Nuclear Physics B231 (1984) 269–295

RENORMALIZATION AND EFFECTIVE LAGRANGIANS

Joseph POLCHINSKI*

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

Received 27 April 1983

There is a strong intuitive understanding of renormalization, due to Wilson, in terms of the scaling of effective lagrangians. We show that this can be made the basis for a proof of perturbative renormalization. We first study renormalizability in the language of renormalization group flows for a toy renormalization group equation. We then derive an exact renormalization group equation for a four-dimensional $\lambda\phi^4$ theory with a momentum cutoff. We organize the cutoff dependence of the effective lagrangian into relevant and irrelevant parts, and derive a linear equation for the irrelevant part. A lengthy but straightforward argument establishes that the piece identified as irrelevant actually is so in perturbation theory. This implies renormalizability. The method extends immediately to any system in which a momentum-space cutoff can be used, but the principle is more general and should apply for any physical cutoff. Neither Weinberg's theorem nor arguments based on the topology of graphs are needed.

1. Introduction

The understanding of renormalization has advanced greatly in the past two decades. Originally it was just a means of removing infinities from perturbative calculations. The question of why nature should be described by a renormalizable theory was not addressed. These were simply the only theories in which calculations could be done.

A great improvement comes when one takes seriously the idea of a physical cutoff at a very large energy scale Λ. The theory at energies above Λ could be another field theory, a lattice, spacetime foam, or anything else. The theory just below Λ should be represented by a very general lagrangian in which the various terms have coefficients of the order of Λ to the appropriate power to make the dimensions correct. Consider the physics at an energy E far below Λ. The non-renormalizable terms, those with coefficients of Λ to negative powers, typically give contributions that are suppressed by powers of Λ. This is true unless the non-renormalizable term is embedded in a Feynman graph sufficiently divergent to make up for the small coefficient. Power counting shows that the only n-point functions sufficiently divergent are those which would be divergent even if they contained only renormal-

* Research was supported in part by the National Science Foundation under grant no. PHY82-15249.

izable interactions. We know, however, that the latter divergences can be reabsorbed in redefinitions of the renormalizable couplings. Thus, to accuracy E/Λ the entire effect of the non-renormalizable terms can be absorbed in the initial values of the renormalizable ones, and all quantities can be calculated in the resulting effective field theory with renormalizable interactions only.

This idea is very nice, but still perturbative and graphical in nature. Also, it does not attempt to address the question of why arguments based on naive power counting are correct, that is, why renormalization actually works. A further improvement comes with the idea of smoothly lowering the cutoff. As this is done, the effective lagrangian changes. The effective lagrangian at lower scales is given in terms of its form at a given scale, and its change with scale is governed by a scaling or renormalization group equation. Typically, as we scale down to smaller momenta, the lagrangian converges toward a finite-dimensional submanifold in the space of possible lagrangians. That is, the scaling transformation has only a finite number of non-negative eigenvalues, with deviations in the orthogonal directions damped as we travel to low momenta. These orthogonal directions are therefore termed "irrelevant". In the zero-coupling limit, the negative, irrelevant, eigenvalues correspond to precisely the non-renormalizable interactions. Since there is nothing discontinuous about the scaling transformation as the couplings are changed, those eigenvalues which were negative at zero coupling should remain negative at sufficiently small coupling. This is equivalent to renormalizability, a connection which will be developed further in sect. 2. This understanding of renormalization is due primarily to Wilson [1]. Ref. [2] lists an assortment of discussions from related points of view.

The classic proofs of renormalization in perturbation theory [3–5] are based on the old idea of removing infinities. They involve detailed graphical arguments and convergence theorems [6] that are rather far removed from the present intuitive picture*. The understanding in terms of renormalization group flows is so compelling that one must wonder whether it can be justified only by appealing to the existing proofs. In fact, we will find that this is not so. Once we learn to discuss renormalization in the language of renormalization group flows, a proof follows in a straightforward way. We are concerned here only with perturbation theory, but the proof follows the outline we would expect for a non-perturbative argument.

In sect. 2 we study a toy renormalization group equation, showing how to divide the effective action into relevant** and irrelevant parts. We show that the toy equation describes a renormalizable theory. In sect. 3 we make concrete the idea of a

* We should mention in particular the proof of Callan and of Blaer and Young [5], which uses the renormalization group to simplify greatly the graphical arguments. This is still essentially a graphical proof, as it is based on a skeleton expansion, and is global rather than local in momenta, as it uses Weinberg's theorem [6].

** In standard usage what we call "relevant" is usually divided further into "relevant" and "marginal". Throughout this paper, "relevant" should be understood to mean "relevant or marginal".

differential scaling of a cutoff, obtaining an exact renormalization group equation for a $\lambda\phi^4$ theory with a momentum-space cutoff. Guided by the results of sect. 2 we organize the action into relevant and irrelevant parts. Very simple bounds on momentum integrals are sufficient to prove that in perturbation theory, the terms identified as irrelevant really are. The detailed order-by-order argument is given in sect. 4. It is then a short step to show that correlation functions depend on the cutoff only through inverse powers.

The method extends automatically to any theory in which a momentum cutoff can be used, and to composite operator renormalization. The idea is far more general, however, and should apply for any physical cutoff. In the conclusions, we discuss the extension to gauge theories and beyond perturbation theory.

The exact renormalization group equation for $\lambda\phi^4$, as well as the understanding of renormalization in terms of relevant and irrelevant operators, can all be found in the work of Wilson [1]. Our contribution here is to note that these ideas can be used to give a self-contained proof of renormalization in perturbation theory.

It might cause some confusion that the $\lambda\phi^4$ theory which we will be studying does not actually have a continuum limit [7], in the sense that the bare coupling $\lambda^0(\Lambda_0)$ diverges when the cutoff Λ_0 is still finite. The point is that $\lambda^0(\Lambda_0)$ considered as a formal power series in the renormalized coupling λ^{R} is perfectly well defined for *all* finite Λ_0, as are all other quantities, and our results are concerned strictly with this expansion. It should be emphasized that two distinct properties of the renormalization group flow are involved here. One is purely local: is the flow converging in all but a finite number of directions? This is the crux of our present work. The second concerns global properties of the flow: is there a starting point, for a given Λ_0, which gives the desired value of λ^{R} at some fixed lower scale? The second question is an interesting one, but the answer is always "yes" in perturbation theory.

2. Renormalization group flows

In this section we study a toy renormalization group equation which describes one relevant and one irrelevant coupling. This section is not needed for the proof in sect. 3, but is intended to illustrate the ideas in a simpler setting. The two couplings are g_4, which is dimensionless, and g_6, which has dimensions of $(\text{mass})^{-2}$. These have been chosen to correspond dimensionally to a 4-point and a 6-point function in the effective action of a 4-dimensional scalar field theory. These couplings depend on the scale Λ according to

$$\Lambda \frac{\mathrm{d}g_4}{\mathrm{d}\Lambda} = \beta_4(g_4, \Lambda^2 g_6), \tag{1a}$$

$$\Lambda \frac{\mathrm{d}g_6}{\mathrm{d}\Lambda} = \Lambda^{-2}\beta_6(g_4, \Lambda^2 g_6), \tag{1b}$$

where factors of Λ have been inserted by dimensional analysis. Define the dimensionless variables $\lambda_4 = g_4$, $\lambda_6 = \Lambda^2 g_6$, so that

$$\Lambda \frac{d\lambda_4}{d\Lambda} = \beta_4(\lambda_4, \lambda_6), \tag{2a}$$

$$\Lambda \frac{d\lambda_6}{d\Lambda} - 2\lambda_6 = \beta_6(\lambda_4, \lambda_6). \tag{2b}$$

Take a particular solution of (2), $(\bar{\lambda}_4, \bar{\lambda}_6)$, and consider small deviations $\varepsilon_i = \lambda_i - \bar{\lambda}_i$. Then, to order ε,

$$\Lambda \frac{d\varepsilon_4}{d\Lambda} = \overline{\frac{\partial \beta_4}{\partial \lambda_4}} \varepsilon_4 + \overline{\frac{\partial \beta_4}{\partial \lambda_6}} \varepsilon_6, \tag{3a}$$

$$\Lambda \frac{d\varepsilon_6}{d\Lambda} - 2\varepsilon_6 = \overline{\frac{\partial \beta_6}{\partial \lambda_4}} \varepsilon_4 + \overline{\frac{\partial \beta_6}{\partial \lambda_6}} \varepsilon_6, \tag{3b}$$

where the bar means the quantity is evaluated at $(\bar{\lambda}_4, \bar{\lambda}_6)$. The term $-2\varepsilon_6$ in (3b) suggests that deviations in λ_6 are strongly damped (by $O(\Lambda^2/\Lambda_0^2)$) as we evolve from a scale Λ_0 to a much lower scale Λ. However, before a deviation in λ_6 is damped away, it will, through (3a), cause λ_4 to run a little faster or slower than it otherwise would have done. (Compare the discussion in the second paragraph of the introduction of the low-energy effects of non-renormalizable terms: λ_6 enters only through the effective value of λ_4). Fig. 1 shows the situation in the $\lambda_4 - \lambda_6$ plane. A_1 and A_2, initially separated by a small amount in the λ_6 direction, have evolved to B_1 and B_2. The point B_2 is ahead and remains so as we move along the trajectories. However, there are other points on the B_2 trajectory, such as B_2', which are much closer to B_1. To expose this we write, again to order ε,

$$\xi_4 = \varepsilon_4 - \frac{d\bar{\lambda}_4}{d\Lambda}\left(\varepsilon_4 \Big/ \frac{d\bar{\lambda}_4}{d\Lambda}\right) = 0, \tag{4a}$$

$$\xi_6 = \varepsilon_6 - \frac{d\bar{\lambda}_6}{d\Lambda}\left(\varepsilon_4 \Big/ \frac{d\bar{\lambda}_4}{d\Lambda}\right), \tag{4b}$$

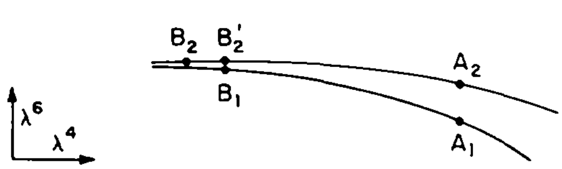


Fig. 1. Neighboring trajectories in the $\lambda_4 - \lambda_6$ plane.

so that (ξ_4, ξ_6) is the vector from B_1 to the point vertically above on the B_2 trajectory. (This is a convenient definition as long as $\bar{\beta}_4 \neq 0$, so that the trajectories do not turn vertical in the $\lambda_4 - \lambda_6$ plane.) Thus we find

$$\Lambda \frac{\mathrm{d}\xi_6}{\mathrm{d}\Lambda} - 2\xi_6 = \left\{ \overline{\frac{\partial\beta_6}{\partial\lambda_6}} + \overline{\frac{\partial\beta_4}{\partial\lambda_4}} - \Lambda \frac{\mathrm{d}}{\mathrm{d}\Lambda} \ln \bar{\beta}_4 \right\} \xi_6, \tag{5}$$

which integrates to

$$\xi_6(\Lambda) = \xi_6(\Lambda_0)\left(\frac{\Lambda^2}{\Lambda_0^2}\right)\left(\frac{\bar{\beta}_4(\Lambda_0)}{\bar{\beta}_4(\Lambda)}\right)\exp\int_{\Lambda_0}^{\Lambda} \frac{\mathrm{d}\Lambda'}{\Lambda'}\left(\overline{\frac{\partial\beta_6}{\partial\lambda_6}}(\Lambda') + \overline{\frac{\partial\beta_4}{\partial\lambda_4}}(\Lambda')\right). \tag{6}$$

As long as we remain at couplings sufficiently small that the integrand in (6) remains small and that the ratio of $\bar{\beta}_4$'s runs sufficiently slowly*, the behavior of ξ_6 at small Λ is dominated by Λ^2/Λ_0^2. Thus one would say that ξ_6 is an irrelevant parameter. Two nearby trajectories approach each other strongly in the infrared, the separation going as a power of Λ/Λ_0.

Fig. 2 shows the trajectories in the $\lambda_4 - \lambda_6$ plane. As the initial conditions vary in a two-dimensional space, the theory in the infrared lies very near a one-dimensional subspace: as soon as we know λ_4, we know λ_6, to accuracy Λ^2/Λ_0^2. Actually, even this one relevant parameter just marks where we are along a single trajectory: it is transmuted into a scale.

Now let us relate this to the usual language of renormalization. Imagine starting with a bare theory at a scale Λ_0, with λ_4 set to a particular value and λ_6 to zero; this is the point C_1 in fig. 2. At a much lower scale Λ_R, we are at D_1, at which λ_4 is defined to have the value λ_4^R. Now consider a larger cutoff Λ_0'. We can find another

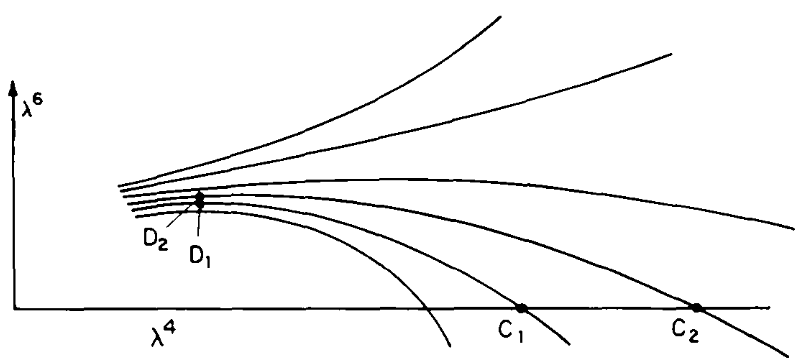


Fig. 2. Convergence of trajectories in the $\lambda_4 - \lambda_6$ plane.

* This is just the requirement that the anomalous dimensions do not overwhelm the canonical ones.

point C_2, lying on a longer trajectory, which at Λ_R arrives at D_2 where again $\lambda_4 = \lambda_4^R$ *. We can proceed in this way, thus defining the bare coupling λ_4^0 as a function of λ_4^R, Λ_R, and Λ_0. Now take $\Lambda_0 \to \infty$ holding Λ_R and λ_4^R fixed. We know that λ_6 depends on Λ_0 only as Λ_R^2/Λ_0^2, so it approaches a limit. But this is just what we mean by renormalizability: as we take the cutoff to infinity, holding the renormalized couplings fixed, all other quantities in theory (in this case λ_6 is all there is) approach limits as inverse powers of the cutoff. Thus, renormalizability follows in a very general way when dimensional analysis is applied to the renormalization group equation for an effective lagrangian.

Two points should be made clear. First, it is not being said that λ_6 goes to zero in the infrared – merely that its value is determined in terms of λ_4. Second, there is nothing special about the λ_4-axis in fig. 2, it simply corresponds to the way that calculations are usually done, with only the relevant bare couplings non-zero. We could use another curve in place of the λ_4-axis to define our bare theory, corresponding to bare non-renormalizable terms with coefficients characterized by Λ_0 to negative powers, and the result would be the same. The one thing we *cannot* do is to take Λ_0 to infinity while keeping bare non-renormalizable terms with coefficients of inverse powers of some smaller scale.

Let us outline, in the context of eq. (2), the strategy to be followed in sect. 3 for scalar field theory. Starting from initial conditions $\lambda_4 = \lambda_4^0$, $\lambda_6 = 0$ at Λ_0, eq. (2) defines the functions $\lambda_i(\Lambda, \Lambda_0, \lambda_4^0)$. The vector

$$\Lambda_0 \frac{\partial}{\partial \Lambda_0} \lambda_i(\Lambda, \Lambda_0, \lambda_4^0), \tag{7}$$

satisfies the linearized eq. (3), because $\Lambda_0 \partial/\partial\Lambda_0$ commutes with $\Lambda\partial/\partial\Lambda$. Thus, as with ε_i before, it should be nearly parallel in the infrared to the relevant trajectory. We want to subtract off the relevant part, the part parallel to the trajectory, but will do this differently from before. Rather than subtract off a multiple of the tangent vector to the trajectory, we will subtract off an appropriate (Λ-dependent) amount of

$$\frac{\partial}{\partial \lambda_4^0} \lambda_i(\Lambda, \Lambda_0, \lambda_4^0). \tag{8}$$

The vector (8) also satisfies the linear eq. (3), so it too should be nearly parallel to the trajectory in the infrared. Define, then,

$$v_i(\Lambda) = \Lambda_0 \frac{\partial \lambda_i(\Lambda)}{\partial \Lambda_0} - \frac{\partial \lambda_i(\Lambda)}{\partial \lambda_4^0} \left(\frac{\partial \lambda_4(\Lambda)}{\partial \lambda_4^0} \right)^{-1} \Lambda_0 \frac{\partial \lambda_4(\Lambda)}{\partial \Lambda_0}. \tag{9}$$

* In fact, it is not always possible to do this, sometimes λ_4^0 diverges for finite Λ_0. However, in perturbation theory, our eventual interest, λ_4^0 always exists as a power series in λ_4^R (see the end of sect. 1).

so that $v_4(\Lambda)=0$. $v_i(\Lambda)$ satisfies a linear equation similar to (5), and we can again conclude that it is driven small in the infrared. The point of the definition (9) is that

$$v_i(\Lambda)=\Lambda_0\frac{\mathrm{d}}{\mathrm{d}\Lambda_0}\lambda_i\big(\Lambda,\Lambda_0,\lambda_4^0(\Lambda,\lambda_4,\Lambda_0)\big). \tag{10}$$

That is, it is the total derivative of $\lambda_i(\Lambda)$ holding $\lambda_4(\Lambda)$ fixed. For $\Lambda=\Lambda_R$ it is the derivative of the point D in fig. 2 as the bare cutoff is changed. Once we conclude that $v_i(\Lambda_R)$ is of order Λ_R^2/Λ_0^2 times slowly varying functions, eq. (10) may be integrated to conclude that $\lambda_i(\Lambda_R,\Lambda_0,\lambda_4^0(\Lambda_R,\lambda_4,\Lambda_0))$ has a limit as the cutoff is taken to infinity. This is the usual statement of renormalizability. The quantity to be studied in the next two sections is the analog of $v_i(\Lambda)$.

There are many more useful exercises with these toy equations. Eq. (1) may be solved order by order in g_4^0. Divergences appear, just as in field theory, which cancel magically in renormalized quantities (those expressed in terms of g_4, not g_4^0). Of course, from the point of view of fig. 2, this is not mysterious at all. One may prove this to all orders, starting from the linearized equation for $v_i(\Lambda)$, and one is led to the same steps as will be taken in sect. 4. Finally, additional relevant and irrelevant couplings may be added, with similar conclusions.

Let us summarize the relation between the flow of the effective lagrangian and renormalizability. Suppose it is known that the effective lagrangian at low scales is strongly attracted towards an n-dimensional submanifold, where n is the number of renormalizable couplings. Consider the $n+1$ parameter theory defined by the values of the n bare couplings and the bare cutoff Λ^0. Then in general each point on the n-dimensional submanifold is the image of a one-dimensional curve in the set of bare theories. It is then possible (modulo the global questions discussed at the end of sect. 1) to take Λ^0 to infinity along such a curve, with the physics at a given scale thus remaining fixed to accuracy of inverse powers of Λ^0.

3. Scalar field theory

We will consider a scalar field theory in four euclidean dimensions, with a momentum space cutoff. The propagator is

$$(p^2+m^2)^{-1}K(p^2/\Lambda_0^2). \tag{11}$$

Here $K(p^2/\Lambda^2)$ is a general cutoff function which we will take to have the value 1 for $p^2<\Lambda^2$ and to vanish rapidly at infinity. We define the theory by the propagator (11) and vertices given by the interaction lagrangian

$$L_{\text{int}}=\int \mathrm{d}^4x\left(-\tfrac{1}{2}\rho_1^0\phi^2(x)-\tfrac{1}{2}\rho_2^0\big(\partial_\mu\phi(x)\big)^2-\frac{1}{4!}\rho_3^0\phi^4(x)\right). \tag{12}$$

The bare couplings, which we label ρ_a^0, are usually called

$$\rho_1^0 = \delta m^2, \qquad \rho_2^0 = Z - 1, \qquad \rho_3^0 = \lambda^0. \tag{13}$$

The system may be represented by a functional integral

$$Z(J) = \int \mathrm{d}\phi \exp\left(\int \frac{\mathrm{d}^4 p}{(2\pi)^4} \left[-\tfrac{1}{2}\phi(p)\phi(-p)(p^2+m^2)K^{-1}(p^2/\Lambda_0^2) + J(p)\phi(-p) \right] + L_{\mathrm{int}}(\phi) \right), \tag{14}$$

where $J(p)$ is the external source used to obtain the n-point functions.

We wish to integrate out the high-momentum components of ϕ, so that we reduce the cutoff in (11) to a much lower scale Λ_{R}. We will take

$$|m^2| < \Lambda_{\mathrm{R}}, \tag{15a}$$

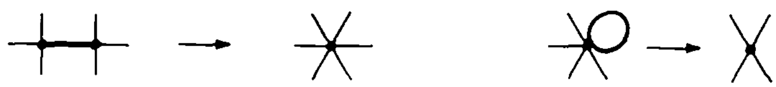
$$J(p) = 0, \qquad \text{for } p^2 > \Lambda_{\mathrm{R}}^2. \tag{15b}$$

That is, Λ_{R} is kept above the scale at which we are probing the physics. When we integrate modes out, new effective interactions are generated. To see this, write the functional integral with a general interaction lagrangian $L(\phi, \Lambda)$:

$$\begin{aligned} Z(J, L, \Lambda) &= \int \mathrm{d}\phi \exp\left(\int \frac{\mathrm{d}^4 p}{(2\pi)^4} \left[-\tfrac{1}{2}\phi(p)\phi(-p)(p^2+m^2)K^{-1}(p^2/\Lambda^2) + J(p)\phi(-p) \right] + L(\phi, \Lambda) \right) \\ &\equiv \int \mathrm{d}\phi \exp S(\phi, \Lambda), \end{aligned} \tag{16}$$

so that

$$\Lambda \frac{\mathrm{d}Z}{\mathrm{d}\Lambda} = \int \mathrm{d}\phi \left(\int \frac{\mathrm{d}^4 p}{(2\pi)^4} \left[-\tfrac{1}{2}\phi(p)\phi(-p)(p^2+m^2)\Lambda \frac{\partial K^{-1}(p^2/\Lambda^2)}{\partial \Lambda} \right] + \Lambda \frac{\partial}{\partial \Lambda} L(\phi, \Lambda) \right) \exp S(\phi, \Lambda). \tag{17}$$



(a) (b)

Fig. 3. (a) Replacement of propagator with $(\partial L/\partial\phi)^2$ vertex. (b) Replacement of propagator with $(\partial^2 L/\partial\phi^2)$ vertex.

If we choose

$$\Lambda\frac{\partial L}{\partial\Lambda} = -\tfrac{1}{2}\int \mathrm{d}^4p\,(2\pi)^4(p^2+m^2)^{-1}\Lambda\frac{\partial K(p^2/\Lambda^2)}{\partial\Lambda} \times\left\{\frac{\partial L}{\partial\phi(-p)}\frac{\partial L}{\partial\phi(p)} + \frac{\partial^2 L}{\partial\phi(p)\,\partial\phi(-p)}\right\}, \tag{18}$$

then eq. (17) becomes

$$\Lambda\frac{\mathrm{d}Z}{\mathrm{d}\Lambda} = \int \mathrm{d}^4p\,\Lambda\frac{\partial K(p^2/\Lambda^2)}{\partial\Lambda}\int \mathrm{d}\phi\,\frac{\partial}{\partial\phi(p)} \times\left\{\left(\phi(p)K^{-1}(p^2/\Lambda^2) + \tfrac{1}{2}(2\pi)^4(p^2+m^2)^{-1}\frac{\partial}{\partial\phi(-p)}\right) \times\exp S(\phi,\Lambda)\right\} = 0. \tag{19}$$

(Recall that, due to (15b), $J(p)$ has no overlap with $\partial K/\partial\Lambda$.) This naive manipulation is justified because there is a cutoff. We have neglected field-independent terms which just change $Z(J, L, \Lambda)$ by an overall factor.

Eq. (19) says that if the cutoff Λ is reduced and simultaneously the lagrangian changed according to (18), $Z(J)$ and its functional derivatives, the n-point functions, are left unchanged. Eq. (18) has a simple graphical interpretation. As modes are removed from the propagator, compensating terms must be added in the interaction lagrangian. Graphs where the differentiated propagator connects two different vertices, as in fig. 3a, produce the first term in the bracket in (18), while graphs where both ends of the propagator connect to a single vertex, as in fig. 3b, produce the second term.

Although the lagrangian might start with a simple form such as (12), at lower scales it becomes quite complicated. However, at scales Λ far below Λ_0, we expect that a great simplification will occur. That is, no matter what initial lagrangian we start with (within limits) the lagrangian will be strongly attracted toward a three-dimensional submanifold in the infinite-dimensional space of possible lagrangians. Three, of course, is the number of relevant or renormalizable operators, namely ϕ^2, $\phi\partial^2\phi$, and ϕ^4 *. A convenient set of coordinates for the submanifold is obtained by

* For simplicity we keep the symmetry $\phi\to-\phi$.

expanding $L(\phi, \Lambda)$:

$$L(\phi, \Lambda) = \sum_{m=1}^{\infty} \frac{1}{(2m)!} \int \frac{\mathrm{d}^4 p_1 \ldots \mathrm{d}^4 p_{2m}}{(2\pi)^{8m-4}} L_{2m}(p_1, \ldots, p_{2m}, \Lambda)$$

$$\times \delta^4\Big(\sum_i p_i\Big) \phi(p_1) \ldots \phi(p_{2m}). \tag{20}$$

As always, we ignore the $m = 0$ field-independent piece. Then define

$$\rho_1(\Lambda) = -L_2(0, 0, \Lambda), \tag{21a}$$

$$\rho_2(\Lambda) = -\frac{1}{8} \frac{\partial^2}{\partial p_1^2} L_2(p_1, -p_1, \Lambda)|_{p_1=0}, \tag{21b}$$

$$\rho_3(\Lambda) = -L_4(0, 0, 0, 0, \Lambda). \tag{21c}$$

Note that in the approach we are taking, massless particles are no complication: (21) will be suitable even if $m = 0$. We emphasize that on the submanifold, the lagrangian is *not* expected to have any simple form such as (12), but that once the three "coordinates" (21a–21c) are given, the rest of the action is fixed.

We need to divide the flow of the lagrangian into relevant and irrelevant parts. Consider the vector (in the space of lagrangians)

$$\Lambda_0 \frac{\partial}{\partial \Lambda_0} L(\phi, \Lambda, \Lambda_0, \rho^0), \tag{22}$$

where the dependence on the initial condition (12) is written explicitly. This vector satisfies the linearized equation

$$\Lambda \frac{\partial}{\partial \Lambda}\left(\Lambda_0 \frac{\partial L}{\partial \Lambda_0}\right) = -\tfrac{1}{2} \int \mathrm{d}^4 p (2\pi)^4 \Lambda \frac{\partial K}{\partial \Lambda} (p^2 + m^2)^{-1}$$

$$\times \left\{ 2 \frac{\partial L}{\partial \phi(-p)} \frac{\partial}{\partial \phi(p)} \left(\Lambda_0 \frac{\partial L}{\partial \Lambda_0} \right) + \frac{\partial^2}{\partial \phi(p) \partial \phi(-p)} \left(\Lambda_0 \frac{\partial L}{\partial \Lambda_0} \right) \right\}, \tag{23}$$

which we will abbreviate:

$$\Lambda \frac{\partial}{\partial \Lambda}\left(\Lambda_0 \frac{\partial L}{\partial \Lambda_0}\right) = M\left(\Lambda_0 \frac{\partial L}{\partial \Lambda_0}\right), \tag{24a}$$

and in particular for the parameters (21),

$$\Lambda \frac{\partial}{\partial \Lambda}\left(\Lambda_0 \frac{\partial \rho_a}{\partial \Lambda_0}\right) = M_a\left(\Lambda_0 \frac{\partial L}{\partial \Lambda_0}\right). \tag{24b}$$

At small Λ, the vector (22) is expected to be very nearly parallel to the three-dimensional relevant submanifold. So are the three vectors

$$\frac{\partial}{\partial \rho_a^0} L(\phi, \Lambda, \Lambda_0, \rho^0), \tag{25}$$

which also satisfy eq. (24). Thus, if we form the linear combination of $\Lambda_0 \partial L/\partial \Lambda_0$ and $\partial L/\partial \rho_a^0$ which vanishes in the three directions (21), it should be driven very small in the infrared:

$$V(\Lambda) = \Lambda_0 \frac{\partial L(\Lambda)}{\partial \Lambda_0} - \sum_{a,b} \frac{\partial L(\Lambda)}{\partial \rho_a^0} \frac{\partial \rho_a^0}{\partial \rho_b(\Lambda)} \Lambda_0 \frac{\partial \rho_b(\Lambda)}{\partial \Lambda_0}. \tag{26}$$

Here $\partial \rho_a^0/\partial \rho_b(\Lambda)$ is the matrix inverse to $\partial \rho_b(\Lambda, \Lambda_0, \rho^0)/\partial \rho_a^0$. This inverse always exists in perturbation theory, as the zeroth order term is just δ_{ab}. $V(\Lambda)$ satisfies the linear equation

$$\begin{aligned}
\Lambda \frac{\partial V(\Lambda)}{\partial \Lambda} &= M\left(\Lambda_0 \frac{\partial L(\Lambda)}{\partial \Lambda_0}\right) - \sum_{a,b} M\left(\frac{\partial L(\Lambda)}{\partial \rho_a^0}\right) \frac{\partial \rho_a^0}{\partial \rho_b(\Lambda)} \Lambda_0 \frac{\partial \rho_b(\Lambda)}{\partial \Lambda_0} \\
&\quad + \sum_{a,b,c,d} \frac{\partial L(\Lambda)}{\partial \rho_a^0} \frac{\partial \rho_a^0}{\partial \rho_b(\Lambda)} M_b\left(\frac{\partial L}{\partial \rho_c^0}\right) \frac{\partial \rho_c^0}{\partial \rho_d(\Lambda)} \Lambda_0 \frac{\partial \rho_d(\Lambda)}{\partial \Lambda_0} \\
&\quad - \sum_{a,b} \frac{\partial L(\Lambda)}{\partial \rho_a^0} \frac{\partial \rho_a^0}{\partial \rho_b(\Lambda)} M_b\left(\Lambda_0 \frac{\partial L(\Lambda)}{\partial \Lambda_0}\right) \\
&= M(V(\Lambda)) - \sum_{a,b} \frac{\partial L(\Lambda)}{\partial \rho_a^0} \frac{\partial \rho_a^0}{\partial \rho_b(\Lambda)} M_b(V(\Lambda)),
\end{aligned} \tag{27}$$

using the linearity of M. Eq. (27) is our key. It is an equation linear in the quantity we wish to prove irrelevant, analogous to eq. (5). When (27) is written out explicitly, it will contain complicated non-linear terms involving products of V with L. The exact form of these terms no longer matters, however, for these terms are the ones which vanish fastest when the coupling is made small. Thus, at sufficiently small coupling, the infrared behaviour of (27) will be dominated by the purely linear terms. These terms drive V to zero according to its canonical dimensions, since we have constructed V so that all pieces with dimensions of mass to non-negative powers are absent. We will prove that this is true in perturbation theory. The point is to show that the anomalous dimensions are couplings times powers of logarithms, so the integer canonical dimensions dominate.

In order to put bounds on V, we will need bounds on the running lagrangian (20) and its first four-momentum derivatives, on the quantity

$$\frac{\partial L(\Lambda)}{\partial\rho_b(\Lambda)} \equiv \sum_a \frac{\partial L(\Lambda)}{\partial\rho_a^0}\,\frac{\partial\rho_a^0}{\partial\rho_b(\Lambda)}\,, \tag{28}$$

which appears in (27), and its first four-momentum derivatives, and finally on V and its first four-momentum derivatives. We will obtain some further results about these quantities, and then proceed in sect. 4 with the order-by-order proof. The cutoff function $K(p^2/\Lambda^2)$ will be taken to have the form such as

$$K(p^2/\Lambda^2) = \begin{cases} 1\,, & p^2 \leqslant \Lambda^2 \\ \exp\left[(1-p^2/\Lambda^2)^{-1}\exp\left((4-p^2/\Lambda^2)^{-1}\right)\right], & \Lambda^2 < p^2 < 4\Lambda^2 \\ 0\,, & p^2 \geqslant 4\Lambda^2\,, \end{cases} \tag{29}$$

which has been chosen because it is infinitely differentiable, but its derivative vanishes except for $\Lambda^2 < p^2 < 4\Lambda^2$ *. Define

$$Q(p,\Lambda,m^2) = \frac{1}{(p^2+m^2)}\Lambda^3\frac{\partial}{\partial\Lambda}K(p^2/\Lambda^2)\,. \tag{30}$$

One may easily verify that there are constants C and D_n such that for $|m^2| \leqslant \Lambda_R \leqslant \Lambda$,

$$\int\frac{\mathrm{d}^4p}{(2\pi)^4}|Q(p,\Lambda,m^2)| < C\Lambda^4\,, \tag{31}$$

$$\left\|\frac{\partial^n}{\partial p^n}Q(p,\Lambda,m^2)\right\| < D_n\Lambda^{-n}\,. \tag{32}$$

* It might cause some concern that eqs. (16), (17) and (19) refer to $K^{-1}(p)$, yet $K(p)$ is taken to vanish for $p^2 > 4\Lambda^2$. In fact the n-point functions are still invariant under eq. (18), as may be seen by considering $K(p)$ as the limit of a $\tilde{K}(p)$ which nowhere vanishes. Since the propagator (11) and the effective lagrangian (18) involve only $K(p)$ and not $K^{-1}(p)$, this limit is smooth and the invariance of the n-point functions under eq. (18) for $\tilde{K}(p)$ implies the same for $K(p)$. It is also possible to carry through the whole proof with $\tilde{K}(p)$, but then a different norm from (33), with a non-zero weight function falling off as $p \to \infty$, must be used.

The notation $\| \; \|$ will be used often. It is defined

$$\|f(p_1,\ldots,p_m,\Lambda)\| = \max_{p_i^2<4\Lambda^2} |f(p_1,\ldots,p_m,\Lambda)|, \tag{33}$$

for a function of one or more momenta. This is a useful definition because the values of the effective vertices for $p_i^2 > 4\Lambda^2$ are meaningless, the propagator vanishing in this range.

Now write the renormalization group equation (18) in terms of the dimensionless functions $A_{2m} = \Lambda^{2m-4}L_{2m}$, where L_{2m} are the component functions in (20):

$$\begin{aligned}
&\left(\Lambda\frac{\partial}{\partial\Lambda} + 4 - 2m\right)A_{2m}(p_1,\ldots,p_{2m},\Lambda)\\
&\quad = -\sum_{l=1}^{m}\left\{Q(P,\Lambda,m^2)A_{2l}(p_1,\ldots,p_{2l-1},P,\Lambda)\right.\\
&\qquad\qquad \left.\times A_{2m+2-2l}(p_{2l},\ldots,p_{2m},-P,\Lambda) + \frac{1}{2}\binom{2m}{2l-1} - 1 \text{ permutations}\right\}\\
&\qquad -\tfrac{1}{2}\int\frac{\mathrm{d}^4p}{(2\pi\Lambda)^4}A_{2m+2}(p_1,\ldots,p_{2m},p,-p,\Lambda)Q(p,\Lambda,m^2),
\end{aligned} \tag{34}$$

where $P = \sum_{i=1}^{2l-1}p_i$. Referring to (31) and (32) gives

$$\begin{aligned}
&\left\|\left(\Lambda\frac{\partial}{\partial\Lambda} + 4 - 2m\right)A_{2m}(p_1,\ldots,p_{2m},\Lambda)\right\|\\
&\quad\leqslant \sum_{l=1}^{m}\left\{\frac{1}{2}\binom{2m}{2l-1}D_0\|A_{2l}(\Lambda)\|\cdot\|A_{2m+2-2l}(\Lambda)\|\right\} + \tfrac{1}{2}C\|A_{2m+2}(\Lambda)\|.
\end{aligned} \tag{35}$$

Eq. (35) shows the simplicity of the present approach. Because the momentum integrals encountered are always restricted in range, they may be estimated naively, replacing $A_{2m+2}(p_i)$ with its maximum value. Thus, the detailed momentum dependence no longer enters and eq. (35) is nearly as simple as the equations in sect. 2. From (35) one can see the strategy that will be followed in the inductive proof. At any order in perturbation theory, the non-linear terms on the right-hand side of (35) will already be known. At each order it is then necessary to proceed inductively downward from large m, so the linear term on the right will be controlled (at any given order A_{2m} vanishes for large enough m).

Now consider

$$\left(\frac{\partial}{\partial p_i^\mu}-\frac{\partial}{\partial p_j^\mu}\right)A_{2m}(p_1,\ldots,p_{2m},\Lambda)\equiv\partial^\mu_{i,j}A_{2m}. \tag{36}$$

Only such a difference of derivatives makes sense, as $A_{2m}(p_1,\ldots,p_{2m},\Lambda)$ is defined only for $\Sigma_i p_i=0$. Acting on eq. (34) with $\partial^\mu_{i,j}$ and proceeding as in (35), it follows that

$$\begin{aligned}&\left\|\left(\Lambda\frac{\partial}{\partial\Lambda}+4-2m\right)\partial^\mu_{i,j}A_{2m}(p_1,\ldots,p_{2m},\Lambda)\right\|\\&\leqslant\sum_{l=1}^{m}\left\{\frac{1}{2}\binom{2m}{2l-1}\left(D_1\Lambda^{-1}\|A_{2l}(\Lambda)\|\cdot\|A_{2m+2-2l}(\Lambda)\|\right.\right.\\&\qquad\left.\left.+2D_0\|\partial^\mu_{i,j}A_{2l}(\Lambda)\|\cdot\|A_{2m+2-2l}(\Lambda)\|\right)\right\}+\tfrac{1}{2}C\|\partial^\mu_{i,j}A_{2m+2}(\Lambda)\|.\end{aligned} \tag{37}$$

Again this will involve only lower orders or larger m. Similar results hold for any number of derivatives acting on A_{2m}.

Turning now to the quantity $\partial L(\Lambda)/\partial\rho_b(\Lambda)$, eq. (28), one finds, using the fact that M is a linear operator,

$$\Lambda\frac{\partial}{\partial\Lambda}\left(\frac{\partial L(\Lambda)}{\partial\rho_b(\Lambda)}\right)=M\left(\frac{\partial L(\Lambda)}{\partial\rho_b(\Lambda)}\right)-\sum_a\frac{\partial L(\Lambda)}{\partial\rho_a(\Lambda)}M_a\left(\frac{\partial L(\Lambda)}{\partial\rho_b(\Lambda)}\right). \tag{38}$$

Define an expansion in terms of dimensionless functions $B_{b,2m}$:

$$\begin{aligned}\frac{\partial L(\Lambda)}{\partial\rho_b(\Lambda)}=\sum_{m=1}^{\infty}\frac{\Lambda^{4-2m-2\delta_{b1}}}{(2m)!}\int\frac{\mathrm{d}^4p_1\ldots\mathrm{d}^4p_{2m}}{(2\pi)^{8m-4}}B_{b,2m}(p_1,\ldots,p_{2m},\Lambda)\\\times\delta^4\left(\sum_i p_i\right)\phi(p_1)\ldots\phi(p_{2m}).\end{aligned} \tag{39}$$

Eq. (38) becomes

$$\begin{aligned}
&\left(\Lambda\frac{\partial}{\partial\Lambda}+4-2m-2\delta_{b1}\right)B_{b,2m}(p_1,\dots,p_{2m},\Lambda)\\
&\quad=-\sum_{l=1}^{m}\Big\{Q(P,\Lambda,m^2)A_{2l}(p_1,\dots,p_{2l-1},P,\Lambda)\\
&\qquad\qquad\times B_{b,2m+2-2l}(p_{2l},\dots,p_{2m},-P,\Lambda)+\binom{2m}{2l-1}-1\text{ permutations}\Big\}\\
&\quad-\tfrac{1}{2}\int\frac{\mathrm{d}^4p}{(2\pi\Lambda)^4}B_{b,2m+2}(p_1,\dots,p_{2m},p,-p,\Lambda)Q(p,\Lambda,m^2)\\
&\quad+B_{1,2m}(p_1,\dots,p_{2m},\Lambda)\tfrac{1}{2}\int\frac{\mathrm{d}^4q}{(2\pi\Lambda)^4}B_{b,4}(0,0,q,-q,\Lambda)Q(q,\Lambda,m^2)\\
&\quad+B_{2,2m}(p_1,\dots,p_{2m},\Lambda)\frac{\Lambda^2}{16}\\
&\quad\times\int\frac{\mathrm{d}^4q}{(2\pi\Lambda)^4}\frac{\partial^2}{\partial q_1^2}B_{b,4}(q_1,-q_1,q,-q,\Lambda)|_{q_1=0}Q(q,\Lambda,m^2)\\
&\quad+B_{3,2m}(p_1,\dots,p_{2m},\Lambda)\tfrac{1}{2}\\
&\quad\times\int\frac{\mathrm{d}^4q}{(2\pi\Lambda)^4}B_{b,6}(0,0,0,0,q,-q,\Lambda)Q(q,\Lambda,m^2),
\end{aligned}\tag{40}$$

so that

$$\begin{aligned}
&\left\|\left(\Lambda\frac{\partial}{\partial\Lambda}+4-2m-2\delta_{b1}\right)B_{b,2m}(p_1,\dots,p_{2m},\Lambda)\right\|\\
&\quad\leqslant\sum_{l=1}^{m}\left\{\binom{2m}{2l-1}D_0\|A_{2l}(\Lambda)\|\cdot\|B_{b,2m+2-2l}(\Lambda)\|\right\}\\
&\qquad+\tfrac{1}{2}C\|B_{b,2m+2}(\Lambda)\|+\tfrac{1}{2}C\|B_{1,2m}(\Lambda)\|\cdot\|B_{b,4}(\Lambda)\|\\
&\qquad+\tfrac{1}{16}C\Lambda^2\|B_{2,2m}(\Lambda)\|\cdot\|\partial^{\mu}_{1,2}\partial^{\mu}_{1,2}B_{b,4}(\Lambda)\|\\
&\qquad+\tfrac{1}{2}C\|B_{3,2m}(\Lambda)\|\cdot\|B_{b,6}(\Lambda)\|.
\end{aligned}\tag{41}$$

Similar results, parallel to (37), hold for any number of derivatives acting on $B_{b,2m}$.

Finally, expand $V(\Lambda)$ in dimensionless functions V_{2m}:

$$V(\Lambda)=\sum_{m=1}^{\infty}\frac{\Lambda^{4-2m}}{(2m)!}\int\frac{\mathrm{d}^4p_1\dots\mathrm{d}^4p_{2m}}{(2\pi)^{8m-4}}V_{2m}(p_1,\dots,p_{2m},\Lambda)$$
$$\times\delta^4\Big(\sum_i p_i\Big)\phi(p_1)\dots\phi(p_{2m}). \tag{42}$$

Eq. (27) becomes

$$\Big(\Lambda\frac{\partial}{\partial\Lambda}+4-2m\Big)V_{2m}(p_1,\dots,p_{2m},\Lambda)$$
$$=-\sum_{l=1}^{m}\Big\{Q(P,\Lambda,m^2)A_{2l}(p_1,\dots,p_{2l-1},P,\Lambda)$$
$$\times V_{2m+2-2l}(p_{2l},\dots,p_{2m},-P,\Lambda)+\binom{2m}{2l-1}-1\text{ permutations}\Big\}$$
$$-\tfrac{1}{2}\int\frac{\mathrm{d}^4p}{(2\pi\Lambda)^4}V_{2m+2}(p_1,\dots,p_{2m},p,-p,\Lambda)Q(p,\Lambda,m^2)$$
$$+B_{1,2m}(p_1,\dots,p_{2m},\Lambda)\tfrac{1}{2}\int\frac{\mathrm{d}^4q}{(2\pi\Lambda)^4}V_4(0,0,q,-q,\Lambda)Q(q,\Lambda,m^2)$$
$$+B_{2,2m}(p_1,\dots,p_{2m},\Lambda)\frac{\Lambda^2}{16}\int\frac{\mathrm{d}^4q}{(2\pi\Lambda)^4}$$
$$\times\frac{\partial^2}{\partial q_1^2}V_4(q_1,-q_1,q,-q,\Lambda)|_{q_1=0}Q(q,\Lambda,m^2)$$
$$+B_{3,2m}(p_1,\dots,p_{2m},\Lambda)\tfrac{1}{2}\int\frac{\mathrm{d}^4q}{(2\pi\Lambda)^4}$$
$$\times V_6(0,0,0,0,q,-q,\Lambda)Q(q,\Lambda,m^2). \tag{43}$$

From this

$$\Big\|\Big(\Lambda\frac{\partial}{\partial\Lambda}+4-2m\Big)V_{2m}(p_1,\dots,p_{2m},\Lambda)\Big\|$$
$$\leqslant\sum_{l=1}^{m}\Big\{\binom{2m}{2l-1}D_0\|A_{2l}(\Lambda)\|\cdot\|V_{2m+2-2l}(\Lambda)\|\Big\}$$
$$+\tfrac{1}{2}C\|V_{2m+2}(\Lambda)\|+\tfrac{1}{2}C\|B_{1,2m}(\Lambda)\|\cdot\|V_4(\Lambda)\|$$
$$+\tfrac{1}{16}C\Lambda^2\|B_{2,2m}(\Lambda)\|\cdot\|\partial^{\mu}_{1,2}\partial^{\mu}_{1,2}V_4(\Lambda)\|$$
$$+\tfrac{1}{2}C\|B_{3,2m}(\Lambda)\|\cdot\|V_6(\Lambda)\|, \tag{44}$$

and so on for any number of derivatives.

4. Perturbative renormalization

In this section we will prove, in perturbation theory, that $V(\Lambda)$ is driven to zero in the infrared. At the end of the section we give the steps needed to convert this into a statement of the cutoff-independence of n-point functions.

Theorem. Consider the field theory defined by the propagator (11) and vertices (12). Define the effective lagrangian $L(\phi, \Lambda, \Lambda_0, \rho^0)$ and the relevant parameters $\rho_b(\Lambda, \Lambda_0, \rho^0)$ as before. Define $\rho_a^0(\Lambda_R, \lambda^R, \Lambda_0)$ implicitly by

$$\rho_1(\Lambda_R, \Lambda_0, \rho^0) = 0, \tag{45a}$$

$$\rho_2(\Lambda_R, \Lambda_0, \rho^0) = 0, \tag{45b}$$

$$\rho_3(\Lambda_R, \Lambda_0, \rho^0) = \lambda^R. \tag{45c}$$

Then order by order in perturbation theory in λ^R, the limit

$$\lim_{\Lambda_0 \to \infty} \tilde{L}(\phi, \Lambda_R, \lambda^R, \Lambda_0) = \tilde{L}(\phi, \Lambda_R, \lambda^R, \infty), \tag{46}$$

exists, where $\tilde{L}(\phi, \Lambda_R, \lambda^R, \Lambda_0) \equiv L(\phi, \Lambda_R, \Lambda_0, \rho^0(\Lambda_R, \lambda^R, \Lambda_0))$, and at rth order in λ_R,

$$\|\tilde{L}^{(r)}_{2m}(\Lambda_R, \Lambda_0) - \tilde{L}^{(r)}_{2m}(\Lambda_R, \infty)\| \times \begin{cases} \leqslant \Lambda_R^{4-2m}\left(\dfrac{\Lambda_R}{\Lambda_0}\right)^2 P^{2r-m}(\ln(\Lambda_0/\Lambda_R)), & r+1-m \geqslant 0 \\ = 0, & r+1-m < 0, \end{cases} \tag{47}$$

where P^{2r-m} is a polynomial of degree $2r - m$ whose coefficients, taken to be non-negative, are pure numbers*. The degree of the polynomial is not essential to the proof and the reader is free to ignore the superscript on P throughout. We retain the degree only because it is easy to do so; in fact it can be reduced to $r + 1 - m$, but this requires distracting additional steps.

Note that eqs. (45) actually do define ρ_a^0 implicitly in perturbation theory, for at order r in λ_R the left-hand side of (45) is $\rho_a^{0(r)}$ plus terms which are already known.

Lemma (i). At order r in λ^R,

$$\|\partial^{\mu_1}_{i_1, j_1} \cdots \partial^{\mu_p}_{i_p, j_p} A^{(r)}_{2m}(p_1, \ldots, p_{2m}, \Lambda)\| \leqslant \Lambda^{-p} P^{2r-m}(\ln(\Lambda_0/\Lambda_R)), \qquad r+1-m \geqslant 0,$$
$$= 0, \qquad r+1-m < 0. \tag{48}$$

* P^n is defined to be zero for $n < 0$.

Proof of lemma (i). Note that the initial value of A_{2m}, at Λ_0, is defined by eq. (12), so the LHS of (48) vanishes at Λ_0 for $m \geqslant 3$, for $m = 2$, $p \geqslant 1$, and for $m = 0$, $p \geqslant 3$. The lemma is trivially true for $r = 0$, as (45) implies $A^{(0)}_{2m} = 0$. Suppose it to be true for $r = s - 1$, for some $s \geqslant 1$. We now proceed downward in m. The lemma is true for $m \geqslant s + 2$, that is $A^{(s)}_{2m}$ then vanishes, for the operations shown in fig. 3 could have linked together at most s vertices, leaving $2s + 2$ external fields*. Suppose it is true for $m \geqslant n + 1$. The right-hand side of (35) at order s is

$$\sum_{l=1}^{n}\sum_{t=1}^{s-1}\left\{\tfrac{1}{2}\binom{2n}{2l-1} D_0 \|A^{(t)}_{2l}(\Lambda)\| \cdot \|A^{(s-t)}_{2n+2-2l}(\Lambda)\|\right\} + \tfrac{1}{2} C \|A^{(s)}_{2n+2}(\Lambda)\| . \tag{49}$$

All quantities in (49) are bounded by the induction hypothesis, so (49) is less than or equal to

$$\begin{aligned}&\sum_{l=1}^{n}\sum_{t=1}^{s-1} P^{2t-l}(\ln(\Lambda_0/\lambda_R))P^{2s-2t-n+l-1}(\ln(\Lambda_0/\Lambda_R)) + P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R))\\ &\quad = P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R)) .\end{aligned} \tag{50}$$

In a similar way, all quantities on the right-hand side of (37) and the corresponding equations for higher derivatives are bounded by the induction hypothesis, so

$$\left\|\left(\Lambda\frac{\partial}{\partial\Lambda} + 4 - 2n\right)\partial^{\mu_1}_{i_1, j_1}\cdots\partial^{\mu_p}_{i_p, j_p} A^{(s)}_{2n}(p_1,\dots,p_{2n},\Lambda)\right\| \leqslant \Lambda^{-p}P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R)) . \tag{51}$$

Integrate (51). For $n \geqslant 3$

$$\begin{aligned}&\|\partial^{\mu_1}_{i_1, j_1}\cdots\partial^{\mu_p}_{i_p, j_p} A^{(s)}_{2n}(p_1,\dots,p_{2n},\Lambda)\|\\ &\quad \leqslant \Lambda^{-p}P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R))\int_{\Lambda}^{\Lambda_0}\frac{\mathrm{d}\Lambda'}{\Lambda'}\left(\frac{\Lambda}{\Lambda'}\right)^{2n+p-4}\\ &\quad \leqslant \Lambda^{-p}P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R))\\ &\quad \leqslant \Lambda^{-p}P^{2s-n}(\ln(\Lambda_0/\Lambda_R)) .\end{aligned} \tag{52}$$

* It is curious that this one simple fact cannot be determined from eq. (34). To see the problem, consider the simpler system $\mathrm{d}f_i(x)/\mathrm{d}x = f_{i+1}(x)$, $i = 1,\dots,\infty$, $f_i(0) = 0$. This has many solutions, such as $f_i(x) = 0$, or $f_1(x) = \exp(-1/x^2)$, $f_2(x) = 2\exp(-1/x^2)/x^3, \dots$. Eq. (34) becomes complete with additional information, such as the vanishing of $A^{(s)}_{2m}$ for large m.

This establishes the induction step down to $n = 3$. In the same way it carries to $n = 2$, $p \geqslant 1$, but for $n = 2$, $p = 0$ there is an extra term from the initial value of A_4, and we proceed differently. By (51),

$$\left| \Lambda \frac{\partial}{\partial \Lambda} A_4^{(s)}(0,0,0,0,\Lambda) \right| \leqslant P^{2s-3}(\ln(\Lambda_0/\Lambda_R)). \tag{53}$$

Integrate (53) *up* from the initial value $A_4^{(s)} = -\delta^{s1}$ at $\Lambda = \Lambda_R$, eq. (45c), to obtain

$$|A_4^{(s)}(0,0,0,0,\Lambda)| \leqslant \delta^{s1} + P^{2s-2}(\ln(\Lambda_0/\Lambda_R)) = P^{2s-2}(\ln(\Lambda_0/\Lambda_R)). \tag{54}$$

Now reconstruct $A_4^{(s)}$ via Taylor's theorem:

$$A_4^{(s)}(p_1, p_2, p_3, p_4, \Lambda) = A_4^{(s)}(0,0,0,0,\Lambda) + \sum_{i,j=1}^{3} p_i^\mu p_i^\nu \int_0^1 \mathrm{d}\lambda (1-\lambda)$$

$$\times \left(\partial^\mu_{i,4} \partial^\nu_{j,4} A_4^{(s)}(p_1', p_2', p_3', p_4', \Lambda) \right)\Big|_{p'=\lambda p}. \tag{55}$$

Both terms in (55) are bounded by $P^{2s-2}(\ln(\Lambda_0/\Lambda_R))$, the first from eq. (54) and the second from eq. (52) for $n = 2$, $p = 2$, so the induction step follows for all of $A_4^{(s)}$. For $n = 1$, $p \geqslant 3$, eqs. (51) and (52) follow immediately, while from (51),

$$\left| \left(\Lambda \frac{\partial}{\partial \Lambda} + 2 \right) A_2^{(s)}(0,0,\Lambda) \right| \leqslant P^{2s-2}(\ln(\Lambda_0/\Lambda_R)),$$

$$\left| \left(\Lambda \frac{\partial}{\partial \Lambda} + 2 \right) \partial^\mu_{1,2} \partial^\nu_{1,2} A_2^{(s)}(p_1', p_2', \Lambda)_{p_1'=p_2'=0} \right| \leqslant \Lambda^{-2} P^{2s-2}(\ln(\Lambda_0/\Lambda_R)). \tag{56}$$

Again integrating up from the initial conditions (45a, 45b) one obtains

$$|A_2^{(s)}(0,0,\Lambda)| \leqslant P^{2s-2}(\ln(\Lambda_0/\Lambda_R)),$$

$$|\partial^\mu_{1,2} \partial^\nu_{1,2} A_2^{(s)}(p_1', p_2', \Lambda)_{p_1'=p_2'=0}| \leqslant \Lambda^{-2} P^{2s-1}(\ln(\Lambda_0/\Lambda_R)). \tag{57}$$

Now use Taylor's theorem to reconstruct $A_2^{(s)}$:

$$A_2^{(s)}(p_1, -p_1, \Lambda) = A_2^{(s)}(0,0,\Lambda) + \tfrac{1}{2} p_1^\mu p_1^\nu \left(\partial^\mu_{1,2}\partial^\nu_{1,2} A_2^{(s)}(p_1', p_2', \Lambda)\right)_{p_1'=p_2'=0}$$
$$+ \frac{1}{3!} p_1^\mu p_1^\nu p_1^\sigma p_1^\omega \int_0^1 \mathrm{d}\lambda\,(1-\lambda)^3$$
$$\times\left(\partial^\mu_{1,2}\partial^\nu_{1,2}\partial^\sigma_{1,2}\partial^\omega_{1,2} A_2^{(s)}(p_1', p_2', \Lambda)\right)_{p_1'=-p_2'=\lambda p_1}. \tag{58}$$

Eq. (52) for $n = 1$, $p = 4$ and eq. (57) establish the induction step for $n = 1$, and the lemma is proven.

It is somewhat contrary to the spirit of our proof that we have had to integrate up from Λ_R and not just down from Λ_0, and this deserves some comment. In the case of $A_4(0)$ and $\partial^2 A_2(0)$, this was just a convenience. The initial conditions for these at Λ_0 involve ρ_2^0 and ρ_3^0, so proceeding as we have, we avoid carrying ρ_2^0 and ρ_3^0 around. (This is also why $\ln \Lambda_R$ appears so early: it comes entirely from the dependence of ρ_2^0 and ρ_3^0 on $\ln(\Lambda_0/\Lambda_R)$.) In the case of $A_2(0)$, however, integrating from Λ_R was essential. The point is that our theory is not actually on a typical trajectory. On a typical trajectory, $A_2(0) \sim \Lambda_0^2/\Lambda^2$, that is, the scalar mass is of order Λ_0, and this is all we would have learned integrating down from Λ_0. By imposing (45a) we have forced the initial values of ρ_a^0 to be finely adjusted so as to produce a scalar with $m \ll \Lambda_0$. We must therefore integrate up from (45a) to produce the needed bound on A_2. This is of course the famous naturalness problem for light scalars [8].

Lemma (ii). At order r in λ^R,

$$\|\partial^{\mu_1}_{i_1,j_1}\cdots\partial^{\mu_p}_{i_p,j_p} B^{(r)}_{b,2m}(p_1,\ldots,p_{2m},\Lambda)\| \leqslant \Lambda^{\;\;p} P^{2r\;\;\;m-1+\delta_{b3}}(\ln(\Lambda_0/\Lambda_R)).$$
$$r+2-m \geqslant 0,$$
$$=0, \qquad r+2-m<0. \tag{59}$$

Proof of lemma (ii). From the definition (28), the initial condition is

$$B_{b,2m}(p_1,\ldots,p_{2m},\Lambda_0) = -\delta_{b1}\delta_{m1} - \delta_{b2}\delta_{m1}\frac{p_1^2}{\Lambda_0^2} - \delta_{b3}\delta_{m2}. \tag{60}$$

This also gives the full zeroth-order term,

$$B^{(0)}_{b,2m}(p_1,\ldots,p_{2m},\Lambda) = -\delta_{b1}\delta_{m1} - \delta_{b2}\delta_{m1}\frac{p_1^2}{\Lambda^2} - \delta_{b3}\delta_{m2}, \tag{61}$$

as may be verified by inserting (61) into (40). Thus the lemma is established for

$r = 0$. Note that the definition (28) gives certain parts of B_b exactly:

$$B^{(r)}_{b,2}(0,0,\Lambda) = -\delta^{r0}\delta_{b1}, \tag{62a}$$

$$\partial^{\mu}_{1,2}\partial^{\nu}_{1,2}B^{(r)}_{b,2}(p_1,p_2,\Lambda)_{p_1=p_2=0} = -2\delta^{r0}\delta^{\mu\nu}\delta_{b2}, \tag{62b}$$

$$B^{(r)}_{b,4}(0,0,0,0,\Lambda) = -\delta^{r0}\delta_{b3}. \tag{62c}$$

Suppose now that the lemma is true for $r = s-1$, for some $s \geqslant 1$. Again proceed downward in m. The vanishing of $B^{(s)}_{b,2m}$ for $m \geqslant s+3$ follows from the definition (28) and the vanishing of $A^{(s)}_{2m}$ for $m \geqslant s+2$. Suppose the lemma is true for $m = n+1$. Eq. (41) becomes

$$\begin{aligned}
&\left\|\left(\Lambda\frac{\partial}{\partial\Lambda} + 4 - 2n - 2\delta_{b1}\right)B^{(s)}_{b,2n}(p_1,\ldots,p_{2n},\Lambda)\right\| \\
&\quad\leqslant \sum_{l=1}^{n}\sum_{t=1}^{s}\left\{\binom{2n}{2l-1}D_0\|A^{(t)}_{2l}(\Lambda)\|\cdot\|B^{(s-t)}_{b,2n+2-2l}(\Lambda)\|\right\} \\
&\qquad + \tfrac{1}{2}C\|B^{(s)}_{b,2n+2}(\Lambda)\| + \tfrac{1}{2}C\sum_{t=0}^{s}\|B^{(t)}_{1,2n}(\Lambda)\|\cdot\|B^{(s-t)}_{b,4}(\Lambda)\| \\
&\qquad + \tfrac{1}{16}C\Lambda^2\sum_{t=0}^{s-1}\|B^{(t)}_{2,2n}(\Lambda)\|\cdot\|\partial^{\mu}_{1,2}\partial^{\mu}_{1,2}B^{(s-t)}_{b,4}(\Lambda)\| \\
&\qquad + \tfrac{1}{2}C\sum_{t=0}^{s-1}\|B^{(t)}_{3,2n}(\Lambda)\|\cdot\|B^{(s-t)}_{b,6}(\Lambda)\|.
\end{aligned} \tag{63}$$

Because of the zeroth-order term (61), there are some terms on the right side of (63) which involve $B^{(s)}$ and not just $B^{(s-1)}$, and so are potentially not bounded by the induction hypothesis. These are the last three terms when $t = 0$ and the third to last term when $t = s$. The latter problem involves only $b = 3$ and the unknown bound is on $B^{(s)}_{1,2n}$, so there is no problem if at each value of s and n we bound $B^{(s)}_{1,2n}$ before $B^{(s)}_{3,2n}$. The former problem is not present if $n \geqslant 3$, because of the form of (61), so for $n \geqslant 3$

$$\begin{aligned}
&\left\|\left(\Lambda\frac{\partial}{\partial\Lambda} + 4 - 2n - 2\delta_{b1}\right)B^{(s)}_{b,2n}(p_1,\ldots,p_{2n},\Lambda)\right\| \\
&\quad\leqslant \sum_{l=1}^{n}\sum_{t=1}^{s}\left\{P^{2t-l}(\ln(\Lambda_0/\Lambda_R))P^{2s-2t-n+l+\delta_{b3}}(\ln(\Lambda_0/\Lambda_R))\right\} \\
&\qquad + P^{2s-n+\delta_{b3}}(\ln(\Lambda_0/\Lambda_R)) \\
&\qquad + \sum_{t=0}^{s}P^{2t-n+1}(\ln(\Lambda_0/\Lambda_R))P^{2s-2t-1+\delta_{b3}}(\ln(\Lambda_0/\Lambda_R)) \\
&\qquad + \text{like terms} \leqslant P^{2s-n+\delta_{b3}}(\ln(\Lambda_0/\Lambda_R)).
\end{aligned} \tag{64}$$

In a similar way,

$$\left\| \left(\Lambda \frac{\partial}{\partial \Lambda} + 4 - 2n - 2\delta_{h1} \right) \partial^{\mu_1}_{i_1, j_1} \cdots \partial^{\mu_p}_{i_p, j_p} B^{(s)}_{b,2n}(p_1, \ldots, p_{2n}, \Lambda) \right\|$$
$$\leqslant \Lambda^{-p} P^{2s-n+\delta_{b1}} (\ln(\Lambda_0/\Lambda_R)) . \tag{65}$$

Integrating

$$\| \partial^{\mu_1}_{i_1, j_1} \cdots \partial^{\mu_p}_{i_p, j_p} B^{(s)}_{b,2n}(p_1, \ldots, p_{2n}, \Lambda) \|$$
$$\leqslant \Lambda^{-p} P^{2s-n+\delta_{b1}} (\ln(\Lambda_0/\Lambda_R)) \int_{\Lambda}^{\Lambda_0} \frac{\mathrm{d}\Lambda'}{\Lambda'} \left(\frac{\Lambda}{\Lambda'} \right)^{2n+2\delta_{b1}-p-4}$$
$$\leqslant \Lambda^{-p} P^{2s-n-1+\delta_{b1}} (\ln(\Lambda_0/\Lambda_R)) \tag{66}$$

which establishes the induction step down to $n = 3$. Once (66) is known for $B^{(s)}_{b,6}$, the right-hand side of (63) is bounded for $n = 2$ and (64) and (66) then follow for $n = 2$. Then with (66) for $B^{(s)}_{b,4}$, (63) is bounded for $n = 1$ and (64) follows; eq. (66) now follows for $n = 1$, $p \geqslant 2$ and with (62a) the rest of $B^{(s)}_{b,2}$ can be reconstructed using Taylor's theorem, completing the lemma.

Lemma (iii). At order r in λ^R

$$\| \partial^{\mu_1}_{i_1, j_1} \cdots \partial^{\mu_p}_{i_p, j_p} V^{(r)}_{2m}(p_1, \ldots, p_{2m}, \Lambda) \| \leqslant \Lambda^{-p} \left(\frac{\Lambda}{\Lambda_0} \right)^2 P^{2r-m} (\ln(\Lambda_0/\Lambda_R)) ,$$
$$r + 1 - m \geqslant 0 ,$$
$$= 0 , \qquad r + 1 - m < 0 . \tag{67}$$

Proof of lemma (iii). To obtain the initial condition on V, given the definition (26), use

$$\Lambda_0 \frac{\partial}{\partial \Lambda_0} L(\Lambda, \Lambda_0, \rho^0)|_{\Lambda = \Lambda_0} + \Lambda \frac{\partial}{\partial \Lambda} L(\Lambda, \Lambda_0, \rho^0)|_{\Lambda = \Lambda_0}$$
$$= \Lambda_0 \frac{\partial}{\partial \Lambda_0} L(\Lambda_0, \Lambda_0, \rho^0) = 0 , \tag{68}$$

so that with eq. (51) of lemma (i), the initial condition satisfies

$$\| \partial^{\mu_1}_{i_1, j_1} \cdots \partial^{\mu_p}_{i_p, j_p} V^{(r)}_{2m}(p_1, \ldots, p_{2m}, \Lambda_0) \| \leqslant \Lambda_0^{-p} P^{2r-m-1} (\ln(\Lambda_0/\Lambda_R)) ,$$
$$r + 1 - m \geqslant 0 ,$$
$$= 0 , \qquad r + 1 - m < 0 . \tag{69}$$

Also from the definition (26), it follows that $V_{2m}^{(0)} = 0$. Suppose the lemma to be true for $r = s - 1$, some $s \geqslant 1$. Again we argue inductively downward in m. The vanishing of $V_{2m}^{(s)}$ for $m \geqslant s + 2$ follows by the same argument as for $A_{2m}^{(s)}$. Suppose the lemma to be true for $m \geqslant n + 1$. Then from (44):

$$
\begin{aligned}
&\left\| \left(\Lambda \frac{\partial}{\partial \Lambda} + 4 - 2n \right) V_{2n}^{(r)}(p_1, \ldots, p_{2n}, \Lambda) \right\| \\
&\quad \leqslant \sum_{l=1}^{n} \sum_{t=1}^{s} \left\{ \binom{2n}{2l-1} D_0 \|A_{2l}^{(t)}(\Lambda)\| \cdot \|V_{2n+2-2l}^{(s-t)}(\Lambda)\| \right\} + \tfrac{1}{2} C \|V_{2n+2}^{(s)}(\Lambda)\| \\
&\quad + \tfrac{1}{2} C \sum_{t=0}^{s-1} \Big\{ \|B_{1,2n}^{(t)}(\Lambda)\| \cdot \|V_4^{(s-t)}(\Lambda)\| \\
&\quad + \tfrac{1}{8} \Lambda^2 \|B_{2,2n}^{(t)}(\Lambda)\| \cdot \|\partial_{1,2}^{\mu} \partial_{1,2}^{\mu} V_4^{(s-t)}(\Lambda)\| \\
&\quad + \|B_{3,2m}^{(t)}(\Lambda)\| \cdot \|V_6^{(s-t)}(\Lambda)\| \Big\} .
\end{aligned} \tag{70}
$$

As in lemma (ii), the induction hypothesis bounds the right-hand side of (70) if the $t = 0$ terms vanish. In particular, for $n \geqslant 3$,

$$
\left\| \left(\Lambda \frac{\partial}{\partial \Lambda} + 4 - 2n \right) V_{2n}^{(s)}(p_1, \ldots, p_{2n}, \Lambda) \right\| \leqslant \left(\frac{\Lambda}{\Lambda_0} \right)^2 P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R)) \tag{71}
$$

and with derivatives

$$
\begin{aligned}
&\left\| \left(\Lambda \frac{\partial}{\partial \Lambda} + 4 - 2n \right) \partial_{i_1, j_1}^{\mu_1} \ldots \partial_{i_p, j_p}^{\mu_p} V_{2n}^{(s)}(p_1, \ldots, p_{2n}, \Lambda) \right\| \\
&\quad \leqslant \Lambda^{-p} \left(\frac{\Lambda}{\Lambda_0} \right)^2 P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R)) .
\end{aligned} \tag{72}
$$

Integrating (72) gives

$$
\begin{aligned}
\|\partial_{i_1, j_1}^{\mu_1} \ldots \partial_{i_p, j_p}^{\mu_p} V_{2n}^{(s)}(p_1, \ldots, p_{2n}, \Lambda)\| \leqslant{}& \Lambda^{-p} \left(\frac{\Lambda}{\Lambda_0} \right)^2 P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R)) \\
& \times \int_{\Lambda}^{\Lambda_0} \frac{\mathrm{d}\Lambda'}{\Lambda'} \left(\frac{\Lambda}{\Lambda'} \right)^{2n-p-6} \\
& + \Lambda^{-p} \left(\frac{\Lambda}{\Lambda_0} \right)^{2m-p-4} P^{2s-n-1}(\ln(\Lambda_0/\Lambda_R))
\end{aligned}
$$

(the last term is from the initial value (69))

$$\leqslant \Lambda^{-p}\left(\frac{\Lambda}{\Lambda_0}\right)^2 P^{2s-n}(\ln(\Lambda_0/\Lambda_R)), \tag{73}$$

establishing the induction step down to $n = 3$. As in lemma (ii), eq. (73) for $n = 3$ gives us eq. (71) and (72) for $n = 2$. Eq. (73) then follows for $n = 2$, $p \geqslant 2$ but $V_4(0,0,0,0,\Lambda)$ is known (it is zero by construction) and Taylor's theorem establishes the induction step for all of $n = 2$. Eq. (71) and (72) now follow for $n = 1$; eq. (73) follows for $n = 1$, $p \geqslant 4$. Again, $n = 1$, $p = 0$ and $p = 2$, vanish by construction at zero momentum and we use Taylor's theorem to establish (73) for all of $n = 1$ and the lemma is proven.

Proof of theorem. By definition of V,

$$V(\phi, \Lambda_R, \lambda^R, \Lambda_0) = \Lambda_0 \frac{d}{d\Lambda_0} L(\phi, \Lambda_R, \Lambda_0, \rho^0(\Lambda_R, \lambda^R, \Lambda_0)), \tag{74}$$

so that integrating (74) and using lemma (iii),

$$\begin{aligned} &\|\tilde{L}^{(r)}_{2m}(\Lambda_R, \Lambda_0) - \tilde{L}^{(r)}_{2m}(\Lambda_R, \Lambda'_0)\| \\ &= \Lambda_R^{4-2m}\left\| \int_{\Lambda_0}^{\Lambda'_0} \frac{d\Lambda''_0}{\Lambda''_0} V^{(r)}_{2m}(\Lambda_R, \Lambda''_0) \right\| \qquad (75)\\ &\times \begin{cases} \leqslant \Lambda_R^{4-2m}\left\{\left(\dfrac{\Lambda_R}{\Lambda_0}\right)^2 P^{2r-m}(\ln(\Lambda_0/\Lambda_R)) - \left(\dfrac{\Lambda_R}{\Lambda'_0}\right)^2 P^{2r-m}(\ln(\Lambda'_0/\Lambda_R))\right\}, & r+1-m \geqslant 0, \\ = 0, & r+1-m < 0. \end{cases} \end{aligned}$$

The existence of a limit (46) with the property (47) follows from elementary properties of limits (Cauchy's criterion), completing the theorem.

Ultimately we are interested not in the effective lagrangian but in the n-point functions. A general n-point function $G(\Lambda_0, \Lambda_R, \lambda^R)$ is defined initially with the propagator (11) cut off at Λ_0 and the bare vertices (12), with ρ^0_a defined as a function of λ^R by the renormalization conditions (45). The point of integrating out modes was that we get the identical answer if we use $P(p, \Lambda_R)$ cut off at Λ_R, and the vertices of the effective lagrangian at Λ_R. Thus at order r, $G^{(r)}(\Lambda_0, \Lambda_R, \lambda^R)$ is given by a finite number of terms of the form

$$\int d^{4l}p\, \tilde{L}^{(r_1)}_{2m_1}(\Lambda_R, \Lambda_0)\dots\tilde{L}^{(r_k)}_{2m_k}(\Lambda_R, \Lambda_0) P(p_1, \Lambda_R)\dots P(p_n, \Lambda_R), \tag{76}$$

where $\sum r_i = r$. Thus, at order r in λ^{R},

$$|G^{(r)}(\Lambda_0) - G^{(r)}(\infty)| \leqslant \left(\int \mathrm{d}^{4l}p\, P(p_1, \Lambda_{\mathrm{R}}) \dots P(p_n, \Lambda_{\mathrm{R}}) \right)$$

$$\times \| \tilde{L}^{(r_1)}_{2m_1}(\Lambda_{\mathrm{R}}, \Lambda_0) \dots \tilde{L}^{(r_k)}_{2m_k}(\Lambda_{\mathrm{R}}, \Lambda_0)$$

$$- \tilde{L}^{(r_1)}_{2m_1}(\Lambda_{\mathrm{R}}, \infty) \dots \tilde{L}^{(r_k)}_{2m_k}(\Lambda_{\mathrm{R}}, \infty) \| . \tag{77}$$

The separate integral over the propagators has no Λ^0 dependence and is obviously convergent for $m^2 > 0$, converges at $m = 0$ if we avoid the IR divergent point $p_{\mathrm{external}} = 0$, and converges for $m^2 < 0$ if we shift to the correct minimum. The shift in the last case is no problem, as the theorem for $\tilde{L}(\phi, \Lambda_{\mathrm{R}}, \lambda^{\mathrm{R}}, \Lambda_0)$ implies the same result for $\tilde{L}(\phi + v, \Lambda_{\mathrm{R}}, \lambda^{\mathrm{R}}, \Lambda_0)$. The difference in lagrangians in (77) is bounded by the theorem, so

$$|G^{(r)}(\Lambda_0) - G^{(r)}(\infty)| \leqslant \Lambda_0{}^{2} \times \text{polynomial in } \ln(\Lambda_0/\Lambda_{\mathrm{R}})$$

$$\times \text{finite, } \Lambda_0\text{-independent quantities} \tag{78}$$

and so the n-point functions have limits as $\Lambda_0 \to \infty$, which they approach as $1/\Lambda_0^2$.

5. Discussion and conclusions

The argument leading from the linear equation for $V(\Lambda)$, eq. (27), to renormalizability, eq. (78), is lengthy. We claim, however, that the result was a foregone conclusion, for the reasons discussed following (27). Once loop integrals can be treated naively, as in going from eq. (34) to eq. (35), the difference between the field theory and the toy equations such as in sect. 2 all but disappears. Since the toy equations can be understood even for small finite coupling, we always had a map through the perturbative woods.

Two things that we have come to expect are missing from this proof. The first is Weinberg's theorem, which justifies naive power counting for multiloop Feynman graphs with many potential subdivergences. The second is a discussion of such ideas from graph topology as skeleton expansions and overlapping divergences. The point we have tried to make is that renormalizability is a general property which does not hinge on these particular technical points. In our case, naive power counting for integrals was justified because they were always over the limited range $\Lambda^2 < p^2 < 4\Lambda^2$. The renormalization group equation automatically disentangles the overlapping divergences, for in building an arbitrary Feynman graph with the operations shown in fig. 3, it always constructs the subgraphs with the largest momentum first.

The method here can be applied to composite operator renormalization as well, by including $\varepsilon_i O_i$ (O_i = composite operator) and identifying the terms of order ε in the effective lagrangian. Operators of any dimension can be studied, since by including in the bare L all operators (of given symmetry) up to dimension d, we may separate from V not just the relevant parts, but the subleading parts up to Λ_0^{4-d}.

The proof here extends immediately to any system whose symmetries are preserved by a momentum-space cutoff. The principle is very general, however, and should apply to any physical cutoff, such as Pauli-Villars for abelian gauge theories and the lattice (probably with discrete rescaling) or higher covariant derivatives for non-abelian theories. The main problem is to obtain the renormalization group equation appropriate to the cutoff. It is unclear whether dimensional regularization can be adapted to this purpose, as in that case it is $d-4$ and not the scale μ which plays the role of cutoff. It may actually be possible to treat gauge theories with a momentum-space cutoff*. The point is that a momentum cutoff need not change the physics, as the missing modes can be replaced by effective terms in the action. In terms of the new action, the gauge invariance should still be present in some disguised form.

Our key equation (53) and its component form (60) were derived without using perturbation theory in an essential way. It may be possible to make some progress understanding these at finite coupling, starting perhaps with truncated versions. Of course, for a $\lambda\phi^4$ theory at finite coupling the continuum limit cannot actually be taken, as discussed at the end of sect. 1, but it may be possible to establish more limited results.

I am indebted to C. Arabica, D. Friedan, and J. Preskill for inspiration. I would like to thank S. Coleman, S. della Pietra and N. Sakai for suggestions on the manuscript.

References

[1] K.G. Wilson, Phys. Rev. B4 (1971) 3174, 3184;
K.G. Wilson and J.G. Kogut, Phys. Reports 12 (1974) 75

[2] L.P. Kadanoff, Physics 2 (1966) 263;
J. Glimm and A. Jaffe, Fort. Phys. 21 (1973) 327;
E. Brézin, J.C. Le Guillou and J. Zinn-Justin in Phase transitions and critical phenomena, eds. C. Domb and M.S. Green (Academic Press, London, New York) 1975;
T. Appelquist and J. Carazzone, Phys. Rev. D11 (1975) 2856;
S. Weinberg, Physica 96A (1979) 327

[3] F.J. Dyson, Phys. Rev. 75 (1949) 486, 1736;
A. Salam, Phys. Rev. 82 (1951) 217; 84 (1951) 426

[4] N.N. Bogoliubov and O.S. Parasiuk, Acta Math. 97 (1957) 227;
K. Hepp, Comm. Math. Phys. 2 (1966) 301;
W. Zimmermann, Comm. Math. Phys. 15 (1969) 208

* I would like to thank D. Friedan and S. Shenker for discussions of this point.

[5] C.G. Callan, Jr., Phys. Rev. D2 (1970) 1541;
A. Blaer and K. Young, Nucl. Phys. B83 (1974) 493;
C.G. Callan, Jr., in Methods in field theory, eds. R. Balian and J. Zinn-Justin (North-Holland, 1976)
[6] S. Weinberg, Phys. Rev. 118 (1960) 838;
Y. Hahn and W. Zimmermann, Comm. Math. Phys. 10 (1968) 330
[7] J. Fröhlich, Nucl. Phys. B200[FS4] (1982) 281
[8] S. Weinberg, Phys. Rev. D13 (1975) 974; D19 (1979) 1277;
L. Susskind, Phys. Rev. D20 (1979) 2619;
K. Wilson, Phys. Rev. D3 (1971) 1818