

CHAPTER VIII

Special ergodic theory problems in nonchaotic dynamics

§8.1 Theory of quasi-periodic Hamiltonian motions

A very natural and important question that one can ask about stability in Hamiltonian systems is what becomes of the simple foliation of phase space into invariant tori when a perturbing force is switched on.

The simplest case is when the *unperturbed Hamiltonian* has the form

$$\varepsilon_{8.1.1} \quad \mathcal{H}_0(\underline{A}, \underline{\alpha}) \stackrel{\text{def}}{=} K(\underline{A}) = \frac{1}{2J} \underline{A}^2, \quad \begin{cases} \underline{A} = (A_1, \dots, A_\ell) \in \mathbb{R}^\ell, \\ \underline{\alpha} = (\alpha_1, \dots, \alpha_\ell) \in \mathbb{T}^\ell, \end{cases} \quad (8.1.1)$$

where $J > 0$ and $(\underline{A}, \underline{\alpha}) \in \mathbb{R}^\ell \times \mathbb{T}^\ell$ are ℓ canonical momenta and their ℓ conjugate angles. This is the Hamiltonian representing the motion of ℓ point masses (*rotators*) rotating on ℓ unit circles with moment of inertia J , angular momenta A_j and positions α_j : $K(\underline{A})$ is their kinetic energy.

The motion trivially takes place on ℓ -dimensional invariant tori parameterized by the momenta \underline{A} : calling $\underline{\omega} = (\omega_1, \dots, \omega_\ell) = J^{-1}\underline{A}$, the point $(\underline{A}, \underline{\alpha})$ evolves in time t into $S_t(\underline{A}, \underline{\alpha}) = (\underline{A}, \underline{\alpha} + \underline{\omega}t)$.

If $V(\underline{\alpha})$ is an analytic function (*perturbing potential*) one might expect that the perturbed system described by the Hamiltonian function $K(\underline{A}) + \varepsilon V(\underline{\alpha})$ still has, at least for small ε , quasi-periodic motions developing on tori which foliate the entire phase space and are close to the unperturbed ones in the sense that their parametric equations have the form

$$\varepsilon_{8.1.2} \quad \begin{cases} \underline{\alpha} = \underline{\psi} + \underline{h}_\varepsilon(\underline{B}, \underline{\psi}), \\ \underline{A} = \underline{B} + \underline{H}_\varepsilon(\underline{B}, \underline{\psi}), \end{cases} \quad \underline{\psi} \in \mathbb{T}^\ell, \quad (8.1.2)$$

where $\underline{H}_\varepsilon, \underline{h}_\varepsilon$ are 2ℓ analytic functions of ε small as $\varepsilon \rightarrow 0$ and \underline{B} are ℓ parameters that are introduced to parameterize the various tori. On the tori the motion should be $(\underline{A}, \underline{\alpha}) \rightarrow S_t^\varepsilon(\underline{A}, \underline{\alpha}) = (\underline{A}(t), \underline{\alpha}(t))$ with $(\underline{A}(t), \underline{\alpha}(t))$ obtained by replacing $\underline{\psi}$ with $\underline{\psi} + \underline{\omega}'t$ in equation (8.1.2), *i.e.* the motion should be given by

$$e8.1.3 \quad \begin{cases} \underline{\alpha}(t) = \underline{\alpha} + \underline{\omega}'t + \underline{h}_\varepsilon(\underline{B}, \underline{\alpha} + \underline{\omega}'t), \\ \underline{A}(t) = \underline{B} + \underline{H}_\varepsilon(\underline{B}, \underline{\alpha} + \underline{\omega}'t), \end{cases} \quad (8.1.3)$$

with \underline{B} determined by the condition that the initial data \underline{A} should be contained in the torus, $\underline{A} = \underline{B} + \underline{H}_\varepsilon(\underline{B}, \underline{\alpha})$. Moreover the velocities $\underline{\omega}'$ (*rotation vectors*) should be suitable functions of \underline{B} , *i.e.* of the torus containing the initial data $(\underline{A}, \underline{\alpha})$.

In fact the parameters \underline{B} could be defined quite arbitrarily (by changing correspondingly the definition of the function $\underline{H}_\varepsilon(\underline{B}, \underline{\alpha})$). However in the unperturbed case there is a direct proportionality between velocities and angular momenta so that a natural choice of the parameters \underline{B} could be simply $\underline{B} = J \underline{\omega}'$. This would mean choosing to parameterize the perturbed invariant tori by the rotation vector of the quasi-periodic motion on them.

Another natural choice could be that of determining the invariant tori by their *average action*:

$$e8.1.4 \quad \underline{B} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \underline{A}(t) dt; \quad (8.1.4)$$

in this case $\underline{\omega}'$ would be a suitable function of \underline{B} , not necessarily equal to $J^{-1}\underline{B}$. However as \underline{B} varies the velocities $\underline{\omega}'$ should assume all possible values, because in the unperturbed case they take all possible values as functions of the angular momenta. In other words the perturbed motion would exhibit all unperturbed motions with the only difference that they take place on slightly deformed trajectories.

The above *naive* picture representing the perturbed evolution as taking place on a phase space smoothly foliated into invariant tori on which a peaceful quasi-periodic motion takes place is clearly impossible as the simple case $\ell = 2$, $V(\underline{\alpha}) = \cos(\alpha_1 - \alpha_2)$ shows. In this case the perturbed motions with velocities $\omega_1 = \omega_2 = \omega$ do not form a family of closed trajectories filling up a 2-dimensional torus, as equation (8.1.2) would imply. It is easy to check that while the unperturbed motions with $\omega_1 = \omega_2 = \omega$ consist in a *continuum family of periodic motions* covering a torus the perturbed motions *only contain two* such motions which are close to the unperturbed ones; see problem [8.1.1].

In other words the perturbation “breaks” the invariant torus with $\omega_1 = \omega_2 = \omega$ and the only trace left of it are two isolated periodic orbits and no motion near them is periodic with period $2\pi/\omega$. All this can be easily checked because the perturbed system is still integrable in the sense of Hamiltonian mechanics in most of phase space and its theory reduces essentially to that of the classical pendulum.

It is very interesting that the above example in a way describes “all that can go wrong” and points out that essentially the only exceptions occur at *resonances*, *i.e.* the only motions that may be “absent” among the perturbed motions and which were present among the unperturbed ones are motions with “resonant frequencies” or with frequencies “too close” to resonant ones. A resonance is defined as a quasi-periodic motion with rotation vector $\underline{\omega} \in \mathbb{R}^\ell$ such that

$$e8.1.5 \quad \underline{\omega} \cdot \underline{\nu} = 0 \quad \text{for some } \underline{\nu} = (\nu_1, \dots, \nu_\ell) \in \mathbb{Z}^\ell, \quad \underline{\nu} \neq \underline{0}. \quad (8.1.5)$$

A rotation vector $\underline{\omega}$ can be “far” from resonant in various senses. Here we shall consider, for simplicity, one of the strongest senses in which “far” can be understood and select the following vectors.

D8.1.1 **(8.1.1) Definition:** (Diophantine vectors)

A vector $\underline{\omega} \in \mathbb{R}^\ell$ such that there exist two constants $C, \tau > 0$ for which

$$e8.1.6 \quad C|\underline{\omega} \cdot \underline{\nu}| > |\underline{\nu}|^{-\tau}, \quad \text{for all } \underline{\nu} \in \mathbb{Z}^\ell, \quad \underline{\nu} \neq \underline{0}, \quad (8.1.6)$$

where $|\underline{\nu}| \stackrel{\text{def}}{=} \sum_{i=1}^\ell |\nu_i|$, is called a Diophantine vector with constant C and exponent τ .

Remark: It can be shown, see problem [8.1.7], that if $\tau > \ell - 1$ almost all $\underline{\omega} \in \mathbb{R}^\ell$ verify (8.1.6) for some suitable C (depending on $\underline{\omega}$), while in any ball of radius r in \mathbb{R}^ℓ the set of $\underline{\omega}$'s which do not satisfy (8.1.6) for given C and τ has volume $< B_{\tau, \ell} C^{-1} r^{\ell-1}$, for a suitable $B_{\tau, \ell} < \infty$.

In other words “most vectors $\underline{\omega} \in \mathbb{R}^\ell$ are far from resonances and satisfy a Diophantine condition” (hence they are rationally independent and enjoy the properties discussed in Section §2.2).

The key result is the following one.

P8.1.1 **(8.1.1) Proposition:** (KAM theorem)

Let $V(\underline{\alpha})$ be even and analytic, let $K(\underline{A}) = \underline{A}^2/2J$, $J > 0$, and let $\underline{\omega}_0 = \underline{A}_0/J$ be a Diophantine vector with constant C and exponent τ . Then there exists $\varepsilon_0 > 0$ depending on C and τ and two functions, denoted $\underline{H}(\underline{\psi}, \varepsilon), \underline{h}(\underline{\psi}, \varepsilon)$, analytic for $(\varepsilon, \underline{\psi}) \in (-\varepsilon_0, \varepsilon_0) \times \mathbb{T}^\ell$ and divisible by ε , such that the points $(\underline{A}, \underline{\alpha})$ that lie on the tori

$$e8.1.7 \quad \begin{cases} \underline{\alpha} = \underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon), \\ \underline{A} = \underline{A}_0 + \underline{H}(\underline{\psi}, \varepsilon), \end{cases} \quad \underline{\psi} \in \mathbb{T}^\ell, \quad (8.1.7)$$

form invariant surfaces for the evolution generated by the Hamiltonian

$$e8.1.8 \quad H(\underline{A}, \underline{\alpha}) = K(\underline{A}) + \varepsilon V(\underline{\alpha}), \quad (8.1.8)$$

and their motion is simply $\underline{\psi} \rightarrow \underline{\psi} + \underline{\omega}_0 t$. Furthermore the time average of $\underline{A}(t) = \underline{A}_0 + \underline{H}(\underline{\psi} + \underline{\omega}_0 t, \varepsilon)$ is precisely \underline{A}_0 .

Remarks: (1) This means that the Diophantine tori of the unperturbed system¹ are deformed into Diophantine tori of the perturbed system with the same rotation vector and with the same average action.

N8.1.1

(2) The coincidence between the average action and the rotation vector times the moment of inertia J is a special property of Hamiltonians of the form (8.1.8). This property has been qualified with the adjective *twistless* in [Ga94a]. The interest of analyzing the KAM theorem for the special models in (8.1.8) has been remarked by Thirring, see p. 133 in [Th83], and for this reason they have been called *Thirring models*.

(3) The parity assumption on $V(\underline{\alpha})$, *i.e.*

$$e8.1.9 \quad V(\underline{\alpha}) = \sum_{\underline{\nu} \in \mathbb{Z}^{\ell}} f_{\underline{\nu}} e^{i\underline{\nu} \cdot \underline{\alpha}}, \quad \text{with } f_{\underline{\nu}} = f_{-\underline{\nu}}, \quad (8.1.9)$$

is a simplicity assumption that can be released as discussed in the problems for Section §(8.4).

(4) The proof presented here is elementary and it is based on the proof of Eliasson, [El96]. It is an interpretation of [El96] and provides a detailed analysis of the “cancellations” following [Ga94a].

(5) The proof will be discussed under slightly stronger assumptions in order to clarify its conceptual structure: the two extra assumptions that will be made below can be released as discussed in Section (8.2). The first assumption will be that $V(\underline{\alpha})$ is a *trigonometric polynomial*, *i.e.* the sum in (8.1.9) is, for some $N < \infty$, restricted to

$$e8.1.10 \quad |\underline{\nu}| \leq N < \infty. \quad (8.1.10)$$

The second assumption will be that the rotation vector $\underline{\omega}_0$ satisfies a property slightly stronger than the Diophantine inequality (8.1.7) that we shall introduce later (see definition (8.3.2)).

The proof will be split into several parts:

(i) We first derive the expression for the Taylor coefficients of the function \underline{h} in powers of ε .

Calling $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$, $k \geq 1$, the k -th order coefficients of the Taylor expansion of $\underline{H}(\underline{\psi}, \varepsilon), \underline{h}(\underline{\psi}, \varepsilon)$ in powers of ε we derive recursive expressions for such coefficients supposing that $\underline{h}, \underline{H}$ exist, are analytic at $\varepsilon = 0$ and have zero average over $\underline{\psi}$: we show that they are uniquely determined. In

N8.1.2

this way we obtain a formal power series² and we have to show its convergence to prove the proposition. Existence of the coefficients is a result due to Lindstedt and Newcomb in special cases and to Poincaré, [Po93] Ch. IX, in our case as well as in more general cases. The formal series is sometimes

¹ *i.e.* the tori on which run quasi-periodic motions with Diophantine rotation vectors.

² A formal power series of a field F is an infinite sequence $\{a_k\}_{k=1}^{\infty}$ over F ; it is often written as $a_1\varepsilon + a_2\varepsilon^2 + a_3\varepsilon^3 + \dots$, but with the understanding that no value is assigned to the symbol ε .

called the *Lindstedt series* although the same name is also given to a particular construction of the coefficients $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$ as sums of suitably defined “values” (cf. Section §(8.2)).

(ii) Subsequently, in Section §(8.2), we show that the coefficients of the Lindstedt series can be represented as sums of numerical values assigned to certain diagrams or graphs. Furthermore we check that the difficulty in proving convergence of the formal power series is only due to certain classes of diagrams whose members can give contributions to the coefficients of order k whose size could attain the order of a power of $k!$. This of course causes a problem, known as the *small divisors problem*.

(iii) In the only part which requires some work, Section §(8.3), we identify the diagrams which give too large contributions to the coefficients of order k and the way in which they can be collected together so that cancellations between their values become manifest. Finally we check that their sums is indeed far smaller than feared *a priori* so that the final bound on the coefficients of order k will be only of the order of an exponential in k , Section §(8.4), thus achieving the proof of proposition (8.1.1).

The first step is to show that the formal series for $\underline{h}, \underline{H}$ in powers of ε can be constructed and are well defined to all orders in ε .

L8.1.1 **(8.1.1) Lemma:** (Definition of the Lindstedt series)

There exists at most one solution $\underline{h}(\underline{\psi}), \underline{H}(\underline{\psi})$ of (8.1.7) which is analytic in ε near $\varepsilon = 0$, divisible by ε and with zero average over $\underline{\psi}$. The Taylor coefficients at $\varepsilon = 0$ of order $k \geq 1$ of such a solution are uniquely determined by (8.1.7). The resulting formal power series in ε will be called the Lindstedt series.

Proof: We suppose that $V(\underline{\alpha})$ is a *trigonometric polynomial* of some degree N : this means that the sum in (8.1.9) is restricted by (8.1.10).

Call $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi}), k \geq 1$, the k -th order coefficients of the Taylor expansion of $\underline{H}(\underline{\psi}, \varepsilon), \underline{h}(\underline{\psi}, \varepsilon)$ in powers of ε . By imposing that (8.1.7) with $\underline{\psi} = \underline{\omega}_0 t$ solve the equations of motion by substitution into the equations of motion $\dot{\underline{\alpha}} = J^{-1} \underline{A}, \dot{\underline{A}} = -\varepsilon \partial_{\underline{\alpha}} f(\underline{\alpha})$ we find the necessary (and sufficient) condition

$$e8.1.11 \quad \begin{cases} (\underline{\omega}_0 \cdot \partial_{\underline{\psi}}) \underline{h}(\underline{\psi}, \varepsilon) = J^{-1} \underline{H}(\underline{\psi}, \varepsilon), \\ (\underline{\omega}_0 \cdot \partial_{\underline{\psi}}) \underline{H}(\underline{\psi}, \varepsilon) = -\varepsilon \partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)), \end{cases} \quad (8.1.11)$$

where $\partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)) = \partial_{\underline{\alpha}} V(\underline{\alpha})|_{\underline{\alpha}=\underline{\psi}+\underline{h}(\underline{\psi}, \varepsilon)}$. By equating the coefficients of the expansion of both sides in powers of ε we get immediately recursion relations for $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$.

In fact the second set of equations in (8.1.11) shows that the only unknown is $\underline{h}(\underline{\psi}, \varepsilon)$ and that it is a solution of the “Hamilton–Jacobi equation”:

$$e8.1.12 \quad (\underline{\omega}_0 \cdot \partial_{\underline{\psi}})^2 \underline{h}(\underline{\psi}, \varepsilon) = -\varepsilon J^{-1} \partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)), \quad (8.1.12)$$

because $\underline{H}(\underline{\psi}, \varepsilon)$ is immediately determined from the first equation in the

second set in (8.1.11). Clearly $\underline{h}(\underline{\psi}, \varepsilon)$ is determined up to a constant which we choose to fix by requiring that $\underline{h}(\underline{\psi}, \varepsilon)$ has zero average over $\underline{\psi}$.³

N8.1.3

Writing $\underline{h}(\underline{\psi}, \varepsilon) = \varepsilon \underline{h}^{(1)}(\underline{\psi}) + \varepsilon^2 \underline{h}^{(2)}(\underline{\psi}) + \dots$ and substituting into (8.1.12) we find for $k = 1$

$$e8.1.13 \quad \underline{h}^{(1)}(\underline{\psi}) = - \sum_{\underline{\nu} \neq \underline{0}} \frac{iJ^{-1}\underline{\nu}}{(i\underline{\omega}_0 \cdot \underline{\nu})^2} f_{\underline{\nu}} e^{i\underline{\nu} \cdot \underline{\psi}}. \quad (8.1.13)$$

And more generally one finds the k -th order from (8.1.11) simply by expanding in powers of $\underline{h}(\underline{\psi}, \varepsilon)$ the function $\partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon))$ and, subsequently, by expanding $\underline{h}(\underline{\psi}, \varepsilon)$ in powers of ε . The result is, for $k > 1$,

$$e8.1.14 \quad \begin{aligned} & (\underline{\omega}_0 \cdot \underline{\partial}_{\underline{\psi}})^2 \underline{h}^{(k)}(\underline{\psi}) = \\ & = -J^{-1} \sum_{m=1}^{\infty} \frac{1}{m!} \partial_{\underline{\alpha}}^{m+1} V(\underline{\psi}) \sum_{\substack{k_1, \dots, k_m \geq 1 \\ k_1 + \dots + k_m = k-1}} \prod_{r=1}^m \underline{h}^{(k_r)}(\underline{\psi}). \end{aligned} \quad (8.1.14)$$

Here we made an effort in trying to reduce the quantity of indices: therefore $\partial_{\underline{\alpha}}^{m+1}$ means a tensor with $m+1$ indices j_0, j_1, \dots, j_ℓ in $\{1, 2, \dots, \ell\}$. If we imagine to use (8.1.14) to evaluate the component j_0 of the vector $\underline{h}^{(k)}(\underline{\psi})$ and set $\partial_j \equiv \partial_{\alpha_j}$, then $\partial_{\underline{\alpha}}^{1+m} V(\underline{\psi})$ denotes the tensor $\partial_{j_0} \partial_{j_1} \dots \partial_{j_m} V(\underline{\alpha})$ evaluated at $\underline{\alpha} = \underline{\psi}$; the indices j_1, \dots, j_m of the latter tensor must be contracted with the indices j_1, \dots, j_m determining the components of the vectors $\underline{h}^{(k_r)}(\underline{\psi})$, *i.e.* if $\partial_{j_1, \dots, j_m} \stackrel{def}{=} \partial_{j_1} \dots \partial_{j_m}$ and $k_r \geq 1$

$$e8.1.15 \quad \begin{aligned} & (\partial_{\underline{\alpha}}^{m+1} V(\underline{\psi}) \prod_{r=1}^m \underline{h}^{(k_r)}(\underline{\psi}))_{j_0} = \\ & = \sum_{j_1, \dots, j_m=1}^{\ell} \sum_{k_1 + \dots + k_m = k-1} \partial_{j_0 j_1 \dots j_m} V(\underline{\psi}) \prod_{r=1}^m h_{j_r}^{(k_r)}(\underline{\psi}), \end{aligned} \quad (8.1.15)$$

while naturally the first index j_0 remains uncontracted as it represents the component of $\underline{h}^{(k)}(\underline{\psi})$ that we compute.

The equation (8.1.14) can be solved by Fourier transform: calling $\underline{\Phi}^{(k)}(\underline{\psi})$ the r.h.s. we see that the Fourier transforms of $\underline{h}^{(k)}$ and $\underline{\Phi}^{(k)}$ are related by

$$e8.1.16 \quad -(\underline{\omega}_0 \cdot \underline{\nu})^2 \underline{h}_{\underline{\nu}}^{(k)} = \underline{\Phi}_{\underline{\nu}}^{(k)}, \quad \underline{\nu} \in \mathbb{Z}^\ell, \quad (8.1.16)$$

which can be solved if (and only if) $\underline{\Phi}_{\underline{0}}^{(k)} = \underline{0}$, *i.e.* if $\underline{\Phi}^{(k)}(\underline{\psi})$ has zero average over $\underline{\psi}$. The latter property holds for $k = 1$ because in this case $\underline{\Phi}^{(1)} \equiv \underline{\partial}_{\underline{\alpha}} V$

³ In fact the equation for $\underline{h}(\underline{\psi}, \varepsilon)$ in (8.1.11) can be solved by Fourier transform only if $\underline{H}(\underline{\psi}, \varepsilon)$ has zero average over $\underline{\psi}$ and, if it can be solved, it is then determined up to an additive constant equal to its average over $\underline{\psi}$.

is a gradient (cf. (8.1.13)). In general suppose that $\underline{h}^{(k)}(\underline{\psi})$ is trigonometric polynomial in $\underline{\psi}$ of degree $\leq kN$, odd in $\underline{\psi}$, for $1 \leq k < k_0 - 1$. Then we see immediately that the r.h.s. of (8.1.14), with $k = k_0$, is odd in $\underline{\psi}$ and of degree $\leq k_0N$. Therefore the r.h.s. of (8.1.14) has zero average in $\underline{\psi}$, and (8.1.14) can be also solved for $k = k_0$. This means that the equation for $\underline{h}^{(k_0)}(\underline{\psi})$ can be solved for all $k_0 \geq 1$, and its solution is a trigonometric polynomial of degree $\leq k_0N$ which is odd in $\underline{\psi}$, if $\underline{h}^{(k_0)}(\underline{\psi})$ is uniquely determined by imposing that its average over $\underline{\psi}$ vanishes. ■

Problems for §8.1

Q8.1.1 [8.1.1]: (*A resonance phenomenon*)
 Show that for any $\omega \in \mathbb{R}$ the two-dimensional system described by the Hamiltonian $K(\underline{A}) + \varepsilon V(\underline{\alpha})$, with $K(\underline{A}) = \frac{1}{2}\underline{A}^2, V(\alpha) = \cos(\alpha_1 - \alpha_2)$, admits only two periodic motions with rotation vector (ω, ω) corresponding to the stationary points of a suitable function (this is an example of a *resonance*). (*Hint*: Use the (canonical) change of variable $\beta_1 = \alpha_1 + \alpha_2, \beta_2 = \alpha_1 - \alpha_2$ and similar for \underline{A} . Note that this is a simple example of the general set-up of the forthcoming Section §(9.1) about low-dimensional tori, of which the periodic orbits are an (easy) subcase.)

Q8.1.2 [8.1.2]: (*Rationally independent 3-vectors*, [Ch99])
 Suppose that the two equations $n^3 = an \pm b$, with a, b integers, do not admit integer solutions, and let ω be real and a root of the equation $\omega^3 = a\omega + b$. Show that the vector $\underline{w} = (1, \omega, \omega^2)$ has rationally independent components. The case $a = b = 1$ defines a real root ω which is called the *spiral mean*. (*Hint*: If not $\underline{v} \cdot \underline{w} = \nu_1 + \nu_2\omega + \nu_3\omega^2 = 0$ with $\underline{0} \neq \underline{v} \in \mathbb{Z}^3$ and ω would be a quadratic number, i.e. a number of the form $\omega = (x + \sqrt{y})$ with x, y rational. Since ω is also a solution of the third order equation \sqrt{y} must be rational: in fact if \sqrt{y} is irrational then the equation $\omega^3 = a\omega + b$ implies

$$x^3 + 3xy - ax - b + \sqrt{y}(y + 3x^2 - a) = 0,$$

and necessarily $y + 3x^2 - a = 0$ and $x^3 + 3xy - ax - b = 0$, which means that $8x^3 - 2ax + b = 0$ admits a rational root $x = p/q$, with p, q relatively prime, i.e. $8p^3 - 2apq^2 + bq^3 = 0$. This says that $(2p)^3$ is divisible by q^2 so that $q = 1$ or $q = 2$ and $n = \frac{2p}{q}$ is integer: therefore $n^3 = an - b$ admits an integer solution, in contradiction with the hypothesis. Hence ω has to be rational. If $\omega = p/q, p, q$ relatively prime integers, then $p^3 = apq^2 + q^3$ and q^2 divides p^3 , hence $q = 1$ and p is an integer solution of $n^3 = an + b$, which again contradicts the hypothesis.)

Q8.1.3 [8.1.3]: (*Tartaglia's formula and rational independence*)
 Let $3p, 2q$ be integers with $q^2 + p^3 > 0$ and suppose that the equations $z^3 = 3pz \pm 2q$ do not admit integer solutions. Show that $\underline{w} = (1, \omega, \omega^2)$ is a rationally independent vector if $\omega = (q + (p^3 + q^2)^{\frac{1}{2}})^{\frac{1}{3}} + (q - (p^3 + q^2)^{\frac{1}{2}})^{\frac{1}{3}}$. Show that, therefore, $\underline{w} = (1, 2^{\frac{1}{3}}, 4^{\frac{1}{3}})$ is a rationally independent vector. Check that the spiral mean is

$$\omega = \sqrt[3]{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{23}{27}}} + \sqrt[3]{\frac{1}{2} - \frac{1}{2}\sqrt{\frac{23}{27}}}.$$

(*Hint*: By Tartaglia's formula, divulged by Cardano, ω is a real root of $\omega^3 = -3p\omega + 2q$: hence one applies problem [8.1.2]. Note that $2^{1/3}$ is a root of $\omega^3 = 2$.)

Q8.1.4 [8.1.4]: (*An example of a Diophantine 3-vector*, [Ch99])
 To obtain an extension of problem [2.2.3] let

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad N^k \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} pk \\ qk \\ rk \end{pmatrix} \stackrel{def}{=} \underline{v}_k$$

Show that $\rho_j = p_j, q_j, r_j$ verify the recursion $\rho_j = \rho_{j-2} + \rho_{j-3}$, $j \geq 3$. Let $\lambda > 1$, λ_+, λ_- be the three eigenvalues of N , $|\lambda_+| = |\lambda_-| = \lambda^{-1/2}$ and let $\underline{w}_0, \underline{w}_+, \underline{w}_- = (1, \lambda_j, \lambda_j^2)$, $j = 0, \pm$ be the respective eigenvectors (note that $\lambda_j^3 = \lambda_j + 1$). Let N^T be the transposed of N . Given $\underline{v} \in \mathbb{Z}^3$ define, if it exists, k so that for $0 \leq h \leq k$

$$|(N^T)^h \underline{v} \cdot \underline{w}_0| \leq \max_{\pm} |(N^T)^h \underline{v} \cdot \underline{w}_{\pm}|, \quad |(N^T)^{(k+1)} \underline{v} \cdot \underline{w}_0| > \max_{\pm} |(N^T)^{(k+1)} \underline{v} \cdot \underline{w}_{\pm}|,$$

otherwise set $k = -1$. Check that $|\underline{v}| > B|\lambda_{\pm}|^{-k} = B\lambda^{\frac{k}{2}}$ for some $B > 0$ (which only depends on the matrix N); deduce that therefore \underline{w}_0 is a Diophantine vector with exponent $\tau = 2$. (*Hint*: The recursion can be derived from the remark that $N^3 \underline{v}_0 = \underline{v}_0 + N \underline{v}_0$. To check the Diophantine property remark that $1 \leq |(N^T)^k \underline{v}|$ because N has no zero eigenvalue; hence by the definition of k and for a suitable $b > 0$ depending on the basis $\underline{w}_0, \underline{w}_{\pm}$, one has $1 \leq |(N^T)^k \underline{v}| \leq b \max_{\pm} |\underline{w}_{\pm} \cdot (N^T)^k \underline{v}| \leq b' |\underline{v}| \lambda^{-\frac{k}{2}} \max_{\pm} |\underline{w}_{\pm}|$ (in fact in \mathbb{R}^3 all metrics are equivalent). Therefore, by the definition of k , one has $|\underline{w}_0 \cdot \underline{v}| = \lambda^{-(k+1)} |\underline{w}_0 \cdot (N^T)^{(k+1)} \underline{v}| \geq b'' \lambda^{-(k+1)} |(N^T)^{(k+1)} \underline{v}| \geq b''' \lambda^{-(k+1)} \geq b'''' |\underline{v}|^{-2}$.)

Q8.1.5 [8.1.5]: (*Hyperbolicity, Anosov maps of \mathbb{T}^3 and quasi periodicity*)

Show that the matrix T in problem [4.3.6] yields a natural example of an algebraic hyperbolic map of the torus \mathbb{T}^3 and of a Diophantine 3 vector.

Q8.1.6 [8.1.6]: (*More examples of Diophantine 3-vectors, [Ch99]: an extension of problem [2.2.3]*)

Show that the example of problem [8.1.4] can be extended by considering 3×3 integer entries matrices N with only one eigenvalue with modulus greater than 1 and two others with modulus less than 1 and with eigenvectors with rationally independent components. Find a few examples by studying matrices whose characteristic equation has roots that can be discussed by the techniques of problems [8.1.2] and [8.1.3]. The vectors constructed in this way are all in a class called *Pisot–Vijayaraghavan Diophantine 3-vectors*. (*Hint*: For instance consider matrices obtained from N in problem [8.1.4] by replacing the last row $(0, 1, 0)$ by $(0, p, q)$.)

Q8.1.7 [8.1.7]: (*Genericity of Diophantine vectors*)

Given C and $\tau > \ell - 1$ show that the volume of the points $\underline{w} \in \mathbb{R}^{\ell}$ with $|\underline{w}| < r$ that do not verify property (8.1.6) is bounded by $B_{\tau} C^{-1} r^{\ell-1}$, for a suitable constant B_{τ} depending on τ . Show that this implies that the set of \underline{w} which do not satisfy a Diophantine property with a prefixed exponent $\tau > \ell - 1$ and some constant $C < \infty$ has complement of zero volume in \mathbb{R}^{ℓ} . (*Hint*: Consider the slab $|\underline{w}_0 \cdot \underline{v}| < C^{-1} |\underline{v}|^{\tau}$, $\underline{v} \neq \underline{0}$, in the sphere $|\underline{w}_0| < r$. Note that its volume is bounded proportionally to $C^{-1} |\underline{v}|^{-\tau-1} r^{\ell-1}$ hence the set of points *not* verifying (8.1.6) has volume bounded proportionally to $C^{-1} r^{\ell-1}$. The theorem of Borel–Cantelli, for instance, then implies the second statement.)

Q8.1.8 [8.1.8]: (*Examples of isochronous systems, [CF02]*)

Consider in \mathbb{C}^n the equation $\dot{\underline{z}} = \Lambda \underline{z} + \underline{Q}(\underline{z})$, where Λ is a diagonal matrix with diagonal $\lambda_1, \dots, \lambda_n$ and \underline{Q} is a polynomial with a second order 0 at the origin. Suppose $\lambda_j \equiv i$. Prove that all solutions with initial data \underline{z}_0 small enough are periodic with period 2π .

(*Hint*: Define $\tau = (e^{it} - 1)/i$ and set $\underline{\zeta}(\tau) \stackrel{def}{=} e^{-it} \underline{z}(t)$: then $\underline{\zeta}(\tau)$ satisfies the equation $\frac{d\underline{\zeta}(\tau)}{d\tau} = (1 + i\tau)^{-2} \underline{Q}((1 + i\tau) \underline{\zeta}(\tau))$. The latter equation has solutions holomorphic in τ for $|\tau| < R$ provided the initial data are small enough. Therefore choosing $R > 2$ we see that the solution is $z(t) = e^{it} \zeta((e^{it} - 1)/i)$ which is periodic in t .)

Q8.1.9 [8.1.9]: Find the condition under which the equation for $\underline{\zeta}$ in problem [8.1.8] is autonomous.

Q8.1.10 [8.1.10]: (*Perturbation expansion without divisor problems and isochrony*)

Derive the result of problem [8.1.8] by a perturbation expansion showing the existence of a function $\underline{h}(\underline{\zeta})$ holomorphic near $\underline{0}$ and vanishing to second order such that the change

of variables $\underline{z} = \underline{\zeta} + \underline{h}(\underline{\zeta})$ transforms the differential equation into $\dot{\underline{z}} = \underline{\Lambda}\underline{z}$. (*Hint*: No small divisors problems arise in the perturbative determination of \underline{h} .)

Q8.1.11

[8.1.11]: Consider in \mathbb{C}^n the equation $\dot{\underline{z}} = \underline{\Lambda}\underline{z} + \underline{Q}(\underline{z})$, where $\underline{\Lambda}$ is a diagonal matrix with diagonal $\lambda_1, \dots, \lambda_n$ and \underline{Q} is a polynomial with a second order 0 at the origin. Suppose that $\min |\underline{\nu} \cdot \underline{\lambda} - \lambda_j| > c > 0$ with the minimum evaluated over all *non-negative integer components* vectors $\underline{\nu}$ with $\sum_j \nu_j \geq 2$ and over all $j = 1, \dots, n$. Show the existence of a function $\underline{h}(\underline{\zeta})$ holomorphic near $\underline{0}$ and vanishing to second order such that the change of variables $\underline{z} = \underline{\zeta} + \underline{h}(\underline{\zeta})$ transforms the differential equation into $\dot{\underline{z}} = \underline{\Lambda}\underline{z}$. Hence if $\text{Re } \lambda_j \leq 0$ all its solutions with small enough initial data are asymptotic to a quasi-periodic solution. Exhibit cases in which all solutions are quasi-periodic. (*Hint*: No small divisors problems arise in the perturbative determination of \underline{h} . Consider in particular the cases $\lambda_j = i + \varepsilon_j$, with $\varepsilon_j \leq 0$, and the cases $\lambda_j = i + i\varepsilon_j$, with $|\varepsilon_j|$ small enough.)

§8.2 Graphs and diagrams for the Lindstedt series

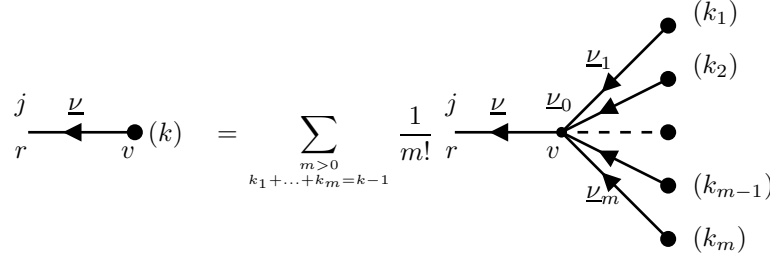
The analysis of the previous section provides us with a recursive algorithm to evaluate a formal power series solution to our problem. However it is very convenient to interpret the algorithm in terms of graphs and diagrams. An explicit construction of the values of the coefficients $\underline{h}^{(k)}(\underline{\psi})$ of the Lindstedt series can be obtained by simply “iterating” (8.1.15) until only $\underline{h}^{(1)}(\underline{\psi})$, cf. (8.1.13), appears. We get what we shall call a *tree representation* of $\underline{h}^{(k)}(\underline{\psi})$.

The construction is most easily performed through the diagrammatic expansion and therefore we dedicate some care to its detailed discussion. Note that the computation of $\underline{h}^{(k)}$ involves the inversion of the operator $(\underline{\omega}_0 \cdot \partial_{\underline{\psi}})^2$ which is easily done if instead of $\underline{h}^{(k)}(\underline{\psi})$ we study its Fourier transform $\underline{h}_{\underline{\nu}}^{(k)}$ as this involves simply a division by $(\underline{\omega}_0 \cdot \underline{\nu})^2$, cf. (8.1.16). Hence we rewrite (8.1.14) as

$$\underline{h}_{\underline{\nu}}^{(k)} = \frac{1}{J(\underline{\omega}_0 \cdot \underline{\nu})^2} \sum_{m=1}^{\infty} \sum_{\nu_0 + \nu_1 + \dots + \nu_m = \underline{\nu}} \frac{1}{m!} i^{\nu_0} f_{\nu_0} \sum_{\substack{k_1, \dots, k_m \geq 1 \\ k_1 + \dots + k_m = k-1}} \prod_{s=1}^m i^{\nu_s} \cdot \underline{h}_{\nu_s}^{(k_s)}. \tag{8.2.1}$$

e8.2.1

And the latter relation can be represented by Fig.(8.2.1), where the l.h.s is a symbol for $\underline{h}_{\underline{\nu}}^{(k)}$ and in the r.h.s. we have a “simple tree” consisting of a “root” r , a “root branch” $\lambda_v \equiv rv$ coming from the “node” (or “vertex”) v and $m = m_v$ branches “entering v ”.



F8.2.1 **Fig.(8.2.1)** Diagrammatic interpretation of (8.2.1). The tensor labels are not marked.

The length of the branches and the angles at which they are drawn (with respect to the sides of the present sheet of paper) are irrelevant.

The node v symbolizes the tensor with entries

$$e8.2.2 \quad \Phi_{v;j_0,j_1,\dots,j_m} = \frac{1}{m_v!} J^{-1}(i_{\underline{v}_0})_{j_0} f_{\underline{v}_0} \prod_{s=1}^m (i_{\underline{v}_0})_{j_s}, \quad (8.2.2)$$

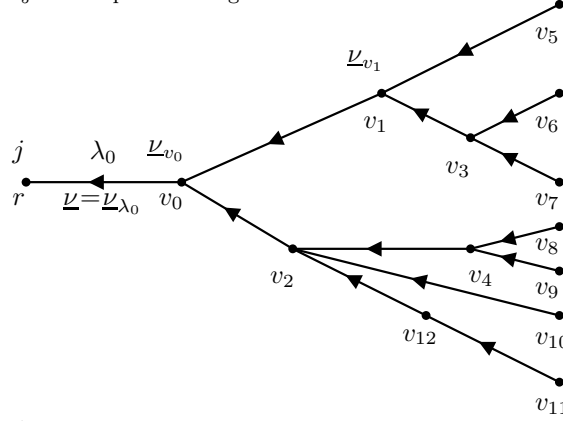
The line entering from the node v in the r.h.s. represents the *propagator*, *i.e.* the matrix

$$e8.2.3 \quad \frac{\delta_{jj_0}}{(\underline{\omega}_0 \cdot \underline{v})^2} \quad (8.2.3)$$

The line exiting from the bullet of the l.h.s represents $\underline{h}_{\underline{v},j}^{(k)}$; the branches exiting from the bullets of the r.h.s. with label (k_s) , see figure (8.2.1), represent $\underline{h}_{\underline{v}_s,j'_s}^{(k_s)}$.

We should imagine that the labels j_0, j_1, \dots, j_m are affixed on the line exiting the node v and, respectively, on the lines entering the node v (and close to v). The labels j'_1, \dots, j'_m should be imagined affixed on the lines exiting the bullets on the r.h.s. of the drawings in figure (8.2.1). And the root label j can be imagined affixed over the root of the tree. Note that in such a way we associate a pair of labels (j'_v, j_v) to each branch $v'v$ of the tree.

The labels $j_0, j_1, \dots, j_m, j'_1, \dots, j'_m$ will be called *tensor labels* and the label j *root label*. However the drawing in the figure does not carry the tensor labels explicitly: the tensor labels on the same line are meant to be “contracted” (*i.e.* set to equal values) and summed over their possible values in $\{1, \dots, \ell\}$.



F8.2.2

Fig.(8.2.2) A tree ϑ with $m_{v_0} = 2, m_{v_1} = 2, m_{v_2} = 3, m_{v_3} = 2, m_{v_4} = 2, m_{v_{12}} = 1$ and $k = 13$, and some decorations. Only two mode labels and two momentum labels are explicitly marked on the lines λ_0, λ_1 ; the number labels, distinguishing the branches, are not shown. The arrows represent the partial ordering on the tree.

At this point, if one notes that (8.1.14) is multilinear in the $\underline{h}_{\underline{\nu}, j'_s}^{(k_s)}$ (i.e. of degree 1 in each of them), it is clear that we can just replace each of the branches exiting from a bullet with the same graphical expression in the r.h.s. of the above figure. And so on, until the labels (k) on all the branches exiting from a bullet (“top branches”) become equal to 1. In this case the branches will represent $\underline{h}_{\underline{\nu}, j'}^{(1)} = (i\underline{\nu}_{j'}) J^{-1} (\underline{\omega}_0 \cdot \underline{\nu}_{j'})^{-2} f_{\underline{\nu}_{j'}}$ for some j' .

Thus we have represented our $\underline{h}_{\underline{\nu}}^{(k)}$ as a “sum over trees”, with k branches and k nodes (we shall not regard the root as a node), of suitable “tree values”.

The combinatorial factors (coming from the factorial in (8.2.2)) can be simplified if we decide to distinguish the lines of the trees with k branches by affixing on them a *number label* taking values from 1 to k . The set of labels attached to the trees, other than the tensor labels, will be called *decorations*.

We shall regard as identical two decorated trees which can be overlapped (with all labels matching) by adjusting the lengths of the lines and by pivoting the branches merging into a node around it.

The rule to construct $\underline{h}_{\underline{\nu}}^{(k)}$ can be summarized as follows.

D8.2.1

(8.2.1) Definition: (Trees and tree values)

To evaluate the Fourier transform $h_{\underline{\nu}, j}^{(k)}$ of $h_j^{(k)}(\underline{\psi})$ the following definitions are given.

(i) A tree ϑ consists of a family of oriented branches (i.e. oriented segments or arrows) $\Lambda(\vartheta)$ numbered from 1 to k arranged to form a (rooted) tree as in Fig.(8.2.2) (i.e. all oriented towards the root r); the initial points of the branches are the nodes $V(\vartheta)$. A partial ordering relation, denoted \preceq , is induced by the lines orientations between the lines and between the nodes.

(ii) At each node v we attach a mode label $\underline{\nu}_v \in \mathbb{Z}^\ell, |\underline{\nu}_v| \leq N$ (cfr. (8.1.10)). At the root r we attach a unit vector $\underline{\nu}_r$ in a selected direction $j \in (1, \dots, \ell)$. The order of the tree will be the number k of nodes or

- N8.2.1 the (equal) number of branches.¹
 (iii) At each branch $\lambda_v \equiv v'v$, connecting a node v with the following node v' , we attach a momentum

$$e8.2.4 \quad \underline{\nu}_{\lambda_v} \equiv \sum_{w \preceq v} \underline{\nu}_w, \quad (8.2.4)$$

- i.e. we can think that into each node “enters” a momentum $\underline{\nu}_v$ which then “flows” toward the root so that the momentum flowing through the branch emerging from v is $\underline{\nu}_{\lambda_v}$.²

- N8.2.2 (iv) Near each node v we write $m_v + 1$ tensor labels $j_{v,0}, j_{v,1}, \dots, j_{v,m_v}$ in $1, \dots, \ell$: the first on the line exiting the node v and the remaining m_v on the lines entering the node v . We imagine such labels written very close to v .³

- N8.2.3 (v) With each line λ joining two nodes $v'v$ or joining the root r and the first node before it v_0 we associate a propagator matrix $\delta_{ij}(\underline{\omega} \cdot \underline{\nu}_\lambda)^{-2}$, cf. (8.2.3), whose indices i, j are contracted, respectively, with the labels $j_{v'}$ and j_v on the line and close to the nodes $v'v$ (defined in (ii) above). In the case of the root line the label j_r is taken equal to j , the index of the component of $h_j^{(k)}$ that we want to study (and no sum on the root label is performed).

(vi) The mode labels and the number labels “decorating” the tree will be called also decorations (i.e. the labels attached to the tree and the decorations are synonymous below). In particular the component indices j of points (iv) and (v) are not decoration since they are contracted, i.e. summed over.

(vii) With each k -th order tree we associate a combinatorial factor $k!^{-1}$. We shall call $\Theta_{k, \underline{\nu}, j}$ the set of all decorated trees of order k with momentum $\underline{\nu}_{\lambda_0} = \underline{\nu}$ flowing through the root branch and such that for each branch λ one has $\underline{\nu}_\lambda \neq \underline{0}$ and carrying a root label j .

The above combinatorial analysis has led us to a final result which is, perhaps surprisingly, very simple.

- L8.2.1 **(8.2.1) Lemma:** (Tree representation of the Lindstedt series)

Given the above definition $\underline{h}_{\underline{\nu}, j}^{(k)}$ will have the following representation for the j -th component of $\underline{h}^{(k)}(\underline{\psi}) = \sum_{\underline{\nu} \in \mathbb{Z}^\ell} \underline{h}_{\underline{\nu}}^{(k)} e^{i\underline{\nu} \cdot \underline{\psi}}$:

$$e8.2.5 \quad h_{\underline{\nu}, j}^{(k)} = -i \frac{1}{k!} \sum_{\vartheta \in \Theta_{k, \underline{\nu}, j}} \left(\prod_{v \in V(\vartheta)} J^{-1} f_{\underline{\nu}_v} \right) \prod_{\lambda = (v'v) \in L(\vartheta)} \frac{\underline{\nu}_{v'} \cdot \underline{\nu}_v}{(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_v})^2}, \quad (8.2.5)$$

where $\Theta_{k, \underline{\nu}, j}$ is defined in item (vii) of definition (8.2.1).

Remarks: (1) The factor $\underline{\nu}_r \cdot \underline{\nu}_0$ in the contribution from the factor corresponding to the root line $\lambda_0 \equiv rv_0$ has to be interpreted by taking $\underline{\nu}_r$ to be

¹ In Fig.(8.2.2) the mode labels are marked only above the nodes v_0 and v_1 ; the number labels and the other labels are not marked explicitly, for simplicity.

² In Fig.(8.2.2) the momentum labels are marked only on the root branch λ_0 and on the branch λ_1 .

³ In Fig.(8.2.2) the tensor labels are not marked.

the unit vector in the direction j (see item (v) of definition (8.2.1)).

(2) The introduction of the number labels that distinguish the tree lines of course greatly increases the number of terms contributing to a given order k because there are many terms that give the same contribution simply because the value of a tree, defined as

$$e8.2.6 \quad \text{Val}(\vartheta) = -i \frac{1}{k!} \left(\prod_{v \in V(\vartheta)} J^{-1} f_{\underline{\nu}_v} \right) \prod_{\lambda = (v'v) \in L(\vartheta)} \frac{\underline{\nu}_{v'} \cdot \underline{\nu}_v}{(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_v})^2}, \quad (8.2.6)$$

does not depend on the number labels. However the advantage is that the combinatorial factor is $k!^{-1}$ for all trees, while using trees without number labels would require a combinatorial factor $\prod_v m_v!^{-1}$ which depends on the structure of each tree: the advantage being that one would need to consider far less trees.

The above formulae remain valid also in the case of a perturbation $V(\underline{a})$ which is not even: however in such case one needs to check (a result by Poincaré, [Po93]) that this is correct because it does not follow immediately from a parity property that the graphs in which a zero current flows have to be discarded; see problem [8.4.8] in Section §(8.4). The formulae also extend to the case in which $V(\underline{a})$ is analytic in \underline{a} .

It is sometimes useful to rewrite (8.2.6) more compactly (however somewhat more symbolically) as

$$e8.2.7 \quad \text{Val}(\vartheta) = \frac{1}{k!} \left(\prod_{v \in V(\vartheta)} F_v \right) \left(\prod_{\lambda \in \Lambda(\vartheta)} G_\lambda \right), \quad h_{\underline{\nu}, j}^{(k)} = \sum_{\vartheta \in \Theta_{k, \underline{\nu}, j}} \text{Val}(\vartheta) \quad (8.2.7)$$

where the symbols are introduced in the following definition.

(8.2.2) Definition: (Propagator and node factors)

D8.2.2 Given a node v and a line λ the node factor F_v (a tensor) and the propagator G_λ (a $\ell \times \ell$ -matrix proportional to the identity) will be defined, respectively as

$$e8.2.8 \quad F_v = J^{-1} (i \underline{\nu}_v)^{m_v+1} f_{\underline{\nu}_v}, \quad G_\lambda = \frac{1}{(\underline{\omega}_0 \cdot \underline{\nu}_\lambda)^2}. \quad (8.2.8)$$

Remarks: (1) The above analysis is well known in graph theory. Since the order of the tree is k , equal to the number of nodes, the above trees are analogous to Feynman graphs of a perturbation theory and the evaluation of the values “corresponds” to the Feynman rules for the KAM problem (in fact a literal meaning could be given to the mentioned analogy, [Ga01]).

(2) In the general case the representation in (8.2.5) is known as the *Lindstedt series* and it was introduced, at low orders only, in Celestial Mechanics problems by Lindstedt and Newcomb (independently); it has been extended and studied to all orders by Poincaré: about the name see comment following (8.1.10). The difficulty in the extension, solved by Poincaré, is that of proving that the algorithm makes sense to all orders, *i.e.* the equations for

$\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$ can be solved by Fourier transform “without ever dividing by 0”. The latter difficulty is absent here because of the parity assumption on $V(\underline{\alpha})$ which excludes that $(\underline{\omega}_0 \cdot \partial_{\underline{\psi}})\underline{H}^{(k)}(\underline{\psi})$, as given by (8.1.11), has non-zero average over $\underline{\psi}$ (being odd), see problems [8.4.8] and [8.4.9] for the non-even cases.

Formula (8.2.5) would provide immediately a proof of proposition (8.1.1), the KAM theorem, if certain trees were not present. The idea is that the values of the “unwanted” trees cancel each other to the extent that their sums behave well enough not to spoil the bounds. *See problem [8.2.4] for an example of an “unwanted” tree with value of extremely large size.*

To see this, *for the purpose of illustration*, we shall first restrict the sum in (8.2.7) to a sum over trees which, besides the property that for all branches one has $\underline{\nu}_\lambda \neq \underline{0}$ (as expressed by the definition of the set $\Theta_{k,\underline{\nu},j}$ in equation (8.2.7)), 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 $v' \preceq v \preceq v_0$.

This property is remarkable because of the following lemma.

L8.2.2

(8.2.2) Lemma: (Summation of values of trees with property [P])
Consider the formal power series in ε obtained by considering the sum of the trees values of the trees in $\Theta_{k,\underline{\nu},j}$ which verify property [P] above. Then the series has a positive convergence radius.

Remark: The purpose of this lemma is to make clear where the problem of proving proposition (8.1.1), *i.e.* the KAM theorem, is really located. To prove the theorem we shall have to understand how to bound the sum of the values of tree graphs which do not satisfy property [P].

Proof: There are at most $2^{2k}k!$ trees (see problem [8.2.2]) and the contraction of the tensors labels F_v , cf. (8.2.8), gives at most ℓ^k terms, while the node momenta $\underline{\nu}_v$ can be chosen in a number of ways bounded by $(2N + 1)^{\ell k} < (3N)^{\ell k}$. Therefore, if $f_0 = \max_{\underline{\nu}} |f_{\underline{\nu}}|$,

$$\begin{aligned} \sum_{\underline{\nu}} \left| \underline{h}_{\underline{\nu}}^{(k)} \right| &\leq (3N)^{\ell k} 2^{2k} \ell^k \frac{f_0^k C^{2k}}{J^k} N^{2k-1} \max_{\vartheta \in \Theta_{k,\underline{\nu},j}} \prod_{\lambda \in V(\vartheta)} (C\underline{\omega}_0 \cdot \underline{\nu}_\lambda)^{-2} \\ &\leq (f_0 C^2 J^{-1})^k N^{(\ell+2)k-1} (4\ell 3^\ell)^k M, \end{aligned} \tag{8.2.9}$$

where M is an estimate of the indicated maximum which is over the k -th order trees ϑ verifying property [P] above. Hence the whole problem is reduced to find a value for M , a “small divisors problem”.

Let q be large: then by the Diophantine condition in (8.1.6) 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 $\underline{\nu} \in \mathbb{Z}^\ell$ 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 branches” (*i.e.* of

branches corresponding to q -singular harmonics) in a tree ϑ with k nodes. Then we can have recourse to the following lemma

(8.2.3) Lemma: (Simple Bryuno–Siegel bound)
The number $N(k, q)$ of “ q -singular branches” of a tree of order k in the above sense is equal to

$$N(k, q) \leq \text{const} \frac{k}{q^{1/\tau}}, \tag{8.2.10}$$

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

Remark: (1) We shall not prove here the lemma as the present argument is being carried over only for illustration purposes (in any event (8.2.10) and the value of the constant are an immediate corollary of the proof of the lemma discussed and proved below).

(2) The intuition behind (8.2.10) is very simple. In order to achieve a branch momentum $\underline{\nu} = \underline{\nu}_{\lambda_v}$ with $C\underline{\omega}_0 \cdot \underline{\nu}$ of size q^{-1} one needs at least $|\underline{\nu}| \geq q^{1/\tau}$, i.e. the node v must be preceded by *at least* $N^{-1}q^{1/\tau}$ nodes. Once a q -singular branch has been generated, the branches following it will have non- q -singular momentum and we must collect *about as many new nodes* to generate a second q -singular branch and so on. Since the total number of nodes is k it follows that the number of q -singular branches is bounded proportionally to $k/q^{1/\tau}$. The actual estimate of the constant in (8.2.10) is irrelevant for our immediate purposes.

Assuming lemma (8.2.3) we can continue the proof of lemma (8.2.2). We fix an exponentially decreasing sequence γ^n , $n = 1, 0, -1, -2, \dots$; we choose $\gamma = 2$. The number of 2^{-n} -singular harmonics among the line momenta which are not $2^{-(n-1)}$ -singular is bounded by $2N2^{3/\tau} k 2^{-n}$, (being trivially bounded by the number of 2^{-n} -singular harmonics!). Hence we can take as a bound on M in (8.2.9) the quantity

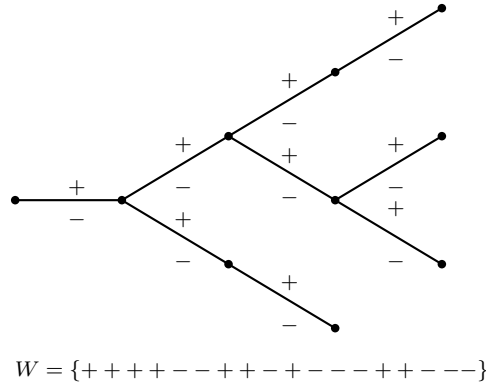
$$\prod_{\lambda \in \Lambda(\vartheta)} \frac{1}{(C\underline{\omega}_0 \cdot \underline{\nu}_\lambda)^2} \leq \prod_{n=-\infty}^{-1} 2^{-(n-1)4N2^{3/\tau}2^{n/\tau}k} = e^{cN\tau k} \equiv M, \tag{8.2.11}$$

where $c > 0$ is a suitable constant (see problem [8.2.5]); therefore the series for the approximation to $h_{\underline{\nu}}^{(k)}$, that we are considering because of the extra restriction in the sum equation (8.2.4), has radius of convergence in ε bounded below by ε'_0 given by

$$(\varepsilon'_0)^{-1} = (f_0 C^2 J^{-1})^k N^{(\ell+2)} (4\ell 3^\ell) e^{cN\tau}. \tag{8.2.12}$$

Hence the proof of the lemma (8.2.2) is concluded. ■

Lemma (8.1.1) shows that we must understand the contributions to the value of $h^{(k)}$ coming from the tree graphs that violate property [P] above.



F8.2.3 Fig.(8.2.3) A rooted tree and the corresponding random walk W .

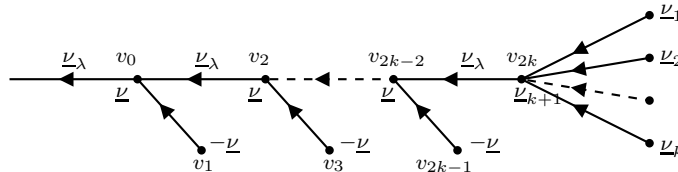
Problems for §8.2

Q8.2.1 [8.2.1]: Prove that the number of trees of order k with assigned labels $\{m_v\}$ is bounded by $(k - 1)! / \prod_v m_v!$ (*Hint*: It is just problem [7.1.4]).

Q8.2.2 [8.2.2]: (*Counting tree graphs*) Show that the number of elements in $\Theta_{k,\nu}$ is bounded by $k!2^{2k}$. (*Hint*: The number of non-numbered trees is bounded by the number of random walks on the one-dimensional lattice with $2k - 1$ steps, that is by 2^{2k} . In fact we can imagine walking on the tree starting at the root and marking $+1$, i.e. a step forward, every time we pass a node; once we reach an endpoint we imagine reversing the direction and continue walking on the “other side” of the branch (we imagine that each branch has an upper and a lower side) marking a -1 , i.e. a step backward until we reach a non trivial node where we reverse the walk direction etc.; see Fig.(8.2.3). At the end we reach the root and the sum of the number of $+1$'s equals that of the -1 and the sequence of ± 1 's determines uniquely the non-numbered tree. Furthermore the number of ways of assigning the numbers $1, \dots, k$ to the branches is $k!$.)

Q8.2.3 [8.2.3]: Prove inductively, by using the result of [8.2.2], that if in (8.2.7) we replace G_λ with 1 then the bound $|h_\nu^{(k)}| < C^k$ does follows for a suitable constant C . (*Hint*: Prove that one has $|h_\nu^{(k)}| < D^k N(k)$, where D is a suitable constant and $N(k)$ is the number of not numbered trees of order k , then use the bound on $N(k)$ proved in problem [8.2.2].)

Q8.2.4 [8.2.4]: (*Factorial growth of (some) trees*) Consider the tree θ :



Show that one can have $\text{Val}(\theta) \simeq C^k k^{\alpha k}$ for suitable positive constants C and α , where the order of θ is $3k + 1$. (*Hint*: : The momentum $\underline{\lambda}$ flowing through the line $v_{2k}v_{2k-2}$ is bounded by Nk , and it can happen that $C_{\underline{\omega}_0} \cdot \underline{\lambda}$ is of the order of $(Nk)^{-\tau}$. Then choose $\underline{\nu}_{2p} + \underline{\nu}_{2p+1} = 0$ for $p = 0, \dots, k - 1$ (hence obtaining $k - 1$ self-energy graphs with the terminology introduced in the forthcoming Section §(8.3)), and use that there are k line carrying a momentum $\underline{\lambda}$.)

Q8.2.5 [8.2.5]: Estimate the constant c in (8.2.11).

§8.3 Cancellations

The key remark in order to take into account the values of the trees that we have excluded by imposing the *unphysical property* [P] in lemma (8.2.2) is that they cancel *almost exactly*. This is the core of the theorem: indeed the previous sections basically contain only definitions necessary to set up the formalism and the following contains bounds based on the maximum principle and the analyticity properties of tree values.

The reason behind the cancellation is very simple. If $\underline{\nu}_{\lambda_{v'}} = \underline{\nu}_{\lambda_v}$ for two comparable nodes $v' \succ v$, and if we denote with ϑ the tree with $\lambda_{v'}$ as root line, we imagine to detach from the tree ϑ the subtree ϑ_2 with last node v . Then attach it to all the *remaining* nodes $w \preceq v'$, $w \in \vartheta/\vartheta_2$, (this means the tree ϑ with the subtree ϑ_2 removed). The simplest case is illustrated in Fig.(8.3.1). We can, furthermore, imagine changing all the signs, simultaneously, of the node momenta of the nodes $w \in \vartheta/\vartheta_2$. Then we add together all the values of the family \mathcal{D} of trees constructed in this way.

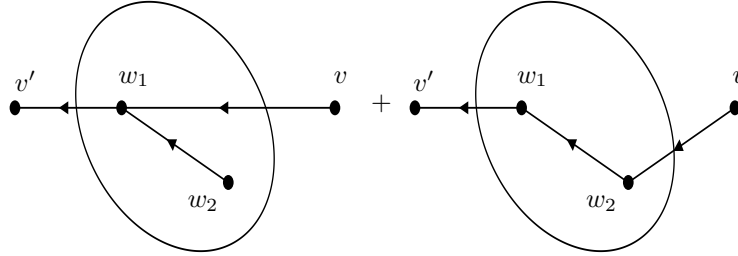
Note that the line momenta of the lines λ_w emerging from $w \in \vartheta/\vartheta_2$ have either the form $\underline{\nu}_{\lambda_w} = \underline{\nu}_{\lambda_w}^0 + \underline{\nu}$ with $\underline{\nu} = \underline{\nu}_{\lambda_v}$ and $\underline{\nu}_{\lambda_w}^0 = \sum_{w' \preceq w, w' \in \vartheta_2} \underline{\nu}_{w'}$ or the form $\underline{\nu}_{\lambda_w} = \underline{\nu}_{\lambda_w}^0$: the first case arises when the path from v to v' passes through w .

Therefore if $\delta = \underline{\omega}_0 \cdot \underline{\nu}$ the denominators of the propagators of the lines $\lambda_w \in \vartheta_2$ (*i.e.* preceding v' but not preceding v) will have the form $(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}^0 + \delta)^{-2}$ or, respectively, $(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}^0)^{-2}$. Hence the above trees can be considered only if no one among the divisors vanishes: which cannot happen *as we suppose temporarily* if $\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}$ is large compared to δ . By construction if we sum the values of all trees in the family \mathcal{D} we get an even function of δ .

The family of trees \mathcal{D} consists of trees whose contributions to $\underline{h}_{\underline{\nu}}^{(k)}$ differ because

(1) some of the branches below v' have changed total momentum 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 $(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w} + \delta)^{-2}$ or $-(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w} + \delta)^{-2}$ (see the branch $\lambda_{w_2} \equiv w_1 w_2$ in Fig.(8.3.2));

(2) there is one of the node factors which changes by taking successively the



F8.3.1

Fig. (8.3.1) The simplest cancellation: the circle encloses a subgraph which violates of property [P] (which we shall call later a *self-energy graph*), provided $\underline{\nu}_{w_1} + \underline{\nu}_{w_2} = 0$. The parts of the tree ϑ above v' and below v are not drawn. Imagine that the branch momentum $\underline{\nu}$ of the branch coming out of v is very large and that $\delta \equiv \underline{\omega}_0 \cdot \underline{\nu}$ is very small and note that in the two trees one has $\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_{w_2}} = \underline{\omega}_0 \cdot \underline{\nu}_{w_2}$ and $\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_{w_2}} = \underline{\omega}_0 \cdot \underline{\nu}_{w_2} + \delta$, respectively.

value $\underline{\nu}_w$, where $w \in \vartheta/\vartheta_2$ is the node to which such a branch is reattached.

Hence if $\delta = \underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}$ is replaced by 0 and if we sum the values of all the trees considered we would build, in this *resummation*, a quantity proportional to $\sum \underline{\nu}_w = \underline{\nu}_{\lambda_{v'}} - \underline{\nu}_{\lambda_v}$ which is zero. Since $\delta \neq 0$ we can expect to see a sum of order δ^2 , as the sum that we are considering is even in δ .

However this can be advantageous only if $|\delta| \ll |\underline{\omega}_0 \cdot \underline{\nu}_{\lambda}^0|$ for any branch λ in $\vartheta \setminus \vartheta_2$. If the latter property does not hold then $\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}$ must be small of order δ at least for some $w \in \vartheta/\vartheta_2$ and $(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}^0 + \delta)^{-2}$ may be an extremely small divisor spoiling the gain due to the fact that the sum vanishes for $\delta = 0$ to second order in δ .

By applying carefully the argument above one sees that the extreme case in which all lines λ_w are such that $\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}$ is close to δ would be essentially also treatable. Therefore the problem is to show that the two regimes just envisaged (and their “combinations”) do exhaust all possibilities.

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 argue here that Eliasson’s method can be interpreted in the same way. The following definition of scale of a line with momentum $\underline{\nu}$ will play a key role.

D8.3.1

(8.3.1) Definition: (Scale of a propagator)

Given a scaling factor γ and setting $\underline{\omega} = C\underline{\omega}_0$ we say that a propagator $G_\lambda = (\underline{\omega}_0 \cdot \underline{\nu}_\lambda)^{-2}$ of a line λ has scale n if

e8.3.1

$$\gamma^{n-1} \leq |\underline{\omega} \cdot \underline{\nu}_\lambda| < \gamma^n, \quad \text{for } n \leq 0, \quad (8.3.1)$$

and we set $n = 1$ if $1 \leq |\underline{\omega} \cdot \underline{\nu}_\lambda|$.

Remark: (1) The notion of scale of a line of a graph was introduced in the theory of (overlapping) divergences in the renormalization of the perturbation series arising in quantum field theory by Hepp, [He61].

(2) In the following we fix $\gamma = 2$: this is an arbitrary choice which recommends itself. However any other value $\gamma > 1$ would be suitable in order to carry the analysis that we need.

We make at this point a second simplifying assumption (which can be removed quite easily as discussed in Section §(8.4)). Namely we want to suppose more than the Diophantine condition (8.1.6), *i.e.* that the vector $\underline{\omega}_0$ is *strongly Diophantine* in the sense of the following definition.

D8.3.2 **(8.3.2) Definition:** (Strongly Diophantine vectors)

Let $\underline{\omega}_0 \in \mathbb{Z}^\ell$ and suppose that there exists $\gamma > 1$ and $C, \tau > 0$ such that for all integers $n \leq 0$

$$\begin{aligned}
 \text{e8.3.2} \quad (1) \quad & C |\underline{\omega}_0 \cdot \underline{\nu}| \geq |\underline{\nu}|^{-\tau}, & \underline{0} \neq \underline{\nu} \in \mathbb{Z}^\ell, & \quad (8.3.2) \\
 (2) \quad & \min_{0 \geq p \geq n} |C |\underline{\omega}_0 \cdot \underline{\nu}| - \gamma^p| > \gamma^{n+1} & \quad 0 < |\underline{\nu}|^\tau \leq (\gamma^{-(n+3)}). &
 \end{aligned}$$

We shall say that $\underline{\omega}_0$ satisfies a strong Diophantine property with scale factor γ , constant C and exponent τ .

Remark: (1) The notion means that the values of the numbers $C|\underline{\omega}_0 \cdot \underline{\nu}|$ are “far” from the prescribed sequence of scale factors γ^p for all $n \leq p \leq 0$ provided $|\underline{\nu}|$ is not too large, *i.e.* it does not exceed a constant times $\gamma^{-n\tau-1}$. The constant is chosen $\gamma^{-3/\tau}$ for later convenience. In other words, taking $\gamma = 2$ a line of momentum $\underline{\nu}$ and scale $n + 3$ is such that the value $C|\underline{\omega}_0 \cdot \underline{\nu}|$ is far on the scale 2^n from the “marks” 2^p , for $n \geq p \geq 0$.

(2) For a given γ one can prove that the set of strong Diophantine vectors contained in any ball of radius r in \mathbb{R}^ℓ has relative measure which tends to 1 for $C \rightarrow \infty$; see problem [8.3.1].

In view of the last remark (2) and for simplicity we shall take $\gamma = 2$ and suppose that $\underline{\omega}_0$ is strongly Diophantine with scale factor 2, constant C and exponent τ , (see problems for Section §(8.4) for a discussion on releasing the assumption).

Proceeding as in quantum field theory, given a tree ϑ we can attach a *scale label* to each branch $\lambda \in \Lambda(\vartheta)$: it is equal to n if n is the scale of the branch propagator. Note that the labels thus attached to a tree are *uniquely determined by the tree*: they will have only the role of helping to visualize the orders of magnitude of the divisors associated with the various tree branches. The scale labels allow us to organize naturally the lines of a tree into “clusters” defined formally as follows.

D8.3.3 **(8.3.3) Definition:** (Clusters of lines and nodes)

Given a tree in which each line carries its scale label one can identify the largest connected clusters T of nodes that are linked by continuous paths of branches carrying the same scale label n_T or a higher one. 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 branches connecting them; by extension we shall say that such branches are contained, or internal, in T . We also denote

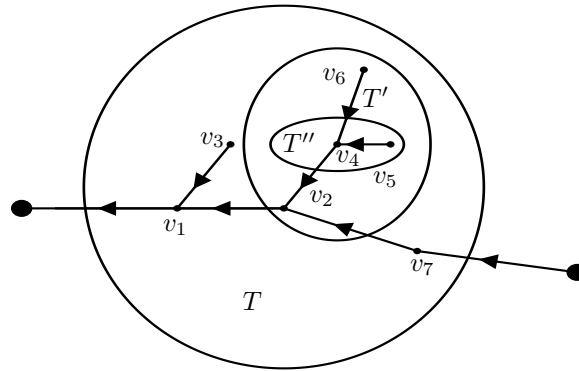
by $\Lambda_1(T)$ the set of branches in $\Lambda(T)$ plus the exiting branch of T (if any). Finally call $\mathcal{T}(\vartheta)$ the set of all clusters in ϑ .

Remark: (1) This can be visualized by drawing a box enclosing all the nodes and lines internal to a cluster T . In this way the boxes are hierarchically ordered by inclusion, cf. Figure (8.3.2).

(2) The definition implies that a cluster T of scale n_T is *maximal*, i.e. it cannot be enlarged by adding new lines of scale $n \geq n_T$.

(3) The branches of the tree carry an arrow pointing to the root: this gives a meaning to the expressions “entering” or “exiting a cluster”.

(4) Each cluster can have at most one exiting line. One can also imagine to define the entering lines (if any). Note that the number of entering lines can be arbitrary: hence figure (8.3.2) is a rather special case, so chosen for its convenience in the later illustration of other features of clusters of a graph.



F8.3.2

Fig.(8.3.2) An example of three clusters symbolically delimited by circles, as visual aids, inside a tree (whose remaining branches and clusters are not drawn and are indicated by the bullets); not all labels are explicitly shown. The scales (not marked) of the branches increase as one crosses inward the circles boundaries: recall, however, that the scale labels are integers ≤ 1 (hence typically ≤ 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_2, v_4, v_5, v_6) add up to $\underline{0}$ the cluster T' is a self-energy graph and such is T if the mode labels of $(v_1, v_2, v_3, v_4, v_5, v_6, v_7)$ add up to $\underline{0}$. The cluster T' is maximal in T .

Among the clusters we consider the ones with the property that there is only one tree branch entering them and only one exiting and both carry the same momentum. We set the following definition.

D8.3.4

(8.3.4) Definition: (Self-energy clusters and divergence seeds)

Let T be a cluster with just one entering and one exiting lines. Denote λ_T the entering branch: its scale $n = n_{\lambda_T}$ is smaller than the smallest scale n_T of the branches inside T . We call w_1 the node into which the branch λ_T ends inside T . We say that such a T is a self-energy subgraph if the following conditions are verified.

(i) $\sum_{w \in T} \underline{v}_w = \underline{0}$: hence the entering and exiting branches carry the same momentum.

(ii) If $n = n_{\lambda_T}$, and E, η are defined by $E \equiv 2^{-3\eta} N^{-1}$, $\eta = \tau^{-1}$ then the

number $M(T)$ of branches contained in T is not too large:

$$e8.3.3 \quad M(T) \stackrel{def}{=} \text{number of branches contained in } T \leq E 2^{-n\eta}, \quad (8.3.3)$$

We call n_{λ_T} the self-energy-scale of T , and λ_T a self-energy branch. A graph containing a self-energy graph will be said to contain a divergence seed.

Remarks: (1) The self-energy graphs are called *resonances* in Eliasson’s terminology, [El96].

(2) Note that the self-energy-scale n of a self-energy graph T is different from the scale n_T of T as a cluster: one has $n < n_T$.

(3) The reason for the name of divergence seed that we give to self-energy subgraphs is due to the fact that the values of graphs of order k which contain many divergence seeds can only be, in general, bounded proportionally to a power of $k!$ as the key example in problem [8.2.4] shows. The appropriateness of the name is even more clear when one considers the theory of invariant tori of dimension lower than the maximal, *i.e.* lower than the number of degrees of freedom, [GG02], of the approach to KAM theory based on “resummations” of the Lindstedt series; see Sections §(9.1) and §(9.1).

Let us consider a tree ϑ and its clusters. We wish to estimate the number $N_n(\vartheta)$ of branches 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 graphs of self-energy-scale n contained in T (hence with entering branches of scale n). Recalling that $\Theta_{k,\underline{\nu},j}$ is defined in item (vii) of definition (8.2.1), we have the following inequality.

L8.3.1 **(8.3.1) Lemma:** For all trees $\vartheta \in \Theta_{k,\underline{\nu},j}$ one has

$$e8.3.4 \quad N_n(\vartheta) \leq \frac{4k}{E 2^{-\eta n}} + \sum_{\substack{T \in \mathcal{T}(\vartheta) \\ n_T = n}} (-1 + q_T), \quad (8.3.4)$$

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

Remarks: (1) This is a version of Bryuno’s lemma which goes back to Bryuno’s work on a theorem by Siegel (in fact Siegel found in a seminal paper the first solution of a small divisor problem, cf. problem [8.3.2]); a proof of (8.3.1) is deferred to Appendix (8.3) below.

(2) Note that a cluster T of scale n can contain self-energy graphs of self-energy scale n which in turn contain other self-energy graphs (necessarily of higher self-energy scale): the number q_T in lemma (8.3.1) counts *only* the self-energy graphs of self-energy scale n .

Besides an estimate of the number of self-energy graphs that can be found in a tree in $\Theta_{k,\underline{\nu},j}$ we shall need a precise description of the families of trees containing divergence seeds whose values we want to sum together to exhibit

the cancellations that show that their sums are not as large as they could *a priori* be. The families are defined below: a definition whose usefulness relies upon the validity of the following lemma (8.3.2). The formulation is quite involved and the reader is advised to follow it on a drawing of a graph, *e.g.* on the graph of figure (8.3.2).

D8.3.5 **(8.3.5) Definition:** (Resummation families)

Given a labeled tree $\vartheta \in \Theta_{k, \underline{\nu}, j}$ and a self-energy graph T of ϑ detach the part of ϑ which has λ_T as root line and attach it successively to the points $w \in \tilde{T}$, where \tilde{T} is the set of nodes of $V(T)$ outside the self-energy graphs contained in T if any (note that the endpoint $w_1 \in V(T)$ of λ_T is among them). This defines a family $\mathcal{F}_1(\vartheta)$ of trees that have the same self energy graphs. Denote by $\Lambda(\tilde{T})$ the set of branches λ contained in T and with at least one point in \tilde{T} , and by $\Lambda_1(\tilde{T})$ the set of branches in $\Lambda(\tilde{T})$ plus the self-energy branches λ_T ; note that all branches $\lambda \in \Lambda(\tilde{T})$ have a scale $n_\lambda \geq n_T$. Select another self energy graph T^1 of ϑ and repeat the above operations of detaching and attaching the line λ_{T^1} on all the trees $\vartheta^1 \in \mathcal{F}_1(\vartheta)$ and thus define a larger family of trees $\mathcal{F}_2(\vartheta)$. Imagine the procedure repeated for all self-energy graphs in ϑ . For each self-energy graph T of ϑ we shall call V_T the number of nodes in \tilde{T} , *i.e.* $V_T = |V(\tilde{T})|$. 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 $w \in \tilde{T}$: the change of sign is performed independently for the various self-energy graphs. This defines a family of $\prod 2V_T$ trees that we call $\mathcal{F}(\vartheta)$ (the product is over all self-energy graphs in ϑ).

Remarks: (1) The number $\prod 2V_T$ is bounded by $\exp \sum 2V_T \leq e^{2k}$.

(2) 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 branch entering the self-energy graph is attached) has the property that it cannot change too much the sizes of the propagators of the branches inside the self-energy graphs. The reason is simply that inside a self-energy graph of self-energy-scale n the number of branches is not very large, being $\leq \bar{N}_n \equiv E 2^{-nn}$ and the number E has been chosen to make the property true.

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

$$e8.3.5 \quad 2^{n_1} + 2^{n_2} + \dots < 2^{n+1}. \quad (8.3.5)$$

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

Therefore not only $|\underline{\omega}_0 \cdot \underline{\nu}_\lambda^0| \geq 2^{n+3}$ (because $\underline{\nu}_\lambda^0$ is a sum of $\leq \bar{N}_n$ node modes, so that $|\underline{\nu}_\lambda^0| \leq N \bar{N}_n$) but $\underline{\omega}_0 \cdot \underline{\nu}_\lambda^0$ is “in the middle” of the dyadic interval containing it and by (8.3.2) does not get out of it if we add a quantity bounded by 2^{n+1} (like $\sigma_\lambda \underline{\omega}_0 \cdot \underline{\nu}_{\lambda_T}$), *cf.* remark (1) to definition

(8.3.2). Hence no branch changes scale as ϑ varies in $\mathcal{F}(\vartheta^1)$, if $\underline{\omega}_0$ verifies (8.3.2).

By the strong Diophantine property (8.3.2) imposed on the rotation vector ω_0 we have therefore achieved a proof of the following lemma.

L8.3.2

(8.3.2) Lemma: (Invariance of the scales of the lines of trees in the resummation families)

The self-energy graphs of the trees ϑ in $\mathcal{F}(\vartheta^1)$ all contain the same sets of branches, and the same branches enter or exit each self-energy graph (although they are attached to generally distinct nodes inside the self-energy graphs: the identity of the branches is here defined by the number label that each of them carries in ϑ^1). Furthermore the scales of the self-energy graphs, and of all the branches, do not change as ϑ varies in $\mathcal{F}(\vartheta^1)$.

Remarks: (1) The above proof of this key lemma relies in an essential way on the strong Diophantine property assumed for the rotation vector $\underline{\omega}_0$. It is therefore desirable to show that one can release the latter assumption. In the problems for Section §(8.4) we shall see that if $\underline{\omega}_0$ does not satisfy the strong Diophantine property for a sequence of scale factors γ^n with $\gamma = 2$ it essentially satisfies it for another sequence of scaling factors also exponentially shrinking to 0 (as $n \rightarrow -\infty$) and this is all one really needs to achieve a proof of the proposition (8.1.1).

(2) The lemma allows us to consider the collection of the trees in $\Theta_{k,\underline{\nu},j}$ organized as follows. Let ϑ^2 be a tree not in $\mathcal{F}(\vartheta^1)$ and construct $\mathcal{F}(\vartheta^2)$, etc. We define a collection $\{\mathcal{F}(\vartheta^i)\}_{i=1,2,\dots}$ of pairwise disjoint families of trees. We shall sum all the contributions to $h_{\underline{\nu},j}^{(k)}$ coming from the individual members of each family and we shall use that no line “changes scale” as the tree varies in the collection in the sense of lemma (8.3.2).

(3) Collecting tree values in this way is a realization of *Eliasson’s resummation*: it is more detailed than the original one in [El96], where no subdivision of the trees in classes was considered and the cancellation that we exhibit below 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 in the problem and the cancellation takes place in a more detailed fashion.

Appendix 8.3: Siegel-Bryuno bound on the number of self-energy graphs.

We give here a proof of the lemma (8.3.1). The argument followed below is a minor adaptation of Bryuno’s proof of Siegel’s theorem, as remarkably exposed by Pöschel, [Po86].

Proof of lemma (8.3.1). Call $N_n^*(\vartheta)$ the number of non-self-energy branches carrying a scale label $\leq n$ in a tree ϑ with k nodes. We shall prove first that $N_n^*(\vartheta) \leq 2k(E2^{-\eta n})^{-1} - 1$ 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 branch 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 $k_j > E 2^{-n\eta}$ branches, 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 $k < E 2^{-n\eta}$, which is certainly the case if E is chosen as in equation (8.3.4).¹

N8.3.1

N8.3.2

In the other case, call $\lambda_1, \dots, \lambda_m$ the $m \geq 0$ branches on scale $\leq n$ which are the nearest to λ_0 :² such branches are the entering branches of a cluster T on scale $n_T > n$. If ϑ_i is the tree with λ_i as root branch one has $N^*(\vartheta) \leq 1 + \sum_{i=1}^m N^*(\vartheta_i)$, and if $m = 0$ the statement is trivial, while if $m \geq 2$ 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 branch of ϑ_1 is a self-energy branch or it cannot have scale $\leq n$.

To see this, note that $|\underline{\omega} \cdot \underline{\nu}_{\lambda_0}| \leq 2^n$ and $|\underline{\omega} \cdot \underline{\nu}_{\lambda_1}| \leq 2^n$, hence $\delta \equiv |(\underline{\omega} \cdot (\underline{\nu}_{\lambda_0} - \underline{\nu}_{\lambda_1}))| \leq 2^{n+1}$, 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 graphs), 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 modes and therefore $|\underline{\nu}_{\lambda_0} - \underline{\nu}_{\lambda_1}| < N E 2^{-n\eta}/2$, 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 $p_n(\vartheta) \leq 2k (E 2^{-n\eta})^{-1} - 1$; see problem [8.3.3]. Thus equation (8.3.4) is proved. ■

Problems for §8.3

Q8.3.1

[8.3.1]: (*Full measure of strongly Diophantine vectors*)

Show that the volume of the points $\underline{\omega} \in \mathbb{R}^\ell$ that do not verify property (8.3.2) for some $C > 0$ has zero volume if $\tau > \ell - 1$. (*Hint:* Follow a path similar to the hint for problem [8.1.7].)

Q8.3.2

[8.3.2]: (*Siegel's theorem*)

Given a function $\underline{F}(\underline{z})$ of ℓ complex variables holomorphic in $|z_i| < R$ and with $\underline{F}(\underline{0}) = \underline{0}$, $\partial_{z_i} F_j(\underline{0}) = \delta_{ij} e^{i\omega_j}$, consider the map

$$z'_j = F_j(\underline{z}) = e^{i\omega_j} z_j + O(|\underline{z}|^2)$$

and the problem of finding new variables $\underline{\zeta} = \underline{\Phi}(\underline{z}) = \underline{z} + O(|\underline{z}|^2)$ such that near the origin the map takes the form

$$\zeta'_j = e^{i\omega_j} \zeta_j.$$

¹ Note that if $k \leq E 2^{-n\eta}$ one has, for all momenta $\underline{\nu}$ of the branches, $|\underline{\nu}| \leq N E 2^{-n\eta}$, i.e. $|\underline{\omega} \cdot \underline{\nu}| \geq (N E 2^{-n\eta})^{-\tau} = 2^3 2^n$ 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 graph concept.

² i.e. such that no other branch along the paths connecting the branches ℓ_1, \dots, ℓ_m to the root is on scale $\leq n$.

Show that if there exist $C, \tau > 0$ such that $|\underline{\omega} \cdot \underline{\nu} + 2\pi n| > C^{-1}(|\underline{\nu}| + |n|)^{-\tau}$ for all $\underline{\nu} \in \mathbb{Z}^\ell, n \in \mathbb{Z}$ with $|\underline{\nu}| + |n| > 0$ then Φ exists and is analytic near the origin $\underline{z} = \underline{0}$. (*Hint*: Setting $\underline{m} = (m_1, \dots, m_\ell) \in \mathbb{Z}_+^\ell$ and $|\underline{m}| = m_1 + \dots + m_\ell$, write $\Phi(\underline{z}) = \underline{z} + \sum_{|\underline{m}| > 1} c_{\underline{m}} \underline{z}^{\underline{m}}$ in Taylor series and determine $c_{\underline{m}}$ recursively via a tree expansion: note that $\underline{m} \in \mathbb{Z}_+^\ell$ plays here the role of $\underline{\nu}$ in the KAM theory: however no self-energy graphs can arise because $m_j \geq 0$. Therefore the result follows from a simpler version of lemma (8.3.1), *i.e.* from lemma (8.2.2) which is also implicit in lemma (8.3.1).)

Q8.3.3

[8.3.3]: (*Estimate of the number of clusters of scale n*)

Show that the number $p_n(\vartheta)$ of clusters on scale n contained in a tree ϑ verifies the bound $p_n(\vartheta) \leq 2k(E2^{-\eta n})^{-1} - 1$, with $\eta = \tau^{-1}$ and $E = N^{-1}2^{-3\eta}$. (*Hint*: The bound is true for $k \leq E2^{-\eta n}$. If the last tree node v_0 is not in a cluster of scale n one has $p_n(\vartheta) = p_n(\vartheta_1) + \dots + p_n(\vartheta_m)$, with the notation in Appendix 8.3, and the statement follows by induction. If v_0 is in a cluster of scale n we call $\vartheta_1, \dots, \vartheta_m$ the subtrees entering the cluster containing v_0 and with orders $k_j > E2^{-\eta n}$: then one has $p_n(\vartheta) = 1 + p_n(\vartheta_1) + \dots + p_n(\vartheta_m)$. As in Appendix 8.3 we can assume $m = 1$, the other cases being trivial. But in such a case there will be only one branch entering the cluster V of scale n containing v_0 and it will have a momentum of scale $\leq n - 1$. Therefore the cluster V must contain at least $E2^{-\eta n}$ nodes. This means that $k_1 \leq k - E2^{-\varepsilon n}$.)

Q8.3.4

[8.3.4]: In the context of problem [8.1.11] replace the condition that the minimum (evaluated over all non-negative integer components vectors $\underline{\nu}$ with $\sum_j \nu_j \geq 2$ and over all $j = 1, \dots, n$) $\min |\underline{\nu} \cdot \underline{\lambda} - \lambda_j|$ is $> c > 0$ with the condition $\min |\underline{\nu} \cdot \underline{\lambda} - \lambda_j| > C|\underline{\nu}|^{-\tau}$ for suitable constants $C, \tau > 0$ and prove that the same conclusions hold. (*Hint*: Follow the proof of Siegel's theorem in problem [8.3.2].)

§8.4 Convergence and KAM theorem

Having exhibited, cf. remark (2) to lemma (8.3.2) and definition (8.3.5), the class of graphs whose values we want to add up before attempting an estimate we turn to proving that the value of their sums admits, at fixed order k , a much better bound than the one they satisfy individually.

Let $\eta_T = \underline{\omega} \cdot \underline{\nu}_{\lambda_T}$ if $\underline{\nu}_{\lambda_T}$ is the momentum of the line λ_T of scale n_{λ_T} , $n_{\lambda_T} < n_T$ and $\underline{\omega} = C\underline{\omega}_0$. If λ is a branch in $\Lambda(\tilde{T})$, *i.e.* by definition (8.3.3) a branch of a line outside the inner self-energy clusters in T if any, we can imagine to write the quantity $\underline{\omega} \cdot \underline{\nu}_\lambda$ as $\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T$, with $\sigma_\lambda = 0, 1$ (cf. figure (8.3.2)). The product of the propagators of the branches in \tilde{T} is $C^{2|\Lambda(\tilde{T})|}$ times

e8.4.1

$$\prod_{\lambda \in \Lambda(\tilde{T})} \frac{1}{(\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T)^2}. \tag{8.4.1}$$

D8.4.1

(8.4.1) Definition: (Height of tress and of self-energy graphs)

If the tree does not contain any self-energy graphs, we say that it has height 0; if the only self-energy graphs do not contain other self-energy graphs, we say that the tree has height 1; more generally if the maximum number of self-energy graphs that contain a given self-energy graph is p , we say that the tree has height p .

Likewise we say that a self-energy graph has height 0 if it does not contain other self-energy graphs. Recursively we say that a self-energy graph has

height p if it contains only self-energy graph of height $\leq p-1$ and at least one self-energy graph of height $p-1$. Given a tree ϑ , call $\mathcal{V}(\vartheta)$ the set of all self-energy graphs in ϑ , and set

$$e8.4.2 \quad \Lambda(\mathcal{V}(\vartheta)) = \bigcup_{T \in \mathcal{V}(\vartheta)} \Lambda(T), \quad \Lambda_1(\mathcal{V}(\vartheta)) = \bigcup_{T \in \mathcal{V}(\vartheta)} \Lambda_1(T). \quad (8.4.2)$$

Remark: Of course in (8.4.2) the union could just be restricted to the maximal self-energy graphs (the sets $\Lambda(T)$ and $\Lambda_1(T)$ have been introduced in definition (8.3.5)).

Consider first the case of a tree ϑ of height 1 and let us denote by T any of its self-energy graphs: if we regard the quantities η_T as independent variables we see that (8.4.1) is holomorphic in η_T for $|\eta_T| < 2^{n_T-3}$. While η_T varies in such complex disk the quantity $|\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T|$ does not become smaller than 2^{n_T-3} .¹ The main point here is that the quantity 2^{n_T-3} will usually be $\gg 2^{n_\lambda}$ which is the value that η_T actually can reach in every tree in $\mathcal{F}(\vartheta)$; this 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 can factor out of the sum of the values of the graphs in $\mathcal{F}(\vartheta)$ the product $\prod_{T \in \mathcal{V}(\vartheta)} \eta_T^{-4}$ (because each self-energy graph has one branch entering and one exiting with the same propagator) and write the product of the propagators of any tree as $C^{2|\Lambda(\text{widetilde}\vartheta)|}$

$$e8.4.3 \quad \left(\prod_{\lambda \in \Lambda(\vartheta) \setminus \Lambda_1(\mathcal{V}(\vartheta))} \frac{1}{(\underline{\omega} \cdot \underline{\nu}_\lambda)^2} \right) \cdot \left(\prod_{\substack{T \in \mathcal{V}(\vartheta) \\ \lambda \in \Lambda(T)}} \frac{1}{(\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T)^2} \right) \cdot \left(\prod_{T \in \mathcal{V}(\vartheta)} \frac{1}{\eta_T^2} \right), \quad (8.4.3)$$

where the first product is over the branches λ which neither enter nor are inside a self-energy graph of ϑ (so that their momentum is the same in all trees of the family $\mathcal{F}(\vartheta)$), the second product is over the branches λ contained in V , and the third product is over the self-energy graphs $T \in \mathcal{V}(\vartheta)$ and takes into account the branches entering T , *i.e.* the self-energy branches. As said above the last product factors out of the sum of the values of the trees in $\mathcal{F}(\vartheta)$ at fixed ϑ .

Consider the sum of the $\prod 2V_T \leq e^{2k}$ products of propagators of the members of the family $\mathcal{F}(\vartheta)$ *divided by the last factor in (8.4.3)*. Each such product relative to the tree ϑ is holomorphic in the region $|\eta_T| < 2^{n_T-3}$ and it is bounded there by $\prod 2^{-2(n_\lambda-3)} \leq 2^{6k} \prod_\lambda 2^{-2n_\lambda}$, if n_λ the scale of the branch λ in ϑ and if the product is over the branches neither entering nor exiting a self-energy graph. This even holds if the η_T are regarded as independent complex parameters.

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

By construction it is clear that the just considered sum of the $\prod 2V_T \leq e^{2k}$ terms from the trees in $\mathcal{F}(\vartheta)$, vanishes to second order in the η_T parameters (by the approximate cancellation discussed above). *By the maximum principle* this means that if we bound the sum by the number of terms times the maximum among them (easy to estimate because the propagators have all well defined seizes) 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 $\underline{\nu}_v$ in (8.2.8) can be bounded by N , and that (as it is straightforward to check) one has

$$e8.4.4 \quad \sum_{v \in V(\vartheta)} m_v = k - 1, \quad (8.4.4)$$

we can bound the contribution to $\underline{h}_{\underline{\nu}}^{(k)}$ from the family $\mathcal{F}(\vartheta)$ by

$$e8.4.5 \quad \left[\frac{1}{k!} \left(\frac{f_0 C^2 N^2}{J} \right)^k 2^{6k} e^{2k} \prod_{n \leq 0} 2^{-2nN_n} \right] \left[\prod_{n \leq 0} \prod_{\substack{T \in \mathcal{T}(\vartheta) \\ n_T = n}} \prod_{i=1}^{q_T} 2^{2(n-n_i+3)} \right], \quad (8.4.5)$$

where

- (1) $N_n = N_n(\vartheta)$ is the number of propagators of scale n in ϑ ($n = 1$ does not appear as $|\underline{\omega} \cdot \underline{\nu}| \geq 1$ in such cases);
- (2) the first square bracket is the bound on the product of individual elements in the family $\mathcal{F}(\vartheta)$ times the bound e^{2k} on their number: this takes into account *also* the last product in (8.4.3);
- (3) the second square bracket is the part coming from the maximum principle, applied to bound the resummations, and is explained as follows.
 - (i) The dependence on the variables $\eta_{T_i} \equiv \eta_i$ relative to self-energy graphs $T_i \subset T$ 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).
 - (ii) 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 (8.4.3) by the product of factors $(|\eta_i| 2^{-n_i+3})^2$ if $n_i > n + 3$. If $n_i \leq n + 3$ we cannot gain anything: but since the contribution to the bound from such terms in (8.4.3) is > 1 we can leave them in it to simplify the notation, (of course this means that the gain factor can be important only when $\ll 1$).

We shall shortly see that the above would be sufficient if there were no trees of height higher than 1. To treat the general case we can proceed inductively and suppose that the bound in (8.4.5) holds for trees of height $1, 2, \dots, p-1$ and for values of the $\eta_T = \underline{\omega} \cdot \underline{\nu}_{\lambda_T}$ of the branches that enter the maximal self-energy graphs T which are in the complex disk $|\eta_T| < 2^{n_T-3}$.

Let ϑ be a tree with height p : then each of its maximal self-energy graphs V contains a tree of height $< p$. We imagine that all the resummations relative to the branches that enter the self-energy graphs that are not maximal have been performed so that we only have to consider the trees that are obtained

by attaching the branches that enter the maximal self-energy graphs T to the nodes in \tilde{T} .

Suppose for simplicity that there is only one maximal self-energy graph 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 graphs of lower height will have the form

$$e8.4.6 \quad \left(\prod_{\lambda \in \Lambda(\vartheta) \setminus \Lambda_1(T)} \frac{\underline{\nu}_{v'} \cdot \underline{\nu}_v}{(\underline{\omega} \cdot \underline{\nu}_\lambda)^2} \right) \cdot \left(\prod_{\lambda \in \Lambda(\tilde{T})} \frac{\underline{\nu}_{v'} \cdot \underline{\nu}_v}{(\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T)^2} \right) \cdot \frac{1}{\eta_T^2} \cdot \left(\prod_{T_i} F(T_i, \underline{\nu}_{v_i}, \underline{\nu}_{v'_i}) \right), \quad (8.4.6)$$

where the last product is over all the maximal self-energy graphs T_i contained in T , v'_i and v_i are the nodes in \tilde{T}_i from which exits (or enters, respectively) the branch that enters (or exits) the self-energy graph 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 branches that enter the self-energy graphs of lower order inside T_i . Note that the branches λ in the first product are the branches external to T (but not entering T), while the ones in the second product are the branches internal to T (*i.e.* branches connecting to nodes in \tilde{T}).

We can then remark that when η_T varies in the complex disk $|\eta_T| \leq 2^{n_T}$ the divisors of the branches that enter the inner self-energy graphs T' (of any height) do not exceed in modulus $2^{n_{T'}}$. Therefore we can bound the quantities $F(T_i, \underline{\nu}_{v_i}, \underline{\nu}_{v'_i})$ via the inductive bound and obtain that (8.4.5) is valid also for trees of height p which contain only one maximal self-energy graph.

Since the case in which there are many self-energy graphs of height p is clearly reducible to the case in which there is only one such cluster the conclusion is the validity of the inequality (8.4.5) in general. It would also be 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.

Hence substituting (8.3.2) into (8.4.5) 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 (8.3.2), taken from the factors with $T = T_i$, (note that there are always enough -1 's).

Hence the product (8.4.5) is bounded by

$$e8.4.7 \quad \frac{1}{k!} (C^2 J^{-1} f_0 N^2)^k e^{2k} 2^{12k} \prod_{n \leq 0} 2^{-8nk} E^{-1} 2^{n_n} \leq \frac{1}{k!} B_0^k, \quad (8.4.7)$$

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 (8.4.7) by the number of tree shapes for ϑ , ($\leq 2^{2k} k!$), by the number of ways of attaching mode labels, ($\leq (3N)^{\ell k}$),

by the number of ways of contacting the tensor labels, ($\leq \ell^k$), so that we can bound $|\underline{h}_{\underline{\nu}}^{(k)}|$ by

$$\varepsilon_0^{-k} \equiv (b_\ell J^{-1} C^2 f_0 N^{2+\ell} e^{cN} e^{\xi N})^k, \quad (8.4.8)$$

with b_ℓ suitably chosen. ■

Problems for §8.4 (*Complements to KAM theory*)

Q8.4.1 [8.4.1]: (*Diophantine spacing*)

Let $\underline{\omega}$ be a Diophantine vector, cf. (8.1.6), with constants C and exponent τ . Consider the set B_n of the values of $|\underline{\omega} \cdot \underline{\nu}|$ as $\underline{\nu} \in \mathbb{Z}^\ell$ varies in the set $0 \leq |\underline{\nu}| < (2^{n+3})^{-1/\tau}$, with $n = 0, -1, -2, \dots$. The sets B_n verify the inclusion relation $B_n \subset B_m$ if $m < n$. Check that the spacing between the points of each of the sets B_n is at least $2^\tau (2(2^{n+3})^{-1/\tau})^{-\tau} \geq 2^{n+3}$. Check also that $x \in B_n$ is such that $x > 2^{n+3}$, if $x \neq 0$.

Q8.4.2 [8.4.2]: In the context of problem [8.4.1] let, more abstractly, B_n , $n = 0, -1, \dots$, be a sequence of sets such that (i) $0 \in B_n$, (ii) $B_n \subset B_m$ if $m < n$ and (iii) the spacing between the points in B_n is at least 2^{n+3} (the latter will be the *spacing property*). Show that there exists a sequence $\gamma_0, \gamma_{-1}, \dots$ with $\gamma_p \in [2^{p-1}, 2^p]$ such that

$$|x - \gamma_p| \geq 2^{n+1} \quad \text{if } n \leq p \leq 0 \quad \text{and } x \in B_n, \quad (8.4.9)$$

for all $n \leq 0$. (*Hint*: The reader who does not want to build the proof can read it in the appendix (8.4).)

Q8.4.3 [8.4.3]: (*Arithmetic properties related to the Diophantine property*)

In the context of problem [8.4.2] let $C_0 = 2^\tau C$ and $\underline{\omega} = C_0 \underline{\omega}_0$ where $\underline{\omega}_0$ is a Diophantine vector with constant C and exponent τ ; show that it is possible to find a sequence γ_p with $\gamma_p \in [2^{p-1}, 2^p]$, $p \leq 0$, such that

$$|\underline{\omega} \cdot \underline{\nu}| - \gamma_p \geq 2^{n+1} \quad \text{if } 0 < |\underline{\nu}| \leq (2^{n+3})^{-\tau^{-1}} \quad (8.4.10)$$

for all $n \leq 0$ and for all $p \geq n$. Furthermore $|\underline{\omega} \cdot \underline{\nu}| \neq \gamma_n$, for all $n \leq 0$. (*Hint*: This is implied by the arithmetic result of problem [8.4.2].)

Q8.4.4 [8.4.4]: (*Elimination of the assumption of strongly Diophantine property*, [GG95])

Starting from the result of problem [8.4.3] consider the notion of self-energy subgraph introduced in Section 8.3 assuming rotation vectors $\underline{\omega}$ satisfying the conditions (8.3.2). Adapt the notion to the case in which only the Diophantine property (8.1.6) is assumed. To define the scales let (as above) $\underline{\omega} = 2^\tau C \underline{\omega}_0$; a propagator will be said to be on scale n if one has $\gamma_{n-1} \leq |\underline{\omega} \cdot \underline{\nu}| < \gamma_n$, for $n \leq 0$, and set $n = 1$ if $\gamma_0 < |\underline{\omega} \cdot \underline{\nu}|$. Show that the line scales do not change as a line entering a self-energy graph is detached and reattached to another node of the subgraph. (*Hint*: The shift of the lines exemplified in Fig.(8.3.1) causes a change of the size of the propagator by at most $\gamma_{n_1} + \gamma_{n_2} + \dots$ (instead of $2^{n_1} + 2^{n_2} + \dots$, see (8.3.5)), but such a quantity is still bounded by 2^{n+1} . By proceeding as in Section 8.4, we see that the product of the propagators is holomorphic in $\eta_T \equiv \underline{\omega} \cdot \underline{\nu}_{\lambda_T}$ for $|\eta_T| < \gamma_{n_T-3 \dots}$)

Q8.4.5 [8.4.5]: (*Extension of Siegel-Bryuno bound*)

Imagine that the perturbation $f(\underline{\alpha})$ is not a trigonometric polynomial but it is just analytic in $\underline{\alpha}$. Extend the definition of self-energy graph by changing the condition that the number of branches is $\leq E 2^{-n\eta}$, cf. (8.3.3), into

$$M(T) \stackrel{\text{def}}{=} \sum_{v \in V(T)} |\underline{\nu}_v| \leq E 2^{-n\eta}, \quad (8.4.11)$$

where $E = 2^{-3\eta}$. Show that a lemma analogous to lemma (8.3.1) holds in the form: for all trees ϑ one has

$$e_{8.4.12} \quad N_n(\vartheta) \leq \frac{4M(\vartheta)}{E 2^{-\eta n}} + \sum_{\substack{T \in \mathcal{T}(\vartheta) \\ n_T = n}} (-1 + q_T), \quad M(\vartheta) = \sum_{v \in V(\vartheta)} |\underline{\nu}_v|, \quad (8.4.12)$$

with $E = 2^{-3\eta}$ and $\eta = 1/\tau$. Check that the above statement implies as a particular case lemma (8.3.1) if the perturbation is a trigonometric polynomial. (*Hint*: The proof is again by induction, and it is essentially identical to that of lemma (8.3.1), apart from the obvious changes in the notations.)

Q8.4.6 [8.4.6]: (*Non-divergent graphs in the case of analytic perturbations*)

Check that the bound in problem [8.4.5] can be directly used to obtain the result of proposition (8.1.1) under the only assumption that the perturbation is analytic. (*Hint*: In (8.4.3) the bound f_0^k has to be replaced by the bound $F^k \prod_{v \in V(\vartheta)} e^{-\kappa|\underline{\nu}_v|}$ for some $F, \kappa > 0$ (which comes from the analyticity assumption on V). The product $\prod_{n \leq 0} 2^{-2nN_n}$ will be replaced by $2^{-2n_0k} \prod_{n \leq n_0} 2^{-2nN_n}$ where n_0 is an arbitrary negative integer. Of course it is no longer possible to bound the product $\prod_{v \in V(\vartheta)} |\underline{\nu}_v|^{m_v+1}$ with N^{2k} as in (8.4.3), but we can use that $\frac{1}{m!} |\underline{\nu}|^m \leq \left(\frac{8}{\kappa}\right)^m$ and that the number of trees with given $\{m_v\}$ and without labels is bounded by $k! / \prod_v m_v!$; see problem [8.2.1]. From the above bound by $F^k \prod_{v \in V(\vartheta)} e^{-\kappa|\underline{\nu}_v|}$ we can extract a factor $\leq \exp[-\kappa|\underline{\nu}|/2]$, with $\underline{\nu} = \underline{\nu}_{\lambda_0}$, if λ_0 is the root branch, while the remaining factor $\exp[-\kappa M(\vartheta)/2]$ (with $M(\vartheta) \equiv \sum_v |\underline{\nu}_v|$, see (8.4.12)), will be kept as such. Collecting together the bounds the sum over all non divergent graphs (*i.e.* without self-energy subgraphs) is

$$F^k e^{-\frac{1}{2}\kappa|\underline{\nu}|} \sum_{\{\underline{\nu}_v\}_{v \in V(\vartheta)}} \left(\prod_v e^{-\frac{1}{2}\kappa|\underline{\nu}_v|} |\underline{\nu}_v| \right) 2^{-2n_0k} \left(\frac{8}{\kappa}\right)^{2k} \prod_{n \leq n_0} 2^{\sum_v |\underline{\nu}_v|} 4E^{-1} 2^{n/\tau}$$

can be bounded by a constant to the power k if n_0 is so chosen that $\sum_{n \leq n_0} 4(\log 2) E^{-1} 2^{n/\tau} < \kappa/4$, which assures the summability over the Fourier labels.)

Q8.4.7 [8.4.7]: (*Cancellation in the case of analytic perturbations*)

Prove that, by considering the quantities $\eta_T \stackrel{def}{=} \underline{\omega} \cdot \underline{\nu}_{\lambda_T}$, cf. lines preceding (8.4.1), as independent variables, the product of propagators is holomorphic in η_T for $|\eta_T| < \gamma_{n_T-3}$. Check that this can be combined with the results of problems [8.4.5] and [8.4.6], to proceed as in Section §8.4 and to achieve a proof of proposition (8.1.1) under the only assumption the perturbation $V(\underline{\alpha})$ is analytic in $\underline{\alpha}$. (*Hint*: If λ is a branch on scale n_T , one has $\gamma_{n_T-1} \leq |\underline{\omega} \cdot \underline{\nu}_\lambda| < \gamma_{n_T}$; then $|\underline{\omega} \cdot \underline{\nu}_\lambda^0| > 2^{n+3}$ implies $|\underline{\omega} \cdot \underline{\nu}_\lambda| > 2^{n+3} - 2^n > 2^{n+2}$, so that $n_T \geq n+3$. On the other hand, if $n_T > n+3$, *i.e.* $n_T = n+m$, for some $m > 3$, one has $|\underline{\omega} \cdot \underline{\nu}_\lambda^0| > \gamma_{n_T-1} - \gamma_n$, because the scales of all the branches do not change, so that it follows that, for $|\eta_T| < \gamma_{n_T-3}$, one has $|\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T| \geq \gamma_{n_T-1} - \gamma_n - \gamma_{n_T-3} \geq (2^{n_T-2} - 2^{n_T-m}) - \gamma_{n_T-3} \geq (2^{n_T-3} + 2^{n_T-4} + \dots + 2^{n_T-m+1}) - 2^{n_T-3} \geq 2^{n_T-4}$; otherwise, if $n_T = n+3$, one has $|\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T| > 2^{n+3} - \gamma_{n_T-3} \geq 2^{n_T-1}$, for $|\eta_T| < \gamma_{n_T-3}$. Therefore we can conclude that, for $|\eta_T| < \gamma_{n_T-3}$, the quantity $|\underline{\omega} \cdot \underline{\nu}_\lambda^0 + \sigma_\lambda \eta_T|$ does not become smaller than 2^{n_T-4} .)

Q8.4.8 [8.4.8]: (*Lindstedt series for non-even perturbations*)

Check that the assumption $f_{\underline{\nu}} = f_{-\underline{\nu}}$ can be avoided in order to obtain the formal solubility of the equations of motions (8.1.12). (*Hint*: We can prove by induction on the tree order that no branch can have a vanishing momentum. Consider all contributions arising from the trees $\vartheta \in \Theta_{k,0}$ deprived of the root branch² (so that $\text{Val}(\vartheta)$ is a well-defined quantity by the inductive hypothesis): we group together all trees obtained from

N8.4.2

² This means that the propagator of the root line is replaced with 1.

each other by shifting the root branch, *i.e.* by changing the node which the root branch exits and orienting the arrows in such a way that they still point toward the root. We call $\mathcal{F}(\vartheta)$ such a class of trees (here ϑ is any element inside the class). The values $\text{Val}(\vartheta')$ of such trees $\vartheta' \in \mathcal{F}(\vartheta)$ differ because (1) there is a factor $i\underline{\nu}_v$ depending on the node v to which the root branch is attached, and (2) some arrows change their directions. More precisely, when the root branch is detached from the node v_0 and reattached to the node v , if $\mathcal{P}(v_0, v) = \{w \in V(\vartheta) : v_0 \succeq w \succeq v\}$ denotes the path joining the node v_0 to the node v , all the momenta flowing through the branches λ along the path $\mathcal{P}(v_0, v)$ change their signs: nevertheless the propagators do not change. Then by summing the values of all possible trees inside the class $\mathcal{F}(\vartheta)$ we obtain a common value times i times $\sum_{v \in V(T)} \underline{\nu}_v$, and the sum gives zero.)

Q8.4.9

[8.4.9]: (*KAM theorem for non-even perturbations*)

Check that the assumption $f_{\underline{\nu}} = f_{-\underline{\nu}}$ can be avoided in the proof of the KAM theorem. (*Hint:* Given a tree ϑ with a self-energy graph T , set $\delta = \underline{\omega} \cdot \underline{\nu}$, with $\underline{\nu} = \underline{\nu}_{\lambda_T^2}$. Consider all trees which can be obtained by shifting the entrance node of the entering branch λ_T^2 : the sum of the values of all the so obtained trees is zero when the propagators of the branches in $\Lambda(T)$ are computed at $\delta = 0$, by the same argument as the one used in Section 8.1. Then consider the corresponding contributions to first order in δ : one has to derive the product of propagators of the branches $\lambda \in \Lambda(T)$ and compute it at $\delta = 0$. The derivative gives $|\Lambda(T)|$ terms, each of which has the derivative acting on a propagator G_λ , while all the other propagators $G_{\lambda'}$, $\lambda' \neq \lambda$, are not derived. The branch λ divides $V(T)$ into two disjoint set of nodes V_1 and V_2 , such that λ_T^1 exits from a node inside V_1 and λ_T^2 enters a node inside V_2 : if $\lambda = \lambda_v$ one has $V_2 = \{w \in V(T) : w \preceq v\}$ and $V_1 = V(T) \setminus V_2$. By setting $\underline{\nu}_1 = \sum_{v \in V_1} \underline{\nu}_v$ and $\underline{\nu}_2 = \sum_{v \in V_2} \underline{\nu}_v$, one has $\underline{\nu}_1 + \underline{\nu}_2 = \underline{0}$. Then consider the families $\mathcal{F}_1(\vartheta)$ and $\mathcal{F}_2(\vartheta)$ of trees obtained as follows: $\mathcal{F}_1(\vartheta)$ is obtained from ϑ by detaching λ_T^1 then reattaching to all the nodes $w \in V_1$ and by detaching λ_T^2 then reattaching to all the nodes $w \in V_2$, while $\mathcal{F}_2(\vartheta)$ is obtained from ϑ by reattaching the branch λ_T^1 to all the nodes $w \in V_2$ and by reattaching the branch λ_T^2 to all the nodes $w \in V_1$. As a consequence of such an operation the arrows of some branches $\lambda \in \Lambda(T)$ change their directions: this means that for some branch λ the momentum $\underline{\nu}_\lambda$ is replaced with $-\underline{\nu}_\lambda$. The derived propagator G_λ can change sign, while the non-derived propagators $G_{\lambda'}$, with $\lambda' \neq \lambda$, do not change by parity: one has a different sign for the trees in $\mathcal{F}_1(\vartheta)$ with respect to the trees in $\mathcal{F}_2(\vartheta)$. Then by summing over all the possible trees in $\mathcal{F}_1(\vartheta)$ we obtain a value $i^2 \underline{\nu}_1 \underline{\nu}_2$ times a common factor, while by summing over all the possible trees in $\mathcal{F}_2(\vartheta)$ we obtain $-i^2 \underline{\nu}_1 \underline{\nu}_2$ times the same common factor, so that the sum of two sums gives zero.)

Appendix 8.4: Weakening the strong Diophantine condition.

Here we prove the statement in problem [8.4.2]. Note that (8.4.10) is very similar to the second condition introduced in (8.3.2) to define the *strong Diophantine condition*. Our point is that all that is really needed (see the following discussion) to prove the KAM theorem are (8.1.6) (Diophantine property) and (8.4.10): the latter is a simple arithmetic property which is in fact a consequence of the first.

Proof: Note that if $\gamma_p \in [2^{p-1}, 2^p] \equiv I_p$ the equations (8.4.9) are obviously verified for $n > p - 3$, hence we can suppose $p \geq n + 3$.

Fix $p \leq 0$ and let $G = [a, b]$ be an interval verifying what we shall call

below the *property* \mathcal{P}_n :

$$\begin{aligned} \mathcal{P}_n : \quad & |x - \gamma| \geq 2^{m+1} \quad \text{for all } n \leq m \leq -3, \\ \text{eA8.4.1} \quad & \text{for all } x \in B_n \text{ and for all } \gamma \in G, |G| \geq 2^{n+1}. \end{aligned} \tag{A8.4.1}$$

Let $G_{p-3} \equiv [2^{p-1}, b_{p-3}]$, with $b_{p-3} \in [2^{p-1}, 2^p]$, be a *maximal* interval verifying property \mathcal{P}_{p-3} (note that G_{p-3} exists because $x \in B_{p-3}$, $x \neq 0$ implies $x \geq 2^p$ by the spacing property, and it is $b_{p-3} \geq 2^{p-1} + 2^{p-2}$).

Assume inductively that the intervals $[a_n, b_n] = G_n$ can be so chosen that $G_{n'} \subseteq G_{n''}$ if $n' < n''$ and G_n is maximal among the intervals contained in G_{n+1} and verifying the property \mathcal{P}_n .

If we can check that the hypothesis implies the existence of an interval $G \subseteq G_n$ verifying \mathcal{P}_{n-1} we shall be able to define G_{n-1} to be a maximal interval among the ones contained in G_n and with the property \mathcal{P}_{n-1} : in case of ambiguity we shall take G_{n-1} to coincide with the rightmost possible choice. And as a consequence we shall be able to define $\gamma_p = \lim_{n \rightarrow -\infty} b_n$, which will verify (8.4.10).

To check the existence of $G \subseteq G_n$ verifying \mathcal{P}_{n-1} we consider first the case in which B_{n-1} has one and only one point x in G_n . If $|G_n| \geq 2^{n+2}$ and x is in the first half of G_n we can take, by the spacing property, $G = [x + 2^n, \min\{b_n, x + 2^{n+2} - 2^n\}]$; if it is in the second half we take $G = [\max\{a_n, x - 2^{n+2} + 2^n\}, x - 2^n]$.

If, on the other hand, $|G_n| < 2^{n+2}$ it is $G_n \subset G_{n+1}$ strictly and, furthermore, the interval $(x - 2^{n+2}, x + 2^{n+2})$ does not contain points of B_{n-1} other than x itself (by the spacing property). The strict inclusion implies that there is a point $y \in B_n$ at distance exactly 2^{n+1} from G_n (recall the maximality of G_n).

Suppose that x is in the first half of G_n and $y < a_n$, *i.e.* $y = a_n - 2^{n+1}$; then $x - y < 2^{n+1} + 2^{n+1}$ contradicting the spacing property. Hence $y > b_n$, *i.e.* $y = b_n + 2^{n+1}$: in such case it cannot be, again by the spacing property, that $x + 2^{n+2} > y = b_n + 2^{n+1}$, so that $b_n - x \geq 2^{n+1}$ and we can take $G = [b_n - 2^n, b_n]$. If x is in the second half the roles of left and right are exchanged.

This completes the analysis of the case in which only one point of B_{n-1} falls in G_n . The cases in which either no point or at least two points of B_n fall in G_n are analogous but easier. If two consecutive points $x < y$ of B_{n-1} fall inside G_n we must have $y - x \geq 2^{n+2}$ by the spacing property: hence $G = [x + 2^n, y - 2^n] \subset G_n$ enjoys the property \mathcal{P}_{n-1} . If no point of B_{n-1} falls in G_n let $y \in B_{n-1}$ be the closest point to G_n ; if its distance to G_n exceeds 2^n we take $G = G_n$. Otherwise suppose that $y > b_n$: the spacing property implies that the interval $(y - 2^{n+2}, y)$ is free of points of B_{n-1} . Hence if $a = \max\{a_n, y - 2^{n+2} + 2^n\}$, $b = y - 2^n$ then $G = [a, b]$ has the property \mathcal{P}_{n-1} . If, instead, $y < a_n$ we set $a = y + 2^n > a_n$ and $b = \min\{b_n, y + 2^{n+2} - 2^n\}$ and $G = [a, b]$ enjoys the property \mathcal{P}_{n-1} . ■

Bibliographical note for §8.1, §8.2, §8.3 and §8.4

The stability problem of Hamiltonian motions is ancient; introductions to

this problem are the first two chapters of the book by Moser, [Mo63], the article [Mo78] and the article by Arnold, [Ar63], from p. 85 to p. 99. For a “classical” proof of the KAM theorem see Appendix 34 of [AA68] or Section 12 in [Ga82]. More recently a new proof of the theorem has been derived by Eliasson, [El96]: we have followed the interpretation given to it in [Ga94]. The proof presented in this section deals with a special case but it can be done in full generality along the same lines, see [GM96] and Section §8.4. The technique used is inspired to the perturbation theory in quantum field theory and renormalization group, [Ga01].

The analytic case considered in this section has been studied, in great generality, by Kolmogorov and then applied to celestial mechanics by Arnold. The case in which $V(\underline{\alpha})$ is only assumed differentiable has been studied by Moser: the methods of this section can also be applied to some differentiable cases, see [BGGM98].

As remarked in problem [8.4.4] almost all $\underline{\omega}$ verify (8.3.2) for some C and some $\tau > \ell - 1$, with a sequence γ_p that can be prescribed *a priori* as $\gamma_p = 2^p$. This, however, leaves out important cases like a quadratic irrational as rotation number in $\ell = 2$. And it has the very unfortunate drawback of being non-constructive, as the set of full measure of the $\underline{\omega}$ verifying the strong Diophantine condition is obtained by abstract arguments (*e.g.* the Borel-Cantelli lemma), see problem [8.1.7]. Nevertheless considering strongly Diophantine vectors is natural as it leads to the simplified proof of the previous section, eliminating the “secondary” difficulty we have discussed here.

The Siegel–Bryuno bound is not the only way to approach the problem of bounds on products of small divisors. Instead one can use that, for trees ϑ without self-energy graphs one has

$$\left| \prod_{\lambda \in \Lambda(\vartheta)} G_\lambda \right| \leq D^k \left(\sum_{v \in V(\vartheta)} |\mathcal{L}_v| \right)^{-1} \prod_{v \in V(\vartheta)} |\mathcal{L}_v|^{3\tau},$$

for a suitable constant D . This is an extension of a bound due to Siegel and Eliasson. and it leads to the same results while it can be fruitfully applied also to related problem for which the bound in lemma (8.3.1) is not well suited. Such a bound would allow us to obtain the same conclusions: for further details we refer to the original papers (see [El96] and [BGGM98]).

The “tree expansion technique” can be applied to most perturbation theory problems providing a simple and unifying way of seeing them (however simplicity is not objective and this point of view does not seem to be shared by everybody).

A further natural question could be what happens if the perturbation is not an even function or, even more, if it depends also on the action variables.

If the function $V(\underline{\alpha})$ is not even, than the formal solubility of the equations of motions and the second order cancellation of the self-energy graphs does not follow anymore from parity considerations. In such a case the cancellation mechanism is a little more complicated, and it requires also

the shifting of the exiting branch of the self-energy graphs (as originally remarked in [CF94]; see [GM96] for an implementation within a formalism closer to the one described here).

The generalization to perturbations depending also on the action variables requires some minor extensions of the cancellations described in Section §8.1, which can be found in [CF96] and in [GM96]; also the case of unperturbed Hamiltonians $K(\underline{A})$ with a more general dependence on the action variables can be dealt with provided that one has $\det \partial_{\underline{A}}^2 K(\underline{A}) \neq 0$ (*anisochrony condition*). But no real new difficulty arises; we shall not discuss further such a case here, for which we refer to the original papers. For a discussion of the optimal conditions under which the KAM theorem can be formulated we refer to a recent exhaustive paper by Rüssmann [Ru01].