Séminaire Poincaré 2 (2002) 31-54 **Séminaire Poincaré**

Exact Renormalization Group

Giovanni Gallavotti
Fisica
Università di Roma "La Sapienza"
P.le Moro 2
00185 Roma, Italia

Abstract. Renormalization group methods are illustrated via examples

Introduction

In theoretical physics, problems in which singular quantities appear in the basic equations, or problems in which the basic equations are apparently free of singularities but their solutions are nevertheless singular are among the most interesting. For instance

(a) The Coulomb potential is singular at contact (*i.e.* at zero distance: an "ultraviolet singularity")
(b) The electric potential is singular also at infinite distance because it decays too slowly to zero (an "infrared singularity").
(c) The gravitational potential is as singular.
(d) In an incompressible Euler fluid the singularity manifests itself because any nontrivial motion generates fluid velocity fields that involve phenomena observable on length and time scales as small as wished (again an ultraviolet singularity).
(e) In relativistic quantum field theory the relativistic covariance is implemented through a requirement of a "local interaction", which implies that physical quantities show interesting phenomena on all short length and time scales, again an ultraviolet singularity.
(f) In statistical mechanics even short range interactions generate phenomena that involve many long length scales (critical phenomena: an infrared singularity as it concerns large space scales).
(g) In mechanics systems able to oscillate with few frequencies generate, when interacting, motions in which all harmonics of the basic frequencies are present and in which some may become so important to change completely the behavior in comparison with the unperturbed one.

The list could continue for a while (to include Fermi liquids and superconductivity, Bose condensation and superfluids, for instance). Similar questions appear also in other fields of Science; an egregious example is the mathematical theory of Fourier transforms: a function can be reconstructed from its Fourier transform via a convolution with a singular kenel $K(x-y)$ (typically the Dirichlet kernel) which is singular at $x=y$. Therefore the convolution probes the behavior of the function on all scales and the question of the (pointwise) convergence of the Fourier series becomes an analysis of an ultraviolet singularity.

The reason for the success of the "renormalization group" is that it is a method that attempts to study problems of the above kinds from a *unified viewpoint*. It is remarkable that sometimes the attempts really solve the problems or, when the problem reamins open, at least provide new insights into it.

The method can be loosely defined as follows: one finds an explicit solution of the problem which, of course, involves quantities that one cannot really compute. A typical case is when the problem admits a perturbative solution, although this is not always the case: in any event the formal solution involves the analysis of singular integrals, *i.e.* of integrals involving functions with singularities.

The singular functions, we call them generically $C(x)$, are then expressed as sums of many

very regular nonsingular functions each of which "lives on a fixed scale', which means that the regular "piece" of the singular function that lives on a scale l_0 is a smooth function of the form $l_0^\alpha \overline{C}(x\, l_0^{-1})$ where $\overline{C}$ is a smooth function rapidly decaying at infinity and α is a constant. The renormalization method is effective when the regular pieces into which $C(x)$ is decomposed differ only by the scale on which they live, possibly apart from a finite number of them. One says that in such cases the singlularity is "scale invariant" and the singularity is a power law behavior in $|x|$, and the exponent α is related to the value of the power. For instance this is for the Coulomb potential which can be written ("resolution of the *ultraviolet singularity*")

$$\frac{1}{|x|} = \frac{(1-e^{-|x|^2})}{|x|} + \sum_{k=1}^{\infty} \frac{(e^{-(\gamma^h|x|)^2} - e^{-(\gamma^{h+1}|x|)^2})}{|x|} = \frac{(1-e^{-|x|^2})}{|x|} + \sum_{k=1}^{\infty} \gamma^h \overline{C}(\gamma^h x)$$

where $\overline{C}(x) = \frac{e^{-|x|^2} - e^{-(\gamma|x|)^2}}{|x|}$; here $l_0 = 1$ and $\gamma > 1$ can be (arbitrarily) taken to be $\gamma = 2$.[1] Then one tries to show that the formal solution can be studied by breaking it into a sum of terms that have all the same structure apart form a change of scale: by summing infinitely many functions which are regular on infinitely many scales one can construct (hopefully in a controllable way) the singularities and the properties of the quantities that are formally expressed in terms of the original singular (often *a priori* even possibly meaningless) integrals.

Breaking a singular expression into many (infinitely many) parts allows to disentangle and to exhibit delicate cancellation phenomena which may allow us to give a meaning too expressions that seem meaningless at first. Sometimes the cancellations may be so effective that the apparent singularities are in fact not there.

A typical example is the KAM theory where in the end no singularities are really there. Another archetypal example of success of the approach is the quantum theory of fields in dimensions $d = 2, 3$.

Here I select the latter two examples as their similarity is striking. Both are problems that arise in systems that are perturbations of simple systems (integrable systems in the first case and free fields in the second). However the singularities do not allow us to proceed straightforwardly.

The interest of the problems and their physical relevance is well known. For instance the first arose historically from the famous remark by Poincaré that the pertubation analysis, used in astronomy since Laplace and crowned by the well known successes of the theory of precessions, compilation of ephemeral tables, discovery of asteroids (Ceres) or major planets (Neptune) ..., could not be an approximation in the naive sense of the term because strictly speaking the series used could not possibly be convergent. It surfaced again in Fermi's juvenile work on the equipartition problem and again in his last work (the "incompiuto" and postumously published Fermi,Pasta Ulam experiment) on the same matter; and many still recall the frustration felt in trying to understand seemingly simple problems such as the computation of the error in the small oscillations thoery or in motions that are small perturbations of simple integrable ones: like two point masses on a circle interacting via a small mutual potential and subject to an external small potential, *i.e.* a system described by two angles $\underline{\alpha} = (\alpha_1, \alpha_2) \in T^2 \equiv [0, 2\pi]^2$ and a total potential function $\varepsilon f(\underline{\alpha})$ with equation of motion

$$\ddot{\underline{\alpha}} = -\varepsilon\, \underline{\partial}_{\underline{\alpha}} f(\underline{\alpha})$$

The motions of assigned angular velocities $\underline{\omega}_0 = (\omega_1, \omega_2)$ of such a system exist and remain similar to the corresponding free motions if the equation (3.1) below has a solution, as an elementary check would allows us to see. An existence proof is, however, rich of conceptual difficulties.

The second problem addresses the basic question of the very possibility of existence of a quantum theory of interacting particles which is, at the same time, relativistically covariant.

[1] The reason for this (almost universal) choice seems to be, according to G. Parisi, that it is the only choice which does not generate the question "why is γ chosen $= 2$?": the choice $\gamma = \pi$ would be equally good but it would inevitaly raise uninteresting questions.

The approach allows us to solve completely the first problem (Sec. 3-6) at least for a large class of motions (the "non resonant" ones). For the second problem in the cases of space-time models of dimension $d = 2, 3, 4$ even the existence of a formal perturbative solution is not clear: but it can be established by the renormalization group method (Sec. 8,9). In the cases $d = 2, 3$ the renormalization approach becomes the basis for completing the solution of the problem (*i.e.* to go beyond the formal level) and thus it is a fundamental building block of the proof that, *at least in dimensions $d = 2, 3$, quantum fields and special relativity are compatible even in presence of nontrivial interactions.*[2]

We choose the above two problems because of the elegance of their solution and of their pedagogical value: however they were not originally solved by the method discussed here. There are a number of other problems which have been first solved via a renormalization group method of the type we considered in the present review: the critical point of various classes of statistical mechanics models or the theory of the ground state of one dimensional spinless Fermi systems, the theory of the convergence of Fourier series, dipole gases, Anderson localization to mention a few.

Much larger is the set of problems that have been studied only heuristically in the physics literature: a permanent challenge is to understand them fully.

We shall introduce the KAM problem and the field theory renormalization in $d = 2, 3$ for scalar fields (typical multiscale problems) by first discussing their single scale counterparts (Sec. 1,2,7): this should induce appreciation of the power of a method to reduce a multiscale problem to a single scale one.

Sometimes the problems that are studied at a heuristic level involve drastic and uncontrolled approximations: therefore many physicists consider important to gain some control on what one would like to neglect. For this reason the applications of the renormalization group in which the results are obtained without concessions to uncontrolled approximations are called "*exact renormalization group*" results while the others do not receive the qualification of "exact" even though they are considered "better" than the results of perturbation theory (when possible) which in the Physics literature seems to be regarded with undeserved contempt. They are called "*non perturbative*": a name well deserved because they usually are (considered) reliable and are certainly remarkably different from predictions obtained by naively truncating perturbation series. The latter fact is in itself a really non trivial achievement as those working on the subject before the work of Wilson, Fisher, Kadanoff, Jona-Di Castro immediately realized.

Consistently I try here to keep the exposition essential but complete and self contained; certain really technical details are in Sec.6 and in the appendices. Commented references to the (immediately relevant) literature can be found in the final pages.

1. Nonsingular perturbation theory

Examples of perturbation analysis abound: the simplest are the "single scale" problems. These are problems in which no "singularities" appear and, as a consequence, the perturbation expansions converge, or are at least asymptotic, for small pertubations. An example is the following implicit functions equations

$$\underline{h}(\underline{\psi}) = \varepsilon \frac{\partial f}{\partial \underline{\alpha}}(\underline{\psi} + \underline{h}(\underline{\psi})) \tag{1.1}$$

where $\underline{\psi} \in T^\ell$ is a point on the ℓ–dimensional torus $T^\ell = [0, 2\pi]^\ell$ and $f(\underline{\alpha})$ is a trigonometric polynomial $f(\underline{\alpha}) = \sum_{|\underline{\nu}| \le N} e^{i\underline{\nu}\cdot\underline{\alpha}} f_{\underline{\nu}}$ of degree N, $|\underline{\nu}| \stackrel{def}{=} \sum |\nu_j|$. The problem posed is to show the existence of a solution $\underline{h}$ analytic for $|\varepsilon|$ small and in $\underline{\psi}$. It is not the simplest of its kind but it is general enough to be useful also in the case of harder problems.[3]

[2] In dimension $d = 4$ the problem is still open, although via the renormalization group method one can show the existence of a well defined perturbative solution.

[3] The simplest equation of the kind would be $h = \varepsilon f(h)$ with $h \in R$: *i.e.* a "zero" dimensional version of (1.1), which could be studied by the same methods that we discuss below but which is too simple for our purposes.

A second example of the same type is the functional integral

$$E_\lambda(\varepsilon f) = \frac{\int P(d\varphi) e^{-\int_\Lambda d^d \underline{x}\,(\lambda\varphi^4_{\underline{x}} + \mu\varphi^2_{\underline{x}} + \varepsilon\varphi_{\underline{x}} f(\underline{x}))}}{\int P(d\varphi) e^{-\int_\Lambda d^d \underline{x}\,(\lambda\varphi^4_{\underline{x}} + \mu\varphi^2_{\underline{x}})}} \tag{1.2}$$

where $f(\underline{x})$ is a generic smooth test function, P is the Gaussian probability distribution on R^d, $d = 2, 3$, with covariance

$$C(\underline{x} - \underline{y}) \stackrel{def}{=} \langle \varphi_{\underline{x}} \varphi_{\underline{y}} \rangle_P = \frac{1}{(2\pi)^d} \int e^{i\,\underline{p}\cdot(\underline{x}-\underline{y})} \frac{1}{(\underline{p}^2 + 1)^2}\, d^d \underline{p} \tag{1.3}$$

and Λ is a finite cubic box.

The problem is to show that E_λ is a smooth function of $\lambda, \mu, \varepsilon$ for *all* test functions f and for $\lambda \geq 0, \mu, \varepsilon$ small enough. The functional derivatives with respect to $\varepsilon f(\underline{x})$ of $\log E_\lambda(\varepsilon f)$ are called the "*Schwinger functions*" of the functional integral in (1.2).

2. Tree expansions. Cancellations

Here we illustrate a technique to study the above problems. The technique is called "renormalization method": the appropriateness of the name is made manifest by its applications to the less trivial problems that will be discussed after Sec. 3 below.

Consider equation (1.1): we write the solution as

$$\underline{h}(\underline{\psi}) = \varepsilon\, \underline{h}^{(1)}(\underline{\psi}) + \varepsilon^2\, \underline{h}^{(2)}(\underline{\psi}) + \varepsilon^3\, \underline{h}^{(3)}(\underline{\psi}) + \dots$$

and note that the coefficients $\underline{h}^{(k)}$ satisfy an equation like $\underline{h}^{(k)} = \left[\partial_{\underline{\alpha}} f(\underline{\psi} + \underline{h}(\underline{\psi}))\right]^{(k-1)}$ where $[\cdot]^{(k)}$ denotes the k-th Taylor coefficient of an expansion in powers of ε of the function inside the square brackets. Hence

$$\begin{aligned} \underline{h}^{(k)}(\underline{\psi}) &= \Big[\sum_{\underline{s} \geq 0} \frac{1}{\underline{s}!}\, \underline{\partial}_{\underline{\psi}}\, \underline{\partial}^{\underline{s}}_{\underline{\psi}} f(\underline{\psi})\, \underline{h}^{\underline{s}} \Big]^{(k-1)} = \\ &= \sum_{\underline{s} \geq 0} \frac{1}{\underline{s}!} \sum_{\sum k_{ij} = k-1} \underline{\partial}_{\underline{\psi}}\, \underline{\partial}^{\underline{s}}_{\underline{\psi}} f(\underline{\psi}) \prod_{i=1}^{\ell} \prod_{j=1}^{s_i} \underline{h}^{(k_{ij})} \end{aligned} \tag{2.1}$$

where $\underline{s} = (s_1, \dots, s_\ell)$ is a multi–index with $s_i \geq 0$ integer, and we define

$$\underline{s}! \stackrel{def}{=} \prod_i s_i!\,, \quad \underline{\partial}^{\underline{s}}_{\underline{\psi}} \equiv (\partial^{s_1}_{\psi_1} \partial^{s_2}_{\psi_2} \dots)\,;$$

the indices of the components of $\underline{h}$ are contracted with the corresponding indices of the components of $\underline{\partial}^{\underline{s}}_{\underline{\psi}}$ as usual in a Taylor expansion.

Clearly the expression (2.1) is intricate: however we can find a quick graphical representation for it: the function $\underline{h}^{(k)}$ will be represented by

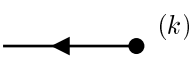

Fig.1: Representation of $\underline{h}^{(k)}$: adding a label $j = 1, \dots, \ell$ on the line will indicate the j-th component of $\underline{h}^{(k)}$.

Therefore we shall represent the relation (2.1) as

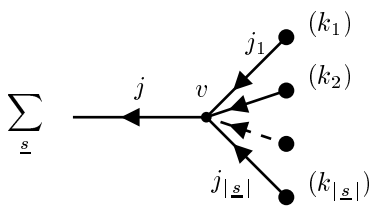

Fig.2: Representation of (2.1); here $k-1=k_1+\ldots+k_{|\underline{s}|}$. The "root" line carries the label j and the other lines carry the labels j_q corresponding to the components of $h^{(k_q)}_{j_q}$ for $q=1,\ldots,|\underline{s}|$. The latter labels will not, in the following figures, continue to be marked (being dummy labels). The "node" v represents the derivatives $\partial_{\underline{\psi}}\partial^{\underline{s}}_{\underline{\psi}}$. Double counting is avoided by using the convention of assigning the label 1 to the first s_1 lines from top to bottom, the label 2 to the next s_2, and so on until all the $|\underline{s}|$ derivation label are considered

The recursive nature of (2.1) is clear and it is also quite clearly reflected in Fig.2 above. The iteration of (2.1) will terminate in finitely many steps and the result can be naturally expressed graphically as in Fig.3 below.

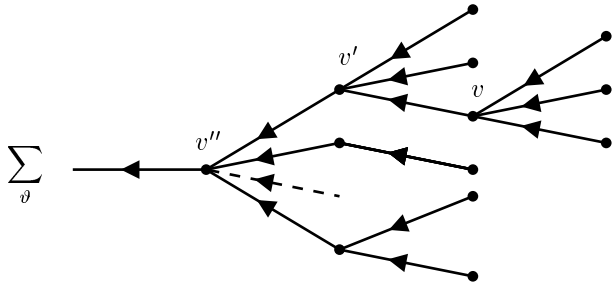

Fig.3: Representation of $\underline{h}^{(k)}$ as a sum of tree graphs ϑ with k nodes. Two pairs of consecutive nodes $v<v'$ and $v'<v''$ are also represented; the node v'' in the picture happens to be what we shall call the first or highest node of ϑ. With each tree graph ϑ a "value" $h_j(\vartheta)$ is assigned (see below) and the sum of the values of all trees with k nodes and root line bearing the label j yields $h^{(k)}_j$. A pair of consecutive nodes will be called a "line".

We imagine to draw the trees in Fig.3 with lines of equal length and "coherently oriented" (*i.e.* the endpoint of an oriented line can only merge into the initial point of another oriented line) by assigning the line labels (not marked in the figures) as explained in the caption to Fig.2: in this way one sees that the number of unlabeled distinct trees ϑ with k nodes does not exceed the number of closed $2k$–steps paths starting at the origin of a one dimensional lattice Z^1, *i.e.* it is $\leq 2^{2k}$. A pair of consecutive (in the partial order fixed by the lines orientations) nodes $\lambda=(v'v)$ will also be called a "line" and it will also be denoted by λ_v if $v<v'$: the orientation of the lines allows us to say that a line follows another or that two lines are comparable, or that a node precedes another node or a line. On each line λ we attach a label $j_\lambda=1,\ldots,\ell$ that we call a *component label*.

We call the highest line (*i.e.* the leftmost in the above figures) the "root" line but do not count the highest extreme of the highest line as a node and we may call it the "root": we shall call the number of nodes (hence of lines) k the "degree" $\deg(\vartheta)$ of ϑ.

Since the function $\underline{h}(\underline{\psi})$ is periodic it is convenient to look for its Fourier coefficients $\underline{h}_{\underline{\nu}}$ with $\underline{\nu}=(\nu_1,\ldots,\nu_\ell)\in Z^\ell$. A simple graphical representation can be given to the coefficients. It suffices to add to each node v of the trees in Fig.3 a label $\underline{\nu}_v$, which we call a "node momentum", and to attribute to each line $\lambda=(v'v)$ a "line current" $\underline{\nu}(\lambda)$ defined as the sum of all the node momenta of the nodes $w<v$:

$$\underline{\nu}(\lambda)=\sum_{w<v}\underline{\nu}_w \tag{2.2}$$

Then if the root line start at the node denoted v_0 and carries a label j we define "*value*" of such labeled trees as

$$i(\underline{\nu}_{v_0})_j\prod_{\lambda\equiv(v'v)\in\vartheta}\left((i\underline{\nu}_v)_{j_\lambda}\,i(\underline{\nu}_{v'})_{j_\lambda}\right)\prod_{v\in\vartheta}\frac{f_{\underline{\nu}_v}}{s_v!} \tag{2.3}$$

and we obtain $h^{(k)}_{j,\underline{\nu}}$ by summing over all trees with k nodes and with current $\underline{\nu}$ flowing on the highest line. In (2.3) we can imagine to have performed the summations over the labels j_λ assigned

to the internal lines, thereby conveniently suppressing all of them except the label j on the root line, so that the summation of the values (2.3) is performed over the trees with k nodes v into which s_v lines merge and which carry a momentum label $\underline{\nu}_v$ over every node v.[4] Double counting is avoided by considering distinct any pair of trees that cannot be trivially superposed by pivoting the lines around the nodes into which they merge but *avoiding*, in the pivoting operations, line crossings.

However it will be convenient to remark that we may imagine that all the lines of such a tree of degree k bear a "*number label*" $1,\dots,k$ (which will never be explicitly marked in the figures) that distinguishes them. Regard as different two trees that cannot be superposed (all labels included) by the operation of pivoting the lines around the nodes into which they merge while, this time, *allowing* crossing of the lines merging into the same node. Then we get many more trees and since we still require that the sum of the values of all trees is $h^{(k)}_{j\,\underline{\nu}}$ the definition of value has to be modified to avoid double countings into

$$\mathrm{Val}(\vartheta) = \frac{i(\underline{\nu}_{v_0})_j}{k!} \prod_{\lambda\equiv(v'v)\in\vartheta} (i\,\underline{\nu}_v \cdot i\,\underline{\nu}_{v'}) \prod_{v\in\vartheta} f_{\underline{\nu}_v} \tag{2.4}$$

With the new way of labeling the number of trees greatly increases, by a factor of order $k!$ being now bounded by $k!2^{2k}$, but the combinatorics becomes simpler for our purposes (even though the sum of the values of all trees is performed with great redundancy). Therefore we shall define the labeling of the trees by imagining that each line carries a number label that distinguishes it from the others.

Note that the node momenta can be supposed *bounded* by $|\underline{\nu}| \le N$ *i.e.* by the degree of f, *cfr* the factors $f_{\underline{\nu}_v}$ in (2.4). However the line currents $\underline{\nu}(\lambda)$ can only be bounded by $|\underline{\nu}(\lambda)| \le kN$ in a tree of degree k. This means that the number of trees with non zero value and degree k is finite and bounded by $2^{2k}(2N+1)^{k\ell}$; the momentum that flows in each line can be as large as kN.

We can say that the value of a tree is the product of node factors (the $f_{\underline{\nu}_v}$ in (2.3)) or "couplings" and of line factors (the $-\underline{\nu}_\nu \cdot \underline{\nu}_{v'}$ and $(i\,\underline{\nu}_{v_0})_j$) or "propagators". The perturbative series for $\underline{h}^{(k)}_{\underline{\nu}}$ acquires in this way the flavor of a Feynman graphs expansion, see Sec.7.

Convergence for small $|\varepsilon|$ of the expansion for $\underline{h}_{\underline{\nu}}$ is immediately proved by bounding the sum of the values of the trees of degree k by

$$\sum_{\vartheta,\deg(\vartheta)=k} |\mathrm{Val}(\vartheta)| \le 2^{2k}(2N+1)^{\ell k} F^k (\ell N^2)^k \tag{2.5}$$

in fact there are at most $2^{2k}k!$ trees and the scalar products $\underline{\nu}_{v'} \cdot \underline{\nu}_v$ give at most ℓ^k terms of size N^2, while the $\underline{\nu}_v$ can be chosen in a number of ways bounded by $(2N+1)^{\ell k} < (3N)^{\ell k}$.

Taking into account that there are at most $(2Nk+1)^\ell$ harmonics (all of which $\le kN$) for which $\underline{h}^{(k)}_{\underline{\nu}}$ is not obviously zero we see that

$$\sum_{\underline{\nu}} e^{\kappa|\underline{\nu}|} |\underline{h}^{(k)}_{\underline{\nu}}| \le (2Nk+1)^\ell ((2N+1)^\ell (\ell N^2) F e^{\kappa N})^k \tag{2.6}$$

so that the function $\underline{h}$ is holomorphic in ε and in $\Im\psi_j$ for $|\Im\psi_j| < \kappa$ and for $|\varepsilon| < \varepsilon_0 = ((2N+1)^\ell \ell N^2 e^{\kappa N} F)^{-1}$. This concludes the "theory" of (1.1) by a "renormalization group approach".

Before passing to study less trivial problems it is worth remarking and stressing that there are important cancellations that occur in summing the trees values. A cancellation, noted in a different context by Lindstedt and Newcomb in particular cases and by Poincaré in general, shows that we can "just" consider trees *in which no line carries a zero current*. In fact if the current flowing through the root line of a tree ϑ is zero we can consider the collection of trees obtained from the given one by "detaching" the root line from the node v_0 ("*root node*" or "*first node*") from which it emerges and by attaching it successively to the other $k-1$ nodes: in this way we form a

[4] Note that now s_v is just an integer ≥ 0 rather than a multiindex: this is due to the summation and to the consequent elimination of the component labels.

collection of k trees whose values differ only because the factor $i(\underline{\nu}_{v_0})_j$ changes as the root node v_0 varies among the tree nodes; therefore the sum of their values is proportional to $i\sum_v(\underline{\nu}_v)_j \equiv \underline{0}$.

The just exhibited cancellation implies that $\underline{h}_{\underline{0}} = \underline{0}$, *i.e.* $\underline{h}$ has zero average. In fact by a similar argument we see that $\underline{h}^{(k)}$ can be computed by considering only the sum of the values of trees in which no line is crossed by a zero current (in a different context this result was established by Poincaré, see below).

Other cancellations are possible: for instance consider a tree ϑ which contains two comparable lines $\lambda_+ < \lambda_-$ (*i.e.* two lines on the same path to the root) on which the same current flows. This means that if $v_1, \dots, v_s$ are the nodes that precede λ_- but which do not precede λ_+ it is $\sum_{p=1}^s \underline{\nu}_{v_p} = \underline{0}$. Then we form the collection of trees obtained by detaching the entering line λ_+ and attaching it successively to the nodes $v_1, \dots, v_s$: we see that the values of all the trees differ only because they contain the factor associated with the propagator of the line λ_+, *i.e.* $-\underline{\nu}_{v_p} \cdot \underline{\nu}_+$ (if v_p is the node to which the line λ_+ is attached and $\underline{\nu}_+$ is the node momentum of the other node of λ_+). Therefore the sum of the values of the collection of trees is proportional to $\sum_j \underline{\nu}_{v_p} \cdot \underline{\nu}_- \equiv \underline{0}$. Of course the same argument applies if we use the line λ_-.

The latter cancellations do not imply that we can compute $\underline{h}$ by summing only the values of trees in which no pair of comparable lines carry zero current: the reason is that the same tree may be necessary to achieve the cancellation relative to "overlapping pairs" of comparable lines, *i.e.* pairs of lines such that the paths joining them along the tree lines overlap. In the case of more difficult problems this is an important obstacle whose proper understanding has been one of the central problems of renormalization theory: in the above case it is not necessary to understand how to disentangle and turn into a useful tool the "overlapping cancellations".

3. Infrared singularities: the problem of KAM theory

Using the terminology of field theory the above is a "one scale problem" because the propagators $-\underline{\nu}_v \cdot \underline{\nu}_{v'}$ are bounded.

The matter becomes much more interesting if one studies what we shall call "Lindstedt equation"

$$(\underline{\omega}_0 \cdot \underline{\partial}_{\underline{\psi}})^2 \underline{h}(\underline{\psi}) = -\varepsilon(\partial_{\underline{\alpha}} f)(\underline{\psi} + \underline{h}(\underline{\psi})) \tag{3.1}$$

where $\underline{h}, f, \varepsilon$ are as in Sec. 1,2 above and $\underline{\omega}_0 = (\omega_1, \dots, \omega_\ell) \in R^\ell$ is a *Diophantine vector*, *i.e.* a vector with the property that there exist two constants $C, \tau > 0$ such that for all non zero integer components vectors $\underline{\nu} \in Z^\ell$ it is

$$|\underline{\omega}_0 \cdot \underline{\nu}| > \frac{1}{C|\underline{\nu}|^\tau}, \qquad \underline{0} \neq \underline{\nu} \in Z^\ell \tag{3.2}$$

Equation (3.1) is substantially more difficult than its "naive" version (1.1). It admits, however, a very similar formal solution: namely $\underline{h}^{(k)}_{\underline{\nu}}$ is given by a "tree expansion" in terms of all the trees (with the same labels and the same counting) considered in the previous case *discarding the trees which contain a line with zero current*; the difference just consists in a different definition of the propagators which change, if $\lambda = (v'v)$ is a line carrying a current $\underline{\nu}(\lambda)$, so that

$$-\underline{\nu}_v \cdot \underline{\nu}_{v'} \;\to\; -\frac{\underline{\nu}_v \cdot \underline{\nu}_{v'}}{(\underline{\omega}_0 \cdot \underline{\nu}(\lambda))^2}, \qquad i(\underline{\nu}_{v_0})_j \;\to\; \frac{i(\underline{\nu}_{v_0})_j}{(\underline{\omega}_0 \cdot \underline{\nu})^2} \tag{3.3}$$

where $\underline{\nu}$ is the the root line current. The new "value" of the labeled trees is, therefore,

$$\mathrm{Val}(\vartheta) = \frac{1}{k!}\frac{i(\underline{\nu}_{v_0})_j}{(\underline{\omega}_0 \cdot \underline{\nu})^2} \prod_{\lambda \equiv (v'v) \in \vartheta} \frac{-\underline{\nu}_v \cdot \underline{\nu}'_v}{(\underline{\omega}_0 \cdot \underline{\nu}(\lambda))^2} \prod_{v \in \vartheta} f_{\underline{\nu}_v} \tag{3.4}$$

and the new difficulty is easily seen. The Diophantine inequality could be "saturated" for large values of the currents $\underline{\nu}(\lambda)$ without *any* of the couplings $f_{\underline{\nu}_v}$ vanishing (*i.e.* $|\underline{\nu}(\lambda)|$ can be very large, of order kN in trees of degree k): therefore if many, say bk for some $b > 0$, lines "resonate"

in the sense that $\underline{\omega}_0 \cdot \underline{\nu}(\lambda)$ is of the order of a power $k^{-\tau}$ the bound on the tree value can become greater, by a factor $k^{\tau b' k} \simeq k!^{b'\tau}$ for some $b' > 0$, than the estimate in (2.5) and the bound is no longer sufficient to achieve a convergence proof no matter how small ε is.

The just described difficulty is called an "*infrared*" problem as it arises from propagators with denominators being *too close to zero at small frequencies*: the quantities $\underline{\omega} \cdot \underline{\nu}(\lambda)/2\pi$ have the interpretation of frequencies that can arise in the Fourier transform of solutions of the mechanical problem that is behind the equation (3.1), *cfr* introduction. It contrasts with *ultraviolet* problems which arise, on the contrary, from propagators with denominators which are away from zero but *not large enough at large frequencies*: an example of such problems is the theory of the boundedness of the functional integral in (1.2) and it will be discussed later.

The key idea for showing that the sum of the tree values in (3.4) over all trees of degree k can still be bounded by B^k for some $B > 0$ is to show that whenever a tree graph appears which contains too many small divisors then cancellations similar to the ones exhibited at the conclusion of Sec.2 take place almost exactly and make the value of the graph small enough for convergence to follow.

It is interesting to check first that the problem really exists. For the purpose it is sufficient to exhibit a single graph of degree k whose value has actually size of order of a power of a factorial of k. The graph is drawn in Fig.4.

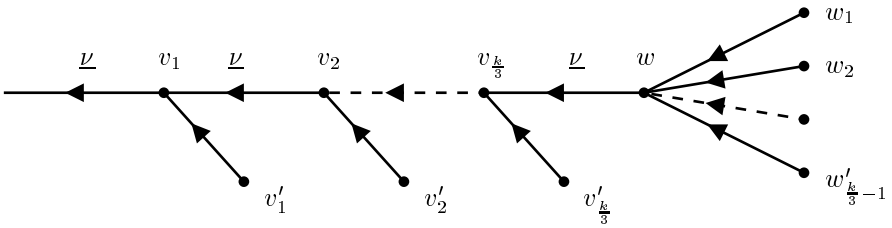

Fig.4: An example of a resonant graph. The graph consists of $\frac{k}{3}$ nodes $v_1, \dots, v_{\frac{k}{3}}$, drawn on a horizontal line, each attached by a line to a side node, $v'_1, \dots, v'_{\frac{k}{3}}$ respectively, and the last $\frac{k}{3}$ nodes are the initial nodes $w_1, w_2, \dots$ of a bunch of $\frac{k}{3} - 1$ lines merging into the node w where the horizontal lines begin. The last $\frac{k}{3} - 1$ nodes carry momenta $\underline{\nu}_{w_1}, \dots, \underline{\nu}_{w_{\frac{k}{3}-1}}$ (not marked in the figure) which together with the momentum $\underline{\nu}_w$ generate at the beginning of the horizontal strectch a current $\underline{\nu}$ suitably constructed to resonate maximally in the sense that $\underline{\omega}_0 \cdot \underline{\nu} \simeq a k^{-\tau/3}$ for some $a > 0$. The nodes $v_1, \dots, v_{\frac{k}{3}}$ carry a small momentum $\underline{\nu}_0$ while the corresponding nodes $v'_1, \dots, v'_{\frac{k}{3}}$ carry momentum $-\underline{\nu}_0$ so that the current flowing in the $\frac{k}{3} + 1$ horizontal lines is steadily $\underline{\nu}$, *i.e.* steadily resonant.

The value of the tree in Fig.4 can be immediately written down from (3.4) (a useful exercise) and one readily sees that there are so many small divisors that the value has size of the order of a power of $k!$.

Formula (3.4) would provide immediately a proof of convergence for small ε if certain trees, among which the one in Fig.4, were not present. The idea, realized in the later sections, is that the "unwanted" trees cancel each other to the extent that their sums behave well enough for not spoiling the bounds.

4. Exhibiting cancellations. The overlapping problem.

To clarify the last paragraph of Sec. 3, *for the purpose of illustration*, we shall first restrict the sum of the contributions (3.4) to $\underline{h}^{(k)}_{\underline{\nu}}$ to a sum over trees which, besides the property that for all lines one has $\underline{\nu}(\lambda) \neq \underline{0}$ (as discussed above), satisfy the property

(P) $\underline{\nu}(\lambda_v) \neq \underline{\nu}(\lambda_{v'})$ *for all pairs of comparable nodes* v, v' *(not necessarily next to each other in the tree order), with* $v' > v$.

There are at most $2^{2k} k!$ trees (as in the simple case of Sec. 2) and the scalar products $\underline{\nu}_{v'} \cdot \underline{\nu}_v$ give at most ℓ^k terms of size N^2, while the $\underline{\nu}_v$ can be chosen in a number of ways bounded by

$(2N+1)^{\ell k} < (3N)^{\ell k}$. Therefore, if $F = \max_{\underline{\nu}} |f_{\underline{\nu}}|$,

$$\begin{aligned} \left| \underline{h}^{(k)}_{\underline{\nu}} \right| &\leq (3N)^{\ell k} 2^{2k} \ell^k F^k C^{2k} N^{2k-1} \max_{\vartheta \in \Theta_{k,\underline{\nu}}} \prod_{\lambda \in V(\vartheta)} (C\,\underline{\omega}_0 \cdot \underline{\nu}(\lambda))^{-2} \\ &\leq (FC^2)^k N^{(\ell+2)k-1} (4\ell 3^\ell)^k M, \end{aligned} \tag{4.1}$$

where M is an estimate of the indicated maximum which is over the k-th degree trees ϑ verifying property (P) above. Hence the whole problem is reduced to find an estimate for M.

Let q be large. By the Diophantine condition in (3.2) one has $C|\underline{\omega}_0 \cdot \underline{\nu}| \geq q^{-1}$ if $0 < |\underline{\nu}| \leq q^{1/\tau}$: we say that the harmonic with Fourier label $\underline{\nu} \in Z^\ell$ is "*q–singular*" if $C|\underline{\omega}_0 \cdot \underline{\nu}| < q^{-1}$ and the following (extension) of a lemma by Bryuno holds for trees of degree k verifying property (P) above:

Fixed $q \geq 1$ let $N(k,q)$ be the number of "q–singular lines" (i.e. of lines corresponding to q–singular harmonics) in a tree ϑ with k nodes. Then

$$N(k,q) \leq \text{const}\, \frac{k}{q^{1/\tau}}, \tag{4.2}$$

and the constant could be taken $2N2^{3/\tau}$.

Remark. The intuition behind (4.2) is very simple. In order to achieve a current $\underline{\nu} = \underline{\nu}(\lambda_v)$ with $C\,\underline{\omega}_0 \cdot \underline{\nu}$ of size q^{-1} one needs at least $|\underline{\nu}| \geq q^{1/\tau}$, *i.e.* by (3.2) the node v must be preceded by *at least* $N^{-1}q^{1/\tau}$ nodes. Once a q–singular line λ has been generated the following lines λ' will have non-q–singular momentum until the number of lines not preceding λ and preceding λ' has grown large of the order of $q^{-1/\tau}$, *i.e.* we must collect *about as many new nodes* (*i.e.* $O(q^{1/\tau})$) to generate a second q–singular line and so on, *at least if* the new singular line λ' does not have the *same* momentum (a case excluded by hypothesis). The latter event would mean that the nodes that precede λ' but do not precede λ have node momenta adding exactly to $\underline{0}$; their presence would invalidate the argument as this is a situation which can be realized *already* with *just* 2 intermediate nodes as in the case of Fig.4. Since the total number of nodes is k it follows that the number of q–singular lines is bounded proportionally to $k/q^{1/\tau}$. The actual estimate of the constant in (4.2) is irrelevant for our immediate purposes.

To proceed we shall assume, for simplicity, that $f_{\underline{\nu}} = f_{-\underline{\nu}}$, *i.e.* that f is an even function. Fix an exponentially decreasing sequence γ^n, $n = 1, 0, -1, -2, \ldots$; we shall make the choice $\gamma = 2$, which recommends itself. The number of 2^{-n}–singular harmonics which are not also $2^{-(n-1)}$–singular is bounded by $2N2^{3/\tau}\, k\, 2^{-n}$, (by (4.2), being trivially bounded by the number of 2^{-n}–singular harmonics!). Hence

$$\prod_{\lambda \in \vartheta} \frac{1}{(C\,\underline{\omega}_0 \cdot \underline{\nu}(\lambda))^2} \leq \prod_{n=-\infty}^{1} 2^{-(n-1)4N2^{3/\tau}2^{n/\tau}k} = e^{cN\tau k} \equiv M, \tag{4.3}$$

where $c > 0$ is a suitable constant (τ–independent); therefore the series for the approximation to $\underline{h}^{(k)}_{\underline{\nu}}$, that we are considering because of the extra restriction (P) on the sum, has radius of convergence in ε bounded below by ε'_0 given by

$$(\varepsilon'_0)^{-1} = (FC^2J^{-1})^k N^{(\ell+2)} (4\ell 3^\ell) e^{cN\tau}. \tag{4.4}$$

The key remark in order to take into account the trees that we have excluded by imposing the *unphysical property* (P) above is that they cancel *almost exactly*. The reason is very simple. Let $\underline{\nu}(\lambda_v) = \underline{\nu}(\lambda')$ with λ' coming out of a node following v and ending in the node v' then we can imagine to detach from the tree ϑ the subtree ϑ_2 with last node v. Then attach it, successively, to

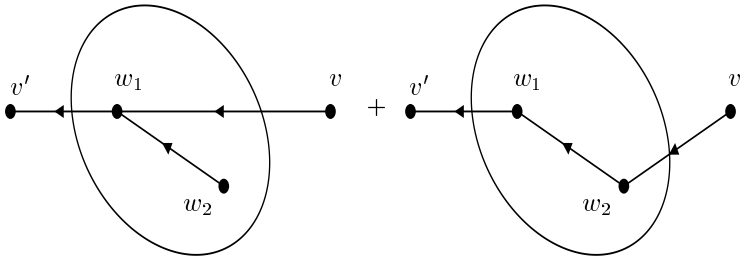

Fig.5: The simplest cancellation: the circle represents a violation of property (P) (which we shall call later a *self-energy graph*), provided $\underline{\nu}_{w_1} + \underline{\nu}_{w_2} = \underline{0}$. The parts of the tree ϑ above v' and below v are not drawn. Imagine that the line momentum $\underline{\nu}$ of the line coming out of v is very large so that $\delta \equiv \underline{\omega}_0 \cdot \underline{\nu}$ is very small and note that in the two trees one has $(\underline{\omega}_0 \cdot \underline{\nu}(\lambda_{w_2}))^2 = (\underline{\omega}_0 \cdot \underline{\nu}_{w_2})^2$ and $(\underline{\omega}_0 \cdot \underline{\nu}(\lambda_{w_2}))^2 = (\underline{\omega}_0 \cdot \underline{\nu}_{w_2} + \delta)^2$, respectively. If the signs of the node momenta of w_1, w_2 are *simultaneously* changed and the values of the four trees obtained in this way are summed we obtain an even function of δ.

all the remaining nodes w which precede λ' but do not precede v. The simplest case is illustrated in Fig.5 with $\lambda' = (v', w_1)$.

We obtain a family of trees whose contributions to $\underline{h}^{(k)}_{\underline{\nu}}$ differ because
(1) some of the lines below λ' changed the current by the amount $\underline{\nu} \equiv \underline{\nu}(\lambda_v)$: this means that some of the denominators $(\underline{\omega}_0 \cdot \underline{\nu}(\lambda_w))^{-2}$ have become $(\underline{\omega}_0 \cdot \underline{\nu}(\lambda_w) + \delta)^{-2}$ if $\delta \equiv \underline{\omega}_0 \cdot \underline{\nu}$ (see the line $\lambda_{w_2} \equiv w_1 w_2$ in Fig.5) and:
(2) the scalar product $\underline{\nu}_v \cdot \underline{\nu}_w$ changes because of the successive changes of the factor $\underline{\nu}_w$, where $w \in \vartheta/\vartheta_2$ is the node to which the line λ_v is reattached.

Hence the sum of the values of all the trees considered plus those obtained by a simultaneous change of the signs of the node momenta of the nodes w preceding λ' but not preceding v would build a quantity which is even in δ. Factoring the common δ^{-4} due to the propagators of the lines λ_v and λ' the remaining sum is a function of δ which for $\delta = 0$ would be proportional to: $\sum \underline{\nu}_w = \underline{\nu}(\lambda') - \underline{\nu}(\lambda_v)$ which is zero (note that the simplifying parity in $\underline{\nu}$ assumed on $f_{\underline{\nu}}$ has to be used here). Since $\delta \neq 0$ we can expect to see that it has order δ^2 which would "cancel" one of the divisors of the lines λ', λ_v.

This is indeed true in the case of Fig.4: by performing the operation depicted in Fig.5 *for each* of the $\frac{k}{3}$ pairs of nodes following the initial bunch of $k/3$ nodes one checks that the result of the sum of the values of the $2\,2^{\frac{k}{3}}$ trees thus obtained is bounded proportionally to $\delta^{2\frac{k}{3}}$ (*i.e.* we get a δ^2 from the cancellation in Fig.5 for each pair of nodes) which compensates the division by $\delta^{2\frac{k}{3}+1}$ due to the small divisors at the cost of adding a factor exponential in k (harmless for the purposes of convergence as it affects only the size of the convergence radius estimate). Therefore although there are too many small divisors *things go as if the whole chain in Fig.4 had only one!*

In general this can be true only if $|\delta| \ll |\underline{\omega}_0 \cdot \underline{\nu}(\lambda)|$ for all lines λ between λ_v and λ'. If the latter property is not true then δ must be small of order $\underline{\omega}_0 \cdot \underline{\nu}(\lambda)$ at least and, hence, this means that there are many nodes w with $v' < w$ but not $\leq v$: indeed in a number of the order needed to create a momentum with small divisors of order $\underline{\omega}_0 \cdot \underline{\nu}(\lambda)$.

The intuitive argument about Bryuno's lemma following (4.2) shows that such an extreme case would be also treatable: after all also in this case the repetition of the small divisor in the lines λ', λ_v is accompanied by a great number of nodes w between v and v' so that the argument given in the remark for estimate (4.2) remains valid. Therefore the problem is to show that the two regimes just envisaged (and their "combinations") do exhaust all possibilities.

5. Multiscale decomposition. Clusters and self energy graphs. Hierarchical organization of cancellations and overlapping control.

Such problems are very common in renormalization theory where they are called *overlapping divergences* problems. Their systematic analysis is made through the "renormalization group methods".

We fix a *scaling* parameter γ, and we take $\gamma = 2$ for consistency with (4.3) (see also the

footnote [1] in the introduction); we also define $\underline{\omega} \equiv C\,\underline{\omega}_0$: it is a dimensionless frequency. Then we say that a propagator $-\underline{\nu}_{v'} \cdot \underline{\nu}_v / (\underline{\omega}_0 \cdot \underline{\nu}(\lambda))^{-2}$ is *on scale* n if $2^{n-1} \leq |\underline{\omega} \cdot \underline{\nu}(\lambda)| < 2^n$, for $n \leq 0$, and we set $n = 1$ if $1 \leq |\underline{\omega} \cdot \underline{\nu}(\lambda)|$.

We make at this point a second simplifying assumption (which can be removed easily, as discussed in the literature quoted below). Namely we want to suppose more than the Diophantine condition (3.2): the condition will be the existence of constants $C, \tau, \gamma > 1$ such that

$$\begin{aligned} &(1) \qquad C\,|\underline{\omega}_0 \cdot \underline{\nu}| \geq |\underline{\nu}|^{-\tau}, \qquad\qquad \underline{0} \neq \underline{\nu} \in Z^l, \\ &(2) \qquad \min_{0 \geq p \geq n} \left| C\,|\underline{\omega}_0 \cdot \underline{\nu}| - \gamma^p \right| > \gamma^{n+1} \qquad \text{if } n \leq 0,\ 0 < |\underline{\nu}| \leq (\gamma^{n+3})^{-\tau^{-1}}, \end{aligned} \tag{5.1}$$

where we shall again take $\gamma = 2$. The property in (5.1) will called the *strong Diophantine condition.* One can check that the set of strongly Diophantine vectors contained in any ball Σ_r of radius r in R^ℓ has measure which tends to $volume(\Sigma_r)$ for $C \to \infty$, *i.e.* the set of strongly Diophantine vectors has full volume.

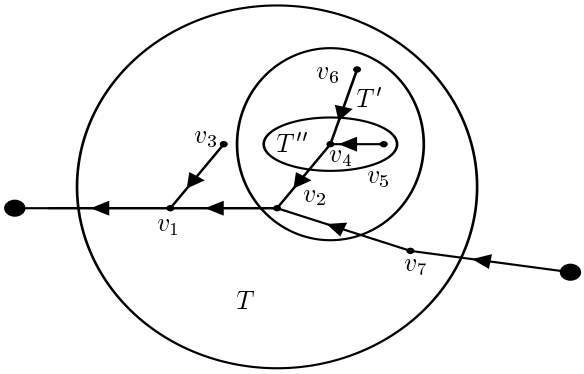

Fig.6: An example of three clusters symbolically delimited by circles, as visual aids, inside a tree (whose remaining lines and clusters are not drawn and are indicated by the bullets); not all labels are explicitly shown. The scales (not marked) of the lines increase as one crosses inward the circles boundaries: recall, however, that the scale labels are ≤ 0. If the mode labels of (v_4, v_5) add up to $\underline{0}$ the cluster T'' is a self-energy graph. If the mode labels of (v_4, v_5, v_2, v_6) add up to $\underline{0}$ the cluster T' is a self-energy graph and such is T if the mode labels of $(v_1, v_2, v_7, v_4, v_5, v_2, v_6)$ add up to $\underline{0}$. The cluster T' is maximal in T.

Given a tree ϑ and calling $\Lambda(\vartheta)$ the set of its lines including the line ending in the root, we can attach a *scale label* to each line $\lambda \in \Lambda(\vartheta)$: it is equal to n if n is the scale of the line propagator. Note that the scale labels thus attached to a tree are *uniquely determined* by the other tree labels: they will have only the function of help in visualizing the orders of magnitude of the divisors associated with the various tree lines.

Looking at such labels we identify the connected *clusters* T *of scale* n_T formed by a set of lines
(i) connected by a continuous path in the tree consisting of lines with scale labels $\geq n_T$,
(ii) which contain at least one line of scale n_T
(iii) and which are maximal with the latter two properties.
We shall say that the "*cluster* T *has scale* n_T".

We shall denote by $V(T)$ the set of nodes in T, and by $\Lambda(T)$ the set of lines connecting them. We also denote by $\Lambda_1(T)$ the set of lines in $\Lambda(T)$ plus the entering and exiting lines of T. Finally call $\mathcal{T}(\vartheta)$ the set of all clusters in ϑ.

Among the clusters we consider the ones with the property that there is only one tree line entering them and only one exiting and both carry the same momentum. Here we recall that the tree lines carry an arrow pointing to the root to give a meaning to the words "entering" and "exiting".

If T is one such cluster and λ_T is the line entering it we say that T is a *self-energy cluster* if the number $M(T)$ of lines contained in T is "not too large"

$$M(T) \stackrel{def}{=} \text{number of lines contained in } T \leq E\,2^{-n\eta}, \tag{5.2}$$

where $n = n_{\lambda_T}$, and E, η are defined by:[5] $E \equiv 2^{-3\eta} N^{-1}$, $\eta = \tau^{-1}$. We call n_{λ_T} the *self-energy-scale* of T, and λ_T a *self-energy line*.

To refer to self-energy clusters T we need some terminology
(a) we denote $\lambda_T \equiv (w_{-T} v_{-T})$ the entering line: its scale $n = n_{\lambda_T}$ is smaller than the smallest scale n_T of the lines inside T; likewise we denote $\lambda_T^+ \equiv (v_{+T} w_{+T})$ the exiting line. Hence w_{-T} is the node inside T into which the entering line λ_T ends.
(b) Let $\widetilde{T}$ be the set of nodes of $V(T)$ *outside* the self-energy clusters *contained* in T (if any).
(c) Denote by $\Lambda(\widetilde{T})$ the set of lines λ contained in T and with at least one point in $\widetilde{T}$, and by $\Lambda_1(\widetilde{T})$ the set of lines in $\Lambda(\widetilde{T})$ *plus* the lines entering and exiting T; note that all lines $\lambda \in \Lambda(\widetilde{T})$ have a scale $n_\lambda \geq n_T$.

Remarks. (1) The self-energy clusters are called *resonances* in Eliasson's terminology, see references.
(2) Note that the self-energy-scale n of a self-energy cluster T (*i.e.* the scale of the entering line) is different from the scale n_T of T as a cluster (*i.e.* the lowest scale of the lines inside the cluster): one has $n < n_T$.

Let us consider a tree ϑ and its clusters. We wish to estimate the number $N_n(\vartheta)$ of lines in $\Lambda(\vartheta)$ with scale $n \leq 0$.

Denoting by T a cluster of scale n let q_T be the number of self-energy clusters of self-energy-scale n contained in T (hence with entering lines of scale n), we have the following inequality.

For all trees $\vartheta \in \Theta_{k,\underline{\nu}}$ one has

$$N_n(\vartheta) \leq \frac{4k}{E\, 2^{-\eta n}} + \sum_{\substack{T \in \mathcal{T}(\vartheta) \\ n_T = n}} (-1 + q_T), \tag{5.3}$$

with $E = N^{-1} 2^{-3\eta}, \eta = \tau^{-1}$.

Remark. This is a version of Bryuno's lemma; a proof is given for completeness in the Appendix below. Intuitively the above inequality has the same content as (4.2): if there are self energy clusters one simply adds to the bound (4.2) (first term in the r.h.s. of (5.3)) the number of such graphs (*i.e.* the sum in (5.3)). For the apparently "extra -1" see appendix A1.

Consider a tree ϑ^1; we define the family $\mathcal{F}(\vartheta^1)$ generated by ϑ^1 as follows. Given a self-energy cluster T of ϑ^1 we detach the part of ϑ^1 which has λ_T as root line and attach it successively to the points $w \in \widetilde{T}$ (note that the endpoint $w_1 \in V(T)$ of λ_T is necessarily among them).

The above procedure is then repeated for all self-energy clusters in ϑ. For each self-energy cluster T of ϑ^1 we shall call V_T the number of nodes in $\widetilde{T}$, *i.e.* $V_T = |V(\widetilde{T})|$. To the just defined set of trees we add the trees obtained by reversing simultaneously the signs of the node modes $\underline{\nu}_w$, for $w \in \widetilde{T}$: the change of sign is performed independently on the various self-energy clusters. This defines a family of $\prod 2V_T$ trees that we call $\mathcal{F}(\vartheta_1)$ (the product is over all self-energy clusters in ϑ). The number $\prod 2V_T$ will be bounded by $\exp \sum 2V_T \leq e^{2k}$.

It is important to note that the definition of self-energy graph is such that the above operation (of shift of the node to which the line entering the self-energy cluster is attached) cannot change too much the sizes of the propagators of the lines inside the self-energy clusters.

This is called the "non overlapping lemma" and the reason behind its validity is simply that inside a self-energy cluster of self-energy-scale n the number of lines is not very large, being $\leq \overline{N}_n \equiv E\, 2^{-n\eta}$.

Indeed let λ be a line contained inside the self-energy clusters $T = T_1 \subset T_2 \subset \ldots$ of self-energy-scales $n = n_1 > n_2 > \ldots$; then the shifting of the lines λ_{T_i} can cause a change in the size of the propagator of λ by at most

[5] This is just a convenient definite choice.

$$2^{n_1} + 2^{n_2} + \ldots < 2^{n+1}. \tag{5.4}$$

For any line λ in $\Lambda(T)$ the quantity $\underline{\omega}_0 \cdot \underline{\nu}_\lambda$ has the form $\underline{\omega}_0 \cdot \underline{\nu}^0_\lambda + \sigma_\lambda \underline{\omega}_0 \cdot \underline{\nu}(\lambda_T)$ if $\underline{\nu}^0_\lambda$ is the momentum of the line λ "inside the self-energy cluster T", *i.e.* it is the sum of all the node momenta of the nodes preceding λ in the sense of the line arrows, but contained in T; and $\sigma_\lambda = 0, 1$.

Therefore not only $|\underline{\omega} \cdot \underline{\nu}^0(\lambda)| \geq 2^{n+3}$ (because $\underline{\nu}^0(\lambda)$ is a sum of $\leq \overline{N}_n$ node momenta, so that $|\underline{\nu}^0(\lambda)| \leq N\overline{N}_n$) but $\underline{\omega} \cdot \underline{\nu}^0(\lambda)$ is "in the middle" of the diadic interval containing it and, by the strong Diophantine property(5.1), does not get out of it if we add a quantity bounded by 2^{n+1} (like $\sigma_\lambda \underline{\omega}_0 \cdot \underline{\nu}(\lambda_T)$). Hence no line changes scale as ϑ varies in $\mathcal{F}(\vartheta^1)$, if $\underline{\omega}_0$ verifies (5.1).

By the strong Diophantine condition (5.1) on $\underline{\omega}_0$ the self-energy clusters of the trees in $\mathcal{F}(\vartheta^1)$ all contain the same sets of lines, and the same lines enter or exit each self-energy cluster (although they are attached to generally distinct nodes inside the self-energy clusters: the identity of the lines is here defined by the number label that each of them carries in ϑ^1). Furthermore the scales of the self-energy clusters, and in fact of all the lines, do not change.

Let ϑ^2 be a tree not in $\mathcal{F}(\vartheta^1)$ and construct $\mathcal{F}(\vartheta^2)$, *etc*, obtaining in this way a collection $\{\mathcal{F}(\vartheta^i)\}_{i=1,2,\ldots}$ of pairwise disjoint families of trees. We shall sum all the contributions to $\underline{h}^{(k)}_{\underline{\nu}}$ coming from the individual members of each family and then sum over the families. This is a realization of *Eliasson's resummation*: it is more detailed than his original one, where no subdivision of the trees in classes was considered and the cancellation implied by the one that we exhibit in Sec. 6 was derived from an argument involving all graphs at the same time. Thus the Eliasson cancellation can be regarded as a cancellation due to a special symmetry of the problem (analogous to the Ward identities of field theory) and the above analysis shows that more symmetry is present as the cancellation takes place already at a lower level in which less trees are added together.

This completes the organization of the tree values which makes evident the cancellations necessary to show, see Sec.6, that not only the problem associated with the tree in Fig.4 but also the analogous problem in the most general graph can be solved by the above considerations.

6. Cancellations and dimensional bounds.

The above hierarchical organization of the sum of the terms giving rise to the k–th order contribution $h^{(k)}_j$ is sufficient for our purposes. One can proceed to bound the sum of the contributions from each collection of terms straightforwardly by using repeatedly the maximum principle (namely the bound of the value of an analytic function at a point by its maximum modulus in a (complex) region around the point divided by the distance to the boundary of the region).

Referring to the notions associated with the self energy clusters, see items a,b,c following (5.2), we call η_T the quantity $\underline{\omega}_0 \cdot \underline{\nu}(\lambda_T)$ associated with the self-energy cluster T. If λ is a line in $\Lambda(\widetilde{T})$, defined after (5.2), we can imagine to write the quantity $\underline{\omega}_0 \cdot \underline{\nu}(\lambda)$ as $\underline{\omega}_0 \cdot \underline{\nu}^0(\lambda) + \sigma_\lambda \eta_T$, with $\sigma_\lambda = 0, 1$: the product of the propagators of the lines *inside* $\widetilde{T}$ is

$$\prod_{\lambda \equiv (v'v) \in \Lambda(\widetilde{T})} \frac{-\underline{\nu}_v \cdot \underline{\nu}_{v'}}{(\underline{\omega}_0 \cdot \underline{\nu}^0(\lambda) + \sigma_\lambda \eta_T)^2}. \tag{6.1}$$

For simplicity we do not explicitly distinguish the possibility that λ is the root line: in that case the corresponding factor in (6.1) has the slightly different form $-i(\underline{\nu}_{v_0})_j/(\underline{\omega}_0 \cdot \underline{\nu}(\lambda))^2$.

If the tree does not contain any self-energy clusters, we say that it has *height* 0; if the only self-energy clusters do not contain other self-energy clusters, we say that the tree has height 1, see Fig.4 for an example; more generally if the maximum number of self-energy clusters that contain a given self-energy cluster is p, we say that the *tree has height p*. Similarly we say that a *self-energy cluster has height p* if it contains at least one self-energy cluster that is contained in exactly p self-energy clusters and none which is contained in more ($p = 0$ corresponds to a self-energy cluster

which does not contain any other self-energy clusters). Given a tree ϑ, call $V(\vartheta)$ the set of all self-energy clusters in ϑ, and set

$$\Lambda(V(\vartheta)) = \cup_{T\in V(\vartheta)}\Lambda(T), \qquad \Lambda_1(V(\vartheta)) = \cup_{T\in V(\vartheta)}\Lambda_1(T). \tag{6.2}$$

Of course in (6.2) the union could be restricted only to the maximal self-energy clusters.

First consider the simple case of a tree ϑ of height 1 and let us denote by T any of its self-energy clusters: if we regard the quantities η_T as independent variables we see that (6.1) is holomorphic in η_T for $|\eta_T| < 2^{n_T-3}$. While η_T varies in such complex disk the quantity $|\underline{\omega}_0\cdot\underline{\nu}^0(\lambda)+\sigma_\lambda\eta_T|$ does not become smaller than 2^{n_T-3}.[6] The main point here is that the quantity 2^{n_T-3} will usually be $\gg 2^{n_{\lambda_T}}$ which is the value that η_T actually can reach in every tree in $\mathcal{F}(\vartheta)$; this is what happens in the special case of Fig.4 and it can be exploited in applying the maximum principle, as done below.

Note that the quantities η_T do not depend on the element of the family $\mathcal{F}(\vartheta)$ so that we could factor out of the sum of the values of the graphs in $\mathcal{F}(\vartheta)$ the factors η_T^{-2}; we can write the product of the propagators of any tree as

$$\begin{aligned}&\Big(\prod_{\substack{\lambda\in\Lambda(\vartheta)\setminus\Lambda_1(V(\vartheta))\\ \lambda\equiv(v'v)}}\frac{-\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}_0\cdot\underline{\nu}(\lambda))^2}\Big)\cdot\Big(\prod_{T\in V(\vartheta)}\prod_{\lambda\in\Lambda(T)}\frac{-\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}_0\cdot\underline{\nu}^0(\lambda)+\sigma_\lambda\eta_T)^2}\Big)\cdot\\ &\cdot\Big(\prod\nolimits^+_{T\in V(\vartheta)}\frac{-\underline{\nu}_{v_T^+}\cdot\underline{\nu}_{w_T^+}}{\eta_T^2}\Big)\cdot\Big(\prod\nolimits^-_{T\in V(\vartheta)}\frac{-\underline{\nu}_{w_T^-}\cdot\underline{\nu}_{v_T^-}}{\eta_T^2}\Big)\end{aligned} \tag{6.3}$$

where the first product is over the lines λ which neither enter nor exit nor are inside a self-energy cluster of ϑ (so that their momentum is the same in all trees of the family $\mathcal{F}(\vartheta)$), the second product is over the lines λ contained in $V(\vartheta)$, the third product is over the self-energy clusters $T\in V(\vartheta)$ and takes into account the lines exiting T *but not entering another self energy cluster* and the last product is over the lines that *enter* the self energy clusters.

As said above the denominators η_T^2 factor out of the sum of the values of the trees in $\mathcal{F}(\vartheta)$ at fixed ϑ. We can therefore consider the sum of the $\prod 2V_T \le e^{2k}$ values of the graphs members of the family $\mathcal{F}(\vartheta)$ *divided by the product of the factors* η_T^{-2} *associated with the lines entering or exiting the self energy clusters*, *i.e.* we consider the sum of the values in (6.3) computed without the denominators in the last two products.

Each such sum is holomorphic in the region $|\eta_T| < 2^{n_T-3}$ and in the latter region it is bounded by $\prod 2^{-2(n_\lambda-3)} \le 2^{6k}\prod_\lambda 2^{-2n_\lambda}$, if n_λ the scale of the line λ in ϑ and if the product is over the lines neither entering nor exiting a self-energy cluster. This even holds if the η_T are regarded as independent complex parameters.

By construction the just considered sum of the $\prod 2V_T \le e^{2k}$ terms from the trees in $\mathcal{F}(\vartheta)$, vanishes to second order in *each* of the η_T parameters (by the approximate cancellation discussed above due to the fact that the sum $\sum_{w\in T}\underline{\nu}_w = \underline{0}$ and to the parity property supposed for $f_{\underline{\nu}}$). *By the maximum principle* this means that if we bound the sum by the number of terms times the maximum among them (which is easy to estimate because the propagators have all well defined seizes fixed by their scales) we can multiply the result by a further factor of the order of $2^{2n_{\lambda_T}}/2^{2(n_T-3)}$ and still obtain a valid bound.

Hence by the maximum principle and, recalling that each $\underline{\nu}_v$ can be bounded by N, we can bound the contribution to $\underline{h}^{(k)}_{\underline{\nu}}$ from the family $\mathcal{F}(\vartheta^1)$ by

$$\Big[\frac{1}{k!}\Big(FC^2N^2\Big)^k 2^{6k}e^{2k}\prod_{n\le 0}2^{-2nN_n}\Big]\Big[\prod_{n\le 0}\prod_{\substack{T\in\mathcal{T}(\vartheta)\\ n_T=n}}\prod_{i=1}^{q_T}2^{2(n-n_i+3)}\Big], \tag{6.4}$$

[6] In fact $|\underline{\omega}_0\cdot\underline{\nu}^0(\lambda)| \ge 2^{n+3}$ because T is a self-energy cluster; therefore $|\underline{\omega}_0\cdot\underline{\nu}(\lambda)| \ge 2^{n+3}-2^{n+1} > 2^{n+2}$ so that $n_T \ge n+3$. On the other hand we note that $|\underline{\omega}_0\cdot\underline{\nu}^0(\lambda)| > 2^{n_T-1}-2^{n+1}$, so that it follows that $|\underline{\omega}_0\cdot\underline{\nu}^0(\lambda)+\sigma_\lambda\eta_T| \ge 2^{n_T-1}-2^{n+1}-2^{n_T-3} \ge 2^{n_T-3}$, for $|\eta_T| < 2^{n_T-3}$.

where
(1) $N_n = N_n(\vartheta)$ is the number of propagators of scale n in ϑ^1 ($n = 1$ does not appear because $|\underline{\omega} \cdot \underline{\nu}| \geq 1$ in such cases);
(2) the first square bracket is the bound on the product of individual elements in the family $\mathcal{F}(\vartheta^1)$ times the bound e^{2k} on their number: this takes into account *also* the last product in (6.3);
(3) the second square bracket is the part coming from the maximum principle, applied to bound the resummations, and is explained as follows.
(4) The dependence on the variables $\eta_{T_i} \equiv \eta_i$ relative to self-energy clusters $T_i \subset T$ with self-energy-scale $n_{\lambda_{T_i}} = n$ is holomorphic for $|\eta_i| < 2^{n_i - 3}$, if $n_i \equiv n_{T_i}$, provided $n_i > n + 3$ (see above).
(5) The resummation says that the dependence on the η_i's has a second order zero in each. Hence the maximum principle tells us that we can improve the bound given by the third factor in (6.3) by the product of factors $(|\eta_i|\, 2^{-n_i+3})^2$ as $n_i \geq n+3$ which yield the product in the second square bracket.

The above would be sufficient if there were no trees of height higher than 1. In fact substituting (5.3) into (6.4) we see that the q_T is taken away by the first factor in $2^{2n}2^{-2n_i}$, while the remaining 2^{-2n_i} are compensated by the -1 before the $+q_T$ in (5.3), taken from the factors with $T = T_i$, (note that there are always enough -1's). It follows that the product (6.4) is bounded by

$$\frac{1}{k!}\,(C^2FN^2)^k e^{2k} 2^{12k} \prod_{n \leq 0} 2^{-8nkE^{-1}\, 2^{\eta n}} \leq \frac{1}{k!}\, B_0^k, \tag{6.5}$$

with B_0 suitably chosen.

To sum over the trees we note that fixed ϑ the collection of clusters is fixed. Therefore we only have to multiply (6.5) by the number of tree shapes for ϑ, ($\leq 2^{2k}k!$), by the number of ways of attaching momentum labels, ($\leq (3N)^{\ell k}$), by the number of ways of contracting the tensor labels, ($\leq \ell^k$), so that we can bound $|\underline{h}\,\underline{\nu}^{(k)}|$ by

$$\varepsilon_0^{-k} \equiv (b_\ell C^2 F N^{2+\ell} e^{cN})^k, \tag{6.6}$$

with b_ℓ suitably chosen.

To treat the general case we can proceed inductively and suppose that the bound in (6.4) holds for trees of height $1, 2, \ldots, p-1$ and for values of the $\eta_T = \underline{\omega}_0 \cdot \underline{\nu}(\lambda_T)$ of the lines that enter the maximal self-energy clusters T which are in the complex disk $|\eta_T| < 2^{n_T - 3}$, see Appendix.

7. Feynman graphs for the integral in (1.2).

The analysis of the problem (1.2),(1.3) is started by checking the existence of a formal series expansion in λ: which, of course, has to be followed by the study of the convergence (in fact of the asymptoticity) properties of the series.

The first problem is an easy one and its solution is classical: it is most easily described in terms of graphs: consider the following *graphical elements*

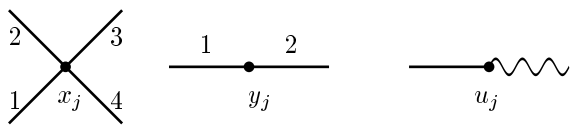

Fig.7: The three graphical elements for the Feynman graphs expansion of $\log E_{\lambda,\mu}(\varepsilon f)$. The labels $1, 2 \ldots$ signify that the lines of a single graph element must be considered as distinct.

The coefficient of $\lambda^{k_4}\mu^{k_2}\varepsilon^{k_1}$ in the expansion of $\log E_{\lambda,\mu}(\varepsilon f)$ is obtained by considering the *connected* graphs Γ that can be formed with k_4 graph elements of the first type in Fig.7, *i.e.* with 4 lines, and with k_2 graph elements of the second type and k_1 elements of the third type and then merging pairwise the solid lines to form a connected graph with k_1 wiggly lines left unpaired. One often refers to lines obtained by merging a pair of lines by calling it a "*contraction*" so that the

various graphs are obtained by contracting pairwise the non wavy lines of the graph elements in Fig.7. With each such Feynman graph we associate an "*amplitude*" which is simply

$$\int (-\lambda)^{k_4}(-\mu)^{k_2}(-\varepsilon)^{k_1} \prod_{(\xi_i\xi_j)\in\Gamma} C(\xi_i-\xi_j)\frac{\prod d\xi_j}{k_4!k_2!k_1!} \tag{7.1}$$

where $(\xi_1,\dots,\xi_{k_4+k_2+k_1}) = (x_1,\dots,x_{k_4},y_1,\dots,y_{k_2},u_1,\dots,u_{k_1})$.

For instance for $k_2=1,k_1=2,k_4=0$ and $k_2=0,k_1=2,k_4=2$ we get, respectively, graphs like

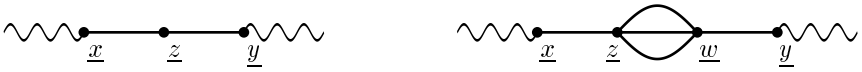

Fig.8: two graphs contributing to the orders $\mu\varepsilon^2$ and $\lambda^2\varepsilon^2$, respectively, to $\log E_{\lambda,\mu}(\varepsilon f)$.

Many graphs which differ only by the identity of the lines that are contracted yield the same value. If we do not write the identity labels on the lines then each graph Γ has to be multiplied by a suitable combinatorial factor $n(\Gamma)$ for an appropriate count. The total number $\sum_{\Gamma\ labeled} 1 = \sum_{\Gamma\ unlabeled} n(\Gamma)$ is of the order of $(2(k_4+k_2+k_1))!$.

Since the propagator $C(\underline{x}-\underline{y})$ in (1.2) is a continuous function which decays exponentially as $|\underline{x}-\underline{y}|\to\infty$ it is clear that the integrals in (7.1) are finite for all graphs: hence the formal power series for the generating function $\log E_\lambda(\varepsilon f)$ of the Schwinger functions is well defined to all orders.

In the present case we certainly cannot have convergence of the latter well defined expansion for the obvious reason that the integral in (1.2) is (almost) obviously divergent for $\lambda<0,\mu,\varepsilon=0$ (divergence can be established because in this case all integrals are non negative and admit a lower bound B^k that grows exponentially in k so that the order k coefficient grows as $B^k(2k)!/k!$, *i.e.* too fast).

Nevertheless convergence of the integral can be proved as well as the asymptoticity of the formal power series to which it is formally equal: this is a consequence of an important inequality due to Nelson. However here we shall not discuss this point further.

8. An ultraviolet problem: φ^4_d for $d=2,3$. Multiscale decomposition and dimensional estimates.

One of the most studied problems in renormalization theory is the analysis of the integral (1.2) in which the probability distribution $P(d\varphi)$ is Gaussian with a covariance different from (1.3) and given by

$$C(\underline{x}-\underline{y}) \stackrel{def}{=} \langle\varphi_{\underline{x}}\varphi_{\underline{y}}\rangle_P = \frac{1}{(2\pi)^d}\int e^{i\underline{p}\cdot(\underline{x}-\underline{y})}\frac{1}{\underline{p}^2+1}\,d^d\underline{p} \tag{8.1}$$

We see that $C(\underline{x})$ decays exponentially as $|\underline{x}|\to\infty$ *but* $C(0)=\infty$ which means that with P-probability 1 the functions $\varphi_{\underline{x}}$ are in fact rather singular and, more precisely, are distributions. The rate of divergence of $C(\underline{x})$ as $\underline{x}\to 0$ is $\simeq -\log|\underline{x}|$ if $d=2$ and $\simeq|\underline{x}|^{-1}$ if $d=3$.

Therefore not only the problem is harder, but it is not even clear whether it makes sense at all since the function in the exponent in (1.2) is no longer meaningful.

The logarithmic divergence in $d=2$ is "very weak" that one checks that all graphs without self contractions, *i.e.* without any pairing of lines emerging from the same graph element, are finite. However in general such contractions occur and therefore yield factors $C(\underline{0})=+\infty$. Hence, clearly, the problem is not well posed.

The physical interpretation of the integral (1.2) with P with covariance (8.1) does not require that the integral be well defined for all λ,μ small but "just" that there is a function $\mu(\lambda)$ (possibly depending also on $\underline{x}$) such that the integral is meaningful. In other words one asks whether one can find $\mu(\lambda)$ such that the $\log E_\lambda(\varepsilon f)$ is well defined and smooth in λ,ε for $\lambda>0$ small and $\mu=\mu(\lambda)$. It is not surprising that the $\mu(\lambda)$ if at all existent should be *infinite*! Of course the fault can be attributed to the fact that the integrand itself is not well defined.

One tries to attach a meaning to the integral by replacing the probability distribution $P(d\varphi)$ by a *regularized* distribution $P_N(d\varphi)$ where N is a "*cut–off*" parameter and P_N is a Gaussian functional integral with covariance

$$C^{(N)}(\underline{x}-\underline{y}) = \frac{1}{(2\pi)^d}\int \frac{e^{i\,\underline{p}\cdot(\underline{x}-\underline{y})}}{1+\underline{p}^{\,2}}\chi_N(\underline{p})\,d^d\,\underline{p} \tag{8.2}$$

where $\chi_N(\underline{p}) \xrightarrow[N\to\infty]{} 1$. Possible choices are

$$\chi_N(\underline{p}) = \begin{cases} 1 & \text{if } |\underline{p}| \le 2^N \\ 0 & \text{if } |\underline{p}| > 2^N \end{cases} \qquad \text{or} \qquad \chi_N(\underline{p}) = \frac{2^{2N}-1}{2^{2N}+\underline{p}^{\,2}} \tag{8.3}$$

where the first choice is is perhaps the most natural while the second might be the easiest technically. The matter is debated: the two choices however lead to the same result in the limit as $N\to\infty$ (*i.e.* to the same $N=\infty$ limit value for the integral (1.2)). Here I shall follow the traditional approach that uses a smooth cut–off, *e.g.* the second choice in (8.3), for expository reasons (brevity): however the first choice is gaining grounds in the recent research works on "exact renormalization group".

If C is replaced by $C^{(N)}$ as defined by the second of (8.3) the integral becomes well defined essentially because it becomes like the one studied in Sec. 4: the Fourier transform of the propagator goes to 0 as $\underline{p}^{-4}$ and $C^{(N)}(\underline{x})$ is a continuous exponentially decreasing function so that all integrals in the perturbative expansion for $\log Z$ are finite and Z itself is well defined thanks to the Nelson inequality.

The possibility of taking advantage of the freedom of the choice of $\mu(\lambda)$ then leads to consider the integral

$$\begin{aligned} E_{\lambda,N}(\varepsilon f) &= \frac{\int P_N(d\varphi) e^{-\int (V_N(\varphi_x)+\varepsilon\varphi_{\underline{x}} f(\underline{x}))\,d^d\,\underline{x}}}{\int P_N(d\varphi) e^{-\int V_N(\varphi_x)\,d^d\,\underline{x}}} \\ V_N(\varphi_x) &= \lambda\varphi^4_{\underline{x}} + \mu_N(\underline{x})\varphi^2_{\underline{x}} \end{aligned} \tag{8.4}$$

and the problem is to find μ_N so that the limit as $N\to\infty$ of (8.4) exists and is smooth in λ,ε,f. The quantity μ is allowed to depend on $\lambda, N, \underline{x}$ (but not $\varphi_{\underline{x}}$) with the only condition that it should be bounded uniformly in $\underline{x}$ *at fixed* N so that at fixed N the formal perturbation expansion is well defined.

If $d=2$ the above remark that the only divergent (as $N\to\infty$) graphs are the ones with self contractions leads immediately to try to determine μ_N in such a way that all such graphs cancel each other.

From the theory of Gaussian integrals it is well known that elimination of the graphs with self contractions is possible simply by requiring that V_N be a linear combination of "Wick monomials" defined as $:\varphi^n_{\underline{x}}: := \sqrt{2C^{(N)}(0)}^{\,n} H_n(\frac{\varphi_{\underline{x}}}{\sqrt{2C^{(N)}(0)}})$ where $H_n(x)$ is the n–th Hermite polynomial ($H_4(z)=z^4-3z^2+\frac{3}{2}$, $H_2(z)=z^2-\frac{1}{2}$, ...) and $C^{(N)}(0)=\langle\varphi^2_{\underline{x}}\rangle_{P_N}$.

It is therefore very convenient to start with a V_N of the form

$$\begin{aligned} &V_N(\varphi) = \lambda :\varphi^4_{\underline{x}}: + \overline{\mu}_N(\underline{x}) :\varphi^2_{\underline{x}}: \equiv \\ &\equiv \lambda\,(\varphi^4_{\underline{x}} - 6\,C^{(N)}(0)\,\varphi^2_{\underline{x}}) + \overline{\mu}_N(\underline{x})\,\varphi^2_{\underline{x}} + const \end{aligned} \tag{8.5}$$

and in the case $d=2$ one can simply take $\overline{\mu}_N\equiv 0$ which in terms of the notation in (8.4) means $\mu_N(\underline{x}) = -6\,\lambda\,C^{(N)}(0)$ (which diverges as $N\to\infty$ as expected and depends on λ). The perturbative analysis is complete and gives a finite result (uniformly in N) order by order: it remains the hard part of the job which is to prove the existence of the limit as $N\to\infty$ of $E_{\lambda,N}(\varepsilon f)$, (8.4). This is not discussed here because we only want to show the analogy between the KAM and the field theory problems. Therefore we shall eventually concentrate attention on the much more interesting problem of the perturbation analysis of (8.4) in the $d=3$ case.

Of course, even with the choice in (8.5) if $d=3$ all integrals depend on N with divergences occurring as $N\to\infty$ at least if one does not attempt to use the freedom in the choice of $\overline{\mu}(\lambda)$.

Nevertheless at fixed N we can write the perturbation expansion, or better its formal coefficients, just as in the case discussed in Sec. 4 and we therefore imagine to have written the expansion of $\log E_{\lambda,\mu}(\varepsilon f)$ in powers of $\lambda, \mu, \varepsilon$ via exactly the Feynman graphs of Sec. 4 with the new propagator $C^{(N)}$ and with the constant $\overline{\mu}$ allowed to depend on $N, \underline{x}, \lambda$ and *no Feynman graphs with self contractions.*

Since $C^{(N)}$ is the Fourier transform of the function $\frac{2^{2N}-1}{(2^{2N}+\underline{p}^2)(1+\underline{p}^2)} = \frac{1}{1+\underline{p}^2} - \frac{1}{2^{2N}+\underline{p}^2}$ we can write

$$\begin{aligned} C^{(N)}(\underline{x}-\underline{y}) &= \frac{1}{(2\pi)^d}\int d^d\,\underline{p}\, e^{i\,\underline{p}\cdot(\underline{x}-\underline{y})}\sum_{h=0}^{N-1}\Big(\frac{1}{2^{2h}+\underline{p}^2}-\frac{1}{2^{2(h+1)}+\underline{p}^2}\Big) = \\ &= \frac{1}{(2\pi)^d}\int d^d\,\underline{p}\, e^{i\,\underline{p}\cdot(\underline{x}-\underline{y})}\sum_{h=0}^{N-1} 2^{2h}\frac{3}{(2^{2h}+\underline{p}^2)(2^{2(h+1)}+\underline{p}^2)} = \\ &= \sum_{h=0}^{N-1} 2^{(d-2)h}\overline{C}(2^h(\underline{x}-\underline{y})) \stackrel{def}{=} \sum_{h=0}^{N-1} C_h(\underline{x}-\underline{y}) \end{aligned} \tag{8.6}$$

where $\overline{C}$ has a Fourier transform $3/(1+\underline{p}^2)(4+\underline{p}^2)$, which gives a "*multiscale decomposition*" of the "ultraviolet" singularity of the propagator in the limit as $N \to \infty$.

The "scale covariant" representation of the propagator achieved by the decomposition in (8.6) can be used to decompose finely the integrals corresponding to the Feynman graphs and defining the perturbation expansion coefficients and to rearrange the sums of the terms thus obtained in a way that exhibits the remarkable cancellations that will allow us to show that *if* $\mu_N(\lambda)$ *is suitably defined* then the perturbation expansion of $E_\lambda(\varepsilon f)$ (defined as the limit of $E_{\lambda,N}(\varepsilon f)$) is well defined order by order in λ, ε thus proving renormalizability (and performing renormalization).

The latter result will not be sufficient yet to show that the generating function of the Schwinger functions is actually well defined and defines a new non trivial (*i.e.* non Gaussian) probability distribution over the fields $\varphi_{\underline{x}}$: again the reason is that we cannot expect that the perturbative series be convergent since the expressions that they should define are not defined for $\lambda < 0$ already for $N < \infty$. Nevertheless renormalizability is a key ingredient and a first step in the proof of existence of the no cut–off limit of the generating function for the Schwinger functions.

We imagine to associate with each line $\ell = (\underline{x}, \underline{y})$ of the Feynman graphs a "*scale label*" $h_\ell = N, N-1, \dots, 0$ and we shall define its value by replacing the propagator $C^{(N)}(\underline{x}-\underline{y})$ with $C_{h_\ell}(\underline{x}-\underline{y}) = 2^{(d-2)h_\ell}\overline{C}(2^{h_\ell}(\underline{x}-\underline{y}))$, *i.e.* with the propagator on scale h_ℓ.

Fixed a Feynman graph Γ of the latter type we can associate with it a tree graph ϑ: we define first the "*clusters*" of lines in Γ: a "cluster of scale h" will be a connected subset of lines whose scales are $\geq h$ which, furthermore, contains at least one line of scale h and which is a maximal set with the latter two properties.

It will be convenient and natural to regard each vertex of the Feynman graph as a cluster, in fact as a cluster of scale $h+1$ if h is the highest scale of the graph lines that merge into the vertex, even though they contain no lines. In this way the number of vertices of a Feynman graph and the number of nodes of the corresponding tree are equal.

The clusters are by definition arranged hierarchically and are naturally partially ordered by inclusion: therefore they can be represented by the nodes of a tree ϑ. The set of all graphs which give rise to the same tree will be called the set of graphs Γ compatible with ϑ and we shall write, somewhat improperly, $\Gamma \subset \vartheta$.

Note that the lines of the tree have nothing to do with the lines of the Feynman graph. An illustration of the above definitions is provided by the following Fig.9 (with a Feynman graph with 7 vertices).

Instead of thinking the coefficient of orders k_4, k_2, k_1 in $\lambda, \mu, \varepsilon$ as a sum of values of Feynman graphs we can think of it as a double sum over all trees with $k_4+k_2+k_1$ nodes and over all graphs compatible with each tree.

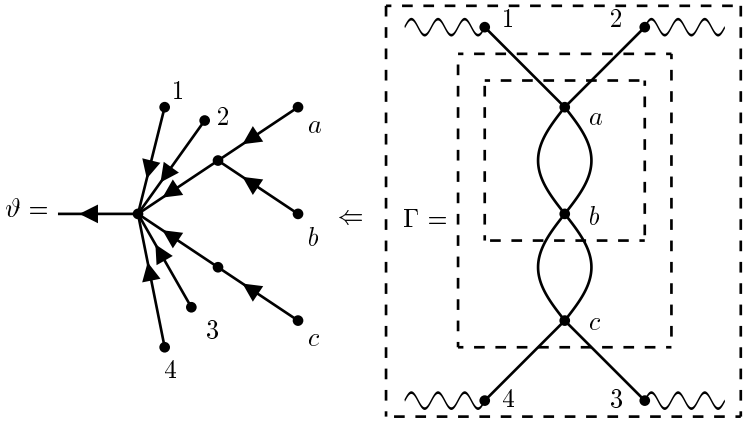

Fig.9: An example of a Feynman graph Γ with its clusters. The cluster structure uniquely identifies a tree ϑ. The nodes a, b, c are supposed to represent the first graph element of Fig.7 while the nodes $1, 2, 3, 4$ represent the third graph element: altogheter such seven vertices correspond to the seven lowest nodes of the tree, *i.e.* they are the innermost clusters. The clusters are here represented by dashed rectangles (rather than by ellipses as in Fig.5,6, to avoid confusion with the graph loops). The three unlabeled nodes of ϑ correspond to the three dashed rectangles. It is instructive to draw the corresponding, quite different, picture in the case the line $1a$ (for instance) has a scale larger than all the others.

Given a tree ϑ and a graph Γ compatible with it
(1) we say that a graph vertex v belongs to a cluster if at least one of the lines that end in it is part of the cluster;
(2) we say that a node $v \in \vartheta$ corresponding to a cluster of scale h_v has degree p_v if there are p_v lines "emerging" from it, *i.e.* p_v lines of the graph have an extreme vertex in the cluster v but have scale $h < h_v$. It is possible that both ends of the line are vertices in a cluster but the line has scale lower than that of the cluster: in this case the line counts twice in the definition of p_v (because we imagine that two lines emerge from the cluster and are then contracted on a lower scale).
(3) the number p_v of lines external to the cluster v should not be confused with the number s_v defined as the number of tree nodes that precede v or, equivalently, as the number of clusters contained in the cluster v but not in smaller clusters.

The cluster structure sets a natural order in the integrations necessary to evaluate the graph value. Taking into account the expression of the graph values in (7.1) and the form $2^{(d-2)h}C_1(2^h(\underline{x}-\underline{y}))$ of the propagator on scale h we can bound the graph value simply by bounding the propagators on scale h by $B_0 2^{(d-2)h} e^{-\kappa 2^h|\underline{x}-\underline{y}|}$, for some $B_0, \kappa > 0$, and use the exponential decay to bound the results of the integrals.

Let h_0 be the scale of the root line. Let $M_{4,v}, M_{2,v}, M_{1,v}$ be the numbers of vertices of the graph elements in the cluster v which correspond, respectively, to the first, second and third graph element in Fig.7 and let $m_{4,v}, m_{2,v}, m_{1,v}$ be the numbers of vertices in the cluster v of scale h_v *which are not contained inside inner clusters* and which correspond to the mentioned graph elements. One remarks that

$$\sum_v (h_v - h_0)(s_v - 1) = \sum_v (h_v - h_{v'})(M_{4,v} + M_{2,v} + M_{1,v} - 1) \tag{8.7}$$

and it follows that the bound on the value of a graph Γ compatible with the given tree ϑ is $|\Lambda| B^{k-1} |\lambda^{M_4} \overline{\mu}_{max}^{M_2} \varepsilon^{M_1}|$, with $k = M_4 + M_2 + M_1$, $\mu_{max} = \max_{\underline{x}} |\mu_N(\underline{x})|$ and B depending on B_0 and $\int e^{-\kappa|\underline{x}|} d^d \underline{x}$, times

$$\begin{aligned} &\prod_v 2^{h_v(d-2)\frac{1}{2}(4m_{4,v}+2m_{2,v}+m_{1,v}-n_v^e+\sum_{j=1}^{s_v} n_{v_j}^e)} 2^{-h_v d(s_v-1)} = \\ &= 2^{h_0((d-4)M_4 - 2M_2 - (d-\frac{d-2}{2})M_1 + d)} \cdot \\ &\cdot \prod_v 2^{(h_v - h_{v'})((d-4)M_{4,v} - 2M_{2,v} - (d-\frac{d-2}{2})M_{1,v} - n_v^e \frac{d-2}{2} + d)} \end{aligned} \tag{8.8}$$

where M_4, M_2, M_1 are the total number of graph elements of the first,second and third type in Fig.7, $M_{4,v}, M_{2,v}, M_{1,v}$ are the total number of graph elements of the first,second and third type in Fig.7 contained in the cluster v, and n^e_v are the total number of lines emerging from the cluster v.

Having obtained the above *dimensional estimates* we must sum over the graphs and trees, which essentially means summing over the scales h_v, and show that the sum converges as $N \to \infty$: this will be true but only for certain choices of $\overline{\mu}_N$, as dicussed in Sec.9 below. The case $d = 2$ is very simple (and in any event we have already shown the existence of the perturbation analysis so that discussing the bound (8.8) is not necessary): therefore we concentrate attention on the more interesting $d = 3$ case.

9. Determination of the counterterms and renormalizability ($d = 3$).

Since we consider contributions with $M_1 > 0$ ($M_1 = 0$ corresponds to "vacuum graphs" which do not contribute to the Schwinger functions) the exponent in the first factor is always < 0 so that the sum over h_0 (which has to be performed together with the sum over the other scales to take into account all graphs) converges; noting that $M_{1,v}$ has to be even (otherwise not all lines can be paired) the exponents in the product are also < 0 unless
(1) $n^e_v = 0$ which also corresponds to a vacuum graph
(2) $n^e_v = 2$ if $M_{2,v} = 1, M_{1,v} = M_{4,v} = 0$ which would give an exponent 0 but which cannot arise in a cluster
(3) $n^e_v = 4$ if $M_{4,v} = 1, M_{1,v} = M_{2,v} = 0$ which would give an exponent 0 but which cannot arise in a cluster
(4) $n^e_v = 2$ if $M_{4,v} = 1, M_{1,v} = M_{2,v} = 0$ which gives an exponent 1
(5) $n^e_v = 2$ if $M_{4,v} = 2, M_{1,v} = M_{2,v} = 0$ which gives an exponent 0.

Therefore we only have to study the cases (4) and (5) which correspond to graphs containing a subgraph like the ones in Fig.10 below.

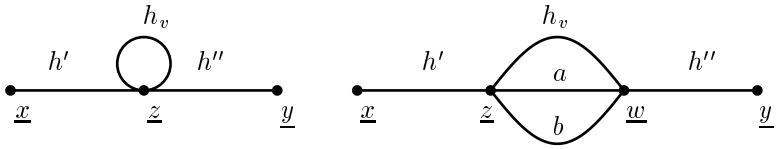

Fig.10: The two subgraphs for which the estimate above diverges, as $N \to \infty$, when summed over the scales: we suppose $h' \le h'' < h_v \le a, b$.

By our choice (8.5) of the interaction V_N the first graph in Fig.10 contains a self contraction and therefore does not arise (as commented above). Furthermore we can associate the second graphs with the subgraph

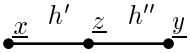

Fig.11: The subgraph whose contribution will be summed to the one of the second subgraph in Fig.10.

and their contribution to a graph value will be respectively

$$\begin{aligned} &\overline{\mu}\, C_{h'}(\underline{x} - \underline{z}) C_{h''}(\underline{z} - \underline{y}) \\ &6\,\lambda^2\, C_{h_v}(\underline{z} - \underline{w}) C_a(\underline{z} - \underline{w}) C_b(\underline{z} - \underline{w}) C_{h'}(\underline{x} - \underline{z}) C_{h''}(\underline{w} - \underline{y}) \end{aligned} \tag{9.1}$$

which, if summed together, exonerate us from considering Feynman graphs containing clusters like the second one in Fig.10.

It will be useful to rewrite the sum of the two terms in (9.1) as the sum of

$$\Big(\overline{\mu} + 6\,\lambda^2 \sum_{h_v > h'', a, b \geq h_v} \int d^d\,\underline{w}\; C_{h_v}(\underline{z} - \underline{w}) C_a(\underline{z} - \underline{w}) C_b(\underline{z} - \underline{w})\Big) \cdot \\ \cdot\, C_{h'}(\underline{x} - \underline{z})\, C_{h''}(\underline{z} - \underline{y}) \tag{9.2}$$

and of

$$6\,\lambda^2 \sum_{h_v > h'', a, b \geq h_v} \int d^d\; w\, C_{h_v}(\underline{z} - \underline{w}) C_a(\underline{z} - \underline{w}) C_b(\underline{z} - \underline{w}) \cdot \\ \cdot\, C_{h'}(\underline{x} - \underline{z}) \Big(C_{h''}(\underline{w} - \underline{y}) - C_{h''}(\underline{z} - \underline{y})\Big) \tag{9.3}$$

Therefore we can imagine to consider graphs without subgraphs like the ones in Fig.10 but which can instead contain subgraphs of the two forms

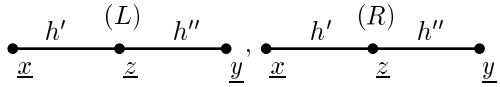

Fig.12: The two subgraphs into which the sum of the contributions from the second subgraph in Fig.9 and from the one in Fig.11 is decomposed.

which contribute to the value of a graph in which they appear, respectively, a factor given by (9.2) or by (9.3).

Note that in (9.3) the difference $\Delta \stackrel{def}{=} \big(C_{h''}(\underline{w} - \underline{y}) - C_{h''}(\underline{z} - \underline{y})\big)$ appears multiplied by the factor $C_{h_v}(\underline{z} - \underline{w})$: therefore the points $\underline{z}$, $\underline{w}$ can be considered to be at distance $O(2^{-h_v})$ typical of the scale of the lines linking $\underline{z}$, $\underline{w}$. However the covariance $C_{h''}$ "lives" on scale h'' so that the difference Δ can be bounded proportionally to $2^{-(h_v - h'')}$.

The factor (9.3) will yield in the bound (8.8) *an extra factor* $2^{-(h'' - h_v)}$ while if $\overline{\mu}_N(\lambda, \underline{x})$ is *defined* as

$$\overline{\mu}_N(\lambda, \underline{x}) = -\Big(6\,\lambda^2 \sum_{h=0, a, b \geq h}^{N} \int d^d\,\underline{w}\; C_h(\underline{z} - \underline{w}) C_a(\underline{z} - \underline{w}) C_b(\underline{z} - \underline{w})\Big) \tag{9.4}$$

then the first graph in Fig.12 will contribute to the bound (8.8) an extra factor

$$-\Big(6\,\lambda^2 \sum_{h=0, a, b \geq h}^{h''} \int d^d\,\underline{w}\; C_h(\underline{z} - \underline{w}) C_a(\underline{z} - \underline{w}) C_b(\underline{z} - \underline{w})\Big) \tag{9.5}$$

which is bounded proportionally to h'': *i.e.* in the bound corresponding to (8.8) there will be an extra power of $h_v^{m_{2,v}}$ which does not affect convergence of the sums over the scales.

Thus we see that formal perturbation theory can be well defined at each order so that the theory is "*renormalizable*", *i.e.* it admits a formal power series in λ, ε for the Schwinger functions generator $E_\lambda(\varepsilon f)$ with coefficients uniformly bounded as $N \to \infty$: the *harder problem* of showing that, with the choice (9.4) for $\overline{\mu}(\underline{x}, \lambda)$, the limit as $N \to \infty$ exists and the perturbation series is asymptotic to it is of course far more interesting: the above analysis is an essential tool for obtaining the result but new ideas need to be introduced, see references.

Note that the dependence of $\overline{\mu}$ on $\underline{x}$ is essential (it can be avoided only if periodic boundary conditions on Λ are adopted): a point that is often not mentioned in the literature.

The case $d = 4$ can only be studied in perturbation theory and often it is conjectured that there is no way to find a probability distribution over the fields $\varphi_{\underline{x}}$ which can be associated with the formal perturbation series (however the problem is wide open).

It becomes necessary, however, to allow also the coefficient of $\varphi^4_{\underline{x}}$ to depend on the cut–off N and to add another "counterterm" $\alpha_N : (\partial_{\underline{x}} \varphi_{\underline{x}})^2$:

$$V_N(\varphi_{\underline{x}}) = \lambda_N(\underline{x}) : \varphi^4_{\underline{x}} : + \mu_N(\underline{x}) : \varphi^2_{\underline{x}} : + \alpha_N(\underline{x}) : (\partial_{\underline{x}} \varphi_{\underline{x}})^2 : \tag{9.6}$$

and the question is whether one can find functions $\lambda_N, \mu_N, \alpha_N$ of a parameter λ such that the limit of $E_{\lambda,N}(\varepsilon f)$ as $N \to \infty$ exists and is smooth in λ, ε as well as *non trivial*, *i.e.* not quadratic in f. The problem remains physically interesting because the physical interpretation of the theory as a quantum field model would allow such an extension of the problem.

The method followed in dimension $d = 2, 3$ can also be applied to show renormalizability in the case $d = 4$ with, however, some rather major modifications that we cannot discuss here: see the references.

The reader should not be surprised that the analysis of the quantum fields models is apparently simpler than the one met in studying the KAM theory. The reason is simply that while in the KAM case we have presented a complete discussion in the case of quantum fields we only presented a complete solution to problem of the existence of a *formal perturbation series*. As already stressed another very important (and not easy) part of the work remains to be done and it is to show that the functions $E_{\lambda,N}(\varepsilon f)$ really have a limit as $N \to \infty$ and that the limit admits the formal pertirbation series as an asymptotic series: this problem is often called the *large fields problem* and it can be solved in dimension $d = 2, 3$ by a deeper use of the multiscale analysis, see the literature below.

Appendix A1: Siegel-Bryuno bound on the number of self-energy clusters.

Call $N_n^*(\vartheta)$ the number of non-self-energy lines carrying a scale label $\le n$ in a tree ϑ with k nodes. We shall prove first that $N_n^*(\vartheta) \le 2k(E2^{-\eta n})^{-1} - 1$ if $N_n(\vartheta) > 0$ (recall that $E = N^{-1}2^{-3\eta}$ and $\eta = 1/\tau$). We fix n and denote $N_n^*(\vartheta)$ as $N^*(\vartheta)$.

If ϑ has the root line λ_0 with scale $> n$ then calling $\vartheta_1, \vartheta_2, \dots, \vartheta_m$ the subtrees of ϑ emerging from the last node of ϑ and with $k_j > E\,2^{-\eta n}$ lines, one has $N^*(\vartheta) = N^*(\vartheta_1) + \dots + N^*(\vartheta_m)$ and the statement is inductively implied from its validity for $k' < k$ provided it is true that $N^*(\vartheta) = 0$ if $k < E2^{-\eta n}$, which is is certainly the case if E is chosen as in equation (5.3).[7]

In the other case, call $\lambda_1, \dots, \lambda_m$ the $m \ge 0$ lines on scale $\le n$ which are the nearest to λ_0:[8] such lines are the entering lines of a cluster T on scale $n_T > n$. If ϑ_i is the tree with λ_i as root line one has $N^*(\vartheta) \le 1 + \sum_{i=1}^m N^*(\vartheta_i)$, and if $m = 0$ the statement is trivial, while if $m \ge 2$ the statement is again inductively implied by its validity for $k' < k$.

If $m = 1$ we once more have a trivial case unless the order k_1 of ϑ_1 is $k_1 > k - E\,2^{-n\eta}/2$. Finally, and this is the real problem as the analysis of a few examples shows, we claim that in the latter case either the root line of ϑ_1 is a self-energy line or it cannot have scale $\le n$.

To see this, note that $|\underline{\omega}_0 \cdot \underline{\nu}(\lambda_0)| \le 2^n$ and $|\underline{\omega}_0 \cdot \underline{\nu}(\lambda_1)| \le 2^n$, hence $\delta \equiv |(\underline{\omega}_0 \cdot (\underline{\nu}(\lambda_0) - \underline{\nu}(\lambda_1))| \le 2^{n+1}$, and the Diophantine condition implies that either $|\underline{\nu}(\lambda_0) - \underline{\nu}(\lambda_1)| > 2^{-(n+1)\eta}$ or $\underline{\nu}(\lambda_0) = \underline{\nu}(\lambda_1)$. The latter case being discarded as $k - k_1 < E\,2^{-n\eta}/2$ (and we are not considering the self-energy clusters), it follows that $k - k_1 < E\,2^{-n\eta}/2$ is inconsistent: it would in fact imply that $\underline{\nu}(\lambda_0) - \underline{\nu}(\lambda_1))$ is a sum of $k - k_1$ node momenta and therefore $|\underline{\nu}(\lambda_0) - \underline{\nu}(\lambda_1)| < NE\,2^{-n\eta}/2$, hence $\delta > 2^3\,2^n$ which contradicts the above opposite inequality.

A similar, far easier, induction can be used to prove that if $N_n^*(\vartheta) > 0$ then the number $p_n(\vartheta)$ of clusters of scale n verifies the bound $p_n(\vartheta) \le 2k\,(E2^{-\eta n})^{-1} - 1$. Thus equation (5.3) is proved.

Remark. The above argument is a minor adaptation of Bryuno's proof of Siegel's theorem, as remarkably exposed by Pöschel.

Appendix A2: The KAM bound for graphs containing overlapping self energy graphs.

Let ϑ be a tree with height p: then each of its maximal self-energy clusters V contains a tree

[7] Note that if $k \le E\,2^{-n\eta}$ one has, for all momenta $\underline{\nu}$ of the lines, $|\underline{\nu}| \le NE\,2^{-n\eta}$, *i.e.* $|\underline{\omega}_0 \cdot \underline{\nu}| \ge (NE\,2^{-n\eta})^{-\tau} = 2^3\,2^n$ so that there are *no* clusters T with $n_T = n$ and $N^*(\vartheta) = 0$. The choice $E = N^{-1}2^{-3\eta}$ is convenient: but this, as well as the whole lemma, remains true if 3 is replaced by any number larger than 1. The choice of 3 is made only to simplify some of the arguments based on the self-energy cluster concept.

[8] *i.e.* such that no other lin along the paths connecting the lines $\ell_1, \dots, \ell_m$ to the root is on scale $\le n$.

of height $< p$. We imagine that all the resummations relative to the lines that enter the self-energy clusters that are not maximal have been performed so that we only have to consider the trees that are obtained by attaching the lines that enter the maximal self-energy clusters T to the nodes in $\widetilde{T}$.

Suppose for simplicity that there is only one maximal self-energy cluster T of height p. Then the sum of the values of the trees of the family $\mathcal{F}(\vartheta)$ obtained by shifting the entrance node into the self-energy clusters of lower height will have the form

$$\Big(\prod_{\lambda\in\Lambda(\vartheta)\setminus\Lambda_1(T)}\frac{\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}\cdot\underline{\nu}(\lambda))^2}\Big)\cdot \Big(\prod_{\lambda\in\Lambda(\widetilde{T})}\frac{\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}\cdot\underline{\nu}^0(\lambda)+\sigma_\lambda\eta_T)^2}\Big)\cdot\frac{1}{\eta_T^4}\cdot\Big(\prod_{T_i}F(T_i,\underline{\nu}_{v_i},\underline{\nu}_{v_i'})\Big), \tag{A2.1}$$

where α is 1 if the lines entering and exiting the cluster the last product is over all the maximal self-energy clusters T_i contained in T, v_i' and v_i are the nodes in $\widetilde{T}_i$ from which exits (or enters, respectively) the line that enters (or exits) the self-energy cluster T_i, and $F(T_i,\underline{\nu}_{v_i},\underline{\nu}_{v_i'})$ is the sum of the values of all the trees that we have to sum in shifting the entrance node of the lines that enter the self-energy clusters of lower order inside T_i. Note that the lines λ in the first product are the lines external to T (but neither entering or exiting T), while the ones in the second product are the lines internal to T (*i.e.* lines connecting to nodes in $\widetilde{T}$).

We can then remark that when η_T varies in the complex disk $|\eta_T|\le 2^{n_T}$ the divisors of the lines that enter the inner self-energy clusters T' (of any height) do not exceed in modulus $2^{n_{\lambda_{T'}}}$. Therefore we can bound the quantities $F(T_i,\underline{\nu}_{v_i},\underline{\nu}_{v_i'})$ via the inductive bound and obtain that (6.4) is valid also for trees of height p which contain only one maximal self-energy cluster.

The case in which there are many self-energy clusters of height p is reducible to the case in which there is only one such cluster, see references, the conclusion is the validity of the inequality (6.4) in general. It would also possible to give a proof of the inequality that is not based on an inductive argument but we leave it as a problem for the reader.

Acknowledgment. I am grateful to G.Gentile, A. Giuliani and V. Mastropietro for their critical comments. Support from IHES and INFN is also acknowledged.

Bibliographical notes

The KAM analysis is taken from

[Ga94] Gallavotti, G.: *Twistless KAM tori*, Communications in Mathematical Physics, **164**, 145–156, 1994.

and it is a reinterpretation of the original work of Eliasson

[El96] Eliasson, H.: *Absolutely convergent series expansions for quasi-periodic motions*, Mathematical Physics Electronic Journal, **2**, 1996, (preprint published in 1988).

In [Ga94] the cancellations are shown to occur by collecting suitable families of terms (each containing a rather small number of terms) appearing in the Lindstedt series for $\underline{h}$ while in [El96] the necessary cancellations are shown to occur if one sums all the Lindstedt terms as a consequence of a general symmetry argument. The extension of Brjuno's lemma used here is taken from

[Pö86] Pöschel, J.: *Invariant manifolds of complex analytic mappings*, Les Houches, XLIII (1984), Vol. II, p. 949– 964, Eds. K. Osterwalder & R. Stora, North Holland, 1986.

Closely related developments (and generalizations) can be found in

[GG95] Gallavotti, G., Gentile, G.: *Majorant series convergence for twistless KAM tori*, Ergodic theory and dynamical systems, **15**, p. 857–869, 1995. See also Gallavotti, G.: *Invariant tori: a*

field theoretic point of view on Eliasson's work, in *Advances in Dynamical Systems and Quantum Physics*, 117–132, Ed. R. Figari, World Scientific, 1995; and *Renormalization group in Statistical Mechanics and Mechanics: gauge symmetries and vanishing beta functions*, Physics Reports, **352**, 251–272, 2001; and Gentile, G., Mastropietro,V.: *KAM theorem revisited*, Physica D, **90**, 225–234, 1996. And *Methods of analysis of the Lindstedt series for KAM tori and renormalizability in classical mechanics. A review with some applications*, Reviews in Mathematical Physics, **8**, 393–444, 1996

The field theory analysis is based on Wilson's approach to renormalization and it is taken from

[BCGNPOS78] Benfatto, G., Cassandro, M., Gallavotti, G., Nicolò, F., Presutti, E., Olivieri, E., and Scacciatelli, E. *Some probabilistic techniques in field theory*, Communications in Mathematical Physics **59**, 143–166, 1978 and **71**, 95–130, 1980.

For the "exact renormalization group" analysis in the case $d = 4$ see

[Pl84] Polchinski, J.: *Renormalization group and effective lagrangians*, Nuclear Physics **B231**, 269–295, 1984.
[GN85] Gallavotti, G., and Nicolò, F. *Renormalization theory for four dimensional scalar fields*, Communications in Mathematical Physics **100**, 545–590, 1985 and **101**, 1–36, 1985.

The two (independent) approaches differ only because in [Pl84] a continuous variation of the cut–off is considered while in [GN85] the variation takes place discretely: one obtains a continuous flow of the effective potential in the first case described by differential equations and a discrete evolution descibed by a map, respectively.

For reviews on the approach followed here see

[Ga85] Gallavotti, G.: *Renormalization theory and ultraviolet stability via renormalization group methods*, Reviews of Modern Physics **57**, 471–569, 1985. And Gallavotti, G., *The structure of renormalization theory: renormalization, form factors, and resummations in scalar field theory*, ed. K. Osterwalder, R. Stora, Les Houches, XLIII, 1984, Phenomenes critiques, Systemes aleatoires, Theories de jauge, Elseviers Science, 1986, p. 467-492. *The beta function method for resummations in field theory*, Constructive quantum field theory, II, Scuola di Erice, 1988, ed. G. Velo, A. Wightman, NATO ASI series B, vol. **234**, 69–88, 1990

For a general review on the exact renormalization group see

[BG95] Benfatto, G., Gallavotti, G.: *Renormalization group*, p. 1–144, Princeton University Press, 1995.