Renormalization

In this chapter we face the ultraviolet divergences that we have found in perturbative quantum field theory. These divergences are not simply a technical nuicance to be disposed of and forgotten. As we will explain, they parameterize the dependence on quantum fluctuations at short distance scales (or equivalently, high momenta).

Historically, it took a long time to reach this understanding. In the 1930's, when the ultraviolet divergences were first discovered in qunatum electrodynamics, many physicists believed that fundamental principles of physics had to be changed to eliminate the divergences. In the late 1940's Bethe, Feynman, Schwinger, Tomonaga, and Dyson, and others proposed a program of 'renormalization' that gave finite and physically sensible results by absorbing the divergences into redefinitions of physical quantities. This leads to calculations that agree with experiment to 8 significant digits in QED, the most accurate calculations in all of science.

Even after the technical aspects of renormalization were understood, conceptual difficulties remained. It was widely believed that only a limited class of 'renormalizable' theories made physical sense. (The fact that general relativity is not renormalizable in this sense was therefore considered a deep problem.) Also, the renormalization program was viewed by many physicists as an *ad hoc* procedure justified only by the fact that it yields physically sensible results. This was changed by the profound work of K. Wilson in the 1970's, which laid the foundation for the modern understanding of renormalization. According to the present view, renormalization is nothing more than parameterizing the sensitivity of low-energy physics to high-energy physics. This viewpoint allows one to make sense out of 'non-renormalizable' theories as effective field theories describing physics at low energies. We now understand that even 'renormalizable' theories are effective field theories in this sense, and this viewpoint explains why nature is (approximately) described by renormalizable theories. This modern point of view is the one we will take in this chapter.

1 Renormalization in Quantum Mechanics

Ultraviolet divergences and the need for renormalization appear not only in field theory, but also in simple quantum mechanical models. We will study these first to understand these phenomena in a simpler setting, and hopefully dispell the air of mystery that often surrounds the subject of renormalization.

1.1 1 Dimension

We begin with 1-dimensional quantum mechanics, described by the Hamiltonian

$$\hat{H} = \tfrac{1}{2}\hat{p}^2 + \hat{V}. \tag{1.1}$$

We are using units where $\hbar = 1$, $m = 1$. In these units, all quantities have dimensions of length to some power. Since $\hat{p} = -id/dx$ acting on position space wavefunctions, we have dimensions

$$[p] = \frac{1}{L}, \qquad [E] = \frac{1}{L^2}. \tag{1.2}$$

Suppose that the potential $V(x)$ is centered at the origin and has a range of order a and a height of order V_0. We will focus on scattering, so we consider an incident plane wave to the left of $x = 0$ with momentum $p > 0$. That is, we assume that the position-space wavefunction has the asymptotic forms

$$\Psi(x) \to \begin{cases} Ae^{ipx} + Be^{-ipx} & x \to -\infty \\ Ce^{ipx} & x \to +\infty, \end{cases} \tag{1.3}$$

where the contributions proportional to A (B) [C] represent the incoming (reflected) [transmitted] waves (See Fig. 2.1).

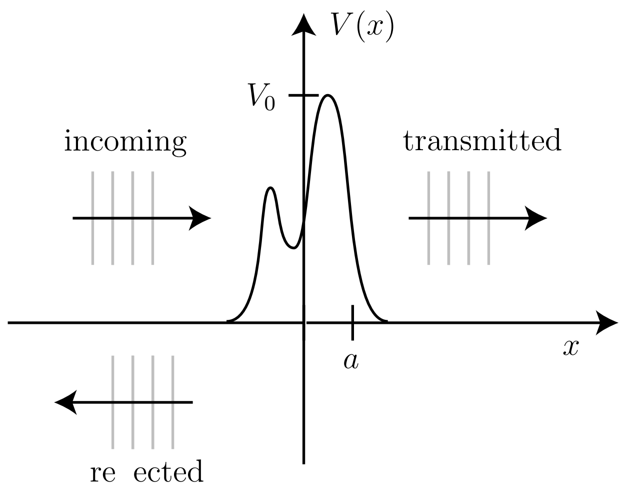


Fig. 1. Scattering from a local potential in one-dimensional quantum mechanics.

Suppose that the range of the potential a is small compared to the de Broglie wavelength $\lambda = 2\pi/p$. This means that the incoming wavefunction is approximately constant over the range of the potential, and we expect the details of the potential to be unimportant. We can then obtain a good approximation by approximating the potential by a delta function:

$$V(x) \simeq c\,\delta(x), \tag{1.4}$$

where c is a phenomenological parameter ('coupling constant') chosen to reproduce the results of the true theory. Note that $\delta(x)$ has units of $1/L$ (since $\int dx\,\delta(x) = 1$), so the dimension of c is

$$[c] = \frac{1}{L}. \tag{1.5}$$

Approximating the potential by a delta function can be justified by considering trial wavefunctions $\psi(x)$ and $\chi(x)$ that vary on a length scale $\lambda \gg a$. Consider matrix elements of the potential between such states:

$$\langle\chi|\hat{V}|\psi\rangle = \int dx\,\chi^*(x)\psi(x)V(x). \tag{1.6}$$

The wavefunctions $\chi(x)$ and $\psi(x)$ are approximately constant in the region where the potential is nonvanishing, so we can write

$$\langle\chi|\hat{V}|\psi\rangle \simeq \chi^*(0)\psi(0)\int dx\,V(x). \tag{1.7}$$

This is equivalent to the approximation Eq. (1.4) with

$$c = \int dx\,V(x). \tag{1.8}$$

We can systematically correct this approximation by expanding the wavefunctions in a Taylor series around $x = 0$:

$$f(x) \stackrel{\text{def}}{=} \chi^*(x)\psi(x) = f(0) + f'(0)x + \mathcal{O}(1/\lambda^2). \tag{1.9}$$

Substituting into Eq. (1.6), we obtain

$$\langle\chi|\hat{V}|\psi\rangle = f(0)\int dx\,V(x) + f'(0)\int dx\,xV(x) + \cdots. \tag{1.10}$$

The first few terms of this series can be reproduced by approximating the potential as

$$V(x) = c_0\delta(x) + c_1\delta'(x) + \mathcal{O}(V_0 a^2/\lambda^2), \tag{1.11}$$

where

$$c_0 = \int dx\, V(x), \quad c_1 = -\int dx\, x\, V(x), \tag{1.12}$$

etc. Note that this expansion is closely analogous to the multipole expansion in electrostatics. In that case, a complicated charge distribution can be replaced by a simpler one (monopole, dipole, ...) for purposes of finding the potential far away. In the present case, we see that multipole moments of the potential are sufficient to approximate the matrix elements of the potential for low-momentum states. This is true no matter how complicated the potential $V(x)$ is, as long as it has short range.

The solution of the scattering problem for a delta function potential is an elementary exercise done in many quantum mechanics books. The solution of the time-independent Schrödinger equation

$$-\tfrac{1}{2}\Psi''(x) + c_0\delta(x)\Psi(x) = E\Psi(x) \tag{1.13}$$

has the form

$$\Psi(x) = \begin{cases} Ae^{ipx} + Be^{-ipx} & x < 0 \\ Ce^{ipx} & x > 0, \end{cases} \tag{1.14}$$

with $p = \sqrt{2E}$. Integrating the equation over a small interval $(-\epsilon, \epsilon)$ containing the origin, we obtain

$$-\tfrac{1}{2}\left[\Psi'(\epsilon) - \Psi'(-\epsilon)\right] + c_0\Psi(0) = \mathcal{O}(\epsilon). \tag{1.15}$$

Taking $\epsilon \to 0$ gives the condition

$$\frac{ip}{2}[C - A + B] + c_0 C = 0. \tag{1.16}$$

In order for $\Psi(0)$ to be well-defined, we require that the solution is continuous at $x = 0$, which gives

$$A + B = C. \tag{1.17}$$

We can solve these equations up to the overall normalization of the wavefunction, which has no physical meaning. We obtain

$$\begin{aligned} T &\stackrel{\text{def}}{=} \frac{C}{A} = \text{transmission amplitude} \\ &= \frac{p}{p + ic_0}, \end{aligned} \tag{1.18}$$

$$\begin{aligned} R &\stackrel{\text{def}}{=} \frac{B}{A} = \text{reflection amplitude} \\ &= -\frac{ic_0}{p + ic_0}. \end{aligned} \tag{1.19}$$

Note that

$$|T|^2 + |R|^2 = 1, \tag{1.20}$$

as required by unitarity (conservation of probability).

From the discussion above, we expect this to be an accurate result for any short-range potential as long as $p \ll 1/a$. The leading behavior in this limit is

$$T \simeq -\frac{ip}{c_0}. \tag{1.21}$$

Note that T is dimensionless, so this answer is consistent with dimensional analysis. Eq. (1.21) is the 'low-energy theorem' for scattering from a short-range potential in one-dimensional quantum mechanics.

Suppose, however, that the microscopic potential is an odd function of x:

$$V(-x) = -V(x). \tag{1.22}$$

Then the first nonzero term in Eq. (1.11) is

$$V(x) \simeq c_1\, \delta'(x). \tag{1.23}$$

Note that c_1 is dimensionless. We must then solve the Schrödinger equation

$$-\tfrac{1}{2}\Psi''(x) + c_1 \delta'(x)\Psi(x) = E\Psi(x). \tag{1.24}$$

This is *not* a textbook exercise, for the very good reason that no solution exists! To see this, look at the jump condition:

$$\tfrac{1}{2}\left[\Psi'(\epsilon) - \Psi'(-\epsilon)\right] - c_1 \Psi'(0) = \mathcal{O}(\epsilon). \tag{1.25}$$

The problem is that $\Psi'(0)$ is not well-defined, because the jump condition tells us that Ψ' is discontinuous at $x = 0$.

In fact, this inconsistency is a symptom of an ultraviolet divergence precisely analogous to the ones encountered in quantum field theory. To see this, let us formulate this problem perturbatively, as we do in quantum field theory. We write the Dyson series for the interaction-picture time-evolution operator

$$\hat{U}_I(t_f, t_i) = \text{Texp}\left\{ -i \int_{t_i}^{t_f} dt\, \hat{H}_I(t) \right\}, \tag{1.26}$$

where

$$\hat{H}_I(t) = e^{+i\hat{H}_0 t} \hat{V} e^{-i\hat{H}_0 t}, \quad \hat{H}_0 = \hat{p}^2 \tag{1.27}$$

defines the interaction-picture Hamiltonian. The S-matrix is given by

$$\hat{S} = \lim_{T \to \infty} \hat{U}_I(+T, -T), \tag{1.28}$$

so the Dyson series directly gives an expansion of the S-matrix. The first few terms in the expansion are

$$\begin{aligned} \langle p_f | \hat{S} | p_i \rangle = \delta(p_f - p_i) + 2\pi\delta(E_f - E_i) \bigg[&\langle p_f | \hat{V} | p_f \rangle \\ &+ \int dp \, \langle p_f | \hat{V} | p \rangle \frac{i}{E - \frac{1}{2}p^2 + i\epsilon} \langle p | \hat{V} | p_i \rangle + \cdots \bigg]. \end{aligned} \tag{1.29}$$

This perturbation series is precisely analogous to the perturbation series used in quantum field theory. This series can be interpreted (*à la* Feynman) as describing the amplitude as a sum of terms where the particle goes from its initial state $|p_i\rangle$ to the final state $|p_f\rangle$ in 'all possible ways.' The first term represents the possibility that the particle does not interact at all; the higher terms represent the possibility that the particle interacts once, twice, ... with the potential. The interaction with the potential does not conserve the particle momentum (the potential does not 'recoil'), so the momentum of the particle between interactions with the potential takes on all possible values, as evidenced by the momentum integral in the third term. It is this momentum integral that brings in intermediate states of arbitrarily high momentum, and gives the possibility for ultraviolet divergences.

For $V(x) = c_0\delta(x)$ we have

$$\langle p' | \hat{V} | p \rangle = \frac{c_0}{2\pi}, \tag{1.30}$$

and the second-order term in Eq. (1.29) is given by

$$\left(\frac{c_0}{2\pi} \right)^2 \int dp \, \frac{i}{E - \frac{1}{2}p^2 + i\epsilon}. \tag{1.31}$$

Note that this integral is convergent for large p. Higher order terms contain additional momentum integrals, but for each momentum integral dp there is an energy denominator $\sim 1/p^2$, so all terms in the perturbation series are convergent.

On the other hand, for $V(x) = c_1\delta'(x)$ (setting $c_0 = 0$ for the moment), we have

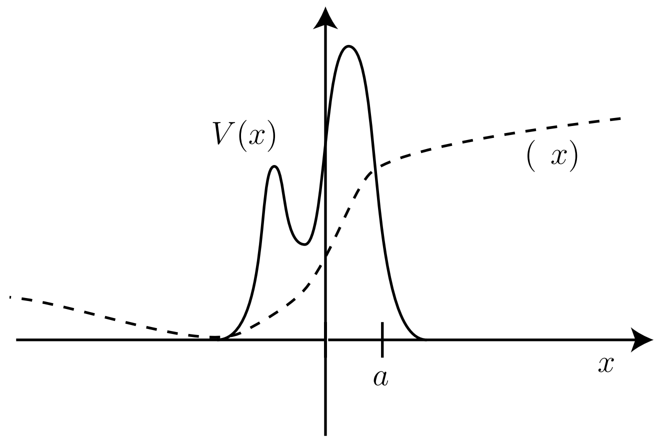
$$\langle p' | \hat{V} | p \rangle = \frac{ic_1(p' - p)}{2\pi}. \tag{1.32}$$

The second-order contribution is then given by

$$\left(\frac{c_1}{2\pi} \right)^2 \int dp \, \frac{(p - p_i)(p - p_f)}{E - \frac{1}{2}p^2 + i\epsilon}, \tag{1.33}$$

which has a linear ultravoilet divergence. Higher orders in the perturbation series are also linearly divergent. This is an indication that the inconsistency found above is due to an ultraviolet divergence due to sensitivity to high-momentum modes.

We can understand the sensitivity to high-momentum modes in another way by going back to the Schrödinger equation. If we look at the solution of the Schrödinger equation for the true potential, it will look schematically as follows:



The wavefunction varies slowly (on a length scale of order λ) in the region where the potential is nonzero, but it in general varies rapidly (on a length scale of order a) inside the potential. In momentum space, we see that the true solution involves high-momentum as well as low-momentum modes. The determination of c_0, $c_1, \ldots$ in Eq. (1.12) was done to match the matrix elements of the potential for long wavelength (low-momentum) states. We now see that this is insufficient, because the true solution involves high-momentum modes.

One might be tempted to conclude from this that a phenomenological description is simply not possible beyond the delta function approximation. After all, if the underlying theory is known, one can always compute the corrections without any approximation, and it might be argued that the inconsistencies found here imply that this is the only consistent way to proceed beyond leading order. However, this point of view is unsatisfactory. If the low-energy behavior were not governed by universal low-energy theorems such as Eq. (1.21), it would mean that we can obtain detailed information about physics at arbitrarily short distances from measurements at long distances. This would be an experimentalist's dream, but a theorist's nightmare: it would mean that experimentalists can probe features of the short-distance physics with long-distance experiments, but theorists cannot make predictions for long-distance experiments without knowing the *exact* short-distance theory. This is certainly counter to our experience and intuition that experiments at low energies cannot resolve the detailed short-distance features of physical systems. We therefore expect that the low-energy behavior is governed by low-energy theorems, and we must face up to the problem of working them out.

If there is a universal low-energy form for scattering from a short-range potential of this form, we can work it out with an arbitrary odd short-range potential. We will use

$$V(x) = c_1 \frac{\delta(x+a) - \delta(x-a)}{2a}. \tag{1.34}$$

As $a \to 0$, $V(x) \to c_1 \delta'(x)$, so this can be viewed as a discretization of the potential. For $a \neq 0$, we have a well-defined potential with width of order a. The parameter a is our first example of a **short-distance cutoff**. The reason for this terminology is that the discretized theory is less sensitive to short distance modes, so these can be viewed as being 'cut off' from the theory.

We can compute the transmission amplitude by writing a solution of the form

$$\Psi(x) = \begin{cases} Ae^{ipx} + Be^{-ipx} & x < -a \\ A'e^{ipx} + B'e^{-ipx} & -a < x < a \\ Ce^{ipx} & x > a, \end{cases} \tag{1.35}$$

and imposing continuity and the jump conditions at $x = -a$ and $x = a$. The solution is

$$\frac{1}{T} = \frac{c_1^2}{4a^2 p^2}(1 - e^{4iap}). \tag{1.36}$$

Note that this diverges (as $1/a$) as $a \to 0$. For $p \ll a$, the leading behavior is

$$T \simeq -\frac{iap}{c_1^2}. \tag{1.37}$$

Note that the cutoff theory naïvely depends on 2 parameters, namely c_1 and the cutoff a. However, this result shows that the leading low-energy behavior depends only on the combination

$$c_R = \frac{c_1^2}{a}. \tag{1.38}$$

Therefore, the low-energy theorem can be written as

$$T \simeq -\frac{ip}{c_R}, \tag{1.39}$$

which depends on the single phenomenological paramer c_R. The final result is independent of the cutoff a in the sense that we can compensate for a change in a by changing c_1.

Note that the result Eq. (1.39) is surprising from the point of view of dimensional analysis, because c_1 and c_R have different dimensions. In fact, c_1 is dimensionless, so dimensional analysis would tell us that the dimensionless transition amplitude cannot depend on the momentum! However, because the cutoff parameter a has dimension, the renormalized parameter can have a different dimension than the coupling in the Hamiltonian. We say that the renormalized coupling has an **anomalous dimension**. This is a general feature of theories with ultraviolet divergences.

The steps followed above to obtain the low-energy theorem for this simple system are precisely those we will follow in quantum field theory. Let us restate the main features for emphasis.

(i) *Ultraviolet divergences:* When we perform a naïve calculation using local interactions (delta functions and their derivatives), we find that the results are generally inconsistent due to short-distance divergences. The origin of these divergences is in the fact that quantum mechanics involves sums over a complete set of states, so quantum corrections are sensitive to the properties of high-momentum intermediate states.

(ii) *Regularization:* To parameterize the sensitivity to short distance, we modify the theory at a distance scale of order a (the cutoff) so that it is well-defined. We say that the theory has been **regularized**. In the theory with the cutoff, the ultraviolet divergences are replaced by sensitivity to a, in the sense that the physical quantities diverge in the limit $a \to 0$ with the couplings held fixed.

(iii) *Renormalization:* The regulated theory apparently has one more parameter than the naïve continuum theory, namely the cutoff. However, when we compute physical quantities, we find that they depend only on a combination of the cutoff and the other parameters. In other words, a change in the cutoff can be compensated by a change in the couplings so that all physical quantities are left invariant. We therefore finally obtain well-defined finite results that depend on the same number of parameters as the original local formulation.

1.2 2 Dimensions

We now consider a short-range potential in 2 spatial dimensions. This example will illustrate some additional features of renormalization. If we approximate the potential by a delta function, the Schrödinger equation is

$$-\tfrac{1}{2}\nabla^2\Psi + c\delta^2(x)\Psi(x) = E\Psi(x). \tag{1.40}$$

Note that the 'coupling constant' c is dimensionless (in units where $\hbar = 1$, $m = 1$). A straightforward attempt to solve this equation leads to inconsistencies. We can see that these are due to ultraviolet divergences by looking at the second-order term in the perturbative expansion Eq. (1.29). It is

$$\left(\frac{c}{2\pi}\right)^2 \int d^2p \, \frac{i}{E - \frac{1}{2}p^2 + i\epsilon}. \tag{1.41}$$

This is logarithmically divergent for large p.

To solve this problem, we must again regulate the delta function. For simplicity, we restrict attention to spherically symmetric solutions. We then have

$$\nabla^2\Psi = \frac{1}{r}\frac{d}{dr}\left(r\frac{d\Psi}{dr}\right), \tag{1.42}$$

and

$$\delta^2(x) = \frac{1}{2\pi r}\delta(r), \tag{1.43}$$

so that $\int d^2x \, \delta^2(x) = 1$. We regulate the delta function by replacing

$$\delta(r) \to \delta(r-a), \tag{1.44}$$

where a is a cutoff. We must then solve the equation

$$-\frac{1}{2r}\frac{d}{dr}\left(r\frac{d\Psi}{dr}\right) + \frac{c}{2\pi r}\delta(r-a)\Psi(r) = E\Psi(r). \tag{1.45}$$

For $r \neq a$, the general is a linear combination of the Bessel functions $J_0(pr)$ and $Y_0(pr)$. The solution for $r < a$ involves only the Bessel function that is regular at the origin: for small x,

$$J_0(x) = 1 - \frac{x^2}{4} + \mathcal{O}(x^4), \qquad Y_0(x) = \frac{2}{\pi}\left(\ln\frac{x}{2} + \gamma\right)\left(1 + \mathcal{O}(x^2)\right), \tag{1.46}$$

where $\gamma = 0.577216\ldots$ is the Euler constant. We therefore have

$$\Psi(r) = \begin{cases} CJ_0(pr) & r < a, \\ AJ_0(pr) + BY_0(pr) & r > a, \end{cases} \tag{1.47}$$

where $p = \sqrt{2E}$. Multiplying Eq. (1.45) by r and integrating the equation from $a-\epsilon$ to $a+\epsilon$ and letting $\epsilon \to 0$ gives the discontinuity condition

$$\Psi'(a+\epsilon) - \Psi'(a-\epsilon) = \frac{c}{\pi}\Psi(a). \tag{1.48}$$

In addition, we demand that Ψ is continuous at $r = a$. This gives two equations that determine the wavefunction Eq. (1.47) up to overall normalization. Solving for A and B and expanding for $p \ll 1/a$, we obtain

$$\begin{aligned} A &= \left[1 - \frac{c}{\pi}\left(\ln\frac{pa}{2} + \gamma\right) + \mathcal{O}(p^2a^2)\right] C, \\ B &= \left[\frac{c}{2} + \mathcal{O}(p^2a^2)\right] C. \end{aligned} \tag{1.49}$$

To see what this means for scattering, note that for $r \gg a$ we have

$$\begin{aligned} \Psi(r) &\to A\cos\left(pr - \frac{\pi}{4}\right) + B\sin\left(pr - \frac{\pi}{4}\right) \\ &\propto \cos\left(pr - \frac{\pi}{4} + \delta_0\right), \end{aligned} \tag{1.50}$$

where δ_0 is the s-wave scattering phase shift, given by

$$\tan\delta_0 = -\frac{B}{A}. \tag{1.51}$$

All observable quantities for s-wave scattering can be expressed in terms of δ_0, so we can think of it as a physical observable.[1] For small p, we have

$$\cot\delta_0 = -\frac{2}{c} + \frac{2}{\pi}\left(\ln\frac{p}{2\Lambda} + \gamma\right) + \mathcal{O}(p^2/\Lambda^2), \tag{1.52}$$

where we have introduced

$$\Lambda \stackrel{\text{def}}{=} \frac{1}{a}, \tag{1.53}$$

which can be thought of as a momenum cutoff. We work in terms of Λ in order to make our discussion more parallel with the quantum field theory case.

Although Eq. (1.52) is a function of both c and Λ, it is not hard to show that it is a function only of one combination of these variables. To do this we must find a way to change Λ and c simultaneously while keeping the phase shift invariant. That is, we look for a function $c(\Lambda)$ such that Eq. (1.52) is independent of Λ (up to the $\mathcal{O}(1/\Lambda^2)$ corrections). Imposing

$$0 = \frac{d}{d\Lambda}\cot\delta_0 = \left(\frac{\partial}{\partial\Lambda} + \frac{dc}{d\Lambda}\frac{\partial}{\partial c}\right)\cot\delta_0, \tag{1.54}$$

---

[1] Note the formal similarity between $\tan\delta_0$ and the reflection coefficient for the 1-dimensional case.

we obtain

$$\Lambda \frac{d}{d\Lambda}\left(\frac{1}{c}\right) = -\frac{1}{\pi}. \tag{1.55}$$

This is our first example of a **renormalization group equation**.[2] The couplings $c(a)$ defined in this way are called **running couplings**. They define a family of effective theories with different cutoffs, such that the low-energy physics is the same *for all momenta p*. This shows that the result Eq. (1.52) depends only on a single parameter.

We can express this in terms of a renormalized coupling c_R defined by

$$\frac{1}{c_R(\mu)} \stackrel{\text{def}}{=} \frac{1}{c} - \frac{1}{\pi}\left(\ln \frac{\mu}{2\Lambda} + \gamma\right), \tag{1.56}$$

where μ is a momentum scale required to write a dimensionally consistent definition for the renormalized coupling. Since μ is arbitrary, physical results must be independent of μ. We have absorbed the constant term $-\gamma/\pi$ into the renormalized just to get simpler expressions. Then we can write

$$\cot \delta_0 = -\frac{2}{c_R(\mu)} + \frac{2}{\pi} \ln \frac{p}{\mu} + \mathcal{O}(p^2/\Lambda^2), \tag{1.57}$$

which shows that the phase shift is a function only of c_R. From these formulas, it is clear that the value of μ in Eq. (1.57) is arbitrary: if we change μ, the function $c_R(\mu)$ changes in such a way as to leave the phase shift invariant. This is also ensured by the equation

$$\mu \frac{\partial}{\partial \mu}\left(\frac{1}{c_R}\right) = -\frac{1}{\pi} \tag{1.58}$$

that is satisfied by the definition of c_R. We see that the bare and renormalized couplings obey exactly the same renormalization group equation. This is not a coincidence. From Eq. (1.56), we see that

$$c_R(\mu = 2e^{-\gamma}\Lambda) = c(\Lambda). \tag{1.59}$$

Since this equation is true for any Λ, we see that the bare coupling can be thought of as the renormalized coupling evaluated at a scale $\mu \sim \Lambda$.

The advantage of the renormalized coupling is that it is very closely related to physical quantities. In fact, from Eq. (1.57) we have

$$\cot \delta_0 = -\frac{2}{c_R(\mu = p)} + \mathcal{O}(p^2/\Lambda^2). \tag{1.60}$$

[2] The 'renormalization group' is not a group in the mathematical sense. The name is historical.

We see that for this simple problem, the renormalized coupling contains *all* information about the physics of the problem, *i.e.* the p dependence of the phase shift. In general, physical quantities depend on more than one dimensionful quantity, and the relation between renormalized couplings and physical amplitudes is not so simple, but we will see that they are still more closely related to physical quantities.

There is a beautiful physical picture that underlies these results, due to K. Wilson. The renormalized coupling is defined by decreasing the momentum cutoff Λ while changing the couplings to keep the low-energy physics the same. Because the theory with a lower cutoff has fewer degrees of freedom, this can be thought of as coarse-graining, or integrating out high-momentum fluctuations. In this way, we obtain a family of effective field theories that describe the same long-distance physics. The reason we can define such effective field theories is that physics at low momentum is sensitive to short-distance physics only through the value of the effective coupling. We can continue lowering the cutoff until it becomes of order the physical momentum p. At this point, almost all of the fluctuations have been integrated out, and the renormalized coupling contains essentially all the dynamical information in the theory.

The least intuitive part of this picture is that we can lower the cutoff Λ all the way to the physical scale p. In fact, if the cutoff and the physical scale are the same, there is no longer a small parameter that can make the effective field theory description valid. The reason we can make take $\mu \sim p$ is that evolving the couplings from the scale Λ to the scale μ using the renormalization group equation Eq. (1.58) and boundary condition Eq. (1.59) does not affect the $\mathcal{O}(1/\Lambda^2)$ corrections. These remain $\mathcal{O}(1/\Lambda^2)$ even when expressed in terms of the renormalized couplings at the scale $\mu \sim p$.

This means that the renormalization group discussed here does not precisely correspond to the idea of 'lowering the cutoff.' However, there is a great deal of intuitive power in this analogy, and we will explore these ideas further in the context of quantum field theory.

These results sufficiently important that it is worth repeating the main points for emphasis.

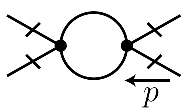
• *Renormalization group:* The fact that the low-energy physics is not directly sensitive to the cutoff can be summarized by the statement that it is possible to change the cutoff and simultaneously change the couplings so that the low-energy physics is unchanged. This can be summarized by renormalization group equations such as Eq. (1.55).

• *Renormalized couplings:* We can define renormalized couplings by evolving the couplings down to the physical scales using the renormalization group equations. The resulting renormalized couplings resum a large class of corrections because the theory with a lower cutoff contains fewer fluctuations.

Finally, note that a direct application of dimensional analysis to this system would tell us that δ_0 must be independent of p, since δ_0 and c are dimensionless, while p is dimensionful. However, the introduction of the regulator a spoils this argument. Eq. (1.57) shows explicitly that the phase shift has a nontrivial dependence on p. Thus we see again that renormalization introduces nontrivial modifications of scaling behavior, a feature that will recur in quantum field theory.

2 Regularization

We now return to quantum field theory. It is a fact of life that general loop diagrams in quantum field theories diverge, and are therefore ill-defined. For example, in ϕ^4 theory in $3+1$ dimensions, the 1-loop correction to the 4-point function is



$$= \frac{(-i\lambda)^2}{2} \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m^2 + i\epsilon} \frac{i}{(k+p)^2 - m^2 + i\epsilon}, \tag{2.1}$$

where p is the total momentum flowing into the left of the diagram. For large values of k, we can neglect the mass and external momenta, and the integral behaves as

$$\sim \int \frac{d^4k}{k^4} \tag{2.2}$$

which diverges as $k \to \infty$. We see that the integral is ill-defined due to an ultraviolet divergence.

This divergence is precisely analogous to the ultraviolet divergences found in the quantum mechanical models above. To see this, note that the loop corrections are the $\mathcal{O}(\hbar)$ (and higher) corrections in the semiclassical expansion, and therefore parameterize the quantum fluctuations around a classical trajectory. The integral over the loop momentum k adds up the fluctuations at all momentum scales. The fact that the integral diverges for large momenta means that the quantum corrections are sensitive to quantum fluctuations at high momenta.

The ultraviolet divergences in both quantum mechanics and quantum field theory can be viewed as a consequence of the fact that we took the continuum limit 'too soon.' In quantum mechanics, we fixed this by going back to a smooth potential where everything is well-defined. In quantum field theory, the situation is a little different because the principles of relativity, unitarity, and causality require us to write a theory with local interactions. This generally leads to theories with ultraviolet divergences, such as QED or ϕ^4 theory discussed above. It is very difficult to write a physically sensible local quantum field theory that corresponds to the theory of interest at low momenta, and is free from ultraviolet divergences.[3]

Nonetheless, we must modify the high-momentum behavior of quantum field theory somehow in order to make the theory well-defined. This is called **regularization**. All of the regulators we will discuss have unphysical features that show up at high momentum. However, this will not matter in the end, since the main result of renormalization theory is that physical results are sensitive to physics at very short distances only through the values of renormalized coupling constants.

We now briefly describe some regulators for ϕ^4 theory in $3+1$ spacetime dimensions.

2.1 Momentum Cutoff

Perhaps the simplest regulator for perturbation theory is simply to restrict the momentum integrals so that we do not integrate over all momenta. The Lorentz-invariant restriction $k^2 < \Lambda^2$ does not prevent some of the components of k from becoming large, but we can impose the cutoff after Wick-rotation to Euclidean loop momenta:

$$\int d^4 k_E \to \int_{k_E^2 < \Lambda^2} d^4 k_E. \tag{2.3}$$

This cutoff is 'unphysical' in that it violates unitarity. The integral over momenta originates in a sum over a complete set of intermediate states, and simply omitting some of the momenta will violate unitarity. However, this shows up only for external momenta near Λ, and we are interested in $p \ll \Lambda$.

2.2 Lattice Regularization

Perhaps the most 'physical' regulator is to replace continuous spacetime with a lattice. We encountered the lattice formulation as a way of discretizing the path integral to make it well-defined. On a lattice, a quantum field theory becomes a quantum system whose degrees of freedom consist of one field variable ϕ_x at each lattice point x. The lattice structure does not do violence to the quantum-mechanical structure of the theory, but it does not preserve the spacetime symmetries.

[3] It is not impossible: there are supersymmetric quantum field theories that are free from ultraviolet divergences. However, theories of this type are so complicated that they are of little practical value.

Consider a cubic lattice consisting of the points

$$x = a \cdot (n_0, n_1, n_2, n_3), \tag{2.4}$$

where a is the lattice spacing and n_0, n_1, n_2, n_3 are integers. Instead of continous Lorentz and translation symmetry we have the discrete symmetries of the lattice. This need not be a problem, since we can recover the continuous symmetries at low energies.

If we define Fourier transformed fields

$$\tilde{\phi}_k \stackrel{\text{def}}{=} \sum_x e^{ik\cdot x} \phi_x, \tag{2.5}$$

we see that the momenta

$$k,\ k + \frac{2\pi}{\ell}(1,0,0,0), \ldots,\ k + \frac{2\pi}{\ell}(0,0,0,1), \tag{2.6}$$

are equivalent. Therefore, a complete set of momenta k lies in a 'Brillouin zone' that is a 4-dimensional cube with sides of length $2\pi/a$. When we write the momentum-space Feynman rules for this theory, the momentum integrals are therefore integrals over a finite range of momenta. This shows that the lattice regulator can be viewed as a sophisticated version of the momentum-space cutoff. The latice regulator is very cumbersome for analytic calculations, but it is very convenient for computer calculations.

2.3 Higher Derivative Regulator

Another possible regulator for ϕ^4 theory is obtained by adding to the Lagrangian density a term

$$\delta\mathcal{L} = -\frac{1}{2\Lambda^2}\phi\Box^2\phi. \tag{2.7}$$

If we view this as part of the kinetic term, the propagator is modified to

$$\Delta(k) = \frac{1}{k^2 - m^2 + i\epsilon} \to \frac{1}{k^2 - m^2 - k^4/\Lambda^2 + i\epsilon}. \tag{2.8}$$

Note that this propagator behaves as $1/k^4$ at large k, improving the convergence of Feynman diagrams. (We can add even higher powers of $\Box$ in the Lagrangian if this is not sufficient.)

Since we are simply adding a local term to the Lagrangian, it is tempting to think that this defines a completely physical finite theory. However, notice that the propagator has a poles at

$$m_{\pm} = \frac{\Lambda^2}{2}\left[1 \pm \left(1 - \frac{4m^2}{\Lambda^2}\right)^{1/2}\right], \tag{2.9}$$

where

$$m_+^2 = \Lambda^2[1 + \mathcal{O}(m^2/\Lambda^2)], \quad m_-^2 = m^2[1 + \mathcal{O}(m^2/\Lambda^2)]. \tag{2.10}$$

are both positive. However, the pole at m_+ has a negative residue. This can be easily seen from the factorized form

$$\Delta(k) = -\frac{1}{(k^2 - m_-^2 + i\epsilon)(k^2 - m_-^2 - i\epsilon)}. \tag{2.11}$$

(As usual, the $i\epsilon$ factors can be reconstructed by giving m^2 a small negative imaginary part.) The residue of the pole at $k^2 = m_-^2$ is positive, but the residue at $k^2 = m_-^2$ is negative. We have seen that the residue of a pole due to 1-particle intermediate state $|\vec{k}\rangle$ is $|\langle 0|\hat{\phi}(0)|\vec{k} = 0\rangle|^2 > 0$. This means that there is an unphysical pole at $k^2 \sim \Lambda^2$. It can be shown that there is always a pole with negative residue no matter how the higher-derivative terms are chosen. Closely related to this is the fact that the pole at m_+^2 has the wrong $i\epsilon$ prescription, which means that negative energy is propagating forward in time.

Again, this is not a problem if we choose Λ to be large compared to m and the physical energy scales we are interested in probing.

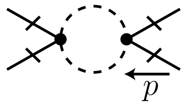
2.4 Pauli–Villars

The idea of the Pauli–Villars regulator is to add unphysical scalar Grassmann fields to the theory. Since loops of fermionic fields have an extra minus sign compared to bosonic fields, we can choose the interactions of the Pauli–Villars fields to make loop diagrams finite.

For example, to regularize the diagram in Eq. (2.1) above, we introduce the Pauli–Villars fields χ with Lagrangian

$$\delta\mathcal{L} = \frac{1}{2}\partial^\mu\chi^\dagger\partial_\mu\chi - \Lambda^2\bar{\chi}\chi - \frac{\lambda}{2}\phi^2\chi^\dagger\chi \tag{2.12}$$

(We cannot use a real spinor field, since $\chi^2 = 0$.) At 1-loop order, the scalar 4-point function has the additional contribution



$$= -\frac{(-i\lambda)^2}{2} \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - \Lambda^2 + i\epsilon} \frac{i}{(k+p)^2 - \Lambda^2 + i\epsilon}, \tag{2.13}$$

where dashed line denotes the χ propagator, and the minus sign is due to the fermion loop. The coefficient of the $\phi^2\chi^\dagger\chi$ interaction term has been chosen so that for $k \gg \Lambda$ the integrand is equal and opposite to the integrand Eq. (2.1). This is enough to ensure that the sum of Eq. (2.13) and Eq. (2.1) is convergent. For $p, m \ll k \ll \Lambda$, the integrand of the Pauli–Villars loop behaves like a constant, so Λ acts like a momentum-space cutoff.

A scalar fermion such as χ violates the spin-statistics theorem, and is therefore unphysical. (Specifically, it propagates outside the lightcone.) Since we interpret χ as part of the regulator, we do not compute diagrams with χ external lines. Even so, like the higher-derivative regulator, the Pauli–Villars regulator gives rise to unphysical singularities in amplitudes involving only the scalar fields. However, as long as the mass Λ of the Pauli–Villars field is large compared to the physical scales of interest, we expect this regulator to only affect the high-energy behavior of the theory.

The term in Eq. (2.12) does not regulate all diagrams, and in general additional Pauli–Villars fields must to be introduced at each loop order. Nonetheless, the Pauli–Villars regulator is actually quite convenient for some purposes, as we will see later.

2.5 Dimensional Regularization

We have seen in some examples that quantum field theories are less divergent in lower spacetime dimensions. Motivated by this, one can regulate Feynman diagrams by taking the spacetime dimension d to be a *continuous* parameter. A momentum integral of the form

$$\int d^d k \prod_{j=1}^{n} \frac{1}{(k+p_j)^2 - m^2 + i\epsilon} \tag{2.14}$$

is therefore convergent for a sufficiently small d, and we can define its value as $d \to 4$ by analytic continuation in d. When this is done, one finds that there are poles of the form $1/(d-4)$ that signal the presence of ultraviolet divergences. This regulator is the most useful one for most practical calculations. However, it is rather formal and unintuitive. (For example, it is not known how to formulate dimensional regularization outside of perturbation theory.) We will therefore discuss this regulator only after we have a general understanding of renormalization.

2.6 Overview

The regulators we have discussed are quite different from each other, but they share several important common features. In each case, the regulator introduces a new energy scale Λ into the theory. In the regulated theory, momenta with components small compared to Λ behave as they do in the unregulated theory, while the contribution from momentum modes with components of order Λ or larger are suppressed. All of the regulators have unphysical features that emerge at momentum scales of order Λ, but these do not bother us as long as Λ is larger than the physical scales of interest.

3 One Loop Renormalization

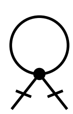
We now get our hands dirty with some 1-loop renormalization calculations.

We begin with ϕ^4 theory in $3+1$ dimensions. We write the Lagrangian as

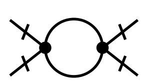
$$\mathcal{L}_0 = \frac{1}{2}\partial^\mu \phi \partial_\mu \phi - \frac{m_0^2}{2}\phi^2 - \frac{\lambda_0}{4!}\phi^4. \tag{3.1}$$

The subscripts remind us that the quantities refer to the parameters that appear in the regulated lagrangian. The couplings are called **bare couplings**, and the Lagrangian is called the **bare Lagrangian**. It is important to keep in mind that the 'bare' Lagrangian is simply the Lagrangian that appears in the path integral.

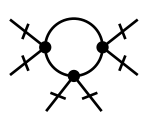
The first step in renormalizing this theory is to identify all the divergent diagrams that appear at one loop. Since the ultraviolet divergences appear for large loop momenta, we can simply ignore all external masses and external momenta to determine whether a diagram diverges. To quantify the divergences, we use a momentum-space cutoff Λ. The 1-loop 1PI diagrams are



$$\sim \int^\Lambda \frac{d^4k}{k^2} \sim \Lambda^2,$$



$$\sim \int^\Lambda \frac{d^4k}{k^4} \sim \ln\Lambda,$$



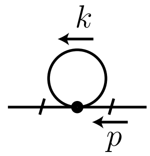
$$\sim \int^\Lambda \frac{d^4k}{k^6} = \text{finite}, \tag{3.2}$$

$$\vdots$$

All diagrams with 6 or more external legs converge, because there are more propagators and hence more powers of k in the denominator. We see that there are only two divergent diagrams at this order. Let us evaluate them.

3.1 Mass Renormalization

The 2-point function is



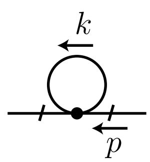
$$= \frac{-i\lambda_0}{2} \int \frac{d^4k}{(2\pi)^4} \frac{i}{k^2 - m_0^2 + i\epsilon}$$

$$= -\frac{i\lambda_0}{2} \int \frac{d^4k_E}{(2\pi)^4} \frac{1}{k_E^2 + m_0^2}$$

$$= -\frac{i\lambda_0}{2} \frac{1}{16\pi^2} \int_0^{\Lambda^2} dk_E^2 \, \frac{k_E^2}{k_E^2 + m_0^2}$$

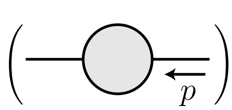
$$= -\frac{i\lambda_0}{32\pi^2} \left[\Lambda^2 - m_0^2 \ln \frac{\Lambda^2 + m_0^2}{m_0^2} \right]. \tag{3.3}$$

Here we have continued to Euclidean momentum in the second line, and imposed a momentum space cutoff in the third line. Since we are interested in taking Λ large compared to the mass m, we expand the result to obtain



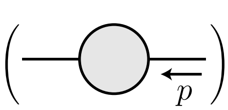
$$= -\frac{i\lambda_0}{32\pi^2} \left[\Lambda^2 - m_0^2 \ln \frac{\Lambda^2}{m_0^2} + \mathcal{O}(m_0^4/\Lambda^2) \right]. \tag{3.4}$$

This result depends on Λ, but we will show that we can change the cutoff and simultaneously change the couplings of the theory to keep the correlation functions invariant. Specifically, we must keep invariant the full inverse propagator, which is given at one loop by



$$\left(\cdots \right)^{-1} = p^2 - m_0^2 - \Sigma(p^2), \tag{3.5}$$

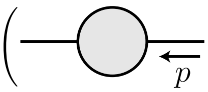
where $-i\Sigma(p^2)$ is the sum of the 1PI diagrams. At this order, we therefore have

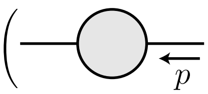


$$\left(\cdots \right)^{-1} = p^2 - m_0^2 - \frac{\hbar\lambda_0}{32\pi^2} \left[\Lambda^2 - m_0^2 \ln \frac{\Lambda^2}{m_0^2} + \mathcal{O}(m_0^4/\Lambda^2) \right] + \mathcal{O}(\hbar^2), \tag{3.6}$$

where we have explicitly included the loop-counting parameter $\hbar$.

We want to show that we can change Λ and simultaneously change the couplings so that the inverse propagator remains invariant, up to the $\mathcal{O}(1/\Lambda^2)$ corrections. We therefore impose

$$0 = \frac{d}{d\Lambda} \left(\vphantom{\frac{d}{d\Lambda}} \right.$$

$$\left. \vphantom{\frac{d}{d\Lambda}} \right)^{-1}$$

$$= \left(\frac{\partial}{\partial\Lambda} + \frac{dm_0^2}{d\Lambda}\frac{\partial}{\partial m_0^2} + \frac{d\lambda_0}{d\Lambda}\frac{\partial}{\partial\lambda_0} \right) \left(\vphantom{\frac{d}{d\Lambda}} \right.$$

$$\left. \vphantom{\frac{d}{d\Lambda}} \right)^{-1}. \tag{3.7}$$

Since m_0^2 and λ_0 do not depend on Λ in the absence of loop corrections, we have

$$\frac{dm_0^2}{d\Lambda} = \mathcal{O}(\hbar), \qquad \frac{d\lambda_0}{d\Lambda} = \mathcal{O}(\hbar). \tag{3.8}$$

This is important for systematically working order-by-order in the loop expansion. Substituting Eq. (3.6) into Eq. (3.7), we obtain

$$0 = -\frac{dm_0^2}{d\Lambda} - \frac{\hbar\lambda_0}{32\pi^2}\left[2\Lambda - \frac{2}{\Lambda}m_0^2\right] + \mathcal{O}(\hbar^2). \tag{3.9}$$

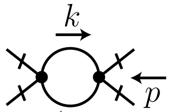
Note that the terms where $\partial/\partial\Lambda$ acts on m_0^2 and λ_0 in the second term of Eq. (3.6) are $\mathcal{O}(\hbar^2)$, and therefore negligible at this order. We therefore have

$$\boxed{\Lambda\frac{dm_0^2}{d\Lambda} = -\frac{\hbar\lambda_0}{16\pi^2}\left(\Lambda^2 - m_0^2\right) + \mathcal{O}(\hbar^2).} \tag{3.10}$$

This equation tells us that in order to keep the inverse propagator invariant to $\mathcal{O}(\hbar)$, we must change the mass term in the Lagrangian. We say that the mass is **renormalized**.

3.2 Coupling Constant Renormalization

We now turn to the diagrams with four external legs. There are three crossed diagrams that contribute to the 1PI 4-point function. Each one has the form



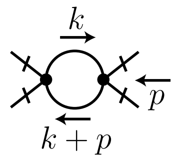
$$= \frac{(-i\lambda_0)^2}{2}\int\frac{d^4k}{(2\pi)^4}\,\frac{i}{(k+p)^2 - m_0^2 + i\epsilon}\,\frac{i}{k^2 - m_0^2 + i\epsilon} \tag{3.11}$$

$$= \frac{\lambda^2}{2}\int_0^1 dx \int\frac{d^4K}{(2\pi)^4}\,\frac{1}{(K^2 - M_0^2 + i\epsilon)^2}, \tag{3.12}$$

where we have introduced Feynman parameters and

$$K = k + xp, \qquad M_0^2 = m_0^2 - x(1-x)p^2. \tag{3.13}$$

Continuing to Euclidean momenta and imposing a cutoff, we have

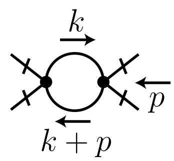


$$= \frac{i\lambda_0^2}{2} \int_0^1 dx \, \frac{1}{16\pi^2} \int_0^{\Lambda^2} dK_E^2 \, \frac{K_E^2}{(K_E^2 + M_0^2)^2}. \tag{3.14}$$

The K_E^2 integral is elementary:

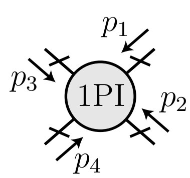
$$\int_0^{\Lambda^2} dK_E^2 \, \frac{K_E^2}{(K_E^2 + M_0^2)^2} = 1 + \ln \frac{\Lambda^2}{M_0^2} + \mathcal{O}(M_0^2/\Lambda^2). \tag{3.15}$$

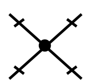
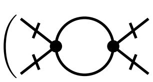
We therefore obtain



$$= \frac{i\lambda_0^2}{32\pi^2} \int_0^1 dx \left[1 + \ln \frac{\Lambda^2}{M_0^2} + \mathcal{O}(M_0^2/\Lambda^2) \right]. \tag{3.16}$$

(Remember that M_0 depends on x, so the remaining integral is nontrivial.) The full 1PI 4-point function at this order is therefore



$= $  $+ \Big($  $+$ crossed$\Big)$

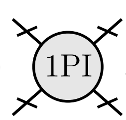
$$= -i\lambda_0 + \frac{i\hbar\lambda_0^2}{32\pi^2} \int_0^1 dx \left[3 + \ln \frac{\Lambda^2}{M_0^2(s)} + \ln \frac{\Lambda^2}{M_0^2(t)} \ln \frac{\Lambda^2}{M_0^2(u)} + \mathcal{O}(1/\Lambda^2) \right] + \mathcal{O}(\hbar^2), \tag{3.17}$$

where

$$s \stackrel{\text{def}}{=} (p_1 + p_2)^2, \quad t \stackrel{\text{def}}{=} (p_1 + p_3)^2, \quad u \stackrel{\text{def}}{=} (p_1 + p_4)^2, \tag{3.18}$$

and we have included powers of $\hbar$ to make the loop-counting explicit.

We want to show that a change in Λ can be compensated by changing the bare couplings, up to $\mathcal{O}(1/\Lambda^2)$ corrections. We thererfore impose

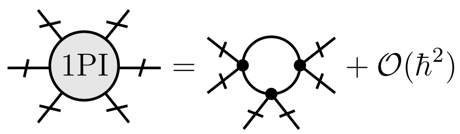
$$0 = \left(\frac{\partial}{\partial\Lambda} + \frac{dm_0^2}{d\Lambda} \frac{\partial}{\partial m_0^2} + \frac{d\lambda_0}{d\Lambda} \frac{\partial}{\partial\lambda_0} \right)$$  $. \quad (3.19)$

Since we are working only to $\mathcal{O}(\hbar)$, we can neglect terms where the derivatives act on λ_0 and m_0^2 in the second term in Eq. (3.55). We therefore obtain simply

$$\boxed{\Lambda \frac{d\lambda_0}{d\Lambda} = \frac{3\hbar\lambda_0^2}{16\pi^2} + \mathcal{O}(\hbar^2).} \tag{3.20}$$

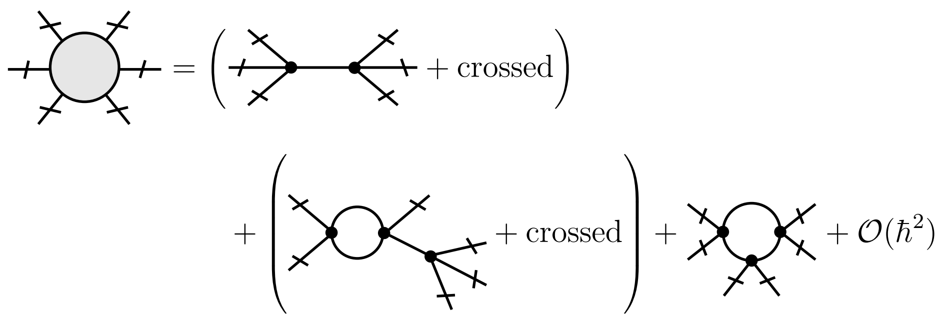
This tells us that, at least for the 4-point function, changing the cutoff is equivalent to changing λ_0. We therefore say that the coupling is renormalized.

If we impose Eqs. (3.10) and (3.20), all the other 1PI correlation functions are also invariant up to corrections of order $\mathcal{O}(1/\Lambda^2)$ and $\mathcal{O}(\hbar)$. The reason is simply that the 1PI correlation functions with 6 or more external legs have no tree-level contribution (since there are no terms in the Lagrangian proportional to ϕ^6 or higher), and the loop contributions are finite, hence independent of Λ up to $\mathcal{O}(1/\Lambda^2)$ corrections:



(3.21)

Here we see the utility of the 1PI correlation functions. The full 6-point function at one loop is more nontrivial:



(3.22)

The first two terms have nontrivial Λ dependence at this order in the loop expansion. The Λ dependence in the first loop diagram on the second line is cancelled by the Λ dependence of λ in the first term. We know that the full 6-point function is invariant because it can be reconstructed from the 1PI correlation functions, and we have seen that they are invariant.

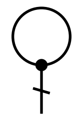
3.3 Wavefunction Renormalization

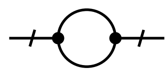
The example above does not illustrate all of the types of renormalization required at higher orders (and in more general theories). To illustrate these, we turn to another example: ϕ^3 theory in $5+1$ dimensions, with bare Lagrangian

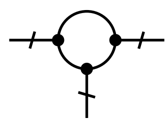
$$\mathcal{L}_0 = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{m_0^2}{2}\phi^2 - \frac{\lambda_0}{3!}\phi^3. \tag{3.23}$$

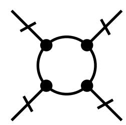
The reason for considering this theory in $5+1$ dimensions will become clearer below.

The divergent 1-loop diagrams are easily enumerated:

 $\sim \int^\Lambda \frac{d^6k}{k^2} \sim \Lambda^4,$

 $\sim \int^\Lambda \frac{d^6k}{k^4} \sim \Lambda^2,$

 $\sim \int^\Lambda \frac{d^6k}{k^6} \sim \ln\Lambda,$ (3.24)

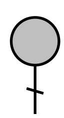
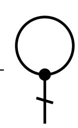
 $\sim \int^\Lambda \frac{d^6k}{k^8} = \text{finite},$

$\vdots$

The existence of a nonzero 1-point function is a new feature of this theory, and requires comment. Such graphs are often called **tadpole** graphs. (Tadpoles vanish in ϕ^4 theory because of the symmetry $\phi \mapsto -\phi$.) By momentum conservation, the 1-point function is nonzero only for zero external momentum. As we will learn later, a nonvanishing 1-point function is a sign that the theory is not at the minimum of its potential. For now we will avoid this issue by noting that the 1-point function can be cancelled exactly by adding a 'bare' linear term

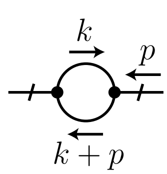
$$\Delta\mathcal{L}_0 = \kappa_0\phi, \tag{3.25}$$

where κ_0 is a coupling with mass dimension $+4$. We impose the condition

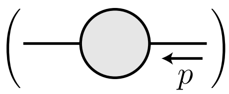
$0 =$  $= -i\kappa_0 +$  $+ \cdots$ (3.26)

which can always be solved for κ_0. Note that with this condition imposed, we can ignore all graphs with tadpole subgraphs.

Computing the 2-point function with a momenum-space cutoff gives



$= \frac{i\lambda_0^2}{256\pi^3}\int_0^1 dx \left[\Lambda^2 + M_0^2\left(1 - 2\ln\frac{\Lambda^2}{M_0^2}\right) + \mathcal{O}(M_0^2/\Lambda^2)\right],$ (3.27) where $M_0^2 = m_0^2 - x(1-x)p^2$. Note that unlike the 2-point function in ϕ^4 theory computed above, this diagram has nontrivial dependence on p^2 because the external momentum runs through the loop. The inverse propagator is



$$\left(\cdots \right)^{-1} = p^2 - m_0^2 + \frac{\hbar\lambda_0^2}{256\pi^3}\left[\Lambda^2 + \left(2m_0^2 - \tfrac{1}{3}p^2\right)\ln\Lambda^2\right] + (\Lambda \text{ independent}) + \mathcal{O}(M_0^2/\Lambda^2) + \mathcal{O}(\hbar^2). \tag{3.28}$$

Because of the $p^2 \ln\Lambda^2$ term, this can be independent of Λ for all p only if $\lambda_0 = 0$ (free field theory).

However, the Λ dependence in Eq. (3.28) can be absorbed by rescaling the fields. Specifically, we write the Lagrangian in terms of **bare fields** ϕ_0:

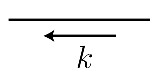
$$\mathcal{L}_0 = \frac{1}{2}\partial^\mu\phi_0\partial_\mu\phi_0 - \frac{m_0^2}{2}\phi_0^2 - \frac{\lambda_0}{3!}\phi_0^3, \tag{3.29}$$

where

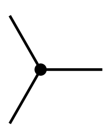
$$\phi_0 = \sqrt{Z_0}\phi. \tag{3.30}$$

Physical quantities (such as S-matrix elements or physical masses) are independent of the scale of the fields, but rescaling the fields is necessary if we want to make correlation functions independent of the cutoff. One could in principle work directly in terms of physical quantities where the dependence on the scale of the fields cancels out, but it is more convenient for most purposes to work with correlation functions and renormalize the scale of the fields.

With $Z_0 \neq 1$, the Feynman rules for the theory are modified by powers of Z_0:

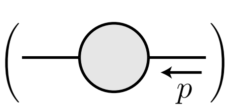


$$= \frac{iZ_0^{-1}}{p^2 - m_0^2 + i\epsilon}, \tag{3.31}$$



$$= -i(Z_0)^{3/2}\lambda_0. \tag{3.32}$$

We therefore obtain



$$\left(\cdots \right)^{-1} = Z_0\left\{p^2 - m_0^2 + \frac{\hbar\lambda_0^2}{256\pi^3}\left[\Lambda^2 + \left(2m_0^2 - \tfrac{1}{3}p^2\right)\ln\Lambda^2\right]\right\} + (\Lambda \text{ independent}) + \mathcal{O}(M_0^2/\Lambda^2) + \mathcal{O}(\hbar^2). \tag{3.33}$$

Note that the full inverse propagator including the 1-loop corrections is proportional to Z_0; it is not hard to show that this exact.

We can now make the propagator independent of Λ by allowing both Z_0 and m_0^2 to depend on Λ. Demanding that the coefficient of the p^2 term is independent of Λ gives

$$\boxed{\Lambda \frac{d \ln Z_0}{d\Lambda} = -\frac{\hbar \lambda_0^2}{384\pi^3} + \mathcal{O}(\hbar^2).} \tag{3.34}$$

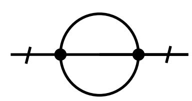
In order for the p-independent term to be independent of Λ, we require

$$\Lambda \frac{d}{d\Lambda}\left(Z_0 m_0^2\right) = \frac{\hbar Z_0 \lambda_0^2}{256\pi^3}\left(2\Lambda^2 - 4m_0^2\right). \tag{3.35}$$

Using the result Eq. (3.34), this gives

$$\boxed{\Lambda \frac{d m_0^2}{d\Lambda} = \frac{\hbar \lambda_0^2}{128\pi^3}\left(\Lambda^2 - \tfrac{5}{3} m_0^2\right).} \tag{3.36}$$

Note that if we had gone to 2-loop order in ϕ^4 theory, we would also require wavefunction renormalization. This is due to diagrams such as



in which external momentum flows through the loops. These diagrams have nontrivial dependence on the external momentum, and require wavefunction renormalization to make them Λ independent.

3.4 Locality of Divergences

Let us summarize what we have found so far. We computed various 1-loop divergent diagrams using a cutoff regulator and found that the Λ dependence could be absorbed into redefinitions of the couplings in the Lagrangian. This was possible because of the simple structure of the terms that diverge as $\Lambda \to \infty$ with the bare couplings held fixed. (We call this the **divergent part** of the diagram.) The divergent part of all of the diagrams above always has the structure of a polynomial in momenta. This is exactly the form of a tree-level contribution from a local Lagrangian, so we call such contributions **local**. In particular, we found that the divergent parts had precisely the form of local terms that are already present in the Lagrangian, and this was the reason we were able to absorb the Λ dependence into a redefinition of the couplings.

Theories where the Λ dependence can be absorbed into redefinitions of the coupling constants are called **renormalizable**. This structure is very general, as we will see.

For example, let us show that the divergent part of *any* 1-loop diagram is local. Consider any divergent 1-loop diagram, defined by imposing some regulator. The key observation is that the diagram can be made more convergent by differentiating with respect to external momenta. To see this, note that loop propagators have the general form $i/((k+p)^2 - m^2)$, where k is the loop momentum and p is some linear combination of external momenta. Because

$$\begin{aligned} \frac{\partial}{\partial p^\mu}\left(\frac{i}{(k+p)^2 - m^2}\right) &= -\frac{2i(k_\mu + p_\mu)}{[(k+p)^2 - m^2]^2} \\ &\sim \frac{1}{k^3} \quad \text{as } k \to \infty, \end{aligned} \tag{3.37}$$

we see that differentiating with respect to external momenta always increases the number of powers of k in the denominator. This is sufficient to conclude that if we differentiate the diagram sufficiently many times with respect to external momenta, the diagram will converge.

Schematically, if $I(p)$ is the integral corresponding to a particular Feynman diagram with external momenta p, then

$$\left(\frac{\partial}{\partial p}\right)^n I(p) = f(p) + \mathcal{O}(p/\Lambda), \tag{3.38}$$

where $f(p)$ is finite and independent of Λ. ($f(p)$ is the limit of the left-hand side as $\Lambda \to \infty$.) Integrating this equation n times with respect to the external momenta, we obtain

$$I(p) = \tilde{f}(p) + P_n(p) + \mathcal{O}(p/\Lambda). \tag{3.39}$$

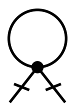
where $\tilde{f}(p)$ is finite and independent of Λ, and $P_n(p)$ is an n^{th} order polynomial in the external momenta that contains the n constants of integration. We see that all of the terms that diverge as $\Lambda \to \infty$ must be in the constants of integration, and are therefore polynomials in the momenta. The finite part $\tilde{f}(p)$ contains all of the nontrivial non-analytic structure of the diagram.

This discussion also explains why the choice of regulator is unimportant. If we evaluate the same diagram using two different regulators, both regulators will give the same function $f(p)$ in Eq. (3.38) simply because they must give the same value for convergent diagrams in the limit $\Lambda \to \infty$. This means that after integration, the results will be the same up to the integration constants in the polynomial P_n in

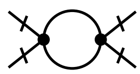
Eq. (3.39). Thus, two different regulators will differ only by local terms, which can be absorbed in redefinitions of the couplings. It is reassuring that the non-trivial non-analytic structure of loop diagrams discussed earlier unaffected by the regulator.

3.5 Power Counting

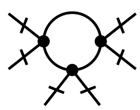
We now consider the question of which diagrams diverge at one loop in general scalar field theories in d spacetime dimensions. For any 1PI diagram, define the **superficial degree of divergence** D to be the total number of powers of loop momenta in the numerator minus the number of powers in the denominator. For example, in ϕ^4 theory in $d = 4$, we have



$$\sim \int \frac{d^4k}{k^2} \qquad D = 2,$$



$$\sim \int \frac{d^4k}{k^4} \qquad D = 0, \tag{3.40}$$



$$\sim \int \frac{d^4k}{k^6} \qquad D = -2,$$

etc. A 1-loop diagram can have a ultraviolet divergence only if $D \geq 0$. However, there are important cases where ultraviolet divergences in different diagrams contributing to the same correlation function cancel as a result of symmetries, so the divergences may be less severe than indicated by the value of D. Also, ultraviolet divergences in higher-loop diagrams are more subtle because they gave more than one loop momentum, and there can be **subdivergences** where one loop momentum gets large while the others remain finite. These are the reasons for calling D the 'superficial' degree of divergence.

D can be determined simply by dimensional analysis because it is obtained by neglecting all external momenta and masses in the propagators. For any diagram contributing to the 1PI correlation function $\Gamma^{(n)}$, we have

$$D = [\Gamma^{(n)}] - [\text{couplings}], \tag{3.41}$$

where $[\cdot]$ denotes the mass dimension, and 'couplings' denotes the product of couplings that appear in the diagram. The divergence structure therefore depends crucially on whether the couplings have positive or negative mass dimension.

Let us do the dimensional analysis for couplings and 1PI correlation functions for an arbitrary scalar field theory in d spacetime dimensions. The Lagragian has mass dimension d, so demanding that the free field theory Lagrangian has the correct dimensions gives

$$[\phi] = \frac{d-2}{2}. \tag{3.42}$$

The most general interaction Lagrangian can be written schematically as

$$\mathcal{L}_{\text{int}} \sim \sum_{n \geq 3} \sum_{p} \lambda_{(n,p)} \partial^p \phi^n. \tag{3.43}$$

That is, the most general term contains $n \geq 3$ powers of ϕ and p derivatives. The mass dimensions of the couplings are therefore

$$[\lambda_{(n,p)}] = d - n\frac{d-2}{2} - p. \tag{3.44}$$

The dimension of the 1PI correlation function $\Gamma^{(n)}$ is the same as $\lambda_{(n,0)}$, as can be seen from the fact that if $\lambda_{(n,0)} \neq 0$, there is a tree-level contribution $\Gamma^{(n)} \sim \lambda_{(n,0)}$. Thus,

$$[\Gamma^{(n)}] = [\lambda_{(n,0)}] = d - n\frac{d-2}{2}. \tag{3.45}$$

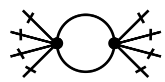
Now consider a theory where all couplings have mass dimension ≥ 0. From Eq. (3.41), we see that the dimension of the couplings cannot increase the degree of divergence, and the only divergent n-point functions are those with $[\Gamma^{(n)}] \geq 0$. But these are in one-to-one correspondence with the couplings in the theory as long as we include *all* couplings (consistent with the symmetries of the theory) that have dimension ≥ 0. In these theories, all 1-loop divergences can be absorbed into redefinitions of the couplings in the lagrangian, and we say that these theories are **renormalizable** at one loop.

Renormalizability at one loop order is clearly a necessary condition for renormalizability at all orders. We can therefore already conclude that the number of renormalizable theories is quite limited. The following is a list of all interactions with dimension ≥ 0 for all $d \geq 2$:

$$\begin{aligned}
d = 2: &\quad \sum_{n \geq 3} \phi^n + \left(\sum_{n \geq 3} \partial^2 \phi^n \right), \\
d = 3: &\quad \phi^3 + \phi^4 + \phi^5 + \left(\phi^6\right), \\
d = 4: &\quad \phi^3 + \left(\phi^4\right), \\
d = 5: &\quad \phi^3, \\
d = 6: &\quad \left(\phi^3\right),
\end{aligned} \tag{3.46}$$

where the terms in parentheses have dimensionless couplings. For $d \geq 7$, there are no renormalizable scalar field theories.

Now consider theories that have at least one coupling with negative mass dimension. Then it is not hard to see that *all* $\Gamma^{(n)}$ have contributions with $D \geq 0$ if we go to sufficiently high order in perturbation theory. The reason is that we can always find a diagram that involves enough powers of the negative-dimension coupling to make $D \geq 0$ (see Eq. (3.41)). If we only consider 1-loop diagrams the conclusion is not quite so dramatic, but one still finds in general that there are divergent $\Gamma^{(n)}$ that do not correspond to any coupling in the Lagrangian. For example, consider adding a ϕ^6 term to scalar field theory in $d = 4$. Then at 1 loop, we have the diagram



$$\sim \int d^4k \, \frac{1}{k^4} \qquad D = 0 \tag{3.47}$$

that gives a logarithmically divergent contribution to the 8-point function. If we include a ϕ^8 coupling to absorb this divergence, we find that the 10- and 12-point functions also diverge at one loop, and so on. In theories such as this, we cannot absorb the Λ dependence without introducing an *infinite* number of interactions. These theories appear to have no predictive power, and are therefore called **non-renormalizable**. We will see later that these can also be renormalized in an appropriate sense.

3.6 Renormalized Perturbation Theory

Let us consider first a theory that is renormalizable in the sense defined above. We have found that in ϕ^4 theory in $3+1$ dimensions and in ϕ^3 theory in $5+1$ dimensions, all 1PI correlation functions can be written

$$\Gamma^{(n)}(p) = f^{(n)}(p; Z_0, m_0^2, \lambda_0, \Lambda) + \mathcal{O}(p/\Lambda), \tag{3.48}$$

where p are the external momenta. The function $f^{(n)}$ contains all the terms that grow with Λ or are independent of Λ. We have seen that (at one loop order) we can keep $f^{(n)}$ fixed by changing Λ and simultaneously changing the bare couplings. The rule for changing the couplings is given by renormalization group equations

$$\begin{aligned}
\frac{\Lambda}{Z_0}\frac{dZ_0}{d\Lambda} &= \gamma(\lambda_0), \\
\frac{\Lambda}{m_0^2}\frac{dm_0^2}{d\Lambda} &= \gamma_m(\lambda_0), \\
\Lambda\frac{d\lambda_0}{d\Lambda} &= \beta(\lambda_0).
\end{aligned} \tag{3.49}$$

The function β is called the **beta function** of the theory, and γ is called the **anomalous dimension**.[4]

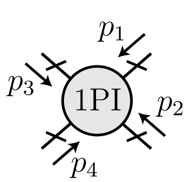
Following the 2-dimensional quantum mechanics example discussed in Subsection 1.2, we introduce **renormalized couplings** $Z_R(\mu)$, $m_R^2(\mu)$, and $\lambda_R(\mu)$ by

$$\begin{aligned} \frac{\mu}{Z_R}\frac{dZ_R}{d\mu} &= \gamma(\lambda_R), \qquad & Z_R(\mu = \Lambda) &= Z_0, \\ \frac{\mu}{m_R^2}\frac{dm_R^2}{d\mu} &= \gamma_m(\lambda_R), \qquad & m_R^2(\mu = \Lambda) &= m_0^2, \\ \mu\frac{d\lambda_R}{d\mu} &= \beta(\lambda_R), \qquad & \lambda_R(\mu = \Lambda) &= \lambda_0. \end{aligned} \tag{3.50}$$

Here μ is an arbitrary **renormalization scale**. In other words, the renormalized couplings are defined by evolving the bare couplings from Λ to the scale μ using the renormalization group equations. The renormalization group equations leave $f^{(n)}$ in Eq. (3.48) invariant, so we can write

$$\Gamma^{(n)}(p) = f^{(n)}(p; Z_R(\mu), m_R^2(\mu), \lambda_R(\mu), \mu) + \mathcal{O}(p/\Lambda). \tag{3.51}$$

The point of introducing the renormalized couplings is that the scale μ can be taken to be close to the scale of the external momenta p. If we do this, there is no longer a large scale appearing in the 1-loop corrections in $f^{(n)}$, and so the corrections expressed in terms of the renormalized couplings are under control. For example, the full 1PI 4-point function in ϕ^4 theory in $3 + 1$ dimensions becomes (see Eq. (3.17))



$$= -i\lambda_R(\mu) + \frac{i\hbar\lambda_R^2(\mu)}{32\pi^2}\int_0^1 dx \left[3 + \ln\frac{\mu^2}{M^2(s)} + \ln\frac{\mu^2}{M^2(t)}\ln\frac{\mu^2}{M^2(u)} + \mathcal{O}(1/\Lambda^2)\right] + \mathcal{O}(\hbar^2), \tag{3.52}$$

where $M(p^2) = m_R^2(\mu) - x(1-x)p^2$. The perturbative expansion expressed in terms of the renormalized couplings is called **renormalized perturbation theory**.

The elimination of the cutoff dependence in Eq. (3.51) may appear somewhat magical, but it is actually quite simple. The perturbative expansion Eq. (3.48) expressed

[4]The function γ_m is sometimes called the **mass anomalous dimension**. Conventions for β, γ, and γ_m in the literature sometimes differ by signs and factors of 2.

in terms of bare quantities has large corrections proportional to Λ^2 or $\ln \Lambda$. The renormalization group equation tells us that these large corrections can be absorbed into a redefinition of the couplings. Expressing the series in terms of renormalized couplings defined using the renormalization group 'resums' these large corrections and gives a Λ independent result.

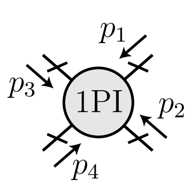
The renormalized results depend on a renormalization scale μ. The scale μ is arbitrary in the sense that if we could sum the entire perturbative expansion (or better: solve the theory exactly) the result would be independent of μ. However, the choice $\mu \sim p$ is special because ensures that there are no large logarithms in the loop corrections. Changing μ changes the relative size of different terms in the perturbative expansion, and choosing $\mu \sim p$ can be thought of as optimizing the convergence of the perturbative expansion.

In fact, the renormalized couplings defined above include corrections to all orders $\hbar$, since the beta function and anomalous dimensions are proportional to $\hbar$.[5] This means that it is crucial to understand the corrections that are higher order in $\hbar$ to understand whether these higher-order corrections really do have the structure predicted by the renormalization group equations. We will see that renormalization beyond one loop involves highly nontrivial issues that are best understood using the 'exact' renormalization group of K. Wilson. We will show that renormalization group equations of the form Eq. (3.49) can be defined to all orders in perturbation theory, fully justifying the arguments above.

3.7 Counterterm Renormalization

We now relate the discussion of renormalization above to the language usually used in practical calculations (and most textbooks). The key observation is that in passing from the bare to renormalized expressions, we only need to know the relation between the bare and renormalized couplings to $\mathcal{O}(\hbar)$.

Consider for example the expression for the 1PI 4-point function at one loop expressed in terms of bare quantities (see Eq. (3.17)):



$$= -i\lambda_0 + \frac{i\hbar\lambda_0^2}{32\pi^2}\int_0^1 dx \left[3 + \ln\frac{\Lambda^2}{M_0^2(s)} + \ln\frac{\Lambda^2}{M_0^2(t)}\ln\frac{\Lambda^2}{M_0^2(u)}\right.$$

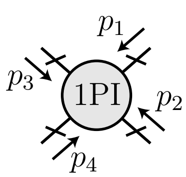
[5] This will become completely obvious when we solve the renormalization group equations below.

$$+ \mathcal{O}(1/\Lambda^2)\Big] + \mathcal{O}(\hbar^2), \tag{3.53}$$

where $M_0(p^2) = m_0^2 - x(1-x)p^2$. To express this in terms of renormalized couplings, note that the renormalization group guarantees that the bare couplings can be re-expressed in terms of renormalized couplings to absorb the Λ dependence. Since we are neglecting terms of order $\hbar^2$ and higher, we only need to know this relation to order $\hbar$. This is easily read off from Eq. (3.53):

$$\lambda_R(\mu) = \lambda_0 - \frac{3i\hbar\lambda_0^2}{32\pi^2} \ln\frac{\Lambda^2}{\mu^2} + \mathcal{O}(\hbar^2). \tag{3.54}$$

(Note that this is equivalent solving the renormalization group equation Eq. (3.20) to linear order in $\hbar$.) Substituting into Eq. (3.53) we obtain



$$= -i\lambda_R(\mu) + \frac{i\hbar\lambda_0^2}{32\pi^2}\int_0^1 dx \left[3 + \ln\frac{\mu^2}{M_0^2(s)} + \ln\frac{\mu^2}{M_0^2(t)} \ln\frac{\mu^2}{M_0^2(u)} + \mathcal{O}(1/\Lambda^2)\right] + \mathcal{O}(\hbar^2). \tag{3.55}$$

The large Λ-dependent terms have disappeared, but the expression still depends on the bare couplings. However, since λ_0 and λ_R differ only by terms of order $\hbar$ (see Eq. (3.54)), we simply replace λ_0 by λ_R in the $\mathcal{O}(\hbar)$ term of Eq. (3.55). In this way, we exactly reproduce Eq. (3.52), obtained by renormalization group arguments above.

The steps just described have caused unease in innumerable students of quantum field theory and more than a few researchers. The reason is that λ_0 and λ_R differ by a large 'divergent' quantity, and it is far from clear that it is correct to neglect the higher-order corrections in Eq. (3.54). However, the renormalization group argument given above justifies this (provided that we can prove the validity of the renormalization group to all orders in $\hbar$.)

The steps above can be reformulated in a way that is very useful for practical calculation. We write the bare Lagrangian as

$$\mathcal{L}_0 = \mathcal{L}_R + \Delta\mathcal{L}, \tag{3.56}$$

where (for ϕ^4 theory)

$$\mathcal{L}_0 = \frac{Z_0}{2}\partial^\mu\phi\partial_\mu\phi - \frac{Z_0 m_0^2}{2} m_0^2\phi^2 - \frac{Z_0^2\lambda_0}{4!}\phi^4 \tag{3.57}$$

is the bare Lagrangian, and

$$\mathcal{L}_R = \frac{1}{2}\partial^\mu \phi \partial_\mu \phi - \frac{m_R^2}{2}\phi^2 - \frac{\lambda_R}{4!}\phi^4 \tag{3.58}$$

is the **renormalized Lagrangian**, and

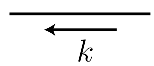
$$\Delta\mathcal{L} = \frac{\Delta Z}{2}\partial^\mu \phi \partial_\mu \phi - \frac{\Delta m^2}{2} m_0^2 \phi^2 - \frac{\Delta\lambda}{4!}\phi^4 \tag{3.59}$$

is the **counterterm Lagrangian**. In other words, we have written the bare couplings as

$$Z_0 = 1 + \Delta Z, \quad Z_0 m_0^2 = m_R^2 + \Delta m^2, \quad Z_0^2 \lambda_0 = \lambda_R + \Delta\lambda. \tag{3.60}$$

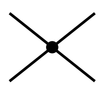
The general results above tell us that we can choose the counterterms ΔZ, Δm^2 and $\Delta\lambda$ as functions of the renormalized couplings m_R^2 and λ_R order by order in perturbation theory to cancel the large Λ dependent terms.

Since the counterterms start at $\mathcal{O}(\hbar)$, the tree-level Feynman rules are written in terms of the renormalized parameters. We therefore have propagator



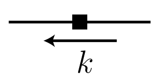
$$= \frac{i}{k^2 - m_R^2 + i\epsilon}, \tag{3.61}$$

and vertex

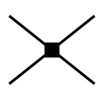


$$= -i\lambda_R. \tag{3.62}$$

The counterterms are treated as additional vertices that will be determined order-by-order in $\hbar$ to cancel the divergences:

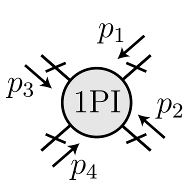


$$= i(\Delta Z k^2 - \Delta m^2),$$

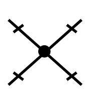


$$= -i\Delta\lambda. \tag{3.63}$$

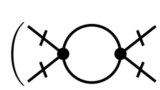
In this language, the 1-loop corrections to the 4-point function are



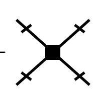
$=$



$+ \Big($



$+ \text{crossed}\Big) +$



$$+ \mathcal{O}(\hbar^2). \tag{3.64}$$

Choosing

$$\Delta\lambda = -\frac{3\hbar\lambda_R^2}{32\pi^2}\ln\frac{\Lambda^2}{\mu^2} + \mathcal{O}(\hbar^2) \tag{3.65}$$

cancels the divergence and gives Eq. (3.52).

We now show how to properly interpret couplings with negative mass dimension. The key point is that such couplings become less important at low momenta. If κ is a coupling with negative mass dimension $-n$, we write

$$\kappa = \frac{1}{M^n}, \tag{3.66}$$

where M is a mass scale that gives the strength of the coupling. At tree level, dimensional analysis tells us that the dimensionless quantity that characterizes the importance of the coupling κ as a perurbation is $(p/M)^n$, where p is a physical momentum (or mass) scale. For sufficiently small p, this is a small perturbation.

These simple arguments appear to be invalidated by loop corrections, which can give effects proportional to $(\Lambda/M)^n$, where Λ is the cutoff. However, the argument of Subsection 3.4 tells us that the divergent part of these diagrams is local, *i.e.* it is a polynomial in the external momenta. The Λ dependence can therefore be absorbed into redefinitions of the couplings as before. We can therefore express the results in terms of a renormalized expansion with no Λ dependence. If all the renormalized higher-dimensional couplings are of order $1/M$ raised to the appropriate power, then for external momenta $p \ll M$ their effects can be included in a systematic expansion in p/M.

In Subsection 3.5 we noted that a theory that has couplings with negative mass dimension requires an infinite number of couplings to absorb all of the Λ dependence. However, if we are satisfied with a fixed order in the p/M expansion, there are only a finite number of couplings that contribute. In this sense the theory is predictive.

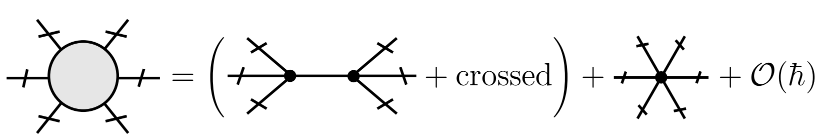
Let us illustrate these points with an example. We add a bare ϕ^6 term to a scalar field theory in $d = 4$:

$$\Delta\mathcal{L}_0 = -\frac{\lambda_{6,0}}{6!}\phi^6, \tag{3.67}$$

with

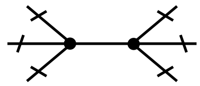
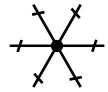
$$\lambda_{6,0} = \frac{1}{M^2}. \tag{3.68}$$

At tree level, this gives a new contribution to the full 6-point function:



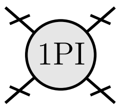
(3.69)

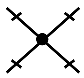
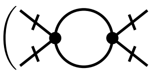
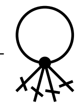
By dimensional analysis,

 $\sim \dfrac{\lambda_{4,0}^2}{p^2}$,  $\sim \dfrac{1}{M^2}$, (3.70)

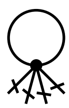
so we see that the ϕ^6 coupling gives a small perturbation if $p \ll M$.

The situation becomes a bit more subtle if we include loop corrections. For example, the ϕ^6 term gives a new contribution to the 1PI 4-point function at one loop:



$=$  $+$ $\Big($  $+$ crossed $\Big)$ $+$  $+\, \mathcal{O}(\hbar^2)$. (3.71)

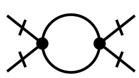
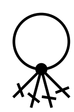
The new contribution is easily evaluated:



$$= \frac{-i\lambda_{6,0}}{2} \int \frac{d^4 k}{(2\pi)^4} \frac{i}{k^2 - m_0^2}$$

$$= -\frac{i}{32\pi^2 M^2} \left[\Lambda^2 + m^2 \ln \frac{\Lambda^2}{m_0^2} + \mathcal{O}(m_0^2/\Lambda^2) \right]. \tag{3.72}$$

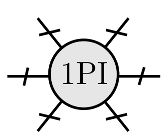
Compare this to the loop correction from the graph involving only the 4-point function (see Eq. (3.17) for the precise expression):

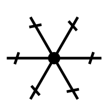
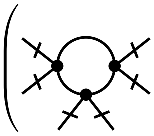
 $\sim \dfrac{\lambda_{4,0}^2}{16\pi^2} \ln \Lambda$,  $\sim \dfrac{\Lambda^2}{16\pi^2 M^2}$. (3.73)

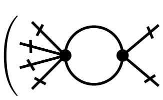
We see that the new contribution Eq. (3.72) cannot be treated as a perturbation if $\Lambda \gtrsim M$. However, the large Λ-dependent contribution is independent of momentum, and can therefore be absorbed into a redefinition of the 4-point coupling. Specifically, the renormalization group equation for $\lambda_{4,0}$ is modified to

$$\Lambda \frac{d\lambda_{4,0}}{d\Lambda} = \frac{3\hbar \lambda_{4,0}^2}{16\pi^2} - \frac{\hbar}{16\pi^2 M^2} \left(\Lambda^2 + m_0^2 \right) + \mathcal{O}(\hbar^2). \tag{3.74}$$

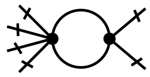
Let us consider the 6-point function. At one loop, the 1PI 6-point function is given by



$=$  $+$ $\Bigg($  $+$ crossed $\Bigg)$

$+$ $\Big($  $+$ crossed $\Big)$ $+\, \mathcal{O}(\hbar^2)$. (3.75)

The only divergent graph is the new graph involving the ϕ^6 coupling:



$$\sim \frac{1}{M^2}\int\frac{d^4k}{(2\pi)^4}\,\frac{1}{k^2} \sim \frac{1}{16\pi^2 M^2}\ln\Lambda. \tag{3.76}$$

The momentum integral is exactly the same as in Eq. (3.11); we will not bother to keep track of symmetry factors and crossed diagrams. The important point for our present discussion is that the divergence is independent of momentum. This can be seen by direct calculation (see Eq. (3.16)) or from the general argument of Subsection 3.4. This divergence can therefore be absorbed into a redefinition of $\lambda_{6,0}$. We obtain a renormalization group equation

$$\Lambda\frac{d\lambda_{6,0}}{d\Lambda} = \frac{\hbar b_6 \lambda_{4,0}\lambda_{6,0}}{16\pi^2} + \mathcal{O}(\hbar^2), \tag{3.77}$$

where b_6 is an order-1 constant that we will not compute here.

The renormalization group equations above ensure that the 1PI correlation functions $\Gamma^{(n)}$ for $n = 2, 4, 6$ can be made independent of Λ by adjusting the couplings. We can define renormalized couplings using the same procedure followed for the theory without the ϕ^6 coupling:[6]

$$\begin{aligned}
\frac{\mu}{m_R^2}\frac{dm_R^2}{d\mu} &= \gamma_m(\lambda_{4,R},\lambda_{6,R}), \qquad & m_R^2(\mu=\Lambda) &= m_0^2,\\
\mu\frac{d\lambda_{4,R}}{d\mu} &= \beta_4(\lambda_{4,R},\lambda_{6,R}), \qquad & \lambda_{4,R}(\mu=\Lambda) &= \lambda_{4,0},\\
\mu\frac{d\lambda_{6,R}}{d\mu} &= \beta_6(\lambda_{4,R},\lambda_{6,R}), \qquad & \lambda_{6,R}(\mu=\Lambda) &= \lambda_{6,0}.
\end{aligned} \tag{3.78}$$

If we take $\mu \sim p$, we obtain renormalized predictions for the 2-, 4-, and 6-point functions, with no large corrections.

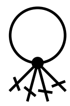
Let us estimate the size of the corrections from the ϕ^6 coupling in the renormalized expansion. An important point is that the renormalization group equation for $\lambda_{6,R}$ gives $\mu d\lambda_{6,R}/d\mu \propto \lambda_{6,R}$, so

$$\lambda_{6,R} \sim \lambda_{6,0}\ln\Lambda. \tag{3.79}$$

In the remainder of this Subsection, we will be interested in keeping track of power corrections, so we ignore the logarithmic corrections and write $\lambda_{6,R} \sim 1/M^2$.

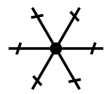
[6] Note that at this order there is no need for wavefunction renormalization in this theory.

In the renormalized expansion defined above, the correction to the 1PI 4-point function is of order



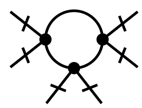
$$\sim \frac{\lambda_{6,R}}{16\pi^2} m_R^2 \sim \frac{1}{16\pi^2} \frac{m_R^2}{M^2}. \tag{3.80}$$

This is a small correction as long as $m_R \ll M$. The 1PI 6-point function has a new tree-level contribution



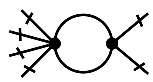
$$\sim \lambda_{6,R} \sim \frac{1}{M^2}. \tag{3.81}$$

This should be compared to the (finite) 1-loop contribution



$$\sim \frac{\lambda_{4,R}^3}{16\pi^2} \frac{1}{p^2}. \tag{3.82}$$

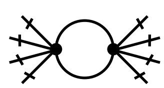
Again we see that the contribution due to the ϕ^6 term is small if $p \ll M$. The loop correction involving the ϕ^6 term is even smaller:



$$\sim \frac{\lambda_{4,R}\lambda_{6,R}}{16\pi^2} \sim \frac{\lambda_{4,R}}{16\pi^2} \frac{1}{M^2}. \tag{3.83}$$

We conclude that in the renormalized expansion, the ϕ^6 coupling gives corrections to the 2-, 4-, and 6-point functions that are suppressed by p^2/M^2 (times logs), even if 1-loop effects are included.

We now come to the 8-point function and the apparent difficulty discussed at the end of Subsection 3.5. The 1PI 8-point function now has a divergent contribution



$$\sim \frac{1}{M^4} \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^4} \sim \frac{1}{16\pi^2 M^4} \ln \Lambda. \tag{3.84}$$

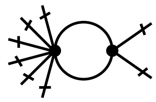
The Λ dependence can only be cancelled only if we introduce a ϕ^8 term into the Lagrangian:

$$\Delta \mathcal{L}_0 = \frac{\lambda_{8,0}}{8!} \phi^8. \tag{3.85}$$

If we do this, then the Λ dependence in the 8-point function can be absorbed into a redefinition of the 8-point coupling, defined by the renormalization group equation

$$\Lambda \frac{d\lambda_{8,0}}{d\Lambda} = \frac{\hbar b_8 \lambda_{6,0}^2}{16\pi^2} + \frac{\hbar b_8' \lambda_{8,0} \lambda_{4,0}}{16\pi^2} + \mathcal{O}(\hbar^2), \tag{3.86}$$

where b_8 and b'_8 are order-1 constants that we will not compute. The contribution proportional to $\lambda_{8,0}\lambda_{4,0}$ comes from the diagram



$$\sim \lambda_{4,0}\lambda_{8,0}\int\frac{d^4k}{(2\pi)^4}\,\frac{1}{k^4} \sim \frac{\hbar\lambda_{4,0}\lambda_{8,0}}{16\pi^2}\ln\Lambda. \tag{3.87}$$

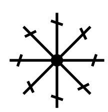
Having introduced the 8-point function, we obtain sensible predictions for the 2-, 4-, 6-, and 8-point functions. To estimate these corrections, we will assume that for $\Lambda \sim M$,

$$\lambda_{6,0} \sim \frac{1}{M^2}, \qquad \lambda_{8,0} \sim \frac{1}{M^4}. \tag{3.88}$$

The idea behind this assumption is that the effects parameterized by the couplings $\lambda_{6,0}$ and $\lambda_{8,0}$ arise from a more fundamental theory at the scale M in which the dimensionless couplings are of order 1. Following the discussion above, we can then define renormalized couplings $\lambda_{4,R}$, $\lambda_{6,R}$, and $\lambda_{8,R}$ using the renormalization group equations above. Because $\beta_6 \propto 1/M^2$, $\beta_8 \propto 1/M^4$, we have (ignoring logarithms)

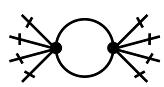
$$\lambda_{6,R} \sim \frac{1}{M^2}, \qquad \lambda_{8,R} \sim \frac{1}{M^4}. \tag{3.89}$$

The ϕ^8 coupling gives a tree-level contribution to the 8-point function



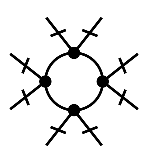
$$\sim \lambda_{8,R} \sim \frac{1}{M^4}, \tag{3.90}$$

and the ϕ^6 coupling gives a 1-loop contribution



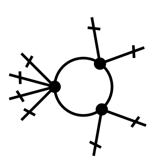
$$\sim \frac{\lambda_{6,R}^2}{16\pi^2} \sim \frac{1}{16\pi^2}\,\frac{1}{M^4}. \tag{3.91}$$

These should be compared to the (finite) loop contribution from the ϕ^4 coupling:



$$\sim \frac{\lambda_{4,R}^4}{16\pi^2}\,\frac{1}{p^4}. \tag{3.92}$$

There is also a new finite contribution to the 8-point function



$$\sim \frac{\lambda_{4,R}^2\lambda_{6,R}}{16\pi^2}\,\frac{1}{p^2} \sim \frac{\lambda_{4,R}^2}{16\pi^2}\,\frac{1}{M^2p^2}. \tag{3.93}$$

We see that all the contributions from the ϕ^8 term are suppressed by p^4/M^4 compared to the contributions of the renormalizable terms. If we are satisfied with accuracy p^2/M^2, we can ignore *all* diagrams involving the ϕ^8 couplings.

All of this is really just dimensional analysis. The coupling of a ϕ^n term has dimension $4 - n$, and if we assume

$$\lambda_{n,0} \sim \frac{1}{M^{4-n}}, \tag{3.94}$$

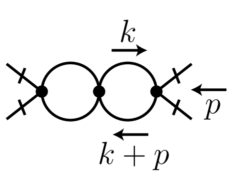
for $\Lambda \sim M$, then the effects of many insertions of $\lambda_{n,0}$ (for $n > 4$) will be suppressed by many powers of p/M. The only subtlety is that divergent loop effects can overcome this suppression by giving rise to terms of the form Λ/M. But these can be absorbed by redefining the couplings, and in the resulting renormalized perturbative expansion, powers of Λ do not appear. If we are interested in keeping only terms of order p^2/M^2 (say) in this expansion, we must only keep couplings with mass dimension -2 or greater. There are only a finite number of these couplings, so the expansion is predictive.

4 Beyond One Loop

The above discussion has been limited to 1-loop corrections. We now show that new subtleties appear at 2-loop order (and higher), but it can be shown that they do not upset the features found at 1-loop order in the previous Section.

4.1 Subdivergences

An important ingredients of the discussion above was that the divergent part of any 1-loop diagram is local, *i.e.* a polynomial in the external momenta. This breaks down at 2-loop order and higher, as can be seen from simple examples. For example, consider the following 2-loop diagram in ϕ^4 theory:



$$\propto \lambda_0^3 \left(\int \frac{d^4k}{(2\pi)^4} \frac{1}{(k+p)^2 - m_0^2} \frac{1}{k^2 - m_0^2} \right)^2$$

$$\propto \lambda_0^3 \left(\int_0^1 dx \left[\ln \frac{\Lambda^2}{M_0^2(s)} + \ln \frac{\Lambda^2}{M_0^2(s)} + \ln \frac{\Lambda^2}{M_0^2(s)} \right] + \cdots \right)^2 \tag{4.1}$$

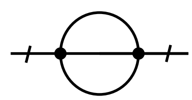
Each of the factors in parentheses can be written as a local divergent term plus a finite term, but the product contains terms such as

$$\ln \Lambda \int_0^1 dx \, \ln M_0^2(s) \tag{4.2}$$

from the cross terms. These terms are not polynomials in momenta, so the divergent part of this diagram is not local. Terms such as Eq. (4.2) arise from regions of loop momentum integration where one of the loop momenta is getting large and the other remains finite. This is called a **subdivergence**.

It is instructive to understand how the general argument of Subsection 3.4 fails for diagrams such as Eq. (4.1). Differentiating the diagram Eq. (4.1) with respect to external momenta does not make the diagram converge. This is because no matter how many times we differentiate, we get terms where all the derivatives act on one of the factors, while the other still diverges.

More generally, multiloop diagrams such as



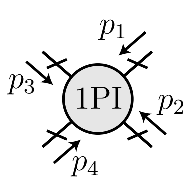
(4.3)

have subdivergences arising from regions of loop momentum integation where one momentum is large and the other is small. In general higher-loop diagrams, the integral is not a simple product, and the subdivergences are said to be **overlapping divergences**.

4.2 Cancelation of Subdivergences

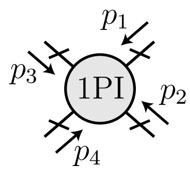
We now explain why nonlocal subdivergences such as those discussed above do not lead to a breakdown of renormalizability. We will not give a rigorous argument, but we simply sketch the main idea. We will obtain a better understanding for the cancellation mechanism discussed below when we discuss the 'exact' renormalization group.

Let us go back once again to the expression for the 1PI 4-point function in ϕ^4 theory in terms of bare quantities (see Eq. (3.17))



$$= -i\lambda_0 + \frac{i\hbar\lambda_0^2}{32\pi^2}\int_0^1 dx \left[3 + \ln\frac{\Lambda^2}{M_0^2(s)} + \ln\frac{\Lambda^2}{M_0^2(t)} \ln\frac{\Lambda^2}{M_0^2(u)} + \mathcal{O}(1/\Lambda^2) \right] + \mathcal{O}(\hbar^2). \tag{4.4}$$

We have checked that to $\mathcal{O}(\hbar)$ the Λ dependence can be absorbed into a redefinition of the bare couplings:

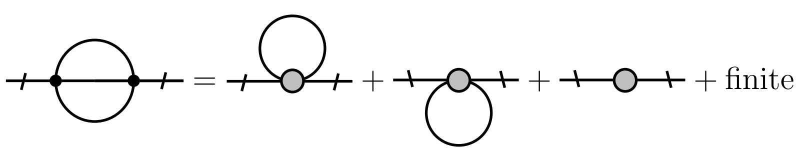
$$0 = \Lambda \frac{d}{d\Lambda}$$  (4.5)

Suppose we wish to check this at $\mathcal{O}(\hbar^2)$. To do this, we would have to compute all the 2-loop diagrams. However, from Eq. (4.4) we can see a non-local $\mathcal{O}(\hbar^2)$ term arising when $\Lambda d/d\Lambda$ acts on λ_0 in the $\mathcal{O}(\hbar)$ term of Eq. (4.4):

$$0 = \cdots - \frac{i\hbar\lambda_0}{16\pi^2}\, \Lambda \frac{d\lambda_0}{d\Lambda} \int_0^1 dx \ln M_0^2(s) + \cdots \tag{4.6}$$

Since $\Lambda d\lambda_0/d\Lambda \sim \hbar\lambda_0^2$, this term has exactly the same form as the term arising when $\Lambda d/d\Lambda$ acts on $\ln\Lambda$ in 2-loop subdivergences such as Eq. (4.2) above. In fact, the claim is that the nonlocal terms arising from the 1-loop expression above exactly cancel the non-local terms from 2-loop subdivergences.

We will prove this statement using the exact renormalization group below. For now we will simply give a plausibility argument. As discussed above, the subdivergences that arise at 2 loops come from a region of loop momenta where one loop momentum is larger than the other. When the momentum in one loop becomes large, we can approximate that loop by a local term, which we indicate diagramatically by shrinking the loop to a point. Therefore, we can write *e.g.*

 (4.7)

where the shaded blobs denote subdivergences. (The third diagram indicates the **overal divergence** that occurs when all loop momenta become large.) Each of the blobs is given by a local 1-loop expression, so the 2-loop subdivergence is the product of 1-loop expressions. The Λ dependence of this product of 1-loop expressions exactly cancels the higher-order Λ dependence in 1-loop expressions such as Eq. (4.4) above. This ensures that at 2-loop order, the only new Λ dependence comes from the overall divergence. This is local, so the Λ dependence can be cancelled by a local counterterm, even at 2-loop order.

The main result of perturbative renormalization theory is that this cancelation of subdivergences persists to all orders in the loop expansion. This ensures that the Λ dependence can be cancelled by local counterterms to all orders in the loop expansion.

5 The Exact Renormalization Group

We now introduce another version of the renormalization group, due to K. Wilson. It is useful because it gives a powerful and intuitive way of looking at renormalization in general. In particular, it clarifies the cancelation of subdivergences discussed above.

5.1 The Wilson Effective Action

Wilson's idea is to consider the operation of lowering the cutoff Λ while keeping the physics at scales below Λ fixed. We will discuss this idea in the context of Euclidean scalar field theory with a momentum space cutoff. We can write the path integral as

$$Z[J] = \int d[\phi]\, e^{-S[\phi] - \int J\phi}, \tag{5.1}$$

where we integrate only over Euclidean momentum modes with momenta below Λ:

$$d[\phi] = \prod_{k^2 < \Lambda^2} d\tilde{\phi}_k, \tag{5.2}$$

where

$$\tilde{\phi}_k \stackrel{\text{def}}{=} \int d^4x\, e^{ik\cdot x} \phi(x) \tag{5.3}$$

are the momentum-space fields. (We drop the subscript 'E' for 'Euclidean' in this Section.) Since we are interested in probing only momenta far below Λ, we assume that

$$\tilde{J}_k \neq 0 \text{ only for } k^2 \ll \Lambda^2. \tag{5.4}$$

We want to write a different action with a different value of the cutoff $\Lambda' < \Lambda$ that gives rise to the same low-momentum physics. We can do this simply by 'integrating out' the modes with momenta $\Lambda'^2 < k^2 < \Lambda^2$. That is, we define

$$\boxed{e^{-S'[\phi]} \stackrel{\text{def}}{=} \int \left(\prod_{\Lambda'^2 < k^2 < \Lambda^2} d\tilde{\phi}_k \right) e^{-S[\phi]}.} \tag{5.5}$$

Then we have

$$Z[J] = \int \left(\prod_{k^2 < \Lambda'^2} d\tilde{\phi}_k \right) e^{-S'[\phi] - \int J\phi}. \tag{5.6}$$

This is nothing more than splitting up the integral into the contribution from modes with $0 \leq k^2 < \Lambda'^2$ and $\Lambda'^2 < k^2 < \Lambda^2$. (Note that the condition Eq. (5.4) ensures that the source term involves only momenta with $k^2 < \Lambda'^2$.) $S'[\phi]$ is the **Wilson effective action**, defined by integrating out the modes above the new cutoff Λ'. It has the property that the path integral with cutoff Λ' and action $S'[\phi]$ gives the same physics as the original theory with the cutoff Λ.

To get a feel for what $S'[\phi]$ is, let us compute it in perturbation theory. We can do this by writing

$$\phi = \phi' + \phi'', \tag{5.7}$$

where

$$\begin{aligned} \tilde{\phi}'_k &= \begin{cases} \tilde{\phi}_k & 0 \leq k^2 < \Lambda'^2, \\ 0 & k^2 > \Lambda'^2, \end{cases} \\ \tilde{\phi}''_k &= \begin{cases} 0 & 0 \leq k^2 < \Lambda'^2, \\ \tilde{\phi}_k & \Lambda'^2 < k^2 < \Lambda^2. \end{cases} \end{aligned} \tag{5.8}$$

That is, ϕ' consists of the modes we are keeping in the Wilsonian effective action, and ϕ'' consists of the modes we are integrating out. Let us suppose that the original action $S[\phi]$ is invariant under $\phi \mapsto -\phi$ and contains all renormalizable terms in 4 dimensions:

$$S[\phi] = \int d^4x \left[\frac{1}{2} \partial_\mu \phi \partial_\mu \phi + \frac{m^2}{2} \phi^2 + \frac{\lambda}{4!} \phi^4 \right]. \tag{5.9}$$

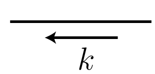
Substituting Eq. (5.7) for ϕ, the kinetic term becomes

$$S_{\text{free}}[\phi] = \int d^4x \left[\frac{1}{2} \partial_\mu \phi' \partial_\mu \phi' + \frac{m^2}{2} \phi'^2 + \frac{1}{2} \partial_\mu \phi'' \partial_\mu \phi'' + \frac{m^2}{2} \phi''^2 \right]. \tag{5.10}$$

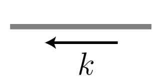
There are no cross terms because of the orthogonality of different Fourier modes. However, the interaction term does contain cross terms:

$$S_{\text{int}}[\phi] = \int d^4x \left[\frac{\lambda}{4!} \phi'^4 + \frac{\lambda}{3!} \phi'^3 \phi'' + \frac{\lambda}{4} \phi'^2 \phi''^2 + \frac{\lambda}{3!} \phi' \phi''^3 + \frac{\lambda}{4!} \phi''^4 \right]. \tag{5.11}$$

The Feynman rules for the theory in terms of the fields ϕ' and ϕ'' is easy to work out. We denote ϕ' propagators as ordinary solid lines, and ϕ'' propagators as thick grey lines:



$$= \frac{\theta(\Lambda^2 - k^2)}{k^2 + m^2},$$

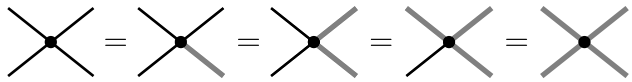


$$= \frac{\theta(k^2 - \Lambda^2)}{k^2 + m^2}, \tag{5.12}$$

where the θ functions

$$\theta(x) = \begin{cases} 1 & x < 1 \\ 0 & x > 1 \end{cases} \tag{5.13}$$

enforce the restriction on the momenta. The vertices are

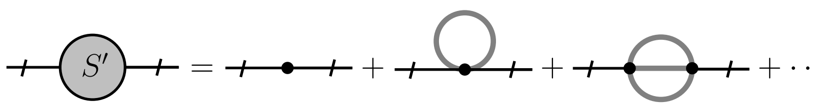


$$= -\lambda. \tag{5.14}$$

The diagrams that contribute to $S'[\phi]$ in Eq. (5.5) contain only ϕ'' internal lines, since only ϕ'' is integrated over. The field ϕ' is not integrated over, and therefore acts like an external source. From the exponentiation of connected diagrams contributing Eq. (5.5), we see that

$$-S'[\phi] = \sum (\text{connected diagrams with } \phi' \text{ sources}) \,. \tag{5.15}$$

For example, the term in $S'[\phi]$ proportional to ϕ'^2 is given by the sum of all connected diagrams with 2 external amputated ϕ' lines:



$$(5.16)$$

Because all internal lines are ϕ'' lines, the momentum in all internal lines is larger than Λ'. This means that we can expand the diagrams in powers of the mass and external momentum without encountering any non-analyticity. (Contrast this with the full diagrams: there the presence of subdivergences means that even the divergent part of individual diagrams is non-analytic in the external momenta.) This means that we can expand the action $S'[\phi]$ as a sum of local terms (polynomials in momenta) if we are willing to keep only a finite number of terms in the $1/\Lambda'$ expansion. We see that integrating out the modes above Λ' defines a *local* action with the same low-momentum physics as $S[\phi]$. This defines the **exact renormalization group**.[7] Note that the action $S'[\phi]$ contains *all* couplings allowed by the symmetries, even if we start with an action $S[\phi]$ that contains only a finite number of terms.

We can now consider taking the infinitesmal limit $\Lambda' \to \Lambda - d\Lambda$ and writing a differential version of the exact renormalization group. Let us write the result for the restricted set of couplings

$$\Delta S[\phi] = \int d^4x \left[\frac{\lambda_4}{4!}\phi^4 + \frac{\lambda_6}{6!}\phi^6 + \cdots \right]. \tag{5.17}$$

[7] In the literature this is often called the 'Wilson renormalization group.'

In the differential limit, the exact renormalization group becomes an infinite coupled set of differential equations for all of the couplings in the theory:

$$\begin{aligned} \Lambda \frac{d\lambda_4}{d\Lambda} &= \beta_4(\lambda_4, \Lambda^2\lambda_6, \ldots), \\ \Lambda \frac{d\lambda_6}{d\Lambda} &= \frac{1}{\Lambda^2}\beta_6(\lambda_4, \Lambda^2\lambda_6, \ldots), \\ &\vdots \end{aligned} \tag{5.18}$$

where we have inserted powers of Λ so that the functions β_4, β_6, … are dimensionless. In terms of the dimensionless couplings

$$\hat{\lambda}_4 = \lambda_4, \quad \hat{\lambda}_6 = \Lambda^2\lambda_6, \quad \ldots, \tag{5.19}$$

we have

$$\begin{aligned} \Lambda \frac{d\hat{\lambda}_4}{d\Lambda} &= \beta_4(\hat{\lambda}_4, \hat{\lambda}_6, \ldots), \\ \Lambda \frac{d\hat{\lambda}_6}{d\Lambda} &= 2\hat{\lambda}_6 + \beta_6(\hat{\lambda}_4, \hat{\lambda}_6, \ldots), \\ &\vdots \end{aligned} \tag{5.20}$$

The inhomogeneous $2\hat{\lambda}_6$ term simply reflects the fact that λ_6 has mass dimension -2. We will assume for simplicity that for the initial cutoff Λ, all the dimensionless couplings $\hat{\lambda}_4$, $\hat{\lambda}_6$, … are order 1.

In perturbation theory, the beta functions have the form

$$\begin{aligned} \beta_4 &\sim \frac{\hbar}{16\pi^2}\left[\hat{\lambda}_4^2 + \hat{\lambda}_6^2\right] + \mathcal{O}(\hbar^2) \\ \beta_6 &\sim \frac{\hbar}{16\pi^2}\left[\hat{\lambda}_4^3 + \hat{\lambda}_4\hat{\lambda}_6 + \hat{\lambda}_8\right] + \mathcal{O}(\hbar^2) \\ &\vdots \end{aligned} \tag{5.21}$$

For small values of the couplings, we see that the inhomogeneous term in $\hat{\lambda}_6$ dominates, and for $\Lambda' \ll \Lambda$ we have

$$\hat{\lambda}_6(\Lambda') \simeq \left(\frac{\Lambda'}{\Lambda}\right)^2 \hat{\lambda}_6(\Lambda). \tag{5.22}$$

This is just a restatement of the fact that λ_6 has dimension -2, and therefore the effects of λ_6 are suppressed at low energies. On the other hand, dimensionless couplings such as λ_4 have only a quantum contribution to their scaling that is much smaller if the couplings are order 1.

As we lower the cutoff from Λ, before higher-dimension couplings such as λ_6 die away, they will influence the running of dimensionless couplings like λ_4. This effect is as large as the effects from the renormalizable couplings. We see that λ_6 has large effects, but these are precisely of the type that can be absorbed into a redefinition of the renormalizable couplings. This is exactly what we concluded earlier when we considered the renormalization of 'non-renormalizable' theories.

5.2 Scalars and Naturalness

Up until now, we have suppressed the dependence on the mass-squared term. As we now explain, the properties of the mass term in scalar field theories makes these theories technically unnatural.

The exact renormalization group equation for the scalar mass-squared term has the form

$$\Lambda \frac{dm^2}{d\Lambda} = \Lambda^2 \beta_m(m^2/\Lambda^2, \lambda_4, \Lambda^2\lambda_6, \ldots), \tag{5.23}$$

where in perturbation theory

$$\beta_m \sim \frac{\hbar}{16\pi^2}\left[\lambda_4\Lambda^2 + \lambda_4 m^2\right] + \mathcal{O}(\hbar^2). \tag{5.24}$$

For general initial conditions, the 1-loop term proportional to Λ^2 means that m^2 decreases rapidly as we lower Λ.[8] For initial conditions $m^2 \sim \Lambda^2$, the mass-squared term for lower values of the cutoff will be of order $+\Lambda^2$ (since $\beta_m \ll m^2$ in this case). If we take $\beta_5 \sim \lambda_4/(16\pi^2)$, then the beta function becomes important, and the value of the mass at lower values of the cutoff will be of order $\pm\lambda_4\Lambda^2/(16\pi^2)$. However, this is only $\sim 10^{-2}\Lambda$, and the mass of the scalar particle is not much smaller than the cutoff Λ.

Note that there is a critical value of the initial mass so that the value of the mass for $\Lambda \to 0$ vanishes. We can make the scalar mass small compared to the cutoff Λ only by carefully choosing the initial value of the scalar mass close to this critical value.

[8] We assume that the coefficient of the Λ^2 term in the mass-squared beta function is positive, as is the case in ϕ^4 theory in $3+1$ dimensions. (See Eq. (3.36).)

The accuracy with which the initial value of the scalar coupling must be chosen is of order

$$\frac{m_{\text{phys}}^2}{\lambda_4 \Lambda^2/(16\pi^2)}. \tag{5.25}$$

Our point of view is that the initial value of the mass-squared term is determined by a more fundamental theory that describes physics above the initial cutoff Λ. Since there is nothing special about the critical value of the mass-squared (other than the fact that it gives a small scalar mass) it would appear to be miraculous if the fundamental theory gave precisely this value. We therefore say that theories with scalar particles are **fine-tuned** or **unnatural**.

These considerations are very important for particle physics. There is now overwhelming experimental evidence that the weak interactions are described by a spontaneously broken gauge theory, but there is currently no direct evidence about then nature of the dynamics that breaks the electroweak gauge theory.[9] The simplest possibility is that the symmetry breaking is due to the dynamics of a scalar field, called the **Higgs boson**. In order to describe the observed masses of the W and Z bosons (massive analogs of the photon), the mass-squared term of the Higgs boson must be of order (100 GeV)^2. The considerations above suggest that if this theory is right, then the absence of fine-tuning tells us that the cutoff of the theory should not be much greater than $\Lambda \sim 4\pi \times 100$ GeV ~ 1 TeV. This is within reach of upcoming accelerator experiments.

We therefore have an interesting 'no lose' proposition. Either a theory based on a scalar Higgs is correct, and there is new physics at the TeV scale that renders the existence of the Higgs natural; or something more exotic than a scalar Higgs is responsible for electroweak symmetry breaking. Speculative models that embody both types of proposals exist in the literature. The fact that the mechanism of symmetry breaking must be discovered in the energy range 100 GeV to 1 TeV is the main driving force behind both theoretical and experimental reseach in elementary particle theory.

[9]Gauge symmetry and spontaneous breaking will hopefully be studied later in your career.