5

Renormalization

In this chapter we come to the general theory of renormalization. The basic difficulty is that a graph may not only possess an overall divergence. It may have in addition many subdivergences which can be nested or can overlap in very complicated ways. Most of our effort must go to disentangling these complications.

We will begin by investigating some simple graphs. These will show us how to set up the formalism in the general case. The ultimate result is the forest formula of Zimmermann (1969). Contrary to its reputation, this is not an esoteric procedure, designed for pedantically rigorous treatments. Rather, the forest formula is merely a general way of writing down what is in fact the natural and obvious way of extracting the divergences from any integral. Its power is demonstrated by the ease of treating overlapping divergences, the handling of which is normally considered the *bête noire* of renormalization theory.

The forest formula is applied to individual Feynman graphs. It extracts the finite part of a graph by subtracting its overall divergence and its subdivergences. We will, of course, need to show that the subtractions can be implemented as actual counterterms in the Lagrangian. We will also show that the counterterms are local, i.e., polynomial in momentum.

An important advantage of using the forest formula to obtain the finite part of each graph, rather than working directly with counterterms in the Lagrangian, is that the procedure applies to more general situations. As we will see in Chapter 6, it will enable us to renormalize composite operators. A more important case is the computation of asymptotic behavior as external momenta of a Green's function get large. For this case, the forest formula permits a good derivation of Wilson's operator-product expansion, which we will discuss in Chapter 10.

Let the value of a Feynman graph be written as:

$$U(G)(p_1,\ldots,p_N) = \int \mathrm{d}^d k_1 \ldots \mathrm{d}^d k_L \, I(p_1,\ldots,p_N;k_1,\ldots k_L). \qquad (5.0.1)$$

Here $p_1,\ldots,p_N$ are the external momenta, and $k_1,\ldots,k_L$ are the loop

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

momenta. Renormalization is removal of that part of the large-k behavior that causes divergences. Moreover, the very same techniques can be used to extract the behavior for large p – as we will see when we treat the operator product expansion in Chapter 10.

Although Weinberg's (1960) theorem tells us the power-counting applicable to either kind of asymptotic behavior, it does not tell us how to organize it. In particular it was only much later that Wilson (1969) formulated his operator product expansion, which is the important tool in computing asymptotic behavior, for example in deep-inelastic scattering – see Chapter 14. Many generalizations have been made – see Mueller (1981) for a review. These are phenomenologically very important, and the method by which they are proved is close to that for Wilson's expansion.

5.1 Divergences and subdivergences

The idea of renormalization theory is that ultra-violet divergences of a field theory are to be cancelled by renormalizations of the parameters of the theory. We propose to prove this in perturbation theory. The use of perturbation theory implies that we expand the counterterms in the action in powers of the renormalized coupling, g, thereby generating extra graphs with these counterterms as some of the vertices. To avoid superfluous technicalities, we will consider the case of ϕ^3 theory in six-dimensional space-time.

A very efficient way to understand renormalization was discovered by Bogoliubov & Shirkov (1955, 1956, 1980) and Bogoliubov & Parasiuk (1957), and we shall follow their approach. The first step is to decompose the Lagrangian as follows:

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_b + \mathscr{L}_{ct}. \tag{5.1.1}$$

Here $\mathscr{L}_0$ is the free Lagrangian used to generate the free propagator $i/(p^2 - m^2 + i\varepsilon)$ in perturbation theory:

$$\mathscr{L}_0 = (\partial\phi)^2/2 - m^2\phi^2/2, \tag{5.1.2}$$

with m being the renormalized mass. The rest of the Lagrangian, $\mathscr{L}_I = \mathscr{L}_b + \mathscr{L}_{ct}$, is the interaction, and consists of two terms. The first, which we will call the basic interaction, is

$$\mathscr{L}_b = -g\phi^3/3!, \tag{5.1.3}$$

where g is the renormalized coupling. The second term, $\mathscr{L}_{ct}$, we will call the counterterm Lagrangian.

Consider graphs generated by the basic interaction. These have UV

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

divergences which are to be cancelled by graphs with some of their interaction vertices taken from the counterterm Lagrangian

$$\mathscr{L}_{\text{ct}} = \delta Z(\partial\phi)^2/2 - \delta m^2\phi^2/2 - \delta g\phi^3/3! - \delta h\phi. \tag{5.1.4}$$

(The term linear in ϕ is needed to cancel tadpole graphs – see Figs. 5.1.4 and 5.1.5 below.) In order to give meaning to δZ, δm^2, δg, and δh, we must impose an ultra-violet cut-off. We will use dimensional regularization in the following sections.

The key to the method that we use is to realize that each of the three terms in the counterterm Lagrangian should not be considered as a single quantity. Rather it is to be considered as a sum of terms, each of them cancelling the overall divergence in one particular graph generated by the basic interaction. For example, the self-energy graph, Fig. 3.1.1, gives a contribution $\delta_1 Z$ to δZ, and a term $\delta_1 m^2$ to δm^2. Our calculation of this graph led to (3.5.7), so with minimal subtraction we have

$$\left.\begin{aligned} \delta_1 Z &= g^2/[384\pi^3(d-6)], \\ \delta_1 m^2 &= g^2 m^2/[64\pi^3(d-6)]. \end{aligned}\right\} \tag{5.1.5}$$

Then

$$\delta Z = \sum_{\text{graphs } G} \delta_G Z = \delta_1 Z + \cdots,$$

with similar formulae for the other counterterms.

We saw the utility of this idea by examining graphs like those in Fig. 3.2.1 and Fig. 3.2.2. Graphs like Fig. 3.2.2 contain vertices corresponding to the counterterm $\delta_1 m^2$ (and $\delta_1 Z$). Such graphs are all generated by taking graphs like Fig. 3.2.1 with no counterterms and finding where Fig. 3.1.1 occurs as a subgraph. Substitution of the counterterm for one or more of these subgraphs gives the graphs with counterterm vertices.

This leads to the idea that we consider by itself the renormalization of a single graph generated from the basic Lagrangian. We add to it a set of counterterm graphs to give a finite result. Only as a separate step do we recognize that the counterterm vertices are, in fact, generated from a piece of an interaction Lagrangian.

The graph-by-graph method is probably the most powerful approach to understanding not only the problem of ultra-violet divergences but also many other problems in asymptotic behavior. Even so, it is not at all trivial to ensure that the renormalization program can be carried out. The essential steps are:

(1) To find the regions in the space of loop momenta of a graph that give ultra-violet divergences.

(2) To show how to generate a series of counterterm graphs for a given basic graph.
(3) To show that the counterterm vertices are local (i.e., polynomial in momenta).
(4) To find the conditions under which the counterterm vertices amount only to renormalizations of the parameters of the Lagrangian.

The complications in carrying out this program arise when one treats the case of the divergence of a graph which has a divergent subgraph. To understand why there is a difficulty, we will examine the graphs of order g^4 for the full propagator – Figs. 5.1.1–5.1.3.

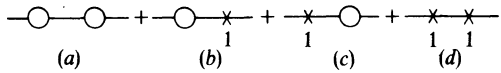

Fig. 5.1.1. A two-loop graph for the propagator in ϕ^3 theory, together with its counterterm graphs.

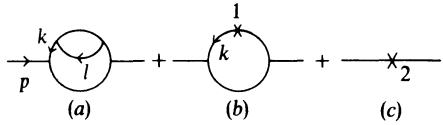

Fig. 5.1.2. A two-loop graph for the propagator in ϕ^3 theory, together with its counterterm graphs.

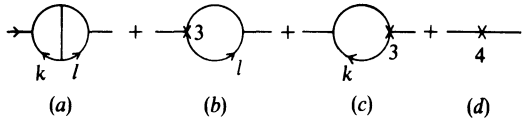

Fig. 5.1.3. A two-loop graph for the propagator in ϕ^3 theory, together with its counterterm graphs.

We ignore the graphs with tadpoles, such as Fig. 5.1.4. These are divergent and need a counterterm $\delta h\phi$. We can use a renormalization condition that $\langle 0|\phi|0\rangle$ vanishes. Then the total of the tadpole graphs is zero (e.g., Fig. 5.1.5), so we omit any graphs containing them.

Let us return to the sets of graphs listed in Figs. 5.1.1–5.1.3. In each set there is one basic graph and a set of counterterm graphs. Ultra-violet

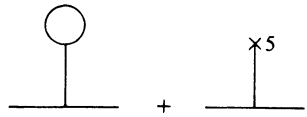

Fig. 5.1.4. A tadpole graph, together with its counterterm.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

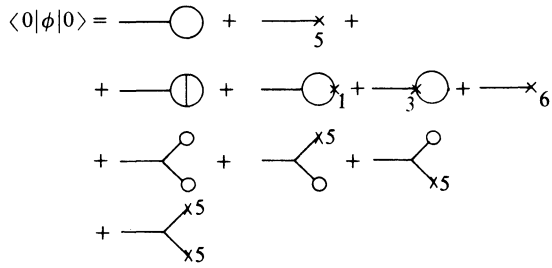

Fig. 5.1.5. Graphs to $O(g^3)$ for $\langle 0|\phi|0\rangle$.

divergences involve a loop momentum that gets large, so the divergences are always confined to one-particle-irreducible subgraphs. The simplest case is Fig. 5.1.1, where the basic graph has two insertions of the one-loop self-energy. It is made finite by adding the graphs with one or both of the self-energy subgraphs replaced by a counterterm. (We use a cross to denote a counterterm in a graph, and we use the label '1' for the counterterm of the one-loop self-energy.)

In Fig. 5.1.2, the basic graph is more complicated. We will treat it in detail in Section 5.2, and we merely summarize the results here. It has two UV divergences. The first comes from letting both loop momenta k and l go to infinity; we call this the overall divergence. But there is also a divergence where the momentum in the outer loop stays finite. This is an example of a subdivergence. Its existence, as we will see, implies that there is a term proportional to $p^2 \ln(p^2)$ in the divergence of the basic self-energy graph. This cannot be cancelled by any local counterterm. However there is also a graph with a counterterm to the subgraph. This graph is Fig. 5.1.2(*b*). After we add the two graphs, the non-local divergence cancels. The overall divergence in the result is then cancelled by a local counterterm, for which we use the label '2'. We implement this as a counterterm in the action by inserting terms $\delta_2 Z$ and $\delta_2 m^2$ into the complete counterterms δZ and δm^2.

The pattern is simple. We consider as a single finite entity one basic graph together with counterterms for its subdivergences and for the overall divergence.

Another case is shown in Fig. 5.1.3. There are two subdivergences, each corresponding to a vertex subgraph. The one-loop vertex is logarithmically divergent and is made finite by renormalizing the coupling (Fig. 3.6.1). Since the two divergent subgraphs overlap, the counterterm graphs are generated by replacing one but not both of the vertex graphs by its counterterm. The overall divergence is then local and is cancelled by counterterms $\delta_4 Z$ and $\delta_4 m^2$.

Our first task in Section 5.2 will be to verify the above statements. To generalize the argument we will then observe that power-counting as in Section 3.3 determines the strength of the overall divergence. To prove that the presence of subdivergences does not affect the form of the overall counterterm, we will differentiate with respect to external momenta to remove the overall divergence. Then we will be able to construct an inductive proof that if subdivergences have been cancelled by counterterms then the overall divergence is local and its strength is determined by simple power-counting.

We will also show how to disentangle the combinatoric problems when subdivergences are nested. Finally, we will discuss Weinberg's theorem. This theorem tells us exactly which regions of momentum we must consider. In practice one is very simple-minded about locating UV divergences. For example, we stated that the regions giving divergences for Fig. 5.1.2 are: (*a*) k and l going to infinity together, and (*b*) l going to infinity, with k fixed. In each region, all the momenta get large in a particular 1PI subgraph that is divergent by power-counting. Weinberg's theorem tells us that these are the only regions we have to consider explicitly. In the case of Fig. 5.1.2 there is another region that is important, where l goes to infinity with k also going to infinity, but much more slowly. This region interpolates between the other two, but in fact does not need to be treated as a separate case.

5.2 Two-loop self-energy in $(\phi^3)_6$

In this section we will explain the properties of overall divergences and subdivergences by computing the two-loop self-energy graphs, Figs 5.1.2 and 5.1.3, in ϕ^3 theory at space-time dimension $d = 6$. We will again use dimensional regularization, and will need the values of the one-loop counterterms in order to cancel subdivergences.

The one-loop self-energy was considered in Section 3.6.2, where we found that the counterterms needed were given by (5.1.5). We can also compute the one-loop vertex graph, Fig. 3.6.1, with the resulting counterterm being (cf., (3.6.13))

$$\delta_3 g = \mu^{3-d/2} g^3 / [64\pi^3 (d-6)]. \tag{5.2.1}$$

It is worth noting that this implies a value for the one-loop term in the bare coupling:

$$\begin{aligned} g_0 &= [\mu^{3-d/2} g + \delta_3 g + O(g^5)] Z^{-3/2} \\ &= g\mu^{3-d/2} \{1 + \tfrac{3}{4} g^2 / [64\pi^3 (d-6)] + O(g^4)\}. \end{aligned} \tag{5.2.2}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

5.2.1 *Fig.* 5.1.2

In order to be able to compute Fig. 5.1.2 in closed form we work with the massless theory. The value of the graph is then

$$\Sigma_{2a} = \frac{-g^4\mu^{12-2d}}{(2\pi)^{2d}} \times$$

$$\times \tfrac{1}{2}\int \mathrm{d}^d k \mathrm{d}^d l \frac{1}{[(p+k)^2+\mathrm{i}\varepsilon](k^2+\mathrm{i}\varepsilon)^2(l^2+\mathrm{i}\varepsilon)[(k-l)^2+\mathrm{i}\varepsilon]}. \quad (5.2.3)$$

The inner loop is easily computed in terms of Γ-functions:

$$\int \mathrm{d}^d l \frac{1}{l^2(k-l)^2} = \mathrm{i}\pi^{d/2}\Gamma(2-d/2)\int_0^1 \mathrm{d}x[-k^2x(1-x)]^{d/2-2}$$

$$= \mathrm{i}\pi^{d/2}\Gamma(2-d/2)\frac{\Gamma(d/2-1)^2}{\Gamma(d-2)}(-k^2)^{d/2-2}. \quad (5.2.4)$$

We now have

$$\Sigma_{2a} = \frac{\mathrm{i}g^4}{2^{13}\pi^9}(16\pi^3\mu^4)^{3-d/2}\Gamma(2-d/2)\frac{\Gamma(d/2-1)^2}{\Gamma(d-2)} \times$$

$$\times \int \mathrm{d}^d k \frac{1}{(-k^2)^{4-d/2}[-(p+k)^2]}. \quad (5.2.5)$$

The denominators can be combined by a Feynman parameter:

$$\frac{1}{A^{4-d/2}B} = \frac{\Gamma(5-d/2)}{\Gamma(4-d/2)}\int_0^1 \mathrm{d}x \frac{x^{3-d/2}}{[Ax+B(1-x)]^{5-d/2}}. \quad (5.2.6)$$

after which the k-integral can be performed. The result is

$$\Sigma_{2a} = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{2}\left(\frac{-p^2}{4\pi\mu^2}\right)^{d-6} \times$$

$$\times \frac{\Gamma(2-d/2)\Gamma(5-d)\Gamma(d/2-1)^3\Gamma(d-4)}{\Gamma(d-2)\Gamma(4-d/2)\Gamma(3d/2-5)}$$

$$\equiv \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{2}\left(\frac{-p^2}{4\pi\mu^2}\right)^{d-6} \Gamma(2-d/2)\Gamma(5-d)K(d). \quad (5.2.7)$$

The overall ultra-violet divergence is contained in the factor $\Gamma(5-d)$. Observe that the argument of this Γ-function is exactly minus half times the degree of divergence. The subdivergence is contained in the factor $\Gamma(2-d/2)$; this is the same as we calculated in Chapter 3.

Before we discuss further the UV divergences we should observe that there are also infra-red divergences. These come from the existence of long-range forces in a theory with massless fields. In momentum space, they

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

appear as divergences when some momenta go to zero. For example, if $d \leq 2$ the integral over the momentum through any propagator has a divergence at zero momentum:

$$\int_{q \sim 0} d^d q/q^2 \simeq \text{constant}/(d-2) \text{ at } d \text{ close to } 2.$$

This accounts for the factor $\Gamma(d/2-1)^3$. When $d \leq 4$ there is also a divergence at $k=0$ with l and p fixed. Our only concern is with UV problems, so we ignore the IR divergence. If we used a massive field, the IR divergences would go away, but we would not have an explicit formula for Σ_{2a}.

Now let us expand Σ_{2a} in powers of $d-6$ to exhibit its divergences, and its dependence on p^2:

$$\Sigma_{2a} = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\frac{1}{(d-6)^2} + \frac{1}{d-6}\left[\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\right] + \right.$$
$$\left. + \tfrac{1}{2}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant} + O(d-6)\right\}. \tag{5.2.8}$$

The double pole at $d=6$ and the double logarithm in the finite part are both reflections of the fact of having a subdivergence. The p-dependence is a power of p^2 times a polynomial in $\ln(-p^2)$. This is a characteristic feature of massless theories.

The simple pole has a coefficient that is not polynomial in p. Consequently, it cannot be cancelled by any local counterterm. It is easy to see that this is caused by the presence of the subdivergence. The subdivergence comes from the region where the loop momentum of the inner loop goes to infinity while the momentum k in the outer loop remains finite. Integrating over finite k gives a logarithm of p times the divergent part of the inner loop. We have already introduced into the Lagrangian a counterterm for the inner loop, so that there is a graph, Fig. 5.1.2(b), in which this counterterm appears in such a way as to cancel the subdivergence.

Therefore the sum of Figs. 5.1.2(a) and (b) should have no subdivergence, but only an overall divergence. This can be cancelled by a local counterterm (i.e., a polynomial in p). We will prove this in Section 5.2.2 by differentiating three times with respect to the external momentum p^μ; this gives a result which has negative degree of divergence, i.e., there is no overall divergence. Since the subdivergence is cancelled, there is no subdivergence whatever, so the counterterm must be quadratic in p. We represent this by Fig. 5.1.2(c).

In Section 5.2.2 we will make explicit this proof of locality of the counterterms.

Here we will verify the above statements by explicit calculations. In our case that $m = 0$, the value of Fig. 5.1.2(b) is

$$\begin{aligned}\Sigma_{2b} &= -i\delta_1 Z \frac{g^2\mu^{6-d}}{(2\pi)^d}\int d^d k \frac{k^2}{(k^2)^2(p+k)^2} \\ &= \left(\frac{g^2}{64\pi^3}\right)^2 \left(\frac{-p^2}{6}\right)\frac{\Gamma(2-d/2)\Gamma(d/2-1)^2}{(d-6)\Gamma(d-2)}\left(\frac{-p^2}{4\pi\mu^2}\right)^{d/2-3} \\ &= \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\frac{-2}{(d-6)^2} + \frac{1}{d-6}\left[-\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\right]\right. \\ &\quad \left. -\tfrac{1}{4}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant} + O(d-6)\right\}.\end{aligned} \tag{5.2.9}$$

The non-local divergence disappears when we add this graph to Σ_{2a}, with the result

$$\begin{aligned}\Sigma_{2a} + \Sigma_{2b} = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}&\left\{\frac{-1}{(d-6)^2} + \frac{\text{constant}}{(d-6)}\right. \\ &\left. +\tfrac{1}{4}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant} + O(d-6)\right\}.\end{aligned} \tag{5.2.10}$$

The non-local divergence has cancelled, as promised. However, the double pole and, in the finite part, the double logarithm have not cancelled, even though it is evident from our calculation that they are associated with the subdivergence nested inside the overall divergence. This is a general phenomenon. Indeed we will see in Chapter 7, where we discuss the renormalization group, that the coefficients of the double pole and of the double logarithm could have been predicted from the one-loop counterterms without any explicit two-loop calculations.

Since the non-local divergences have now cancelled, the overall divergence can be cancelled by choosing a wave-function counterterm

$$\delta_2 Z = \left(\frac{g^2}{64\pi^3}\right)^2 \frac{1}{36}\left\{\frac{-1}{(d-6)^2} + \frac{\text{constant}}{(d-6)}\right\}. \tag{5.2.11}$$

Then we obtain at $d = 6$ a finite result, which we term the renormalized

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

value of Fig. 5.1.2:

$$\Sigma_{2R}^{(MS)} = \Sigma_{2a} + \Sigma_{2b} + (\Sigma_{2c} = -p^2\delta_2 Z)$$
$$= \left(\frac{g^2}{64\pi^3}\right)^2 \frac{p^2}{36}\left\{\tfrac{1}{4}\ln^2\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant } \ln\left(\frac{-p^2}{4\pi\mu^2}\right) + \text{constant}\right\}. \tag{5.2.12}$$

5.2.2 *Differentiation with respect to external momenta*

We saw that the graph Fig. 5.1.2 has an overall divergence which is local, but that it is local only after we have subtracted the subdivergence. In general we will need to show that the counterterm of a 1PI graph is a polynomial in its external momenta with degree equal to the degree of divergence. Our argument (following Caswell & Kennedy (1982)) depends on differentiating with respect to external momenta.

In this subsection we will apply the argument to Fig. 5.1.2, emphasizing its generality. Then in the next subsection we will apply it to Fig. 5.1.3. Even though that graph has an overlapping divergence, traditionally considered a hard problem, we will see that our method works as easily for this graph as for Fig. 5.1.2.

We first differentiate Fig. 5.1.2(*a*) three times with respect to p^μ, to make its degree of divergence negative. We represent the result pictorially by Fig. 5.2.1, where each dot indicates one differentiation with respect to p. Similarly, differentiating Fig. 5.1.2(*b*) three times gives Fig. 5.2.2. Now Fig. 5.2.2 cancels the subdivergence in Fig. 5.2.1, and there is no overall divergence, so their sum is finite. Thus the third derivative of the sum of Figs. 5.1.2(*a*) and (*b*) is finite. So the overall counterterm is quadratic in p. Lorentz invariance forces it to be of the form $A(d)p^2 + B(d)$.

We glibly asserted that Fig. 5.2.1 plus Fig. 5.2.2 is finite. This statement is not as obvious as it seems. Let us prove it. We Wick-rotate the integrations over k and l in Fig. 5.2.1, and consider regions of the integral that might give a UV divergence. If k and l go to infinity at the same rate, then there is no

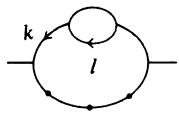

Fig. 5.2.1. Result of differentiating Fig. 5.1.2(*a*) three times with respect to its external momentum.

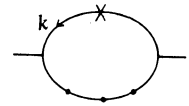

Fig. 5.2.2. Result of differentiating Fig. 5.1.2(*b*) three times with respect to its external momentum.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

divergence, because the degree of divergence is negative. If l goes to infinity with k fixed, there is a divergence, but it is cancelled by the counterterm graph, Fig. 5.2.2.

The remaining significant possibility is that both k and l go to infinity, but that k is much less than l. The ratios of different components of either one of k or l are finite, so we may summarize the order of magnitude of the contribution from this region as:

$$\text{finite} \int^{\infty} dk\, k^{-4} \int_{l \gg k} dl\, l. \tag{5.2.13}$$

This gives a divergent contribution, if l is of order $k^{3/2}$. We must add Fig. 5.2.2, which, as we will show, cancels this new divergence. Observe that the counterterm was arranged to cancel the divergence when l goes to infinity with k fixed, rather than when k is large, as we now have.

Let us expand the inner loop of Fig. 5.2.1 in powers of k when $l \gg k$, up to its degree of divergence, which is quadratic. The coefficients of these powers are integrals over all l, restricted to $l \gg k$. The divergences in the coefficients are cancelled by Fig. 5.2.2, and we have the following estimates of the sizes of the coefficients:

$$\begin{aligned}
\text{coefficient of } k^0 &= \text{finite}\left\{\int_k^{\infty} dl\, l - \text{divergence}\right\} \\
&= \text{finite}\left\{\int_1^{\infty} dl\, l - \text{divergence}\right\} + \text{finite}\int_1^{k} dl\, l \\
&= O(k^2), \\
\text{coefficient of } k^1 &= O(k), \\
\text{coefficient of } k^2 &= \text{finite}\left\{\int_k^{\infty} dl/l - \text{divergence}\right\} \\
&= O(\ln(k)).
\end{aligned} \tag{5.2.14}$$

The sum of Figs. 5.2.1 and 5.2.2 in the region $k \to \infty$, with l possibly much bigger than k, is then of order

$$\int^{\infty} dk\, k^{-2} \ln(k). \tag{5.2.15}$$

The power of k is the same as is given by the overall degree of divergence, but there is an extra logarithm. We get a finite result, as claimed. The higher-than-quadratic terms in the expansion of the loop in powers of k give no divergence at all.

What has happened? The divergence for $l \gg k \gg 1$ could only occur

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

because the interior loop was itself divergent. The fact that we sent k to infinity merely suppressed this divergence somewhat. Suppose we neglect k in the integral for the inside loop. Then the counterterm is in effect the negative of the integral over l of the loop from finite l to infinity. But in the region we are considering for Fig. 5.2.1, we are restricting l to be much bigger than k, which is itself getting large. So if we neglect k in the loop, then we are left with

$$\text{finite} + \int_{l<k} \mathrm{d}l \text{ (integrand of loop with } k \text{ neglected).} \qquad (5.2.16)$$

Furthermore, we expand the loop in powers of k, to uncover the sub-leading divergences. Each extra explicit power of k in the expansion compensates for the lowering of the divergence. The quadratic term multiplies $\int \mathrm{d}l/l$, so giving an extra logarithm (but *not* a power).

The key step in the proof is to perform the integral with the larger momentum l first. We have shown that, for the purpose of determining whether or not a divergence occurs, we need only consider as distinct regions: (1) $k, l \to \infty$ at the same rate, and (2) $l \to \infty$ with k finite. (We might also try $k \to \infty$ with l finite, but the subgraph with the lines carrying the loop momentum k has negative degree of divergence, so we get no divergence from that region.) The region $k, l \to \infty$, with $k \ll l$, is schizoid: it can be considered as essentially part of either of the two regions (1) and (2) that we have just defined. As region (2), the divergence is cancelled by a counterterm when $l \to \infty$, with k large but fixed. As region (1), the final integral over k is finite, and the only sign of this intermediate case is the extra logarithm in the integrand.

5.2.3 *Fig.* 5.1.3

We conclude this section by considering the example of Fig. 5.1.3. At $d = 6$ the graph (*a*) has an overall quadratic divergence. It also has a logarithmic subdivergence when either of the loop momenta k or l gets large. The subdivergences are cancelled by vertex corrections, which are shown in graphs (*b*) and (*c*). We must prove that the overall counterterm (*d*) is quadratic in p.

Conventionally, this graph is regarded as a difficulty in the theory of renormalization, for it contains an overlapping divergence. That is to say, one of the lines is common to both subdivergences. This is seen as a problem (Bjorken & Drell (1966)) if one tries to write the graph as an insertion of a renormalized vertex, Fig. 5.2.3, in the one-loop self-energy. The graph (*b*)

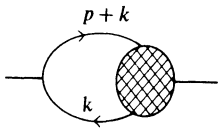

Fig. 5.2.3. Illustrating the problem of overlapping divergences, as seen in Fig. 5.1.3.

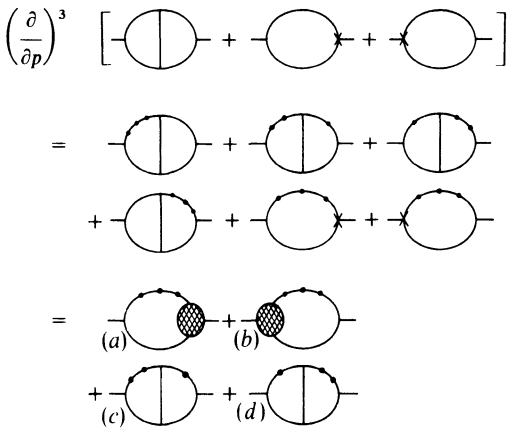

Fig. 5.2.4. Result of differentiating Fig. 5.1.3 three times with respect to its external momentum.

for the counterterm to one of the subdivergences is not of this form. The corresponding difficulty does not happen in our first example, Fig. 5.1.2.

However, our trick of differentiating three times with respect to p works as well for Fig. 5.1.3 as it did for Fig. 5.1.2. For the sum of (*a*), (*b*), and (*c*), we find Fig. 5.2.4. The point is that differentiating either of the subgraphs makes it convergent, while the counterterms for the subdivergences are independent of momenta. We get terms (*a*) and (*b*), which have renormalized subgraphs, and graphs (*c*) and (*d*), which have no subdivergences at all. None of the graphs has an overall divergence. The calculation of the overall counterterms is left as an exercise for the reader. The correct result is (Macfarlane & Woo (1974)):

$$\delta_4 Z = \left(\frac{g^2}{64\pi^3}\right)^2 \left[\frac{1}{6}\frac{1}{(d-6)^2} + \frac{1}{3}\frac{1}{(d-6)}\right],$$

$$\delta_4 m^2 = \left(\frac{g^2}{64\pi^3}\right)^2 m^2 \left[-\frac{1}{(d-6)^2} - \frac{1}{2}\frac{1}{(d-6)}\right], \tag{5.2.17}$$

if we use minimal subtraction.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

5.3 Renormalization of Feynman graphs

We have seen that, in order to construct a sensible (i.e., local) counterterm for a (1PI) Feynman graph, one must first subtract off its subdivergences. This is natural since the subtractions for subdivergences are automatically generated from having the counterterms be definite pieces of the interaction Lagrangian. Without subtraction of the subdivergences, the divergence of a graph need not be local. It may even have a power of momentum greater than the degree of divergence; an obvious case of this is a graph that is finite according to naive power-counting but that has a subdivergence.

It is therefore useful to devise a procedure for starting with a basic Feynman graph G, constructing a set of counterterm graphs, and thereby obtaining a finite renormalized value $R(G)$:

$$R(G) = U(G) + S(G). \tag{5.3.1}$$

Here $U(G)$ is the 'unrenormalized' value of the basic graph (which diverges as the UV cut-off is removed), and $S(G)$ is the subtraction – the sum of the counterterm graphs.

The strategy we use to construct $S(G)$ is very general. It applies to the asymptotic behavior of any integral as one or more parameters approach a limiting value. In field theory it can be applied not only to the renormalization problem but also to the calculation of the asymptotic behavior of a Green's function as some but not all of its external momenta get large. (A standard example which we will treat in Chapter 10 is the operator product expansion of Wilson (1969)).

The procedure that we use for renormalization was first developed by Bogoliubov and Parasiuk (see Bogoliubov & Shirkov (1980)), with corrections due to Hepp (1966). Their construction was recursive and has the acronym BPH. Zimmermann (1969) showed how to solve the recursion – the result being called the forest formula. All these authors used zero-momentum subtractions. Zimmermann (1970, 1973a) showed moreover that there is then no need to use an explicit UV cut-off. He applied the algorithm for computing $R(G)$ directly to the integrand rather than to the integral; this formulation has the title BPHZ. It is not necessary to use zero-momentum subtractions. For example Speer (1974), Collins (1975b), Breitenlohner & Maison (1977a, b, c) showed how to make the same ideas work using minimal subtraction.

Our treatment will aim at showing the underlying simplicity of the methods and their power to demystify renormalization theory. We will see that the methods do not depend on use of a particular renormalization prescription, even though we will often use minimal subtraction.

We will examine the structure of the subtractions for a graph G. (A graph we define by specifying its set of vertices and lines, each line joining two vertices and each vertex attached to at least one line.) We write the graph's unrenormalized value as

$$U_G(p_1,\ldots,p_N) = \int \mathrm{d}^d k_1 \ldots \mathrm{d}^d k_L I_G(p_1,\ldots,p_N;k_1,\ldots,k_L). \tag{5.3.2}$$

Here we let L be the number of loops and N be the number of vertices. The external momenta at the vertices are p_i. In a Feynman graph for a Green's function there is an external momentum at the vertices for the external fields, but at an interaction vertex, we have $p_i = 0$.

5.3.1 *One-particle-irreducible graph with no subdivergences*

The simplest case is a one-particle-irreducible (1PI) graph with no subdivergences. Then the only possible divergence is an overall divergence where the momenta on all the lines get large simultaneously. We may renormalize the graph by subtracting an overall counterterm:

$$R(G) = U(G) - T \circ U(G). \tag{5.3.3}$$

Here T denotes some operation that extracts the divergence of $U(G)$. It implements whatever renormalization prescription that we choose to use. For example, we might use minimal subtraction. In that case T takes the Laurent expansion of $U(G)$ about $d = d_0$, and picks out the pole terms. (We let the physical space-time dimension be d_0; i.e., $d_0 = 4$ for the real world, or $d_0 = 6$ for the ϕ^3 model we used in the previous sections.) We will use either of two notations for the action of T on an unrenormalized object: $T \circ U(G)$ or $T(G)$. Both will mean the same.

We could use zero-momentum subtractions. In that case T picks out the terms up to order $\delta(G)$ in the Taylor expansion of $U(G)$ about zero momentum. Here $\delta(G)$ is, as usual, the degree of divergence. There are many other possibilities. In our work, we will use the minimal subtraction scheme.

Then, for example, the one-loop self energy in $(\phi^3)_6$ gives

$$\begin{aligned} T \circ \left\{ \frac{ig^2}{2} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{\mu^{6-d}}{(k^2 - m^2)[(p+k)^2 - m^2]} \right\} \\ = \text{pole part of } \left\{ \frac{-g^2}{128\pi^3} \Gamma(2 - d/2) \times \right. \\ \left. \times \int_0^1 \mathrm{d}x \left[\frac{m^2 - p^2 x(1-x)}{4\pi\mu^2} \right]^{d/2-3} [m^2 - p^2 x(1-x)] \right\} \\ = \frac{g^2}{64\pi^3} \frac{(p^2/6 - m^2)}{(d-6)}. \end{aligned} \tag{5.3.4}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

The one-loop vertex, Fig. 3.6.1, gives

$$T\left\{g^3\int\frac{d^dk}{(2\pi)^d}\frac{\mu^{9-3d/2}}{(k^2-m^2)[(p+k)^2-m^2][(p+q+k)^2-m^2]}\right\}$$
$$=\text{pole part of}\left\{\frac{-ig^3\mu^{3-d/2}}{64\pi^3}\Gamma(3-d/2)\times\right.$$
$$\left.\times\int_0^1 dx\int_0^{1-x}dy\left[\frac{m^2-q^2xy-(p^2x+(p+q)^2y)(1-x-y)}{4\pi\mu^2}\right]^{d/2-3}\right\}$$
$$=i\frac{g^3\mu^{3-d/2}}{64\pi^3(d-6)}. \tag{5.3.5}$$

Observe that in this last case we define the pole to come with a factor $\mu^{3-d/2}$. This is an example of a general rule that one must define the pole part of $U(G)$ to have the same dimension as $U(G)$, for all d.

5.3.2 *General case*

In general we not only have to handle the case of an overall divergence, but also the case that subdivergences are nested within the overall divergences. Another case is exemplified by the propagator with two self-energy insertions (Fig. 5.1.1), where within one graph there are two subgraphs which can diverge independently.

As we saw from examples, we must subtract off subdivergences before finding the overall divergence. In view of the complications when the subdivergences themselves have subdivergences, etc. (*ad nauseam*), we must be extremely precise as to what is to be done. This is what we will now do. It is helpful to have a specific non-trivial example in mind, to make sense of the mathematics. Such examples are treated in subsection 5.3.3 and in Section 5.4. The reader should try to read these sections concurrently with the general treatment in this section.

First, let us define a specific divergence as being the divergence occurring when the loop momenta on a certain set of lines get big, with the momenta on other lines and the external momenta staying finite. Whether or not a given set of lines has a divergence associated with it is determined by power-counting. A divergence is thus associated with a certain subgraph. (At this stage, we do not require that the subgraph be connected.)

If a graph G diverges when all its lines get large loop momenta, it is said to have an overall divergence. A one-particle-reducible graph (like any of Fig. 3.2.1) cannot have an overall divergence – some lines are not a part of any loop. All other divergences involve a smaller subset of the lines. They, of course, are called subdivergences. Every subdivergence of a graph is the overall divergence of one of its subgraphs.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

Observe that Fig. 5.1.1(*a*) has no overall divergence, but has three subdivergences. These come from the regions in the integration over loop momenta where: (1) the left-hand loop has large momentum, (2) the right-hand loop has large momentum, and (3) both loops have large momenta.

To renormalize a graph G we assume that we know how to renormalize its subdivergences, and we then let $\bar{R}(G)$ be the unrenormalized value of G with subtractions made to cancel the subdivergences. Then the only remaining divergence that is possible is an overall divergence. So we define an overall counterterm:

$$C(G) = -T\circ\bar{R}(G) \tag{5.3.6}$$

by applying to $\bar{R}(G)$ the same subtraction operator T as we discussed earlier; if there is no overall divergence (e.g., if G is one-particle-reducible) then $C(G)$ is zero. In any event the renormalized value of G is defined as

$$R(G) = \bar{R}(G) + C(G). \tag{5.3.7}$$

The definitions (5.3.6) and (5.3.7) give us $R(G)$ provided that we know how to subtract subdivergences. This is essentially a matter of renormalizing smaller graphs; we will construct $\bar{R}(G)$ in a moment. Once we have done this, we will have a recursive definition of $R(G)$: successive application of (5.3.7) to smaller and smaller subgraphs ultimately brings us to graphs with no subdivergences. These we know how to renormalize.

Now let us define $\bar{R}(G)$, which is to be $U(G)$ with subdivergences subtracted. For the case of a graph with no subdivergences we must define

$$\bar{R}(G) = U(G) \quad \text{(if } G \text{ has no subdivergences)}. \tag{5.3.8a}$$

For a larger graph we define

$$\bar{R}(G) = U(G) + \sum_{\gamma \subsetneq G} C_\gamma(G). \tag{5.3.8b}$$

We sum over all subgraphs γ of G, other than G itself, as indicated by the notation $\gamma \subsetneq G$. The other new notation $C_\gamma(G)$ means that we replace the subgraph γ by its overall counterterm, as defined by (5.3.6), i.e.,

$$C(\gamma) = \begin{cases} -T\circ\bar{R}(\gamma), & \text{if } \gamma \text{ has an overall divergence} \\ 0 & \text{if } \gamma \text{ has no overall divergence} \end{cases}. \tag{5.3.9}$$

To make a simple formula, we write the sum as being over all γ's rather than only over divergent γ's; then the $C_\gamma(G)$ is zero if γ is not overall divergent. We could of course restrict the sum only to those subgraphs that have an overall divergence.

One tricky point in the above equations arises in defining $C(\gamma)$ for a disconnected subgraph γ. An example is the subgraph of Fig. 5.1.1(*a*) consisting of the two self-energy loops. We will discuss this next.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

5.3.3 *Application of general formulae*

Equations (5.3.6) to (5.3.9) give a definition of $R(G)$. Let us see how they apply to simple examples. For a 1PI graph with no subdivergences they just reproduce $R(G) = U(G) - T\circ U(G)$.

Next consider a graph like Fig. 3.2.1(*a*) or (*c*), whose only divergence is a subgraph with no further subdivergences. Then there is no overall divergence, so by (5.3.7)

$$R(G) = \bar{R}(G). \tag{5.3.10}$$

There is only one subdivergence, so (5.3.8) collapses to give

$$\bar{R}(G) = U(G) + C_\gamma(G), \tag{5.3.11}$$

where γ is the divergent subgraph. Here $C_\gamma(G)$ is the full graph with γ replaced by $-T\circ U(\gamma)$. We reproduce the obvious result. There is one counterterm graph like Fig. 3.2.2(*a*) or (*c*).

We now look at a graph with two or more divergent subgraphs which do not intersect and which have no subdivergences. It is sufficient to consider G to be Fig. 5.1.1(*a*). There is no overall divergence, so again $R(G) = \bar{R}(G)$. Let γ_1 and γ_2 be the self-energy bubbles. Then the subdivergences correspond to the three subgraphs γ_1, γ_2, and $\gamma_1 \cup \gamma_2$. (Here $\gamma_1 \cup \gamma_2$ means, as usual, the union of γ_1 and γ_2.) So

$$R(G) = \bar{R}(G) = U(G) + C_{\gamma_1}(G) + C_{\gamma_2}(G) + C_{\gamma_1 \cup \gamma_2}(G). \tag{5.3.12}$$

Evidently $C_{\gamma_1}(G)$ is just $U(G)$ with γ_1 replaced by its counterterm, $-T\circ U(\gamma_1)$; and similarly for γ_2. But what is $C_{\gamma_1\cup\gamma_2}(G)$?

It corresponds to a subtraction for $\gamma_1 \cup \gamma_2$ for the region where all loop momenta are large. But we must subtract from it the counterterms for the regions where only one momentum is large:

$$C(\gamma_1 \cup \gamma_2) = -T\circ[U(\gamma_1)U(\gamma_2) + C(\gamma_1)U(\gamma_2) + U(\gamma_1)C(\gamma_2)]. \tag{5.3.13}$$

Here we used the fact that $\gamma_1 \cup \gamma_2$ is disconnected, so that

$$U(\gamma_1 \cup \gamma_2) = U(\gamma_1)U(\gamma_2). \tag{5.3.14}$$

To work out (5.3.13), we must define T when acting on a disconnected graph to act independently on its components. Thus:

$$T\circ[U(\gamma_1)U(\gamma_2)] = [T\circ U(\gamma_1)][T\circ U(\gamma_2)] = C(\gamma_1)C(\gamma_2), \tag{5.3.15}$$

$$T\circ[C(\gamma_1)U(\gamma_2)] = [T\circ C(\gamma_1)][T\circ U(\gamma_2)] = -C(\gamma_1)C(\gamma_2), \tag{5.3.16}$$

etc.

We used the property that $T\circ U(\gamma_i) = -C(\gamma_i)$. Furthermore, $T\circ[T\circ U(\gamma_i)] = T\circ U(\gamma_i)$, i.e., the pole part of a pole part is itself. We therefore find

that

$$C(\gamma_1 \cup \gamma_2) = [T \circ U(\gamma_1)][T \circ U(\gamma_2)], \tag{5.3.17}$$

so that we reproduce the counterterm graph Fig. 5.1.1(*d*).

The above procedure generalizes to an arbitrary graph. It may appear excessively complicated, but it allows the smoothest way of defining $R(G)$.

Finally, we observe that our definitions (5.3.6)–(5.3.10) exactly reproduce our results for the two-loop graphs like Figs. 5.1.2 and 5.1.3.

5.3.4 *Summary*

In this section we have proved very little. We have set up a series of definitions that state exactly what we mean by the renormalization of a Feynman graph. The notation we have introduced will be important in making proofs. What we will need to prove is that the overall counterterms are local and of a degree in momentum given by naive power-counting. We will also show how to solve the recursion to find an explicit formula due to Zimmermann (1969).

5.4 Three-loop example

The three-loop self energy graph of Fig. 5.4.1 in ϕ^3 theory in six dimensions is an example of a graph with nested and multiply overlapping divergences. We call it G. Its divergent subgraphs are:

$$\begin{aligned}
&\gamma_1 = \{\text{lines carrying loop momentum } k\},\\
&\gamma_2 = \{\text{lines carrying loop momentum } q\},\\
&\gamma_3 = \{\text{lines carrying loop momenta } k \text{ and/or } l\},\\
&\gamma_4 = \{\text{lines carrying loop momenta } q \text{ and/or } l\},\\
&\gamma_5 = \gamma_1 \cup \gamma_2.
\end{aligned} \tag{5.4.1}$$

The first four of these are connected 1PI vertex graphs; the last is a set of two unconnected vertex graphs. According to our definitions of Section 5.3 we have

$$\begin{aligned}
R(G) &= \bar{R}(G) + C(G)\\
&= \bar{R}(G) - T \circ [\bar{R}(G)].
\end{aligned} \tag{5.4.2}$$

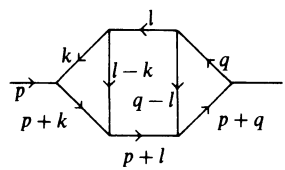

Fig. 5.4.1. Three-loop self-energy graph in ϕ^3 theory.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

This equation states that we first subtract subdivergences to obtain $\bar{R}(G)$, and then take off the overall divergence.

To define $\bar{R}(G)$ we subtract subdivergences:

$$\begin{aligned}\bar{R}(G) &= G + \sum_{i=1}^{5} C_{\gamma_i}(G) \\ &= G + \sum G\Big|_{\gamma_i \to C(\gamma_i)} \end{aligned} \tag{5.4.3}$$

We represent this as Fig. 5.4.2. The notation with the vertical bar in this equation denotes that we take G and replace γ_i by the corresponding counterterm $C(\gamma_i)$. In the figure, the labels $1,\ldots,4$ signify which of the subgraphs $\gamma_1,\ldots,\gamma_4$ has been replaced by its counterterm.

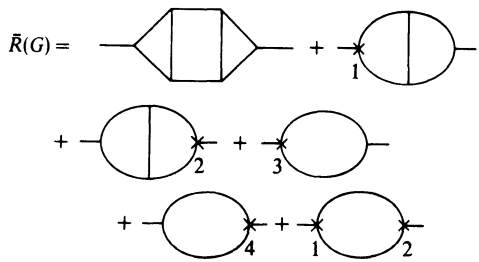

Fig. 5.4.2. Subtraction of subdivergences of Fig. 5.4.1.

Only γ_1 and γ_2 have no further subdivergences, so $C(\gamma_1)$ and $C(\gamma_2)$ are the ordinary one-loop counterterms. But we have still to define $C(\gamma_i)$ for $i = 3, 4, 5$:

$$\begin{aligned} C(\gamma_3) &= -T\circ[\gamma_3 - \gamma_3|_{\gamma_1 \to T(\gamma_1)}], \\ C(\gamma_4) &= -T\circ[\gamma_4 - \gamma_4|_{\gamma_2 \to T(\gamma_2)}], \\ C(\gamma_5) &= -T\circ[\gamma_1\gamma_2 - T(\gamma_1)\gamma_2 - \gamma_1 T(\gamma_2)] \\ &= [-T(\gamma_1)][-T(\gamma_2)]. \end{aligned} \tag{5.4.4}$$

The overall result is obtained by combining (5.4.2)–(5.4.4). If we represent the effect of applying T to a 1PI graph by enclosing it in a box, we can write $R(G)$ as shown in Fig. 5.4.3. There are sixteen terms in all. The first eight represent $U(G)$ minus its subdivergences, and the last eight form the subtraction for the overall divergence. The expansion of $R(G)$ represented in Fig. 5.4.3 is an example of the forest formula, to be discussed in the next section.

As we saw by examining two-loop graphs, the overall counterterm for a graph is non-local unless we first subtract off subdivergences. Otherwise there would be divergent contributions from where some but not all

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

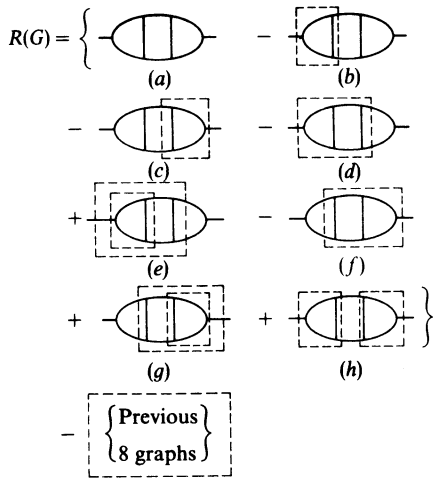

Fig. 5.4.3. Renormalization of Fig. 5.4.1.

subgraphs have large loop momenta. There would also be divergent contributions from the region where all loop momenta get large but at different rates. We can check from Fig. 5.4.3 that none of these problems occur for $R(G)$. Let us do this explicitly.

Let us show that the overall counterterm for G is local. We differentiate $\bar{R}(G)$ three times with respect to the external momentum p and show that the result is finite. Given the momentum routing of Fig. 5.4.1, there are three lines to differentiate: $p + k$, $p + l$, $p + q$. Here we have used the momentum carried by the line as a label for the line. Differentiating the original graph gives ten terms, where the three derivatives are applied to any combination of the three lines. One term is where we differentiate $p + k$ three times (Fig. 5.4.4). Although there is then no overall divergence, there remain subdivergences, so we must examine the corresponding differentiations applied to the counterterm graphs (b)–(h). We must regard the derivatives as acting on these graphs before divergences are computed (by the operation symbolized by the box).

The differentiation makes the subgraphs γ_1 and γ_3 completely finite, by removing both their overall divergence and γ_3's only subdivergence. Hence

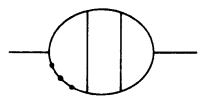

Fig. 5.4.4. One of the terms obtained by differentiating Fig. 5.4.1 three times with respect to its external momentum.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

the counterterm graphs (*b*), (*d*), (*e*), and (*h*) are zero after differentiation. This leaves graphs (*c*), (*f*), and (*g*). These cancel the subdivergences of (*a*) coming from the subgraphs γ_2 and γ_4, which are unaffected by the differentiation.

We may examine the other nine terms in $\partial^3 \bar{R}(G)/\partial p^3$ similarly, and we find that in fact $\partial^3 \bar{R}(G)/\partial p^3$ is finite, as claimed.

5.5 Forest formula

5.5.1 *Formula*

Zimmermann (1969, 1970) gave an explicit solution of the recursive definition of the renormalized value $R(G)$ of a graph G. The general idea can be gleaned from the example we examined in the previous section. There the recursion generated a series of sixteen terms. One was the original graph, and the others had the subtraction operation T applied one or more times. For example, in graph (*e*) we first replace the left-most loop γ_1 by $T(\gamma_1)$, with a result we can write as $T_{\gamma_1}(G)$. We then take the subgraph equivalent to γ_3, viz. $T_{\gamma_1}(\gamma_3)$, and replace it by the result of acting with T. This gives graph (*e*). The sum of the two graphs (*d*) and (*e*) is used as the subtraction for the subdivergences of G associated with γ_3.

Each of the sixteen terms is pictured as the original graph with some number of connected 1PI subgraphs enclosed in boxes to indicate application of T to the subgraph. Each term can be specified by giving its set of boxed subgraphs. Each such set is called a *forest*. The subgraphs which form a particular forest are either disjoint or nested: they are said to be *non-overlapping*. The set of all possible forests for G is called $\mathscr{F}(G)$.

There are sixteen forests occurring in Fig. 5.4.3. The first eight are (in set theory notation):

(*a*) the empty set $\varnothing$,
(*b*) $\{\gamma_1\}$,
(*c*) $\{\gamma_2\}$,
(*d*) $\{\gamma_3\}$,
(*e*) $\{\gamma_1, \gamma_3\}$,
(*f*) $\{\gamma_4\}$,
(*g*) $\{\gamma_2, \gamma_4\}$,
(*h*) $\{\gamma_1, \gamma_2\}$.

These do not contain the whole graph; they are called the *normal* forests. The other eight forests in Fig. 5.4.3 consist of one of the above eight forests to which is added as another element the complete graph G. A forest of G

containing G is called a *full* forest. The distinction between normal and full forests is that the normal forests subtract off the subdivergences, and the full forests combine to subtract the overall divergence.

Not all forests occur in Fig. 5.4.3; for example, the forest

$$U = \{\text{subgraph consisting of lines carrying loop momentum } l\}$$

does not appear. Such forests contain at least one overall convergent subgraph as an element.

Inspection of Fig. 5.4.3 shows that

$$R(G) = \sum_{U \in \mathcal{F}(G)} \prod_{\gamma \in U} (-T_\gamma) G. \tag{5.5.1}$$

Here the sum is over all forests U of G. The operator T_γ replaces γ by $T(\gamma)$. Note that for nested γ's the T_γ's should be applied inside to outside. Equation (5.5.1) is called the forest formula; it is due to Zimmermann (1969). Suppose we compute $R(G)$ for an arbitrary graph G by using (5.5.1). Then, as we will prove shortly, the result is the same as if we used the recursive definition of $R(G)$ given in Section 5.3.

It is convenient to let the sum over forests be over all forests rather than only over those consisting of subgraphs that are superficially divergent; the extra forests give a zero contribution. The reason for doing this is that we will sometimes wish to change the definition of T so that we make subtractions for some convergent graphs, as well as for divergent graphs. For example, such a redefinition will be the key to proving the operator product expansion in Chapter 10.

We now have both a recursive and a non-recursive definition for the renormalization of a Feynman graph. It will prove very useful to have both definitions available. Different proofs will need different forms of the definition. In particular, proofs by induction on the number of loops of a graph will naturally use the recursive definition.

5.5.2 *Proof*

The proof of the forest formula is elementary, but somewhat involved. We first use (5.5.1), and the following equations:

$$\bar{R}(G) = \sum_{U \in \mathcal{F}(G)} \prod_{\gamma \in U} (-T_\gamma) \circ G; \tag{5.5.2}$$

$$C(G) = \left\{ \begin{array}{ll} -T_G \circ \bar{R}(G), & \text{if } G \text{ is 1PI,} \\ \prod [C(\gamma_i)], & \text{if } G \text{ is a disjoint union of 1PI } \gamma_i\text{'s,} \\ 0 & \text{otherwise,} \end{array} \right\} \tag{5.5.3}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

as definitions of $R(G)$, $\bar{R}(G)$, and $C(G)$. Here $\bar{\mathscr{F}}(G)$ is the set of normal forests of G (i.e., those that do not contain G). These definitions are correct for the graph Fig. 5.4.1, as can be seen by inspection of Fig. 5.4.3.

Since the recursive definitions uniquely give $R(G)$, $\bar{R}(G)$, and $C(G)$ in terms of the operation T_γ, it suffices to show that (5.5.1)–(5.5.3) satisfy the recursion relations (5.3.6)–(5.3.9). Notice first that (5.3.6) and (5.3.9) are really the same, except for being applied to different graphs.

If $\bar{R}(G)$ given by (5.5.2) is correct, and if subgraphs are correctly renormalized, then (5.5.3) is equivalent to our original definition (5.3.6) of $C(G)$. Moreover, suppose that G is connected and one-particle-irreducible. Now each forest of G is either a normal forest, that is, a forest of which G is not an element, or it is a normal forest to which is adjoined G. Then the formula (5.3.7) for $R(G)$ is a direct consequence of (5.5.1)–(5.5.3) for such a graph. If G is not a union of 1PI graphs, then there is no overall divergence, and again (5.3.7) holds.

So it remains to prove the following:

(1) $R(G)$ is correct when G is a disjoint union of more than one 1PI graph. (Note that this case occurs in renormalizing the graph of Fig. 3.2.1(*b*), as we saw in Section 5.3.)
(2) $\bar{R}(G)$ is correct, i.e., it satisfies (5.3.8), for a general graph.

If G is a disjoint union of 1PI graphs γ_i, then each forest is a union of forests, one for each component. Then $R(G) = \prod_i R(\gamma_i)$, as we should expect. The problem is that this is not manifestly true in the recursive definition, where we make an overall subtraction for G. We dealt with this problem between (5.3.11) and (5.3.17).

Our proof of (5.3.8) is by induction on the size of G. Now a one-loop 1PI graph has no non-trivial subgraphs, so its only normal forest is the empty set. Then formula (5.5.2) collapses to $\bar{R}(G) = U(G)$, just as it should. This enables us to start the induction.

It remains to prove (5.3.8*b*). For this, observe that each forest U has a unique set of biggest subgraphs $M_1,\ldots,M_j$. Each M_i is contained in no bigger subgraph in U, and each $\gamma \in U$ is contained in some M_i. The existence and uniqueness of this set of M's is seen by considering pairs γ_i, γ_k of elements of U. Since γ_i and γ_k are non-overlapping, there are three possibilities:

(1) $\gamma_i \subset \gamma_k$, in which case remove γ_i from further consideration.
(2) $\gamma_k \subset \gamma_i$, in which case remove γ_k from further consideration.
(3) $\gamma_i \cap \gamma_k = \varnothing$, in which case leave both in.

Repeat until no further eliminations are possible; then the result is the set of M_i's.

The forest U is the union of a full forest, one for each M_i. We can write our definition (5.5.2) of $\bar{R}(G)$ as

$$\bar{R}(G) = G + \sum_{M_1,\ldots,M_n} \left\{ \prod_{i=1}^{n} (-T_{M_i}) \sum_{U_1 \in \mathfrak{F}(M_1)} \cdots \sum_{U_n \in \mathfrak{F}(M_n)} \times \right.$$
$$\left. \times \left[\prod_{\gamma_1 \in U_1} (-T_{\gamma_1}) \cdots \prod_{\gamma_n \in U_n} (-T_{\gamma_n}) \right] G \right\}. \tag{5.5.4}$$

Here the first term comes from the case in (5.5.2) that $U = \varnothing$, and the sum in the second term is over non-empty sets of disjoint 1PI graphs M_i *excepting the case that* $M_i = G$. By setting $\gamma = M_1 \cup \ldots \cup M_n$ and using (5.5.3) to determine $C(M_1 \cup \ldots \cup M_n)$, we find (5.3.8*b*).

5.6 Relation to $\mathscr{L}$

We have seen how to renormalize an individual Feynman graph by making a series of subtractions. The motivation for doing this came from consideration of examples in which the subtractions were generated by counterterms in the interaction Lagrangian. We will now show that this is true to all orders. We will assume the natural result (to be proved later) that the polynomial degree of the overall counterterm of a graph is given by its degree of divergence, just as for low-order graphs.

First, we must make precise the result that we will prove. For each 1PI graph G, we have constructed its overall counterterm $C(G)$. Since this is a polynomial in the external momenta of G, it can be written as the vertex derived from an interaction term $D(G)/N(G)$ in the Lagrangian $\mathscr{L}$. Here $N(G)$ is a symmetry factor of the same sort as the 3! that appears with the ϕ^3 interaction term in $\mathscr{L}$. Each power of a momentum entering $D(G)$ corresponds to i times a derivative of the corresponding field. If G is an n-point graph and each of its external lines corresponds to the same type of field, then $N(G)$ is $n!$. If there are a number of different fields and n_i is the number of lines of type i entering G then

$$N(G) = \prod_i n_i!. \tag{5.6.1}$$

For each graph for a Green's function, the forest formula gives a set of graphs with counterterms. We will demonstrate that the set of counterterm graphs is generated from the counterterm vertices in the interaction Lagrangian.

Consider ϕ^3 theory. As before, we write the Lagrangian as:

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_b + \mathscr{L}_{ct} = \mathscr{L}_0 + \mathscr{L}_I. \tag{5.6.2}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

The free Lagrangian

$$\mathscr{L}_0 = (\partial\phi)^2/2 - m^2\phi^2/2 \tag{5.6.3}$$

generates the propagator, while the interaction $\mathscr{L}_I$ consists of two terms, $\mathscr{L}_b$ and $\mathscr{L}_{ct}$. The basic interaction is

$$\mathscr{L}_b = -g\mu^{3-d/2}\phi^3/3!, \tag{5.6.4}$$

and $\mathscr{L}_{ct}$ is the counterterm Lagrangian used to cancel the ultra-violet divergences:

$$\mathscr{L}_{ct} = \sum_G D(G)/N(G). \tag{5.6.5}$$

Here the sum is over all 1PI graphs. Those that have no overall divergence generate no counterterm; for these $D(G) = 0$. Each 1PI graph that has an overall divergence generates a term in (5.6.5). The formulae (5.6.2) and (5.6.5) apply in any theory.

Since (5.6.5) applies to any theory, it applies in particular to ϕ^3 in higher than six dimensions. Thus it enables us to renormalize a non-renormalizable theory. But the sum must include counterterms $D(G)$ with an arbitrarily large number of powers of momentum and with an arbitrarily large number of external lines for G. It is only in six or fewer dimensions that the counterterms have the same form as terms in the basic ϕ^3 Lagrangian $\mathscr{L}_0 + \mathscr{L}_b$.

Now that we have developed a convenient notation, the most difficult part of the proof is to ensure that the combinatorial factors come out right. We will prove that the Lagrangian defined by (5.6.2) and (5.6.5) gives the same renormalized Green's functions as those generated by our recursive definition in Section 5.3 (and therefore the identically same Green's functions as given in Section 5.5 by the forest formula). The proof will be given for ϕ^3 theory in six or fewer dimensions, but it easily generalizes.

Consider the full N-point Green's function G_N at order g^P. It is sufficient to work only with connected graphs. If the theory is renormalizable (as we will prove in Section 5.7), then the sum of counterterms has the form:

$$\mathscr{L}_{ct} = \delta Z(\partial\phi)^2/2 - \delta m^2\phi^2/2 - \delta g\phi^3/3!. \tag{5.6.6}$$

with (by (5.6.5))

$$\begin{aligned} -\delta Z &= \sum_{2\text{-point } G} [\text{Coefficient of } -ip^2 \text{ in } C(G)], \\ -\delta m^2 &= \sum_{2\text{-point } G} [\text{Coefficient of } ip^0 \text{ in } C(G)], \\ -\delta g &= \sum_{3\text{-point } G} C(G)/i. \end{aligned} \tag{5.6.7}$$

We ignore the tadpoles, yet again. The term of order g^P in the perturbation

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

expansion of G_N has vertices generated by the different terms in $\mathscr{L}_b + \mathscr{L}_{ct}$. There will be graphs with all of their vertices being the basic interaction $\mathscr{L}_b$. Let the set of these be called B. The other graphs will contain one or more of the counterterm vertices generated by (5.6.6). Each counterterm can then be decomposed into a sum of terms by applying (5.6.7) at each counterterm vertex. Each term has each of the counterterm vertices replaced by the overall divergence of some graph. Then in the result, each term T corresponds to a unique basic graph $b(T) \in B$.

So we have

$$G_N = \sum_G \left(G + \sum_{b(T)=G} T \right). \tag{5.6.8}$$

On the other hand, we have constructed the renormalization of each of the graphs G by writing

$$R(G) = G + \sum_{\gamma \subseteq G} C_\gamma(G). \tag{5.6.9}$$

Each of the terms T in (5.6.8) is constructed by replacing each of a set of one or more disjoint 1PI subgraphs $\gamma_1, \ldots, \gamma_j$ by its counterterm given by $iD_{\gamma_i}(G)$. On identifying γ in (5.6.9) with $\gamma_1 \cup \gamma_2 \cdots \cup \gamma_j$, we expect that

$$G_n = \sum_G R(G). \tag{5.6.10}$$

This result would be obvious, were it not that the symmetry factors do not manifestly match up.

The problem is illustrated by Fig. 5.6.1. There the basic graph is (*a*), and the complete set of subtractions needed to renormalize it consists of (*b*)–(*e*). Now the symmetry factor for (*a*) is 1/8: There is a factor 1/2 for each self-energy graph and an overall 1/2 for the top–bottom symmetry of the whole graph. Each of the subtractions (*b*) and (*c*) has a symmetry factor 1/4, since the remaining factor 1/2 goes into the counterterm for the self-energy. Both graphs (*b*) and (*c*) are equal.

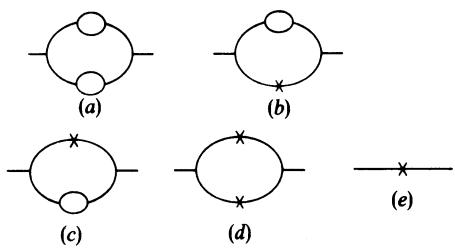

Fig. 5.6.1. Renormalization of a graph, to illustrate symmetry factors.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

Considered as Feynman graphs, these are the same graph, but with symmetry factor 1/2. So they give one term with factor 1/2 in (5.6.8) (derived from the Lagrangian), while in (5.6.9) (from the recursion formula) there are two equal terms with factor 1/4. The end result is the same. We must consider (*b*) and (*c*) as distinct graphs when defining $R(G)$, since they correspond to different regions of loop-momentum space – we must take each momentum variable to be distinguishable.

To construct a general proof is tedious. The symmetry factor of a graph G is $1/N(G)$, where $N(G)$ is the dimension of the graph's symmetry group. So we write

$$G = \bar{G}/N(G), \tag{5.6.11}$$

where the overbar indicates computation ignoring all symmetry factors. Similarly we define $\overline{C(G)}$ by

$$C(G) = \overline{C(G)}/N(G). \tag{5.6.12}$$

Now the renormalized value of a graph G is

$$\begin{aligned} R(G) &= G + \sum_{\gamma \subseteq G} C_\gamma(G) \\ &= \frac{1}{N(G)}\left[\bar{G} + \sum_\gamma \overline{C_\gamma(G)}\right] \\ &= \frac{1}{N(G)}\bar{G} + \sum_\gamma \left(\frac{\prod_i N_i}{N(G)}\right)\left[\frac{1}{\prod_i N_i}\overline{C_\gamma(G)}\right]. \end{aligned} \tag{5.6.13}$$

In the last line we have observed that γ is a disjoint union of 1PI graphs γ_1, $\gamma_2, \ldots$. Moreover, we have explicitly indicated the symmetry factors $1/N_i \equiv 1/N(\gamma_i)$ for each γ_i which is replaced by its overall counterterm. For a given subgraph $\gamma = \cup \gamma_i$ the symmetry groups of the γ_i's are a commuting set of subgroups of the symmetry group of G. Therefore the quantity $N(G)/\prod_i N_i$ must be an integer.

Next, consider the Green's functions generated by the Lagrangian (5.6.2), as in (5.6.8),

$$G_N = \sum_G \left\{\frac{1}{N(G)}\bar{G} + \sum_{\gamma \subseteq G}{}' \frac{1}{N(G/\gamma)}\overline{C_\gamma(G)}\right\}. \tag{5.6.14}$$

Here we have observed that each graph containing one or more counterterms is generated from a basic graph by replacing some 1PI subgraphs $\gamma_1, \ldots, \gamma_j$ by counterterms. We write γ as the union of the γ_i's. Then we let G/γ be the graph resulting from substituting counterterms for the γ_i's. By the

definition of the counterterm Lagrangian, the result is the same as $C_\gamma(G)$, aside from symmetry factors. The prime on the $\sum'$ indicates that only γ's giving distinct Feynman graphs are considered. (Thus, for example, Figs. 5.6.1(*b*) and (*c*) are not counted separately.) Thus, if we define

$$K(G,\gamma) = [\text{number of graphs } \gamma' \text{ for which } G/\gamma = G/\gamma'],$$

then we must prove that

$$K(G,\gamma) = \frac{N(G)}{\left[\prod_i N_i\right][N(G/\gamma)]}. \tag{5.6.15}$$

It is easiest to couch this final step in the language of group theory. The denominator of the right-hand side of (5.6.15) is the product of dimensions of commuting subgroups of the symmetry group of G. (Note that, for example, two ϕ^3 counterterms generated by different self-energy subgraphs are counted as different.) These subgraphs generate another subgroup, of which the set of cosets in the symmetry group of G has exactly the dimension of the right-hand side of (5.6.15). But, concretely, each coset corresponds to one of the graphs counted by $K(G,\gamma)$.

5.7 Renormalizability

5.7.1 *Renormalizability and non-renormalizability*

In this section we explain the properties of renormalizability, non-renormalizability, and super-renormalizability of a field theory. We do this first for every order of perturbation theory, and then we consider to what extent the properties are true beyond perturbation theory, for the complete theory. The method in perturbation theory is power-counting and dimensional analysis.

Consider first ϕ^3 theory in a space-time of integer dimension d_0. We have seen how to renormalize it to get finite Green's functions by adding counterterms (5.6.5) to the Lagrangian. Each counterterm is a polynomial in the field ϕ and its derivatives. The theory is called renormalizable if the only counterterms needed are proportional to the terms $(\partial\phi)^2$, ϕ^2, and ϕ^3 present in the original Lagrangian $\mathscr{L}_0 + \mathscr{L}_b$. This is equivalent to saying that the Lagrangian has the form

$$\mathscr{L} = (\partial\phi_0)^2/2 - m_0^2\phi_0^2/2 - g_0\phi_0^3/3!, \tag{5.7.1}$$

where the bare field ϕ_0 is $Z^{1/2}\phi$. The bare mass m_0, the bare coupling g_0, and the field-strength renormalization Z each have singular behavior as the

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

ultra-violet regulator is removed. A linear term $\delta h\phi$ is needed as well. We may regard it as being present in the original Lagrangian. In any event it is only a single extra coupling. It can be ignored if we impose the renormalization condition that $\langle 0|\phi|0\rangle = 0$ to determine δh, and if we ignore tadpole graphs.

We generalize to an arbitrary theory by calling a theory renormalizable if the Green's functions of its elementary fields can be made finite by rescaling the fields (in a cut-off dependent way) and by making some suitable cut-off dependent change in the couplings and masses.

In perturbation theory we determine whether or not we have renormalizability by examining the possible values of the degree of divergence $\delta(G)$ for the 1PI graphs. For every graph G a counterterm is needed if $\delta(G) \geq 0$. As we will prove in Section 5.8 the counterterm $C(G)$ is a polynomial of degree $\delta(G)$ in the external momenta, and provided we use a scheme like dimensional regularization that preserves Poincare invariance, the counterterms are Poincare invariant.

Let us now determine whether or not ϕ^3 theory in d space-time dimensions is renormalizable. In d space-time dimensions the N-point 1PI graphs have dimension (in momentum space)

$$d(G_N) = N + d - Nd/2.$$

Then (by (3.3.12)) the degree of divergence of a graph for G_N at order g^P is

$$\delta(G_N) = d + (1 - d/2)N + (d/2 - 3)P. \tag{5.7.2}$$

Note that the minimum value of P to have a one-loop connected graph is N.

Inspection of (5.7.2) shows that if $d > 6$ then, for any value of N, there can be made N-point graphs with arbitrarily high degree of divergence by going to large enough order in g. The theory is therefore not renormalizable if $d > 6$, and the non-renormalizability is a direct consequence of the negative dimension of g.

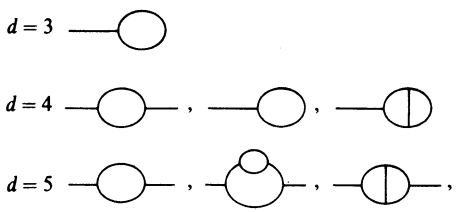

Fig. 5.7.1. All the graphs with overall divergences in ϕ^3 theory at those space-time dimensions where it is super-renormalizable.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

If $d = 6$, only the one-, two-, and three-point functions are divergent, with degree of divergence 4, 2, and 0, respectively. The permissible counterterms are just terms of the form of those in $\mathscr{L}_0 + \mathscr{L}_b$, so the theory is renormalizable if $d = 6$. Moreover, there is a divergence in every order of g (except for tree graphs, of course).

If $d = 3$, 4, or 5, then only a finite set of graphs, illustrated in Fig. 5.7.1, have overall divergences, and renormalization is needed only for the mass and for the tadpole coupling. Again we have renormalizability.

5.7.2 *Cosmological term*

Strictly speaking, we should also consider Feynman graphs with no external lines. These are the vacuum bubbles. They generate the energy density of the vacuum, and normally are ignored. But in gravitational physics, they cannot be ignored. Counterterms for such graphs (present in ϕ^3 theory whenever $d \geq 2$) are proportional to the unit operator. They are a renormalization of what in General Relativity is the cosmological constant. A counterterm is even needed for free-field theory – where the divergence is conventionally removed by normal-ordering (see, for example, Bjorken & Drell (1966)). We see that normal-ordering is nothing but a primitive form of renormalization.

5.7.3 *Degrees of renormalizability*

It is convenient to distinguish three types of renormalizable theory:

(1) Finite: no counterterms needed at all.
(2) Super-renormalizable: only a finite set of graphs need overall counterterms.
(3) Strictly renormalizable: infinitely many graphs need overall counterterms. (But note that they only renormalize a finite set of terms in the basic Lagrangian, since we assumed renormalizability of the theory.)

Finiteness or super-renormalizability normally occur when all the couplings in the basic Lagrangian have positive dimension.

Note that in a super-renormalizable theory, the number of divergent basic graphs is infinite. For example, even if there is only one graph γ with an overall divergence, any graph containing γ as a subgraph is divergent. However all such graphs become finite after adding to γ its counterterm, so only one counterterm, $C(\gamma)$, appears in the Lagrangian.

Mathematical physicists (see Glimm & Jaffe (1981)) have investigated renormalizability beyond perturbation theory. This is important, since

perturbation series are in general asymptotic series rather than convergent series. Thus one cannot simply sum the perturbation series to obtain the complete theory. Even so, it has been proved for many super-renormalizable theories that perturbation theory gives an exactly correct account of the divergences. (A much investigated case is ϕ^4 theory in two and in three space-time dimensions.)

In a super-renormalizable theory, the series for a bare mass or coupling in terms of the renormalizable quantities has a finite number of terms. Therefore the series converges, and one only has to prove that (*a*) perturbation theory is asymptotic to the true theory, and (*b*) there are no terms like $\exp(-1/g)$ in the bare masses or couplings that are smaller than any power of g. The rigorous proof amounts to showing that in summing the perturbation series to a finite order, the error is correctly estimated by the first term omitted. In particular, this applies to the existence of any possible ultra-violet divergence.

Rigorous proofs are not yet available for any strictly renormalizable theory. One difficulty is obvious: the series for, say, the bare coupling, g, is an infinite series, each term of which diverges as the UV cut-off is removed. Since the series is presumably asymptotic rather than convergent, one cannot directly obtain any information about renormalization in the full theory: the error obtained in using a truncated form of the series is of the order of the first term omitted, and that is always divergent.

It might even appear that perturbation theory has no light at all to shed on the question of renormalizability of the full theory. This is in fact not so, as we will see when we discuss the renormalization group in Chapter 7. If the theory has the property called asymptotic freedom then a series of suitable redefinitions of g allows short-distance phenomena to be computed reliably. In particular the UV divergences can be computed in terms of weak coupling series without divergent coefficients. It is sensible to conjecture that a suitably refined analysis can be made to obtain rigorous bounds of the errors so that the perturbative results correctly give the divergences. Monte-Carlo studies of the functional integral (Creutz & Moriarty (1982)) support this conjecture. In four dimensions, only certain non-abelian gauge theories (including QCD) are asymptotically free (Gross (1976)).

We will also see in Chapter 7 that in non-asymptotically free theories, like ϕ^4 and QED in four dimensions, perturbation theory cannot reliably describe short-distance phenomena. There are, in fact, indications (Symanzik (1982)) that the full ϕ^4 theory is not renormalizable, contrary to the situation order-by-order in perturbation theory.

5.7.4 *Non-renormalizability*

For theories which are not renormalizable in perturbation theory, there are many possibilities. Among them are the following:

(1) There is only a finite set of 1PI Green's functions which have overall divergences. A typical case is ϕ^3 theory in six or fewer space-time dimensions when the basic Lagrangian,

$$\mathscr{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - g\phi^3/3!, \tag{5.7.3}$$

has no term linear in ϕ. The one-, two-, and three-point functions have divergences, but there is no term $h\phi$ whose coupling can be renormalized to cancel the divergence of the tadpole graphs. However, addition of such a term generates a renormalizable theory. More generally, suppose we have a finite set of overall-divergent Green's functions. A renormalizable theory is generated by adding a finite set of extra interactions.

(2) There is an infinite set of Green's functions with overall divergences. However, for all but a finite set of the Green's functions, the divergences cancel after summing over all graphs of a given order. (There are no known cases of this.)

(3) As for case 2, except that the divergences cancel only for the S-matrix, rather than for all off-shell Green's functions. An important case is a spontaneously broken gauge theory, when it is quantized in its unitary gauge.

(4) The theory is made renormalizable by going beyond perturbation theory in some systematic and sensible way. One case (as in the Gross–Neveu (1974) model – see Gross (1976)) is of a theory that is strictly renormalizable and asymptotically free for some dimension $d = d_0$, and that is considered in some dimension d slightly greater than d_0.

(5) As for case 1, except that the extra terms make physical nonsense. A case is the Yang–Mills theory with a mass term in Feynman gauge. Then the extra terms destroy unitarity ('t Hooft (1971a)).

(6) None of the above.

Roughly speaking, there are no general rules. Each case must be handled separately. Only for the last two cases (5 and 6) should a theory be called non-renormalizable. A fundamental theory should be renormalizable, for otherwise either physical quantities are actually infinite or they are finite, but an infinite set of parameters is needed to specify the finite parts of the counterterms.

Nevertheless, a statement that a particular theory is non-renormalizable

is really a statement of ignorance: nobody has found a way to construct a physically sensible version of the theory. (Cases 1 to 5 are where somebody has found a way.) In practice, when a theory is labelled non-renormalizable, what is usually meant is that the theory is not renormalizable order-by-order in perturbation theory; such a statement can be proved by calculating a finite number of graphs.

Within the usual functional-integral approach (with a lattice cut-off), not only has the complete ϕ^4 theory been proved renormalizable for $d = 2$ and 3, but it has been proved non-renormalizable for $d > 4$ (Aizenman (1981)).

5.7.5 *Relation of renormalizability to dimension of coupling*

To prove perturbative renormalizability of a theory of scalar fields, we generalize the argument leading to (5.7.2). The argument will apply when no coupling has negative dimension. Renormalizability will hold with possibly the addition of extra interactions (like the $h\phi$ term in ϕ^3 theory) whose coefficients have non-negative dimension. Our proof will easily generalize to theories with fermion and gauge fields. The problems we will encounter in gauge theories will all be to do with the question of whether these extra terms are compatible with the gauge invariance. But we will leave these questions to Chapter 12.

Let a general term in $\mathscr{L}$ or $\mathscr{L}_{\text{ct}}$ be written schematically as

$$(\text{coupling } f)(\text{derivative})^A(\text{field})^N. \tag{5.7.4}$$

The vertex generated by this term is one possible graph for the 1PI Green's function Γ_N with N external lines. Thus the dimension of Γ_N satisfies

$$d(\Gamma_N) = d(f) + A. \tag{5.7.5}$$

Since no coupling has negative dimension, the degree of divergence of any graph for Γ_N is at most $d(\Gamma_N)$, as we saw from examples in Section 3.3.3, and as we will prove in Section 5.8. That is, the degree of divergence $\delta(\Gamma_N)$ satisfies

$$\delta(\Gamma_N) \leq d(\Gamma_N), \tag{5.7.6}$$

with equality only for graphs all of whose couplings have zero dimension.

To renormalize the N-point graphs, we add counterterms of the form (5.7.4) with at most $\delta(\Gamma_N)$ derivatives. So the possible counterterms satisfy

$$d(f) = d(\Gamma_N) - A \geq \delta(\Gamma_N) - A \geq 0. \tag{5.7.7}$$

The last inequality follows since a counterterm with A derivatives is needed only if the degree of divergence is at least A. From (5.7.7) it follows that we

need no couplings of negative dimension, given that none are present in the original Lagrangian.

Some of the couplings generated as counterterms may not be present in the original Lagrangian even if it contains no couplings of negative dimension. But the number of new couplings needed is nevertheless finite, because only a finite set of counterterms satisfy (5.7.7).

5.7.6 *Non-renormalizable theories of physics*

From the discussion above, it is natural to conclude that a theory of physics should be renormalizable. In fact, the strong, electromagnetic, and weak interactions appear to be described by a renormalizable theory. This theory is a combination of quantum chromodynamics for strong interactions and the Weinberg–Salam theory for weak interactions.

Around 1970 there was a revolution in the theory of weak interactions when it was discovered that non-abelian gauge theories are renormalizable. It is precisely one of these theories that was found to be necessary to construct a renormalizable theory of weak interactions in agreement with experiment. See Beg & Sirlin (1982) for a historical review.

Unfortunately, this progress has not extended to gravity. Einstein's theory of general relativity is non-renormalizable, after quantization, and there is no very promising alternative. (This situation exists despite many significant attempts to improve it – Hawking & Israel (1979).)

It is a mistake to suppose that non-renormalizable theories should be banished from consideration. Remember that for many years weak interactions were *successfully* calculated using the 'four-fermion' theory, which is non-renormalizable. For most purposes, weak interactions could be adequately treated in the lowest order of perturbation theory, where no renormalization is needed. But the non-renormalizability of higher-order calculations raised the question of consistency of the theory: is it legitimate to calculate even an approximation to a nonsensical (i.e., non-existent) theory? Will the results of calculations mean anything? The same questions arise in gravity. There, the classical theory of general relativity is very successful, but the quantized theory is badly non-renormalizable.

We must therefore understand how and why we may use non-renormalizable theories in physics.

Now, to perform consistent calculations in any theory which contains ultra-violet divergences, we must impose an ultra-violet cut-off, M, of some sort. In the case of a renormalizable theory we can take M to infinity and obtain finite results that are insensitive to the cut-off. Another related

property of a renormalizable theory is the decoupling theorem of Appelquist and Carazzone, which we will discuss in Chapter 8. This theorem applies to a renormalizable theory which contains fields whose masses are much bigger than the energies of the scattering processes under consideration. The theorem states that the heavy fields can be deleted with only a small effect (suppressed by a power of the heavy mass) on cross-sections, etc. The hallmark of a renormalizable theory is in fact that it is complete in itself. It contains no direct indications of whether it is only part of a larger and more complete theory.

These statements are false for a non-renormalizable theory. Consider the old four-fermion theory of weak interactions. Its coupling is $G \sim 10^{-5}$ GeV^{-2}. We cannot take the UV cut-off arbitrarily large, for an nth order graph has a divergence of order $2n$; it behaves like $(M^2G)^n$ for large cut-off M. Counterterms to make the graph finite need a correspondingly large number of derivatives, but only a finite number of counterterms are available. Hence we cannot take the cut-off to infinity, and if we want insensitivity to the cut-off we must take $M \ll G^{-1/2}$. Moreover, the energy, E, of the process under consideration must be much less than M, otherwise the calculation is dominated by details of the cut-off procedure. In other words, the four-fermion interaction is a good approximation to physics only if $E \ll M \ll G^{-1/2}$. The minimum possible relative error of calculations is of the order of the maximum of M^2G and E^2/M^2.

Now, it is always possible in principle to do experiments at arbitrarily high energy. So the applicability of four-fermion theory at low energies implies that at energies rather below $G^{-1/2} \sim 300\,\mathrm{GeV}$ there is new physics. That is, the four-fermion theory becomes incorrect at that energy. The last fifteen years of weak interaction physics confirms this. (See, for example, Bjorken (1982))

For gravity, the corresponding energy scale is the Planck mass, of the order of $10^{19}\,\mathrm{GeV}$. This is extremely far beyond the range of normal accelerator experiments. Evidence for phenomena on such an energy scale must come from much more esoteric observations. Examples might be found in certain areas of the cosmology of the early universe, or from seeing the decay of a proton (Langacker (1981)).

In any case, a non-renormalizable theory contains indications that it cannot describe all phenomena. It contains the seeds of its own destruction as a viable theory of a field of physics. So, given a successful non-renormalizable theory, one must ask the following questions: (1) 'Of which more complete theory is it a part?' (2) 'How is it related to that theory?' An example is given by the relation between the Weinberg–Salam theory,

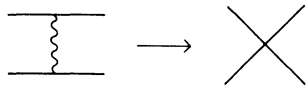

Fig. 5.7.2. W-boson exchange gives an effective four-point interaction at low energies.

which is renormalizable, and the four-fermion theory, which is not. The four-fermion theory arises as an approximation to W-boson exchange at low energy (Fig. 5.7.2). One replaces the propagator

$$\mathrm{i}/(q^2 - m_W^2),$$

for the W-boson, by

$$\mathrm{i}/(-m_W^2).$$

The graph is suppressed by factor of at least E^2/m_W^2 compared to photon exchange. It gives an example of the decoupling theorem: the heavy particles have small effects at low energies.

The only reason we can see such effects is the high degree of symmetry of the strong and electromagnetic interactions. These interactions conserve P, C, T, and the number of each flavor of quark and of each flavor of lepton. Weak-interaction amplitudes are much smaller than strong-interaction or electromagnetic amplitudes for similar processes, and are therefore normally invisible. But there are many processes that are completely forbidden in the absence of weak interactions; for these, any weak-interaction cross-section, no matter how small, is all there is.

So one important way in which a non-renormalizable theory arises is as a low-energy approximation to a renormalizable theory in a process that is forbidden in the absence of the heavy fields. The heavy fields effectively give a cut-off on the non-renormalizable theory. Then, for example, the four-fermion coupling G is computable in terms of the underlying theory via a formula like

$$G = \text{constant } g^2/m_W^2 + \text{higher order corrections in } g.$$

Here g is the dimensionsless coupling of the Weinberg–Salam theory. One manifest characteristic of this non-renormalizable theory is the weakness of its interactions. Also note that a higher power of G is a higher inverse power of m_W^2. We are taking the leading power of m_W as m_W gets large, so it is in general incorrect to calculate in the non-renormalizable theory beyond lowest order. Higher-order calculations must be done in the full theory.

A slightly different situation arises in gravity. There one must perform calculations beyond tree approximation, since gravitationally bound states

like the solar system are formed by multiple exchange of gravitons. Counterterms are generated involving higher-derivative interactions (e.g., R^2, $R_{\mu\nu}^2$. etc.). The ambiguity in the finite parts of these counterterms gives an uncertainty in the Green's functions. However the uncertainty is a power of momentum divided by some large mass scale, and is negligible for low-momentum-transfer processes. In weak interactions, the size of the higher-order corrections is of the same order of magnitude as the intrinsic error in the calculations, but in gravity this is not so because of the zero mass of the graviton.

Another difference is that gravity is actually the strongest of the four fundamental interactions when considered on a large enough scale. In contrast, on atomic or molecular scales, it is the other three interactions that are by far the strongest. However, the strong and weak interactions have a finite range, so that they are essentially zero outside the nucleus. Particles can have both signs of electric charge, so that bulk matter, if charged, tends to attract charge of the opposite sign to it. Bulk matter is therefore generally neutral. But gravity couples to mass or energy, so it is always attractive. Hence gravity wins out as the strongest interaction for large enough assemblages of matter. However at nuclear and atomic scales, it is negligible by a factor of about 10^{40} compared to the other interactions.

Let us summarize by restating the key conclusion about the distinction between renormalizable and non-renormalizable theories. A non-renormalizable theory considered at low energy gives some indications that at high enough energies it must break down, and cannot be a complete theory. A renormalizable theory gives no such indication.

5.8 Proof of locality of counterterms; Weinberg's theorem

In our examples, we saw that the counterterm $C(G)$ of a graph G is a polynomial in its external momenta, of degree equal to its overall degree of divergence $\delta(G)$. This is a general property, as we will now prove.

The original proof of this theorem and some related results is due to Weinberg (1960); a simpler proof was given by Hahn & Zimmermann (1968). It is useful to distinguish three results:

(1) Suppose that a 1PI graph G and all its 1PI subgraphs have negative degree of divergence. Then the graph is finite. That the degrees of divergence of the graph and subgraphs are negative means that there is no divergence when all or some of the loop momenta go to infinity together, with the other momenta finite. The problem is to eliminate the possibility of a divergence from more exotic scalings.

(2) Suppose that a 1PI graph G has negative degree of divergence, but that it might have subdivergences. Then the graph is finite if we first subtract off subdivergences. More simply, if $\delta(G) < 0$ then $\bar{R}(G)$ is finite.
(3) If a 1PI graph G has degree of divergence $\delta(G)$, then its overall counterterm $C(G)$ is polynomial in the external momenta of G of degree $\delta(G)$.

Property (1) is a trivial case of (2). We will reduce (3) to (2) by the same differentiation method as we used in Section 5.2.2. The proofs will be by induction. This naturally suggests that we use the recursive definition of the renormalization $R(G)$ of G.

One generalization is useful. It is that the renormalization prescription may be chosen so that result (3) reads ('t Hooft (1973), Weinberg (1973), and Collins (1974)):

(3′) If a 1PI graph G has degree of divergence $\delta(G)$, then its overall counterterm $C(G)$ is polynomial in the external momenta of G and in the massive parameters in the Lagrangian. (The parameters in question are the masses of fermions and the squared masses of bosons.) The dimensions of the terms in the polynomial are at most $\delta(G)$.

5.8.1 *Degree of counterterms equals degree of divergence*

We first prove Property (3), that the overall counterterm $C(G)$ of a 1PI graph is polynomial in the external momenta of degree $\delta(G)$. We will do this assuming Property (2), that a graph with its subdivergences subtracted is finite if its degree of divergence is negative. Let G be a 1PI graph with degree of divergence $\delta(G) \geq 0$. We will consider $\bar{R}(G)$, which is G plus counterterms for its subdivergences. Following Caswell & Kennedy (1982), let us differentiate the graph $\delta(G)+1$ times with respect to external momenta. This produces a result that has negative degree of divergence. We differentiate not only the graph G, but also its various counterterm graphs $C_\gamma(G)$. The aim is to show that the result is actually convergent. To do this we will show that the differentiated counterterm graphs are the correct counterterm graphs for the differentiated original graph. This may sound obvious, but there are some subtleties, so we will give the details.

Let the external momenta of G be $p_1, \ldots, p_n$. Its renormalized value is

$$\begin{aligned} R(G) &= G + \sum_{\gamma \subseteq G} C_\gamma(G) \\ &= \bar{R}(G) + C(G) \\ &= \bar{R}(G) - T_G \circ \bar{R}(G). \end{aligned} \tag{5.8.1}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

Let ∂ denote differentiation with respect to one of the external momenta, and let ∂^λ denote any λ-fold differentiation with respect to the external momenta. The property to be shown is that $\partial^\lambda \bar{R}(G)$ is finite if $\lambda > \delta(G)$. (It is clear from naive power-counting that $\delta(\partial^\lambda G) = \delta(G) - \lambda$.)

Suppose we ensure that differentiation commutes with the basic subtraction operator T_γ. This amounts to imposing a very natural relation between the finite parts of, for example, $T_\gamma(\partial G)$ and $T_\gamma(G)$. (The relation is satisfied by the pole-part subtractions, but it is possible to choose exotic renormalization prescriptions not satisfying the hypothesis.) Then for any graph γ we have

$$\partial C(\gamma) = C(\partial\gamma). \tag{5.8.2}$$

It follows that, for the original graph, we have

$$\partial^\lambda \bar{R}(G) = \bar{R}(\partial^\lambda G). \tag{5.8.3}$$

The point here is that a differentiation when acting on a graph gives a number of terms, in each of which one of the propagators or vertices is differentiated. It is a simple generalization of the argument given in Sections 5.2.2 and 5.2.3 for specific graphs that the counterterms for subgraphs of G are the correct ones after differentiation in (5.8.3).

Now $\bar{R}(\partial^\lambda G)$ is the sum of a graph $\partial^\lambda G$ that has negative degree of divergence and the counterterm graphs for its subdivergences. Hence by Property (2) it is finite, so that we may choose the subtraction operator T to give zero. Therefore the counterterm in (5.8.1) for the undifferentiated graph is polynomial of degree $\delta(G)$ in the external momenta.

The same argument (Collins (1974)) also shows that counterterms are polynomials in mass. Here it is necessary to note that differentiation with respect to m^2 does not automatically reduce the degree of divergence. This only happens if counterterms for subdivergences are polynomial. If a counterterm has a piece proportional to $\ln(m^2)$ then differentiating with respect to m^2 leaves the degree of divergence unchanged. The proof merely demonstrates that it is always possible to choose counterterms to be polynomial in m^2; it is not compulsory.

5.8.2 $\bar{R}(G)$ *is finite if* $\delta(G)$ *is negative*

It was evident in one-loop examples that a graph with degree of divergence δ is renormalized by a local counterterm of degree δ in the external momenta. To generalize the result to an arbitrary graph, we constructed a renormalization procedure which involved computing the counterterm for

a 1PI graph only after subtracting subdivergences. We differentiated the graph $\delta + 1$ times with respect to its external momenta to prove its counterterm to be local and of degree δ. This proof relied on assuming the following statement:

> If a graph has negative degree of divergence and has its subdivergences subtracted according to the rules, then it is finite. More briefly, if $\delta(\Gamma) < 0$ then $\bar{R}(\Gamma)$ is finite.

This statement sounds extremely plausible. It is nevertheless in need of proof. We have to ensure that the subtraction procedure actually accomplishes its purpose of removing the subdivergences. (That is, there are no spurious divergences induced by the procedure.) In addition, we normally only consider the divergences as arising from regions in which some loop momenta go to infinity, all at the same rate; this generates the usual power-counting. It is necessary to eliminate more exotic possibilities.

The most important problem, which is the one we will examine, is to treat the case that a collection of loop momenta go to infinity, but at different rates. In Section 5.2.2, we examined the special case of Fig. 5.2.1. The general case is very similar. Inductively, we assume that properties (1) to (3), listed at the beginning of Section 5.8, are true for all smaller graphs than the graph G under consideration. We consider regions of the integration over loop momenta where all or some momenta go to infinity, not necessarily at the same rate. We will eliminate them as possible sources of additional divergences. If all the momenta go to infinity together, then the negative overall degree of divergence means that there is no actual divergence from this region.

If some momenta stay finite while the others go to infinity (not necessarily at the same rate), then let γ be the subgraph consisting of all those lines with the large momenta. Our inductive hypothesis ensures that all the resulting divergences are cancelled by counterterms for subgraphs.

The remaining case is that *all* of the loop momenta go to infinity, but again not at the same rate. Let k denote the components of the smallest momenta, and let l denote the rest. (Our notation is meant to copy that used for Fig. 5.2.1, and so is the proof.) Let γ be the subgraph consisting of all those lines carrying the loop momenta l. It may be a single 1PI graph or a disjoint union of 1PI subgraphs. Let these 1PI subgraphs be $\gamma_1, \ldots, \gamma_L$. Expand each subgraph in powers of its external momenta up to its degree of divergence. The remainder for each subgraph is really a graph with negative overall degree of divergence; the contribution vanishes as l goes to infinity, so we should have set $l = O(k)$. The expanded terms contribute just as they

did for Fig. 5.2.1. After subtraction of divergences we have a factor in the integral over k corresponding to the dimension of the subgraph.

We gloss over here some of the important details, notably what happens to the value of a general subgraph when some of its external momenta get large. But the main lines of the argument should be apparent.

The structure of the proof is the same as in subsection 5.2.2 and the result is the same.

5.8.3 *Asymptotic behavior*

Weinberg (1960) not only proved the convergence theorem stated above with more complete rigor, but he also investigated what happens when several of the external momenta $p_1, p_2, \ldots, p_n$ of a graph γ get large in the Euclidean region. They are assumed all to be of an order Q, with the ratios p_j^μ/Q fixed as $Q \to \infty$. None of the sums of subsets of p_j^μ/Q vanish. Weinberg then states how to find the asymptotic behavior:

(1) Consider any subgraph γ connected to all the lines carrying the large momenta. Let all the loop momenta of γ be of order Q. Compute the power of Q: Q^{a_γ}.

(2) Look at all such subgraphs. Let a_γ have a maximum value a.

(*a*) If there is a unique graph with this maximum power, then $\Gamma \propto Q^a$ as $Q \to \infty$.

(*b*) If there are several subgraphs with $a_\gamma = a$, then let N be the number of such subgraphs. The asymptotic behavior is:

$$\Gamma = Q^a[A_0 B_0 + A_1 B_1 \ln Q + A_2 B_2 (\ln Q)^2 + \cdots A_{N-1} B_{N-1} (\ln Q)^{N-1}] + O(Q^{a-1}). \tag{5.8.4}$$

Here the A_i's are functions of those momenta that are fixed as $Q \to \infty$ and the B_i's are functions of the finite quantities p_j^μ/Q.

This theorem is needed inductively in the guts of the convergence theorem proved in the last subsection. Its proof is similar.

It is not obvious that this part of Weinberg's theorem is of much use for physics, other than for its part in this convergence proof, since the asymptotic behavior is of Euclidean momenta. However, in the deep-inelastic scattering of a lepton on a hadron, there is a photon or a weak interaction boson that is far off-shell. The momentum carried by the boson is effectively Euclidean, and Weinberg's theorem applies. We will see this in Chapter 14. There are also generalizations to other intrinsically Minkowskian situations (e.g. Amati, Petronzio & Veneziano (1978), Ellis *et*

al. (1979), Libby & Sterman (1978), Mueller (1978, 1981), Stirling (1978), and Buras (1981)). These are beyond the scope of the present book.

5.9 Oversubtractions

We showed how to renormalize a Feynman graph by making subtractions for the divergent subgraphs and for the overall divergence of the graph. It is possible, however, to make subtractions on graphs that are not divergent. Subtractions can also be made with a higher degree polynomial in the external momenta than called for by the degree of divergence. Either of these cases is called oversubtraction. Now, the general form of the renormalization, either by the recursive method or by the forest formula, did not specify the exact form of the subtraction operator T. So oversubtractions can be made without changing the general formalism.

There are two important uses for oversubtractions. The first is when we wish to use 'physical values' of masses or couplings as the renormalized parameters. We will discuss this in a moment. The second use is to construct operator product expansions. There, subtractions are made not only to cancel UV divergences but also to extract asymptotic behavior as some external momenta get large. We will discuss this later in Chapter 10.

5.9.1 *Mass-shell renormalization and oversubtraction*

We have considered renormalization as the procedure of removing divergences. Another point of view comes from the observation that one cannot observe directly the mass and coupling parameters that appear as coefficients in the Lagrangian. For example, consider a theory where each field has a corresponding single-particle state. Then the masses that are measured are those of the single particles, and it is often sensible to parametrize the theory in terms of these masses. Similar remarks can be applied to couplings. (Thus in QED one normally parametrizes the theory by the electron's mass and charge, defined by the long-range part of its electric field.) It can also be convenient to rescale the fields so that each propagator has a pole of unit residue.

In a simple renormalizable theory like ϕ^3 in six dimensions the renormalizations to accomplish such a mass-shell parametrization are precisely those necessary to cancel the UV divergences. Thus we may define the subtraction operator applied to a self-energy graph $\Sigma(p^2)$ to be

$$T_{(\mathrm{ph})} \circ \Sigma(p^2) = \Sigma(m_{\mathrm{ph}}^2) + (p^2 - m_{\mathrm{ph}}^2)\Sigma'(m_{\mathrm{ph}}^2), \tag{5.9.1}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

so that the inverse propagator satisfies

$$-\mathrm{i}[p^2 - m_{\mathrm{ph}}^2 - \Sigma_{\mathrm{ph}}(p^2)] = -\mathrm{i}(p^2 - m_{\mathrm{ph}}^2) + O(p^2 - m_{\mathrm{ph}}^2)^2, \quad (5.9.2)$$

as $p^2 \to m_{\mathrm{ph}}^2$. We use the subscript 'ph' to indicate renormalization according to the mass-shell scheme.

Of course, mass-shell renormalization is only one out of many renormalization prescriptions that we may use to cancel UV divergences. But we may also choose to renormalize in the absence of divergences. Consider, as an example, ϕ^3 theory again, but now in four dimensions. We may continue to use (5.9.1) and (5.9.2) for the renormalization of the propagator so we have a 'physical' parametrization. But all except the one-loop self-energy graph have no divergence, so all the wave-function counterterms are finite and all but one of the mass counterterms are finite. The combinatorics of the renormalization procedure as described earlier all work unchanged.

5.9.2 *Remarks*

One important technical problem is to check that the oversubtracted and the normally subtracted theories differ only by a reparametrization. This can be done by the methods which we will describe in Sections 7.1 and 7.2.

In the previous subsection 5.9.1, we took the point of view that renormalization is the process of reparametrizing the theory in terms of 'physical' quantities. It should be noted that this is not always a useful point of view. In the first place, other renormalization prescriptions are more convenient for handling certain types of calculation. In the second place, there may be infra-red divergences that make the mass-shell structure of a theory not what one would naively expect: thus in QED the electron's propagator does not have a simple pole. And, finally, in some theories there are many more particles and couplings than independent parameters. This is very common in gauge theories.

5.9.3 *Oversubtraction on IPR graphs*

The aim of oversubtraction, generally, is to impose some condition on Green's functions. So far, we have assumed the condition to be imposed on the 1PI graphs, since those are the ones needing counterterms for divergences. However, consider $\phi^3 + \phi^4$ theory:

$$\mathscr{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - f\phi^3/6 - g\phi^4/24 + \text{counterterms}. \quad (5.9.3)$$

Let us choose to renormalize at zero external momentum. Thus the self-

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

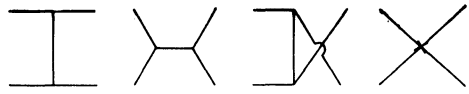

Fig. 5.9.1. Subtraction of one-particle-reducible subgraphs.

energy Σ and the three-point 1PI function $\Gamma_{(3)}$ satisfy

$$\Sigma(p^2 = 0) = \frac{d\Sigma}{dp^2}(p^2 = 0) = 0,$$

$$\Gamma_{(3)}(p_1^2 = p_2^2 = p_3^2 = 0) = \text{lowest order} = -if. \tag{5.9.4}$$

Following from our earlier work we might renormalize the four-point function $\Gamma_{(4)}$ by requiring the sum of the 1PI graphs to be equal to their lowest order value at zero external momentum. However it is also sensible to impose instead the condition on the amputated four-point function $\Gamma_{(4)}^{(a)}$. These graphs are 1PI only in the four external lines. (We should amputate the graphs since the counterterm vertex will have attached to it external propagators.) Thus in addition to the three tree graphs of Fig. 5.9.1 (*a*)–(*c*), we require the counterterm, Fig. 5.9.1(*d*):

Our general method of renormalization tells us that whenever we have a basic graph containing one of the graphs (*a*), (*b*), or (*c*) in Fig. 5.9.1 as a subgraph, there will be counterterm graphs in which this subgraph is replaced by the counterterm vertex (*d*). These counterterm graphs may be divergent even when the basic graph is finite. An example is shown in Fig. 5.9.2. In Fig. 5.9.2(*a*) if we impose the renormalization condition on the 1PI functions only the graph (*b*) occurs as counterterm; (*a*) plus (*b*) is finite. If we impose the condition on amputated graphs we immediately meet graph (*c*) where the line A is replaced by its 1/3 share of the counterterm Fig. 5.9.1(*d*).

But graph (*c*) has a divergence, so we must renormalize it by a three-point counterterm to the subgraph consisting of the line A and the loop B. This is

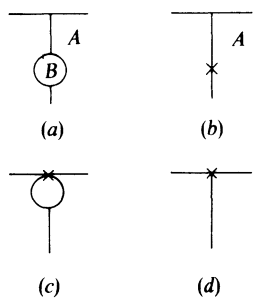

Fig. 5.9.2. The subtractions of Fig. 5.9.1, inside a bigger graph.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

shown in graph (*d*). Note that the graph consisting of line *A* and loop *B* has a subdivergence, but no overall divergence. Even so, the overall counterterm in (*d*) is divergent.

It is not difficult to see that the extra counterterms needed to impose the renormalization condition on the 1PR amputated graphs do not change our results on renormalization. The instructions for renormalization in Sections 5.3 and 5.5 can be used provided only that we replace the term '1PI subgraph' by 'amputated subgraph'.

The use of subtractions on 1PR graphs is too baroque for normal use. However it is a device that is useful for discussing the large mass expansion and the operator-product expansion (Chapters 8 and 10).

5.10 Renormalization without regulators: the BPHZ scheme

In setting up the renormalization procedure in Sections 5.3 and 5.5 we were careful not to use a specific definition of the subtraction operation. This was to allow for the choice of one out of the infinitely many possible renormalization prescriptions. An obvious one is the mass-shell subtraction procedure indicated in the last section. Another is the minimal subtraction procedure to be defined precisely in Section 5.11; we have already made much use of it. In this section we will explain the method of Zimmermann (1969), in which the subtractions are applied directly to the Feynman integrand, so that no regulator need be used.

The starting point is the method due to Bogoliubov & Parasiuk (1957) and Hepp (1966), called the BPH scheme. They observed that the overall counterterm for a graph Γ is a polynomial of degree $\delta(\Gamma)$, its degree of divergence. So they defined the subtraction operator $T(\Gamma)$ to be the terms up to order $\delta(\Gamma)$ in the Taylor expansion of Γ about zero external momentum. For example, consider the one-loop self-energy graph Fig. 3.1.1 in ϕ^3 theory in six dimensions. After dimensional regularization its unrenormalized value is

$$\Sigma_a(p^2,d)=\frac{-g^2}{2(4\pi)^{d/2}}\Gamma(2-d/2)\int_0^1 \mathrm{d}x\,[m^2-p^2x(1-x)]^{d/2-2}. \qquad (5.10.1)$$

The terms up to order p^2 in its Taylor expansion about $p=0$ are

$$T\circ\Sigma_a=\frac{-g^2}{2(4\pi)^{d/2}}\Gamma(2-d/2)\times$$
$$\times\int_0^1 \mathrm{d}x\,m^{d-4}[1-(d/2-2)p^2x(1-x)/m^2]. \qquad (5.10.2)$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

The renormalized value of the graph is $\Gamma_a - T^\circ\Gamma_a$. So at $d=6$ this is

$$-\frac{g^2}{128\pi^3}\int_0^1 dx\{[m^2-p^2x(1-x)]\ln[1-p^2x(1-x)/m^2]+p^2x(1-x)\}. \tag{5.10.3}$$

Zimmermann's (1969) achievement was to realize that this construction can be applied directly to the integrand. Subdivergences are subtracted with the aid of his forest formula. Then the result is an integral which, according to power-counting, has no UV divergences. The integral therefore has in fact no divergences (Hahn & Zimmermann (1968), Zimmermann (1968)). This method is called BPHZ renormalization.

In Section 3.4, we applied this method to the above graph, with the result (3.4.7). It can be explicitly calculated by putting all the terms over a common denominator and then using standard parametric methods. The result agrees with (5.10.3).

An example involving a subdivergence is given by Fig. 5.10.1 for ϕ^4 in four dimensions. Let the renormalized integral be

$$\Gamma_{\text{BPHZ}} = \frac{ig^3}{2(2\pi)^8}\int d^4k d^4l I(p_1,p_2,p_3,p_4,k,l). \tag{5.10.4}$$

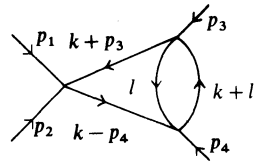

Fig. 5.10.1. A two-loop vertex graph in ϕ^4 theory.

Then we will construct the integrand I.

The unrenormalized integrand is

$$U = \frac{1}{(l^2-m^2)}\frac{1}{[(k+l)^2-m^2]}\frac{1}{[(k+p_3)^2-m^2]}\frac{1}{[(k-p_4)^2-m^2]} \tag{5.10.5}$$

Subtraction of the sole subdivergence gives

$$\begin{aligned}\bar{R}(U) &= U - \frac{1}{(l^2-m^2)^2}\frac{1}{[(k+p_3)^2-m^2]}\frac{1}{[(k-p_4)^2-m^2]} \\ &= \frac{-2k\cdot l - k^2}{(l^2-m^2)^2[(k+l)^2-m^2][(k+p_3)^2-m^2][(k-p_4)^2-m^2]}.\end{aligned} \tag{5.10.6}$$

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press

Then the overall divergence is subtracted to give

$$I = R(U) = \bar{R}(U) - [R(U)]|_{p_1 = p_2 = p_3 = p_4 = 0}$$
$$= \frac{(2k\cdot l + k^2)\{(k^2 - m^2)[p_4^2 + p_3^2 + 2k\cdot(p_3 - p_4)] + (p_3^2 + 2k\cdot p_3)(p_4^2 - 2k\cdot p_4)\}}{(l^2 - m^2)^2(k^2 - m^2)^2[(k+l)^2 - m^2][(k+p_3)^2 - m^2][(k-p_4)^2 - m^2]}. \tag{5.10.7}$$

The BPHZ scheme has a number of advantages:

(1) It is applied to the integrand and generates a convergent integral without requiring any regularization.
(2) Thus it exhibits the fact that the properties of a renormalized field theory do not depend on which UV regulator is used.
(3) Mathematically it is rather elegant. In particular there is no need to discuss directly the divergences of Feynman graphs; it is only required to have a theorem that tells us that a graph that is convergent according to the naive criteria is actually convergent.
(4) It allows a very simple proof of the operator-product expansion.

There are a number of disadvantages:

(1) It is not the best scheme for theories (especially gauge theories) with complicated symmetries, where relations between counterterms have to be preserved; the scheme does not allow direct computation of the value of a divergence.
(2) The subtractions are made at zero momentum and therefore are infrared divergent in a massless theory.
(3) When the scheme is generalized to handle massless theories, it becomes much more complicated (Lowenstein, Weinstein & Zimmermann, 1974a, b).

5.11 Minimal subtraction

5.11.1 *Definition*

It can be proved (Speer (1974) and Breitenlohner & Maison (1977a, b, c)) that, when dimensional regularization is used, the UV divergences of Feynman graphs appear as poles at isolated values of the space-time dimension d. Minimal subtraction ('t Hooft (1973)) – the MS scheme – consists of defining the counterterms to be poles at the physical value of d, $d = 4$. We have already used this scheme, in Chapter 3. Our purpose in this section is to make precise the definition of minimal subtraction.

The main complication is that bare couplings have a dimension that depends on d, so that we must introduce the unit of mass μ, as follows:

(1) Consider in turn the coefficient $g_i + \delta g_i$ of each term in $\mathscr{L}$. Let the dimension of g_i be $a_i + b_i(4-d)$. Then we replace $g_i + \delta g_i$ by $\mu^{b_i(4-d)}(g_i + \delta g_i)$. Thus the renormalized coupling g_i and the counterterm δg_i both have dimension a_i, independently of d.

(2) Let Γ be a 1PI graph to which it is desired to apply a subtraction operator T. Let the dimension of Γ be $A + B(d-4)$, and suppose the couplings all contain powers of μ as just explained. Then we define

$$T(\Gamma) = \mu^{B(d-4)}\{\text{pole part of } (\mu^{B(4-d)}\Gamma) \text{ at } d = 4\}.$$

The pole part is obtained by making a Laurent expansion about $d = 4$. We have arranged to take the pole part of a function whose dimension does not depend on d.

(3) Suppose we are talking about a theory in a different number of physical dimensions than four. For example, we might be in ϕ^3 theory in six dimensions. Then the '4' in the above formulae is replaced by the correct physical value.

For a simple graph with no subdivergences, like the one-loop self-energy in (5.10.1), this prescription amounts to subtracting the pole:

$$\begin{aligned}\Sigma_a^{(\overline{\text{MS}})}(d=6) = \lim_{d\to 6}\Bigg\{ & \frac{-g^2\mu^{6-d}}{2(4\pi)^{d/2}}\Gamma(2-d/2)\int_0^1 \mathrm{d}x[m^2 - p^2x(1-x)]^{d/2-2} \\ & - \left[\text{pole} = \frac{-g^2}{128\pi^3}\frac{1}{d/2-3}(m^2 - \tfrac{1}{6}p^2)\right]\Bigg\} \\ = \frac{-g^2}{128\pi^3}\Bigg\{ & [\gamma_{\mathrm{E}} - 1 - \ln(4\pi)](m^2 - \tfrac{1}{6}p^2) \\ & + \int_0^1 \mathrm{d}x[m^2 - p^2x(1-x)]\ln\left[\frac{m^2 - p^2x(1-x)}{\mu^2}\right]\Bigg\}. \qquad (5.11.1)\end{aligned}$$

For graphs with subdivergences, the subdivergences must of course be subtracted before removing the overall pole.

The advantages of the scheme are:

(1) It automatically preserves complicated symmetries. The exceptions are chiral symmetries and the like, which in general cannot be preserved by quantization – see Chapter 13.

(2) It has no problems with massless theories. In fact, dimensional continuation regulates both IR and UV divergences, thus removing the need for a separate IR cut-off.

(3) Calculations are very convenient.

(4) Computation of the divergent part of a Feynman graph – needed for

renormalization-group calculations – is almost trivial at the one-loop level.

Some disadvantages of minimal subtraction are:

(1) It is unphysical.
(2) The proof of the operator-product expansion is made harder than in the BPHZ scheme.

5.11.2 $\overline{MS}$ renormalization

The MS scheme has found much use especially in work on QCD, where it has become standard. Another disadvantage that then appears is that minimal subtraction tends to produce large coefficients in the perturbation expansion. These are primarily due to the $\ln(4\pi) - \gamma_E \sim 1.95$ term such as appears in (5.11.1). It has become conventional to work with a modified scheme, called the $\overline{\text{MS}}$ scheme (Bardeen, Buras, Duke & Muta (1978)). Here the μ of the MS scheme is written as

$$\mu = \bar{\mu}\left(\frac{e^{\gamma_E}}{4\pi}\right)^{1/2} \simeq 0{\cdot}38\bar{\mu}. \tag{5.11.2}$$

Then we have, instead of (5.11.1), the cleaner form

$$\Sigma_a^{(\overline{\text{MS}})} = \frac{-g^2}{128\pi^3}\left\{\tfrac{1}{6}p^2 - m^2 + \int_0^1 \mathrm{d}x[m^2 - p^2x(1-x)]\ln\left[\frac{m^2 - p^2x(1-x)}{\bar{\mu}^2}\right]\right\}. \tag{5.11.3}$$

5.11.3 Minimal subtraction with other regulators

Minimal subtraction could also be applied with other UV cut-offs. For example, if a lattice of spacing a is used, then the singular $a \to 0$ behavior of graph of degree of divergence δ is

$$a^{-\delta}[\text{polynomial in } \ln(a)].$$

One can therefore define $T(\Gamma)$ as the singular part of Γ, with the general form

$$T(\Gamma) = \sum_{\beta=1}^{\beta_{\max}} [\ln(a\mu)]^\beta A_{0,\beta} + \sum_{\alpha=1}^{\delta} \sum_{\beta=0}^{\beta_{\max}} \frac{1}{a^\alpha}[\ln(a\mu)]^\beta A_{\alpha,\beta}. \tag{5.11.4}$$

After subtraction of subdivergences, the coefficients $A_{\alpha,\beta}$ are polynomials in masses and momenta. Note again the appearance of a unit of mass. This scheme has found little use.

https://doi.org/10.1017/9781009401807.005 Published online by Cambridge University Press