Nuclear Physics B254 (1985) 11–18

THE BIRTH OF ASYMPTOTIC FREEDOM

Gerard 't HOOFT

*Institut voor Theoretische Fysica, Rijksuniversiteit, Utrecht, The Netherlands**

An account is given of the author's personal perception of the historical developments that accumulated into the present understanding of the renormalization group behavior of renormalizable quantum field theories.

When I was asked to speak at the special colloquium at DESY to commemorate Kurt Symanzik the subject "Asymptotic freedom" was a natural choice to make. For my memories of Kurt Symanzik from the time that this notion was developed and understood are filled with admiration and gratitude.

The notion of "renormalization group" was introduced by Peterman and Stückelberg [1] in 1953. What they observed was in modern words the following. When a renormalized amplitude is computed as a perturbative expansion in terms of some set of coupling constants then this can be seen as a recursive procedure. The formally infinite counterterms of the lagrangian that have been introduced to make amplitudes finite up to a certain order are used again in the next order diagrams. The surviving infinities in the new diagrams can then again be cancelled by new local counterterms.

Now the infinite parts of these counterterms are prescribed, but the finite parts can be chosen at will, in principle. Any recipe that prescribes the choice of these finite parts is called a regularization procedure. Reformulating the theory in terms of the formally infinite bare coupling constants, masses and possibly field renormalization factors is called "renormalization". The transition from one prescription to another can be described by substitutions of the form

$$g' = g + a_2 g^2 + a_3 g^3 \dots ,$$

$$m' = m + b_1(m) g + b_2(m) g^2 \dots ,$$

etc.

One formalism may use g and m as the fundamental independent variables, another

* Mailing address: Princetonplein 5, PO Box 80.006, 3508 TA Utrecht, The Netherlands.

uses g' and m' (there could of course be larger sets of parameters as the case may be). As long as we stick to perturbation expansion no one choice is essentially superior to any other. The transformations (1) can be seen as "general coordinate transformations" and of course form a group: the renormalization group. Physically observable phenomena ought to be invariant under the renormalization group transformations.

Since the renormalization group transformations mix terms of different order in perturbation expansion one might hope to obtain new information on non-perturbative features in a theory. Unfortunately the amount of information one gets is quite limited, only one abelian subgroup of the renormalization group is non-trivial in this respect: it is a subgroup that can be mapped onto the dilatation group.

The dilatation group was the starting point of Gell-Mann and Low in their pioneering paper [2] in 1954. They applied it to quantum electrodynamics. What they found is that the regularization procedure is scale dependent, and that a scale transformation is therefore associated with a "renormalization group transformation" of the form

$$\frac{\mu^2\partial}{\partial\mu^2}\alpha=\frac{N}{3\pi}\alpha^2+\mathrm{O}(\alpha^3)\,, \tag{1}$$

where μ is the energy scale and N is the number of charged fermion types. $\alpha=e^2/4\pi$ is now the running coupling constant. At $\mu=m_\mathrm{e}$ it is $\frac{1}{137}$.

Clearly, α increases (more and more rapidly) with increasing energy scale. Gell-Mann and Low speculated that the r.h.s. of eq. (1) would develop a zero at some value $\alpha=\alpha_0$. Then that would be a universal, in principle calculable, value for the fine structure constant. Later, Symanzik [3] would frequently study this possibility. When I first met Kurt Symanzik he was working on an apparently different problem. That was in the Cargèse Summer School [4] of 1970 where he presented his beautiful lectures on renormalization of theories with softly broken symmetries such as the Gell-Mann–Lévy σ-model [5]. The same topic was also discussed there by the late Benjamin W. Lee [6]. The observation which Symanzik carefully formulated was that if a symmetry is broken softly (i.e. with lagrangian terms whose coefficients have the dimension of a positive power of a mass), then also the renormalization counterterms must be soft, and this limits considerably the number of arbitrary terms one has to introduce. It is clear that this is closely related to the small-distance behavior of these theories and quite naturally Symanzik was led to investigate the small-distance structure of renormalizable theories further. By considering infinitesimal mass and coupling constant insertions he refined essentially the work of Peterman and Stückelberg obtaining, as one should, the dilatation group. Independently of Callan [7] he wrote down the modern form of the renormalization group equation [3]:

$$\left[\alpha(g^2)\frac{m^2\partial}{\partial m^2}+\beta(g)\frac{\partial}{\partial g}-2n\gamma(g)\right]\Gamma(p_1,\dots,p_{2n},m^2,g)=\delta\Gamma\,, \tag{2}$$

where $\delta\Gamma$ is rapidly decreasing for large momenta p. The function $\beta(g)$ is the same function as the right-hand side of eq. (1).

For the renormalizable 4-dimensional theories known at that time, β was always found to be positive. Indeed, speculations on zeros of β were always difficult. But Wilson and Fisher [8] found that if the theory is considered in $4-\varepsilon$ dimensions then β starts out negative, with a zero close to the origin. This was a very useful observation enabling him to compute critical coefficients in the scaling limit of theories of statistical systems in 3 dimensions, by expanding in ε.

The word "renormalization group" is nowadays also used in connection with numerical approximations in statistical systems that have to do with scale transformations. However in such systems most often the transformation can only be applied in one direction, and so it would be more correct in these cases to talk about a semigroup rather than a group [22].

Returning to Symanzik's lectures in Cargèse, when I followed them I had another problem in my mind. It was the renormalization problem of Yang-Mills fields [9, 10]. The expert in this area was Veltman, my thesis advisor then. He had taught me his unitarity conditions and Ward identities [9]. It was a beautiful problem but apart from Veltman only very few people were working on it and in Cargèse it was hardly mentioned. It had struck me that the mass term introduced by Feynman [10] and also adopted by Veltman, was not truly soft in the sense of Symanzik. This is because it multiplies the longitudinal part of the gauge field. In the massless limit this longitudinal part would become singular unless a gauge condition is added, but the gauge condition in turn violates unitarity as soon as the mass is switched on again. The way in which mass terms develop in the σ-model with spontaneous symmetry breaking looks much more natural, and indeed the resolution of our problem turned out to lie here: mass in a gauge theory can be due entirely to spontaneous symmetry breaking, and renormalisability needs not be affected by this mechanism. I prefer the words "Higgs mechanism" rather than "spontaneous symmetry breaking" because strictly speaking the vacuum in a gauge theory is never degenerate: all states in Hilbert space are formally invariant under gauge transformations. I use the word "formal" here because in perturbation theory one often expands around the limit in which the gauge coupling constant is switched off, and precisely there one obtains the Goldstone mode, *with* a degenerate vacuum.

Although Symanzik was never really active in the field of gauge theories, he was one of the very first to recognize their importance when the renormalisability was established. He urged phenomenologists and experimentalists to take the new class of models seriously. But he himself was more interested in basic features of field theories and used to constrain himself to $\lambda\phi^4$ theory or sometimes electrodynamics (QED).

In June 1972 I attended a meeting in Marseille on Renormalization of Yang-Mills fields and applications to particles physics [11]. At the Marseille airport I met Symanzik (without knowing we had been on the same plane). He told me about his work on $\lambda\phi^4$ theory with a negative coupling constant [12]. A natural question for me to ask was whether such theories with a non-positive hamiltonian were viable at all. And then he explained that this question of course worried him but that perhaps this disease could be cured by a remarkable property of his theory. The equation

$$\frac{\mu^2\partial}{\partial\mu^2}\lambda = \frac{3}{32\pi^2}\lambda^2 + O(\lambda^3) \tag{3}$$

is solved by

$$\lambda \to \frac{-a}{1+(3a/32\pi^2)\log\mu^2}, \tag{4}$$

so at small distances ($\mu \to \infty$) the coupling strength vanishes and the hamiltonian would be dominated by the kinetic term which is still positive.

Nowadays it is generally agreed that such arguments are not sufficient to cure this theory, but it is nevertheless interesting in its own right. I had great joy in informing Symanzik of my own findings in gauge theories. There the β function (although I did not call it that yet) starts out negative naturally:

$$\beta(g^2) = -\frac{1}{16\pi^2}\left(\tfrac{11}{3}C_1 - \tfrac{1}{6}C_2N_s - \tfrac{2}{3}C_3N_f\right)g^4 + O(g^6), \tag{5}$$

where N_s is the number of scalars and N_f the number of fermions in the elementary representation. For SU(2), $C_1 = 2$, $C_2 = C_3 = 1$ and for SU(3), $C_1 = 3$. So in SU(2) up to 11 fermions are allowed, in SU(3) $16\frac{1}{2}$.

Symanzik's reply was one of interest but skepticism. It looked too good to be true. If true, this result would be very important, he told me, and he advised me to publish it soon. I ignored this sensible advice however, because I found it necessary to first write down elaborately my methods [13] which deviated from what was then conventional. I did mention my result of eq. (5) at the discussion session after Symanzik's talk at the conference.

I must add that at that time I was totally ignorant of strong interaction phenomenology. So I was not aware of an important paper by Parisi [14] in which he explained how remarkably well Symanzik's asymptotically free $-\lambda\phi^4$ theory could explain the Bjorken scaling properties of the fairly successful parton models for deep inelastic scattering.

Soon afterwards the results of two American groups were announced: Politzer [15] had discovered independently the minus sign for gauge theories and stressed in his announcement the importance of now having a calculable perturbative behavior at high energies. Gross and Wilczek [16] had also computed β for gauge theories. I think they first coined the words "asymptotic freedom"*. In any case, they fully

* But, as Iliopoulos would remark later, invariably when someone talks about freedom, what he really means turns out to be something else... [17].

realised how well these observations fit in a parton theory with Bjorken scaling. In a first attempt they tried to construct a gauge model for partons with Higgs mechanism, but in fact the physics community was now prepared for discovering the fully unbroken gauge theory of quantum chromodynamics, a discovery subsequently made by many authors [18].

It was Symanzik who at various occasions had stressed that the discovery of asymptotic freedom had been made first in Europe, and that discussion remarks made at conferences should be recognized as official announcements.

Why are gauge theories asymptotically free, defying the earlier no-go theorems based on the Källen representation of propagators? Of course these no-go theorems do not apply directly because the propagator is not gauge-invariant and may therefore contain ghosts, but still, the large negative coefficient comes as a surprise.

Several heuristic arguments have been put forward to explain the sign, often in terms of a negative screening effect. In my way of computing things by far the largest contribution to β came not from the ghosts but from the large color-magnetic moments of the gluons. It is as if screening takes place in the magnetic rather than the electric sector of the theory. In all, the best argument I can find is that the sign is as it is just because of a fluke in a fairly complicated calculation.

It was also realized [19] that non-abelian gauge theories are the only theories with negative β coefficient(s). This is perhaps why the sign never worried me: my calculation in gauge theories was done before I even considered $\lambda\phi^4$ or QED!.

One often hears the statement that a Higgs mechanism would spoil asymptotic freedom, as a price paid for infrared freedom. This is not at all true, although asymptotic freedom poses severe (and interesting) restrictions on the values of the physical parameters. There exists a universal algebra that gives all β functions to one-loop order for any renormalizable field theory in four dimensions. The last part of this paper will deal with this algebra. We mention the result only; the calculation can be readily made using the algorithm of ref. [13].

The most general renormalizable lagrangian of a field theory is

$$L = -\tfrac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} - \tfrac{1}{2}(D_\mu\phi)^2 - \bar{\psi}\gamma D\psi - V(\phi) - \bar{\psi}W(\phi)\psi. \tag{6}$$

Here ϕ is a real scalar and ψ a Dirac fermion; we write

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g^{abc}A^b_\mu A^c_\nu, \tag{7a}$$

$$D_\mu\phi_i = \partial_\mu\phi_i + T^a_{ij}A^a_\mu\phi_j, \tag{7b}$$

$$D_\mu\psi_i = \partial_\mu\psi_i + U^a_{ij}A^a_\mu\psi_j, \qquad U = U_{\mathrm{s}} + U_{\mathrm{p}}\gamma_5, \tag{7c}$$

$$W_{ij}(\phi) = S_{ij}(\phi) + iP_{ij}(\phi)\gamma_5. \tag{7d}$$

V is a quartic and S and P are linear polynomials in ϕ. We absorbed the gauge coupling constant(s) into the structure constants and generators T and U of the (completely arbitrary) gauge group. Further we write

$$C_1^{ab} = g^{apq} g^{bpq}, \tag{8a}$$

$$C_2^{ab} = -\mathrm{Tr}(T^a T^b), \tag{8b}$$

$$C_3^{ab} = -\mathrm{Tr}\left(U_\mathrm{L}^a U_\mathrm{L}^b + U_\mathrm{R}^a U_\mathrm{R}^b\right) = -2\,\mathrm{Tr}\left(U_\mathrm{s}^a U_\mathrm{s}^b + U_\mathrm{p}^a U_\mathrm{p}^b\right), \tag{8c}$$

$$U_{\substack{\mathrm{L}\\ \mathrm{R}}} = U_\mathrm{s} \pm U_\mathrm{p}, \tag{8d}$$

$$W = S + iP, \qquad V_i = \frac{\partial V}{\partial \phi_i}, \qquad S_{,i} = \frac{\partial S}{\partial \phi_i} \text{ etc.}, \tag{8e}$$

$$W^* = S - iP.$$

We find that the one-loop counterterms of the theory can be cast in the form of a gauge invariant lagrangian:

$$\frac{1}{8\pi^2(n-4)}\,\Delta L, \tag{9}$$

if certain (in general gauge-dependent) field renormalizations are carried out first. Our expression for ΔL is

$$\Delta L = G_{\mu\nu}^a G_{\mu\nu}^b\left[\tfrac{11}{12}C_1^{ab} - \tfrac{1}{24}C_2^{ab} - \tfrac{1}{6}C_3^{ab}\right] - \Delta V - \bar{\psi}\,\Delta W \psi, \tag{10}$$

with

$$\Delta V = -\tfrac{1}{4}V_{ij}^2 - \tfrac{3}{2}V_i(T^2\phi)_i - \tfrac{3}{4}(\phi T^a T^b \phi)^2$$
$$- \phi_i V_j \mathrm{Tr}(S_{,i}S_{,j} + P_{,i}P_{,j}) + \mathrm{Tr}(S^2 + P^2)^2 - \mathrm{Tr}[S, P]^2, \tag{11}$$

and (see (8e))

$$\Delta W = -\tfrac{1}{4}W_{,i}W_{,i}^* W - \tfrac{1}{4}W W_{,i}^* W_{,i} - W_{,i} W^* W_{,i}$$
$$- \tfrac{3}{2}U_\mathrm{R}^2 W - \tfrac{3}{2}W U_\mathrm{L}^2 - W_{,i}\phi_j \mathrm{Tr}(S_{,i}S_{,j} + P_{,i}P_{,j}). \tag{12}$$

Although these expressions may look complicated they are really very compact and easy to use for finding the one-loop β function quickly. One has [13]

$$\frac{\mu^2\,\partial L}{\partial \mu^2} = \beta(L) = -\frac{\Delta L}{16\pi^2}. \tag{13}$$

The β function for the gauge coupling constant(s) can readily be obtained by rescaling the fields A_μ^a to cast the kinetic part of the gauge lagrangian $L + \Delta L$ back into the form $-\frac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a$. One recognizes eq. (5) as the first term in eq. (10).

I would like to concentrate for a moment on the signs and relative magnitudes of the various terms. In general there are three kinds of dimensionless coupling constants:

(i) gauge coupling constants, which we denote by a single symbol g;
(ii) Yukawa coupling constants, to be denoted by y;
(iii) ϕ^4 coupling constants, indicated by a λ.

The signs of the various β coefficients are then indicated in the following shorthand expressions:

$$\beta(g) = (-C_1 + C_2 N_s + C_3 N_f) g^3,$$

$$\beta(y) = C_4 y^3 - C_5 g^2 y,$$

$$\beta(\lambda) = C_6 \lambda^2 - C_7 g^2 \lambda + C_8 g^4 + C_9 y^2 \lambda - C_{10} y^4. \qquad (14)$$

The relative magnitudes of the coefficients C may be very model dependent but their signs are universal. Because of the + signs for C_2, C_3, C_4 and C_6 we can never have asymptotic freedom for abelian gauge theories ($C_1 = 0$) or scalar or Yukawa theories.

But the general gauge theory with fermions and scalars may well have all β negative. The high-energy limit will not be a *stable* solution of the flow equations but I cannot think of a good physical argument requiring such stability. On the contrary, stability should be required if the theory is fixed at the high-energy end of the spectrum and the equations are solved for small μ. Such stability is called "naturalness" and still constitutes an ill-understood problem in field theory.

In the simplest examples of theories with just three coupling constants I found that only one-parameter sets of coupling constants are asymptotically free. That would be a nice contribution to the unification ideas were it not that no suggestive model could be found this way, and, anyhow, asymptotic freedom being only a logarithmic property is not a very physical requirement for theories with weak couplings.

The exercise is good however, and we found that solutions required C_5 to be sufficiently large and C_8 sufficiently small. As a rule of thumb, in searching for asymptotically free sets we found therefore that fermions should be in sufficiently high representation (rather than the elementary one) and scalars sufficiently low. N_f can often be adjusted. Typical algebraic numbers such as [20]

$$\frac{\lambda}{g^2} = \frac{1}{8}(\sqrt{129} - 5)$$

may come out.

Recently the algebraic key formula has been extended to encompass also the two-loop corrections to the β functions [21].

References

[1] E.C.G. Stueckelberg and A. Peterman, Helv. Phys. Acta 26 (1953) 499;
N.N. Bogoliubov and D.V. Shirkov, Introduction to the theory of quantized fields (Interscience, New York, 1959)
[2] M. Gell-Mann and F. Low, Phys. Rev. 95 (1954) 1300
[3] K. Symanzik, Commun. Math. Phys. 18 (1970) 227; 23 (1971) 49
[4] Proc. NATO Summer School, l'Institut d'Etudes Scientifiques de Cargèse, July 1970, ed. M. Lévy and D. Bessis
[5] M. Gell-Mann and M. Lévy, Nuovo Cim. 16 (1960) 705
[6] B.W. Lee, Chiral dynamics (Gordon and Breach, New York, 1972)
[7] C.G. Callan, Phys. Rev. D2 (1970) 1541
[8] K.G. Wilson, and M.E. Fisher, Phys. Rev. Lett. 28 (1972) 240
[9] M. Veltman, Nucl. Phys. B7 (1968) 637; B21 (1970) 288;
J. Reiff and M. Veltman, Nucl. Phys. B13 (1969) 545
[10] R.P. Feynman, Acta Phys. Polon. 24 (1963) 697
[11] Proc. Colloquium on Renormalization of Yang-Mills fields and applications to particle physics, Marseille, June 19–23 1972, ed. C.P. Korthals-Altes et al.
[12] K. Symanzik, Nuovo Cim. Lett. 6 (1973) 77
[13] G. 't Hooft, Nucl. Phys. B61 (1973) 455; B62 (1973) 444
[14] G. Parisi, Nuovo Cim. Lett. 7 (1973) 84
[15] H.D. Politzer, Phys. Rev. Lett. 30 (1973) 1346
[16] D.J. Gross and F. Wilczek, Phys. Rev. Lett. 30 (1973) 1343
[17] J. Iliopoulos, Proc. XVII Int. Conf. on High-energy physics, London, July 1974, ed. J.R. Smith, Rutherford Lab, p. III-89
[18] M. Fritsch, M. Gell-Mann and H. Leutwyler, Phys. Rev. Lett. 47B (1973) 365
S. Weinberg, Phys. Rev. Lett. 31 (1973) 494;
K. Wilson, Phys. Rev. D10 (1974) 2445;
R. Balian, J.M. Drouffe and C. Itzykson, Phys. Rev. D10 (1974) 3376;
K. Kogut and L. Susskind, Phys. Rev. D9 (1974) 3501
[19] S. Coleman and D. Gross, Phys. Rev. Lett. 31 (1973) 851
[20] G. 't Hooft, Cargèse Summer School on Progress in gauge field theory, Sept. 1–15 1983, ed. G. 't Hooft et al.
[21] R. van Damme, Nucl. Phys. B227 (1983) 317, and to be published
[22] N.G. van Kampen, private communication