A hint of renormalization

Bertrand Delamotte[a)]
Laboratoire de Physique Théorique et Hautes Energies, Universités Paris VI, Pierre et Marie Curie, Paris VII, Denis Diderot, 2 Place Jussieu, 75251 Paris Cedex 05, France

(Received 28 January 2003; accepted 15 September 2003)

An elementary introduction to perturbative renormalization and renormalization group is presented. No prior knowledge of field theory is necessary because we do not refer to a particular physical theory. We are thus able to disentangle what is specific to field theory and what is intrinsic to renormalization. We link the general arguments and results to real phenomena encountered in particle physics and statistical mechanics. © *2004 American Association of Physics Teachers.*
[DOI: 10.1119/1.1624112]

I. INTRODUCTION

Hans Bethe in a seminal 1947 paper was the first to calculate the energy gap, known as the Lamb shift, between the $2s$ and $2p$ levels of the hydrogen atom.[1] These levels were found to be degenerate even in Dirac's theory, which includes relativistic corrections. Several authors had suggested that the origin of the shift could be the interaction of the electron with its own radiation field (and not only with the Coulomb field). However, to quote Bethe, ''This shift comes out infinite in all existing theories and has therefore always been ignored.'' Bethe's calculation was the first to lead to a finite, accurate result. Renormalization—in its modern perturbative sense—was born.[2] Since then it has developed into a general algorithm to get rid of infinities that appear at each order of perturbation theory in (almost) all quantum field theories (QFT).[3–7] In the meantime, the physical origin of these divergences has also been explained (see Ref. 8 for many interesting contributions on the history and philosophy of renormalization and renormalization group).

In QFT, as in ordinary quantum mechanics, the perturbative calculation of any physical process involves, at each order, a summation over (virtual) intermediate states. However, if the theory is Lorentz invariant, an infinite number of supplementary states exist compared with the Galilean case and their summation, being generically divergent, produces infinities. The origin of these ''new'' states is deeply rooted in quantum mechanics and special relativity. When these two theories are combined, a new length scale appears, built out of the mass m of the particles: the Compton wavelength $\hbar/mc$. It vanishes in both formal limits $\hbar=0$ and $c=\infty$, corresponding, respectively, to classical and Galilean theories. Because of Heisenberg inequalities, probing distances smaller than this length scale requires energies higher than mc^2 and thus imply the creation of particles. This possibility to create and annihilate particles forbids the localization of the original particle better than the Compton wavelength because the particles that have just been created are strictly identical to the original one. Quantum mechanically, these multi-particle states play a role even when the energy involved in the process under study is lower than mc^2, because they are summed over as virtual states in perturbation theory. Thus, the divergences of perturbation theory in QFT are directly linked to its short distance structure, which is highly nontrivial because its description involves the infinity of multi-particle states.

Removing these divergences has been the nightmare and the delight of many physicists working in particle physics. It seemed hopeless to the non-specialist to understand renormalization because it required prior knowledge of quantum mechanics, relativity, electrodynamics, etc. This state of affairs contributed to the nobility of the subject: studying the ultimate constituents of matter and being incomprehensible fit well together. However, strangely (at least at first sight) the theoretical breakthrough in the understanding of renormalization beyond its algorithmic aspect came from Wilson's work on continuous phase transitions.[9] The phenomena that take place at these transitions are neither quantum mechanical[10] nor relativistic and are non-trivial because of their cooperative behavior, that is, their properties at large distances.[11] Thus neither $\hbar$ nor c are necessary for renormalization. Something else is at work that does not require quantum mechanics, relativity, summation over virtual states, Compton wavelengths, etc., even if in the context of particle physics they are the ingredients that make renormalization necessary. In fact, even divergences that seemed to be the major problem of QFT are now considered only as by-products of the way we have interpreted quantum field theories. We know now that the invisible hand that creates divergences in some theories is actually the existence in these theories of a no man's land in the energy (or length) scales for which cooperative phenomena can take place, more precisely, for which fluctuations can add up coherently.[12] In some cases, they can destabilize the physical picture we were relying on and this manifests itself as divergences. Renormalization, and even more renormalization group, is the right way to deal with these fluctuations.

One of the aims of this article is to disentangle what is specific to field theory and what is intrinsic to the renormalization process. Therefore, we shall not look for a physical model that shows divergences,[13–17] but we shall rather show the general mechanism of perturbative renormalization and the renormalization group without specifying a physical model.

II. A TOY MODEL FOR RENORMALIZATION

In the following, we consider an unspecified theory that involves, by hypothesis, only one free parameter g_0 in terms of which a function $F(x)$, representing a physical quantity, is calculated perturbatively, that is, as a power series. An example in QFT would be quantum electrodynamics (QED), which describes the interaction of charged particles such as electrons with the electromagnetic field. For high energy processes, the mass of the electron is negligible and the only parameter of this theory in this energy regime is its charge,

which is therefore the analog of g_0. F can then represent the cross section of a scattering process as, for instance, the scattering of an electron on a heavy nucleus in which case x is the energy–momentum four-vector of the electron. The coupling constant g_0 is defined by the Hamiltonian of the system, and F is calculated perturbatively using the usual (à la Feynman) approach. Another important example is continuous phase transitions. For fluids, F could represent a density–density correlation function and for magnetism a spin–spin correlation function.[18] Yet another example is the solution of a differential equation that can arise in some physical context and that can show divergences (see the following).

It is convenient for what follows to assume that $F(x)$ has the form:

$$F(x)=g_0+g_0^2F_1(x)+g_0^3F_2(x)+\cdots. \tag{1}$$

Up to a redefinition of F, this form is general and corresponds to what is really encountered in field theory. Let us now assume that the perturbation expansion of $F(x)$ is ill-defined and that the $F_i(x)$ are functions involving divergent quantities. An example of such a function is

$$F_1(x)=\alpha\int_0^\infty \frac{dt}{t+x}, \tag{2}$$

which is logarithmically divergent at the upper limit. This example has been chosen because it shares many common features with divergent integrals encountered in QFT: the integral corresponds to the summation over virtual states and $\alpha(t+x)^{-1}$ represents the probability amplitude associated with each of these states.[19]

A simple although crucial observation is that because there is only one free parameter in the theory by hypothesis, only one "measurement" of $F(x)$, say at the point $x=\mu$, is necessary to fully specify the theory we are studying. Such a measurement is used to fix the value of g_0 so as to reproduce the experimental value of $F(\mu)$. For QED for instance, this procedure would mean that:

(i) We start by writing a general Hamiltonian compatible with basic assumptions, for example, relativity, causality, locality, and gauge invariance.
(ii) We calculate physical processes at a given order of perturbation theory.
(iii) We fix the free parameter(s) of the initial Hamiltonian to reproduce at this order the experimental data.

This last step requires as much data as there are free parameters. Once the parameters are fixed, the theory is completely determined and thus predictive. One could then think that it does not matter whether we parametrize the theory in terms of g_0, which is only useful in intermediate calculations, or with a "physical," that is, a measured quantity $F(\mu)$, because g_0 will be replaced by this quantity anyway. Having this freedom is indeed the generic situation in physics, but the subtlety here is that the perturbation expansion of $F(x)$ is singular, and, thus, so is the relationship between g_0 and $F(\mu)$. Thus, it seems crucial to reparametrize F in terms of $F(\mu)$ when the expansion is ill-defined.

The *renormalizability hypothesis* is that the reparametrization of the theory in terms of a physical quantity, instead of g_0, is enough to turn the perturbation expansion into a well-defined expansion. The hypothesis is therefore that the problem does not come from the perturbation expansion itself, that is, from the functions $F_i(x)$, but from the choice of parameter used to perform it. This hypothesis means that the physical quantity, $F(x)$, initially represented by its ill-defined expansion Eq. (1), should have a well-defined perturbation expansion once it is calculated in terms of the physical parameter $F(\mu)$. This is the simplest hypothesis we can make, because it amounts to preserving the x-dependence of the functions $F_i(x)$ and only modifying the coupling constant g_0. Thus, we assume that $F(x)$ is known at one point μ, and we define g_R by

$$F(\mu)=g_R. \tag{3}$$

In the following, and by analogy with QFT, we call g_R the renormalized coupling constant and Eq. (3) a "renormalization prescription," a barbarian name for such a trivial operation.

We are now in a position to discuss the renormalization program. It consists of reparametrizing the perturbation expansion of F so that it obeys the prescription of Eq. (3). The point here is that we cannot use Eq. (3) together with Eq. (1) because Eq. (1) is ill-defined. We first need to give a well-defined meaning to the perturbation expansion. This is the *regularization* procedure which is the first step of any renormalization.[20,21] The idea is to define the perturbation expansion of F by a limit such that (i) the $F_i(x)$ are well-defined before the limit is taken, and (ii) after the renormalization has been performed, the original formal expansion is recovered when the limit is taken.

We thus introduce a new set of (regularized) functions F_Λ and $F_{i,\Lambda}$, involving a new parameter Λ, which we call the regulator, and such that for Λ finite all these functions are finite. We thus define

$$F_\Lambda(x)=F_\Lambda(x,g_0,\Lambda)=g_0+g_0^2F_{1,\Lambda}(x)+g_0^3F_{2,\Lambda}(x)+\cdots. \tag{4}$$

There are infinitely many ways of regularizing the F_i's and for the example given in Eq. (2), it can consist for instance in introducing a cut-off in the following integral:

$$F_{1,\Lambda}(x)=\alpha\int_0^\Lambda \frac{dt}{t+x}. \tag{5}$$

Different regularization schemes can lead to very different intermediate calculations, but must all lead to identical results.[22] For instance, dimensional regularization is widely used in QFT because it preserves Lorentz and gauge symmetries.[13–15,23] We do not need here to specify a regularization for the function F, because our arguments will be general and the few calculations elementary.

Once a regularization scheme has been chosen, it is possible to use the renormalization prescription, Eq. (3), together with the regularized expansion, Eq. (4), to obtain a well-defined perturbation series for F_Λ in terms of the physical coupling g_R. If this expansion makes sense—this is the renormalizability hypothesis—it must be finite even in the limit $\Lambda\to\infty$, because it expresses a finite physical quantity $F(x)$ in terms of a physical quantity g_R. Thus, the renormalization program consists first in changing $F(x,g_0)$ to $F_\Lambda(x,g_0,\Lambda)$, then in rewriting F_Λ in terms of g_R and μ,

$$F_\Lambda(x,g_0,\Lambda)\to F_\Lambda(x,g_R,\mu), \tag{6}$$

and only then taking the limit $\Lambda\to\infty$ at fixed g_R and μ. If this limit exists, $F_\infty(x)$ is by hypothesis the function $F(x)$:

$$F(x)=F(x,g_R,\mu)\underset{\Lambda\to\infty}{=}F_\Lambda(x,g_R,\mu). \tag{7}$$

Of course, the divergences must still be somewhere, and we shall see that they survive in the relationship between g_0 and g_R; at fixed g_R, g_0 diverges when $\Lambda\to\infty$. In the traditional interpretation of renormalization, this divergence is supposed to be harmless because g_0 is supposed to be a nonphysical quantity. We shall come back to this point later.

The renormalization program is performed recursively, and we now implement it order by order to see how it works and the constraints on the perturbation expansion that it implies. Let us emphasize that the series expansion we shall use in intermediate calculations are highly formal because they are ill-defined in the limit $\Lambda=\infty$. They are justified only by the result we finally obtain: a good perturbation expansion in terms of g_R.[24]

• *Renormalization at order* g_0. At this order $F(x)$ is constant and given by

$$F_\Lambda(x)=g_0+O(g_0^2). \tag{8}$$

Thus the use of Eq. (3) leads to

$$g_0=g_R+O(g_R^2). \tag{9}$$

• *Renormalization at order* g_0^2. Our only freedom to eliminate the divergence of $F_\Lambda(x)$ is to redefine g_0. Because we are working perturbatively, we expand g_0 as a power series in g_R. Thus, we set

$$g_0=g_R+\delta_2 g+\delta_3 g+\cdots, \tag{10}$$

where $\delta_n g\sim O(g_R^n)$. At order g_R^2 we obtain

$$F_\Lambda(x)=g_R+\delta_2 g+g_R^2F_{1,\Lambda}(x)+O(g_R^3), \tag{11}$$

where we have used $g_0^2=g_R^2+O(g_R^3)$. If we impose Eq. (3) at this order, we obtain

$$\delta_2 g=-g_R^2F_{1,\Lambda}(\mu), \tag{12}$$

which diverges when $\Lambda\to\infty$. In our example, Eq. (5), we find

$$\delta_2 g=-\alpha g_R^2\int_0^\Lambda\frac{dt}{t+\mu}=-\alpha g_R^2\log\frac{\Lambda+\mu}{\mu}. \tag{13}$$

If we substitute Eq. (12) into Eq. (11), we obtain F_Λ to this order:

$$F_\Lambda(x)=g_R+g_R^2(F_{1,\Lambda}(x)-F_{1,\Lambda}(\mu))+O(g_R^3). \tag{14}$$

It is clear that this expression for $F_\Lambda(x)$ is finite for all x at this order if and only if the "divergent" part of $F_{1,\Lambda}(x)$ (the part that becomes divergent when $\Lambda\to\infty$) is exactly canceled by that of $F_{1,\Lambda}(\mu)$, that is, if and only if

$$F_{1,\Lambda}(x)-F_{1,\Lambda}(\mu)\ \text{is regular in}\ x\ \text{and}\ \mu\ \text{for}\ \Lambda\to\infty. \tag{15}$$

This condition of course means that the divergent part of $F_{1,\Lambda}(x)$ must be a constant, that is, is x-independent. If this is so, then we define the function $F(x)$—now called renormalized—as the limit of $F_\Lambda(x)$ when $\Lambda\to\infty$. The condition (15) is fulfilled for the example of Eq. (2), and we trivially find that $F(x)$ reads:

$$F(x)=g_R+\alpha(\mu-x)g_R^2\int_0^\infty\frac{dt}{(t+x)(t+\mu)}+O(g_R^3), \tag{16}$$

which is obviously well defined and such that the prescription of Eq. (3) is verified. We say that we have renormalized the theory to this order.

Before going to the next order of perturbation theory, let us note two important facts. First, the renormalization procedure consists of "adding a divergent term" $\delta_2 g$ to F_Λ to remove its divergence. The cancellation takes place between the second term of its expansion and the first one of order g_0. Both lead to a term of order g_R^2, the one coming from the expansion of g_0 in terms of g_R being tuned so as to cancel the divergence of the other. This mechanism of cancellation is a general phenomenon: a divergence coming from the nth term of the perturbation expansion is canceled by the expansion in powers of g_R of the $n-1$ preceding terms. Second, this cancellation is possible for all x only if the divergence of $F_{1,\Lambda}(x)$ is a number, that is, is x-independent. If it is not so, then $F_{1,\Lambda}(x)-F_{1,\Lambda}(\mu)$ would still be divergent $\forall x\neq\mu$. This divergence would require the imposition of at least one more renormalization prescription to be removed and this second prescription would define a second, independent, coupling constant (see Appendix A for two functions, one renormalizable and one that is not). The necessity for a second measurement of $F(x)$ would contradict our assumption that there is only one free parameter in the theory. Thus we conclude that this assumption drastically constrains the x-dependence of the divergences at order g_0^2. We actually show in the following that this constraint propagates to any order of perturbation theory in a nontrivial way. We also will show that together with dimensional analysis and for a very wide and important class of theories, these constraints are sufficient to determine the analytical form of the divergences.

• *Renormalization at order* g_0^3. We suppose that F can be renormalized at order g_R^2, that is, condition (15) is fulfilled. To understand the structure of the renormalization procedure, it is necessary to go one step further. At order g_R^3 we obtain

$$F_\Lambda(x)=g_R+\delta_2 g+\delta_3 g+(g_R^2+2g_R\delta_2 g)F_{1,\Lambda}(x)+g_R^3F_{2,\Lambda}(x)+O(g_R^4), \tag{17}$$

where we have used $g_0^3=g_R^3+O(g_R^4)$ and $g_0^2=g_R^2+2g_R\delta_2 g+O(g_R^4)$. We again impose the prescription Eq. (3) and obtain

$$\delta_3 g=2g_R^3(F_{1,\Lambda}(\mu))^2-g_R^3F_{2,\Lambda}(\mu). \tag{18}$$

If we substitute Eq. (18) in Eq. (17), we obtain

$$F_\Lambda(x)=g_R+g_R^2[F_{1,\Lambda}(x)-F_{1,\Lambda}(\mu)]+g_R^3[F_{2,\Lambda}(x)-F_{2,\Lambda}(\mu)-2F_{1,\Lambda}(\mu)(F_{1,\Lambda}(x)-F_{1,\Lambda}(\mu))]+O(g_R^4). \tag{19}$$

Once again, we require that the divergence has been subtracted for all x which imposes on the x-dependence of the divergent part of $F_{2,\Lambda}(x)$:

$$F_{2,\Lambda}(x)-F_{2,\Lambda}(\mu)-2F_{1,\Lambda}(\mu)(F_{1,\Lambda}(x)-F_{1,\Lambda}(\mu))\ \text{is regular in}\ x\ \text{and}\ \mu\ \text{as}\ \Lambda\to\infty. \tag{20}$$

Note that this constraint involves not only $F_{2,\Lambda}$, but also $F_{1,\Lambda}$. It is convenient to rewrite $F_{1,\Lambda}(x)$ and $F_{2,\Lambda}(x)$ as the sum of a regular and of singular (when $\Lambda\to\infty$) part:

$$F_{i,\Lambda}(x)=F^s_{i,\Lambda}(x)+F^r_{i,\Lambda}(x). \tag{21}$$

Because $\infty+\text{anything finite}=\infty$, this decomposition is not unique: the $F^s_{i,\Lambda}(x)$ are defined up to a regular part. It is convenient to choose $F^s_{1,\Lambda}(x)$ such that

$$F^s_{1,\Lambda}(x)-F^s_{1,\Lambda}(\mu)\underset{\Lambda\to\infty}{\to}0, \tag{22}$$

which, of course, implies condition (15). We show in Appendix B that, reciprocally, this choice is always possible if (15) is fulfilled. As already stated, Eq. (22) means that the divergent part of $F_{1,\Lambda}$ is x-independent. We can actually impose a more stringent condition on $F^s_{1,\Lambda}$ because, by again tuning the regular part of $F_{1,\Lambda}$, we can choose $F^s_{1,\Lambda}$ to be completely independent of x, for any Λ. We thus define

$$F^s_{1,\Lambda}(x)=f_1(\Lambda). \tag{23}$$

In our example, Eq. (5), we can choose

$$f_1(\Lambda)=\alpha\log\Lambda,\quad F^r_{1,\Lambda}(x)=\alpha\log\left(\frac{\Lambda+x}{\Lambda x}\right). \tag{24}$$

We now substitute Eq. (23) into Eq. (20) and, using the same kind of arguments as in Appendix B, we obtain a constraint on the singular part of $F_{2,\Lambda}(x)$ similar to the one on $F^s_{1,\Lambda}(x)$, Eq. (22):

$$F^s_{2,\Lambda}(x)-F^s_{2,\Lambda}(\mu)-2\,f_1(\Lambda)[F^r_{1,\Lambda}(x)-F^r_{1,\Lambda}(\mu)]\underset{\Lambda\to\infty}{\to}0. \tag{25}$$

Equation (25) can be rewritten as

$$[F^s_{2,\Lambda}(x)-2\,f_1(\Lambda)F^r_{1,\Lambda}(x)]-[F^s_{2,\Lambda}(\mu)-2\,f_1(\Lambda)F^r_{1,\Lambda}(\mu)]\underset{\Lambda\to\infty}{\to}0. \tag{26}$$

Equation (26) has the same structure as Eq. (22) up to the replacement: $F^s_{1,\Lambda}\to F^s_{2,\Lambda}-2\,f_1(\Lambda)F^r_{1,\Lambda}$ and therefore has the same kind of solution as Eq. (23):

$$F^s_{2,\Lambda}(x)=2\,f_1(\Lambda)F^r_{1,\Lambda}(x)+f_2(\Lambda), \tag{27}$$

where $f_2(\Lambda)$ is any function of Λ and is independent of x. We see in Eq. (27) that unlike $F_{1,\Lambda}(x)$, the divergent part of $F_{2,\Lambda}(x)$ depends on x. However, this dependence is entirely determined by the first order of the perturbation expansion. The $\delta_2 g$ term, necessary to remove the $O(g_0^2)$ divergence, has produced at order g_R^3 an x-dependent divergent term: $2g_R\delta_2 g F_{1,\Lambda}(x)$. This kind of x-dependence is also a general phenomenon of renormalization: the (counter-)terms that remove divergences at a given order produce divergences at higher orders. If the theory is renormalizable, these divergences contribute to the cancellation of divergences present in the perturbation expansion at higher orders. Thus, perturbative renormalizability, that is, the possibility of eliminating order by order all divergences by the redefinition of the coupling(s), implies a precise structure of (the divergent parts of) the successive terms of the perturbation series. At order n, the singular part of $F_{n,\Lambda}$ involves x-dependent terms entirely determined by the preceding orders plus one new term that is x-independent. In our example of Eqs. (2) and (24) we find

$$F^s_{2,\Lambda}(x)=2\alpha^2\log\Lambda\log\frac{x+\Lambda}{\Lambda x}+f_2(\Lambda). \tag{28}$$

By expanding $\log\Lambda\log(x+\Lambda)/\Lambda x$ in powers of Λ^{-1} and by again redefining the regular part of $F_{2,\Lambda}$, we obtain a simpler form for $F^s_{2,\Lambda}(x)$:

$$F^s_{2,\Lambda}(x)=-2\alpha^2\log\Lambda\log x+f_2(\Lambda). \tag{29}$$

This relation will be important in the following when we shall discuss the renormalization group.

Let us draw our first conclusion. Infinities occur in the perturbation expansion of the theory because we have assumed that it was not regularized. Actually, these divergences have forced us to regularize the expansion and thus to introduce a new scale Λ. Once regularization has been performed, renormalization can be achieved by eliminating g_0. The limit $\Lambda\to\infty$ can then be taken. The process is recursive and can be performed only if the divergences possess, order by order, a very precise structure. This structure ultimately expresses that there is only one coupling constant to be renormalized. This means that imposing only one prescription at $x=\mu$ is enough to subtract the divergences for all x. In general, a theory is said to be renormalizable if all divergences can be recursively subtracted by imposing as many prescriptions as there are independent parameters in the theory. In QFT, these are masses, coupling constants, and the normalization of the fields. An important and non-trivial topic is thus to know which parameters are independent, because symmetries of the theory (like gauge symmetries) can relate different parameters (and Green functions).

Let us once again recall that renormalization is nothing but a reparametrization in terms of the physical quantity g_R.[25] The price to pay for renormalizing F is that g_0 becomes infinite in the limit $\Lambda\to\infty$, see Eq. (12). We again emphasize that if g_0 is believed to be no more than a non-measurable parameter, useful only in intermediate calculations, it is indeed of no consequence that this quantity is infinite in the limit $\Lambda\to\infty$. That g_0 was a divergent non-physical quantity has been common belief for decades in QFT. The physical results given by the renormalized quantities were thought to be calculable only in terms of unphysical quantities like g_0 (called bare quantities) that the renormalization algorithm could only eliminate afterward. It was as if we had to make two mistakes that compensated for each other: first introduce bare quantities in terms of which everything was infinite, and then eliminate them by adding other divergent quantities. Undoubtly, the procedure worked, but, to say the least, the interpretation seemed rather obscure.

Before studying the renormalization group, let us now specialize to a particular class of renormalizable theories.

III. RENORMALIZABLE THEORIES WITH DIMENSIONLESS COUPLINGS

A very important class of field theories corresponds to the situation where g_0 is dimensionless, and x, which in QFT represents coordinates or momenta, has dimensions (or more generally when g_0 and x have independent dimensions). In four-dimensional space–time, quantum electrodynamics is in this class, because the fine structure constant is dimensionless; quantum chromodynamics and the Weinberg–Salam

model of electro-weak interactions are also in this class. In four space dimensions, the ϕ^4 model relevant for the Ginzburg–Landau–Wilson approach to critical phenomena is in this class too. This particular class of renormalizable theories is the cornerstone of renormalization in field theories.

Our main goal in this section is to show that, independently of the underlying physical model, dimensional analysis together with the renormalizability constraint determine almost entirely the structure of the divergences. This underlying simplicity of the nature of the divergences explains that there is no combinatorial miracle of Feynman diagrams in QFT as it might seem at first glance. Let us now see in detail how it works.

Because $F_\Lambda(x)$ has the same dimension as g_0, it also is dimensionless and so are the $F_{i,\Lambda}(x)$. The only possibility for a dimensionless quantity like F to be a function of a dimensional variable like x is that there exists another dimensional variable such that F depends on x only through the ratio of these two variables. Apart from x, the only other quantity on which F depends is Λ, which must therefore have the same dimension as x. This is indeed the case in our example, Eq. (5). Thus, the functions $F_{i,\Lambda}(x)$ depend on the ratio x/Λ only.[26] Let us show that this is enough to prove that the $F^s_{i,\Lambda}(x)$ are sums of powers of logarithms with, for most of them, prescribed prefactors.

Let us start with $F^s_{1,\Lambda}(x)$. On one hand, we have seen that by redefining the regular part of $F_{1,\Lambda}(x)$, we could take its singular part $F^s_{1,\Lambda}(x)$ independent of x, Eq. (23). On the other hand, we know that $F_{1,\Lambda}(x)$ is a function of x/Λ. Thus, by redefining $F^r_{1,\Lambda}(x)$, it must be possible to extract an x-dependent regular part, $r(x)$, of this function so as to build the x/Λ dependence of $F^s_{1,\Lambda}(x)$:

$$F^s_{1,\Lambda}(x)=f\left(\frac{x}{\Lambda}\right)=f_1(\Lambda)+r(x). \tag{30}$$

Hence, $F^s_{1,\Lambda}$ is separable into functions of x only and of Λ only which sum up to a function of x/Λ. We show in Appendix C the well-known fact that only the logarithm obeys this property. We obtain [see Eqs. (C3) and (C4)]

$$F^s_{1,\Lambda}(x)=-f_1\left(\frac{x}{\Lambda}\right)=f_1(\Lambda)-f_1(x)=\alpha\log\frac{\Lambda}{x}. \tag{31}$$

Therefore, for renormalizable theories and for dimensionless functions such as F, only logarithmic divergences are allowed at order g_0^2 (in QFT, this is the so-called one-loop term). This is the reason why logarithms are encountered everywhere in QFT. Note that because of dimensional analysis, the finite part of $F_{1,\Lambda}(x)$ is nothing but $r(x)$, up to an additive constant, at least for $\Lambda\to\infty$. This can be checked for the example given in Eq. (5). Thus, by dimensional analysis, the structure of the divergence determines that of the finite part (up to a constant). Notice that things would not be that simple if $F_\Lambda(x)$ depended on another dimensional parameter, which is the case of massive field theories where masses and momenta have the same dimension. In this case, the finite part is only partially determined by the singular one.

Let us now show that the structure of $F^s_{2,\Lambda}$ also is entirely determined for renormalizable theories with dimensionless couplings both by the renormalizability hypothesis and by dimensional analysis. We have already partially studied this case with the example given in Eq. (5) where $F^s_{1,\Lambda}(x)$ is logarithmically divergent, a characteristic feature of these renormalizable theories. In particular, we have shown that in this case, renormalizability imposes at order g_0^3 that $F^s_{2,\Lambda}$ is of the form given in Eq. (29). Let us now use dimensional analysis that once again imposes that $F^s_{2,\Lambda}$ depends only on x/Λ. The only freedom we have to reconstruct a function of x/Λ from the form of $F^s_{2,\Lambda}$ given in Eq. (29) is to add a regular function to it. It is not difficult to find how to proceed because the only admissible term including $\log\Lambda\log x$ is $\log^2\Lambda/x$:

$$\log^2\frac{\Lambda}{x}=\log^2\Lambda-2\log\Lambda\log x+\log^2 x. \tag{32}$$

Thus, to obtain the dimensionally correct extension of the term $-2\alpha^2\log\Lambda\log x$ in Eq. (29), we extract $\alpha^2\log^2\Lambda$ from $f_2(\Lambda)$ and add the regular term $\alpha^2\log^2 x$:

$$\begin{aligned}
&-2\alpha^2\log\Lambda\log x+f_2(\Lambda)\\
&\quad\to-2\alpha^2\log\Lambda\log x+\alpha^2\log^2\Lambda+[f_2(\Lambda)-\alpha^2\log^2\Lambda]\\
&\quad\to-2\alpha^2\log\Lambda\log x+\alpha^2\log^2\Lambda+\alpha^2\log^2 x+[f_2(\Lambda)\\
&\quad-\alpha^2\log^2\Lambda]\\
&\quad\to\alpha^2\log^2\frac{\Lambda}{x}+[f_2(\Lambda)-\alpha^2\log^2\Lambda].
\end{aligned} \tag{33}$$

Thus, we obtain for the new function $F^s_{2,\Lambda}(x)$:

$$F^s_{2,\Lambda}(x)=\alpha^2\log^2\frac{\Lambda}{x}+f_2(\Lambda)-\alpha^2\log^2\Lambda. \tag{34}$$

Now, for $f_2(\Lambda)-\alpha^2\log^2\Lambda$, we can repeat the same argument as the one used previously for $F^s_{1,\Lambda}(x)$ [which is equal to $f_1(\Lambda)$, Eq. (23)]: it is a function of Λ that must become a function of x/Λ only by adding a function of x. It is thus also a logarithm, see Eqs. (30) and (31) and Appendix C. Therefore, we add a $\log x$ term to $F^s_{2,\Lambda}(x)$ and obtain the final result:

$$F^s_{2,\Lambda}(x)=\alpha^2\log^2\frac{\Lambda}{x}+\beta\log\frac{\Lambda}{x}, \tag{35}$$

where β is a pure number. We emphasize that although it is x-independent, the term $\alpha^2\log^2\Lambda$ involved in $F^s_{2,\Lambda}(x)$ arises from the $\log\Lambda\log x$ term thanks to dimensional analysis. It is thus entirely determined by the term of order g_0^2 of perturbation theory. Only the sub-leading logarithm $\beta\log\Lambda/x$ is new. It is not difficult now to guess the structure of the next order of perturbation: it involves a $\log^3\Lambda/x$ with a prefactor α^3, a $\log^2\Lambda/x$ term with a prefactor which is a function of α and β and a $\log\Lambda/x$ with a prefactor independent of α and β. A precise calculation shows that

$$F^s_\Lambda(x)=\alpha g_0^2\log\frac{\Lambda}{x}+\alpha^2 g_0^3\log^2\frac{\Lambda}{x}+\alpha^3 g_0^4\log^3\frac{\Lambda}{x}+\cdots$$
$$+\beta g_0^3\log\frac{\Lambda}{x}\quad+\frac{5}{2}\alpha\beta g_0^4\log^2\frac{\Lambda}{x}+\cdots$$
$$+\gamma g_0^4\log\frac{\Lambda}{x}+\cdots. \tag{36}$$

We have written the series so as to exhibit its "triangular" nature: the first line corresponds to the leading logarithms, the second to the sub-leading, etc., and the nth column to the $(n+1)$th order of perturbation. The leading logarithms are entirely controlled by the g_0^2 term, the sub-leading logarithms by both the g_0^2 and g_0^3 terms, etc. It is clear that order by order for the divergent terms, only the log term is new, all the $\log^2$, $\log^3$, etc., terms are determined by the preceding orders. This structure strongly suggests that we can, at least partially, resum the perturbation series. We notice that although the leading logarithms form a simple geometric series, this is no longer true for the sub-leading logarithms where, for instance, the factor $5\alpha\beta/2$ of Eq. (36) is non-trivial. Thanks to the renormalization group, there exists a systematic way to perform these resummations[27] (see the following).

We again emphasize that for our simple toy model the divergences together with dimensional analysis determine almost entirely the entire function $F(x)$ in the limit of large Λ. To show this explicitly, we rewrite F as

$$F_\Lambda(x,g_0,\Lambda)=g_0+F^s_\Lambda(x,g_0,\Lambda)+F^r_\Lambda(x,g_0,\Lambda) \tag{37}$$

with $F^s_\Lambda(x,g_0,\Lambda)$ given by Eq. (36) at $O(g_0^4)$ and $F^r_\Lambda(x,g_0,\Lambda)\sim O(g_0^2)$. From dimensional analysis, $F^r_\Lambda(x,g_0,\Lambda)$ is also a function of x/Λ only which, by definition, is finite when $\Lambda\to\infty$. Thus, for large Λ,

$$F^r_\Lambda(x,g_0,\Lambda)=\mathcal{F}\left(\frac{x}{\Lambda},g_0\right)\simeq\mathcal{F}(0,g_0). \tag{38}$$

$F^r_\Lambda(x)$ is therefore almost x-independent for large Λ: it is a (g_0-dependent) number in this limit. For the sake of simplicity, let us consider the case where it is vanishing:

$$F_\Lambda(x,g_0,\Lambda)=g_0+F^s_\Lambda(x,g_0,\Lambda) \tag{39}$$

with $F^s_\Lambda(x,g_0,\Lambda)$ a function of x/Λ only. By using the renormalization prescription, Eq. (3), we can calculate g_R as a function of g_0 and Λ/μ and by formally inverting the series, we obtain at $O(g_R^4)$:

$$g_0=g_R-\alpha g_R^2\log\frac{\Lambda}{\mu}+g_r^3\left[\alpha^2\log^2\frac{\Lambda}{\mu}-\beta\log\frac{\Lambda}{\mu}\right]$$
$$+g_R^4\left[-\gamma\log\frac{\Lambda}{\mu}+\frac{5}{2}\alpha\beta\log^2\frac{\Lambda}{\mu}-\alpha^3\log^3\frac{\Lambda}{\mu}\right]. \tag{40}$$

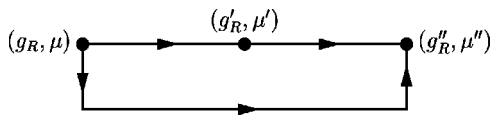

Fig. 1. An illustration of the renormalization group: the two equivalent ways to compose changes of parametrizations.

By substituting this expression in Eqs. (36) and (39), we obtain at $O(g_R^4)$:

$$F_\Lambda(x)=g_R+\alpha g_R^2\log\frac{\mu}{x}+\alpha^2 g_R^3\log^2\frac{\mu}{x}+\alpha^3 g_R^4\log^3\frac{\mu}{x}\cdots$$
$$+\beta g_R^3\log\frac{\mu}{x}\quad+\frac{5}{2}\alpha\beta g_R^4\log^2\frac{\mu}{x}\cdots$$
$$+\gamma g_R^4\log\frac{\mu}{x}\cdots. \tag{41}$$

Thus, we find that the renormalization process leaves unchanged the functional form of F_Λ, Eq. (36), and just consists in replacing (g_0,Λ) by (g_R,μ). This very important fact is related to a self-similarity property that we study in detail from the renormalization group viewpoint. Notice that of course any explicit dependence on Λ and g_0 has been eliminated in Eq. (41) and that the limit $\Lambda\to\infty$ can now be safely taken, if desired.

Note that we have obtained logarithmic divergences because we have studied the renormalization of a dimensionless coupling constant. If g_0 was dimensional, we would have obtained power law divergences. This is for instance what happens in QFT for the mass terms [see also in the following the expansion in Eq. (45)].

IV. RENORMALIZATION GROUP

Although the renormalization group will allow us to partially resum the perturbation expansion, we shall not introduce it in this way. Rather, we want to examine the internal consistency of the renormalization procedure.

We have chosen a renormalization prescription at the point $x=\mu$ where g_R is defined. Obviously, this point is not special, and we could have chosen any other point μ' or μ'' to parametrize the theory. Because there is only one independent coupling constant, the different coupling constants $g_R=g_R(\mu)$, $g'_R=g_R(\mu')$, $g''_R=g_R(\mu'')$ should all be related in such a way that $F(x)=F(x,\mu,g_R)=F(x,\mu',g'_R)=F(x,\mu'',g''_R)$, etc. This means that there should exist an equivalence class of parametrizations of the same theory and that it should not matter in practice which element in the class is chosen. This independence of the physical quantity with respect to the choice of prescription point also means that the changes of parametrizations should be a (renormalization) group law: going from the parametrization given by (μ,g_R) to that given by (μ',g'_R) and then to that given by (μ'',g''_R) or going directly from the first parametrization (μ,g_R) to the last one (μ'',g''_R) should make no difference, see Fig. 1.

Put this way, this statement seems to be void. Actually, it is. More precisely, it would be so if we were performing exact calculations: we would gain no new physical information by implementing the renormalization group law. This is because this group law does not reflect a symmetry of the physics, but only of the parametrization of our solution. This situation is completely analogous to what happens for the solution of a differential equation: we can parametrize it at time t in terms of the initial conditions at time t_0 for instance, or we can use the equation itself to calculate the solution at an intermediate time τ and then use this solution

as a new initial condition to parametrize the solution at time t. The changes of initial conditions that preserve the final solution can be composed thanks to a group law. Let us consider, for example, the following trivial, but illuminating, example:

$$\dot{y}(t)=\epsilon y(t), \quad y(t_0)=r_0, \tag{42}$$

the solution of which is

$$y(t)=f(r_0,t-t_0)=r_0 e^{\epsilon(t-t_0)}. \tag{43}$$

The group law can be written as[28]

$$f(r_0,t-t_0)=f(f(r_0,\tau-t_0),t-\tau) \quad \forall\,\tau, \tag{44}$$

which you can verify using the exact solution, Eq. (43). The non-trivial point with these group laws is that, in general, they are violated at any finite order of the perturbation expansions. In our previous example, we obtain to order ϵ,

$$y(t)\simeq f_1(r_0,t-t_0)=r_0(1+\epsilon(t-t_0)), \tag{45}$$

and

$$f_1(f_1(r_0,\tau-t_0),t-\tau)=r_0(1+\epsilon(t-t_0)) + \epsilon^2 r_0(t-\tau)(\tau-t_0). \tag{46}$$

The group law is verified to order ϵ because the perturbation expansion is exact at this order. However, it is violated by a term of order ϵ^2 that can be arbitrarily large even for small ϵ, provided $t-t_0$ is large enough.

The interest of the group law, Eq. (44), is that it is possible to enforce it and then to improve the perturbation result. Actually, when renormalization is necessary, the group law lets us partially resum the perturbation series of divergent terms.

Let us now see how this improvement of the perturbation series works for the example of the differential equation (42). In this case, the divergence occurs for $t_0\to-\infty$. Thus, t_0 plays the role of the cut-off Λ, $t-t_0$ of $\log\Lambda/\mu$, and $t-\tau$ of $\log\mu'/\mu$. Once t_0 is finite, no divergence remains, but the relics of the divergences occurring for $t_0\to-\infty$ are the large violations of the group law because both the divergences and these violations originate in the fact that the perturbation expansion is performed in powers of $\epsilon(t-t_0)$ and not of ϵ. To further study the relevance of the group law, it is interesting to forget the higher order terms of the perturbation expansion for a while and to look for an improved approximation that coincides at order ϵ with the perturbation result and that obeys the group law at order ϵ^2:

$$f_1^{\mathrm{imp}}(r_0,t-t_0)=r_0(1+\epsilon(t-t_0)+\epsilon^2 G(t-t_0)). \tag{47}$$

By imposing the group law, Eq. (44), to order ϵ^2, we obtain a functional equation for G:

$$G(t-t_0)=G(\tau-t_0)+G(t-\tau)+(\tau-t_0)(t-\tau). \tag{48}$$

If we differentiate Eq. (48) with respect to t_0 and take $t_0=\tau$, we obtain, setting $x=t-\tau$,

$$G'(x)=x+G'(0). \tag{49}$$

Because $G(0)=0$, Eq. (49) implies that

$$G(x)=\frac{x^2}{2}+ax, \tag{50}$$

where a is arbitrary. For $a=0$, this result is actually the perturbation result to order ϵ^2 because

$$y(t)\simeq r_0\left(1+\epsilon(t-t_0)+\frac{\epsilon^2}{2}(t-t_0)^2\right)+O(\epsilon^3). \tag{51}$$

Thus, the first order in the perturbation expansion, together with the group law, determines entirely the term of highest degree in $t-t_0$ at the next order. Of course, to verify exactly the group law, we should pursue the expansion in ϵ to all orders. It is easy to show that to order ϵ^n, the term of highest degree in $t-t_0$ is completely determined by both the first-order result and the group law and coincides with the perturbation result: $\epsilon^n(t-t_0)^n/n!$. Thus, the only information given by the perturbation expansion is that all subdominant terms, $\epsilon^n(t-t_0)^p$ with $p<n$, vanish in this example. We could now show how the implementation of the group law lets us resum the perturbation expansion. Unfortunately, this example is too simple and some important features of the renormalization group are missed in this case. (See Appendix E for a complete discussion of the implementation of the renormalization group on this example.) We therefore go back to our toy model for which we specialize to renormalizable theories with dimensionless couplings.

Renormalization group for renormalizable theories with dimensionless couplings. We now reconsider our toy model, Eqs. (4), (36), and (37), from the point of view of the renormalization group. For the sake of simplicity, we keep only the dominant terms at each order, that is, apart from g_0, the divergent ones in Eq. (39).

First, notice that in the same way g_R is clearly associated with the scale μ, Eq. (3), so is g_0 with the scale Λ because from Eq. (36), we find[29]

$$F_\Lambda(x=\Lambda)=g_0. \tag{52}$$

Let us define a third coupling constant associated with the scale μ',

$$F_\Lambda(\mu')=g_R', \tag{53}$$

and study the relationship between these different coupling constants at order g_0^2. From

$$F_\Lambda(x,g_0,\Lambda)=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{x}\right)+O(g_0^3), \tag{54}$$

we obtain

$$g_R=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{\mu}\right)+O(g_0^3), \tag{55}$$

$$g_R'=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{\mu'}\right)+O(g_0^3). \tag{56}$$

By eliminating g_0 between these two equations, we find

$$g_R'=g_R+\alpha g_R^2\log\left(\frac{\mu}{\mu'}\right)+O(g_R^3), \tag{57}$$

and thus, as expected, the group law controlling the change of prescription point is verified perturbatively. We note that the essential ingredient for this composition law is that Eq. (57) is independent of Λ. This is what ensures that the same form can be used to change (g_0,Λ) into (g_R,μ) and then (g_R,μ) into (g_R',μ'). This independence, in turn, is nothing but the signature of perturbative renormalizability which lets us completely eliminate at each order (g_0,Λ) for (g_R,μ). Perturbatively, everything looks fine. However, the previous calculation relies on a formal step that is not mathematically

correct, at least for large Λ. Indeed, to go from Eq. (56) to Eq. (57), the series $g_R=g_R(g_0)$ must be inverted to find $g_0=g_0(g_R)$ while, for $\Lambda\to\infty$, the series $g_R=g_R(g_0)$ is clearly not convergent and thus not invertible. Thus, the neglected terms of order g_R^3 in Eq. (57) involve a term proportional to $\log\Lambda/\mu\log\mu'/\mu$—analogous to the term $(t-\tau)(\tau-t_0)$ of Eqs. (46) and (48)—which is neglected because it is of order g_R^3, but which is very large for large Λ (see Appendix D). From a practical point of view, the existence at any order of these large terms of higher orders spoil the group law so that the independence of the physical results with respect to the choice of prescription point is not verified.

As in the case of the differential equation (47), we can look for an improved function: F^{imp},

$$F^{\text{imp}}(x,g_0,\Lambda)=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{x}\right)+g_0^3G\left(\frac{\Lambda}{x}\right)+O(g_0^4), \tag{58}$$

for which the group law at order g_0^3 is obeyed. It is shown in Appendix D that this constraint implies that

$$G(x)=\alpha^2\log^2 x+\beta\log x, \tag{59}$$

where β is arbitrary. Thus,

$$F^{\text{imp}}(x,g_0,\Lambda)=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{x}\right)+\alpha^2g_0^3\log^2\left(\frac{\Lambda}{x}\right)+\beta g_0^3\log\left(\frac{\Lambda}{x}\right)+O(g_0^4). \tag{60}$$

Once again, we find that the group law together with the order g_0^2 result determines the leading behavior at the next order, here the $\log^2(\Lambda/x)$ term. Moreover, we find that the group law imposes the existence of the same $\log^2$ term as the one found from the renormalizability constraint, Eqs. (35) and (36), and allows the existence of a sub-leading logarithm. Although nontrivial, this should not be too surprising because the renormalizability constraint means that once F is well defined at $x=\mu$, it also is everywhere and in particular at $x=\mu'$. The renormalizability constraint is therefore certainly necessary for the implementation of the group law. As in the example of the differential equation, Eq. (42), we should pursue the expansion to all orders to obtain an exactly verified group law. It is clear that by doing so, we would find the same expansion as the one obtained from the renormalizability constraint. Thus, if we use perturbation theory to calculate the coefficient in front of the first leading logarithm (of order g_0^2) and impose the group law, we should be able to resum all the leading logarithms. To do the resummation of the sub-leading and sub-sub-leading logarithms, a knowledge of, respectively, the order g_0^3 and g_0^4 terms is required. Clearly, we need to understand how to systematically construct the function f giving g_R' in terms of g_R and μ/μ',[27]

$$g_R'=f\left(g_R,\frac{\mu}{\mu'}\right), \tag{61}$$

such that

- its expansion at order n is given by the nth order of perturbation theory,
- the group law is *exactly* verified:

$$f\left(g_R,\frac{\mu}{\mu''}\right)=f\left(f\left(g_R,\frac{\mu}{\mu'}\right),\frac{\mu'}{\mu''}\right). \tag{62}$$

The function f is then said to be the *self-similar approximation* at order n of the exact relationship between g_R and g_R'.[30] First notice one crucial thing. Our first aim was to study the perturbation expansion of a function F in a power series of a coupling constant g_0. Then we have discovered that the logarithmic divergence at order g_0^2 propagates to all orders so that the expansion is actually performed in $g_0\log\Lambda/\mu$ instead of g_0. Because Λ is the regulator, it is supposed to be very large compared with μ, so that the large logarithmic terms invalidate the use of the perturbation expansion. Reciprocally, it is clear that perturbation theory is perfectly valid if it is performed between two scales μ_1 and μ_2 which are very close. Thus, instead of using perturbation theory to make a big jump between two very distinct scales, say Λ and μ, we should use it to perform a series of very little steps for which it is valid at each of them. In geometrical terms, the fact that the perturbative approach is valid only between two very close scales means that we should not use perturbation theory to approximate the equation of the curve given by the function f, Eq. (61), that joins the points (μ,g_R) and (μ',g_R'), but we should use it to calculate the (field of) tangent vectors to this curve, that is, its envelope. The curve itself should then be reconstructed by integration, see Appendix E. By doing so, the group law will be automatically verified because, by construction, the integration precisely consists in composing infinitesimal changes of reparametrization infinitely many times. Let us consider again Eq. (55). We want to calculate the evolution of $g_R(\mu)$ with μ for a given model specified by (Λ,g_0). Thus we define

$$\beta(g_R)=\mu\left.\frac{\partial g_R}{\partial\mu}\right|_{g_0,\Lambda}, \tag{63}$$

which gives the infinitesimal evolution of the coupling constants with the scale for the model corresponding to (g_0,Λ). We trivially find to this order from Eq. (55),

$$\beta(g_R)=-\alpha g_0^2+O(g_0^3), \tag{64}$$

and thus, by trivially inverting the series of Eq. (55), we obtain

$$\beta(g_R)=-\alpha g_R^2+O(g_R^3). \tag{65}$$

Now, if we integrate Eq. (63) together with Eq. (65), we obtain

$$g_R'=\frac{g_R}{1-\alpha g_R\log\dfrac{\mu}{\mu'}}. \tag{66}$$

This relation has several interesting properties.

(i) When expanded to order g_R^2, the perturbation result to this order is recovered, Eq. (57). This is quite normal because $\beta(g_R)$ has been calculated to this order.

(ii) When expanded to all orders, the whole series of leading logarithms is recovered. This is more interesting because $\beta(g_R)$ has been calculated only to order g^2, but simply means that all the leading logarithms are determined by the first one.

(iii) The group law (62) is obeyed exactly. We have thus found the function f of Eq. (61) to this order. It is very

instructive to check the group law directly from Eq. (66) and to verify that the β function found in Eq. (65) is not modified if we add the leading logarithmic term of order g_0^3 to relation (55):

$$g_R = g_0 + \alpha g_0^2 \log\left(\frac{\Lambda}{\mu}\right) + \alpha^2 g_0^3 \log^2\left(\frac{\Lambda}{\mu}\right) + O(g_0^4). \qquad (67)$$

The independence of the β function with respect to the addition of the successive leading logarithmic terms means that this function is indeed the right object to build self-similar approximations out of the perturbation expansion.

Let us now return to the β function itself. First, we have calculated the logarithmic derivative $\mu \partial g_R/\partial \mu$ instead of the ordinary derivative with respect to μ because we wanted to have a dimensionless β function. Second, even the dimensionless quantity, $\beta(g_R)$, could have depended on Λ/μ. However, the evolution of $g_R(\mu)$ between μ and $\mu + d\mu$ cannot depend in perturbation theory on Λ because the theory is perturbatively renormalizable: the perturbative relation between $g_R(\mu)$ and $g_R(\mu')$ depends only on μ and μ' and not on Λ. Thus, being dimensionless, the β function cannot depend on μ alone and is thus only a function of g_R. This property is general for any renormalizable theory: in the space of coupling constants, the β function is always a *local* function. Third, the β function is the function to be expanded in perturbation theory because it is given by a true series in g_R and not in $g_R \log \Lambda/\mu$. This is clear for our example, Eq. (65), where there is no logarithm, and can be proven formally by the following argument. If we use Eqs. (61) and (63), we find that

$$\beta(g_R) = -\frac{\partial f}{\partial y}(g_R, y)_{|_{y=1}}. \qquad (68)$$

If f is a double series in g and in $\log(\mu/\mu')$,

$$f\left(g_R, \frac{\mu}{\mu'}\right) = \sum_{n,p} \alpha_{n,p} g_R^n \log^p \frac{\mu}{\mu'}, \qquad (69)$$

it is clear from Eq. (68) that only terms with $p=1$ contribute to $\beta(g)$, with the logarithm replaced by -1. Thus we immediately deduce from this argument and from Eq. (36) that

$$\beta(g_R) = -\alpha g_R^2 - \beta g_R^3 - \gamma g_R^4 + O(g_R^5). \qquad (70)$$

It is easy to check that the first two coefficients, $-\alpha$ and $-\beta$, are universal in the sense that for two different theories, parametrized by (g_R, μ) and (g_R', μ), the two β functions have the same first two coefficients in their expansions.

This method of computing the β function also lets us bypass the strange way to calculate it that we have used in Eqs. (64) and (65), where we have first expressed g_R in terms of g_0 to calculate $\beta(g_R)$ as a function of g_0 and then, by inversion of the series, re-obtained a function of g_R. These two steps are *a priori* dangerous because they both involve large logarithms. Actually, they always cancel each other. This can be seen directly for the example of Eq. (67) and the reason for this cancellation comes from Eqs. (68) and (69), which show that no inversion of series is needed to calculate $\beta(g_R)$. There is no miracle here, because only the behavior at $y = \mu/\mu' = 1$, which of course does not involve Λ, matters.

Finally, we mention that the integration of the β function at $O(g_R^3)$—analogous to a two-loop result in QFT—leads to an implicit equation for g_R' that generalizes Eq. (66):

$$\frac{1}{g_R'} - \frac{1}{g_R} + \frac{\beta}{\alpha} \log\left(\frac{g_R}{g_R'} \frac{\alpha + \beta g_R'}{\alpha + \beta g_R}\right) = \alpha \log \frac{\mu'}{\mu}. \qquad (71)$$

There is no simple solution of this transcendental equation. It is however possible to obtain an iterative solution that is valid if the $O(g_R^3)$ term is small compared with the $O(g_R^2)$ one, that is, if $g_R \beta/\alpha \ll 1$. It is obtained by replacing g_R' in the third term of Eq. (71) by its expression obtained to order g_R^2, Eq. (66):

$$g_R' = \frac{g_R}{1 - \alpha g_R \log \frac{\mu}{\mu'} + \frac{\beta}{\alpha} g_R \log\left(1 - \alpha g_R \log \frac{\mu}{\mu'}\right)}. \qquad (72)$$

It is easy to check that Eq. (72) resums exactly all the leading and sub-leading logarithms of the perturbation expansion Eq. (41). Note that contrary to the one-loop result, Eq. (66), which resums only the leading logarithms, the exact expression in Eq. (71) contributes also to the sub-sub-leading logarithms as well as the sub-sub-sub-leading ones and so on and so forth.

V. SUMMARY

(1) The long way of renormalization starts with a theory depending on only one parameter g_0, which is the small parameter in which perturbation series are expanded. In particle physics, this parameter is in general a coupling constant like an electric charge involved in a Hamiltonian (more precisely the fine structure constant for electrodynamics). This parameter is also the first order contribution of a physical quantity F. In particle/statistical physics, F is a Green/correlation function. The first order of perturbation theory neglects fluctuations—quantum or statistical—and thus corresponds to the classical/mean field approximation. The parameter g_0 also is to this order a measurable quantity because it is given by a Green function. Thus, it is natural to interpret it as the unique and physical coupling constant of the problem. If, as we suppose in the following, g_0 is dimensionless, so is F. Moreover, if x is dimensional—it represents momenta in QFT—it is natural that F does not depend on it as is found in the classical theory, that is, at first order of the perturbation expansion.

(2) If F does depend on x, as we suppose it does at second order of perturbation theory, it must depend on another dimensional parameter, Λ, through the ratio of x and Λ. If we have not included this parameter from the beginning in the model, the x-dependent terms are either vanishing, which is what happens at first order, or infinite as they are at second and higher orders. This is the very origin of divergences (from the technical point of view).

(3) These divergences require that we regularize F. This requirement, in turn, requires the introduction of the scale Λ that was missing. In the context of field theory, the divergences occur in Feynman diagrams for high momenta, that is, at short distances. The cut-off Λ suppresses the fluctuations at short distances compared with Λ^{-1}. In statistical physics, this scale, although introduced for formal reasons, has a natural interpretation because the theories are always effective theories built at a given microscopic scale. It corresponds in general to the range of interaction of the constituents of the model, for example, a lattice spacing for spins, the average intermolecular distance for fluids. In particle

physics, things are less simple. At least psychologically. It was indeed natural in the early days of quantum electrodynamics to think that this theory was fundamental, that is, not derived from a more fundamental theory. More precisely, it was believed that QED had to be mathematically internally consistent, even if in the real world new physics had to occur at higher energies. Thus, the regulator scale was introduced only as a trick to perform intermediate calculations. The limit $\Lambda \to \infty$ was supposed to be the right way to eliminate this unwanted scale, which anyway seemed to have no interpretation. We shall see in the following that the community now interprets the renormalization process differently.

(4) Once the theory is regularized, F can be a nontrivial function of x. The price is that different values of x now correspond to different values of the coupling constant (defined as the values of F for these x). Actually, it no longer makes sense to speak of a coupling constant in itself. The only meaningful concept is the pair $(\mu, g_R(\mu))$ of coupling constants at a given scale. The relevant question now is, ''What are the physical reasons in particle/statistical physics that make the coupling constants depend on the scale while they are constants in the classical/mean field approximation?'' As mentioned, for particle physics, the answer is the existence of new quantum fluctuations corresponding to the possibility of creating (and annihilating) particles at energies higher than mc^2. What was scale independent in the classical theory becomes scale dependent in the quantum theory because, as the available energy increases, more and more particles can be created. The pairs of (virtual) particles surrounding an electron are polarized by its presence and thus screen its charge. As a consequence, the charge of an electron depends on the distance (or equivalently the energy) at which it is probed, at least for distances smaller than the Compton wavelength.

Note that the energy scale mc^2 should not be confused with the cut-off scale Λ. mc^2 is the energy scale above which quantum fluctuations start to play a significant role while Λ is the scale where they are cut-off. Thus, although the Compton wavelength is a short distance scale for the classical theory, it is a long distance scale for QFT, the short one being Λ^{-1}. There are thus three domains of length scales in QFT: above the Compton wavelength where the theory behaves classically (up to small quantum corrections coming from high energy virtual processes), between the Compton wavelength and the cut-off scale Λ^{-1} where the relativistic and quantum fluctuations play a great role, and below Λ^{-1} where a new, more fundamental theory has to be invoked.[12] In statistical physics, the analog of the Compton wavelength is the correlation length which is a measure of the distance at which two microscopic constituents of the system are able to influence each other through thermal fluctuations.[31] For the Ising model, for instance, the correlation length away from the critical point is the order of the lattice spacing and the corrections to the mean-field approximation due to fluctuations are small. Unlike particle physics where the masses and therefore the Compton wavelengths are fixed, the correlation lengths in statistical mechanics can be tuned by varying the temperature. Near the critical temperature where the phase transition takes place, the correlation length becomes extremely large and fluctuations on all length scales between the microscopic scale of order Λ^{-1}, a lattice spacing, and the correlation length add up to modify the mean-field behavior (see Refs. 32, 33 and also Ref. 34 for a bibliography on this subject). We see here a key to the relevance of renormalization: two very different scales must exist between which a nontrivial dynamics (quantum or statistical in our examples) can develop. This situation is *a priori* rather unnatural as can be seen for phase transitions, where a fine tuning of temperature must be implemented to obtain correlation lengths much larger than the microscopic scale. Most of the time, physical systems have an intrinsic scale (of time, energy, length, etc.) and all the other relevant scales of the problem are of the same order. All phenomena occurring at very different scales are thus almost completely suppressed. The existence of a unique relevant scale is one of the reasons why renormalization is not necessary in most physical theories. In QFT it is mandatory because the masses of the known particles are much smaller than a hypothetical cut-off scale Λ, still to be discovered, where new physics should take place. This is a rather unnatural situation, because, contrary to phase transitions, there is no analog of a temperature that could be fine-tuned to create a large splitting of energy, that is, mass, scales. The question of naturalness of the models we have at present in particle physics is still largely open, although there has been much effort in this direction using supersymmetry.

(5) The classical theory is valid down to the Compton/correlation length, but cannot be continued naively beyond this scale; otherwise, when mixed with the quantum formalism, it produces divergences. Actually, it is known in QFT that the fields should be considered as distributions and not as ordinary functions. The need for considering distributions comes from the nontrivial structure of the theory at very short length scale where fluctuations are very important. At short distances, functions are not sufficient to describe the field state, which is not smooth but rough, and distributions are necessary. Renormalizing the theory consists actually in building, order by order, the correct ''distributional continuation'' of the classical theory. The fluctuations are then correctly taken into account and depend on the scale at which the theory is probed: this nontrivial scale dependence can only be taken into account theoretically through the dependence of the (analog of the) function F with x and thus of the coupling with the scale μ.

(6) If the theory is perturbatively renormalizable, the pairs $(\mu, g(\mu))$ form an equivalence class of parametrizations of the theory. The change of parametrization from $(\mu, g(\mu))$ to $(\mu', g(\mu'))$, called a renormalization group transformation, is then performed by a law which is self-similar, that is, such that it can be iterated several times while being form-invariant.[27,30] This law is obtained by the integration of

$$\beta(g_R) = \mu \left. \frac{\partial g_R}{\partial \mu} \right|_{g_0, \Lambda}. \tag{73}$$

This function has a true perturbation expansion in terms of g_R unlike the perturbative relation between $g_R(\mu)$ and $g_R(\mu')$ which involves logarithms of μ/μ' that can be large. The integration of Eq. (73) partially resums the perturbation series and is thus semi-nonperturbative even if $\beta(g_R)$ has been calculated perturbatively. The self-similar nature of the group law is encoded in the fact that $\beta(g_R)$ is independent of Λ.[5]

In particle physics, the β function gives the evolution of the strength of the interaction as the energy at which it is probed varies and the integration of the β function resums partially the perturbation expansion. First, as the energy increases, the coupling constant can decrease and eventually

vanish. This is what happens when $\alpha>0$ in Eqs. (65) and (66). In this case, the particles almost cease to interact at very high energies or equivalently when they are very close to each other. The theory is then said to be asymptotically free in the ultraviolet domain.[3,5] Reciprocally, at low energies the coupling increases and perturbation theory can no longer be trusted. A possible scenario is that bound states are created at a sufficiently low energy scale so that the perturbation approach has to be reconsidered in this domain to take into account these new elementary excitations. Non-Abelian gauge theories are the only known theories in four space–time dimensions that are ultraviolet free, and it is widely believed that quantum chromodynamics—which is such a theory—explains quark confinement. The other important behavior of the scale dependence of the coupling constant is obtained for $\alpha<0$ in which case it increases at high energies. This corresponds, for instance, to quantum electrodynamics. For this kind of theory, the dramatic increase of the coupling at high energies is supposed to be a signal that the theory ceases to be valid beyond a certain energy range and that new physics, governed by an asymptotically free theory (like the standard model of electro-weak interactions), has to take place at short distances.

(7) Renormalizability, or its nonperturbative equivalent, self-similarity, ensures that although the theory is initially formulated at the scale Λ, this scale together with g_0 can be entirely eliminated for another scale better adapted to the physics we study. If the theory was solved exactly, it would make no difference which parametrization we used. However, in perturbation theory, this renormalization lets us avoid calculating small numbers as differences of very large ones. It would indeed be very unpleasant, and actually meaningless, to calculate energies of order 100 GeV, for instance—the scale μ of our analysis—in terms of energies of order of the Planck scale $\simeq 10^{19}$ GeV, the analog of the scale Λ. In a renormalizable theory, the possibility to perturbatively eliminate the large scale has a very deep meaning: it is the signature that the physics is short distance insensitive or equivalently that there is a decoupling of the physics at different scales. The only memory of the short distance scale lies in the initial conditions of the renormalization group flow, not in the flow itself: the β function does not depend on Λ. We again emphasize that, usually, the decoupling of the physics at very different scales is trivially related to the existence of a typical scale such that the influence of all phenomena occurring at different scales is almost completely suppressed. Here, the decoupling is much more subtle because there is no typical length in the whole domain of length scales that is very small compared with the Compton wavelength and very large compared with Λ^{-1}. Because interactions among particles correspond to nonlinearities in the theories, we could naively believe that all scales interact with each other—which is true—so that calculating, for instance, the low energy behavior of the theory would require the detailed calculation of all interactions occurring at higher energies. Needless to say that in a field theory, involving infinitely many degrees of freedom—the value of the field at each point—such a calculation would be hopeless, apart from exactly solvable models. Fortunately, such a calculation is not necessary for physical quantities that can be calculated from renormalizable couplings only. Starting at very high energies, typically Λ, where all coupling constants are naturally of order 1, the renormalization group flow drives almost all of them to zero, leaving only, at low energies, the renormalizable couplings. This is the interpretation of nonrenormalizable couplings. They are not terrible monsters that should be forgotten as was believed in the early days of QFT. They are simply couplings that the RG flow eliminates at low energies. If we are lucky, the renormalizable couplings become rather small after their RG evolution between Λ and the scale μ at which we work, and perturbation theory is valid at this scale.

We see here the phenomenon of universality: among the infinitely many coupling constants that are *a priori* necessary to encode the dynamics of the infinitely many degrees of freedom of the theory, only a few ones are finally relevant.[35] All the others are washed out at large distances. This is the reason why, perturbatively, it is not possible to keep these couplings finite at large distance, and it is necessary to set them to zero.[36] The simplest nontrivial example of universality is given by the law of large numbers (the central limit theorem) which is crucial in statistical mechanics.[32] In systems where it can be applied, all the details of the underlying probability distribution of the constituents of the system are irrelevant for the cooperative phenomena which are governed by a Gaussian probability distribution.[37] This drastic reduction of complexity is precisely what is necessary for physics because it lets us build effective theories in which only a few couplings are kept.[12] Renormalizability in statistical field theory is one of the nontrivial generalizations of the central limit theorem.

(8) The cut-off Λ, first introduced as a mathematical trick to regularize integrals, has actually a deep physical meaning: it is the scale beyond which new physics occur and below which the model we study is a good effective description of the physics. In general, it involves only the renormalizable couplings and thus cannot pretend to be an exact description of the physics at all scales. However, if Λ is very large compared with the energy scale in which we are interested, all nonrenormalizable couplings are highly suppressed and the effective model, retaining only renormalizable couplings, is valid and accurate (the Wilson RG formalism is well suited to this study, see Refs. 35 and 38). In some models—the asymptotically free ones—it is possible to formally take the limit $\Lambda\to\infty$ both perturbatively and nonperturbatively, and there is therefore no reason to invoke a more fundamental theory taking over at a finite (but large) Λ. Let us emphasize here several interesting points.

(i) For a theory corresponding to the pair $(\mu,g_R(\mu))$, the limit $\Lambda\to\infty$ must be taken within the equivalence class of parametrizations to which $(\mu,g_R(\mu))$ belongs.[39] A divergent nonregularized perturbation expansion consists in taking $\Lambda=\infty$ while keeping g_0 finite. From this viewpoint, the origin of the divergences is that the pair $(\Lambda=\infty,g_0)$ does not belong to any equivalence class of a sensible theory. Perturbative renormalization consists in computing g_0 as a formal powers series in g_R (at finite Λ), so that (Λ,g_0) corresponds to a mathematically consistent theory; we then take the limit $\Lambda\to\infty$.

(ii) Because of universality, it is physically impossible to know from low energy data if Λ is very large or truly infinite.

(iii) Although mathematically consistent, it seems unnatural to reverse the RG flow while keeping only the renormalizable couplings and thus to imagine that even at asymptotically high energies, Nature has used

only the couplings that we are able to detect at low energies. It seems more natural that a fundamental theory does not suffer from renormalization problems. String theory is a possible candidate.[40]

To conclude, we see that although the renormalization procedure has not evolved much these last thirty years, our interpretation of renormalization has drastically changed:[12] the renormalized theory was assumed to be fundamental, while it is now believed to be only an effective one; Λ was interpreted as an artificial parameter that was only useful in intermediate calculations, while we now believe that it corresponds to a fundamental scale where new physics occurs; nonrenormalizable couplings were thought to be forbidden, while they are now interpreted as the remnants of interaction terms in a more fundamental theory. Renormalization group is now seen as an efficient tool to build effective low energy theories when large fluctuations occur between two very different scales that change the physics qualitatively and quantitatively.

ACKNOWLEDGMENTS

The author thanks S. Dusuel and B. Douçot for many interesting discussions and remarks. He also thanks J. Bartlett for his help in the choice of title of this article, and L. Canet, A. Elkharrat, E. Huguet, A. Laverne, P. Lecheminant, R. Mosseri, D. Mouhanna, M. Tissier, and N. Wschebor for a careful reading of the manuscript and many suggestions. He also thanks all the students of the Ninth Vietnam School of Physics (Hué, January 2003) for having worked hard on this article and for their suggestions. Laboratoire de Physique Théorique et Hautes Energies is a Laboratoire associé au CNRS: UMR 7589.

APPENDIX A: TOY MODELS FOR RENORMALIZABLE AND NONRENORMALIZABLE PERTURBATION EXPANSIONS

We give an example of a nonrenormalizable theory and of a theory which needs two couplings to be renormalized. Let us suppose that

$$F_{1,\Lambda}(x)=\alpha\int_1^\Lambda dt\,\frac{t}{t+x}, \tag{A1}$$

which, unlike the example of Eq. (5), is linearly divergent. To renormalize this function, we have to impose a prescription at one point, and we choose

$$F_\Lambda(0)=g_R. \tag{A2}$$

Note that it was not possible in the example of Eq. (5) to take $\mu=0$, because this choice would have lead to a divergence of the integral at the lower bound. In Eq. (A1) taking $\mu=0$ is possible because the lower bound of the integral is nonvanishing and actually plays somewhat the role of a nonvanishing μ. We have

$$\delta_2 g=-\alpha g_R^2\int_1^\Lambda dt, \tag{A3}$$

so that

$$F_\Lambda(x)=g_R-\alpha g_R^2 x\int_1^\Lambda\frac{dt}{t+x}, \tag{A4}$$

which is still (logarithmically) divergent for all $x\neq 0$. The difference between the two examples given in Eqs. (5) and (A1) is that in the last one, once the linear divergence has been subtracted, the logarithmic sub-divergence remains. Subtracting it would require us to impose a second prescription that would define a new coupling constant. In the absence of this second coupling constant, the logarithmic divergence cannot be subtracted and the model is nonrenormalizable.

Let us examine how a second coupling constant could solve the problem. Generically, this second coupling, which we call λ_0, already contributes at first order. We take as an example

$$F_\Lambda(x)=g_0+\lambda_0 x+\alpha g_0^2\int_1^\Lambda dt\,\frac{t}{t+x}+O(g_0^3). \tag{A5}$$

Let us take as renormalization prescriptions,

$$\frac{\partial F_\Lambda}{\partial x}(x=0)=\lambda_R, \tag{A6}$$

in addition to Eq. (A2). We obtain at first order that $g_0=g_R+O(g_R^2)$ and $\lambda_0=\lambda_R+O(g_R^2)$ and at second order

$$\delta_2 g=-\alpha g_R^2\int_1^\Lambda dt, \tag{A7}$$

$$\delta_2\lambda=\alpha g_R^2\int_1^\Lambda\frac{dt}{t}. \tag{A8}$$

If we substitute these expressions in Eq. (A5), we find

$$F_\Lambda(x)=g_R+\lambda_R x+\alpha g_R^2 x^2\int_1^\Lambda\frac{dt}{t(t+x)}+O(g_R^3). \tag{A9}$$

Obviously, this expression converges when $\Lambda\to\infty$. The two renormalization prescriptions let us subtract the linear divergence as well as the logarithmic sub-divergence. We emphasize that in the previous example we only eliminated the second divergence at order g_0^2. At higher orders, there are two ways a theory can behave, characterized by two different renormalizability properties. The first one is that all divergences can be removed to all orders by renormalizing only the two couplings g_0 and λ_0. A variant of this possibility is that a third coupling—or a finite number of new couplings—turns out to be necessary and sufficient to remove the divergences. In this case, the model is renormalizable at the price of introducing all the necessary couplings. The second possibility is that the new interaction term, which has induced the existence of the λ_0 term in F, generates new divergences at higher orders. In this case, new interaction terms (and coupling constants) are required to remove the new divergences. These new terms can themselves generate new divergences at even higher orders, which require new couplings to be removed and so on and so forth. In this case, infinitely many interaction terms are necessary to remove the divergences and the model is perturbatively nonrenormalizable.

APPENDIX B: DERIVATION OF EQ. (22)

Let us show that it is always possible to make the choice used in Eq. (22). Due to condition (15), we have generally

$$F^s_{1,\Lambda}(x)-F^s_{1,\Lambda}(\mu)\to g'_1(x,\mu)\ \text{ when }\ \Lambda\to\infty, \tag{B1}$$

where the limit g_1' is a well-defined function satisfying $g_1'(x,\mu)=-g_1'(\mu,x)$. If we first evaluate $F^s_{1,\Lambda}(x)$ in Eq. (B1) at $x=1$ for instance and at $\mu=1$ and subtract them, we obtain that g_1' has the following form:

$$g_1'(x,\mu)=g_1(x)-g_1(\mu), \tag{B2}$$

namely, a combination of the same function of x and μ. Then, by redefining $F^s_{1,\Lambda}$: $F^s_{1,\Lambda}\to F^s_{1,\Lambda}-g_1$ we satisfy Eq. (22). Note that the previous choice of singular part is not necessary and is only convenient.

APPENDIX C: LOGARITHMIC DIVERGENCES IN RENORMALIZABLE THEORIES WITH DIMENSIONLESS COUPLINGS

We prove for renormalizable theories with dimensionless couplings that $F^s_{1,\Lambda}(x)$ must be a logarithm. If we use Eq. (23), dimensional analysis, and the freedom to choose the regular part of $F_{1,\Lambda}$, we have

$$F^s_{1,\Lambda}(x)=f\left(\frac{x}{\Lambda}\right)=f_1(\Lambda)+r(x). \tag{C1}$$

Note that in full generality, the regular part we add to $f_1(\Lambda)$ could depend on Λ: $r_\Lambda(x)$. However, because it is regular, we can choose to add only the Λ-independent function corresponding to the $\Lambda\to\infty$ limit of r_Λ: $r(x)=r_\infty(x)$. If we differentiate Eq. (C1) with respect to x and then take $x=1$ and $\Lambda=1/y$, we obtain

$$f'(y)=\frac{r'(1)}{y}, \tag{C2}$$

and thus

$$f(y)=-\alpha\log y, \tag{C3}$$

where the minus sign has been written for convenience. From (C1) and (C3) we conclude that $f(x)=r(x)=-f_1(x)$ and that

$$F^s_{1,\Lambda}(x)=f\left(\frac{x}{\Lambda}\right)=f(x)-f(\Lambda)=\alpha\log\frac{\Lambda}{x}. \tag{C4}$$

APPENDIX D: RENORMALIZATION GROUP IMPROVED EXPANSION

We show how to derive Eq. (59). Consider the definition of $F^{\rm imp}$:

$$F^{\rm imp}(x,g_0,\Lambda)=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{x}\right)+g_0^3G\left(\frac{\Lambda}{x}\right)+O(g_0^4). \tag{D1}$$

We can calculate g_R and g_R' from their definitions (where $F^{\rm imp}$ is used instead of F) and from Eq. (D1):

$$g_R=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{\mu}\right)+g_0^3G\left(\frac{\Lambda}{\mu}\right)+O(g_0^4), \tag{D2}$$

$$g_R'=g_0+\alpha g_0^2\log\left(\frac{\Lambda}{\mu'}\right)+g_0^3G\left(\frac{\Lambda}{\mu'}\right)+O(g_0^4). \tag{D3}$$

If we invert the series $g_R=g_R(g_0)$ of Eq. (D2), we obtain

$$g_0=g_R-\alpha g_R^2\log\left(\frac{\Lambda}{\mu}\right)+2\alpha^2g_R^3\log^2\left(\frac{\Lambda}{\mu}\right)-g_R^3G\left(\frac{\Lambda}{\mu}\right)+O(g_R^4). \tag{D4}$$

Thus, substituting this expression for g_0 in $g_R'=g_R'(g_0)$, Eq. (D2), we obtain

$$g_R'=g_R+\alpha g_R^2\log\left(\frac{\mu}{\mu'}\right)+g_R^3\left[2\alpha^2\left(\log^2\left(\frac{\Lambda}{\mu}\right)-\log\left(\frac{\Lambda}{\mu}\right)\log\left(\frac{\Lambda}{\mu'}\right)\right)+G\left(\frac{\Lambda}{\mu'}\right)-G\left(\frac{\Lambda}{\mu}\right)\right]. \tag{D5}$$

The group law is obeyed at this order if the relation between g_R' and g_R is of the same form as the one between g_R and g_0, Eq. (D3). This condition requires

$$g_R'=g_R+\alpha g_R^2\log\left(\frac{\mu}{\mu'}\right)+g_R^3G\left(\frac{\mu}{\mu'}\right)+O(g_R^4), \tag{D6}$$

and thus

$$2\alpha^2\log\frac{\Lambda}{\mu}\log\frac{\mu'}{\mu}+G\left(\frac{\Lambda}{\mu'}\right)-G\left(\frac{\Lambda}{\mu}\right)=G\left(\frac{\mu}{\mu'}\right). \tag{D7}$$

By differentiating this relation with respect to Λ and by taking $\Lambda=\mu$, we find, setting $x=\mu/\mu'$:

$$G'(x)=2\alpha^2\frac{\log x}{x}+\frac{G'(1)}{x}. \tag{D8}$$

If we take into account that $G(1)=0$, we find by integration

$$G(x)=\alpha^2\log^2x+\beta\log x, \tag{D9}$$

where β is arbitrary.

APPENDIX E: THE RENORMALIZATION GROUP APPLIED TO A DIFFERENTIAL EQUATION

We show how the renormalization program can be implemented for the example of the differential equation (42) whose exact solution is

$$y(t)=f(r_0,t-t_0)=r_0e^{\epsilon(t-t_0)}. \tag{E1}$$

In perturbation theory, we find

$$y(t)=r_0\left(1+\epsilon(t-t_0)+\frac{\epsilon^2}{2}(t-t_0)^2+\cdots\right). \tag{E2}$$

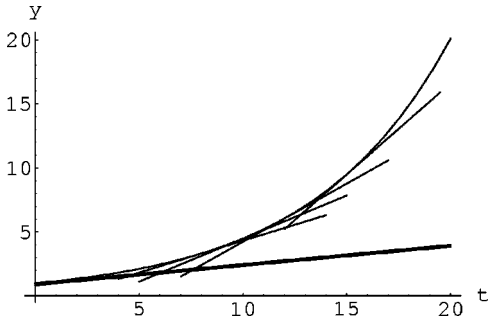

Fig. 2. The curve $y(t)$ as a function of t. The (thick) lower line is the approximation of order ϵ, see Eq. (E2). The other lines represent the (field of) tangent vectors to the curve—the envelope—given by the β function, Eq. (E7).

At order ϵ^0, $y(t)$ is constant and finite, whereas, at any higher order in ϵ, a divergence occurs for $t_0 \to -\infty$. This divergence arises of course from the fact that the expansion turns out to be in powers of $\epsilon(t-t_0)$ and not in powers of ϵ alone (the secular problem). Thus, as shown in Fig. 2, the approximation of order $O(\epsilon)$ becomes worse and worse as t increases. A renormalization prescription consists here in imposing that for a finite τ:

$$y(\tau) = r_\tau. \tag{E3}$$

If we perform the calculation to order ϵ, we find to first order:

$$r_\tau = r_0(1+\epsilon(\tau - t_0)) + O(\epsilon^2), \tag{E4}$$

and thus, as expected,

$$y(t) = r_\tau(1+\epsilon(t-\tau)) + O(\epsilon^2). \tag{E5}$$

The theory is perturbatively renormalizable at this order because by imposing a single renormalization prescription, it is possible to completely eliminate t_0 and r_0. Let us define the β-function for r_τ by

$$\beta(r_\tau) = \left.\frac{\partial r_\tau}{\partial \tau}\right|_{r_0, t_0} = \left.\frac{\partial f}{\partial \zeta}(r_\tau, \zeta)\right|_{\zeta = 0}. \tag{E6}$$

We find

$$\beta(r_\tau) = \epsilon r_\tau + O(\epsilon^2). \tag{E7}$$

It is very instructive to perform this calculation at higher orders because we then find that the $O(\epsilon)$ result of Eq. (E7) is exact [this result is trivially shown using the second equality of Eq. (E6)]. Thus, there is no $O(\epsilon^2)$ corrections to $\beta(r_\tau)$. This result means that, in this example, there is no sub-leading terms such as $\epsilon^n(t-t_0)^p$ with $p<n$ in the perturbation expansion.

Clearly, the β function gives the tangent to the curve $y(t)$. Equation (E7) shows that contrary to $y(t)$, the β function has a true ϵ expansion (involving only one term in this example). This result is reminiscent of what we have already observed in our general discussion, see Eqs. (36) and (70). Of course, this example is too simple because using the β function leads to the same differential equation for r_τ as the one for $y(t)$ that we started with, Eq. (42): the RG does not help us to solve ordinary differential equations. However, although mathematically trivial, our analysis shows that perturbation theory should not be used for large $t-t_0$, but that it is perfectly valid for infinitesimal time steps, see Fig. 2. It also shows that the higher order terms of the perturbation expansion are completely analogous to the series of the leading logarithms we have previously encountered: they are entirely determined by the $O(\epsilon)$ term together with self-similarity (encoded in the β function). Note finally that for partial differential equations (PDE) that describe the dynamics of infinitely many degrees of freedom (as in field theory), the RG techniques do not let us reconstruct the PDE from the first orders of perturbation theory. The β functions lead to ordinary differential equations, the integration of which let us improve the perturbation computation of several quantities thanks to a partial resummation of the perturbation expansion.[27,30]

[a]Electronic mail: delamotte@lpthe.jussieu.fr

[1] H. Bethe, "The electromagnetic shift of energy levels," Phys. Rev. **72**, 339–341 (1947).

[2] S. Weinberg, *The Quantum Theory of Fields* (Cambridge U.P., Cambridge, 1995).

[3] J. Collins, *Renormalization* (Cambridge U.P., Cambridge, 1984).

[4] L. Ryder, *Quantum Field Theory* (Cambridge U.P., Cambridge, 1985).

[5] M. Le Bellac, *Quantum and Statistical Field Theory* (Oxford U.P., Oxford, 1992).

[6] J. Binney, N. Dowrick, A. Fisher, and M. Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Oxford U.P., Oxford, 1992).

[7] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison–Wesley, Reading, MA, 1992).

[8] *Conceptual Foundations of Quantum Field Theory*, edited by T. Cao (Cambridge U.P., Cambridge, 1999).

[9] K. G. Wilson and J. Kogut, "The renormalization group and the ϵ-expansion," Phys. Rep. C **12**, 75–199 (1974).

[10] Some phase transitions are triggered by quantum fluctuations. This subtlety plays no role in what follows.

[11] The short distance physics in statistical systems is given by Hamiltonians describing, for instance, interactions among magnetic ions or molecules of a fluid.

[12] G. Lepage, "What is renormalization?," TASI'89 Summer School, Boulder, ASI, 1989, p. 483.

[13] P. Kraus and D. Griffiths, "Renormalization of a model quantum field theory," Am. J. Phys. **60**, 1013–1023 (1992).

[14] I. Mitra, A. DasGupta, and B. Dutta-Roy, "Regularization and renormalization in scattering from Dirac delta-potentials," Am. J. Phys. **66**, 1101–1109 (1998).

[15] P. Gosdzinsky and R. Tarrach, "Learning quantum field theory from elementary quantum mechanics," Am. J. Phys. **59**, 70–74 (1991).

[16] L. Mead and J. Godines, "An analytical example of renormalization in two-dimensional quantum mechanics," Am. J. Phys. **59**, 935–937 (1991).

[17] K. Adhikari and A. Ghosh, "Renormalization in non-relativistic quantum mechanics," J. Phys. A **30**, 6553–6564 (1997).

[18] Actually, the analog of the function in Eq. (1) would be a correlation function of four density or spin fields taken at four different points. These functions are not easily measurable and thus g_0 does not have in general an intuitive meaning in this case. Because this subtlety plays no role in our discussion, we ignore this difficulty in the following.

[19] Actually for QED it is a four-dimensional integral over four-momenta and the integrand is a product of propagators.

[20] R. Delbourgo, "How to deal with infinite integrals in quantum field theory," Rep. Prog. Phys. **39**, 345–399 (1976).

[21] S. Nyeo, "Regularization methods for delta-function potential in two-dimensional quantum mechanics," Am. J. Phys. **68**, 571–575 (2000).

[22] It is nontrivial to prove in full generality that the results obtained after renormalization are independent of the regularization scheme. However, it is easy to grasp the idea behind it. Because renormalization consists in eliminating parameters like g_0 and in replacing them by measurable couplings like g_R, the renormalized quantities like $F(x)$ are finally expressed only in terms of physical quantities that are independent of the regularization scheme (Ref. 15).

[23] M. Hans, "An electrostatic example to illustrate dimensional regularization and renormalization group technique," Am. J. Phys. **51**, 694–698 (1983).

[24] Note that the (renormalized) series in g_R can themselves be nonconvergent. Most of the time they are at best asymptotic. In some cases they can be resummed using Borel transform and Padé approximants.

[25] In QFT, it is in general also necessary to change the normalization of the analog of the function F—the Green functions—by a factor that diverges in the limit $\Lambda \to \infty$. This procedure is known as field renormalization.

[26] Let us emphasize that there is a subtlety if dimensional regularization is used. Actually, this regularization also introduces a dimensional parameter λ, which is not directly a regulator as is the cut-off Λ in the integral of Eq. (5). The analog of Λ in this regularization is given by $\Lambda = \lambda \exp(1/\epsilon)$, where $\epsilon = 4 - D$ and D is the spatial dimension. It often is convenient to take $\lambda \sim \mu$. We mention that dimensional regularization kills all nonlogarithmic divergences (Ref. 14).

[27] D. Shirkov and V. Kovalev, "Bogoliubov renormalization group and symmetry of solutions in mathematical physics," Phys. Rep. **352**, 219–249 (2001).

[28] The elements of the group are the functions: $g_t = f(\cdot, t)$ for $t \in \mathbb{R}$. They

transform an initial condition r_0 into the solution at a later time interval t of the differential equation we consider [see Eq. (43) in our example]: $g_t(r_0)=f(r_0,t)$. The composition law is thus $g_{t'}.g_t=f(f(\cdot,t),t')$. It obeys trivially the identity: $g_{t'}.g_t=g_{t+t'}$ which is nothing but Eq. (44). The identity is $g_{t=0}$ and the inverse is g_{-t}. The law is associative as it should be for a group.

[29]If we had not omitted in Eq. (52) the finite parts, we would have found $F(x=\Lambda)=g_0+ag_0^2+bg_0^3+\cdots$. Thus g_0 is in general not associated exactly with the scale Λ, but with Λ up to a factor of order unity.

[30]V. Kovalev and D. Shirkov, ''Functional self-similarity and renormalization group symmetry in mathematical physics,'' Theor. Math. Phys. **121**, 1315–1332 (1999).

[31]It is quite similar to the Compton wavelength of the pion which is the typical range of the nuclear force between hadrons like protons and nucleons.

[32]E. Raposo, S. de Oliveira, A. Nemirovsky, and M. Coutinho-Filho, ''Random walks: A pedestrian approach to polymers, critical phenomena and field theory,'' Am. J. Phys. **59**, 633–645 (1991).

[33]H. Stanley, ''Scaling, universality, and renormalization: Three pillars of modern critical phenomena,'' Rev. Mod. Phys. **71**, S358–S366 (1999).

[34]J. Tobochnik, ''Resource Letter CPPPT-1: Critical point phenomena and phase transitions,'' Am. J. Phys. **69**, 255–263 (2001).

[35]C. Bagnuls and C. Bervillier, ''Exact renormalization group equations: An introductory review,'' Phys. Rep. **348**, 91–157 (2001).

[36]More precisely, working with nonrenormalizable couplings would require us to fine-tune infinitely many of them at scale Λ^{-1} to unnatural values. Most of the time, such a finely-tuned model is no longer predictive.

[37]G. Jona-Lasinio, ''Renormalization group and probability theory,'' cond-mat/0009219.

[38]N. Tetradis and C. Wetterich, ''Critical exponents from the effective average action,'' Nucl. Phys. B [FS] **422**, 541–592 (1994).

[39]Nonperturbatively, the existence of the limit $\Lambda\to\infty$ is more subtle than perturbatively. The renormalization group flow must be controlled in this limit and this is achieved if nonperturbatively g_0 has a finite limit, that is, if there exists an ultraviolet fixed point of the RG flow. The case $g_0\to 0$ when $\Lambda\to\infty$, corresponds to asymptotically free theories, that is, in four space–time dimensions, to non-Abelian gauge theories.

[40]C. Schmidhuber, ''On water, steam, and string theory,'' Am. J. Phys. **65**, 1042–1050 (1997).