

3

Dilatations

(1971)

1 Introduction

It is an old idea in particle physics that, in some sense, at sufficiently high energies the masses of the elementary particles should become unimportant. In recent years this somewhat vague hope has acquired a more definite form in the theory of scale transformations, or dilatations. These are transformations that would be exact invariances of the world if all elementary particle masses (more generally, all dimensionful couplings) vanished. The hope is that, by studying these approximate symmetries in Lagrangian field theories, we can gain knowledge about how scale transformations behave in the real world, and learn something about those kinematic realms where the effects of masses are indeed unimportant, or at least simply calculable. This is a hope that has been fulfilled for broken chiral symmetries, by the study of models such as the sigma-model.

The purpose of these lectures is to report on the progress of such investigations. As we shall see, there is still much that is obscure. In particular, the connection between the sorts of things I will be talking about and the famous experimental scaling of deep inelastic electroproduction remains to be unravelled. Nevertheless, what has been done so far has already yielded some remarkable theoretical surprises.

Section 2 is a description of the formal Lagrangian theory of scale transformations. Some of the Ward identities that express broken scale invariance are derived, models of Nambu–Goldstone scale-symmetry breaking are discussed, and the connection between scale invariance and conformal invariance is explained. These last two topics will not be returned to in the remainder of these lectures, and a reader who is only interested in the main line of the argument may skip the Sections (2.3 and 2.4) that discuss them.

Section 3 is an investigation of broken scale invariance in renormalized

perturbation theory. We find that the Ward identities derived in the preceding section are lies (in the polite current parlance, they contain anomalies), and we find the true equations that replace them, the Callan–Symanzik equations.

Section 4 explains how, despite the Callan–Symanzik anomalies (in fact, because of them) scale invariance may still be regained in appropriate kinematic regions. This astonishing magic trick is due to Ken Wilson.

My own contributions to this field have been minor. I have learned most of what I know from conversations with Curtis Callan, John Ellis, Roman Jackiw, Kurt Symanzik, Kenneth Wilson, and Bruno Zumino. I would like to express both my gratitude to them and my hope that I have not distorted their ideas too badly.

2 The formal theory of broken scale invariance

2.1 *Symmetries, currents, and Ward identities*

We will deal in these lectures with Lagrangian field theories:¹ that is to say, theories involving a set of fields, which we will assemble into a big vector, ϕ , and whose dynamics are determined by a function of the fields and their first derivatives, $\mathcal{L}(\phi, \partial_\mu \phi)$, called the Lagrangian, via Hamilton's principle:

$$\delta I \equiv \delta \int d^4x \mathcal{L} = 0, \quad (2.1)$$

for solutions of the equations of motion. I is called the action integral.

Let us consider some infinitesimal transformation of the fields:

$$\phi \rightarrow \phi + \delta\phi. \quad (2.2)$$

(We will always suppress the infinitesimal parameter that should properly multiply $\delta\phi$.) Let us suppose that under this transformation

$$\delta I = \int d^4x \Delta, \quad (2.3)$$

where Δ is some function of the fields and their derivatives. (Eq. (2.3) is supposed to be true for general fields, not just for solutions of the equations of motion.) If Δ vanishes, the transformation has no effect on the dynamics and is called a symmetry.

Then one can show from the general formalism of Lagrangian field theory,² without any assumptions other than those stated that:

(a) It is always possible to define an object, j^μ , called 'the current associated with the transformation' such that

$$\partial_\mu j^\mu = \Delta. \quad (2.4)$$

Do not be misled by the notation into assuming that the current is always

a four-vector. Its Lorentz-transformation properties depend upon those of the transformation (2.2). If the transformation is an internal-symmetry transformation (Lorentz scalar), then the current will indeed be a four-vector, but for more complicated cases, it will have more complicated transformation properties. For example, the four-parameter group of space-time translations leads to a set of four currents, which, together, form the components of the energy-momentum tensor.

(b) It is always possible to define a generalization of the ordinary time-ordered product of a string of fields, called the T^* -ordered product. It has the interesting property that

$$\begin{aligned} \frac{\partial}{\partial y^\mu} T^* \langle 0 | j^\mu(y) \phi(x_1) \dots \phi(x_n) | 0 \rangle \\ = T^* \langle 0 | \Delta(y) \phi(x_1) \dots \phi(x_n) | 0 \rangle \\ - i \delta^{(4)} T^* (x_1 - y) \langle 0 | \delta \phi(x_1) \phi(x_2) \dots \phi(x_n) | 0 \rangle \\ - i \delta^{(4)} T^* (x_2 - y) \langle 0 | \phi(x_1) \delta \phi(x_2) \dots \phi(x_n) | 0 \rangle \\ + \dots \end{aligned} \quad (2.5)$$

These equations are called Ward identities.

In the case of internal symmetries (including chiral symmetries) these equations follow directly from the definition of the time-ordered product and the equal-time commutation relations. In more complicated cases, the direct definition of the time-ordered product is ambiguous when two space-time arguments coincide, and the commutators have peculiar terms (Schwinger terms) in them. Nevertheless, there is always a way of removing the ambiguity such that the Ward identities remain the same as in the simplest case.³

A particularly useful consequence of the Ward identities can be obtained by integrating them over all space with respect to the variable y . The integral of the left hand side vanishes by integration by parts. (If there are no massless particles in the theory – these can give surface terms in the parts integral.) Thus we obtain

$$\begin{aligned} \int d^4 y T^* \langle 0 | \Delta(y) \phi(x_1) \dots \phi(x_n) | 0 \rangle \\ = i T^* \langle 0 | \delta \phi(x_1) \dots \phi(x_n) | 0 \rangle \\ + i T^* \langle 0 | \phi(x_1) \delta \phi(x_2) \dots \phi(x_n) | 0 \rangle \\ + \dots \end{aligned} \quad (2.6)$$

These equations are called zero-energy theorems. The reason for the terminology is clear in Fourier space, where the divergence Δ carries zero energy (and zero momentum). The one-soft-pion theorems of current algebra are special cases of (2.6).

2.2 Scale transformations and scale dimensions

Scale transformations, or dilatations, are transformations on space-time of the form

$$\alpha: x \rightarrow e^\alpha x, \quad (2.7)$$

where x is a space-time point and α is a real number.⁴ Like all other transformations of physical interest (e.g. chiral transformations) scale transformations can be implemented in a wide variety of ways in field theories. We will be most interested in theories in which these transformations act linearly on the fields:

$$\alpha: \phi(x) \rightarrow e^{ad} \phi(e^{\alpha x}), \quad (2.8)$$

with d some matrix (In Sect. 2.4, we will discuss briefly some theories in which scale transformations act non-linearly.) The infinitesimal transformation is

$$\delta\phi = (d + x^\lambda \partial_\lambda) \phi. \quad (2.9)$$

For a large class of theories (including all renormalizable field theories) these transformations are symmetries, if all non-dimensionless coupling constants (including the masses) are set equal to zero, and if d is chosen to be a matrix that multiplies all Bose fields by one and all Fermi fields by $\frac{3}{2}$. In realistic models, of course, the masses are not zero, so the symmetry is broken.

As an example of how this works, let us consider the ever-popular model of a pseudoscalar meson interacting with a spin-one-half nucleon through Yukawa coupling. We will divide the Lagrangian into a scale-symmetric part and a scale-breaking part:

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_B, \quad (2.10)$$

where

$$\mathcal{L}_s = i\bar{\psi}\gamma_\mu\partial^\mu\psi + \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + g_0\bar{\psi}\gamma_5\psi\phi - \frac{\lambda_0}{4!}\phi^4, \quad (2.11)$$

and

$$\mathcal{L}_B = -m_0\bar{\psi}\psi - \frac{1}{2}\mu_0^2\phi^2. \quad (2.12)$$

(The subscripts are to remind you that these are bare masses and coupling constants.) As stated, we choose the transformation laws

$$\delta\psi = (\tfrac{3}{2} + x_\lambda\partial^\lambda)\psi \quad (2.13)$$

and

$$\delta\phi = (1 + x_\lambda\partial^\lambda)\phi. \quad (2.14)$$

From these equations, it follows directly that

$$\delta\mathcal{L}_s = (4 + x_\lambda\partial^\lambda)\mathcal{L}_s$$

this vanishes in the integral (2.1) upon integration by parts. On the other hand,

$$\delta \mathcal{L}_B = -(3 + x_\lambda \partial^\lambda) m_0 \bar{\psi} \psi - \frac{1}{2} (2 + x_\lambda \partial^\lambda) \mu_0^2 \phi^2. \quad (2.15)$$

Upon integration by parts, this does not vanish, but becomes

$$\Delta = m_0 \bar{\psi} \psi + \mu_0^2 \phi^2. \quad (2.16)$$

If you have accepted the dogma of Sect. 2.1, you realize that this means that we can define a scale current, s_μ , such that

$$\partial_\mu s^\mu = \Delta. \quad (2.17)$$

This is the local version of the statement that only the masses break scale invariance. Also, without even bothering to explicitly construct s_μ , we can immediately write down the low-energy theorems that follow from broken scale invariance, the Eq. (2.6). I will not bother to write them out, though, until we will need them, at the beginning of Sect. 3.

You have probably noticed that the numbers $\frac{3}{2}$ and 1, which occur in the transformation laws for the fields, are just the dimensions of the fields, in the sense of ordinary dimensional analysis. You may be wondering, therefore, whether what we have been doing is just dimensional analysis, disguised by a fancy formalism. If this is the case, you are wrong: the transformations of dimensional analysis not only scale the dynamical variables of a physical theory (in our case, the fields), they also scale all non-dimensionless numerical parameters (in our case, the masses). Phrased somewhat more abstractly, the transformations of dimensional analysis turn one physical theory into another, different theory (e.g. with different masses), and are always exact symmetries – given the exact solutions to the first theory, they yield the exact solutions to the second. Scale transformations, as we have defined them, are very different animals: they do not change numerical parameters – that is to say, they stay within a given physical theory – and they are not exact symmetries, except in special cases (vanishing masses).

To emphasize this difference, we will call the numbers that occur in field transformation laws like (2.13) and (2.14), ‘scale dimensions’. Since, at the current stage of our investigation, they appear to be identical with the dimensions of the fields in the sense of dimensional analysis, this may seem mere nitpicking. However, I assure you that the distinction will be important in the future.

2.3 *More about the scale current and a quick look at the conformal group*

In the last Section we introduced the scale current, s^μ , but said nothing about its explicit form. It turns out that for a large class of

theories,⁵ it is possible to define an energy–momentum tensor, $\theta^{\mu\nu}$, such that

$$s^\mu = x_\nu \theta^{\mu\nu}. \quad (2.18)$$

This should not be too much of a surprise. After all, the conserved currents associated with other geometrical transformations, such as translations and Lorentz transformations, are also written in terms of the energy–momentum tensor. Also, Eq. (2.18) implies that exact scale invariance is equivalent to the vanishing of the trace of the energy–momentum tensor:

$$\partial^\mu s_\mu = \theta^\mu_\mu = 0. \quad (2.19)$$

This should strike a familiar chord, if you remember that for free electromagnetism (the prototype of a scale-invariant field theory), the energy–momentum tensor is traceless. (This is why, in scalar theories of gravitation, there is no bending of light by the sun; there is nothing for the scalar graviton to couple to.)

On the other hand, Eq. (2.19) is far from self-evident. The energy–momentum tensor $\theta_{\mu\nu}$ is not, in general, the conventional symmetric energy–momentum tensor of Belinfante; it is another conserved symmetric tensor which differs from the Belinfante tensor by extra terms which do not affect the construction of the total four-momentum nor the Lorentz generators, but which are important in the scale current. Also, (2.19) is not valid for a general field theory; a theory allows the construction of an energy–momentum tensor obeying (2.19) only if a certain condition is met. I will first state this (obscure) condition and then attempt to clarify its meaning.

Let $\Sigma^{\mu\nu}$ be the spin matrix, that matrix that occurs in the transformation law of the fields under infinitesimal Lorentz transformations,

$$\delta^{\mu\nu}\phi = [x^\mu\partial^\nu - x^\nu\partial^\mu + \Sigma^{\mu\nu}]\phi.$$

Then, we can find an energy–momentum tensor such that (2.19) is valid if and only if

$$\frac{\partial \mathcal{L}}{\partial(\partial_\nu\phi)} \cdot [g^{\mu\nu}d + \Sigma^{\mu\nu}]\phi = \partial_\nu\sigma^{\mu\nu}, \quad (2.20)$$

where $\sigma^{\mu\nu}$ is some tensor function of the fields and their derivatives. It is easy to check that all renormalizable field theories, indeed all theories of the interactions of fields of spin ≤ 1 where the derivative interactions are of the same form as in renormalizable theories, satisfy (2.21).

This is, as promised, obscure. To understand what is going on, let us consider the case of exact scale invariance, Eq. (2.10). In this case, a little computation reveals a surprise: we can construct from the energy–

momentum tensor, not only the conserved scale current, but four other conserved currents:

$$K^{\lambda\mu} = x^2 \theta^{\lambda\mu} - 2x^\lambda x_\rho \theta^{\rho\mu}. \quad (2.21)$$

$$\partial_\mu K^{\lambda\mu} = 2x_\mu \theta^{\lambda\mu} - 2x_\rho \theta^{\rho\lambda} - 2x^\lambda \theta^\rho_\rho = 0. \quad (2.22)$$

Thus it appears that, for theories for which (2.19) holds, exact scale invariance implies invariance under four other (at the moment mysterious) infinitesimal transformations. This leads to a conjecture: the left hand side of Eq. (2.21) is the change in the Lagrangian under these mysterious infinitesimal transformations (with the possible additions of some terms which vanish as a consequence of exact scale invariance); the condition that it be a divergence is just the condition that the action integral be unchanged.

What can these mysterious transformations be? To gain some insight, let us study the simplest scale-invariant theory, the theory of a free massless scalar field,

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi. \quad (2.23)$$

The equation of motion is the wave equation

$$\partial_\mu \partial^\mu \phi = 0. \quad (2.24)$$

Since this is a free field theory, its complete quantum dynamics is determined by the two-point function,

$$\langle 0 | \phi(y) \phi(x) | 0 \rangle = \frac{1}{2\pi^2} \frac{1}{(x-y)^2}. \quad (2.25)$$

(The ambiguity at the pole is removed by giving x a small imaginary part, lying in the forward light cone).

Now, the wave equation, (2.25), is very similar to the three-dimensional Laplace equation. This equation, as you know, possesses geometrical symmetries beyond the usual Euclidean transformations; to be precise, it is invariant under the three-dimensional inversion of coordinates:

$$I: \mathbf{x} \rightarrow \frac{\mathbf{x}}{x^2}. \quad (2.26)$$

This suggests that the theory of the free massless scalar field may be invariant under the Minkowski-space inversion:

$$I: x \rightarrow -x/x^2. \quad (2.27)$$

(The reason for the minus sign will become clear shortly.)

Let us check this idea. Under the inversion,

$$I: (x-y)^2 \rightarrow \frac{(x-y)^2}{x^2 y^2}. \quad (2.28)$$

Thus, the two-point function is invariant if we define the field to transform in the following way:

$$I: \phi(x) \rightarrow \frac{1}{x^2} \phi\left(\frac{-x}{x^2}\right). \quad (2.29)$$

We still have to check that the statement that the imaginary part of x is in the forward light cone is invariant under the inversion. With no loss of generality, we can take both the real and the imaginary part of x to be in the 0–1 plane. Let us define

$$x_{\pm} = x^0 \pm x^1. \quad (2.30)$$

Then in these coordinates, the statement that the imaginary part is inside the forward cone becomes

$$\text{Im } x_{\pm} > 0. \quad (2.31)$$

This is obviously invariant under the inversion

$$I: x_{\pm} \rightarrow -1/x_{\mp}. \quad (2.32)$$

The inversion (2.28) is a discrete transformation, and is thus of no use in obtaining conserved currents; for these, continuous transformations are needed. However, we can use the inversion to find new continuous invariances, by applying it to the right and left of old ones. For example, if we take the four-parameter group of space-time translations,

$$a: x \rightarrow x + a, \quad (2.33)$$

and apply (2.28) to the right and left of these transformations, we obtain

$$a: x \rightarrow \frac{x - a \cdot x^2}{1 - 2a \cdot x + a^2 x^2}. \quad (2.34)$$

These are called conformal transformations. The associated four infinitesimal transformations are

$$\delta^{\mu} x^{\nu} = -g^{\mu\nu} x^2 + 2x^{\mu} x^{\nu}. \quad (2.35)$$

However, no new transformations are obtained by applying the inversion to the right and left of an infinitesimal Lorentz transformation or dilatation; one just obtains Lorentz transformations and dilatations again.

It is straightforward to verify that this set of infinitesimal transformations (six Lorentz transformations, four translations, four conformal transformations, and one dilatation) is closed under commutation; that is to say, it forms the basis of a fifteen-dimensional Lie algebra. (The only commutator that is not immediately evident is that of a conformal transformation and a translation; this turns out to be the sum of an infinitesimal Lorentz transformation and an infinitesimal dilatation.) The fifteen-parameter group obtained by exponentiating this algebra is called the

conformal group. As we have shown, it can be defined as the connected part of the smallest group of transformations on space-time containing both the Poincaré group and the inversion. As we have also shown, the whole conformal group can be realized as a group of symmetries of the free massless scalar field.

A Lorentz-invariant field theory that is invariant under the infinitesimal conformal transformations (2.36) is automatically also scale-invariant, since the commutator of an infinitesimal conformal transformation and an infinitesimal translation contains a scale transformation. However, there is no corresponding group theoretical reason for the converse to be true; indeed, it is fairly easy to find examples of field theories that are scale-invariant but not conformally invariant.

It is straightforward (but lengthy and tedious) to work out the transformation properties of a field of arbitrary spin under infinitesimal conformal transformations, and to find the condition that a scale-invariant theory be also conformally invariant. This turns out to be Eq. (2.21). This verifies the conjecture we made at the beginning of this section: the four extra conserved currents that appear in the limit of exact scale invariance for a theory for which (2.19) holds are simply the four currents associated with infinitesimal conformal transformations.⁵

For theories of this type (which, I remind you, include all renormalizable field theories) this has an interesting consequence even when scale invariance is broken. In this case, Eq. (2.23) becomes

$$\partial_\mu K^{\lambda\mu} = -2x^\lambda \theta_\mu^\mu. \quad (2.36)$$

Thus we get two sets of low-energy theorems from the general formula (2.6). From broken scale invariance

$$\partial_\mu S^\mu = \theta_\mu^\mu, \quad (2.37)$$

we obtain formulae for the matrix elements of θ_μ^μ at zero momentum transfer, while from Eq. (2.37) we obtain formulae for the first derivative of this object with respect to momentum transfer at zero momentum transfer. This is, in principle, much more information than we have from the corresponding statements in the case of chiral symmetry, the one-soft-pion theorems. These only give us formulae for the matrix elements of the divergences of the axial currents at zero momentum transfer, and tell us nothing about their derivatives.

However, despite this bright promise, we will have, in the remainder of these lectures, so much trouble keeping straight what really happens with the scale-invariance low-energy theorems that we will have very little to say about the conformal ones.

2.4 *Hidden scale invariance*

It has long been known that there are two ways a symmetry that commutes with the Lorentz group (an internal symmetry or a chiral symmetry) can be realized in a quantum field theory:

(1) The symmetry may be manifest. The ground state of the theory (the vacuum) is invariant under the symmetry group, particles arrange themselves into degenerate multiplets corresponding to irreducible representations of the group, and S -matrix elements are invariant. If a small term is added to the Lagrangian to break the symmetry, the masses of the particles in the multiplets split, and the symmetry of S -matrix elements becomes only approximate. This seems to be the case in nature for isospin.

(2) The symmetry may be hidden. (This case, first investigated by Nambu and Goldstone, is usually called ‘spontaneously broken symmetry’, but this terminology is deceptive; the symmetry is not really broken – the currents associated with the symmetry are still conserved and the Ward identities are still valid – it is just that the consequences of the symmetry are less obvious than in the other case.) The ground state of the theory (the vacuum) is not invariant and particles do not arrange themselves into multiplets: however, there do appear a set of special particles, the ‘Goldstone bosons’. These are massless spinless mesons, scalar or pseudoscalar depending on whether the symmetry current is vector or axial. Instead of a symmetric S -matrix, we obtain low-energy theorems, statements relating an S -matrix element for a process involving one zero-energy, zero-momentum Goldstone boson to an S -matrix element with the boson absent. If a small term is added to the Lagrangian to break the symmetry, the Goldstone bosons acquire a small mass and the low-energy theorems require an off-mass-shell extrapolation, become only approximate on the mass shell. This seems to be the case in nature for the chiral transformations of $SU(2) \times SU(2)$; the would-be Goldstone bosons are the pions.

A cheap way to get some feeling for symmetries realized in the Nambu–Goldstone manner is by the study of ‘phenomenological Lagrangians’. These are Lagrangian field theories that embody the Goldstone phenomenon, at least in lowest order of perturbation theory. Frequently these models are hard to take seriously as real physical theories; for example, the non-linear chiral Lagrangians, which have played such a prominent role in investigations of current algebra, are non-renormalizable, so that it is impossible to proceed beyond lowest-order perturbation theory in an unambiguous way. Nevertheless, they have at least two uses: they can offer clues to general theorems that one might try to establish by more careful methods, and they can offer counter-examples to wrong argu-

ments, by fulfilling all the premises but not obeying the supposed consequence. I would like to briefly discuss here a few simple phenomenological Lagrangians for which scale invariance (and conformal invariance) is realized as a Nambu–Goldstone symmetry, rather than as a manifest symmetry.

The basic trick is very simple. We begin with a scalar field χ that transforms in the conventional linear way under scale transformations,

$$\alpha: \chi(x) \rightarrow e^\alpha \chi(e^\alpha x), \quad (2.38)$$

and define a new field, σ , by

$$f^{-1} e^{f\sigma} = \chi, \quad (2.39)$$

where f is a constant with the dimensions of length. The new field now transforms in a non-linear way

$$\alpha: \sigma(x) \rightarrow \sigma(e^\alpha x) + \alpha/f. \quad (2.40)$$

The corresponding infinitesimal transformation is⁶

$$\delta\sigma = x^\lambda \partial_\lambda \sigma + f^{-1}. \quad (2.41)$$

(If you are familiar with non-linear chiral Lagrangians, this transformation law may remind you of the chiral transformation law for the pions, the Goldstone bosons of chiral symmetry, which also has an inhomogeneous term in it.) A scale (and conformally) invariant Lagrangian for the σ is

$$\mathcal{L} = \frac{1}{2} \partial_\mu \chi \partial^\mu \chi = \frac{1}{2f^2} (\partial_\mu e^{f\sigma}) (\partial^\mu e^{f\sigma}). \quad (2.42)$$

We can now use the σ field to make any Lagrangian into a scale-invariant Lagrangian: we simply multiply the scale-breaking parts by appropriate powers of $\exp[f\sigma]$ to make them scale invariant, and add the free σ Lagrangian (2.43). For example, for the meson–nucleon theory discussed in Sect. 2.2, we write

$$\mathcal{L} = \mathcal{L}_s - m_0 \bar{\psi} \psi e^{f\sigma} - \frac{\mu_0^2}{2} \phi^2 e^{2f\sigma} + \frac{1}{2f^2} \partial_\mu e^{f\sigma} \partial^\mu e^{f\sigma}, \quad (2.43)$$

where \mathcal{L}_s is the unchanged scale-invariant part of the original Lagrangian Eq. (2.11). This Lagrangian is manifestly scale-invariant. (It is easy to check that it is also conformally invariant.) Despite this, the meson and the nucleon have non-zero masses in lowest-order perturbation theory. Only the σ is massless; it is the Goldstone boson of scale invariance. (Somewhat surprisingly, it is also the Goldstone boson for the four conformal transformations. A naive generalization of the situation for internal symmetries would lead us to believe that five Goldstone bosons are necessary if we are to have five hidden symmetries, but as this example shows, this is not so.)

In the case of chiral symmetries, phenomenological Lagrangians are also used to study symmetry breaking. Usually the symmetry-breaking term is chosen such that the divergence of the axial-vector current is proportional to the pion field. This is called naive PCAC. (The acronym stands for 'partially conserved axial current'.) The advantage of doing this is that it guarantees that the model will possess, in lowest-order perturbation theory, pion pole dominance of the matrix elements of the divergence of the axial vector currents, real PCAC.

Likewise, in the sort of model we have been considering, it is obviously useful to choose the scale-symmetry-breaking term such that the divergence of the scale current is proportional to the σ field. This might be called PCDC – partially conserved dilatation current. This is easily arranged. For example, we can choose the symmetry-breaking term

$$\mathcal{L}_B = -\frac{m_\sigma^2}{16f^2} [e^{4f\sigma} - 4f\sigma - 1]. \quad (2.44)$$

If we expand this in powers of the σ field, we find that

$$\mathcal{L}_B = -\frac{m_\sigma^2}{2} \sigma^2 + O(\sigma^3). \quad (2.45)$$

This is in proper form (there are no terms linear in the σ field), and we see that m_σ is indeed the σ mass. (In lowest order perturbation theory only, as always.) Just as in Sect. 2.2, we can compute the contribution of this term to the divergence of the scale current:

$$\delta \mathcal{L}_B = x_\lambda \partial^\lambda \mathcal{L}_B - \frac{m_\sigma^2}{16f^2} [4e^{4f\sigma} - 4]. \quad (2.46)$$

Integrating by parts, we find

$$\partial_\mu s^\mu = \theta_\mu^\mu = -\frac{m_\sigma^2}{f^2} \sigma, \quad (2.47)$$

the desired result.

There is obviously much more that can be done along these lines. For example, I have said nothing about the interplay between scale symmetry and chiral symmetry, nor have I derived any experimental consequences of these ideas. (These are most readily obtained by identifying the σ with some observed particle. The most popular candidate is the evanescent ϵ , the scalar isoscalar dipion resonance, last seen somewhere between 700 and 1000 MeV.) For these matters I refer you to the literature.⁷

3 The death of scale invariance

3.1 Some definitions and technical details

Let us return to the low-energy theorems discussed in Sect. 2.1. For simplicity, I will restrict myself to the theory of a single self-interacting scalar meson,

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{\lambda_0}{4!} \phi^4 - \frac{1}{2} \mu_0^2 \phi^2, \quad (3.1)$$

but I will try to conduct the discussion in such a way that the generalization to more complicated theories, such as the meson–nucleon theory of Sect. 2.2, is obvious.

Let us denote the one-particle-irreducible⁸ renormalized Green's function with n external lines by

$$\Gamma^{(n)}(p_1 \dots p_n)$$

where the p s are the momenta carried by the external lines, oriented so they all go inward. (The p s are, of course, not all independent; their sum must be zero.) Likewise, let us denote the one-particle-irreducible Green's function with n external meson lines and one insertion of Δ , the divergence of the scale current (in this case, $\mu_0^2 \phi^2$) by

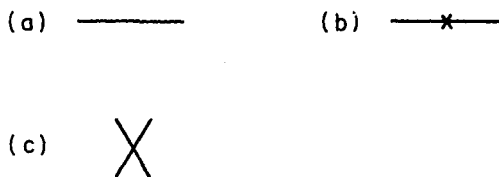
$$\Gamma_\Delta^{(n)}(k; p_1 \dots p_n),$$

where the p s are defined as before, and k is the momentum carried by the insertion. Then it is a trivial exercise in Fourier transforms to see that in this case the low-energy theorem (2.6) becomes

$$\left(\sum_{r=1}^{n-1} p_r \cdot \frac{\partial}{\partial p_r} + nd - 4 \right) \Gamma^{(n)}(p_1 \dots p_n) = -i \Gamma_\Delta^{(n)}(0; p_1 \dots p_n). \quad (3.2)$$

where d is the scale dimension of the scalar field. The analysis of the preceding section has told us that d is one, of course, but I would like to suppress this information momentarily, both to make it easy for you to see the generalization to Fermi fields, for which d is $\frac{3}{2}$, and for another reason, which will become clear shortly.

Equation (3.2) is fairly easy to understand. The first two terms on the left are just the transformation law of the fields, Eq. (2.9) written in momentum space; the sum only runs over $n-1$ momenta because only $n-1$ of the momenta are independent. The factor of four appears because in passing from the Fourier transform of a T^* -product to the conventionally defined $\Gamma^{(n)}$, we must factor out a four-dimensional δ -function, which leaves the four behind as it passes through the differential operator. The right-hand side of the equation is just the left-hand side of (2.6).

Fig. 1

Just to check that we have not made a sign error, let us check (3.2) in lowest-order perturbation theory. To zeroth order, $\Gamma^{(2)}$ is given by Fig. 1(a), and $\Gamma_{\Delta}^{(2)}$ is given by Fig. 1(b), where the cross denotes the mass insertion. Thus Eq. (3.2) becomes

$$\left(p \cdot \frac{\partial}{\partial p} - 2\right)(-i)(p^2 - \mu^2) = -i(2\mu^2), \quad (3.3)$$

which is correct. Likewise, to first order, $\Gamma^{(4)}$ is given by Fig. 1(c), and there is no contribution to $\Gamma_{\Delta}^{(4)}$. Thus Eq. (3.2) becomes

$$\sum_{i=1}^3 \left(p_i \cdot \frac{\partial}{\partial p_i}\right)(-i\lambda) = 0 \quad (3.4)$$

which is again correct.

We can write (3.2) in another form, which will be useful in the sequel. Let us trade the ps for a set of variables consisting of

$$s = \sum_{r=1}^n p_r^2, \quad (3.5)$$

and the dimensionless kinematic variables $p_i \cdot p_j / s$. Then ordinary dimensional analysis tells us that $\Gamma^{(n)}$ is of the form

$$\Gamma^{(n)} = s^{(4-n)/2} F^{(n)}\left(\frac{s}{\mu^2}, \lambda, \frac{p_i \cdot p_j}{s}\right), \quad (3.6)$$

where μ is the renormalized mass, and λ is the renormalized coupling constant. (Remember, λ is dimensionless.) From this it is easy to see that (3.2) can equivalently be written as

$$\left[\mu \frac{\partial}{\partial \mu} + n(1-d)\right] \Gamma^{(n)}(p_1 \dots p_n) = i\Gamma_{\Delta}^{(n)}(0; p_1 \dots p_n). \quad (3.7)$$

3.2 *A disaster in the deep Euclidean region*

The Euclidean region is that region in multi-particle momentum space in which all four-momenta are Euclidean; that is to say, they all have real space parts and imaginary time parts. The deep Euclidean region is that part of the Euclidean region in which the magnitude of s

(defined by Eq. (3.5)) gets very large, while the dimensionless variables $p_i \cdot p_j/s$ stay fixed; furthermore, no partial sum of the ps is zero. The deep Euclidean region is the maximally unphysical limit in which to study Green's functions: all external lines are far off the mass shell, and, furthermore, no matter how the diagram is cut in two, the momentum transferred between the two halves is far off the mass shell.

There are famous bounds on the behaviour of Feynman amplitudes in the deep Euclidean region, first established by Weinberg.⁹ For the functions that appear in Eq. (3.7), these Weinberg bounds say that $\Gamma^{(n)}$ grows no faster than $s^{(4-n)/2}$, times a polynomial in $\ln(s/\mu^2)$, to any finite order in renormalized perturbation theory. (The coefficients in the polynomial are, in general, functions of the coupling constant and of the dimensionless variables $p_i \cdot p_j/s$. Also, the order of the polynomial grows with the order of perturbation theory.) Likewise, $\Gamma_\Delta^{(n)}$ grows no faster than $s^{(2-n)/2}$, again times a polynomial in $\ln(s/\mu^2)$. Crudely, the reason why $\Gamma_\Delta^{(n)}$ grows less rapidly than $\Gamma^{(n)}$ is that, in the deep Euclidean region, all internal momenta are getting large; adding a mass insertion adds an internal propagator, which knocks out one power of s .

We can now combine this with the broken-scale-invariance low-energy theorems, Eq. (3.7), to get a much more powerful statement about the asymptotic behaviour of the Green's functions than is given by the Weinberg bounds alone. For the Weinberg bounds tell us that in the deep Euclidean region, the right-hand side of (3.7) is negligible compared to the individual terms on the left-hand side. Thus, we can neglect it, and obtain an equation for the asymptotic form of $\Gamma^{(n)}$,

$$\left[\mu \frac{\partial}{\partial \mu} + n(1-d) \right] \Gamma_{as}^{(n)} = 0, \quad (3.8)$$

where the subscript indicates the asymptotic form in the deep Euclidean region. Or, since d is one,

$$\mu \frac{\partial}{\partial \mu} \Gamma_{as}^{(n)} = 0. \quad (3.9)$$

That is to say, there are no logarithmic factors in every order of renormalized perturbation theory. This is indeed a powerful statement; unfortunately it is also a false one; anyone who has ever done any Feynman calculation involving a closed loop knows that the logarithms are in fact present. Therefore, Eq. (3.7), from which we deduced the false statement (3.9), must itself be false. In the current technical language the Ward identities of broken scale invariance must contain anomalies. Phrased more straightforwardly, *the entire theoretical structure of Sect. 2 is a lie!*

I want to make the logic of this argument clear. I am not saying that the formal theory of broken scale invariance is wrong because it makes asymptotic predictions that differ from the asymptotic behaviour found in perturbation theory; only a madman would take the asymptotic behaviour of perturbation theory so seriously.¹⁰ I need only assume that perturbation theory properly gives the successive derivatives of the Green's functions with respect to the coupling constant at zero coupling constant. This seems to me to be a very weak and extremely reasonable assumption, at least for renormalizable field theories. Under this assumption, if Eq. (3.7) is generally valid, it must be true order by order in perturbation theory. If the right- and left-hand sides of this equation are equal in a fixed order in perturbation theory, they must be equal in the deep Euclidean limit. They are not.

The technical reason for the occurrence of the anomalies is not difficult to understand: the formal canonical manipulation required to prove Ward identities is justified only if we introduce a cutoff to remove the divergences from the theory. However, this does us no good unless the cut-off is chosen in such a way that the cutoff theory still obeys the Ward identities. For such familiar cases as quantum electrodynamics or the sigma model, for example, this condition presents no difficulties; it is easy to introduce a cutoff in such a way that the relevant equations (gauge invariance in one case and PCAC and current algebra in the other) remain true. For scale invariance, though, the situation is hopeless; any cutoff procedure necessarily involves a large mass, and a large mass necessarily breaks scale invariance in a large way. This argument does not show that the occurrence of anomalies is inevitable, but it does show that there is no reason to believe it is impossible.

3.3 *Anomalous dimensions and other anomalies*

Last year, Roman Jackiw and I got interested in these anomalies and decided to get some information about them by the most simple-minded method imaginable. In meson–nucleon theory (the Lagrangian (2.10)), we simply computed separately the right- and left-hand sides of the low-energy theorem (3.7), to lowest non-trivial order in perturbation theory, order g^2 , to see how they differed. We found² that, to this order, all anomalies could be absorbed in a change in the scale dimension of the fields. In particular, for the meson field, Eq. (3.7) is changed from a falsehood to a truth if we replace the naive value of d (one) by

$$d = 1 + \frac{g^2}{8\pi^2}. \quad (3.10)$$

For the self-energy operator, for example, the solution to Eq. (3.8) becomes

$$\Gamma_{as}^{(2)} \sim s(s/\mu^2)^{1-d}. \quad (3.11)$$

Expanding the exponent and discarding terms of order g^4 and higher, we find

$$\Gamma_{as}^{(2)} \sim s - \frac{g^2}{8\pi^2} s \ln(s/\mu^2). \quad (3.12)$$

This is the correct asymptotic behaviour, to this order in perturbation theory. Similar remarks apply to the nucleon field; here d must be changed from its naive value ($\frac{3}{2}$) to

$$d = 3/2 + \frac{g^2}{32\pi^2}. \quad (3.13)$$

(As it happens, these changes in d not only fix up (3.7), but also take care of all the anomalies in the Ward identities for arbitrary momentum transfers.)

This phenomenon is frequently described by saying that the fields acquire anomalous dimensions. This is a slightly misleading way of putting things, for it tempts us to confuse scale dimensions and dimensions in the sense of dimensional analysis, and to think that something counter to common sense has occurred. It is only the scale dimensions of the fields that have changed as a result of the interactions; the dimensions in the sense of dimensional analysis remain firmly fixed at one and $\frac{3}{2}$. I emphasize again that there is no logical connection between these two sets of numbers; the (discredited) analysis of Sect. 2 is the only thing that ever led us to believe they were equal.

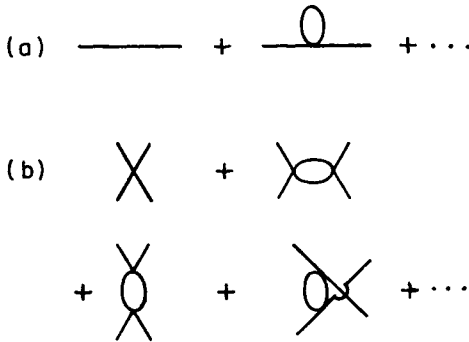
It would be very pleasant if this phenomenon persisted in higher orders of perturbation theory, if all anomalies could be absorbed into a redefinition of the scale dimension. If this were the case, then we would have a correct theory of broken scale invariance with almost as much predictive power as the false theory of Sect. 2; the only price would be the introduction of a new, dynamically determined parameter for each field in the theory, its anomalous dimension.

Unfortunately, this is not what happens. To see this, let us return to the theory of a self-interacting meson field, Eq. (3.1), and assume that Eq. (3.8) is valid in all orders of perturbation theory, except that d is not one. Then it is easy to see that

$$\mu \frac{\partial}{\partial \mu} [\Gamma_{as}^{(4)}/(\Gamma_{as}^{(2)})^2] = 0. \quad (3.14)$$

That is to say, although both the numerator and denominator of this expression have factors of $\ln(s/\mu^2)$ in their asymptotic expansions, the

Fig. 2



logarithms cancel in the ratio, order by order in perturbation theory. Let us check this prediction by calculating the ratio to order λ^2 . The relevant diagrams for $\Gamma^{(2)}$ are shown in Fig. 2(a); the second diagram is just a constant, cancelled by mass renormalization. Thus, we obtain, in the deep Euclidean region,

$$\Gamma_{as}^{(2)} \sim s + O(\lambda^2). \quad (3.15)$$

Figure 2(b) shows the relevant diagrams for $\Gamma^{(4)}$. Every child knows that the three second-order diagrams grow logarithmically for high s ; thus we obtain

$$\Gamma_{as}^{(4)} \sim \lambda + a\lambda^2 \ln(s/\mu^2) + b\lambda^2. \quad (3.16)$$

where a and b are constants. (Actually, b is a function of the dimensionless kinematic variables, but, for our purposes, that is as good as being a constant.) Putting this together we find

$$\Gamma_{as}^{(4)}/(\Gamma_{as}^{(2)})^2 \sim \lambda s^{-2} + a\lambda^2 s^{-2} \ln(s/\mu^2) + b\lambda^2 s^{-2} + O(\lambda^3). \quad (3.17)$$

This is in contradiction to our prediction; thus there must be further anomalies in addition to anomalous dimensions.

3.4 *The last anomalies: the Callan–Symanzik equations*

We are lost in a dark wood. Whenever we look at a more complicated situation, investigate higher orders of perturbation theory, new anomalies appear. Happily, there exists a way to order this chaos, discovered by Curtis Callan and Kurt Symanzik,¹¹ who independently found the true equations that replace the false low-energy theorems (3.7). These equations are surprisingly compact. For the theory of a self-interacting meson field, for example, they involve only two dynamically determined parameters. This is only one more than there would be if the only anoma-

lies were anomalous dimensions.¹² I will now derive the Callan–Symanzik equations for this theory.

The simplest place to begin is in unrenormalized perturbation theory: here one computes the unrenormalized Green's functions (which I will distinguish by a subscript u) in a power series in the bare charge, with the bare mass held fixed. I remind you of the relation between the renormalized Green's functions, which we have been using until now, and the unrenormalized ones

$$\Gamma^{(n)}(p_1 \dots p_n) = (Z_3)^{n/2} \Gamma_u^{(n)}(p_1 \dots p_n). \quad (3.18)$$

where Z_3 is the wave-function renormalization constant. Of course, unlike renormalized perturbation theory, unrenormalized perturbation theory is full of divergences. To control these, we will imagine that we have put a cut-off in the theory in some way, say by modifying the Feynman propagator in the standard manner. Once we re-express things in terms of the physical mass and coupling constant, and perform the multiplicative renormalization (3.18), everything is cutoff-independent in the limit of high cutoff; this is just the statement that (3.1) is a renormalizable Lagrangian. I will denote the unrenormalized Green's function with one mass ($\mu_0^2 \phi^2$) insertion by $\Gamma_{u\Delta}^{(n)}(k; p_1 \dots p_n)$. Then it is easy to see, just by looking at diagrams, that

$$i\Gamma_{u\Delta}^{(n)}(0; p_1 \dots p_n) = \mu_0 \frac{\partial}{\partial \mu_0} \Gamma_u^{(n)}(p_1 \dots p_n). \quad (3.19)$$

This is the unrenormalized version of the low-energy theorem (3.7). As it stands, it is useless, since the differential operator cannot be turned into a scaling operator; the unrenormalized Green's functions depend on another parameter besides μ_0 with the dimensions of a mass, the cutoff.

Simple power counting¹³ shows that if we define

$$\Gamma_{\Delta}^{(n)}(k; p_1 \dots p_n) = Z(Z_3)^{n/2} \Gamma_{u\Delta}^{(n)}(k; p_1 \dots p_n), \quad (3.20)$$

then, for an appropriate (cutoff-dependent) choice of the constant Z , we can make the left-hand side of (3.20) cutoff-independent in the limit of high cutoff. This condition leaves Z undetermined up to a finite (cutoff-independent) multiplicative factor. We will choose this factor later, to make our final equations look as simple as possible.

Putting all this together, we find the consequence of (3.19) for the renormalized Green's functions,

$$i\Gamma_{\Delta}^{(n)}(0, \dots) = Z\mu_0 \frac{\partial}{\partial \mu_0} \Gamma^{(n)}(\dots) - \frac{n}{2} Z\mu_0 \frac{\partial \ln Z_3}{\partial \mu_0} \Gamma^{(n)}(\dots). \quad (3.21)$$

We have not yet finished, since we have the derivative of a renormalized Green's function with respect to a bare mass, and a renormalized Green's

function is cutoff-independent only when expressed in terms of renormalized masses and coupling constants. This is easily taken care of with the aid of the chain rule of differentiation:

$$\left[\left(Z\mu_0 \frac{\partial \mu}{\partial \mu_0} \right) \frac{\partial}{\partial \mu} + \left(Z\mu_0 \frac{\partial \lambda}{\partial \mu_0} \right) \frac{\partial}{\partial \lambda} - \frac{n}{2} \left(Z\mu_0 \frac{\partial \ln Z_3}{\partial \mu_0} \right) \right] \Gamma^{(n)}(\dots) \\ = i\Gamma_{\Delta}^{(n)}(0, \dots). \quad (3.22)$$

where all derivatives are taken with fixed cutoff. The only terms in these equations which retain a possible trace of cutoff-dependence are the three terms in parentheses. These, however, are constants, independent of n and the momenta. Therefore, it is easy to show, by evaluating the equations at three independent points and solving for these terms as functions of cutoff-independent quantities, that they are also cutoff-independent, in the limit of high cutoff. (An especially simple choice is the three points at which the renormalized mass, renormalized charge, and scale of the renormalized field are defined.)

We now choose Z (until now undetermined up to a finite multiplicative constant) such that

$$Z\mu_0 \frac{\partial \mu}{\partial \mu_0} = \mu, \quad (3.23)$$

and define

$$\beta = Z\mu_0 \frac{\partial \lambda}{\partial \mu_0}, \quad (3.24)$$

and

$$\gamma = \frac{1}{2} Z\mu_0 \frac{\partial \ln Z_3}{\partial \mu_0}. \quad (3.25)$$

(Dimensional analysis shows that these functions can only depend on λ .) We thus obtain

$$\left[\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} - n\gamma(\lambda) \right] \Gamma^{(n)}(\dots) = i\Gamma_{\Delta}^{(n)}(0, \dots). \quad (3.26)$$

These are the Callan–Symanzik equations.

Some remarks should be made about these equations:

(1) As promised, these are the equations that replace the naive low-energy theorem (3.7). If β were zero, (3.25) would be identical to (3.7), except that the scale dimension, d , would be anomalous:

$$d = 1 + \gamma. \quad (3.27)$$

Unfortunately, as we have already seen, β is not zero; there are further

anomalies beyond anomalous dimensions. A direct computation¹⁰ shows that

$$\beta = \frac{3\lambda^2}{16\pi^2} + O(\lambda^3). \quad (3.28)$$

(It will be important in the sequel that this first non-zero term in the perturbation expansion of β is positive.)

(2) The generalization to more complicated field theories is obvious: For every dimensionless coupling constant, there is a β -like term, and, for every field, there is a γ -like term. Thus, for example, in meson–nucleon theory, if we denote a Green’s function with n nucleon fields, n antinucleon fields, and m meson fields, by $\Gamma^{(n,n,m)}$, the Callan–Symanzik equations are

$$\left[\mu \frac{\partial}{\partial \mu} + m \frac{\partial}{\partial m} + \beta_1 \frac{\partial}{\partial \lambda} + \beta_2 \frac{\partial}{\partial g} - 2n\gamma_1 - m\gamma_2 \right] \Gamma^{(n,n,m)}(\dots) \\ = i\Gamma_{\Delta}^{(n,n,m)}(0; \dots)$$

where the β s and the γ s are functions of λ , g and μ/m .

(3) Likewise, if we attempt to study Green’s functions for objects other than canonical fields (for example, conserved or partially-conserved currents), this will, in general, change the structure of the γ -terms, which refer to a particular field. The β -terms, though, which do not make reference to a choice of fields, but only to the underlying dynamics, will remain unchanged.

(4) Unlike the low-energy theorems of current algebra, which they so closely resemble, the Callan–Symanzik equations are practically useless for low-energy phenomenology. It is the β -terms that make the difference. A current-algebra low-energy theorem is useful because it expresses one Green’s function in terms of another in a way that does not depend on strong-interaction dynamics. The Callan–Symanzik equations, on the other hand, express one Green’s function in terms of another *and* its derivatives with respect to coupling constants. If you know how to compute *these*, you have already solved the strong-interaction dynamics, and there is no reason for you to be piddling around with low-energy theorems.

(5) Finally, there remains the possibility that β has a zero, and that, for some reason (the bootstrap?) the real value of λ is at this zero. (This is a speculation of Kenneth Wilson, among others). If this is the case, we would regain the naive theory of scale invariance (with anomalous dimensions). There is very little to be said for or against this possibility from a study of perturbation theory alone.

4 The resurrection of scale invariance

4.1 The renormalization group equations and their solution

By the arguments we used in Sect. 3.2, in the deep Euclidean region, we can neglect the right-hand side of the Callan–Symanzik equations. Thus, for the case of a self-interacting meson field, we obtain the following differential equations for the asymptotic forms of the Green's functions:

$$\left[\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} - n\gamma(\lambda) \right] \Gamma_{as}^{(n)} = 0. \quad (4.1)$$

These equations are old friends; they are the equations associated with the so-called renormalization group, first devised by Gell-Mann and Low to study quantum electrodynamics, and later extended to general renormalizable field theories.¹⁴

The renormalization group Eq. (4.1) can be derived much more simply than the Callan–Symanzik equations. As I stated earlier, the deep Euclidean region is the maximally unphysical region; all external lines and all momentum transfers are far from the mass shell. One's first thought would be that in this regime all memory of the actual meson mass would be lost, and all Green's functions would be independent of the mass. However, this is too naive: the objects we are studying are Green's functions for renormalized fields, expressed as functions of the renormalized coupling constant; both the normalization of the field and the value of the renormalized coupling constant are defined on the mass shell; these quantities remember the mass shell no matter how far we flee into the deep Euclidean region. Therefore, the correct statement is that all memory of the actual value of the mass is lost, except for that which is contained in the scale of the fields and the value of λ . In other words, in the deep Euclidean region, a small change in the mass can always be compensated for by an appropriate small change in λ and an appropriate small rescaling of the fields. The equations (4.1) are just the mathematical expression of this statement.

This argument seems to me to owe very little to asymptotic estimates derived from perturbation theory; therefore I will assume from now on that the renormalization group equations are valid independent of perturbation theory, and try to use them to get information about the asymptotic forms of Green's functions in a non-perturbative way.

The renormalization group equations are fairly easy to solve. To motivate the solution, let us consider a similar equation with a direct dynamical interpretation:

$$\frac{\partial \rho}{\partial t} + v(x) \frac{\partial \rho}{\partial x} = L(x)\rho, \quad (4.2)$$

this is an equation of hydrodynamic type. We can think of ρ as the density of a population of bacteria moving with a fluid along a pipe. The fluid has a velocity $v(x)$, a known function of position along the pipe. As the bacteria move along, they are subjected to a changing illumination, $L(x)$, which determines their rate of reproduction.

To solve this equation is now trivial. The solution proceeds in two steps. First we find the position, x' , at time t , of an element of the fluid that is at position x at time zero. That is to say, we solve the ordinary differential equation

$$\frac{dx'(x, t)}{dt} = v(x'), \quad (4.3)$$

with the boundary condition

$$x'(x, 0) = x. \quad (4.4)$$

Because of the time-translation invariance of the equations, this function also tells us the position at time t_1 of an element of the fluid that was at position x at time t_2 , to wit, $x'(x, t_1 - t_2)$. We now find the bacterial density by integrating along a fluid element as it travels down the pipe. Thus we obtain

$$\rho(x, t) = f(x'(x, -t)) \exp \left[\int_{-t}^0 dt' L(x'(x, t')) \right],$$

where f is an arbitrary function, the density of bacteria in the fluid element at the starting time, $t = 0$, as a function of its position at that time, $x'(x, -t)$.

We can now leave our hydrodynamic-bacteriological analogy and return to the renormalization group Eq. (4.2). Making the obvious substitutions, we see that the general solution of (4.2) is obtained by first solving the ordinary differential equation

$$\frac{d\lambda'(\lambda, t)}{dt} = \beta(\lambda'), \quad (4.5)$$

with the boundary condition

$$\lambda'(\lambda, 0) = \lambda. \quad (4.6)$$

The general solution of (4.1) is then

$$\begin{aligned} \Gamma_{as}^{(n)} = & s^{(4-n)/2} f^{(n)} \left(\lambda' \left(\lambda, \frac{1}{2} \ln \frac{s}{\mu^2} \right), \frac{p_i \cdot p_j}{s} \right) \\ & \times \exp \left[-n \int_0^{\frac{1}{2} \ln(s/\mu^2)} dt \gamma(\lambda'(\lambda, t)) \right]. \end{aligned} \quad (4.7)$$

where $f^{(n)}$ is an arbitrary function. Here we have used dimensional analysis, Eq. (3.6), to explicitly restore the dependence on the variables

that do not enter into (4.2), the large variable s and the dimensionless kinematic variables.

Just to check that we have not made a sign error, let us evaluate (4.7) under the assumption that we are actually at a zero of β , as conjectured at the end of Sect. 3. In this case, the solution of the ordinary differential equation is trivial, $\lambda' = \lambda$, and (4.7) reduces to

$$\Gamma_{as}^{(n)} = s^{(4-n)/2} \left(\frac{s}{\mu^2} \right)^{-n\gamma(\lambda)/2} f^{(n)} \left(\lambda, \frac{p_i \cdot p_j}{s} \right). \quad (4.8)$$

This is the correct answer in this special case; everything scales as it would naively, except that the scale dimension is anomalous.

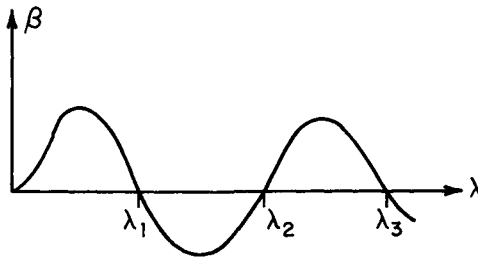
4.2 *The return of scaling in the deep Euclidean region*

The final result of the last section, Eq. (4.7) evidently has a lot of content. It tells us that the asymptotic form of a Green's function in the deep Euclidean region, which *a priori* could depend in an arbitrary way on the dimensionless variables λ and $\ln s/\mu^2$, in fact depends in an arbitrary way only on a certain function of these variables, λ' , which is, in turn, completely determined in terms of the function $\beta(\lambda)$ by Eqs. (4.5) and (4.6). This is useful information as it stands; for example, it has been used to enormously simplify the summation of leading logarithms on the deep Euclidean region. However, as was pointed out by Wilson,¹⁵ if we make some (apparently very mild) additional assumptions, we get a surprisingly greater amount of information.

These additional assumptions are: (1) The functions $f^{(n)}$ appearing in Eq. (4.7) are continuous functions of the variable λ' . Likewise, β and γ are continuous functions of the variable λ . This is certainly mild, and is trivially true in any finite order of perturbation theory, in which these functions are polynomials in these variables. (2) The function $\beta(\lambda)$ has a zero some place along the positive real axis. It is very difficult to say whether this is a plausible or implausible assumption from perturbation theory alone, since any finite order of perturbation theory merely tells us the behaviour of β near the origin, and gives us no information about the presence of zeros away from the origin. It is certainly much weaker than the conjecture we toyed with at the end of Sect. 3, that β had a zero *and* that this zero was the physical coupling constant.

The assumed behaviour of β is shown in Fig. 3. The quadratic zero at the origin is not part of the assumption; that can be established by direct perturbative computation (see eq. (3.27)). The zero at λ_1 is an assumption. The two subsequent zeros at λ_2 and λ_3 are not necessary to the argument, but they lead to interesting results if they are assumed; I have put them in

Fig. 3



the figure so I can talk about them when I have finished the main argument.

Now, let us assume that the physical value of λ lies between zero and λ_1 . Then, when we integrate the differential equation

$$\frac{d\lambda'(\lambda, t)}{dt} = \beta(\lambda'), \quad (4.5)$$

starting from the boundary condition

$$\lambda'(\lambda, 0) = \lambda, \quad (4.6)$$

we will find that λ' is a monotonically increasing function of t , for β is positive in this region. Indeed, λ' stays monotonically increasing for all t , for its derivative can change sign only when λ' exceeds λ_1 , which it cannot do because $\beta(\lambda_1)$ is zero. In other words, λ' asymptotically approaches λ_1 from below as t goes to infinity,

$$\lim_{t \rightarrow \infty} \lambda'(\lambda, t) = \lambda_1. \quad (4.9)$$

This is the essential point: under the stated assumptions, for a range of values of λ , λ' approaches a fixed limit as t goes to infinity. Furthermore, the value of this limit is independent of the value of λ ; as t goes to infinity, λ' loses all memory of where it started.

Likewise, if λ is between λ_1 and λ_2 , the next zero of β (although it is not essential to the argument such a second zero exist), λ' monotonically approaches λ_1 from above. Continuing along the real axis, λ_2 is an exceptional point; if λ equals λ_2 , λ' also equals λ_2 for all t . If λ is between λ_2 and λ_3 , λ' approaches λ_3 from below as t goes to infinity, etc. In terms of the fluid analogy we discussed earlier, the odd zeros of β (λ_1 and λ_3) are sinks; everything is eventually drawn into them.

Now let us return to the case ($0 < \lambda < \lambda_2$) in which the limiting value of λ is λ_1 . Inserting this limit in (4.7), we obtain 'the asymptotic form of the asymptotic form',

$$\lim_{s \rightarrow \infty} \Gamma_{as}^{(n)} = s^{(4-n)/2} f^{(n)} \left(\lambda_1, \frac{p_i p_j}{s} \right) \times \left(\frac{s}{\mu^2} \right)^{-n\gamma(\lambda_1)/2} K^n, \quad (4.10)$$

where the constant K is given by

$$K = \exp \int_0^\infty dt [\gamma(\lambda_1) - \gamma(\lambda'(\lambda, t))]. \quad (4.11)$$

(We will later argue that this integral is convergent.) But this is just simple scaling behavior again! Indeed, it is even simpler scaling behaviour than we thought we had back in Sect. 3.3, when we thought for a while that the only anomalies were anomalous dimensions. For if that were the case, the anomalous dimensions, the exponents that would appear in the asymptotic form (4.10), would depend continuously on λ . In fact, as (4.10) shows, the anomalous dimensions that appear in the asymptotic form are *independent* of λ , for a range of λ ($0 < \lambda < \lambda_2$), as is the whole asymptotic form, aside from the constant K . The β anomalies, which complicate things terribly at low energies, simplify things enormously in the deep Euclidean region. What a wonderful reversal! (This praise of the β -terms is justified, of course, only if our critical assumption – the existence of a zero in β – is true. If it is not, the β -terms remain troublemakers, at high energies as well as low.)

If we are willing to assume that the relevant functions are differentiable as well as continuous, then we can obtain an estimate of the error in (4.10), for then we can see how rapidly λ approaches λ_1 . In the neighborhood of λ_1 , eq. (4.5) can be approximated by

$$\frac{d\lambda'}{dt} = a(\lambda_1 - \lambda'), \quad (4.12)$$

where a is a positive number,

$$a = - \left. \frac{d\beta}{d\lambda} \right|_{\lambda_1}. \quad (4.13)$$

The solution to (4.12) is

$$\lambda' = \lambda_1 + c \exp[-at], \quad (4.14)$$

where c is a constant. This immediately implies the convergence of the integral (4.11). Also, after some straightforward differentiation, it leads to the statement that the error in (4.10) is of the order $(s/\mu^2)^{-a/2}$, compared to the terms retained. Without a knowledge of a , it is impossible to say how small this is, but at least it is a good solid power, not just some logarithms. (All of this presumes, of course, that the zero at λ_1 is a single zero, as shown in Fig. 3. If it were a double or higher-order zero, the terms neglected in (4.10) would just be down by logarithms compared to the terms retained, and the limit would be much less interesting.)

4.3 *Scaling and the operator product expansion*

We have already heard much about the operator product expansion at this school;¹⁶ therefore, rather than explain it in detail here, I will merely summarize its essential features. The operator product expansion was first introduced by Wilson¹⁷ as a conjecture; it is an asymptotic expansion for the product of two local operators as their space-time arguments coincide, as a sum of local operators with *c*-number coefficients. In equations,

$$A(x)B(y) = \sum_n f_n(x-y) O_n\left(\frac{x+y}{2}\right), \quad (4.15)$$

where *A* and *B* are two arbitrary local operators, *x* and *y* are the two arguments that are being brought close together, O_n is the infinite string of local operators, and f_n are the *c*-number coefficients. (We use the equality sign for notational simplicity, even though the series is supposed to be only asymptotic.) Wilson also conjectured that at small distances scale invariance became exact. That is to say that it was possible to assign a dimension $d(A)$, $d(B)$, etc. to every operator occurring in Eq. (4.15), such that for small values of $(x-y)$, f_n became a homogeneous function of $(x-y)$, of order $d(O_n) - d(A) - d(B)$. (We will refer to this statement as Wilson's rule.) In Wilson's original scheme, these dimensions could be anomalous (not equal to the dimensions of the operators in the sense of dimensional analysis), and were not defined in any way outside the operator product expansion; they were simply the numbers that made Wilson's rule work.

The obvious questions to ask are: (1) Is the operator product expansion true? (2) If so, is Wilson's rule true? (3) If so, are the anomalous dimensions that enter into Wilson's rule the same as the anomalous dimensions that govern the asymptotic forms of Green's functions in the deep Euclidean region? I will now give an ingenious argument due to Kurt Symanzik¹⁸ that says the answer to all three questions is yes. The argument does require some speculative assumptions, but they are the same assumptions (the existence of a zero in β , etc.) that we used to get simple scaling behaviour in the deep Euclidean region; no additional assumptions are needed.

For simplicity, let us restrict ourselves to the easiest case, in which the two operators that approach each other are two scalar fields in the self-interacting meson theory we have already discussed so much. In this case, we might expect the first two terms in (4.15) to be

$$\phi(x)\phi(y) = f_0(x-y)I + f_1(x-y)\phi^2\left(\frac{x+y}{2}\right) + \cdots \quad (4.16)$$

where I is the identity operator, and ϕ^2 is the renormalized square of the field. (Let me remind you that this is nothing so simple as the product of ϕ and ϕ ; it is a complicated object defined as a limit in a cut-off theory.)¹³

In free-field theory, it is easy to establish this equation. In this case there are no anomalous dimensions, the renormalized ϕ^2 operator is just the normal-ordered product, and it is easy to check that for small arguments f_0 is an inverse square and f_1 is a constant. For the interacting theory, (4.16) has been established to any finite order in renormalized perturbation theory by Callan and Zimmerman.¹⁹ For our purposes, it is most convenient to state their result in momentum space. (Bringing two space-time arguments together in position space is equivalent to letting a momentum transfer go to Euclidean infinity.) In our notation, the asymptotic expansion found by Callan and Zimmerman is

$$\Gamma^{(n+2)}\left(\frac{p+q}{2}, \frac{p-q}{2}, p_1 \dots p_n\right) = f(q^2)\Gamma_{\phi^2}^{(n)}(p; p_1 \dots p_n) + \dots, \quad (4.17)$$

as q goes to Euclidean infinity, for n greater than zero. The terms neglected are smaller by one power of k^2 than the terms retained. The function $f(q^2)$ is just the Fourier transform of f_1 . The identity term in (4.16) is missing from (4.17) because our Γ s are connected Green's functions; thus it only contributes to the expansion when n is zero. In any finite order of perturbation theory, $f(q^2)$ is not a simple power; it is full of logarithms. This is reminiscent of the situation for scaling behaviour in the deep Euclidean region.

Please note that this limit, the limit that is relevant to the operator product expansion, does *not* take us into the deep Euclidean region. Only one momentum gets large and Euclidean; the others stay close to the mass shell. Therefore, this is not a limit in which the renormalization group equations are applicable. However, the Callan–Symanzik equations are still applicable; they hold at all momenta:

$$\left[\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - n\gamma\right] \Gamma^{(n)}(p_1 \dots) = \Gamma_{\Delta}^{(n)}(0; p_1 \dots). \quad (4.18)$$

In addition to these two equations, we will need two more. One is the generalization of (4.17) when there is a mass insertion, Δ ,

$$\begin{aligned} \Gamma_{\Delta}^{(n+2)}\left(k; \frac{p+q}{2}, \frac{p-q}{2}, p_1 \dots p_n\right) \\ = f(q^2)\Gamma_{\Delta, \phi^2}^{(n)}(k; p; p_1 \dots p_n) + \dots, \end{aligned} \quad (4.19)$$

again, for n greater than zero, as q goes to Euclidean infinity. The function f is the same function that appears in (4.17). This is a reflection of the fact that (4.16) is an *operator* equation; the only purpose of the fixed momenta in (4.17) and (4.19) is to create the initial and final states between which the operators are to be evaluated. Since the equation is true whatever the states are, it does not matter if they are created by ϕ s or Δ s.

The other equation we will need is the Callan–Symanzik equation for a Green’s function built of a string of ϕ s and one ϕ^2 . Simple power counting shows that this can be renormalized by multiplying by the usual factors of Z_3 for the ϕ s, and by an independent renormalization constant for the ϕ^2 . Thus, by the same reasoning as led to the original Callan–Symanzik equations,

$$\left[\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - n\gamma - \gamma_{\phi^2} \right] \Gamma_{\phi^2}^{(n)}(p; p_1 \dots p_n) = \Gamma_{\Delta, \phi^2}^{(n)}(0; p; p_1 \dots p_n), \quad (4.20)$$

where γ_{ϕ^2} is yet another function of λ . (You may think the notation is a bit over-complex, since it ignores the fact that ϕ^2 and Δ are proportional. However, I want to emphasize that this is in no way essential to the argument, which can be extended trivially to other cases than (4.17), for which this is not true.)

Finally, I remind you if we denote by $d(\phi)$ the number that gives the exponent in the deep Euclidean asymptotic form (4.10), then we showed in Sect. 4.2 that

$$d(\phi) = 1 + \gamma(\lambda_1), \quad (4.21)$$

where λ_1 is a zero of β . Likewise, in the same sense of deep Euclidean asymptotic behaviour,

$$d(\phi^2) = 2 + \gamma_{\phi^2}(\lambda_1). \quad (4.22)$$

We are now ready to go. The essential point is simple: $\Gamma_{\Delta, \phi^2}^{(n)}$ can be computed in two ways; either we first take the operator-product limit and then use the Callan–Symanzik equations, or we first use the Callan–Symanzik equations and then take the operator-product limit. Doing things in one order, we find

$$\begin{aligned} & \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - (n+2)\gamma \right) \Gamma^{(n+2)} \left(\frac{p+q}{2}, \frac{p-q}{2}, \dots \right) \\ &= \Gamma_{\Delta}^{(n+2)} \left(0; \frac{p+q}{2}, \frac{p-q}{2}, \dots \right) \\ &= f(q^2) \Gamma_{\Delta, \phi^2}^{(n)}(0; p; \dots), \end{aligned} \quad (4.23)$$

as q goes to Euclidean infinity. Doing things in the other order, we find

$$\begin{aligned}
 & \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - (n+2)\gamma \right) \Gamma^{(n+2)} \left(\frac{p+q}{2}, \frac{p-q}{2}, \dots \right) \\
 &= \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - (n+2)\gamma \right) f(q^2) \Gamma_{\phi^2}^{(n)}(p; \dots) \\
 &= f(q^2) \Gamma_{\Delta, \phi^2}^{(n)}(0; p; \dots) \\
 & \quad + \Gamma_{\phi^2}^{(n)}(p; \dots) \left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - 2\gamma + \gamma_{\phi^2} \right) f(q^2). \quad (4.24)
 \end{aligned}$$

Two things equal to the same thing are equal to each other; thus

$$\left(\mu \frac{\partial}{\partial \mu} + \beta \frac{\partial}{\partial \lambda} - 2\gamma + \gamma_{\phi^2} \right) f(q^2) = 0. \quad (4.25)$$

But this is a homogeneous differential equation, of the same form as the renormalization group eq. (4.1)! Thus, by exactly the same arguments as we used in Sect. (4.2), we find that as q^2 goes to Euclidean infinity,

$$\begin{aligned}
 f(q^2) &\sim \left(\frac{q^2}{\mu^2} \right)^{[2\gamma(\lambda_1) - \gamma_{\phi^2}(\lambda_1)]/2} \\
 &\sim \left(\frac{q^2}{\mu^2} \right)^{[2d(\phi) - d(\phi^2)]/2}, \quad (4.26)
 \end{aligned}$$

verifying Wilson's rule, and showing that the dimensions that occur in Wilson's rule are the same as those that give the exponents in the deep-Euclidean asymptotic form.

5 Conclusions and questions

We have gone through a lot of complicated analysis, so perhaps I should end by summarizing our main results:

(1) The formal theory of broken scale invariance is a pack of lies, hopelessly afflicted with anomalies. In one case only (the low-energy theorems associated with scale invariance), the true equations have been found that replace the false equations. These are the Callan–Symanzik equations. They are simple and elegant, but practically useless for doing low-energy phenomenology, because they involve derivatives with respect to coupling constants.

(2) Under a very strong assumption (that the physical coupling constant is a zero of the function β that appears in the Callan–Symanzik equations), the Callan–Symanzik equations become equivalent to the naive low-energy theorems, except that the dimensions of fields, are, in general, anomalous. Nothing is known about what happens, under this assumption, to the rest of the formal theory. In particular, nothing is known about

what happens to the formal low-energy theorems that follow from broken conformal invariance.²⁰

(3) Under a much weaker assumption (the existence of a zero in β , not necessarily at the physical coupling constant), Green's functions show simple scaling behaviour in the deep Euclidean region, with anomalous dimensions. So do the coefficient functions that occur in Wilson's operator product expansion.

There is obviously much to be done. Here are some live questions:

(1) Perturbation theory tells us nothing about the zeros of β . Is there some other way to investigate this problem? Do the axioms of field theory say anything? Is there any way to get some sort of sum rule for β that can be evaluated experimentally?

(2) Is there any simple and compact way, analogous to the Callan–Symanzik equations, of expressing the anomalies in the low-energy theorems associated with broken conformal invariance? If there is, what do such expressions tell us about asymptotic behaviour?

(3) Is there any way of saving PCDC?

(4) Can the methods we have been playing with here, or extensions of them, be used to get information about the light-cone operator product expansion? Less abstractly, how do we make contact with the experimental data on deep inelastic electron–nucleon scattering?

None of these are easy questions, but progress in this field has been surprisingly rapid in the last few years. Maybe we will get the answers sooner than we expect.

Notes and references

1. Notation: Greek indices run from 0 to 3, with 0 denoting the time axis. Latin indices run from 1 to 3. Space-time points are labelled in three ways: $x \equiv x^\mu \equiv (x^0, \mathbf{x})$. $\partial_\mu = \partial/\partial x^\mu$. The signature of the metric tensor is $(+ - - -)$.
2. For a proof, see, for example, S. Coleman and R. Jackiw: *Ann. of Physics*, **67**, 552 (1971).
3. I should emphasize that when I say these results are perfectly general, I mean only that they can be established in the most general case by formal manipulations of the Lagrangian. Such manipulations are obviously invalid in a quantum field theory, where the divergences that occur when two fields approach the same space-time point make even *defining* the Lagrangian a delicate matter. Nevertheless, the Ward identities are frequently valid, at least to all orders in renormalized perturbation theory. Later on, we will discuss when they can be trusted.
4. Equation (2.7) should be read as follows: Under the transformation labelled by the parameter α , the space-time point x goes into $e^\alpha x$.
5. This, and all other statements in this section, are derived in tedious detail in C. Callan, S. Coleman and R. Jackiw: *Ann. of Phys.*, **59**, 42 (1970). This paper also contains references to the earlier literature on conformal invariance.

6. This equation shows very clearly the lack of connection between scale dimension and dimension in the sense of dimensional analysis. In the sense of dimensional analysis, σ has the dimensions of mass, like any other scalar field; however, because of its inhomogeneous transformation law, its scale dimension is not even defined.
7. Good papers to read are J. Ellis: *Nucl. Phys.* **B22**, 478 (1970), **B26**, 537 (1971), and the lectures of B. Zumino at the 1970 Brandeis Summer School in *Lectures on Elementary Particles and Quantum Field Theory*, edited by S. Deser *et al.* (M.I.T. Press 1970).
8. I remind you that a one-particle-irreducible Green's function is the sum of all connected Feynman diagrams that cannot be cut in two by breaking a single internal line. By convention the diagrams are evaluated with no propagators on the external lines. Thus, $\Gamma^{(2)}$ is the inverse propagator; $\Gamma^{(3)}$ is the proper three-line vertex (zero in the theory (3.1), by parity), $\Gamma^{(4)}$ is the two-into-two off-mass-shell scattering amplitude, less the pole terms (which are absent anyway in (3.1)), etc.
9. S. Weinberg: *Phys. Rev.* **118**, 838 (1960).
10. An instructive example is a recent computation of T. Appelquist and J. Primack: *Phys. Rev.* **D4**, 2454 (1971). In vector-meson-nucleon theory, these authors isolate those diagrams which give the dominant asymptotic behaviour for the nucleon electromagnetic form factor, in each order of perturbation theory (the 'leading logarithms'). They then sum up these leading terms, and find an expression proportional to $\exp[-c \ln^2 s]$, where c is a positive constant, proportional to the square of the coupling constant. This goes to zero more rapidly than any powers of s ; that is to say, the sum of the leading terms is highly damped, unlike the individual leading terms, which grow like powers of logarithms. Also, and even more discouraging, in the asymptotic region the sum of the leading terms is much less important than any one of the terms that were neglected (the non-leading terms) in the original summation.
11. C. G. Callan: *Phys. Rev.*, **D2**, 1541 (1970); K. Symanzik: *Comm. Math. Phys.* **18**, 227 (1970).
12. Unfortunately, for the Ward identities at arbitrary momentum transfers, and, particularly unfortunately for the low-energy theorems that follow from broken conformal invariance, the situation remains obscure.
13. See Chapter 4 in this book.
14. M. Gell-Mann and F. E. Low: *Phys. Rev.* **95**, 1300 (1954); N. N. Bogoliubov and D. V. Shirkov: *Introduction to the Theory of Quantized Fields*, Interscience, 1959.
15. K. Wilson: *Phys. Rev.* **D3**, 1818 (1971). The immediate ancestors of Wilson's arguments can be found in the studies of quantum electrodynamics by M. Baker, K. Johnson and R. Willey: *Phys. Rev.* **136 B**, 1111 (1964); **163**, 1699 (1967); **183**, 1292 (1969); **D 3**, 2516 (1971), and also in the original work on the renormalization group by Gell-Mann and Low¹⁴; however, I believe that the application of these ideas to strong-interaction scaling first appeared in Wilson's work.
16. See, e.g., the lectures of G. Preparata in *Properties of the Fundamental Interactions*. (Editrice Compositori, Bologna, 1973).
17. K. Wilson: *Phys. Rev.* **179**, 1499 (1969).
18. K. Symanzik: *Comm. Math. Phys.* **23**, 49 (1971).
19. C. Callan *Phys. Rev. D* **5**, 3205 (1972). W. Zimmerman: in *Lectures on Elementary Particles and Quantum Field Theory*, edited by S. Deser *et al.*, M.I.T. Press., (1970).
20. It has recently been shown by B. Schroer: *Lettere al Nuovo Cimento* **2**, 867 (1971) that these are also free of anomalies, at a zero of β .