Séminaire Poincaré 2 (2001) 1 – 29

Séminaire Poincaré

An Introduction to Renormalization

Vincent Rivasseau
Laboratoire de Physique Théorique
Université Paris XI
F-91405 Orsay Cedex

Abstract. We review the theory of perturbative renormalization, discuss its limitations, and give a brief introduction to the powerful point of view of the renormalization group, which is necessary to go beyond perturbation theory and to define renormalization in a constructive way.

1 Introduction

The precise quantitative formulation of physical laws usually requires to introduce particular parameters or constants. It was early recognized that interaction with a particular medium or substrate can change the effective value of these constants. For instance Descartes laws for the refraction of light require a medium dependent index n and later Gauss's and Ampère's law introduced electric or magnetic permittivities whose values ϵ and μ in a non-empty medium such as water or glass reflect in a complex way the interaction of light with the atoms of this medium.

Even more simply, Alain Connes's favorite examples of an interaction that changes even the sign of a physical parameter is that of Archimedes: a body such as a ping-pong ball immersed in water acquires a negative effective mass. Although the mass of the ball m may be much smaller than the mass M of the same volume of water, the "effective mass" as experimentally measured from the upwards acceleration of the ball is however limited by friction so that the true "negative effective mass" of the ball measured experimentally is much smaller in modulus than $m - M$.

New effective constants for the often multiplicative laws of physics can be considered as new normalizations of these laws. This is probably the origin of the name "renormalization". But a crisis occurred when physicists of the XXth century realized that this change of constants due to interaction is apparently infinite in the case of quantum field theory. This is disturbing because quantum field theory, which combines quantum mechanics and special relativity, was at that time considered the ultimate framework for the fundamental experimental laws of nature at the microscopic level[1]. Its consistency is therefore a matter of principle, whose importance can hardly be overemphasized.

The way out of this great "renormalization crisis" is a long story which required the efforts of many theoretical and mathematical physicists over the second half of the XXth century. I shall roughly divide it into two main chapters.

First the structure of the infinities or "divergences" in physical quantum field theories such as electrodynamics was elucidated. A recursive process, due to Bogoliubov and followers, was found to hide these infinities into unobservable "bare" parameters that describe the fundamental laws of physics at experimentally inaccessible extremely short distances. Although technically very ingenious, this solution left many physicists and probably most mathematicians under the impression that a real difficulty had been just "pulled under the rug".

It would be unfortunate however to remain under this impression. Indeed the second chapter of the story, known under the curious and slightly inaccurate name of the "renormalization group" (RG), truly solved the difficulty. It was correctly recognized by Wilson and followers that in a quantum theory with many scales involved, the change of parameters from bare to renormalized values is a phenomenon too complex to be described in a single step. Just like the trajectory of a

[1] It is still today to a large extent, although string theory holds great promises for an even more fundamental theory that would encompass gravity and have a natural fundamental ultraviolet length scale, the Planck scale.

complicated dynamical system, it must be instead studied step by step through a local evolution rule.

The change of scale in the RG plays the role of time in dynamical systems. This analogy is deep. There is a natural arrow of time, related to the second principle of thermodynamics, and there is similarly a natural arrow for the RG evolution: microscopic laws are expected to determine macroscopic laws, not the converse. The RG erases unnecessary detailed short scale information or "irrelevant operators". Even cosmology made now everybody familiar with the idea that the passing of time and the change of scale in physics are intimately related.

Apart from these almost philosophical comments, the RG improved point of view lead also concretely to many applications in various domains, some of which are also reviewed here. What seems less known, still today, is that RG also solved in a better way the old problem of infinities in perturbation theory. In the RG, the infinitesimal or discrete evolution under change of scale is perfectly well defined and finite. The old infinities are recognized as artefacts, due to an incorrect interchange of limits. In fact in the non-Abelian gauge theories which are presently at the backbone of the Standard Model, infinities disappear completely. Even after integrating evolution over an infinite sequence of intermediary scales, the RG flow remains perfectly bounded. The bare coupling constant, the ultimate "rug" under which perturbative infinities where supposed to hide, is in fact zero, the most finite of all possible values!

It is this amazing story that I will try to summarize in this note. As a testimony to its central place in recent theoretical physics, let me simply recall the many Nobel prizes awarded for major works on renormalization or related subjects. In 1965, R. Feynman, J. Schwinger and S.-I. Tomonaga received the Nobel prize for their formulation of quantum electrodynamics, the first theory to require renormalization. S. Glashow, S. Weinberg and A. Salam received the 1979 prize for unifying electromagnetic and weak interactions, two renormalizable field theories. In 1999, G. 't Hooft and M. Veltman received the prize for achieving the proof of renormalizability of this electroweak theory and of non-Abelian gauge theories in general. In 1982 the Nobel prize was awarded to K. Wilson for his invention of the renormalization group and its application to critical phenomena. Finally, among other contributions, P.G. de Gennes received the prize in 1991 for applying RG results to polymer physics. Besides these Nobel-winning contributions there have been so many other important works on renormalization that it is truly impossible to give full justice to all of them. So let me apologize in advance and refer to books such as [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] for more complete references.

2 Perturbative (Euclidean) Quantum Field Theory

2.1 Functional Integral and the ϕ^4 Model

Quantum Field Theory is the second quantized formalism appropriate to treat in particular the collision experiments of particle physics, in which particle number is not conserved. Cross sections contain the physical information of the theory. They are the matrix elements of the diffusion matrix $\mathcal{S}$. Under a suitable asymptotic condition, there are "reduction formulae" which express the matrix elements of $\mathcal{S}$ in terms of the Green functions G_N (or time ordered vacuum expectation values) of the field ϕ, which is operator valued and acts on the Fock space:

$$G_N(z_1, ..., z_N) = < \psi_0, T[\phi(z_1), ..., \phi(z_N)]\psi_0 > \ . \tag{2.1}$$

where ψ_0 is the vacuum state and T is an operator, called T-product, that orders a product of operators such as $\phi(z_1), ..., \phi(z_N)$ according to increasing times.

Consider a Lagrangian field theory, and split the total Lagrangian as the sum of a free plus an interacting piece, $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{int}$. The Gell-Mann-Low formula expresses the Green functions as vacuum expectation values of a similar product of free fields with an $e^{i\mathcal{L}_{int}}$ insertion:

$$G_N(z_1, ..., z_N) = \frac{< \psi_0, T\left[\phi(z_1), ..., \phi(z_N) e^{i\int dx \mathcal{L}_{int}(\phi(x))}\right]\psi_0 >}{< \psi_0, T(e^{i\int dx \mathcal{L}_{int}(\phi(x))})\psi_0 >}. \tag{2.2}$$

In the functional integral formalism proposed by Feynman [11], the Gell-Mann-Low formula is itself replaced by a functional integral in terms of an (ill-defined) "integral over histories" which is formally the product of Lebesgue measures over all space time. It is interesting to notice that the integrand appearing in this formalism contains the full Lagrangian $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{int}$, not just the interacting one. The corresponding formula is the Feynman-Kac formula:

$$G_N(z_1, ..., z_N) == \frac{\int \prod_j \phi(z_j) e^{i \int \mathcal{L}(\phi(x)) dx} D\phi}{\int e^{i \int \mathcal{L}(\phi(x)) dx} D\phi}. \tag{2.3}$$

This functional integral has potentially many advantages. First the rules of Gaussian integration make perturbation theory very transparent as shown in the next subsection. The fact that the full Lagrangian appears in (2.3) is interesting when symmetries of the theory are present which are not separate symmetries of the free and interacting Lagrangians, as is the case for non-Abelian gauge theories. It is also well adapted to constrained quantization, and to the study of non-perturbative effects.

There is a deep analogy between the Feynman-Kac formula and the formula which expresses correlation functions in classical statistical mechanics. For instance, the correlation functions for a lattice Ising model are given by

$$\langle \prod_{i=1}^{n} \sigma_{x_i} \rangle = \frac{\sum_{\{\sigma_x = \pm 1\}} e^{-L(\sigma)} \prod_i \sigma_{x_i}}{\sum_{\{\sigma_x = \pm 1\}} e^{-L(\sigma)}}, \tag{2.4}$$

where x labels the discrete sites of the lattice, the sum is over configurations $\{\sigma_x = \pm 1\}$ which associate a "spin" with value +1 or -1 to each such site and $L(\sigma)$ contains usually nearest neighbor interactions and possibly a magnetic field h:

$$L(\sigma) = \sum_{x,y \text{ nearest neighbors}} J \sigma_x \sigma_y + \sum_x h \sigma_x. \tag{2.5}$$

By analytically continuing (2.3) to imaginary time, or Euclidean space, it is possible to complete the analogy with (2.4), hence to establish a firm contact with statistical mechanics [5, 6, 7]. This idea also allows to give much better meaning to the path integral, at least for a free bosonic field. Indeed the corresponding free Euclidean measure $Z^{-1} e^{-\int L_0(\phi(x)) dx} D\phi$, where Z is a normalization factor, can be defined easily as a Gaussian measure. This is simply because L_0 is a quadratic form of positive type[2].

The Green functions continued to Euclidean points are called the Schwinger functions of the model, and are given by the Euclidean Feynman-Kac formula:

$$S_N(z_1, ..., z_N) = Z^{-1} \int \prod_{j=1}^{N} \phi(z_j) e^{-\int \mathcal{L}_i(\phi(x)) dx} d\mu_0(\phi) \tag{2.6}$$

$$Z = \int e^{-\int \mathcal{L}_i(\phi(x)) dx} d\mu_0(\phi). \tag{2.7}$$

The simplest interacting field theory is the theory of a one component scalar bosonic field ϕ with quartic interaction $g\phi^4$ (ϕ^3 which is simpler is unstable). In $\mathbb{R}^d$ it is called the ϕ^4_d model. For $d = 2, 3$ the model is superrenormalizable and has been built by constructive field theory. For $d = 4$ it is renormalizable in perturbation theory. Although the model lacks asymptotic freedom and a non-perturbative version may therefore not exist, it remains a valuable tool for a pedagogical introduction to perturbative renormalization theory.

Formally the Schwinger functions of the ϕ^4_d are the moments of the measure:

[2] However the functional space that supports this measure is not in general a space of smooth functions, but rather of distributions. This was already true for functional integrals such as those of brownian motion, which are supported by continuous but not differentiable paths. Therefore "functional integrals" in quantum field theory should more appropriately be called "distributional integrals".

$$d\nu = \frac{1}{Z} e^{-(g/4!)\int \phi^4 - (m^2/2)\int \phi^2 - (a/2)\int (\partial_\mu \phi \partial^\mu \phi)} D\phi, \tag{2.8}$$

where

- g is the coupling constant, usually assumed positive or complex with positive real part;
- m is the mass; it fixes an energy scale for the theory;
- a is the wave function constant. We often assume it to be 1;
- Z is a normalization factor which makes (2.8) a probability measure;
- $D\phi$ is a formal product $\prod_{x \in \mathbb{R}^d} d\phi(x)$ of Lebesgue measures at every point of $\mathbb{R}^d$.

But such an infinite product of Lebesgue measures is mathematically ill-defined. So it is better to define first the Gaussian part of the measure

$$d\mu(\phi) = \frac{1}{Z_0} e^{-(m^2/2)\int \phi^2 - (a/2)\int (\partial_\mu \phi \partial^\mu \phi)} D\phi. \tag{2.9}$$

More precisely if we consider the translation invariant propagator $C(x,y) \equiv C(x-y)$ (with slight abuse of notation), whose Fourier transform is

$$C(p) = \frac{1}{(2\pi)^d} \frac{1}{p^2 + m^2}, \tag{2.10}$$

we can use Minlos theorem and the general theory of Gaussian processes to define $d\mu_C(\phi)$ as the centered Gaussian measure on the Schwartz space of tempered distributions $S'(\mathbb{R}^d)$ whose covariance is C. A Gaussian measure is uniquely defined by its moments, or the integral of polynomials of fields. Explicitly this integral is zero for a monomial of odd degree, and for $n = 2p$ even it is equal to

$$\int \phi(x_1)...\phi(x_n) d\mu_C(\phi) = \sum_{\gamma} \prod_{l \in \gamma} C(x_{i(l)}, x_{j(l)}), \tag{2.11}$$

where the sum runs over all the pairings γ of the $2p$ arguments into p disjoint pairs $l = (i(l), j(l))$.

Note that since for $d \geq 2$, $C(p)$ is not integrable, $C(x,y)$ must be understood as a distribution. It is therefore convenient to also introduce a regularized kernel, for instance

$$C_\kappa(p) = \frac{1}{(2\pi)^d} \frac{e^{-\kappa(p^2+m^2)}}{p^2 + m^2} \tag{2.12}$$

whose Fourier transform $C_\kappa(x,y)$ is now a smooth function and not a distribution. Such a regularization is called an ultraviolet cutoff, and we have (in the distribution sense) $\lim_{\kappa \to 0} C_\kappa(x,y) = C(x,y)$. Remark that due to the non zero m^2 mass term, the kernel $C_\kappa(x,y)$ decays exponentially at large $|x-y|$ with rate m, taht is for some constant K and $d > 2$ we have:

$$|C_\kappa|(x,y)| \leq K \kappa^{1-d/2} e^{-m|x-y|}. \tag{2.13}$$

It is a standard useful construction to build from the Schwinger functions another class of functions called the connected Schwinger functions (in statistical mechanics connected functions are called Ursell functions or cumulants). These connected Schwinger functions are given by:

$$C_N(z_1, ..., z_N) = \sum_{P_1 \cup ... \cup P_k = \{1,...,N\};\, P_i \cap P_j = 0} (-1)^{k+1} \prod_{i=1}^{k} S_{p_i}(z_{j_1}, ..., z_{j_{p_i}}), \tag{2.14}$$

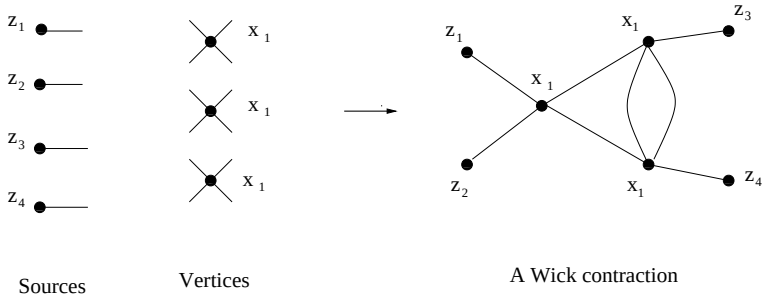

Figure 1: A contraction scheme

where the sum is performed over all distinct partitions of $\{1,...,N\}$ into k subsets $P_1,...,P_k$, P_i being made of p_i elements called $j_1,...,j_{p_i}$. For instance the connected 4-point function, when all odd Schwinger functions vanish due to the unbroken $\phi \to -\phi$ symmetry, is simply given by:

$$\begin{aligned} C_4(z_1,...,z_4) &= S_4(z_1,...,z_4) - S_2(z_1,z_2)S_2(z_3,z_4) \\ &\quad -S_2(z_1,z_3)S_2(z_2,z_4) - S_2(z_1,z_4)S_2(z_2,z_3). \end{aligned} \tag{2.15}$$

2.2 Feynman Rules

The full interacting measure may now be defined as the multiplication of the Gaussian measure $d\mu(\phi)$ by the interaction factor:

$$d\nu = \frac{1}{Z} e^{-(g/4!)\int \phi^4} d\mu(\phi) \tag{2.16}$$

and the Schwinger functions are the normalized moments of this measure:

$$S_N(z_1,...,z_N) = \int \phi(z_1)...\phi(z_N) d\nu(\phi). \tag{2.17}$$

This formula is especially convenient to derive the perturbative expansion and Feynman rules of the theory. Indeed, expanding the exponential as a power series in the coupling constant g, one obtains for the Schwinger functions:

$$S_N(z_1,...,z_N) = \frac{1}{Z} \sum_{n=0}^{\infty} \frac{(-g)^n}{n!} \int [\int \frac{\phi^4(x)}{4!}]^n \phi(z_1)...\phi(z_N) d\mu(\phi) \tag{2.18}$$

It is now possible to perform explicitly the functional integral of the corresponding polynomial. The result gives at any order n a sum over "Wick contractions schemes $\mathcal{W}$", i.e. ways of pairing together $4n+N$ fields into $2n+N/2$ pairs. There are exactly $(4n+N-1)(4n+N-3)...5.3.1 = (4n+N)!!$ such contraction schemes.

Formally at order n the result of perturbation theory is therefore simply the sum over all these schemes $\mathcal{W}$ of the spatial integrals over $x_1,...,x_n$ of the integrand $\prod_{l\in\mathcal{W}} C(x_{i(l)}, x_{j(l)})$ times the factor $\frac{1}{n!}(\frac{-g}{4!})^n$. These integrals are then functions (in fact distributions) of the external positions $z_1,...,z_N$ But they may diverge either because they are integrals over all of $\mathbb{R}^4$ (no volume cutoff) or because of the singularities in the propagator C at coinciding points.

It is convenient to label the n dummy integration variables in (2.18) as $x_1,...,x_n$ and to draw a line for each contraction of two fields. Each position $x_1,...,x_n$ is then associated to a four-legged vertex and each external source z_i to a one-legged vertex, as shown in Figure 1.

For practical computations, it is obviously more convenient to gather all the contractions which lead to the same topological structure, hence the same integral. This leads to the notion of Feynman graphs. To any such graph is associated a contribution or amplitude, which is the sum of

the contributions associated with the corresponding set of Wick contractions. The Feynman rules summarize how to compute this amplitude with its correct combinatoric factor.

We always use the following notations for a graph G:

- $n(G)$ or simply n is the number of internal vertices of G, or the order of the graph.
- $l(G)$ or l is the number of internal lines of G, i.e. lines hooked at both ends to an internal vertex of G.
- $N(G)$ or N is the number of external vertices of G; it corresponds to the order of the Schwinger function one is looking at. When $N = 0$ the graph is a vacuum graph, otherwise it is called an N-point graph.
- $c(G)$ or c is the number of connected components of G,
- $L(G)$ or L is the number of independent loops of G.

For a *regular* ϕ^4 graph, i.e. a graph which has no line hooked at both ends to external vertices, we have the relations:

$$l(G) = 2n(G) - N(G)/2, \tag{2.19}$$

$$L(G) = l(G) - n(G) + c(G) = n(G) + 1 - N(G)/2. \tag{2.20}$$

where in the last equality we assume connectedness of G, hence $c(G) = 1$. We like to define the superficial degree of divergence. For ϕ^4_d it is:

$$\omega(G) = dL(G) - 2l(G), \tag{2.21}$$

so that for a connected graph:

$$\omega(G) = (d-4)n(G) + d - \frac{d-2}{2}N(G). \tag{2.22}$$

It will be important also to define what we call a subgraph. This is not a completely straightforward notion. A *subgraph* F of a graph G is a subset of internal lines of G, together with the corresponding attached vertices. Hence there are exactly $2^{l(G)}$ subgraphs in G. We call the lines in the subset defining F the internal lines of F, and their number is simply $l(F)$, as before. Similarly all the vertices of G hooked to at least one of these internal lines of F are called the internal vertices of F and considered to be in F; their number by definition is $n(F)$. But remark that no external vertex of G can be of this kind. Precisely for this reason, the notion of external vertices does not generalize simply to subgraphs. Nevertheless for power counting we need at least to define a generalization of the number N for subgraphs. A good convention is to call external half-line of F every half-line of G which is not in F but which is hooked to a vertex of F; it is then the number of such external half-lines which we call $N(F)$. With this convention one has for ϕ^4 subgraphs the same relation (2.19) as for regular ϕ^4 graphs.

The definitions of c, L and ω then generalize to subgraphs in a straightforward way.

To compute the amplitude associated to a ϕ^4 graph, we have to add the contributions of the corresponding contraction schemes. This is summarized by the rules:

- To each line l_j with end vertices at positions x_j and y_j, associate a propagator $C(x_j, y_j)$.
- To each internal vertex, associate $(-g)/4!$.
- Count all the contraction schemes giving this diagram. The number should be of the form $(4!)^n n!/S(G)$ where $S(G)$ is an integer called the symmetry factor of the diagram. The 4! represents the permutation of the fields hooked to an internal vertex.
- Multiply all these factors, divide by $n!$ and sum over the position of all internal vertices.

The formula for the bare amplitude of a graph is therefore, as a distribution in $z_1, ...z_N$:

$$A_G(z_1, ..., z_N) \equiv \int \prod_{i=1}^{n} dx_i \prod_{l \in G} C(x_l, y_l). \tag{2.23}$$

This is the "direct" or "x-space" representation of a Feynman integral. As stated above, this integral suffers of possible divergences. But the corresponding quantities with both volume cutoff and ultraviolet cutoff κ are well defined. They are:

$$A^{\kappa}_{G,\Lambda}(z_1, ..., z_N) \equiv \int_{\Lambda^n} \prod_{i=1}^{n} dx_i \prod_{l \in G} C_{\kappa}(x_l, y_l). \tag{2.24}$$

The integrand is indeed bounded and the integration domain is a compact box Λ.

The *unnormalized* Schwinger functions are therefore formally given by the sum over all graphs with the right number of external lines of the corresponding Feynman amplitudes:

$$ZS_N = \sum_{\phi^4 \text{ graphs } G \text{ with } N(G)=N} \frac{(-g)^{n(G)}}{S(G)} A_G. \tag{2.25}$$

Z itself, the normalization, is given by the sum of all vacuum amplitudes:

$$Z = \sum_{\phi^4 \text{ graphs } G \text{ with } N(G)=0} \frac{(-g)^{n(G)}}{S(G)} A_G. \tag{2.26}$$

Let us remark that since the total number of Feynman graphs is $(4n+N)!!$, taking into account Stirling's formula and the symmetry factor $1/n!$ from the exponential we expect perturbation theory at large order to behave as $K^n n!$ for some constant K. Indeed at order n the amplitude of a Feynman graph is a 4n-dimensional integral. It is reasonable to expect that in average it should behave as c^n for some constant c. But this means that one should expect zero radius of convergence for the series (2.25). This is not too surprising. Even the one-dimensional integral

$$F(g) = \int_{-\infty}^{+\infty} e^{-x^2/2 - g x^4} dx \tag{2.27}$$

is well-defined only for $g \geq 0$. We cannot hope infinite dimensional functional integrals of the same kind to behave better than this one dimensional integral. In mathematically precise terms, F is not analytic near $g = 0$, but only Borel summable. A Borel summable function f can be entirely reconstructed from its asymptotic series $\sum_n a_n x^n$, but not by naively adding the terms in the series. One has rather to first define the Borel series

$$B(t) = \sum_n \frac{a_n}{n!} t^n \tag{2.28}$$

and to analytically continue this function B to a neighborhood of the real axis, then recover f through the integral formula

$$f(x) = \frac{1}{x} \int_0^{\infty} e^{-t/y} B(t) dt. \tag{2.29}$$

In the case of the function F, this process is guaranteed to converge (using the obvious analyticity of F for $\Re g > 0$, some uniform Taylor remainder estimates and Nevanlinna's theorem [12]). So we know the integral (2.29) can reconstruct F from the list of its asymptotic coefficients, which in that particular case are nothing but

$$a_n = \frac{(-1)^n}{n!} \int_{-\infty}^{+\infty} x^{4n} e^{-x^2/2} dx = \frac{(-1)^n 4n!!}{n!}. \tag{2.30}$$

In general Bosonic functional integrals require some stability condition for the potential at large field (here e.g. $g \geq 0$), and their perturbation series do not converge. Borel summability is

therefore the best we can hope for the ϕ^4 theory, and it has indeed been proved for the theory in dimensions 2 and 3 [13, 14].

From translation invariance, we do not expect $A^{\kappa}_{G,\Lambda}$ to have a limit as $\Lambda \to \infty$ if there are vacuum subgraphs in G. But we can remark that an amplitude factorizes as the product of the amplitudes of its connected components.

With simple combinatoric verification at the level of contraction schemes we can factorize the sum over all vacuum graphs in the expansion of unnormalized Schwinger functions, hence get for the normalized functions a formula analog to (2.25):

$$S_N = \sum_{\substack{\phi^4 \text{ graphs } G \text{ with } N(G)=N \\ G \text{ without any vacuum subgraph}}} \frac{(-g)^{n(G)}}{S(G)} A_G. \tag{2.31}$$

Now in (2.31) it is possible to pass to the thermodynamic limit (in the sense of formal power series) because using the exponential decrease of the propagator, each individual graph has a limit at fixed external arguments. There is of course no need to divide by the volume for that because each connected component in (2.31) is tied to at least one external source, and they provide the necessary breaking of translation invariance.

Finally one can determine the perturbative expansions for the connected Schwinger functions and the vertex functions. As expected the connected Schwinger functions are given by sums over connected amplitudes:

$$C_N = \sum_{\phi^4 \text{ connected graphs } G \text{ with } N(G)=N} \frac{(-g)^{n(G)}}{S(G)} A_G \tag{2.32}$$

and the vertex functions are the sums of the *amputated* amplitudes for proper graphs, also called one-particle-irreducible. They are the graphs which remain connected even after removal of any given internal line. The amputated amplitudes are defined in momentum space by omitting the Fourier transform of the propagators of the external lines. It is therefore convenient to write these amplitudes in the so-called momentum representation:

$$\Gamma_N(z_1, ..., z_N) = \sum_{\phi^4 \text{ proper graphs } G \text{ with } N(G)=N} \frac{(-g)^{n(G)}}{S(G)} A^T_G(z_1, ..., z_N), \tag{2.33}$$

$$A^T_G(z_1, ..., z_N) \equiv \frac{1}{(2\pi)^{dN/2}} \int dp_1 ... dp_N e^{i \sum p_i z_i} A_G(p_1, ..., p_N), \tag{2.34}$$

$$A_G(p_1, ..., p_N) = \int \prod_{l \text{ internal line of } G} \frac{d^d p_l}{p_l^2 + m^2} \prod_{v \in G} \delta(\sum_l \epsilon_{v,l} p_l). \tag{2.35}$$

Remark in (2.35) the δ functions which ensure momentum conservation at each internal vertex v; the sum inside is over both internal and external momenta; each internal line is oriented in an arbitrary way and each external line is oriented towards the inside of the graph. The incidence matrix $\epsilon(v, l)$ is 1 if the line l arrives at v, -1 if it starts from v and 0 otherwise. Remark also that there is an overall momentum conservation rule $\delta(p_1 + ... + p_N)$ hidden in (2.35). The drawback of the momentum representation lies in the necessity for practical computations to eliminate the δ functions by a "momentum routing" prescription, and there is no canonical choice for that.

2.3 Feynman representation

There are other convenient representations such as the "Feynman parametric representation" which do not need any non canonical choices. To define it we write the α or parametric representation of the propagator:

$$\hat{C}(p) = \frac{1}{(2\pi)^d} \int_0^\infty e^{-\alpha(p^2+m^2)} d\alpha, \tag{2.36}$$

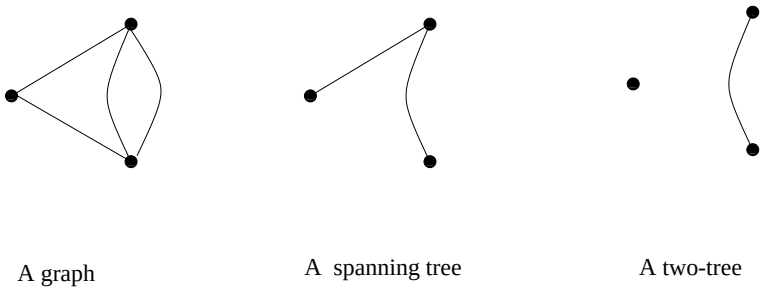

Figure 2: Spanning and two-trees

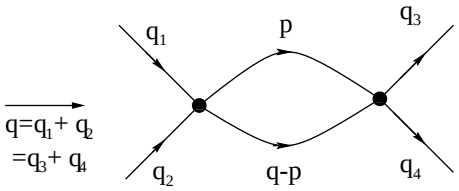

Figure 3: The graph G_0

$$\begin{aligned} C(x,y) &= \frac{1}{(2\pi)^d}\int_0^\infty d\alpha \int e^{ip.(x-y)-\alpha(p^2+m^2)}d^dp \\ &= \frac{1}{(4\pi)^{d/2}}\int_0^\infty \frac{d\alpha}{\alpha^{d/2}} e^{-\alpha m^2-|x-y|^2/(4\alpha)}. \end{aligned} \tag{2.37}$$

The x space or p space integrations can then be explicitly performed in any Feynman amplitude, since they are quadratic. The result is a compact formula with one scalar integration over a parameter α for each internal line of the graph:

$$A_G(p_1,...,p_N) = \delta(\sum_v P_v)\int_0^\infty \prod_l d\alpha_l \; e^{-\sum_l \alpha_l m^2 - V_G(\alpha,p)/U_G(\alpha)} \frac{1}{[U_G(\alpha)]^{d/2}} \tag{2.38}$$

where U_G and V_G are polynomials in α depending on the particular topology of the graph, called the Symanzik polynomials. Their explicit expression is:

$$U_G = \sum_S \prod_{l \text{ not in } S} \alpha_l, \tag{2.39}$$

$$V_G(p,\alpha) = (\sum_T \prod_{l \text{ not in } T} \alpha_l)(\sum_{a\in T_1} p_a)^2. \tag{2.40}$$

In (2.39) the sum runs over the spanning trees S of G. Such a spanning tree is a set of lines without loops connecting all the vertices of the graph. Similarly in (2.40), the sum runs over the two-trees T of G which separate G into two connected components, each containing a non empty set of external lines, one of which is T_1 (by overall momentum conservation, (2.40) does not change if T_1 is replaced by the set of external lines of the other connected component, which is the complementary of T_1) (see Figure 2 for an example).

In this elementary presentation we shall not reproduce the complete proof of these formulas (see [15] or [9]). They rely on a careful analysis of the quadratic form that one obtains in the exponential after rewriting all the propagators in α space. This quadratic form in turn can be deduced form the incidence matrix of the graph.

Remark that the parametric representation is not only "canonical" but also quite economical in large dimensions. In dimension 4, a four point subgraph of order n has $n-1$ loops hence the momentum integration is over a space of dimension $4n-4$; instead the parametric representation is over a space of dimension $l = 2(n-1)$, hence with only half as many scalar components to

integrate. For instance the integral of the graph G_0 of Figure 2 involves only one total external momentum q and can be written formally as

$$\begin{aligned} A_{G_0}(q) &= \int d^4p \frac{1}{(p^2+m^2)((p-q)^2+m^2)} \\ &= \int_0^\infty \int_0^\infty \frac{d\alpha_1 d\alpha_2}{(\alpha_1+\alpha_2)^2} e^{-(\alpha_1+\alpha_2)m^2 - \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}q^2}. \end{aligned} \tag{2.41}$$

However none of these two representations gives convergent integrals because of a divergence at large p or small α's. We return to the structure of these ultraviolet divergences in the next subsection.

The α-representation has also a fundamental interpretation in terms of Brownian motions [16]. In particular, the propagator (2.37) can be written as:

$$C(x,y) = \int_0^\infty d\alpha \exp(-m^2\alpha)\, P(x,y;\alpha) \tag{2.42}$$

where $P(x,y;\alpha) = (4\pi\alpha)^{-d/2} \exp(-|x-y|^2/4\alpha)$ is the Gaussian probability distribution of a Brownian path going from x to y in time α.

The Feynman diagrams can then be understood as made of Brownian paths interacting by Dirac distributions, as in the Edwards model for self-avoiding polymers [17]. This lead P.-G. de Gennes in 1972 to his famous relation between this polymer model and a $[(\phi)^2]^2$ field theory with $O(N)$ symmetry, in the $N \to 0$ limit [18]. This allowed RG results to be applied to polymer physics. A new development appeared when J. des Cloizeaux introduced a simple *direct* (dimensional) renormalization method for the Edwards model [19, 20], working explicitly in the α-representation.

2.4 Ultraviolet Divergences

The amputated amplitudes for a connected graph at finite external momenta are not always finite because of possible ultraviolet divergences. These divergences appear because the momentum integration over the loop variables in (2.35) may not always be absolutely convergent. This can be traced back to the distribution character of the propagator C in direct space, for $d \geq 2$, and the general impossibility to multiply distributions as should be done to define e.g. ϕ^4.

This difficulty, also present in quantum electrodynamics, was the basic puzzle that the founding fathers of quantum field theory were confronted with. Let us explore it, increasing the dimension step by step. The naive global scaling of all internal momenta of the graph explains our definition of the superficial degree of divergence: it measures whether the integral over this global scaling parameter is convergent or not. Therefore graphs with $\omega(G) \geq 0$ are called primitively divergent.

- If $d = 2$, we find $\omega(G) = 2 - 2n$, so the only divergent graphs have $n = 1$, and $N = 0$ or $N = 2$. The only divergence is due to the "tadpole" loop $\int \frac{d^2p}{(p^2+m^2)}$ which is logarithmically divergent.

- If $d = 3$, we find $\omega(G) = 3 - n - N/2$, so the only divergent graphs have $n \leq 3$, $N = 0$, or $n \leq 2$ and $N = 2$. Such a theory with a finite number of "primitively divergent" subgraphs is called superrenormalizable.

- If $d = 4$, $\omega(G) = 4 - N$. Every two point graph is quadratically divergent and every four point graph is logarithmically divergent. This is in agreement with the superficial degree of these graphs being respectively 2 and 0. For instance the graph G_0 at zero momentum without ultraviolet cutoff is logarithmically divergent for large p:

$$A_{G_0}(0) = \int \frac{d^4p}{(p^2+m^2)^2} = +\infty \tag{2.43}$$

and the "*tadpole*" loop $\int \frac{d^dp}{p^2+m^2}$ is quadratically divergent. Theories in which the degree of divergence only depends on the number of external legs are called renormalizable.

- Finally for $d > 4$ we have infinitely many primitively divergent graphs with arbitrarily large number of external legs, and the theory is called non-renormalizable.

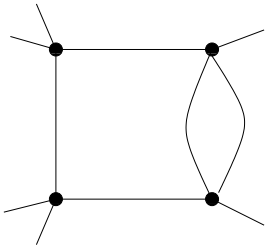

Figure 4: A 6-point subgraph with a divergent subgraph

It was soon recognized that even graphs which have negative superficial degree of divergence, such as the 6-point subgraph of Figure 4 in $d = 4$, are not ultraviolet finite. Indeed they can contain divergent subgraphs, and the corresponding subintegrations do not converge.

The first progress on renormalization came in recognizing that for four-dimensional theories such as ϕ^4 or quantum electrodynamics, the superficially divergent graphs when suitably added to a local counterterm gives rise to a finite contribution. For instance in the case of the graph G_0 the "renormalized" amplitude

$$\begin{aligned} A^R_{G_0}(q) &= \int [\frac{1}{(p^2+m^2)((p+q)^2+m^2)} - \frac{1}{(p^2+m^2)^2}]\, d^4p \\ &= \int_0^\infty \int_0^\infty \frac{d\alpha_1 d\alpha_2 e^{-(\alpha_1+\alpha_2)m^2}}{(\alpha_1+\alpha_2)^2} \big[e^{-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}q^2} - 1\big]. \end{aligned} \tag{2.44}$$

is now finite.

Indeed let us prove finiteness of this amplitude. In the momentum representation, we reduce to the same denominator, and taking advantage of parity we obtain:

$$\begin{aligned} A^R_{G_0}(q) &= \int \frac{-2p.q - q^2}{(p^2+m^2)^2((p+q)^2+m^2)}\, d^4p \\ &= -\int \frac{q^2}{(p^2+m^2)^2((p+q)^2+m^2)}\, d^4p \end{aligned} \tag{2.45}$$

now an obviously convergent integral. In the parametric representation, using $|e^{-x} - 1| \leq x$ for positive x we can bound $A^R_{G_0}(q)$ by

$$\int_0^\infty \int_0^\infty d\alpha_1 d\alpha_2 e^{-(\alpha_1+\alpha_2)m^2} \frac{q^2\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^3} \tag{2.46}$$

which is now a convergent integral. To be more precise, we should make additional remarks:

- the renormalized amplitude is negative
- it behaves as $c \log |q|$ as $|q| \to \infty$
- this large behavior at large q is solely due to the integral over the region $|p| \leq |q|$ of the *counterterm*. Indeed both

$$\int_{|p|\geq|q|} \frac{q^2}{(p^2+m^2)^2((p+q)^2+m^2)}\, d^4p \tag{2.47}$$

and

$$\int_{|p|\leq|q|} \frac{1}{(p^2+m^2)((p+q)^2+m^2)}\, d^4p \tag{2.48}$$

are well defined uniformly bounded integrals as $|q| \to \infty$.

Remark finally that the counterterm, when Fourier transformed, corresponds to a local ϕ^4 term, since the zero momentum value of the graph is nothing but the spatial integral over y of $C^2(x, y)$. This counterterm when added to the bare Lagrangian will renormalize G_0 not only as a

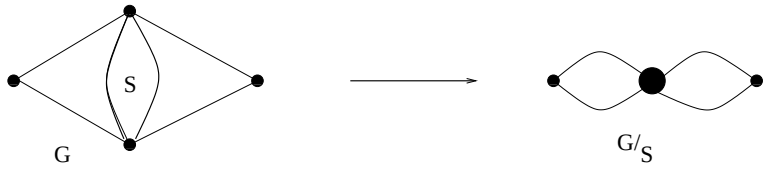

Figure 5: The reduction of a subgraph in a graph

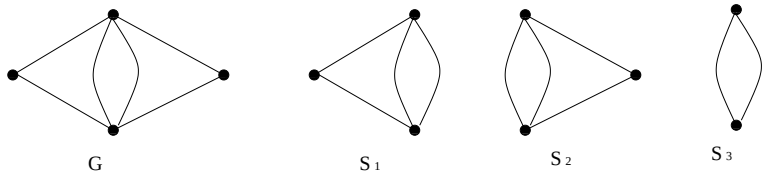

Figure 6: A graph with two overlapping divergent subgraphs

primitive graph, but each time it appears as a subgraph in the expansion, since the combinatoric of inserting a ϕ^4 vertex or a G_0 subgraph at any place in a bigger diagram is clearly the same.

In the same way local counterterms of the ϕ^4, ϕ^2 or $(\nabla\phi)^2$ type for any kind of primitively divergent graph, can be reabsorbed in the parameters of the Lagrangian of (2.8). Such an infinite redefinition which affects only the unobservable "bare" parameters of the theory hence it is not physically inconsistent.

But for a while it was not clear whether one could introduce a proper set of counterterms which is local and remove all the ultraviolet divergences of every graph, not only the main global primitive divergences but also all the divergences associated to subgraphs. This would make all particular submanifolds of the momentum integration convergent. The solution of this problem, by Bogoliubov, Parasiuk, Hepp and Zimmermann [21, 22, 23], and its extension to gauge theories by 'tHooft and Veltman [27] is a first great mathematical triumph of quantum field theory.

2.5 The Bogoliubov Recursion and Zimmermann's Solution

We have now to explain how to organize the set of all subtractions that should be performed in a renormalizable theory to make it ultraviolet finite in perturbation theory. When a local counterterm has been defined for a graph G_1 with N_1 external lines, the modified Lagrangian gives rise to a new vertex with N_1 lines. So for every graph G_2 which contains G_1 as a subgraph, to subtract the subintegration over G_1 corresponds to perform the sum

$$A_{G_1} + c_{G_1} A_{G_2/G_1} \tag{2.49}$$

where G_2/G_1 is the graph obtained by reducing G_1 to a single vertex in G_2 (see Figure 5 for an example). This reduction is an essential operation in renormalization theory. But remark already that if there are several divergent subgraphs in a graph G, we can define a reduced subgraph $G/\mathcal{S}$ only for families $\mathcal{S}$ of *disjoint* subgraphs S.

More generally, if G_2 itself is divergent, it seems clear that the counterterm for G_2 should be defined by taking the local part of (2.49), not of A_{G_1} itself. So the definition of counterterms is inductive, starting with the smaller graphs towards the bigger. This is after all the logic of perturbation theory. This induction was formalized by Bogoliubov. However since a graph G can contain overlapping divergent subgraphs S_1 and S_2 with non-trivial intersection S_3, such as in Figure 6, it is far from clear that this induction actually removes all ultraviolet divergences. The first proof that Bogoliubov's induction actually leads to finite amplitudes is due to Hepp [22], and the first explicit solution of the induction, which involves the notion of "forests" is due to Zimmermann [23].

Suppose we have defined counterterms up to a given order n. Then for a graph G at order

$n+1$ one defines a counterterm c_G and the renormalized amplitude A_G^R by

$$A_G^R = \sum_{\mathcal{S}} (A_{G/\mathcal{S}} \prod_{S \in \mathcal{S}} c_S) + c_G \tag{2.50}$$

where the sum is over all families $\mathcal{S}$ of disjoint primitively divergent subgraphs of G, including the empty one.

The exact definition of c_G contains some arbitrariness if the goal is to make renormalized amplitudes finite. In the BPHZ (Bogoliubov-Parasiuk-Hepp-Zimmermann) renormalization scheme, c_G is the local (or zero momentum part) of the sum in (2.50). More precisely, if we choose a system of loop momenta k for G and call p the external momenta we have

$$A_G(p) = \int dk I_G(p,k) \tag{2.51}$$

and for a primitively divergent graph one defines the counterterm c_G by a subtraction acting directly at the level of the integrand $I_G(p,k)$ in momentum space, to get

$$A_G^R(p) = \int dk (1 - \mathcal{T}^{d(G)}) I_G(p,k) \tag{2.52}$$

where $\mathcal{T}^{d(G)}$, the so-called Taylor "operator" selects the beginning of the Taylor expansion of $I_G(p,k)$ up to order $d(G)$ around the simple point $p=0$. This is in agreement with (2.44).

To generalize to graphs with divergent subgraphs one follows the Bogoliubov recursion. In fact renormalizing proper (i.e. connected one-particle-irreducible) subgraphs is enough, and the explicit solution of the Bogoliubov induction with this subtraction prescription is:

$$A_G^R = \int dk \mathcal{R} I_G(p,k) \tag{2.53}$$

$$\mathcal{R} = \sum_{\mathcal{F}} \prod_{S \in \mathcal{F}} (-\mathcal{T}^{d(S)}) \tag{2.54}$$

where the sum is over all forests of proper divergent subgraphs $S \subset G$, including the empty forest.

Definition 1 *A forest $\mathcal{F}$ is a subset of subgraphs such that for any pair S_1, S_2 of the forest, either $S_1 \subset S_2$ or $S_2 \subset S_1$ or S_1 and S_2 are disjoint.*

This definition ensures that the partial ordering by inclusion in a forest can indeed be pictured as a set of trees, hence the name "forest".

For example the graph G of Figure 5 which has 3 different divergent strict subgraphs, has 12 forests, namely

$$\{\emptyset\}, \{S_1\}, \{S_2\}, \{S_3\}, \{G\}, \{S_3, S_1\}, \{S_3, S_2\}, \{S_3, G\}, \{S_1, G\}, \{S_2, G\}, \{S_3, S_1, G\}, \{S_3, S_2, G\} \tag{2.55}$$

These 12 forests are shown in Figure 7 In formula (2.54) the product of the Taylor operators is taken following the partial ordering of the forest, that is from smaller to bigger graphs. Each Taylor operator selects the beginning of a Taylor expansion in the external momenta of a subgraph S, which can later become internal momenta for G. The definition of $\mathcal{R}$ may therefore depend on the choice of the momentum routing, hence of the loop momenta solving the δ functions in (2.35). This difficulty lead Zimmermann to define particular momentum routing called "admissible". For these choices, Zimmermann could then prove:

Theorem 2.1 *The integrals (2.53) do converge for any G and define amplitudes $A_G^R(p)$ which are tempered distributions when analytically continued to Minkowski space.*

The difficulty linked to momentum routing can be avoided completely by working instead in the parametric representation. It is indeed possible to define an $\mathcal{R}$ operator acting directly in the α-parametric space, equivalent to Zimmermann's operator, but bypassing completely the problem

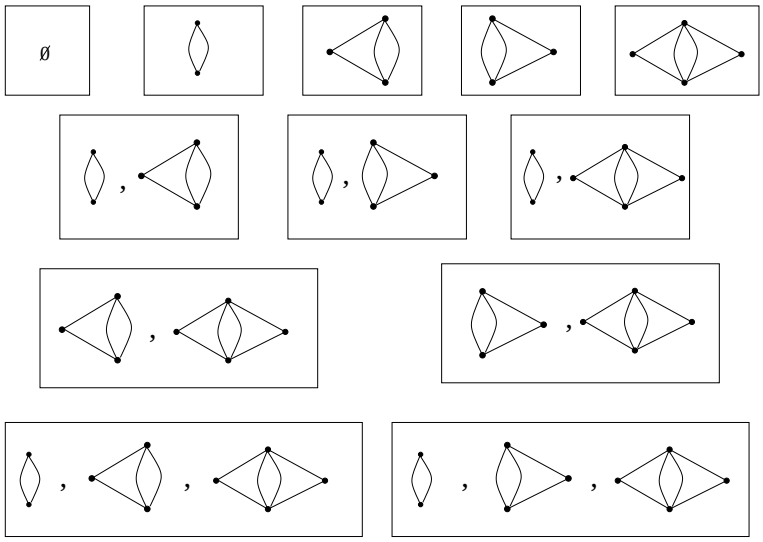

Figure 7: The twelve forests of G

of admissible momentum routing [24, 25]. Then there exists a very explicit proof of finiteness of the renormalized amplitudes. One can divide, for any complete ordering of the parameters α_l, also called a "Hepp sector", the sum over all forests quite naturally into packets, so that each packet gives a finite contribution. The problem is that the packets themselves change when the Hepp sector changes! Nevertheless this method is then sufficiently explicit to not only prove finiteness but also to produce reasonable quantitative estimates of the size of renormalized perturbation theory at large order [26].

The definition of the packets is subtle, but let us try to sketch it. The number of forests in any packet is always a certain power of 2, that is is of the form 2^r for a certain integer r. Indeed the forests which compose any such packet are exactly those containing a fixed forest $\mathcal{F}_0$ and contained in another fixed forest $\mathcal{F}_0 \cup \mathcal{F}_1$. r is simply the number of elements in $\mathcal{F}_1$. So the forests in that packet are those $\mathcal{F}$ that satisfy $\mathcal{F}_0 \subset \mathcal{F} \subset \mathcal{F}_0 \cup \mathcal{F}_1$. Hence the sum of the Taylor subtractions for a given packet always reconstructs an operator

$$\prod_{S \in \mathcal{F}_0} (-\mathcal{T}_S) \prod_{S \in \mathcal{F}_1} (1 - \mathcal{T}_S). \tag{2.56}$$

In a given sector, there is exactly one packet for each forest $\mathcal{F}_0$ with a certain property, which roughly speaking says that $\mathcal{F}_0$ is made of subgraphs with some internal line α-parameter larger than some external line α-parameter in the ordering of the sector considered. Given such an $\mathcal{F}_0$, the forest $\mathcal{F}_1$ then is completely determined by $\mathcal{F}_0$ and the sector. It is made of the subgraphs with the opposite property, that is all α-parameters for the internal lines of these subgraphs of $\mathcal{F}_1$ are smaller than all α-parameters for their external lines in the ordering of the sector [3].

The factorization property (2.56) is what makes each packet finite. Indeed the defining property for the subgraphs of $\mathcal{F}_0$ means that they are not really divergent in the sector considered. This is because the smaller α-parameter for one of their external lines acts as a natural ultraviolet cutoff for the subgraph. In contrast the subgraphs of $\mathcal{F}_1$ are potentially divergent. But for these subgraphs the $1 - \mathcal{T}_S$ operators in (2.56) precisely provide the necessary subtractions! This is the basic mechanism which makes every packet finite.

[3] The true definition is a bit more complicated and inductive, because reduction by the elements of $\mathcal{F}_0$ (as shown in Figure 5) has to be teken into account, starting from the smallest subgraphs in $\mathcal{F}_0$ and working towards the largest.

2.6 Different renormalization schemes

To subtract the value of subgraphs at zero external momenta is obviously a natural but not a canonical choice. It may be even ill-defined if the theory contains massless particles, which is for instance the case of quantum electrodynamics. It is important therefore to have several different sets of renormalization schemes, and to understand how they are related to each other. Two different subsets of counterterms which both make the Feynman amplitudes finite must differ through finite counterterms. In practice one wants usually to fix some physical conditions such as the particular values of some Green functions at some given momenta, and to determine the renormalization scheme corresponding these conditions. It may require two steps: first to use a general scheme to get rid of infinities, then to adjust the scheme through finite counterterms so as to meet the physical conditions.

For instance the BPHZ scheme that we have considered for the massive Euclidean ϕ_4^4 theory corresponds to the following normalization conditions on the connected functions in momentum space:

$$C^4(0,0,0,0) = -g, \tag{2.57}$$

$$C^2(p^2 = 0) = \frac{1}{m^2}, \tag{2.58}$$

$$\frac{d}{dp^2} C^2|_{p^2=0} = -\frac{a}{m^4}. \tag{2.59}$$

Let us say a few words about another popular renormalization scheme, namely dimensional renormalization. The starting idea is that in the parametric representation (2.38) the dimension d can be considered as a complex parameter. The attentive reader can object that external momenta still live in $\mathbb{R}^4$. But since the amplitudes depend only on the Euclidean scalar invariants $(\sum_{a \in T_1} p_a)^2$ built on them (see (2.40)), this is not a major difficulty. Amplitudes such as I_{G_0} in (2.41) become meromorphic functions for $\Re d \leq 4$. They have then a pole at $d = 4$. It is therefore natural to define the finite part of the amplitude as the finite part of the corresponding Laurent series, hence to simply extract the pure pole with its correct residue at $d = 4$. When properly implemented according to Bogoliubov's induction this leads to the notion of dimensional renormalization.

This scheme has many advantages but one major drawback. The advantage is that it preserves the symmetries of the theory such as gauge symmetries. Using it, 't Hooft and Veltman were able to show the renormalizability of the non-Abelian gauge theories at the core of the standard model [27]. For instance although the action $g^{-2} F_{\mu\nu} F^{\mu\nu}$ of a pure non-Abelian gauge theory contains terms of order 2, 3 and 4 in the field A_μ, it is possible with dimensional renormalization to preserve the basic relation between these three terms which make the total Lagrangian a perfect square. In this way the theory remains of the same form after renormalization , but simply with a renormalized parameter g_{ren} instead of g. This success was extremely important to convince physicists to adopt non-Abelian gauge theories for particle physics. As other examples of use of this scheme, let us mention again the renormalization method for the Edwards model of polymers [19, 20] which has been shown to be equivalent to standard (dimensional) field-theoretic renormalization [28]. These works opened the way to the renormalization theory of interacting or self-avoiding crumpled membranes, where the Feynman diagrams are no longer made of lines but of extended surfaces (see, e.g., [29]). Dimensional renormalization is also at the core of the Riemman-Hilbert interpretation of renormalization [30].

But the big drawback of dimensional renormalization is that up to now it remains a purely perturbative technique. Nobody knows how to interpolate correctly in the space-time dimension d the infinite dimensional functional integrals (2.17) which are the basis for the non-perturbative or constructive version of quantum field theory. To solve this difficulty would certainly be a major progress.

2.7 What lies beyond perturbative renormalization?

The theory of perturbative renormalization is a brilliant piece of mathematical physics. The solution of the difficult "overlapping" divergence problem through Bogoliubov's recursion and Zimmermann's forests becomes particularly clear in the parametric representation using Hepp's sectors:

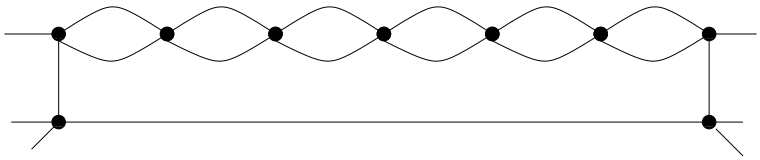

Figure 8: A family of graphs P_n producing a renormalon

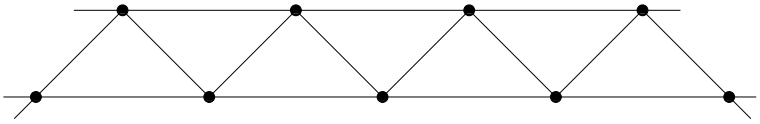

Figure 9: A family of convergent graphs Q_n, that do not produce any renormalon

in each sector there is a different classification of forests into packets so that each packet gives a finite integral.

Dimensional renormalization allows to preserve critical symmetries such as gauge symmetries, hence to prove renormalizability of four dimensional gauge theories, but does not seem adapted to non-perturbative theory. Note however that in this scheme the finite part of the Feynman amplitudes are related to ζ functions. This hints that this theory might be useful for mathematics, particularly number theory. The structure of the forests subtraction has been shown recently to be associated to a Hopf algebra and related to the Riemann-Hilbert problem in the works of Connes and Kreimer [31, 30].

But from the physical point of view we cannot conceal the fact that purely perturbative renormalization theory is also in some sense a conceptual maze. At least two facts already hint at a better theory which lies behind:

- The forest formula seems unnecessarily complicated, with too many terms. For instance if we examine closely the classification of forests into packets, we remark that in any given Hepp sector, only the particular packet corresponding to $\mathcal{F}_0 = \emptyset$ seems absolutely necessary to make the renormalized amplitude finite. The other packets, with non-empty $\mathcal{F}_0$ seem useless, a little bit like "junk DNA": they are there just because they are necessary for other sectors. This does not look optimal.

- The theory makes amplitudes finite, but at which cost! The size of some of these renormalized amplitudes becomes indeed unreasonably large as a size of the graph increases. This phenomenon is called the "renormalon problem". For instance it is easy to check that the renormalized amplitude (at 0 external momenta) of the graphs P_n with 6 external legs and $n+2$ internal vertices in Figure 8 becomes large as $c^n n!$ as $n \to \infty$. Indeed we remarked already that at large q the renormalized amplitude $A^R_{G_0}$ in (2.44) grows like $\log|q|$. Therefore the chain of n such graphs in Figure 8 behaves as $[\log|q|]^n$, and the total amplitude of P_n behaves as

$$\int [\log|q|]^n \frac{d^4q}{[q^2+m^2]^3} \simeq_{n\to\infty} c^n n! \tag{2.60}$$

So there are not only too many Feynman graphs to resum them, but some of them after renormalization also acquire so large values that the corresponding subfamilies of graphs cannot be resummed! These two hints are in fact linked. As their name indicates, renormalons are due to renormalization. Families of completely convergent graphs such as the graphs Q_n of Figure 9, are bounded by c^n, and produce no renormalons. But studying more carefully renormalization in the α parametric representation one can say more. One can check that renormalons are solely due to the forests packets with $\mathcal{F}_0 \neq \emptyset$ and in fact $\mathcal{F}_0$ large. A packet associated to a given $\mathcal{F}_0$ typically grows like $c^n|\mathcal{F}_0|!$ [26]. Recall that the forests $\mathcal{F}_0$ are made of those subgraphs which are not really divergent in the sector considered. So this renormalon analysis generalizes one of our previous remarks. Renormalons are due to subtractions that are not necessary to ensure convergence, just like the strange $\log|q|$ growth of $A^R_{G_0}$ at large q is solely due to the counterterm in the region where

this counterterm is not necessary to make the amplitude finite.

We can therefore conclude that subtractions are not organized in an optimal way by the Bogoliubov recursion. The idea of renormalization itself is not wrong. But to use the size of the graph as the relevant parameter to organize Bogoliubov's induction is not the optimal idea. A better parameter to organize the induction was found in fact for other completely different reasons by Wilson and followers. It is not the size of the graph but rather the size of the line momenta in it that should be used to better organize the renormalization subtractions. This is the point of view of the renormalization group.

3 The Renormalization Group

The renormalization group is a strange name. It is in fact an (ill-defined) semi-group. Its discovery came in two steps: first by thinking about changing the renormalization scheme, field theorists such as Callan and Symanzik discovered a kind of "invariance" of the theory [32, 33]. Two renormalizable theories with two different sets of coupling constants but defined by subtracting at different scales can in fact be the same physical theory if the constants and scales are related through some "renormalization group" equations. It is in fact even possible to prove finiteness of perturbative renormalization, hence to bypass the BPHZ theorem by directly using these renormalization group equations [34].

Then came the conceptual breakthrough of Wilson and followers [35]: instead of renormalizing the theory at once, why not perform this difficult task in a sequence of steps? The evolution of the theory in this sequence of steps is then similar to the evolution of a dynamical system. In dynamical systems we know that it is usually easier (in particular numerically) to perform patiently a large sequence of local steps than to try to guess the global result, or to search for an analytic solution, which is very rare. The same is true in renormalization theory, in which some scale parameter plays the role of time.

Although this was not the historic path, it would have been perfectly possible to arrive also at the same renormalization group concept by simply trying to simplify Zimmermann's formula to get rid of renormalons. Indeed this is exactly what the RG also does!

This note is too short for a complete review of the renormalization group and in particular of its non-perturbative aspects. So we will sketch what it does on the simple example of ϕ^4_4.

3.1 Slicing

One needs first to separate the degrees of freedom of the theory, and to organize them into a sequence of slices, each slice corresponding to a given scale. It is convenient to choose this sequence of scales to form a geometric progression. The idea is then to perform the functional integral only over the modes of the field corresponding to momenta of a given scale and to compute an effective theory for the remaining scales. This should not be done in an arbitrary order: according to the usual scientific philosophy, microscopic laws should determine macroscopic behavior, not the converse [4]. So the "effective" field theory should emerge progressively from the bare theory like an effective picture progressively emerges from averaging the fine pixels in a detailed picture, or like thermodynamics with a few macroscopic parameters such as temperature or pressure should emerge from a very complicated and chaotic microscopic behavior governed by the laws of mechanics.

In a theory such as ϕ^4_4, the mass fixes some particular scale beyond which no interesting physics happens because connected functions decay exponentially just as the propagator itself (2.13). So in this case the renormalization group will be used solely to treat the ultraviolet problem. One can slice the theory by dividing the Euclidean propagator into slices with an index $i \in \mathbb{N}$, and the slice i will correspond to momenta of order roughly M^i, where M is a fixed number, the ratio of the geometric progression (e.g. $M = 2$).

[4] This traditional philosophy is put in question by more holistic points of view such as those based on the dualities of string theory which exchange small and large distances. But in this note I will nevertheless stick to the old-fashioned point of view!

This can be done conveniently with the parametric representation, since α in this representation is roughly like $1/p^2$. So we can define the propagator within a slice as

$$C_i = \int_{M^{-2i}}^{M^{-2(i-1)}} e^{-m^2\alpha - \frac{|x-y|^2}{4\alpha}} \frac{d\alpha}{\alpha^{d/2}}. \tag{3.1}$$

We can intuitively imagine C_i as the piece of the field oscillating with Fourier components essentially only of size roughly M^i. In fact it is easy to prove the bound (for $d > 2$)

$$|C_i(x,y)| \leq K.M^{(d-2)i} e^{-M^i|x-y|} \tag{3.2}$$

where K is some constant.

For the first slice the formula is a little different because

$$C_0 = \int_1^{\infty} e^{-m^2\alpha - \frac{|x-y|^2}{4\alpha}} \frac{d\alpha}{\alpha^{d/2}}. \tag{3.3}$$

Now the full propagator with ultraviolet cutoff M^ρ, ρ being a large integer, may be viewed as a sum of slices:

$$C_{\leq \rho} = \sum_{i=0}^{\rho} C_i \tag{3.4}$$

Then the basic renormalization group step is made of two main operations:

- A functional integration
- The computation of a logarithm to define an effective action

Indeed decomposing a covariance in a Gaussian process corresponds to a decomposition of the field into independent random variables ϕ^i. Let us call

$$\Phi_i = \sum_{j=0}^{i} \phi_j. \tag{3.5}$$

This is the "low-momentum" field for all frequencies lower than i. The RG idea is that starting from scale ρ and performing $\rho - i$ steps, one arrives at an effective action for the remaining field ϕ_i. Then writing $\Phi_i = \phi_i + \Phi_{i-1}$ splits the field into a "fluctuation" field ϕ^i and a "background" field Φ_{i-1}. The first step, functional integration, is performed solely on the fluctuation field, so it computes

$$Z_{i-1}(\Phi_{i-1}) = \int d\mu_{C_i}(\phi_i) e^{-S_i(\phi_i + \Phi_{i-1})}. \tag{3.6}$$

Then the second step rewrites this quantity as the exponential of an effective action, hence simply computes

$$S_{i-1}(\Phi_{i-1}) = -\log[Z_{i-1}(\Phi_{i-1})] \tag{3.7}$$

Now $Z_{i-1} = e^{-S_{i-1}}$ and one can iterate! The flow from the initial bare action $S = S_\rho$ for the full field to an effective renormalized action S_0 for the last "slowly varying" component ϕ_0 of the field is similar to the flow of a dynamical system. Its evolution is decomposed into a sequence of discrete steps from S_i to S_{i-1}.

Of course this program needs many modifications to become a mathematically correct (non-perturbative) prescription. But at least formally it has a non-perturbative potential because it is not formulated at the level of graphs. Integrating over a single "momentum slice" of the field is like computing a field theory with both ultraviolet and infrared cutoff, and should be much easier than a full-fledged ultraviolet or infrared problem.

A key feature of the standard presentation of the renormalization group has been also omitted. Usually one performs a third somewhat confusing operation in a RG step, which is a rescaling of all the lengths of the theory and of the field size. Here it would simply be

$$x \to M^{-1}x, \tag{3.8}$$

$$\phi \to M^{-(d-2)/2}\phi. \tag{3.9}$$

But this rescaling is made to compare more easily the effective action to the former one, just like a "reframing" of our averaged picture to always fit into a frame of fixed size. It is therefore some kind of analogue of changing the reference frame in a dynamical system, from the "laboratory frame" to a "moving frame". We prefer here not to introduce this rescaling, because in many situations the long distance behavior of a theory is not governed by a simple scaling around the point $p = 0$ in momentum space but by more complicated extended singularities. This phenomenon occurs in condensed matter, where the singularity is given by a so called Fermi surface, and in diffusion problems in Minkowski space, where the propagator is singular on a mass-shell. In these cases there is no single simple moving frame (but rather one different moving frame for each limit point of the extended singularity).

Of course there is lot of arbitrariness in the choice of the slicing for the RG. One can use for instance wavelets [36]. A very popular choice is "block-spinning", in which Φ_{i-1} is simply the average of Φ_i over a cube of side size M^{-i}. Again this is a choice which does not generalize easily to extended singularities (and also breaks the rotation invariance of the theory) so (when possible) slicing the covariance of the field seems the best technical tool.

It is clear that the RG strategy is not limited to the study of an ultraviolet problem in field theory. In fact since the renormalization group flows from ultraviolet scales to infrared ones, it is particularly well adapted to the study of critical phenomena in statistical mechanics [35, 5]. The bare critical action leading to an effective massless action corresponds to an initial point at some finite given spatial scale in a RG trajectory, for which a final condition (massless effective theory) is given at very long distance. Similarly "the ultraviolet limit" in field theory corresponds to a sequence of bare actions at smaller and smaller spatial scale which end up on the same renormalized theory at some given fixed spatial scale. So the two problems are very similar. Finally a massless field theory without ultraviolet cutoff is similar to a dynamical system with two boundary conditions one towards $t \to -\infty$ and one towards $t \to +\infty$.

3.2 The Flow

In this section we would like to sketch how the renormalization group deeply changes the way perturbation theory should be organized.

Naive field theory was formulated with a single set of coupling constants, and perturbatively renormalized field theory is formulated with two such sets, the bare and the renormalized constants. The bare couplings become infinite formal power series in the renormalized constants with coefficients which diverge when the ultraviolet cutoff is removed. But the correlation functions when expressed as power series in the renormalized coupling constant have perfectly finite ultraviolet limits order by order. This limit is the sum of the renormalized Feynman amplitudes given by the forest formulas. But in addition to the usual divergence of perturbation theory due to the large number of diagrams this perturbative renormalization theory suffers from a new non-perturbative disease, the renormalons generated by the anomalously large amplitudes of some families of graphs such as those of Figure 8.

How does this change with RG? RG tells us that we should neither use one nor two sets of coupling constants, but an infinite set, one for each scale. All these "running constants" are uniquely related to any one of them because they must lie on a single RG trajectory.

Clearly the RG philosophy means that we should neither compute the correlation functions as series in the bare coupling with diverging coefficients in the ultraviolet limit nor as renormalon-ill series in the renormalized coupling. We should compute them as multi-series in the infinite set of running constants.

Once this big change is accepted, everything falls into place.

The momentum slicing becomes the fundamental tool. The Feynman amplitudes are sliced into "assignments" $\mu = \{i_l\}$ with a slice index i_l for each line. There is also a vertex index i_v for each vertex, namely the highest line index flowing into that vertex. It is a natural convention to consider the true external lines of the graph as having index below all others, for instance here index -1. Then the amplitude for a graph is no longer proportional to the power of a single coupling

but each vertex should be equipped with a running constant g_{i_v} corresponding to its scale in the assignment.

In this way we obtain the "effective expansion" for a given Schwinger function

$$S_N = \sum_{\substack{\phi^4 \text{ graphs } G \text{ with } N(G)=N \\ G \text{ without any vacuum subgraph}}} \sum_{\mu=\{i_l\}} \frac{1}{S(G)} [\prod_{v\in G} g_{i_v}] A^{R,eff}_{G,\mu}, \tag{3.10}$$

where the effectively renormalized amplitude $A^{R,eff}_{G,\mu}$ contains only one subtraction packet, the one associated to $\mathcal{F}_0 = \emptyset$. More precisely the graph G and the assignment μ uniquely define a single "divergent forest" $\mathcal{F}(G,\mu)$ which is made of those divergent subgraphs in G for which the indices of internal lines are all greater than the indices of external lines. Then (for instance in the parametric representation)

$$A^{R,eff}_{G,\mu} = \int d\alpha \ [\prod_{S\in\mathcal{F}(G,\mu)} (1-\mathcal{T}_S) \ ] \ I_{G\mu}(\alpha). \tag{3.11}$$

The Schwinger functions in this "effective expansion" are made of course of exactly the same pieces as the bare or the renormalized expansion. These pieces are simply reshuffled in a different way. Indeed in the effective expansion the subtractions associated to the additional packets responsible for all the complications of Zimmermann's formula have simply disappeared, exactly reabsorbed into the effective constants that equip the vertices. Since these packets were responsible for the renormalons, it is not surprising that the expansion (3.10) is free of the renormalon problem, as expressed by our next Lemma.

Remark that the subgraphs in $\mathcal{F}(G,\mu)$ are indeed exactly those divergent subgraphs which have short spatial scale compared to their external lines. Distances between internal vertices are then shorter than the typical oscillation lengths of the external legs. Since these legs are like sensors through which the subgraph communicates with the external world, subgraphs in $\mathcal{F}(G,\mu)$ look "quasi-point-like" when seen from the outside. It is therefore no surprise that subtracting a truly local counterterm for each such "quasi-local" subgraph, which is what $(1-\mathcal{T}_S)$ does, leaves only a small remainder free of renormalons. More precisely one can prove (putting all external momenta to 0 to simplify):

Lemma 3.1 *There exists a constant K such that for any G*

$$\sum_{\mu} |A^{R,eff}_{G,\mu}| \leq K^{n(G)} \tag{3.12}$$

One can conclude that although in the bare series the amplitudes were not subtracted at all, in the renormalized series they were subtracted too much because lots of useless forests gave rise to renormalons. By abandoning the idea of a single coupling constant, the effective expansion which lies between the bare and renormalized ones has exactly the right amount of subtractions, creating only small contributions.

Of course the attentive reader may object that the lemma has not too much meaning, because each piece $A^{R,eff}_{G,\mu}$ should be multiplied by a different factor $\prod_{v\in G} g_{i_v}$ before being summed over μ in the effective expansion. But let us suppose that all the running constants g_i remain bounded. In this case it is clear that the effective expansion is much better than the renormalized one from the point of view of resummation, since only the usual divergence linked to the large number of graphs remains. And bounded running constants are not uncommon: they occur in asymptotically free theories.

3.3 Asymptotic Freedom

In a just renormalizable theory like ϕ^4_4 the most interesting flow under the renormalization group is the one of the coupling constant. By a simple second order computation this flow is intimately

linked to the sign of the graph G_0 of Figure 3. More precisely, we find that at second order the relation between g_i and g_{i-1} is

$$g_{i-1} \simeq g_i - \beta g_i^2 \tag{3.13}$$

(remember the minus sign in the exponential of the action), where β is a constant, namely the asymptotic value of $\sum_{j,j'/\inf(j,j')=i} \int d^4y C_j(x,y) C_{j'}(x,y)$ when $i \to \infty$. Clearly this constant is positive. So for the normal stable ϕ_4^4 theory, the relation (3.13) inverts into

$$g_i \simeq g_{i-1} + \beta g_{i-1}^2, \tag{3.14}$$

so that fixing the renormalized coupling seems to lead to a large, diverging bare coupling, incompatible with perturbation theory. This is the famous "Landau ghost" problem.

But in non-Abelian gauge theories an extra minus sign is created by the algebra of the Lie brackets. This surprising discovery has deep consequences. The flow relation becomes approximately

$$g_i \simeq g_{i-1} - \beta g_i g_{i-1}, \tag{3.15}$$

with $\beta > 0$, or, dividing by $g_i g_{i-1}$,

$$1/g_i \simeq 1/g_{i-1} + \beta, \tag{3.16}$$

with solution $g_i \simeq \frac{g_0}{1+g_0\beta i}$. A more precise computation to third order in fact leads to

$$g_i \simeq \frac{g_0}{1 + g_0(\beta i + \gamma \log i + O(1))}. \tag{3.17}$$

Such a theory is called asymptotically free (in the ultraviolet limit) because the effective coupling tends to 0 with the cutoff for a finite fixed small renormalized coupling. Physically the interaction is turned off at small distances. This theory is in agreement with scattering experiments which see a collection of almost free particles (quarks and gluons) inside the hadrons at very high energy. This was the main initial argument to adopt quantum chromodynamics, a non-Abelian gauge theory with $SU(3)$ gauge group, as the theory of strong interactions.

Remark that in such asymptotically free theories the flow and all running constants remain bounded (in fact by the renormalized coupling). The initial expectations that infinite Feynman diagrams should lead to infinite bare parameters are clearly wrong in this case since this bare parameter in fact tends to 0 with the ultraviolet cutoff!

Asymptotic freedom is not limited to the rather complicated non-Abelian gauge theories. As is well known, fermion diagrams have an extra minus sign per loop. The Gross-Neveu theory, a theory with quartic coupling and N species of Fermions in two dimensions, has the same power counting as ϕ_4^4, and is also asymptotically free in the ultraviolet limit. This is also the case for instance for the ϕ_4^4 theory with "wrong sign" of the coupling constant, which can be studied at least in the planar limit, which tames the natural instability due to that wrong sign. The "right sign" ϕ_4^4 is not asymptotically free in the ultraviolet but as a consequence it is asymptotically free in the infrared, which means that the corresponding massless critical theory (with fixed ultraviolet cutoff) is almost Gaussian in the long distance limit [35].

3.4 Some Comments on Constructive Renormalization

Constructive field theory has for ambitious goal to define the non-perturbative mathematically correct version of Lagrangian quantum field theory. This may be considered somewhat an academic problem for weakly coupled theories such as quantum electrodynamics, for which perturbative computations up to three loops seem sufficient. But there are strongly coupled theories such as quantum chromodynamics in which a non-perturbative approach is badly needed. Also it would be quite surprising if the patient analysis of the mathematical difficulties related to the summation of quantum perturbation theory did not lead to important new physical insights. After all the difficulties in resumming classical perturbation theory were very important for the modern understanding of dynamical systems [8].

For reviews of constructive theory we refer to [2, 9, 37, 38]. But here let us sketch how the RG has to be modified to become truly a non-perturbative tool, and review briefly the achievements of the theory.

The first difficulty if we try to resum perturbation theory has to do with the large number of Feynman graphs. Convergence of the functional integral itself, and the divergence of perturbation theory can be considered as "large field" problems, because they are related to the fact that a bosonic field is an unbounded variable. Physically a large field corresponds to a large number of excitations or particles being produced, and large field problems are generic in bosonic theories because bosons, in contrast with fermions, can pile up in large numbers at the same place. In Fermionic theories the Pauli principle physically solves that problem: fermions cannot pile up at the same place. Mathematically the corresponding anticommuting functional integrals give rise to determinants. By Gram or Hadamard's inequalities an n by n determinant with elements bounded by 1 can never be of size $n!$ but at most $n^{n/2}$, so that fermionic perturbation theory converges, in sharp contrast with bosonic perturbation theory.

Clearly the RG as initially formulated by Wilson or summarized in (3.6)-(3.7) is not mathematically well-defined. In particular starting form any polynomial action it creates an effective action which is obviously no longer polynomial, and this even after a single step! Therefore the large field problem (integration on ϕ at large ϕ), appears! More precisely, even if the initial bare action is stable, i.e., bounded below, it is not clear that this remains true for $S_{eff}(\phi)$, even after a single RG step. Hence starting from a stable interaction, the second step of the RG may be already ill-defined. This point has to be stressed to physicists!

So constructive theory must modify carefully the two main operations in a RG step to make them well defined. The functional integral in a slice must be treated (in the bosonic case at least) with a tool called a *cluster expansion*. The idea of the cluster expansion is that since perturbation theory diverges we must keep most of it in the form of functional integrals. However one can test whether distant regions of space are joined or not by propagators. So one introduces a lattice of cubes of size comparable to the decay rate of the propagator (here M^{-i}) and one performs a battery of tests to know whether there are vertices or sources in different cubes joined by a propagator. This allows to rewrite the theory as a "polymer gas", the polymers being the sets of cubes joined together as the outcome of the cluster expansion. By construction this polymer gas has hardcore interactions: two connected components are always made of disjoint cubes. But when the coupling constant is small, the activities for the non-trivial polymers (containing more than one cube) are small. Hence the polymer gas is dilute and the statistical mechanics technique of the Mayer expansion, a tool which compares the hardcore gas to a perfect gas, allows to perform the thermodynamic limit. This Mayer expansion is the non-perturbative analog of the computation of the logarithm in the second part of a renormalization group step. In this way the renormalization group can be formulated correctly at the non-perturbative level, as a sequence of intertwinned cluster and Mayer expansions, and the flow of the critical parameters to renormalize, such as the mass, wave function and coupling constant can be computed in this framework.

Using this approach, models of non-trivial interacting field theories have been built over the past thirty years, which satisfy Osterwalder-Schrader's axioms, hence in turn have a continuation to Minkowski space that satisfies Wightman axioms [39, 40]. Such models are unfortunately yet restricted to space-time dimensions 2 or 3 but they include now both superrenormalizable models, such as $P(\phi)_2$ [42, 41, 43], ϕ^4_3 [44, 45, 46, 14] or the Yukawa model in 2 and 3 dimensions, as well as just renormalizable models such as the massive Gross-Neveu model in two dimensions [47, 48]. Most of these models have been built in the weak coupling regime, using expansions such as the cluster and Mayer expansions; the harder models require multiscale versions of these expansions, reshuffled according to the renormalization group philosophy.

In most cases the relationship of the non-perturbative construction to the perturbative one has been elucidated: the non-perturbative Green's functions being the Borel sum of the corresponding perturbative expansion [13, 14, 48]. In this sense one can say that constructive field theory has achieved the goal of resumming all Feynman graphs, although, as explained above, Borel resummation is not a naive ordinary summation but a clever reshuffling of the initial perturbative series.

Unfortunately constructing ϕ_4^4 itself, the initial goal of the constructive program has not been possible since it lacks ultraviolet asymptotic freedom. It has been possible to show numerically and through correlation inequalities that starting from a bare lattice action at short distance with some reasonable assumptions at short distance, the resulting theory is trivial i.e. not interacting [49, 50, 51].

But important partial results have been obtained for the construction of non-Abelian theories in 4 dimensions [52, 53]. New models not perturbatively renormalizable but asymptotically safe are also within reach of these techniques, such as the Gross-Neveu model in three dimensions [54]. In the infrared regime bosonic models of renormalizable power counting such as the critical (massless) ϕ_4^4 with an infrared cutoff [55, 56], or 4 dimensional weakly self-avoiding polymers have been controlled [57], and their asymptotics at large distance have been established. Nonperturbative mass generation has been established in the Gross-Neveu model in two dimensions and in the nonlinear σ model at large number of components with ultraviolet cutoff [58, 59]. Finally the RG when applied to condensed matter give rise to many rigorous results and programs, as sketched in the next section. Altogether this set of results strongly illustrate the power of functional integration in quantum field theory.

3.5 Extended singularities, the new RG frontier

During the last decade one of the main achievements in renormalization theory is the extension of the renormalization group of Wilson (which analyzes long-range behavior governed by simple scaling around the point singularity $p = 0$ in momentum space) to more general extended singularities [60, 61, 62]. This very natural and general idea is susceptible of many applications in various domains, including condensed matter and field theory in Minkowski space. In this section we will discuss the situation for interacting Fermions models such as those describing the conduction electrons in a metal.

The key features which differentiate electrons in condensed matter from Euclidean field theory, and makes the subject in a way mathematically richer, is that space-time rotation invariance is broken, and that particle density is finite. This finite density of particles creates the Fermi sea: particles fill states up to an energy level called the Fermi surface.

The field theory formalism is the best tool to isolate fundamental issues such as the existence of non-perturbative effects). In this formalism the usual Hamiltonian point of view with operators creating electrons or holes is replaced by anticommuting Fermion fields with two spin indices, and propagator

$$C_{ab}(k) = \delta_{ab} \frac{1}{ik_0 - [\epsilon(\vec{k}) - \mu]} \tag{3.18}$$

where $a, b \in \{1, 2\}$ are the spin indices. The momentum vector $\vec{k}$ has d spatial dimensions. and $\epsilon(\vec{k})$ is the energy for a single electron of momentum $\vec{k}$. The parameter μ corresponds to the chemical potential. The (spatial) Fermi surface is the manifold $\epsilon(\vec{k}) = \mu$ [5].

For a jellium isotropic model the energy function is invariant under spatial rotations

$$\epsilon(\vec{k}) = \frac{\vec{k}^2}{2m} \tag{3.19}$$

where m is some effective or "dressed" electron mass. In this case the Fermi surface is simply a sphere. This jellium isotropic model is realistic in the limit of weak electron densities, where the Fermi surface becomes approximately spherical. In general a propagator with a more complicated energy function $e(\vec{k})$ has to be considered. A very interesting case is the two dimensional Hubbard model corresponding to a square lattice. The momenta live on the dual "Brillouin zone" $[-\pi, \pi]^2$, and the energy function is

$$\epsilon(\vec{k}) = \cos k_1 + \cos k_2 \tag{3.20}$$

so that for $\mu = 0$ (the so-called half-filled model), the Fermi surface is a square.

[5] It may be convenient to add also an ultraviolet cut-off to this propagator to make its Fourier transformed kernel in position space well defined. Anyway, very high momenta should be suppressed in this non relativistic theory.

Imaginary (Euclidean) time (in the form of a circle, with antiperiodic boundary conditions for Fermions) corresponds to finite temperature T. When T tends to 0; the imaginary time circle grows to $\mathbb{R}$. At finite temperature, since Fermionic fields have to satisfy antiperiodic boundary conditions, the component k_0 in (3.18) can take only discrete values (called the Matsubara frequencies) :

$$k_0 = \pm \frac{2n+1}{\beta\hbar}\pi \tag{3.21}$$

so the integral over k_0 is really a discrete sum over n. For any n we have $k_0 \neq 0$, so that the denominator in $C(k)$ can never be 0. This is why the temperature provides a natural infrared cut-off. But when $T \to 0$, k_0 becomes a continuous variable and the propagator diverges on the "space-time" Fermi surface, defined by $k_0 = 0$ and $\epsilon(\vec{k}) = \mu$.

The interaction term is defined by:

$$S_\Lambda = \frac{g}{2}\int_\Lambda d^3x \ (\sum_a \bar{\psi}\psi)^2(x) \ . \tag{3.22}$$

Physically this interaction represents an effective interaction due to phonons or other effects. A more realistic interaction would not be completely local to include the short range nature of the phonon propagator, but we can consider the local action (3.22) as an idealization which captures all essential mathematical difficulties.

The basic new feature is that the singularity of the propagator is of codimension 2 in the $d+1$ dimensional space-time. This changes dramatically the power counting of the theory. Instead of changing with dimension, like in ordinary field theory, perturbative power counting is now independent of the dimension, and is the one of a just renormalizable theory. Indeed in a graph with 4 external legs, there are n vertices, $2n-2$ internal lines and $L = n-1$ independent loops. Each independent loop momentum gives rise to two transverse variables, for instance k_0 and $|\vec{k}|$ in the jellium case, and to $d-1$ inessential bounded angular variables. Hence the $2L = 2(n-1)$ dimensions of integration for the loop momenta exactly balance the $2n-2$ singularities of the internal propagators, as is the case in a just renormalizable theory.

In one spatial dimension, hence two space-time dimensions, the Fermi surface reduces to two points, and there is also no proper BCS theory since there is no continuous symmetry breaking in two dimensions (by the "Mermin-Wagner theorem"). Nevertheless the many Fermion system in one spatial dimension gives rise to an interesting non-trivial behavior, called the Luttinger liquid [60].

In two spatial dimensions or more, the key tool to correctly analyze the theory is a decomposition of the propagator analogous to (3.1), but both into discrete slices and in each slice into discrete angular sectors. The slices are defined by:

$$C = \sum_{j=1}^{\infty} C_j \quad ; \quad C_j(k) = \frac{f_j(k)}{ik_0 - e(\vec{k})} \tag{3.23}$$

where the slice function $f_j(k)$ effectively forces $|ik_0 - e(\vec{k})| \sim M^{-j}$, for some fixed parameter $M > 1$. These slices pinch more and more the Fermi surface as $j \to \infty$.

The slice propagator is further decomposed into sectors:

$$C_{(j)}(k) = \sum_{\sigma \in \Sigma_j} C_{j,\sigma}(k) \quad ; \quad C_{j,\sigma}(k) = \frac{f_{j,\sigma}(k)}{ik_0 - e(\vec{k})} \tag{3.24}$$

where Σ_j is a set of angular patches, called sectors, which cover the Fermi sphere. For instance if $d = 2$ we may simply cut the circle into M^j intervals of length $2\pi M^{-j}$, but a better idea is to make the patches as large as possible. What limits really the size of the patches is the curvature of the Fermi surface, so that the optimal number of such patches is really $M^{j/2}$ for the two dimensional jellium model [63], and only j^2 for the two dimensional Hubbard model at half-filling [64].

The RG applied to this problem means as before that higher slices give rise to local effects relatively to lower slices. Integrating the higher slices one obtains effective actions which govern

larger distance physics. These effective actions are however more complicated than in the field theory context. In rotation invariant models, renormalization of the two point function can be absorbed in a change of normalization of the Fermi radius. It removes all infinities from perturbation theory at generic momenta [61]. But the flow for the four point function is a flow for an infinite set of coupling constants describing the momentum zero channel of the Cooper pairs [62]. In the case of an attractive interaction, when the temperature is lowered to zero, this flow diverges at the BCS scale. At this scale the symmetry linked to particle number conservation is spontaneously broken, giving rise to superconductivity, that is to the condensation of Cooper pairs.

This condensation is a nonperturbative phenomenon, like quark confinement. But in contrast with quark confinement, we know in principle how to investigate in a mathematically rigorous way this BCS condensation. Indeed sectors around the Fermi surface play a role analogous to components of a vector field, so that an expansion in $1/N$, where N is the number of such components, could control the BCS regime [65], in which ordinary perturbation is no longer valid. We may call this situation a "dynamical $1/N$" effect. Nevertheless the full mathematical construction of the BCS transition starting from weakly interacting fermions remains a long and difficult program which requires to combine together several ingredients.

The discussion of high temperature superconductivity lead also to some controversy about the nature of interacting fermions systems in the ordinary non-superconducting phase. In particular, validity of the standard Fermi liquid theory (which is essentially defined by the propagator (3.18) up to small corrections) has been questioned in two dimensions. According to a mathematical criterion designed by M. Salmhofer [68], it is now possible to distinguish rigorously between the so-called Fermi liquid behavior and Luttinger liquid behavior above the usual critical BCS temperature. Using renormalization group around the Fermi surface it should be possible to soon complete the proof of the following theorem:

Theorem 3.2 *In two dimensions an interacting fermion system above the condensation temperature can be either a Fermi or a Luttinger liquid, depending on the shape of the Fermi surface. The jellium model with round Fermi surface is a (slightly anomalous) Fermi liquid [67], but the half-filled Hubbard model with a square Fermi surface should be a (slightly anomalous) Luttinger liquid [64].*

The mathematically rigorous construction of a two-dimensional interacting Fermi liquid at zero temperature, corresponding to non-parity invariant Fermi surfaces like those obtained by switching on a generic "magnetic field cutoff", has also been completed recently [69].

Like in the previous section the key to these constructive theorems lies in the resummation of perturbation theory in a single slice, and then in the iteration of renormalization group steps. Curiously, although power counting does not depend on the dimension, momentum conservation in terms of sectors in a fixed slice depends on it. This has dramatic constructive consequences. In $d = 2$ we have the "rhombus rule": four momenta of equal length which add to zero at a given vertex must be roughly two by two parallel. This means that two dimensional condensed matter in a slice is again directly analogous to an N-vector model in which angles on the Fermi surface play the role of colors [66]. This remark is at the core of all rigorous constructions of interacting Fermi liquids [67, 69].

In three dimensions, we expect interacting fermions to behave as regular Fermi liquid above the BCS temperature, but this turns out to be surprisingly difficult to prove non-perturbatively. Indeed there is no longer any analog of the "rhombus rule". Two different momenta at a vertex in a given slice no longer determine the third and fourth: there is an additional torsion angle, since four momenta of same length adding to 0 are not necessarily coplanar. More sophisticated techniques have been designed to deal with this case [70] but until now it is not clear that these techniques allow a full constructive analysis of the model up to the scale where the BCS symmetry breaking takes place.

3.6 Conclusion

If we consider the universal character of the action principle both at the classical and quantum level, and observe that the relation between microscopic and macroscopic laws is perhaps the

most central of all physical questions, it is probably not an exaggeration to conclude that the renormalization group is in some deep sense the "soul" of physics.

3.7 Acknowledgments

I thank B. Duplantier, C. Kopper and J. Magnen for their critical reading of the manuscript.

References

[1] M. Peskin and Daniel V. Schroeder (Contributor), *An Introduction to Quantum Field Theory*, Perseus Publishing, (1995).

[2] J. Glimm and A. Jaffe, *Quantum Physics. A functional integral point of view.* Mc Graw and Hill, New York, 1981.

[3] C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, McGraw Hill, 1980.

[4] P. Ramond, *Field Theory*, Addison-Wesley, 1994.

[5] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, Oxford University Press, 2002.

[6] C. Itzykson and J.M Drouffe, *Statistical Field Theory*, Volumes 1 and 2, Cambridge University Press 1991.

[7] Giorgio Parisi, *Statistical Field Theory*, Perseus Publishing 1998.

[8] G. Benfatto, G. Gallavotti, *Renormalization Group*, Princeton University Press, 1995.

[9] V. Rivasseau, *From Perturbative to Constructive Renormalization*, Princeton University Press, 1991.

[10] Manfred Salmhofer, *Renormalization: An Introduction*, Texts and Monographs in Physics, Springer Verlag, 1999.

[11] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals*, Mc Graw and Hill, New York 1965.

[12] A. Sokal, *An improvement of Watson's theorem on Borel summability*, Journ. Math. Phys. 21, 261 (1980).

[13] J.P. Eckmann, J. Magnen and R. Sénéor, *Decay properties and Borel summability for the Schwinger functions in $P(\phi)_2$ theories*, Comm. Math. Phys. **39**, 251 (1975).

[14] J. Magnen and R. Sénéor, *Phase space cell expansion and Borel summability for the Euclidean ϕ^4_3 theory*, Comm Math. Phys. **56**, 237 (1977).

[15] N. Nakanishi, *Graph Theory and Feynman integrals*, Gordon and Breach, New York 1971.

[16] K. Symanzik, in *Local Quantum Theory*, R. Jost Ed. (Varenna, 1968) Academic Press, New York (1969), 285.

[17] S. F. Edwards, Proc. phys. Soc. Lond. **85** (1965) 613.

[18] P.-G. de Gennes, Phys. Lett. **38 A** (1972) 339.

[19] J. des Cloizeaux, J. Phys. France **42** (1981) 635.

[20] J. des Cloizeaux and G. Jannink, *Polymers in Solution, their Modeling and Structure*, Clarendon Press, Oxford (1990).

[21] N. Bogoliubov and Parasiuk, Acta Math. 97, 227 (1957).

[22] K. Hepp, *Théorie de la renormalisation*, Berlin, Springer Verlag, 1969.

[23] W. Zimmermann, *Convergence of Bogoliubov's method for renormalization in momentum space, Comm. Math. Phys. 15, 208 (1969).*

[24] M. Bergère and Y.M.P. Lam, *Bogoliubov-Parasiuk theorem in the α-parametric representation*, Journ. Math. Phys. 17, 1546 (1976).

[25] M. Bergère and J.B. Zuber, *Renormalization of Feynman amplitudes and parametric integral representation*, Comm. Math. Phys. 35, 113 (1974).

[26] C. de Calan and V. Rivasseau, *Local existence of the Borel transform in Euclidean ϕ^4_4*, Comm. Math. Phys. 82, 69 (1981).

[27] G. 'tHooft and M. Veltman, Nucl. Phys. B50, 318 (1972).

[28] M. Benhamou and G. Mahoux, J. Phys. France **47** (1986) 559.

[29] F. David, B. Duplantier and E. Guitter, Nucl. Phys. **B394** (1993) 555-664.

[30] A. Connes and D. Kreimer, *Renormalization in Quantum Field Theory and the Riemann-Hilbert problem*, J. High Energy Phys. 09, 024 (1999).

[31] A. Connes and D. Kreimer, *Hopf algebras, renormalization and non commutative geometry*, Comm. Math. Phys. 199, 203 (1999).

[32] C. Callan, Phys. Rev D.2, 1541 (1970).

[33] K. Symanzik, *Small distance behaviour in field theory and power counting*, Comm. Math. Phys. 18, 227 (1970).

[34] J. Polchinski, Nucl. Phys. B231, 269 (1984).

[35] K. Wilson, *Renormalization group and critical phenomena, II Phase space cell analysis of critical behavior*, Phys. Rev. B4, 3184 (1974).

[36] G. Battle, *Wavelets and Renormalization*, World Scientific, 1999.

[37] Constructive Physics, Lecture Notes in Physics 446, Springer Verlag, 1995.

[38] V. Rivasseau, *Constructive Field Theory and Applications: Perspectives and Open Problems*, Journ. Math. Phys. 41, 3764 (2000).

[39] A. Wightman, *Quantum field theory in terms of vacuum expectation values*, Phys. Rev. **101**, 860 (1956).

[40] K. Osterwalder and R. Schrader, *Axioms for Euclidean Green's functions*, Comm. Math. Phys. **31**, 83 (1973).

[41] J. Glimm and A. Jaffe, *Quantum Physics. A functional integral point of view*, Mc Graw and Hill, New York, 1981.

[42] J. Glimm, A. Jaffe and T. Spencer, *The Wightman Axioms and Particle Structure in the $P(\phi)_2$ Quantum Field Model*, Ann. Math. **100**, 585 (1974).

[43] B. Simon, *The $P(\phi)_2$ (Euclidean) Quantum field theory*, Princeton University Press, 1974.

[44] J. Glimm and A. Jaffe, *Positivity of the ϕ^4_3 Hamiltonian*, Fortschr. Phys. **21**, 327 (1973).

[45] J. Feldman, *The $\lambda\phi^4_3$ field theory in a finite volume*, Comm. Math. Phys. **37**, 93 (1974).

[46] J. Feldman and K. Osterwalder, *The Wightman axioms and the mass gap for weakly coupled ϕ^4_3 quantum field theories*, Ann. Phys. **97**, 80 (1976).

[47] K. Gawedzki and A. Kupiainen, *Gross-Neveu model through convergent perturbation expansions*, Comm. Math. Phys. **102**, 1 (1985).

[48] J. Feldman, J. Magnen, V. Rivasseau and R. Sénéor, *A renormalizable field theory: the massive Gross-Neveu model in two dimensions*, Comm. Math. Phys. **103**, 67 (1986).

[49] M. Aizenman, *Geometric Analysis of ϕ^4 fields and Ising Models*, Comm. Math. Phys. 86, 1 (1982).

[50] J. Fröhlich. *On the triviality of $\lambda\phi^4_d$ theories and the approach to the critical point in $d \geq 4$ dimensions*, Nucl. Phys. B 200 FS4, 281 (1982).

[51] Roberto Fernandez, Jurg Fröhlich and Alan D. Sokal, *Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory*, Springer Verlag, 1992.

[52] T. Balaban, Comm. Math. Phys. 109, 249 (1987), 119, 243 (1988), 122, 175 (1989), 122, 355 (1989) and references therein.

[53] J. Feldman, J. Magnen, V. Rivasseau and R. Sénéor, *Construction of YM_4 with an infrared cutoff*, Commun. Math. Phys. 155, 325 (1993).

[54] P. Faria da Veiga, *Construction de modèles non-perturbativement renormalisables en théorie quantique des champs*, Thèse, Université de Paris XI, 1992.

[55] K. Gawedzki and A. Kupiainen, *Massless ϕ^4_4 theory: Rigorous control of a renormalizable asymptotically free model*, Comm. Math. Phys. **99**, 197 (1985).

[56] J. Feldman, J. Magnen, V. Rivasseau and R. Sénéor, *Construction of infrared ϕ^4_4 by a phase space expansion*, Comm. Math. Phys. **109**, 437 (1987).

[57] D. Iagolnitzer and J. Magnen, *Polymers in a weak random potential in dimension 4: Rigorous Renormalization Group Analysis*, Commun. Math. Phys. **162**, 85 (1994).

[58] C. Kopper, J. Magnen and V. Rivasseau, *Mass Generation in the Large N Gross-Neveu Model*, Comm. Math. Phys. **169** 121 (1995).

[59] C. Kopper,*Mass Generation in the Large N Non-linear σ Model*, to appear in Comm. Math. Phys. (1999).

[60] G. Benfatto and G. Gallavotti *Perturbation theory of the Fermi surface in a quantum liquid. A general quasi-particle formalism and one dimensional systems*, Journ. Stat. Physics **59**, 541, 1990.

[61] J. Feldman and E. Trubowitz, *Perturbation Theory for Many Fermions Systems*, Helv. Phys. Acta **63**, 156 (1990).

[62] J. Feldman and E. Trubowitz, *The Flow of an Electron-Phonon System to the Superconducting State*, Helv. Phys. Acta **64**, 213 (1991).

[63] J. Feldman, J. Magnen, V. Rivasseau and E. Trubowitz, *An Infinite Volume Expansion for Many Fermion Green's functions*, Helv. Phys. Acta, Vol. 65, 679 (1992).

[64] V. Rivasseau, *The two dimensional Hubbard Model at half-filling: I.Convergent Contributions*, Journ. Stat. Phys. 106, 693 (2002).

[65] J. Feldman, J. Magnen, V. Rivasseau and E. Trubowitz, *An Intrinsic 1/N Expansion for Many Fermion Systems*, Europhys. Letters **24**, 437 (1993).

[66] J. Feldman, J. Magnen, V. Rivasseau and E. Trubowitz, *Two dimensional Many Fermion Systems as Vector Models*, Europhys. Letters 24, 521 (1993).

[67] M. Disertori, V. Rivasseau, *A Rigorous Proof of Fermi Liquid Behavior for Jellium Two-Dimensional Interacting Fermions*, Phys. Rev. Lett. 85, 361 (2000).

[68] M. Salmhofer, *Continuous renormalization for fermions and Fermi liquid theory,* Comm. Math. Phys. **194**, 249 (1998).

[69] J. Feldman, H. Knörrer and E. Trubowitz, *A Two Dimensional Fermi Liquid*, an impressive series of ten papers, all accessible at http://www.math.ubc.ca/ feldman/fl.html.

[70] M. Disertori, J. Magnen and V. Rivasseau, *Interacting Fermi liquid in three dimensions at finite temperature: Part I: Convergent Contributions*, Annales Henri Poincaré Vol. 2, 733-806 (2001).