, \qquad \text{where} \quad x \equiv \frac{Q^2}{2P\cdot q}. \tag{17.24}$$

When (17.23) is expressed as a doubly differential cross section in x and Q^2, it becomes the simple product of a parton-level cross section and a sum of parton distribution functions evaluated at $\xi = x$. In the literature, the symbol x is often used interchangably with ξ, and we will follow this practice from here on.

It is especially convenient to represent the cross section in terms of dimensionless combinations of kinematic variables. One of these should be x; a good choice for the other is

$$y \equiv \frac{2P \cdot q}{2P \cdot k} = \frac{2P \cdot q}{s}. \tag{17.25}$$

In the frame in which the proton is at rest,

$$y = \frac{q^0}{k^0}, \tag{17.26}$$

that is, y is the fraction of the incident electron's energy that is transferred to the hadronic system. On the other hand, since $p = \xi P$, we can evaluate y in terms of parton variables:

$$y = \frac{2p \cdot (k - k')}{2p \cdot k} = \frac{\hat{s} + \hat{u}}{\hat{s}}, \tag{17.27}$$

so that

$$\frac{\hat{u}}{\hat{s}} = -(1 - y). \tag{17.28}$$

From Eq. (17.26) or (17.28), we see that $y \leq 1$. The kinematically allowed region in the (x, y) plane is thus

$$0 \leq x \leq 1, \qquad 0 \leq y \leq 1.$$

To express Eq. (17.23) in terms of x and y, we need the formula

$$Q^2 = xys, \tag{17.29}$$

which follows from Eqs. (17.24) and (17.25), and the change of variables

$$d\xi\, dQ^2 = dx\, dQ^2 = \frac{dQ^2}{dy} dx\, dy = xs\, dx\, dy.$$

Then the differential cross section becomes

$$\frac{d^2\sigma}{dxdy}(e^- p \to e^- X) = \Bigl(\sum_f x f_f(x) Q_f^2\Bigr) \frac{2\pi\alpha^2 s}{Q^4} \bigl[1 + (1 - y)^2\bigr]. \tag{17.30}$$

The factor $1/Q^4$ comes from the square of the virtual photon propagator. Once this factor is removed, the dependence on x and y completely factorizes. Each half of this relation contains physical information. The fact that the parton distributions $f_f(x)$ depend only on x and are independent of Q^2 is the statement of Bjorken scaling. This tells us that the initial distribution of partons is independent of the details of the hard scattering. The y dependence of the cross section comes from the underlying parton cross section. In Chapter 5, we saw that the elementary QED cross sections, viewed in the high-energy limit, reflect the helicities of the interacting particles. The behavior $[1 + (1 - y)^2]$ in (17.30) is known as the *Callan-Gross relation*; it is specific to the scattering of electrons from massless fermions. This relation gave evidence that the partons involved in deep inelastic scattering were fermions, at a time when the relation between partons and quarks was still unclear.

Deep Inelastic Neutrino Scattering

Because the sum over quark flavors factorizes in Eq. (17.30), one cannot determine the individual parton distribution functions $f_f(x)$ from electron scattering experiments alone. One can, however, obtain more detailed information on the structure of the proton through deep inelastic *neutrino* scattering.

Neutrinos have zero electric charge and so do not interact by photon exchange, but they do interact with quarks through weak interactions. We will discuss the weak interactions in detail in Chapter 20; for the moment, let us adopt a simplified description that concentrates on the elementary process shown in Fig. 17.4. In this process, a neutrino converts to the associated charged lepton,‡ exchanging a virtual massive vector boson, the W^+. This boson couples to a quark current that converts a d quark to a u quark. The effect of this exchange process is to provide a different, but completely characterized, method for injecting a hard momentum transfer q. The amplitude for this process is described by the effective Lagrangian

$$\Delta\mathcal{L} = \frac{g^2}{2}\frac{1}{m_W^2}\left[\bar{\ell}\gamma^\mu\left(\frac{1-\gamma^5}{2}\right)\nu\right]\left[\bar{u}\gamma_\mu\left(\frac{1-\gamma^5}{2}\right)d\right] + \text{h.c.}, \tag{17.31}$$

where ℓ, ν, d, u are the fermion fields associated with the charged lepton, the neutrino, and the d and u quarks, and g is the weak interaction coupling constant. The factor $1/m_W^2$ comes from the W boson propagator, considered in the limit $q^2 \ll m_W^2$. The first two factors are often written in terms of the *Fermi constant* G_F, defined as

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}. \tag{17.32}$$

This constant gives the strength of the weak interactions at energies much less than m_W. The crucial property of the weak interactions, shown explicitly in (17.31), is that the W boson couples only to the left-handed helicity states of relativistic fermions. The deeper significance of this property will be discussed in Chapter 20.

For technical reasons, it is easiest to do neutrino deep inelastic scattering using muon neutrinos, which convert to muons after emitting a W boson. It is equally feasible to scatter muon antineutrinos from nuclear targets, and, as we will see, it is interesting to compare the effects of neutrinos and antineutrinos. Since the proton contains a small admixture of the heavier quarks s, c, these also give small contributions to neutrino deep inelastic scattering. However, we will ignore these contributions in our discussion.

‡There is also a *neutral-current* weak interaction in which the neutrino remains a neutrino: see Problem 20.4.

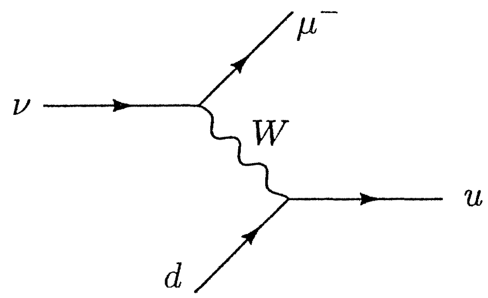


Figure 17.4. The elementary neutrino scattering process mediated by the weak interaction.

The cross section for neutrino deep inelastic scattering is given by a formula analogous to (17.21), with the photon-exchange cross section replaced by one resulting from W exchange. It is straightforward to work out this cross section directly. However, we can also obtain the result from Eq. (17.22), if we look back to Chapter 5 and recall how the structure of this equation arises from the various helicity contributions. In (17.22), the factor $\hat{t}^2$ in the denominator came from the photon propagator; this factor is replaced by m_W^4 in the weak interaction case. The factor $[\hat{s}^2 + \hat{u}^2]$ came from the Dirac matrix algebra. We saw in Section 5.2 that the first term is the contribution of left-handed electrons scattering from left-handed fermions or right-handed electrons scattering from right-handed fermions, and that the second term arises from the other helicity combinations. For the case of neutrino-quark scattering, the interaction (17.31) allows only the scattering of left-handed neutrinos from left-handed quarks, so only the $\hat{s}^2$ term appears. To determine the overall normalization of the cross section, we note that, since the neutrinos are produced in weak interactions, they always have left-handed polarization, so no polarization average should be done. On the other hand, we must still average over the polarization of the initial quark. In all, we find

$$\frac{d\sigma}{d\hat{t}}(\nu d \to \mu^- u) = \frac{\pi g^4}{2(4\pi)^2\hat{s}^2}\left[\frac{\hat{s}^2}{m_W^4}\right] = \frac{G_F^2}{\pi}. \tag{17.33}$$

It is easy to check this formula by explicit computation starting from (17.31).

The cross section for the scattering of antineutrinos from quarks can be worked out in the same way. Note that this reaction involves the exchange of a W^-, and so converts u quarks to d quarks. However, the u quarks must still be left-handed. The only modification from the previous paragraph comes in the fact that the antineutrinos that couple to the interaction (17.31) are right-handed, so the cross section comes from the term in (17.22) proportional to $\hat{u}^2$:

$$\frac{d\sigma}{d\hat{t}}(\bar{\nu} u \to \mu^+ d) = \frac{\pi g^4}{2(4\pi)^2\hat{s}^2}\left[\frac{\hat{u}^2}{m_W^4}\right] = \frac{G_F^2}{\pi}(1-y)^2. \tag{17.34}$$

Again, it is easy to verify this formula directly. The cross section for neutrino-antiquark scattering, converting a $\bar{u}$ into a $\bar{d}$, is also given by Eq. (17.34),

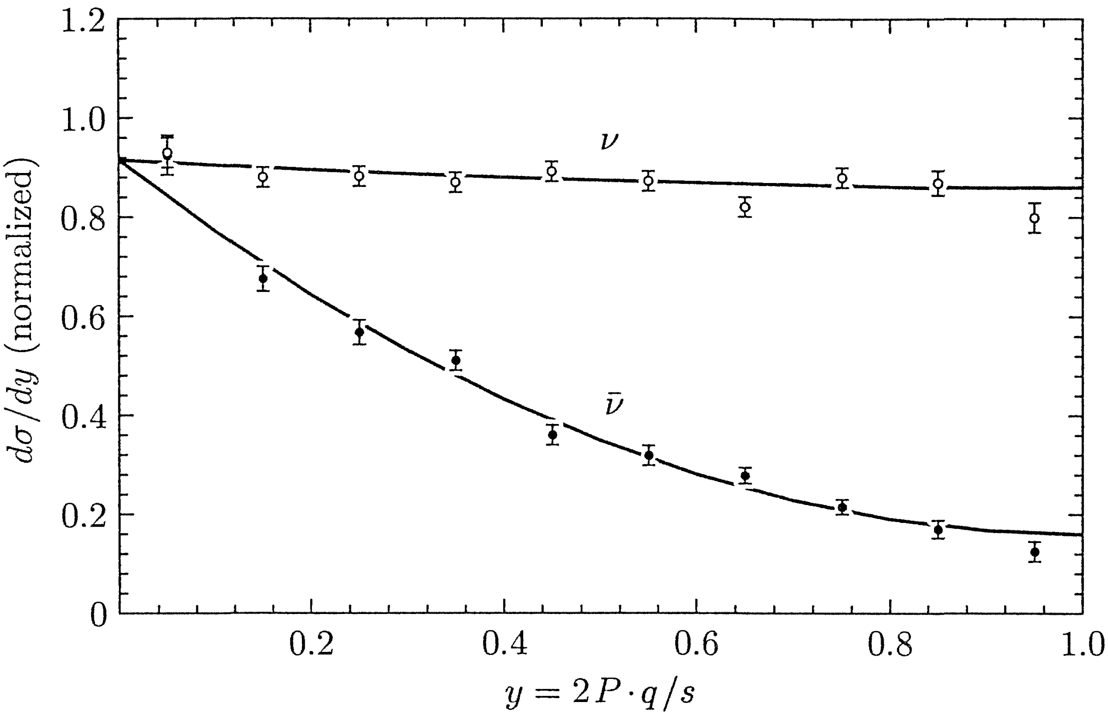


Figure 17.5. The distribution in y of neutrino and anti-neutrino deep inelastic scattering from an iron target, as measured by the CDHS experiment, J. G. H. de Groot, et. al., *Z. Phys.* **C1**, 143 (1979). The solid curves are fits to the form $A + B(1-y)^2$.

while the cross section for antineutrino-antiquark scattering, converting a $\bar{d}$ into a $\bar{u}$, is given by Eq. (17.33).

To convert these parton-level cross sections to physical cross sections, we combine them with the parton distribution functions. The kinematics is exactly the same as in the case of electron scattering. Thus, following the arguments that led to Eq. (17.30), we obtain the expressions

$$\begin{aligned} \frac{d^2\sigma}{dxdy}(\nu p \to \mu^- X) &= \frac{G_F^2 s}{\pi}\left[xf_d(x) + xf_{\bar{u}}(x)\cdot(1-y)^2\right], \\ \frac{d^2\sigma}{dxdy}(\bar{\nu} p \to \mu^+ X) &= \frac{G_F^2 s}{\pi}\left[xf_u(x)\cdot(1-y)^2 + xf_{\bar{d}}(x)\right]. \end{aligned} \tag{17.35}$$

According to these relations, deep inelastic neutrino scattering allows one to map separately the parton distribution functions for u and d quarks and antiquarks in the proton.

In addition, Eq. (17.35) makes a dramatic qualitative prediction: To the extent that a proton (or neutron) is made of quarks with very few additional quark-antiquark pairs, the deep inelastic neutrino scattering cross section should be constant in y, while the antineutrino scattering cross section should fall off as $(1-y)^2$. The measured y dependence of these deep inelastic cross sections is shown in Fig. 17.5. The qualitative behavior predicted by the parton description is clearly evident; the discrepancy from the strict prediction can be accounted for by a small antiquark component in the nucleon wavefunction.

The Parton Distribution Functions

Given that the parton model predictions for electron and neutrino deep inelastic scattering do fit the data, one can make use of these relations to extract the parton distribution functions and so learn something about the structure of the proton.* A set of distribution functions, chosen to fit all available data, is shown in Fig. 17.6. Since all of these distributions, especially those for antiquarks, peak sharply at small x, we have plotted $xf_f(x)$ for each species. As we remarked in Chapter 14, a small violation of Bjorken scaling is observed experimentally, so that these distribution functions change slowly with Q^2. The figure shows these functions at $Q^2 = 4$ GeV^2. We will see in Section 17.5 that this violation of Bjorken scaling is an effect of higher-order corrections in QCD; we will also argue that the measurement of this scaling violation allows one to determine the parton distribution function for gluons, $f_g(x)$. Anticipating this result, we have also plotted this function in the figure. Not surprisingly, one finds that the u and d quarks are most likely to carry a substantial fraction of the proton's momentum, while antiquarks and gluons tend to have small longitudinal fractions.

Since the parton distributions are the probabilities of finding various proton constituents, they must be normalized in a way that reflects the quantum numbers of the proton. The proton is a bound state of uud, plus some admixture of quark-antiquark pairs. Thus it should contain an excess of two u quarks and one d quark over the corresponding antiquarks. These considerations imply the constraints

$$\int_0^1 dx\big[f_u(x) - f_{\bar{u}}(x)\big] = 2, \qquad \int_0^1 dx\big[f_d(x) - f_{\bar{d}}(x)\big] = 1. \tag{17.36}$$

So far we have discussed the parton distributions only for the proton. Similar considerations, however, apply to any other hadron. Each hadron has its own set of parton distribution functions; these obey sum rules analogous to (17.36) but reflecting the particular quantum numbers of the hadron. The parton distribution functions should also reflect the symmetries that link different hadrons. For example, since the neutron can be generated, to a few percent accuracy, by interchanging the role of u and d quarks in the proton, its distribution functions obey

$$f_u^n(x) = f_d(x), \qquad f_d^n(x) = f_u(x), \qquad f_{\bar{u}}^n(x) = f_{\bar{d}}, \qquad \text{etc.} \tag{17.37}$$

In these equations, and henceforth, a distribution function without a special label refers to the proton. The parton distribution functions of the antiproton are given by the exact relations

$$f_u^{\bar{p}}(x) = f_{\bar{u}}(x), \qquad f_{\bar{u}}^{\bar{p}}(x) = f_u(x), \qquad \textit{etc.} \tag{17.38}$$

*A detailed discussion of the extraction of parton distribution functions from data can be found in G. Sterman, et. al., *Rev. Mod. Phys.* **67**, 157 (1995).

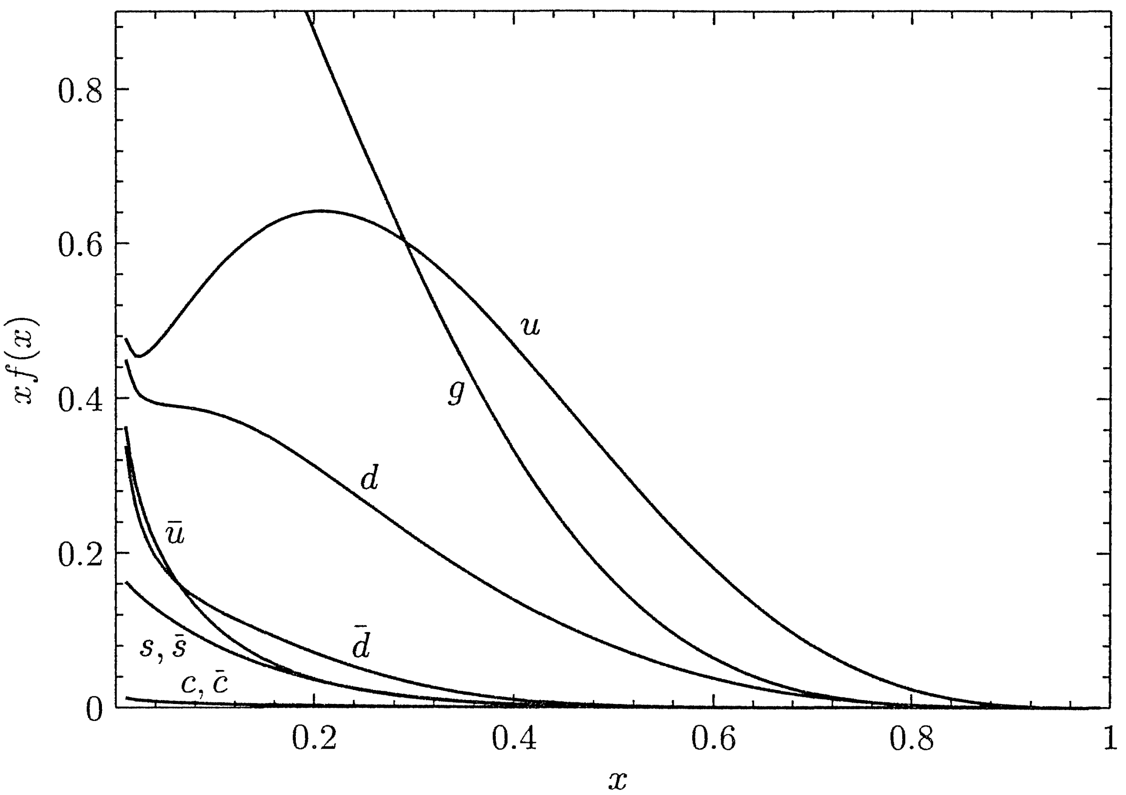


Figure 17.6. Parton distribution functions $xf_f(x)$ for quarks, antiquarks, and gluons in the proton, at $Q^2 = 4$ GeV^2. These distributions are obtained from a fit to deep inelastic scattering data performed by the CTEQ collaboration (CTEQ2L), described in J. Botts, et. al., *Phys. Lett.* **B304**, 159 (1993).

In any case, the total amount of momentum carried by the partons must sum to the total momentum of the hadron. This implies

$$\int_0^1 dx\, x\big[f_u(x) + f_d(x) + f_{\bar{u}}(x) + f_{\bar{d}}(x) + f_g(x)\big] = 1. \tag{17.39}$$

The distribution functions of quarks and antiquarks in the proton, as extracted from deep inelastic scattering data, contribute only about *half* of the total value required for this integral. The remaining energy-momentum must be carried by the gluons.

17.4 Hard-Scattering Processes in Hadron Collisions

If one collides hadrons with other hadrons at very high energy, most of the collisions will involve only soft interactions of the consituent quarks and gluons. Such interactions cannot be treated using perturbative QCD, because α_s is large when the momentum transfer is small. In some collisions, however, two quarks or gluons will exchange a large momentum $p_\perp$ perpendicular to the collision axis. Then, as in deep inelastic scattering, the elementary interaction takes place very rapidly compared to the internal time scale of the hadron wavefunctions, so the lowest-order QCD prediction should accurately describe the process. Again, we should find a parton-model formula that is built from a leading-order subprocess cross section, integrated with parton distribution functions. For the case of proton-proton scattering, these functions will be the same ones that are measured in lepton-proton deep inelastic scattering.

For example, if the hard parton-level process involves quark-antiquark scattering into a final state Y, the leading-order QCD prediction takes the form

$$\begin{aligned}\sigma(p(P_1) + p(P_2) \to Y + X)& \\ = \int_0^1 dx_1 \int_0^1 dx_2 \sum_f f_f(x_1) f_{\bar{f}}(x_2) \cdot \sigma(q_f(x_1 P) + \bar{q}_f(x_2 P) \to Y),&\end{aligned} \tag{17.40}$$

where the sum runs over all species of quarks and antiquarks—u, d, $\bar{u}$, $\bar{d}$, ... (Here again, X denotes any hadronic final state.) The same formula, with appropriately modified distribution functions, applies to any other hadron-hadron collision. This formula will be a good first approximation if, by some invariant measure, a large momentum is transfered in the $q\bar{q}$ reaction. In this section we will discuss several examples of processes of this type.

Lepton Pair Production

The simplest example to analyze is the reaction in which a high-mass lepton pair $\ell^+\ell^-$ emerges from $q\bar{q}$ annihilation in a proton-proton collision. This reaction, called the *Drell-Yan process*, is illustrated in Fig. 17.7. In this case, the underlying $q\bar{q}$ reaction is described by an elementary QED cross section. To the leading order in QCD, the cross section that we require, $\sigma(q\bar{q} \to \ell^+\ell^-)$, is simply related to the cross section $\sigma(e^+e^- \to q\bar{q})$ given in (17.4). The only difference between the two calculations is that we must average rather than sum over the color orientations of the quark and antiquark. This gives two extra factors of 1/3. Thus,

$$\sigma(q_f \bar{q}_f \to \ell^+\ell^-) = \frac{1}{3} Q_f^2 \cdot \frac{4\pi\alpha^2}{3\hat{s}}. \tag{17.41}$$

If both final-state lepton momenta are observed, it is possible to reconstruct the total 4-momentum q of the virtual photon. It is also possible to determine the longitudinal fractions of the initial quark and antiquark, as we will now show. Let

$$M^2 = q^2 \tag{17.42}$$

be the square of the invariant mass of the Drell-Yan pair. (Do not confuse this quantity M with the renormalization scale.) If the initial partons have small transverse momentum, the transverse momentum of the virtual photon will also be small. Its longitudinal momentum, however, will in general be

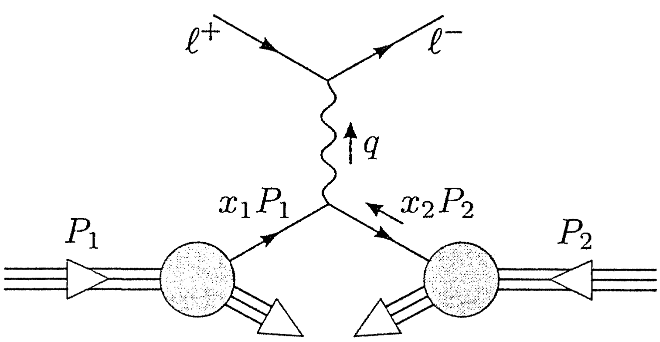


Figure 17.7. The Drell-Yan process: $pp \to \ell^+\ell^- +$ anything.

substantial. We parametrize this using the *rapidity*, Y, of the virtual photon, as defined in Eq. (3.48):

$$q^0 = M \cosh Y, \tag{17.43}$$

where q^0 is measured in the pp center of mass frame. We will express the longitudinal fractions of the quarks, and hence the Drell-Yan cross section, in terms of the observables M^2 and Y.

In the pp center of mass frame, the proton momenta take the explicit forms

$$P_1 = (E, 0, 0, E), \qquad P_2 = (E, 0, 0, -E),$$

where E satisfies $s = 4E^2$. Ignoring their small transverse momenta, we can write the constituent quark and antiquark momenta as x_1 and x_2 times these vectors, so that

$$q = x_1 P_1 + x_2 P_2 = \big((x_1{+}x_2)E, 0, 0, (x_1{-}x_2)E\big). \tag{17.44}$$

By computing the invariant square of this vector we find

$$M^2 = x_1 x_2 s. \tag{17.45}$$

Similarly, comparing (17.43) with (17.44), we find

$$\cosh Y = \frac{x_1 + x_2}{2\sqrt{x_1 x_2}} = \frac{1}{2}\left(\sqrt{\frac{x_1}{x_2}} + \sqrt{\frac{x_2}{x_1}}\right),$$

which implies

$$\exp Y = \sqrt{\frac{x_1}{x_2}}. \tag{17.46}$$

These equations can be inverted to determine x_1 and x_2:

$$x_1 = \frac{M}{\sqrt{s}} e^{Y}, \qquad x_2 = \frac{M}{\sqrt{s}} e^{-Y}. \tag{17.47}$$

Relations (17.45) and (17.46) let us convert the integral in Eq. (17.40) into an integral over the parameters M^2, Y of the produced leptons. The Jacobian of the change of variables is

$$\frac{\partial(M^2, Y)}{\partial(x_1, x_2)} = \begin{vmatrix} x_2 s & x_1 s \\ 1/2x_1 & -1/2x_2 \end{vmatrix} = s = \frac{M^2}{x_1 x_2}.$$

The cross section for lepton pair production is therefore

$$\frac{d^2\sigma}{dM^2dY}(pp \to e^+e^- + X) = \sum_f x_1 f_f(x_1)\, x_2 f_{\bar{f}}(x_2) \cdot \frac{1}{3}Q_f^2 \cdot \frac{4\pi\alpha^2}{3M^4}, \tag{17.48}$$

where x_1 and x_2 are given by Eq. (17.47). It is remarkable that the cross section for the Drell-Yan process is determined point by point by information derivable from deep inelastic scattering. Unfortunately, the relation between the two processes implied by (17.48) receives a correction of order $\alpha_s(M)$ that turns out to be numerically large, and which must be included to check this prediction against experiment.

General Kinematics of Pair Production

In deriving Eq. (17.48), we used the total cross section (17.41) for the parton-level process, integrated over the angular distribution of the outgoing leptons. In principle, we could have retained the angular information and derived a triply differential distribution. This would be the most complete prediction possible for a two-body parton-level reaction. It will be useful to work out the kinematics of such reactions, taking a more general viewpoint. In the generic situation, a parton of type 1 from proton 1 scatters from a parton of type 2 from proton 2, yielding partons of types 3 and 4, with a squared momentum transfer $\hat{t}$. This generic process is shown in Fig. 17.8. In the Drell-Yan process, partons 3 and 4 are leptons. But these partons could also be quarks or gluons, which materialize as hadronic jets. We assume that all partons can be treated as massless. In parton variables, the cross section for this process is

$$\frac{d^3\sigma}{dx_1 dx_2 d\hat{t}}(pp \to 3 + 4 + X) = f_1(x_1) f_2(x_2) \frac{d\sigma}{d\hat{t}}(1 + 2 \to 3 + 4). \tag{17.49}$$

Let us now translate this formula to observable parameters of the final state.

In the leading order of QCD, the transverse momenta of partons 3 and 4 must be equal and opposite, but their longitudinal momenta are not constrained. We will take the three parameters of the final state to be the common magnitude of the parton transverse momenta $p_\perp$ and the *longitudinal rapidities* y_3, y_4 of the final-state partons, defined by the formulae

$$E_i = p_\perp \cosh y_i; \qquad p_{i\|} = p_\perp \sinh y_i. \tag{17.50}$$

The longitudinal rapidity y_i gives the boost of the particle i from the frame where it has zero longitudinal momentum.† Recall from Section 3.3 that rapidities simply add under collinear boosts. The transverse momentum is invariant under longitudinal boosts. Thus, $(y_3, y_4, p_\perp)$ is a set of variables with convenient Lorentz transformation properties with respect to boosts along the

†In the literature on hadron collisions, y_i is usually called simply the rapidity, with the restriction to longitudinal boosts being understood.

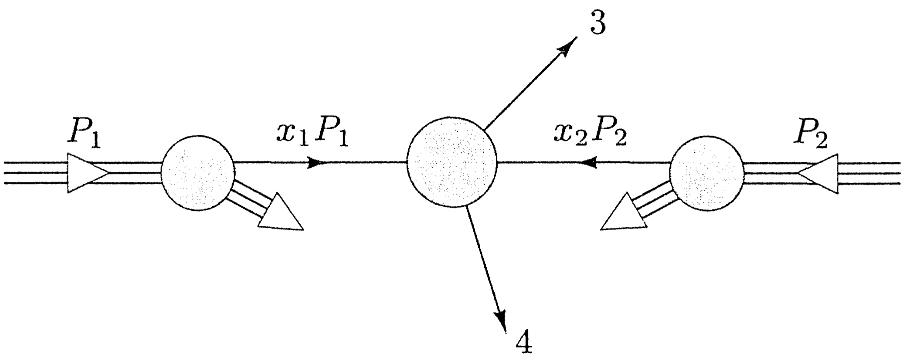


Figure 17.8. A generic two-body parton scattering process.

collision axis. We will now see that these three parameters are related in a straightforward way to the underlying variables x_1, x_2, $\hat{t}$.

Consider the center of mass frame of the colliding partons. The total energy in this frame is $\sqrt{\hat{s}}$. Let us use a subscript $*$ to denote other quantities measured in this frame, for instance, θ_* for the parton scattering angle. Then

$$p_{3\|*} = \tfrac{1}{2}\sqrt{\hat{s}}\cos\theta_*, \qquad p_{3\perp*} = \tfrac{1}{2}\sqrt{\hat{s}}\sin\theta_*, \tag{17.51}$$

and p_{4*} is oriented just oppositely. This frame is also the center of mass frame of partons 3 and 4, so

$$y_{3*} = -y_{4*} \equiv y_*. \tag{17.52}$$

Since rapidities transform by shifts, we can solve for y_* and for the rapidity Y by which one must boost to reach this frame:

$$y_* = \tfrac{1}{2}(y_3 - y_4), \qquad Y = \tfrac{1}{2}(y_3 + y_4). \tag{17.53}$$

The scattering angle θ_* is determined from y_* by combining (17.51) with the relation $E_* = p_\perp \cosh y_*$:

$$\frac{1}{\sin\theta_*} = \cosh y_*. \tag{17.54}$$

Then the Mandelstam variables

$$\hat{s} = \frac{4p_\perp^2}{\sin^2\theta_*}, \qquad \hat{t} = -\tfrac{1}{2}\hat{s}(1-\cos\theta_*) \tag{17.55}$$

can be expressed as

$$\hat{s} = 4p_\perp^2\cosh^2 y_*, \qquad \hat{t} = -2p_\perp^2\cosh y_*\, e^{-y_*}. \tag{17.56}$$

We can combine the first of these expressions with (17.47) to determine x_1 and x_2:

$$x_1 = \frac{2p_\perp}{\sqrt{s}}\cosh y_*\, e^{Y}, \qquad x_2 = \frac{2p_\perp}{\sqrt{s}}\cosh y_*\, e^{-Y}. \tag{17.57}$$

To translate the cross section (17.49) to the final parton observables, we need the Jacobian

$$\frac{\partial(x_1, x_2, \hat{t})}{\partial(y_3, y_4, p_\perp)} = \frac{8p_\perp^3}{s}\cosh^2 y_* = \frac{2p_\perp\hat{s}}{s}. \tag{17.58}$$

Multiplying Eq. (17.49) by this factor gives

$$\frac{d^3\sigma}{dy_3 dy_4 dp_\perp} = f_1(x_1) f_2(x_2) \frac{2p_\perp \hat{s}}{s} \frac{d\sigma}{d\hat{t}}(1+2\to 3+4). \tag{17.59}$$

This can be simplified a bit using the relations $\hat{s} = x_1 x_2 s$ and $p_\perp dp_\perp = d^2p_\perp/2\pi$, yielding the final result:

$$\frac{d^4\sigma}{dy_3 dy_4 d^2p_\perp} = x_1 f_1(x_1)\, x_2 f_2(x_2)\, \frac{1}{\pi}\frac{d\sigma}{d\hat{t}}(1+2\to 3+4). \tag{17.60}$$

In this formula, x_1, x_2, and the Mandelstam variables of the parton subprocess are given by Eqs. (17.57) and (17.56).

This result gives us the complete distribution of final-state leptons or jets in any two-body reaction of partons. For example, to find the distribution of final-state leptons in the Drell-Yan process, we would insert into this formula the differential cross section for quark annihilation into leptons,

$$\frac{d\sigma}{d\hat{t}}(q_f\bar{q}_f \to \ell^+\ell^-) = \frac{1}{3}Q_f^2 \cdot \frac{2\pi\alpha^2}{\hat{s}^2}\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}. \tag{17.61}$$

The formula can be applied equally well to other two-body parton reactions, if we know the relevant parton-level differential cross sections.

Jet Pair Production

The most common two-body parton reactions are those of QCD, involving quarks, gluons, or both. Unfortunately, it is very difficult to distinguish hadronic jets initiated by gluons from those initiated by quarks. It is even more difficult to determine experimentally whether the initial partons in a hard-scattering process were quarks or gluons. Thus, the predictions of QCD for hard-scattering processes are most often quoted as cross sections for jet production in hadronic collisions, summing over all possible reactions of quarks, antiquarks, and gluons. In any event, to derive these predictions, we must work out the basic parton-parton cross sections.

The simple two-body scattering processes of quarks, antiquarks, and gluons are the elementary processes of QCD perturbation theory, in the same sense that the reactions studied in Chapter 5 are the elementary processes of QED perturbation theory. They are the basic hadronic hard-scattering reactions that appear in QCD at the leading order in α_s. In the remainder of this section, we will write down the cross section formulae for the various possible quark and gluon subprocesses. All of these cross sections will be of order α_s^2. In practice, this α_s should be evaluated at a typical momentum transfer of the reaction, for example, at $Q^2 = \hat{t}$.

The simplest subprocess is the scattering of different species of quarks, for example, $u + d \to u + d$. At order α_s^2, this process occurs via the Feynman diagram shown in Fig. 17.9. This process is analogous to electron-muon

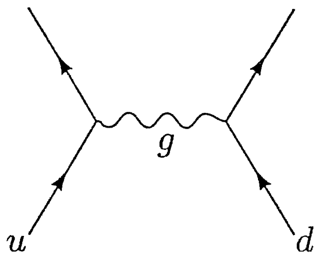


Figure 17.9. Feynman diagram contributing to $ud \to ud$.

scattering in QED, for which we wrote the cross section in Eq. (17.22):

$$\frac{d\sigma}{d\hat{t}}(e^-\mu \to e^-\mu) = \frac{2\pi\alpha^2}{\hat{s}^2}\left[\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}\right]. \tag{17.62}$$

To convert this to the cross section for quark scattering in QCD, we need only replace the QED coupling e^2 by g^2 times an $SU(3)$ group theory factor. The QCD diagram contains the factor

$$(t^a)_{i'i}(t^a)_{j'j},$$

where i, i' are the initial and final colors of the u quark and j, j' are the initial and final colors of the d quark. To compute the cross section, we must square this factor, sum over final colors, and average over initial colors. This gives the factor

$$\frac{1}{3}\cdot\frac{1}{3}\cdot \mathrm{tr}[t^b t^a]\cdot \mathrm{tr}[t^b t^a] = \frac{1}{9}\big[C(r)\big]^2\delta^{ab}\delta^{ab} = \frac{1}{9}\cdot\frac{1}{4}\cdot 8 = \frac{2}{9}, \tag{17.63}$$

where we have used Eq. (15.78) and $C(r) = 1/2$ for the fundmental representation of $SU(3)$. Thus for ud scattering,

$$\frac{d\sigma}{d\hat{t}}(ud \to ud) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}\right]. \tag{17.64}$$

The same formula applies for the scattering of any two different quarks, or, by crossing, to the scattering of a quark and an antiquark of a different species. Crossing from the t to the s channel gives the cross section for $q\bar{q}$ annihilation into a different species:

$$\frac{d\sigma}{d\hat{t}}(u\bar{u} \to d\bar{d}) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}\right]. \tag{17.65}$$

The scattering of a quark with an antiquark of the same species is more complicated, since now there are two Feynman diagrams, shown in Fig. 17.10, which interfere with one another. The analogous QED process is Bhabha scattering, $e^+e^- \to e^+e^-$, for which we worked out the cross section in Problem 5.2:

$$\frac{d\sigma}{d\hat{t}}(e^+e^- \to e^+e^-) = \frac{2\pi\alpha^2}{\hat{s}^2}\left[\left(\frac{\hat{s}}{\hat{t}}\right)^2 + \left(\frac{\hat{t}}{\hat{s}}\right)^2 + \hat{u}^2\left(\frac{1}{\hat{s}}+\frac{1}{\hat{t}}\right)^2\right]. \tag{17.66}$$

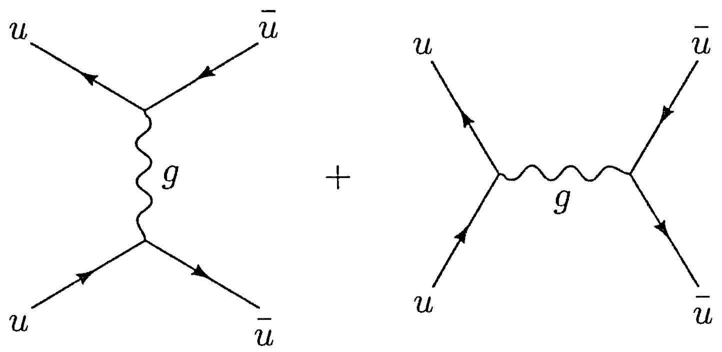


Figure 17.10. Feynman diagrams contributing to $u\bar{u} \to u\bar{u}$.

However, it is not quite straightforward to transcribe this to QCD, because different terms receive different color factors.

This process is most easily analyzed using initial and final states of definite helicity. For massless fermions, helicity is conserved, so the reaction $e^+_R e^-_L \to e^+_L e^-_R$ can receive a contribution only from the s-channel diagram, while $e^+_R e^-_R \to e^+_R e^-_R$ can receive a contribution only from the t-channel diagram. The corresponding cross sections are

$$\begin{aligned}\frac{d\sigma}{d\hat{t}}(e^+_R e^-_L \to e^+_L e^-_R) &= \frac{4\pi\alpha^2}{\hat{s}^2}\left(\frac{\hat{t}}{\hat{s}}\right)^2, \\ \frac{d\sigma}{d\hat{t}}(e^+_R e^-_R \to e^+_R e^-_R) &= \frac{4\pi\alpha^2}{\hat{s}^2}\left(\frac{\hat{s}}{\hat{t}}\right)^2.\end{aligned} \tag{17.67}$$

The cross section for $e^+_R e^-_L \to e^+_L e^-_L$ must vanish. The fourth possible process involving e^+_R receives contributions from both s- and t-channel diagrams. Computing this contribution explicitly, one finds

$$\frac{d\sigma}{d\hat{t}}(e^+_R e^-_L \to e^+_R e^-_L) = \frac{4\pi\alpha^2}{\hat{s}^2}\hat{u}^2\left(\frac{1}{\hat{t}} + \frac{1}{\hat{s}}\right)^2; \tag{17.68}$$

the cross term in the square is the interference term between the two diagrams. The invariance of QED under parity implies that the values of all of these cross sections remain identical when all helicities are reversed. It is easy to check that the spin-averaged cross section is indeed given by (17.66).

To convert Eq. (17.66) to a QCD cross section averaged over colors, we can assign the color factor (17.63) to the square of any individual diagram. However, the cross term between the two diagrams in Fig. 17.10 receives a different color factor:

$$\left(\frac{1}{3}\right)^2 \cdot (t^a)_{i'i}(t^a)_{jj'} \cdot (t^a)_{j'i'}(t^b)_{ij} = \frac{1}{9}\,\mathrm{tr}[t^a t^b t^a t^b]. \tag{17.69}$$

To evaluate this factor, we can make use of Eq. (16.79):

$$t^a t^b t^a t^b = \Big(C_2(r) - \frac{1}{2}C_2(G)\Big)t^a t^a = \left(\frac{4}{3} - \frac{3}{2}\right)\frac{4}{3} = -\frac{2}{9}.$$

So the color factor (17.69) equals $-2/27$.

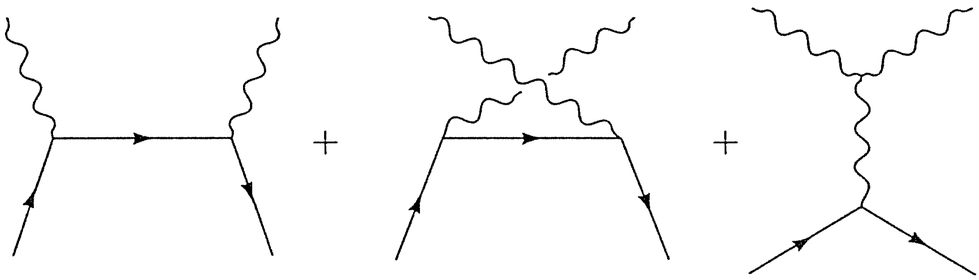


Figure 17.11. Feynman diagrams contributing to $q\bar{q} \to gg$.

Assembling the color factors and the helicity cross sections, we find the following result for the $u\bar{u}$ scattering cross section:

$$\frac{d\sigma}{d\hat{t}}(u\bar{u} \to u\bar{u}) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2} + \frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2} - \frac{2}{3}\frac{\hat{u}^2}{\hat{s}\hat{t}}\right]. \tag{17.70}$$

By crossing between the s and u channels, we find the corresponding cross section for $uu \to uu$:

$$\frac{d\sigma}{d\hat{t}}(uu \to uu) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2} + \frac{\hat{t}^2 + \hat{s}^2}{\hat{u}^2} - \frac{2}{3}\frac{\hat{s}^2}{\hat{u}\hat{t}}\right]. \tag{17.71}$$

The process $\bar{u}\bar{u} \to \bar{u}\bar{u}$ has the same cross section. This completes our catalogue of cross sections for the scattering of quarks and antiquarks.

We turn next to processes that involve both quarks and gluons. We will begin with the reaction $q\bar{q} \to gg$. This is the analogue of the QED annihilation of e^+e^- to $\gamma\gamma$, discussed in Section 5.5. The QED cross section is

$$\frac{d\sigma}{d\hat{t}}(e^+e^- \to \gamma\gamma) = \frac{2\pi\alpha^2}{\hat{s}^2}\left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}}\right]. \tag{17.72}$$

Since the photons are identical particles, this expression should be integrated over only half of the 4π solid angle.

The QCD reaction is considerably more complicated. As we saw in Section 16.1, this process receives contributions from three Feynman diagrams, shown in Fig. 17.11. These contributions must be summed over the transverse polarization states of the gluons. If one chooses instead to evaluate the sum over gluon polarizations by the replacement

$$\sum_{\epsilon} \epsilon^{\mu}\epsilon^{*\nu} \to -g^{\mu\nu}, \tag{17.73}$$

we saw in Section 16.3 that one must also include the (negative) cross section for $q\bar{q}$ annihilation to a ghost-antighost pair.

The leading behavior of the $q\bar{q} \to gg$ cross section as $\hat{t}$ or $\hat{u} \to 0$ is not so hard to evaluate. In either case, only the single diagram with the corresponding kinematic singularity contributes. The color factor associated with the square of either of these diagrams is the square of

$$(t^a)_{ij}(t^b)_{jk},$$

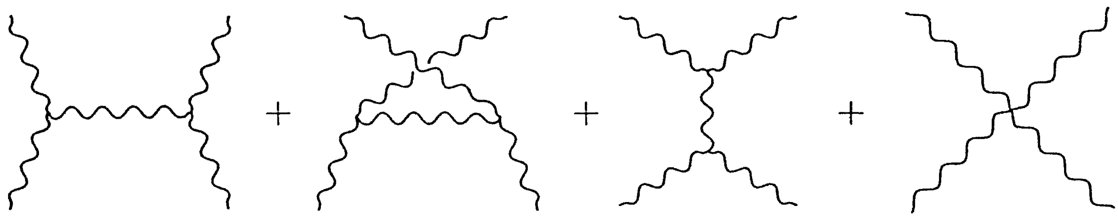


Figure 17.12. Feynman diagrams contributing to $gg \to gg$.

summed over the gluon colors a, b and averaged over the q and $\bar{q}$ colors i, k. This gives

$$\left(\frac{1}{3}\right)^2 \cdot \mathrm{tr}[t^a t^b t^b t^a] = \frac{1}{9} \cdot 3\big(C_2(r)\big)^2 = \frac{16}{27}. \tag{17.74}$$

Thus the most singular terms are given by the QED result, with α replaced by α_s, multiplied by 16/27. The complete evaluation of the cross section is left for Problem 17.3; the result is

$$\frac{d\sigma}{d\hat{t}}(q\bar{q} \to gg) = \frac{32\pi\alpha_s^2}{27\hat{s}^2}\left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}} - \frac{9}{4}\left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\right)\right]. \tag{17.75}$$

The cross sections for the remaining quark-gluon processes can be obtained from this result by crossing. The result for the inverse reaction $gg \to q\bar{q}$ involves the same squared matrix element as (17.75); the only difference is that we average over gluon rather than quark colors, giving a relative factor of $(3/8)^2$. Thus,

$$\frac{d\sigma}{d\hat{t}}(gg \to q\bar{q}) = \frac{\pi\alpha_s^2}{6\hat{s}^2}\left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}} - \frac{9}{4}\left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\right)\right]. \tag{17.76}$$

For the reaction $qg \to qg$, cross the s and t channels in Eq. (17.75) and multiply by 3/8 since there is one gluon color average. This gives

$$\frac{d\sigma}{d\hat{t}}(qg \to qg) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[-\frac{\hat{u}}{\hat{s}} - \frac{\hat{s}}{\hat{u}} + \frac{9}{4}\left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2}\right)\right]. \tag{17.77}$$

The cross section for $\bar{q}g \to \bar{q}g$ is identical.

The final elementary process of QCD is gluon-gluon scattering. This has no QED analogue, and is rather tedious to evaluate. There are four leading-order diagrams, shown in Fig. 17.12. We discuss this process also in Problem 17.3. The final result for the spin- and color-averaged cross section is

$$\frac{d\sigma}{d\hat{t}}(gg \to gg) = \frac{9\pi\alpha_s^2}{2\hat{s}^2}\left[3 - \frac{\hat{t}\hat{u}}{\hat{s}^2} - \frac{\hat{s}\hat{u}}{\hat{t}^2} - \frac{\hat{s}\hat{t}}{\hat{u}^2}\right]. \tag{17.78}$$

The various parton cross sections listed in this section can be combined with the parton distribution functions to predict the cross section for jet production in hadron-hadron collisions. As an example, we show in Fig. 17.13 a comparison of the invariant mass ($\hat{s}$) distribution predicted for parton-parton

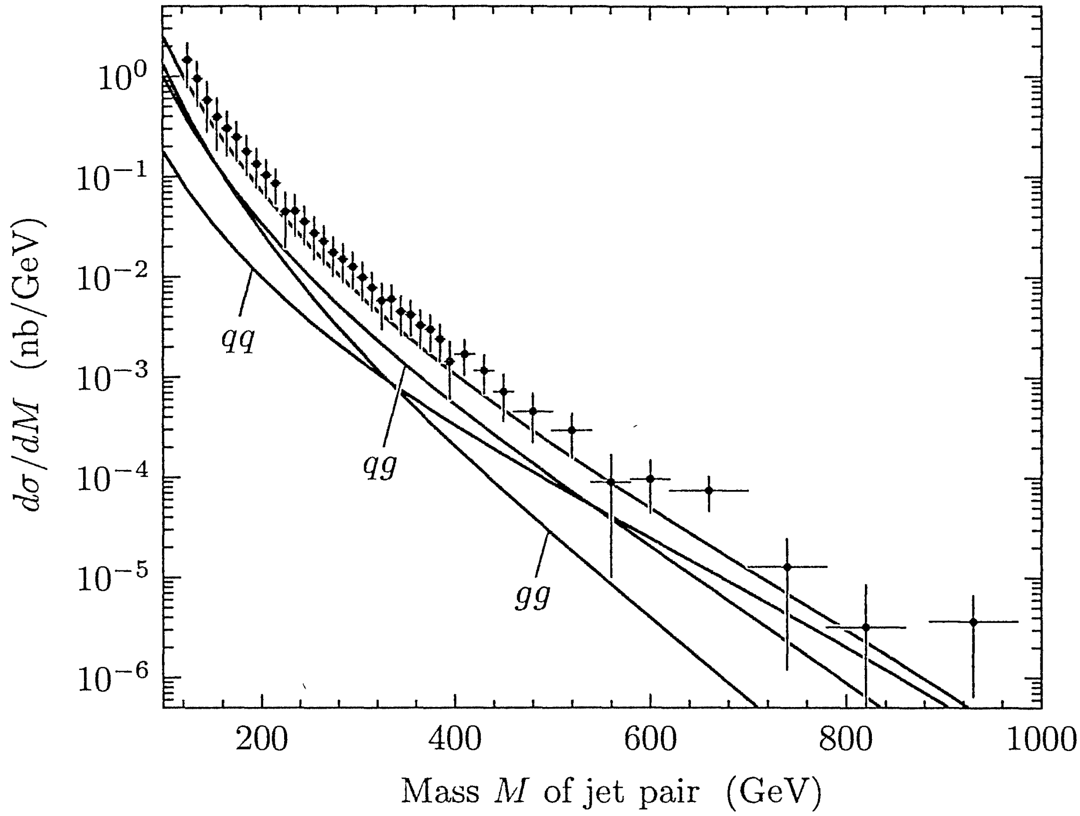


Figure 17.13. Two-jet invariant mass distribution in $p\bar{p}$ collisions at $E_{\rm cm} =$ 1.8 TeV, as measured by the CDF collaboration, F. Abe, et. al., *Phys. Rev.* **D48**, 998 (1993). The measurement is compared to a leading-order QCD calculation using the CTEQ structure functions described in Fig. 17.6. The three lower curves show the invariant mass distributions for the three components of the theoretical prediction: quark-quark (and antiquark) scattering, quark-gluon scattering, and gluon-gluon scattering.

scattering with the invariant mass distribution of two-jet events observed in high-energy $p\bar{p}$ collisions. The overall normalization of the theoretical prediction is uncertain by about a factor of 2 due to the ambiguity of the choice of Q^2 used to evaluate $\alpha_s(Q^2)$ in the parton cross sections, and due to similar ambiguities in deriving parton distributions from deep inelastic scattering cross sections. This uncertainty is reduced to about 30% when corrections of order α_s are included. Still, it is remarkable that the lowest-order QCD prediction tracks the observed distribution as a function of the two-jet invariant mass as it falls by six orders of magnitude. Thus, for the jet production cross section, as for hard processes involving leptons, QCD indeed gives a reasonable description of the behavior of the strong interactions at large momentum transfer.

17.5 Parton Evolution

Now that we have examined the predictions of QCD at the leading order for several strong interaction processes, we should investigate the corrections to these predictions at the next order in α_s. We saw in Section 17.2 that the corrections from individual diagrams may contain mass singularities, singularities associated with collinear emission processes which appear in the limit of zero mass. For the process of e^+e^- annihilation to hadrons, we saw that these mass singularities, and the infrared divergences from soft gluon emission, cancel in the expression for the total cross section. It can be shown that this is a general feature of processes in which quarks and gluons are produced in the collision of leptons or photons. However, when quarks or gluons appear in the initial state of a parton subprocess, the corrections to the process will, in general, have mass singularities that do not cancel. In this section we will demonstrate this effect and work out its physical interpretation. We will find that these singular terms predict a violation of Bjorken scaling by terms depending on the logarithm of the momentum scale. In fact, they lead to a precise set of differential equations that govern the momentum dependence of the parton distributions.

The basic phenomena associated with mass singularities in QCD are already present in the physics of collinear photon emission in QED at high energies, and so it is most straightforward to begin by studying that case. In this section, we will show that collinear photon emission leads to an analogue of a parton distribution function for the electron. We will derive a differential equation describing this distribution function, first constructed by Gribov and Lipatov. Finally, we will generalize this equation to QCD, following the construction of Altarelli and Parisi.‡

In Chapters 5 and 6, we studied several examples of QED processes that involved diagrams with t- or u-channel singularities. In these cases, we found that the total cross section was generally enhanced by an extra factor $\log(s/m^2)$ in the high-energy limit. For example, in Eq. (5.95) we saw that the u-channel exchange diagram in Compton scattering, Fig. 17.14(a), leads to an integral that, in the high-energy limit, takes the form

$$\int \frac{d\cos\theta}{(1+\cos\theta)}.$$

The singularity as $\cos\theta \to -1$ is cut off by the electron mass, leading to the logarithmic enhancement factor. Thus the collinear photon emission costs a factor that is not α but rather $\alpha \log(s/m^2)$. Emission of multiple collinear photons, as in Fig. 17.14(b), gives contributions of order $(\alpha \log(s/m^2))^n$. To improve the accuracy of perturbation theory, it would be useful to find a procedure for summing these terms to all orders in α. In QCD, the corresponding

‡V. N. Gribov and L. Lipatov, *Sov. J. Nucl. Phys.* **15**, 438 (1972); G. Altarelli and G. Parisi, *Nucl. Phys.* **B126**, 298 (1977). We also strongly recommend reading the papers of J. Kogut and L. Susskind, *Phys. Rev.* **D9**, 697, 3391 (1974).

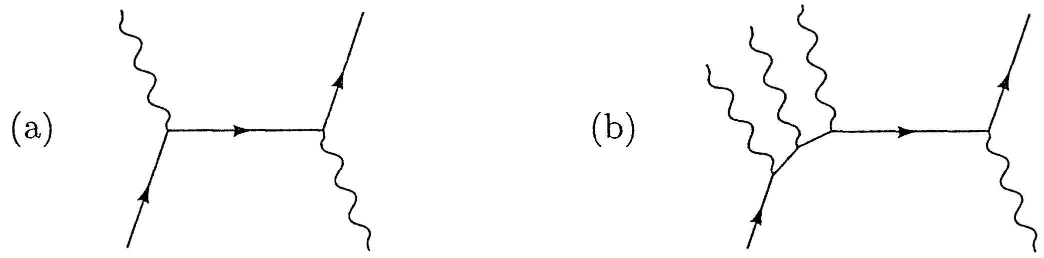


Figure 17.14. Diagrams with mass singularities associated with collinear photon emission: (a) leading order; (b) higher order.

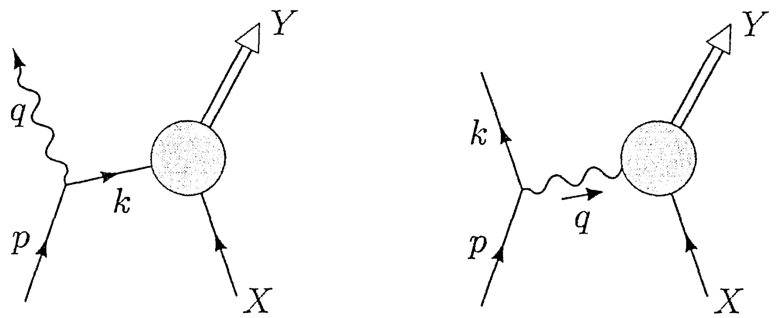


Figure 17.15. General form of diagrams with mass singularities in QED.

factor for collinear gluon emission would be

$$\alpha_s(Q^2)\log\frac{Q^2}{\mu^2},$$

where μ is the momentum scale where nonperturbative QCD effects become important. Comparing with Eq. (17.17), we see that this product is of order 1. Thus, in this case, the resummation of large logarithms is essential if we are to make any quantitative predictions.

In QED, diagrams with mass singularities associated with one collinear emission are of one of the forms shown in Fig. 17.15. In each case, the circle represents a scattering process with large momentum tranfer. The mass singularity appears when the denominator of the intermediate propagator vanishes, that is, when the intermediate state is almost on-shell. Thus, it is natural to consider the first diagram in Fig. 17.15 to be a transition to a real photon and an almost-real electron, followed by the interaction of the electron with the remaining particles in the amplitude. The second diagram should have a similar interpretation with an almost-real photon in the intermediate state.

The only subtlety comes in defining the polarization of the intermediate-state particle. For the case of an intermediate-state electron, the numerator of the propagator is

$$\not{k} = \sum_s u^s(k)\bar{u}^s(k). \tag{17.79}$$

Thus, when $k^2 \to 0$, the photon emission vertex and the remaining part of the amplitude are contracted with on-shell polarization spinors for a massless electron. The analogous statement for the diagram with the photon in the intermediate state would be that the electron emission vertex and the remaining photon amplitude should be contracted with physical transverse polarization vectors for the intermediate-state photon. Since the numerator of the photon propagator is $g^{\mu\nu}$, it is not obvious that the photon propagator reduces in this way. But it is true. To see this, use the expansion for $g^{\mu\nu}$ in terms of massless polarization vectors given in Eq. (16.20):

$$g^{\mu\nu} = \epsilon_+^\mu \epsilon_-^{\nu *} + \epsilon_-^\mu \epsilon_+^{\nu *} - \sum_i \epsilon_{Ti}^\mu \epsilon_{Ti}^{\nu *}. \tag{17.80}$$

Here ϵ_{Ti}^μ are transverse polarization vectors. The forward polarization vector ϵ_+^μ is proportional to the photon momentum q^μ. When we contract ϵ_+^μ with the QED scattering amplitude on the right, we will obtain zero by the Ward identity, and the contraction of $\epsilon_+^{*\nu}$ with the electron emission vertex similarly gives zero. Thus, for the purpose of computing the singular term as the photon momentum q goes on-shell, we may replace

$$\frac{-ig^{\mu\nu}}{q^2} \to \frac{+i}{q^2} \sum_i \epsilon_{Ti}^\mu \epsilon_{Ti}^{\nu *} \tag{17.81}$$

and evaluate the photon emission and absorption amplitudes with transverse polarization vectors.

Matrix Element for Electron Splitting

By replacing the numerator of the intermediate propagator with a sum over polarization vectors, we decouple the photon or electron emission vertex from the rest of the diagram. We will now evaluate this vertex explicitly between physical polarization states of massless particles. The kinematics is shown in Fig. 17.16. The two final particles should be almost collinear, with a small relative transverse momentum. We can choose the incident electron momentum to lie along the $\hat{3}$ axis and the outgoing momenta to lie in the $\hat{1}$-$\hat{3}$ plane. Let z be the fraction of the energy of the initial electron that is carried off by the photon. Then the three 4-momenta can be written as

$$\begin{aligned} p &= (p, 0, 0, p), \\ q &\approx (zp, p_\perp, 0, zp), \\ k &\approx ((1-z)p, -p_\perp, 0, (1-z)p). \end{aligned} \tag{17.82}$$

These three vectors satisfy $p^2 = q^2 = k^2 = 0$, up to terms of order $p_\perp^2$.

In the process where a real photon is emitted, we should have p^2 and q^2 exactly zero, and k^2 slightly off-shell by an amount of order $p_\perp^2$. We will need to know the value of k^2, which appears in the virtual electron propagator. So let us modify Eqs. (17.82) to satisfy the condition $q^2 = 0$ up to terms of order

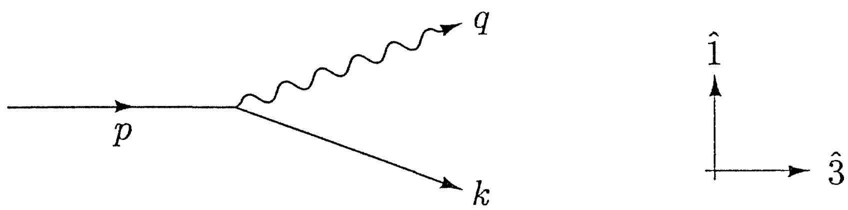


Figure 17.16. Kinematics of the vertex for emission of a collinear electron or photon.

$p_\perp^4$, rewriting q and k as

$$q = (zp, p_\perp, 0, zp - \frac{p_\perp^2}{2zp}),$$
$$k = ((1-z)p, -p_\perp, 0, (1-z)p + \frac{p_\perp^2}{2zp}). \tag{17.83}$$

With this modification,

$$k^2 = -p_\perp^2 - 2(1-z)\frac{p_\perp^2}{2z} + \mathcal{O}(p_\perp^4).$$

Thus, if the photon is real and the electron is virtual, we have

$$q^2 = 0, \qquad k^2 = -\frac{p_\perp^2}{z}. \tag{17.84}$$

Reciprocally, in the process with a real electron and a virtual photon,

$$k^2 = 0, \qquad q^2 = -\frac{p_\perp^2}{(1-z)}. \tag{17.85}$$

These more accurate expressions will be needed only in the propagator of the virtual particle. The matrix element of the electron-photon vertex begins in order $p_\perp$, so it is not significantly affected by the modification of (17.82) to (17.83), and is the same (to lowest order) no matter which particle is virtual.

We now calculate the matrix elements of the QED vertex between massless states of definite helicity. If the initial electron is left-handed, the final electron must also be left-handed, by helicity conservation. Then the photon emission vertex is given by

$$i\mathcal{M} = \bar{u}_L(k)(-ie\gamma_\mu)u_L(p)\epsilon_T^{*\mu}(q), \tag{17.86}$$

where the photon polarization vector may be either left- or right-handed. Recalling the helicity-basis expressions

$$\gamma_\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \bar{\sigma}_\mu & 0 \end{pmatrix}, \qquad u_L(p) = \sqrt{2p^0}\begin{pmatrix} \xi(p) \\ 0 \end{pmatrix} \quad (\text{for } m = 0),$$

we can write more explicitly

$$i\mathcal{M} = -ie\sqrt{2(1-z)p}\sqrt{2p}\,\xi^\dagger(k)\sigma^i\xi(p)\,\epsilon_T^{*i}(q). \tag{17.87}$$

To order $p_\perp$, the left-handed spinors are

$$\xi(p) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \xi(k) = \begin{pmatrix} p_\perp/2(1-z)p \\ 1 \end{pmatrix}. \tag{17.88}$$

The polarization vectors for the photon are

$$\epsilon_L^{*i}(q) = \frac{1}{\sqrt{2}}\left(1, i, -\frac{p_\perp}{zp}\right), \qquad \epsilon_R^{*i}(q) = \frac{1}{\sqrt{2}}\left(1, -i, -\frac{p_\perp}{zp}\right). \tag{17.89}$$

Notice that, when these vectors are contracted with the Pauli matrix in Eq. (17.87), the first two components of the right-handed polarization vector give $(\sigma^1 - i\sigma^2) = 2\sigma^-$, which annihilates $\xi(p)$. The only remaining term comes from the $i = 3$ component, and we find

$$i\mathcal{M}(e_L^- \to e_L^- \gamma_R) = ie\frac{\sqrt{2(1-z)}}{z}p_\perp. \tag{17.90}$$

For the left-handed photon polarization, there is an additional contribution from the first two components of ϵ_L^*. These add to

$$i\mathcal{M}(e_L^- \to e_L^- \gamma_L) = ie\frac{\sqrt{2(1-z)}}{z(1-z)}p_\perp. \tag{17.91}$$

Parity invariance implies that the values of the matrix elements are unchanged if all helicities are flipped; this immediately gives the required matrix elements for the case of an initial e_R^-. The squared matrix element, averaged over initial helicities, is therefore

$$\frac{1}{2}\sum_{\text{pols.}} |\mathcal{M}|^2 = \frac{2e^2p_\perp^2}{z(1-z)}\left[\frac{1 + (1-z)^2}{z}\right]. \tag{17.92}$$

The first term in the brackets comes from a photon with spin parallel to the electron spin; the second term comes from a photon with spin opposite to the electron spin.

The Equivalent Photon Approximation

Now we have all the pieces needed to compute the cross sections for the processes shown in Fig. 17.15. We first consider the process with a virtual photon. Call the initial state on the right-hand side of the diagram X and the final state Y, and let $\mathcal{M}_{\gamma X}$ represent the matrix element for the scattering of the photon from X. We will assume for simplicity that X is unpolarized, so that the scattering cross section does not depend on the virtual photon polarization. Then the complete diagram gives a cross section

$$\sigma = \frac{1}{(1+v_X)2p2E_X}\int\frac{d^3k}{(2\pi)^3}\frac{1}{2k^0}\int d\Pi_Y\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\left(\frac{1}{q^2}\right)^2|\mathcal{M}_{\gamma X}|^2, \tag{17.93}$$

where v_X is the velocity of X and $\int d\Pi_Y$ is the phase space integral over Y.

The integral has a singularity when k is collinear with the incident electron momentum p. To isolate the singularity, substitute for k^0 and q^2 from Eqs.

(17.82) and (17.85) and rewrite the integral over k as

$$d^3k = dk^3 d^2k_\perp = pdz \cdot \pi dp_\perp^2. \tag{17.94}$$

Then the cross section can be expressed as

$$\begin{aligned}\sigma &= \int \frac{pdzdp_\perp^2}{16\pi^2(1-z)p}\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\frac{(1-z)^2}{p_\perp^4}\frac{z}{(1+v_X)2zp2E_X}\int d\Pi_Y|\mathcal{M}_{\gamma X}|^2\\ &= \int \frac{dzdp_\perp^2}{16\pi^2(1-z)}\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\frac{z(1-z)^2}{p_\perp^4}\cdot\sigma(\gamma X\to Y).\end{aligned} \tag{17.95}$$

Finally, insert the spin-averaged electron emission vertex (17.92), to obtain

$$\begin{aligned}\sigma &= \int \frac{dzdp_\perp^2}{16\pi^2}\frac{z(1-z)}{p_\perp^4}\frac{2e^2p_\perp^2}{z(1-z)}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(\gamma X\to Y)\\ &= \int_0^1 dz\int\frac{dp_\perp^2}{p_\perp^2}\frac{\alpha}{2\pi}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(\gamma X\to Y).\end{aligned} \tag{17.96}$$

The integral over $p_\perp^2$ runs from momentum transfers of order s down to the electron mass m^2, which cuts off the singularity. Thus, our final result is

$$\sigma(e^-X\to e^-Y) = \int_0^1 dz\,\frac{\alpha}{2\pi}\log\frac{s}{m^2}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(\gamma X\to Y). \tag{17.97}$$

The cross section on the right-hand side is computed for a real, transversely polarized photon of momentum zp. The factor $\log(s/m^2)$ represents the mass singularity. This formula is the Weizsacker-Williams *equivalent photon approximation*, which we encountered earlier in Problems 5.5 and 6.2.

Formula (17.97) takes on a new significance when it is juxtaposed with the QCD predictions of the previous two sections. This QED formula has just the same form as a parton model expression, with the Weizsacker-Williams distribution function

$$f_\gamma(z) = \frac{\alpha}{2\pi}\log\frac{s}{m^2}\left[\frac{1+(1-z)^2}{z}\right] \tag{17.98}$$

playing the role of the probability to find a photon of longitudinal fraction z in the incident electron.

The Electron Distribution

The first diagram of Fig. 17.5, with an emitted photon and a virtual electron, can be treated in the same way. The analogue of (17.93) is

$$\sigma = \frac{1}{(1+v_X)2p2E_X}\int\frac{d^3q}{(2\pi)^3}\frac{1}{2q^0}\int d\Pi_Y\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\left(\frac{1}{k^2}\right)^2|\mathcal{M}_{e^-X}|^2.$$