Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable hamiltonian systems. A review.*

Giovanni Gallavotti[1]

Abstract: *Rotators interacting with a pendulum via small, velocity independent, potentials are considered. If the interaction potential does not depend on the pendulum position then the pendulum and the rotators are decoupled and we study the invariant tori of the rotators system at fixed rotation numbers: we exhibit cancellations, to all orders of perturbation theory, that allow proving the stability and analyticity of the dipohantine tori. We find in this way a proof of the KAM theorem by direct bounds of the k–th order coefficient of the perturbation expansion of the parametric equations of the tori in terms of their average anomalies: this extends Siegel's approach, from the linearization of analytic maps to the KAM theory; the convergence radius does not depend, in this case, on the twist strength, which could even vanish ("twistless KAM tori"). The same ideas apply to the case in which the potential couples the pendulum and the rotators: in this case the invariant tori with diophantine rotation numbers are unstable and have stable and unstable manifolds ("whiskers"): instead of studying the perturbation theory of the invariant tori we look for the cancellations that must be present because the homoclinic intersections of the whiskers are "quasi flat", if the rotation velocity of the quasi periodic motion on the tori is large. We rederive in this way the result that, under suitable conditions, the homoclinic splitting is smaller than any power in the period of the forcing and find the exact asymptotics in the two dimensional cases (e.g. in the case of a periodically forced pendulum). The technique can be applied to study other quantities: we mention, as another example, the homoclinic scattering phase shifts.*

Key words: *KAM, homoclinic points, cancellations, perturbation theory, classical mechanics, renormalization*

§1 Introduction

We discuss the invariant tori and the splitting of their homoclinic stable and unstable manifolds for a special class of quasi integrable hamiltonian systems. We apply and extend, in the considered class, the important ideas of Melnikov and Eliasson respectively on the theory of low dimensional invariant tori and their manifolds, [Me], and on the cancellations behind the convergence of the formal perturbation series for the invariant tori of maximal dimension, [E], *i.e.* for the KAM tori ([K],[A2],[M]). We also point out the analogy of the methods with those used in quantum field theory, particularly in the renormalization group approaches, [G2].

The ideas of Melnikov and Eliasson have been around since quite a while: but it seems that few realized their importance; a possible explanation is that the original papers are plagued by excessive generality. Assuming that this is the reason it is desirable to show what they imply in a "simple" case.

There are, however, no really simple cases (as it is well known).

In his book on mathematical physics Thirring made an attempt to find a "simple" model to explain the KAM theory, [T]. Although his models are essentially as difficult to treat as the most general hamiltonian system with twist, some aspects of the proofs are somewhat simpler and leave us with the feeling that something more can still be done. The class of models consists of a family of rotators (*i.e.* points on circles or, if one deems this too abstract, cylinders with fixed axis), say $l-1$ in number, interacting with a pendulum via a conservative force. The inertia moments J_j, $j=1,\ldots,l-1$, of the rotators form a matrix J which is diagonal. And they are supposed to be $J_j \ge J_0 > 0$, if J_0 is the inertia of the pendulum. Actually Thirring has no pendulum, just a system of rotators: this can be viewed as a special case in which the pendulum and

---

* This paper is deposited in the archive `mp_arc@math.utexas.edu`: to get a TeX version of it, send an empty E-mail message to this address and instructions will be sent back.

1 E-mail: `gallavotti%40221.hepnet@lbl.gov`: Dipartimento di Fisica, Università di Roma, "La Sapienza", P. Moro 5, 00185 Roma, Italia.

the rotators do not interact. Also he considers a somewhat more general model than model 1), by allowing the matrix J to depend on $\vec{\alpha}$.

More formally: we shall consider the l degrees of freedom hamiltonian $H_\mu \equiv H_0 + \mu f$ given by one of the following two expressions:

$$\begin{aligned}
&1)\quad \vec{\omega}\cdot\vec{A} + \frac{1}{2}J^{-1}\vec{A}\cdot\vec{A} + \mu \sum_{0<|\vec{\nu}|\le N} f_{\vec{\nu}} \cos\vec{\alpha}\cdot\vec{\nu}\\
&2)\quad \vec{\omega}\cdot\vec{A} + \frac{1}{2}J^{-1}\vec{A}\cdot\vec{A} + \frac{I^2}{2J_0} + g^2 J_0(\cos\varphi - 1) + \mu \sum_{\substack{|\nu|\le N\\ \vec{\nu}\neq\vec{0}}} f_\nu \cos(\vec{\alpha}\cdot\vec{\nu} + n\varphi)
\end{aligned}\tag{1.1}$$

where $(I,\varphi) \in R^2, (\vec{A},\vec{\alpha}) \in R^{2(l-1)}$ are canonically conjugated variables, $\vec{\omega} \in R^{l-1}$, $\nu \equiv (n,\vec{\nu}) \in Z^l$, $|\nu| = |n| + |\vec{\nu}| = |n| + \sum_{i=1}^{l} |\nu_i|$ and $J, J_0 > 0$ (respectively, "rotator's inertia moment" and "pendulum inertia moment"), $g > 0$ (g^2 is the "gravity"), f_ν are fixed constants. And $\vec{\omega}, \mu$ are parameters.

To avoid trivial, or lengthy, comments here and there we suppose that $J \ge J_0$ with $J_0 > 0$ fixed forever, setting a scale for the size of the inertia moments, (we do not take it 1 as we are convinced that it is better not to fix units of measure), and that $f_{\vec{0}} = 0$ or $f_{n,\vec{0}} \equiv 0$, for all n, (for what concerns the trivial comments), and $|f_\nu| \le J_0\vec{\omega}^2$ (for the lengthy ones): this will be clearly not restrictive.

Model 2) will be called *rotator–pendulum model*, or *simple resonance* model (see below for motivation) or *Arnold model*, and model 1) will be called *rotator model* or *Thirring model*: they deserve a name, because they are a very interesting class with many properties that cannot be discussed here, the analysis of a few (non trivial) of which was begun in [A] for the first and in [T] for the second.

Model 1) has many applications in plasma physics, and celestial mechanics. Model 2) has great relevance for the analysis of the breakdown of invariant tori and the corresponding universal behaviour.

In this paper we suppose a priori that:

Hypothesis H: *the parameters $\vec{\omega}, \mu$ verify, in general:*

$$\vec{\omega} = \frac{\vec{\omega}_0}{\sqrt{\eta}}, \qquad |\mu| \le b\eta^Q, \qquad \eta \le 1 \tag{1.2}$$

with Q and b^{-1} which will be restricted to be large enough in the course of the analysis. In the case of model 1) we shall also fix $\eta = 1, Q = 0$.

and:

Hypothesis H': *$\vec{\omega}_0$ is a diophantine vector, i.e. :*

$$|\vec{\omega}_0\cdot\vec{\nu}| \ge \tfrac{1}{C_0|\vec{\nu}|^\tau}\ , \qquad \textit{for all } \vec{0} \neq \vec{\nu} \in Z^{l-1} \tag{1.3}$$

for some diophantine constant C_0 and some diophantine exponent $\tau > 0$.

A natural *energy scale* for the model 1) will be $E = J_0\vec{\omega}^2$ and for model 2) one can take $E = J_0 g^2$.

We prove the following theorem for model 1):

Theorem 1 *(Twistless KAM): Consider the solution flow of the hamiltonian equations of the Thirring model 1), under the assumptions H,H' above. Then there exists a constant $b > 0$ and a function (dimensionless) $\vec{x}(\vec{\psi},\mu)$ of $\vec{\psi} \in T^{l-1}, \mu \in R$, holomorphic in μ and $\vec{\psi}$, in the complex domains $|\mu| < e^{-N\xi}b$ and $|\operatorname{Im}\psi_j| < \xi$, parameterized by $\xi > 0$, and bounded by 1, such that the:*

$$\vec{\alpha} = \vec{\psi} - \mu\frac{E}{J}\sum_{\vec{\nu}\neq\vec{0}} \vec{x}_{\vec{\nu}}(\mu)\frac{\sin\vec{\nu}\cdot\vec{\psi}}{(\vec{\omega}\cdot\vec{\nu})^2}, \qquad \vec{A} = -\mu E \sum_{\vec{\nu}\neq\vec{0}} \vec{x}_{\vec{\nu}}(\mu)\frac{\cos\vec{\nu}\cdot\vec{\psi}}{(\vec{\omega}\cdot\vec{\nu})} \tag{1.4}$$

with $\vec{\psi} \in T^{l-1}$, $E \equiv J_0\vec{\omega}^2$, are the parametric equations of an invariant torus (for the flow) on which the flow is quasi periodic with spectrum $\vec{\omega}$ and the angles $\vec{\psi}$ are its average anomalies: i.e. the flow is $\vec{\psi} \to \vec{\psi} + \vec{\omega}t$, so that the angles ψ_j rotate at constant velocity ω_j. The constant b depends only on $J_0, \vec{\omega}^2, C_0, \tau, N, l$; and $\vec{x}_{\vec{\nu}}(\mu) = \vec{x}_{-\vec{\nu}}(\mu)$ denotes the Fourier transform of $\vec{x}(\vec{\psi},\mu)$.

Remarks:

1) The fact that the theorem is *uniform* in the size of $t_w = \det J^{-1}$, provided $J \geq J_0$, means that it holds even if the twist $t_w = 0$. This is the reason we call the above theorem "twistless" KAM theorem. We prove it, under a slightly stronger diophantine hypothesis (see (7.1)), by finding a bound on the k–th order of the μ–expansion of the function $\vec{x}$: see §7 and Appendices A3,A4.
2) If some or all the inertia moments J_j are $+\infty$ the above theeorem is an easy consequence of the classical KAM theorem (or better of its proof); see comments around (4.16) for the case $J_j \equiv +\infty$. However the bound b obtained via the classical proof depends on the twist rate, *i.e.* on the maximum among the J_j *which are not* $+\infty$, and diverges as the rate approaches 0. This non uniformity is quite surprising, but it is an artifact of the classical proof (as a direct careful analysis of the latter also shows).

Our second class of results concerns the simple resonance model, 2) in (1.1).
The $l = 2$ and $J = +\infty$ case will *not* be excluded and corresponds to the "pendulum in a periodic force field"; if $l > 2$ we take $J < +\infty$, for simplicity, to be a constant but our results can be extended, essentially unchanged, to the case in which J is a diagonal matrix with the first element zero and the others positive (in fact one can take J to be a $(l-1) \times (l-1)$ symmetric matrix with the first row and columns zero and with rank $l-2$). This is also implicitly shown, in the framework of the classical proofs, by the proof of KAM theorem for model 1) developed in [T], if particularized to model 1): it yields results uniform in the twist rate. This is probably the case, as well, for the original proof of Kolmogorov, [K], if particularized to model 1).

Remark: To motivate the name of "simple resonance model" and hypothesis H one should recognize that model 2) presents the "basic" structure (after a few simple canonical changes of variables) of a perturbation of a completely integrable system, near a "simple resonance" (see [BG], [CG]). Suppose that the original (*i.e.* before the above mentioned change of variables) unperturbed hamiltonian is non degenerate, and η, $0 < \eta \leq 1$, is a measure of the size of the original perturbation. Then, provided η is small enough, one finds that the parameters $\vec{\omega}$ and μ verify: $\vec{\omega} \equiv \vec{\omega}_0 \eta^{-1/2}$, $|\mu| \leq \eta^{Q_0}$, where $Q_0 > 1$ can be prefixed at the beginning of the construction of the change of coordinates, . This is, perhaps, the strongest motivation to study the hamiltonians (1.1), model 2), with $\vec{\omega}$ given proportional to $1/\sqrt{\eta}$.

For $\mu = 0$, the hamiltonian equations generated by (1.1), (*i.e.* $\dot{I} = -\partial_\varphi H_\mu$, $\dot{\varphi} = \partial_I H_\mu$, $\dot{\vec{A}} = -\partial_{\vec{\alpha}} H_\mu$, $\dot{\vec{\alpha}} = \partial_{\vec{A}} H_\mu$), admit $(l-1)$–dimensional invariant tori:

$$\mathcal{T}_0 \equiv \{I = 0 = \varphi\} \times \{\vec{A} \equiv \vec{A}^0 \ , \vec{\alpha} \in T^{l-1}\} \tag{1.5}$$

possessing homoclinic stable and unstable manifolds, *called "whiskers"*. The manifolds equations are:

$$W_0^{\pm} \equiv W_0 \equiv \{\tfrac{I^2}{2J_0} + g^2 J_0(\cos\varphi - 1) = 0\} \times \{\vec{A} \equiv \vec{A}^0 \ , \vec{\alpha} \in T^{l-1}\} \tag{1.6}$$

The integrability is reflected also by the degeneracy property that $W_0^+ \equiv W_0^-$.
Then, it follows "from KAM theory", [Me],[E],[CG], that "many" unperturbed tori around the torus $\vec{A}^0 = \vec{0}$ (including the one $\vec{A}^0 = \vec{0}$ itself) can be *continued analytically* (in μ), togheter with their whiskers, into invariant tori with the same $\vec{\omega}$, for all $|\mu| < b\eta^Q$ (if b is a suitable constant, explicitly computable in terms of a few parameters associated with H_0, f in (1.1), see for instance[CG]) and for Q large enough; we shall call such tori *persistent*. The determination of b, Q reqires going through an analysis very similar to that of the classical KAM theorem: hence we say that such tori and whiskers are "obtained by KAM analytic continuation".[2].

---

[2] If one wants to get an idea of the kind of numbers that might be involved here (which is *not logically necessary* for reading the present paper) one can look at [CG]. It follows from [CG] that Q can be taken to be 10: see (5.76) of [CG] where $E_0 = O(\eta^{-1/2})$ and all other parameters are η–independent; including C_0, which is not related to the C_0 of the present paper because in [CG] the quoted inequality was obtained under the hypothesis that the constant bearing there the same name, besides being a bound on the diophantine constant for $\vec{\omega}$, was also larger than g^{-1}. In the present case the g^{-1} is of order $O(1)$, while the diophantine constant for $\vec{\omega}$ is, by (1.2), (1.3), $C_0\eta^{1/2}$ so that in applying the quoted result to our case we are forced to take the constant C_0 in (5.76) of [CG] of order $O(1)$.
Note that in applying the results of [CG] to find an estimate for Q we apply the (5.76) and not the final result (5.90), which would give a more stringent condition ($Q = 71$), because we are using here only an intermediary result whose proof was

Remark: If $l=2$, however, one does not need an elaborated method and a direct elementary check of the persistence of the invariant tori, which in this case are periodic orbits, is possible, togheter with a reasonably straightforward estimate of b,Q. One finds, for instance, that for all $J\le+\infty$ the value $Q=1$ is sufficient, (hint: write the persistence condition as a fixed point equation for the Poincaré map of the periodic orbit).

In some applications the (1.2) is not a very strong condition because of the remark following (1.1) whereby the Q_0 mentioned there can be taken as large as wished, by finding suitable coordinates. In particular, if $Q_0=Q+1$ then the analyticity domain in μ contains, well inside, the value η^{Q_0} for η small enough.[3]

We shall always suppose that μ is taken small enough so that the invariant torus that we are considering is persistent: *i.e.* we shall always suppose that $|\mu|<b\eta^Q$ where Q,b are the constants mentioned in the discussion following the hypothesis (1.2), so that the above persistence properties hold (the values of the constants can be determined in terms of properties of H_0,f in (1.1), see for instance [CG] where such estimates are obtained).

We shall denote by $W^\pm_\mu$ and $\mathcal{T}_\mu$ the stable and unstable whiskers and, respectively, the whiskered tori obtained by the KAM analytic continuation (in μ). The stable and unstable whiskers $W^\pm_\mu$ are characterized by the fact that distance$(S^t_\mu X^\pm_\mu,\mathcal{T}_\mu)\to0$ exponentially fast as $t\to\pm\infty$; here S^t_μ is the hamiltonian flow generated by (1.1). The flow on the persistent whiskered tori can be described, in suitable coordinates and for all $|\mu|$ small (*i.e.* $|\mu|\le b\eta^Q$), by $\vec\psi\to\vec\psi+\vec\omega' t$, for suitable diophantine vectors $\vec\omega'$ *independent of* μ (*e.g.* $\vec\omega'=\vec\omega$ in the continuation of the unperturbed torus $\vec A=\vec0$, $I=0=\varphi$).

A legacy of the unperturbed situation (see (1.6)) is that the persistent whiskers $W^\pm_\mu$ arising from (1.6) are, for $|\mu|$ small enough, *graphs over the angles* $\vec\alpha\in T^{l-1}$ and φ, at least if $|\varphi|<2\pi-\delta$ for any prefixed $\delta>0$; hence they can be written as:

$$W^\pm_\mu=\{(I,\vec A,\varphi,\vec\alpha)=(I^\pm_\mu(\vec\alpha,\varphi),\vec A^\pm_\mu(\vec\alpha,\varphi),\varphi,\vec\alpha):\vec\alpha\in T^{l-1},|\varphi|<2\pi-\delta\}\tag{1.7}$$

for suitable real–analytic (in $(\vec\alpha,\varphi)$ *and* μ) functions $\vec A^\pm,I^\pm$. It is also not difficult to check that the *parity* (in $(\vec\alpha,\varphi)$ of (1.1) implies that $(\vec\alpha,\varphi)=(\vec0,\pi)$ is a homoclinic point, *i.e.* $(\vec0,\pi)\in W^+_\mu\cap W^-_\mu$ (see (4.18) below), *for all* μ small enough (so small that the above tori and whiskers can be proved to exist).

In this context, it is natural to *measure the splitting between* W^+_μ *and* W^-_μ at $\varphi=\pi$ and $\vec\alpha=\vec0$ by the quantity:

$$\delta(\vec\alpha)\equiv\det\partial_{\vec\alpha}(\vec A^+_\mu-\vec A^-_\mu)|_{\varphi=\pi}\tag{1.8}$$

and its $\vec\alpha$–derivatives at $\vec\alpha=\vec0$.

Using the theory normal forms ([N], see also [BG]), one can show, see [Nei], that if $\mu\le b\eta^Q$ then δ is *smaller than any power in* η *as* $\eta\to0$.

Here we use an algorithm derived in [CG], (see also §2 below), for the computation of the μ–expansion coefficients of the functions $\vec A^\pm,I^\pm$. And we derive the above smallness result by explicitly checking several interesting *cancellation mechanisms*, operating to all orders of perturbation theory, and which are behind the smallness of the splitting δ when $\mu\le b\eta^Q$, see also [E], [ACKR].

As a byproduct we obtain, and extend, the results of [Nei]. We obtain also the exact asymptotics in the cases $l=2$, (with $0<J\le+\infty$), see §8:

Theorem 2 *(Quasi flat homoclinic intersections): Let $l=2$, $J>J_0$, $J\le+\infty$; let N_0 be the degree in φ of the trigonometric polynomial f in (1.1); then, provided η is small enough:*

$$\delta(0)=\frac{\mu}{\eta^{N_0-1/2}}A_*\,e^{-\frac{\pi\omega_0}{2g\sqrt\eta}}(1+O(\sqrt\eta))\tag{1.9}$$

if $|\mu|\le\eta^q$ with $q>3N_0+5$, if $A_*=-g^{-1}(\frac{\omega_0}{g})^{2N_0-1}(f_{N_0,1}+f_{N_0,-1})\frac{4\pi(-1)^{N_0}}{(2N_0-1)!}$, provided $A_*\ne0$. And in fact $\delta(\alpha)$ is a holomorphic function of α for $|\operatorname{Im}\alpha|<\zeta$, for all $\zeta>0$, if η is small enough (depending on ζ). And:

$$|\delta(\alpha)|\le D_\zeta\frac{|\mu|}{\eta^{N_0-1/2}}e^{-\frac{\pi\omega_0}{2g\sqrt\eta}},\qquad|\operatorname{Im}\alpha|<\zeta\tag{1.10}$$

---

completed under the condition (5.76). Such result is the persistence of a single invariant torus with given rotation vector $\vec\omega$: the paper [CG] was concerned with the persistence of a whole family of invariant tori, and subject to the condition of lying on a surface of prefixed energy (which led to add to (5.76) two further conditions, (5.81), (5.85) in [CG]).

In any event the above bounds should not be taken too seriously from a quantitative viewpoint as they are very likely far from optimal (as no effort at all was devoted to obtaining numerically good bounds).

[3] *e.g.* $Q_0=11$ according to the just mentioned estimates in [CG].

for a suitable $D_\zeta > 0$ and for η small enough (depending on q, ζ).

Remarks:

1) The optimal result is probably $q > N_0 - \frac{1}{2}$. In Appendix A1 we sketch how to get easily a better result: $q > N_0 + \frac{9}{2}$ instead of the pessimistic one $q > 3N_0 + 5$ given above.
2) For completeness we quickly rederive, in §2 and §4, the recursive formulae for the whiskers (see [CG], appendix 13).
3) It will appear that our technique consists in representing the coefficient of the μ^k-th order contribution to $\delta(\vec{\alpha})$ as a time integral from $t = -\infty$ to $t = +\infty$. The estimates should follow if analyticity properties of the integrands, allowing the shift of the t integration to a region where the integrand is small, could be checked.
Such an idea has been present in the literature since a long time, and sometimes it led to errors. It is also behind the available studies of the asymptotic behaviour of $\delta(\vec{\alpha})$ as $\eta \to 0$ (and $\mu = \eta^Q$ with Q large enough).
In our representation of $\delta(\vec{\alpha})$ the above approach is not so straightforward: *we cannot prove the holomorphy of the integrand in the integral representation of the k-th order coefficient of the μ expansion of* $\delta(\vec{\alpha})$, which in fact *is not holomorphic* in any useful region for the above idea to work literally, for $k \geq 2$. Rather we show that the integrand can be written as a sum of a holomorphic term which is small, because of an analyticity argument like the one mentioned above, and of a non holomorphic term which, however, is "as small as needed" if the bounds on the coefficients of the lower orders are "as small as needed". Since the first order trivially has all the necessary properties, the result is obtained by induction.
4) It is worth mentioning a related property, discovered in [CG], that will not be needed nor discussed further in the present work but that is closely related to theorem 2.
In many cases in which a hamiltonian like model 2) arises in the analysis of a resonant motion of a slightly pertubed completely integrable system, it happens that the original unperturbed hamiltonian is strongly degenerate, *i.e.* it does not depend on all the action variables (a feature common in celestial mechanics). In such cases, that we shall call "degenerate", the vector $\vec{\omega}$ in (1.1) has (at least) one component which is of $O(\sqrt{\eta})$.
More physically one can say that the models 2) verifying (1.2),(1.3) contain two basic time scales, of order $O(\eta^{1/2})$ ("fast") and $O(1)$ ("slow") corresponding to the frequences $\vec{\omega}$ and g. But in the degenerate cases there are (at least) three time scales of order $O(\eta^{1/2})$, $O(1)$ and $O(\eta^{-1/2})$ ("secular"), as the very slow secular scale reflects that the perturbation of the original system removes its degeneracy to order $O(\eta)$, relative to the $\vec{\omega}$'s size.
It is remarkable that a second order computation along the lines discussed below, shows that *generically the splitting δ, with $\mu \equiv \eta^Q$, is of $O(\eta^M)$ for some $M > 0$: hence it is not smaller than any power in η.* The latter fact can be used to show, at least in some cases, the existence of *Arnold's drift and diffusion* for any $\eta > 0$ small enough for systems which are small perturbations of degenerate completely integrable systems, see [CG].
The latter problem is still completely open in the non degenerate case and probably requires some really new ideas for its solution.
5) The proof that we give in the case $l = 2$ and in the general case have different nature: in both cases we use explicitly some general results on the existence of $(l-1)$–dimensional tori, [M], in the form derived in [CG]. However we shall show that the $l = 2$ case is not really relying on such results. We think that also the $l > 2$ cases can be freed from its dependence on [CG]: the reason is the validity of theorem 1 and its proof: it should be possible to carry out a similar proof also for theorem 2, if $l > 2$.

§2 Recursive formulae

In this section we recall some basic facts from the KAM theory mentioned in the introduction and we derive from such facts simple recursive formulae for the functions $I^{\pm}_{\mu}$, $\vec{A}^{\pm}_{\mu}$ in (1.7) and their time evolution.

Let us consider, for concreteness, the μ dependence of the above mentioned (see comments following (1.6)) analytic continuation in the parameter μ of the unperturbed whiskers having $\vec{A}^0 = \vec{0}$. The unperturbed motion is simply:

$$X^0(t) \equiv (I^0(t), \vec{0}, \varphi^0(t), \vec{\alpha} + \vec{\omega} t) \tag{2.1}$$

where $(I^0(t),\varphi^0(t))$ is the separatrix motion, generated by the pendulum in (1.1) starting at, say, $\varphi=\pi$. If we call $P(I,\varphi)$ the unperturbed pendulum energy in (1.1), it is $P(I^0,\varphi^0)\equiv 0$ (see (1.6)).

As mentioned in the introduction, under the hypotheses H,H' (see §1) the unperturbed whiskers $W_0^\pm\equiv W_0$ persist for $|\mu|\le\mu_0\equiv b\eta^Q$ and can be analytically continued into whiskers $W_\mu^\pm$.

Let $X_\mu^\sigma(t;\alpha)$, $\sigma=\pm$, be the evolution, under the flow generated by (1.1), of the point on W_μ^σ given by $(I_\mu^\sigma(\vec\alpha,\pi),\vec A_\mu^\sigma(\vec\alpha,\pi),\pi,\vec\alpha)$ (see (1.7); from now on we shall fix $\varphi=\pi$, which amounts to studying the whiskers at the "Poincaré section" $\{\varphi=\pi\}$).

The mentioned analyticity in μ (see [CG], for instance, §5,§6) allows us to consider the Taylor series expansions of the whiskers equations; let:

$$X_\mu^\sigma(t)\equiv X_\mu^\sigma(t;\vec\alpha)\equiv\sum_{k\ge0}X^{k\sigma}(t;\vec\alpha)\mu^k=\sum_{k\ge0}X^{k\sigma}(t)\mu^k,\qquad\sigma=\pm\tag{2.2}$$

be the power series in μ of X_μ^σ, (convergent for μ small); note that $X^{0\sigma}\equiv X^0$ is the unperturbed whisker. We shall often not write explicitly the $\vec\alpha$ variable among the arguments of various $\vec\alpha$ dependent functions, to simplify the notations.

¿From KAM theory, it follows that the t–dependence of $X_\mu^\sigma(t)$ has the form:

$$X_\mu^\sigma(t)=X_\mu^\sigma(\vec\omega t,t)\equiv X_\mu^\sigma(\vec\omega t,t;\vec\alpha)\tag{2.3}$$

where $X_\mu^\sigma(\vec\psi,t;\vec\alpha)$ is a real analytic function, of *all* its arguments (μ included), which is periodic in $\vec\psi$ and $\vec\alpha$; furthermore it has a holomorphy domain:

$$\mathcal D\equiv\mathcal D_{\xi,\xi_0,K,\mu_0}\equiv\{|\,\mathrm{Im}\,\psi_i|<\xi,\quad|\,\mathrm{Im}\,\alpha_i|<\xi_0,\quad|\,\mathrm{Im}\,t|<Kg^{-1},\quad|\mu|<b\eta^Q\}\tag{2.4}$$

where ξ,ξ_0,K,b,Q are suitable positive parameters. And on $\mathcal D$ the following bound holds (see (6.28) in [CG], and [CG] for a complete proof):

$$\sup\nolimits_{\mathcal D}|X_\mu^\sigma(\vec\psi,t;\vec\alpha)-X_\mu^\sigma(\vec\psi,\sigma\infty;\vec\alpha)|\le De^{-|\,\mathrm{Re}\,gt|\kappa},\qquad\sup\nolimits_{\mathcal D}|X_\mu^\sigma(\vec\psi,t,\vec\alpha)|\le D\tag{2.5}$$

where $\sigma=\ \mathrm{sign}\,(\,\mathrm{Re}\,t)$ and $|X|=|X_-|+|X_\downarrow|+(J_0g)^{-1}\big(|X_+|+|X_\uparrow|\big)$.

In particular, if $X^{k\sigma}(\vec\psi,t;\vec\alpha)$ is the $k^{\rm th}$ Taylor coefficient (in the μ expansion) of the function $X_\mu^\sigma(\vec\psi,t;\vec\alpha)$ and if $\mu_0=b\eta^Q$, one has for all $k\ge0$:

$$\begin{aligned}&\sup_{\mathcal D}|X^{k\sigma}(\vec\psi,t;\vec\alpha)-X^{k\sigma}(\vec\psi,\sigma\infty;\vec\alpha)|\le D\ \mu_0^{-k}\varepsilon^{-|\,\mathrm{Re}\,gt|\kappa},\qquad\mu_0=b\eta^Q\\&\sup_{\mathcal D}|X^{k\sigma}(\vec\psi,t,\vec\alpha)|<D\ \mu_0^{-k}\end{aligned}\tag{2.6}$$

and $X^{k\sigma}(t)$ in (2.2) is recognized to coincide with $X^{k\sigma}(\vec\omega t,t;\vec\alpha)$.

We number the components of X with a label j, $j=0,\ldots,2l-1$, with the convention that:

$$X_0=X_-,\quad(X_j)_{j=1,\ldots,l-1}=\vec X_\downarrow,\quad X_l=X_+,\quad(X_j)_{j=l,\ldots,2l-1}=\vec X_\uparrow\tag{2.7}$$

i.e. we write first the angle and then the action components; first the pendulum and then the rotators.

Inserting (2.2) into the Hamilton equation associated with (1.1), we see that the coefficients $X^{k\sigma}(t)\equiv X^{k\sigma}(\vec\omega t,t;\vec\alpha)$ satisfy the hierarchy of equations:

$$\frac{d}{dt}X^{k\sigma}\equiv\dot X^{k\sigma}=LX^{k\sigma}+F^{k\sigma}\tag{2.8}$$

where:

$$L\equiv L(t)=\begin{pmatrix}0&\vec0&J_0^{-1}&\vec0\\\vec0&0&\vec0&J^{-1}\\g^2J_0\sin\varphi^0(t)&\vec0&0&\vec0\\\vec0&0&\vec0&0\end{pmatrix},\quad F^1(t)=\begin{pmatrix}0\\0\\-\partial_\varphi f(\varphi^0(t),\vec\alpha+\vec\omega t)\\-\partial_{\vec\alpha}f(\varphi^0(t),\vec\alpha+\vec\omega t)\end{pmatrix}\tag{2.9}$$

and where $F^{k\sigma}$ depends upon $X^0,...,X^{k-1\sigma}$ but not on $X^{k\sigma}$; here (as everywhere else) the arrows denote $(l-1)$–vectors. Note that the entries of the $(2l\times 2l)$ matrix L have different meaning according to their position: the $\vec{0}$'s in the first and third row are $(l-1)$ (row) vectors, the $\vec{0}$'s in the first and third column are $(l-1)$ (column) vectors, and (even more confusing) the 0's and J^{-1} in the second and fourth column are $(l-1)\times(l-1)$ matrices, while the 0's in the first and third columns are scalars.

More explicitly, and more generally, if we write the Hamilton equations for $X = X^{k\sigma}_{\mu}(t)$, with hamiltonian (1.1), as:

$$\dot{X} = G_0(X) + \mu G(X) \tag{2.10}$$

(*i.e.* $G_0 = E\partial H_0$, $G = E\partial f$, $E \equiv$ standard symplectic matrix, $\partial \equiv (\partial_I, \partial_{\vec{A}}, \partial_\varphi, \partial_{\vec{\alpha}})$) then, by Taylor expansion, we can rewrite (2.8) for $k \geq 1$ as:

$$\begin{aligned}\dot{X}^{k\sigma}_r = &\sum_{j=1}^{2l} (\partial_j G_{0r})(X^0(t)) X^{k\sigma}_j + \sum_{|\vec{m}|>1} (G_{0r})_{\vec{m}}(X^0(t)) \sum_{(k^i_j)_{\vec{m},k}} \prod_{i=0}^{2l-1}\prod_{j=1}^{m_i} X_i^{k^i_j\sigma} + \\ &+ \sum_{|\vec{m}|\geq 0} (G_r)_{\vec{m}}(X^0(t)) \sum_{(k^i_j)_{\vec{m},k-1}} \prod_{i=0}^{2l-1}\prod_{j=1}^{m_i} X_i^{k^i_j}\end{aligned} \tag{2.11}$$

where we have used the notation ($k^i_j \geq 1, m_i > 0$):

$$\begin{aligned}&(G)_{\vec{m}}(\cdot) \equiv \Big(\frac{\partial_I^{m_1}\partial_{A_1}^{m_2}\ldots\partial_{A_{l-1}}^{m_l}\partial_\varphi^{m_{l+1}}\partial_{\alpha_1}^{m_{l+2}}\ldots\partial_{\alpha_{l-1}}^{m_{2l}}\, G}{m_1!\,m_2!\,m_3!\ldots m_{l+2}!\ldots m_{2l}!}\Big)(\cdot)\\ &(k^i_j)_{\vec{m},p} \equiv (k^1_1,\ldots,k^1_{m_1},k^2_1,\ldots,k^2_{m_2},\ldots,k^{2l}_1,\ldots,k^{2l}_{m_{2l}}) \qquad \text{s.t.} \sum k^i_j = p\end{aligned} \tag{2.12}$$

and $r = 1, ..., 2l$, and the case $k = 1$ requires a convenient interpretation (that the reader can easily work out, as this is the very easiest case, and as the result must be consistent with (2.8)).

Expression (2.11) gives in particular a formula for $F^{k\sigma}$ in terms of the coefficients $X^0, ..., X^{k-1\sigma}$ and of the derivatives of H_0 and f. Therefore, denoting, as above, by indices $+, \uparrow, -, \downarrow$ the components $I, \vec{A}, \varphi, \vec{\alpha}$, we see that, in our special case, (2.8) takes the particularly simple form:

$$\begin{aligned}&\frac{d}{dt}X^{k\sigma}_+ = (g^2 J_0 \sin\varphi^0)X^{k\sigma}_- + F^{k\sigma}_+ \ , \quad &\frac{d}{dt}X^{k\sigma}_\uparrow = \vec{F}^{k\sigma}_\uparrow\\ &\frac{d}{dt}X^{k\sigma}_- = J_0^{-1}X^{k\sigma}_+ \ , &\frac{d}{dt}\vec{X}^{k\sigma}_\downarrow = J^{-1}\vec{X}^{k\sigma}_\uparrow\end{aligned} \tag{2.13}$$

as $F^{k\sigma}_-$, $\vec{F}^{k\sigma}_\downarrow$ vanish identically, for $k \geq 1$. And if $X_- \equiv X_0$, for all $k \geq 1$ it is:

$$\begin{aligned}&F^{k\sigma}_- \equiv 0 \ , \ \vec{F}^{k\sigma}_\downarrow \equiv \vec{0} \ , \ \vec{F}^{k\sigma}_\uparrow = -\sum_{|\vec{m}|\geq 0} (\partial_{\vec{\alpha}} f)_{\vec{m}}(\varphi^0, \vec{\omega}t) \sum_{(k^i_j)_{\vec{m},k-1}} \prod_{i=0}^{l-1}\prod_{j=1}^{m_i} X_i^{k^i_j\sigma}\\ &F^{k\sigma}_+ \equiv \sum_{|\vec{m}|\geq 2} (g^2 J_0 \sin\varphi)_{\vec{m}}(\varphi^0) \sum_{(k_j)_k} \prod_{j=1}^{m} X_-^{k_j\sigma} - \sum_{|\vec{m}|\geq 0} (\partial_\varphi f)_{\vec{m}}(\varphi^0, \vec{\omega}t) \sum_{(k^i_j)_{\vec{m},k-1}} \prod_{i=0}^{l-1}\prod_{j=1}^{m_i} X_i^{k^i_j\sigma}\end{aligned} \tag{2.14}$$

where $(k^i_j)_{\vec{m},k}, (k^i_j)_{\vec{m},k-1}$ (are defined in (2.12)). Note that the first sum in the expression for $\vec{F}^h_+$ can only involve vectors $\vec{m}$ with $m_j = 0$ if $j \geq 1$, because the function $J_0 g^2 \cos\varphi$ depends only on φ and not on $\vec{\alpha}$, (hence also $k^i_j = 0$ if $i > 0$). We use here the above notation to uniformize the notations.

The evolution of X^k is determined by integrating (2.13), if the initial data are known. The $k = 1$ case requires a suitable interpretation of the symbols, which we leave to the reader (and the result has to be (2.9),(2.8)).

We recall that the *wronskian matrix* $W(t)$ of a solution $t \to x(t)$ of a differential equation $\dot{x} = f(x)$ in R^n is a $n \times n$ matrix whose columns are formed by n linearly independent solutions of the linear differential equation obtained by linearizing f around the solution x and assuming $W(0) =$ identity.

The solubility by elementary quadrature of the free pendulum equations on the separatrix (see for instance (3.3) below), leads after a well known classical calculation to the following expression for the wronskian

$\overline{W}(t)$ of the separatrix motion of the pendulum appearing in (1.1), with initial data at $t = 0$ given by $\varphi = \pi, I = 2gJ_0$:

$$\overline{W}(t) = \begin{pmatrix} \frac{1}{\cosh gt} & \frac{\bar{w}}{4J_0 g} \\ -J_0 g \frac{\sinh gt}{\cosh^2 gt} & (1 - \frac{\bar{w}}{4}\frac{\sinh gt}{\cosh^2 gt}) \cosh gt \end{pmatrix}, \qquad \bar{w} \equiv \frac{2gt + \sinh 2gt}{\cosh gt} \tag{2.15}$$

And the evolution of the $\pm$ (*i.e.* I, φ) components can be determined by using the above wronskian:

$$\begin{pmatrix} X_-^{k\sigma} \\ X_+^{k\sigma} \end{pmatrix} = \overline{W}(t) \begin{pmatrix} 0 \\ X_+^{k\sigma}(0) \end{pmatrix} + \overline{W}(t) \int_0^t \overline{W}^{-1}(\tau) \begin{pmatrix} 0 \\ F_+^{k\sigma}(\tau) \end{pmatrix} d\tau \tag{2.16}$$

Thus, denoting by w_{ij} $(i, j = 0, l)$ the entries of $\overline{W}$ we see immediately that:

$$\begin{aligned} X_+^{k\sigma}(t) &= w_{ll}(t) X_+^{k\sigma}(0) + w_{ll}(t) \int_0^t w_{00}(\tau) F_+^{k\sigma}(\tau) d\tau - w_{l0}(t) \int_0^t \bar{w}_{0l}(\tau) F_+^{k\sigma}(\tau)\, d\tau \\ X_-^{k\sigma}(t) &= \bar{w}_{0l}(t) X_+^{k\sigma}(0) + w_{0l}(t) \int_0^t w_{00}(\tau) F_+^{k\sigma}(\tau) d\tau - w_{00}(t) \int_0^t w_{0l}(\tau) F_+^{k\sigma}(\tau)\, d\tau \end{aligned} \tag{2.17}$$

The integration of the equations (2.13) for the $\uparrow, \downarrow$ components is "easier" yielding:

$$\begin{aligned} \vec{X}_\uparrow^{k\sigma}(t) =& \vec{X}_\uparrow^{k\sigma}(0) + \int_0^t \vec{F}_\uparrow^{k\sigma}(\tau) d\tau \\ \vec{X}_\downarrow^{k\sigma}(t) =& J^{-1} \Big(t\vec{X}_\uparrow^{k\sigma}(0) + \int_0^t d\tau\, (t - \tau) F_\uparrow^{k\sigma}(\tau) \Big) \end{aligned} \tag{2.18}$$

having used that the $\vec{X}_\downarrow^{k\sigma} \equiv \vec{0}$ because the initial datum is fixed and μ independent; and (2.17), (2.18) can be used to find a reasonably simple algorithm to represent the whiskers equations to all orders $k \geq 1$ of the perturbation expansion.

We shall regard the two functions $X^{k\sigma}(t)$, as forming a single function $X^k(t)$:

$$X^k(t) = \begin{cases} X^{k+}(t) & \text{if } \sigma = \operatorname{sign} t = + \\ X^{k-}(t) & \text{if } \sigma = \operatorname{sign} t = - \end{cases} \tag{2.19}$$

and at this point it is useful to open a parenthesis to define some integration operations that can be performed on such functions; such operations arise as soon as one tries to determine the initial data (still unknown) in (2.17),(2.18).

§3 The improper integration $\mathcal{I}$.

The operation is simply the integration over t from $\sigma\infty$ to t, $\sigma = \operatorname{sign} t$. In general such operation cannot be defined as an ordinary integral of a summable function, because the functions on which it has to operate (typically the integrands in (2.17),(2.18)) do not, in general, tend to 0 as $t \to \infty$.

But the simplicity of the initial hamiltonian has the consequence that the functions $X^k(t)$, and the matrix elements w_{ij} in (2.15), belong to a very special class of analytic functions on which the integration operations that we need can be given a meaning.

To describe such class we introduce various spaces of functions; all of them are subspaces of the space $\hat{\mathcal{M}}$ of the functions of t defined as follows.

Definition 1: *Let $\hat{\mathcal{M}}$ be the space of the functions of t which can be represented, for some $k \geq 0$, as:*

$$M(t) = \sum_{j=0}^{k} \frac{(\sigma t g)^j}{j!} M_j^\sigma(x, \vec{\omega} t) , \quad x \equiv e^{-\sigma g t} , \quad \sigma = \operatorname{sign} t \tag{3.1}$$

with $M_j^\sigma(x, \vec{\psi})$ a trigonometric polynomial in $\vec{\psi}$ with coefficients holomorphic in the x-plane in the annulus $0 < |x| < 1$, with: 1) possible singularities, outside the open unit disk, in a closed cone centered at the origin, *with axis of symmetry on the imaginary axis and half opening* $< \frac{\pi}{2}$; *2) possible polar singularities at* $x = 0$; *3)* $M_k \neq 0$. *The number* k *will be called the* t*–degree of* M. *The smallest cone containing the singularities will be called the singularity cone of* M.

It is not difficult to see that if a function admits a representation like (3.1), with the above properties, then such a representation is unique, (see [CG], eq. (10.15), note that in [CG] $\hat{\mathcal{M}}$ is called $\mathcal{M}$).
It is convenient to consider the functions of t defined by the monomials:

$$\sigma^\chi \, \tfrac{(\sigma t g)^j}{j!} x^h e^{i\vec{\omega}\cdot\vec{\nu}\, t} \tag{3.2}$$

where $\chi = 0, 1$ and we use the notations in (3.1). We remark that the functions in $\hat{\mathcal{M}}$ can be expanded in the above monomials (*i.e.* the (3.2) "span" the space $\hat{\mathcal{M}}$). The functions $M \in \hat{\mathcal{M}}$ such that the residuum at $x = 0$ of $x^{-1}\langle M^\sigma_j(x, \cdot)\rangle$ is zero, (here the average is over $\vec{\psi}$, *i.e.* it is an "angle average"), form a subspace $\hat{\mathcal{M}}_0$ of $\hat{\mathcal{M}}$.
The functions in $\hat{\mathcal{M}}, \hat{\mathcal{M}}_0$ are not bounded near $x = 0$, in general. We denote $\mathcal{M}$ and $\mathcal{M}_0$ the respective subspaces of the functions bounded near $x = 0$: which means that the M_j have no pole at $x = 0$ and, furthermore, that $M_j(0, \vec{\psi}) = 0$ if $j > 0$.
The symbols $\hat{\mathcal{M}}^k, \hat{\mathcal{M}}^k_0, \mathcal{M}^k, \mathcal{M}^k_0$ will denote the subspaces of $\hat{\mathcal{M}}, \hat{\mathcal{M}}_0, \mathcal{M}, \mathcal{M}_0$, respectively, containing the functions of t–degree $\le k$.
Note that $M \in \hat{\mathcal{M}}$ can be written as $M = P + M'$ with P being a polynomial in σt (with σ dependent coefficients) and with $M' \in \hat{\mathcal{M}}_0$: this can be done in only one way and we call P the "polynomial component" of M. Likewise $M \in \mathcal{M}$ can be written as $M = p + M'$ with p being a constant function (with constant value depending on σ) and $M' \in \mathcal{M}_0$: p will be called the "constant component" of M. In both cases M' will be the "non singular" component of M.
The coefficients of the above mentioned expansions and polynomials depend on $\sigma = \pm$, *i.e.* each $M \in \hat{\mathcal{M}}$ is, in general, a pair of functions M^σ defined and holomorphic for $t > 0$ and $t < 0$, respectively (and, more specifically, in a domain including $\{\sigma\,\mathrm{Re}\,t > 0,\, |\mathrm{Im}\, gt| < \pi/2\}$).
The functions $M^\sigma(t)$ might sometimes (as in our cases below) be continued analytically in t but in general $M^+(-t) \neq M^-(-t)$ even when it makes sense (by analytic continuation) to ask whether equality holds. For the purpose of comparison with [CG] we note that the only spaces introduced there are $\hat{\mathcal{M}}$ and $\hat{\mathcal{M}}_0$, and in [CG] they are called $\mathcal{M}$ and $\mathcal{M}_0$ respectively.
Note however that if $M \in \mathcal{M}$ the points with $\mathrm{Re}\,t = 0$ and $|\mathrm{Im}\,t| < \pi/(2g)$ ($gt = \pm i\pi/2$ corresponds to $x = \mp i$) are, (by our hypothesis on the location of the singularities of the M_j functions), regularity points so that the values at $t^\pm$, "to the right" and "to the left" of t, will be regarded as well defined and given by $M(t^\pm) \equiv \lim_{t'\to t,\, \mathrm{Re}\,t' \to \mathrm{Re}\,t^\pm} M(t')$, in particular $M^\pm(0^\pm) \equiv M^\pm_0(1^-, \vec{0})$.
It must be remarked (as this will be essential later) that, since $\varphi^0(t) = 4 \arctan e^{-gt}$:

$$\cos\varphi^0 = 1 - \tfrac{2}{(\cosh gt)^2} = 1 - 8\tfrac{x^2}{(1+x^2)^2}\ , \quad \sin\varphi^0 = 2\tfrac{\sinh gt}{(\cosh gt)^2} = 4\sigma x \tfrac{1-x^2}{(1+x^2)^2} \tag{3.3}$$

and since f in (1.1) is a trigonometric polynomial, the function F^1, see (2.9), belongs to $\mathcal{M}$ and, in fact, the component $\vec{F}^1_\uparrow$ belongs to $\mathcal{M}_0$ (as accidentally does F^1_+ as well).
In general if a function $M \in \hat{\mathcal{M}}$ is *holomorphic in* t in the strip $|\mathrm{Im}\, gt| < \frac{\pi}{2}$ and it has a given parity, then it follows that: $M^+_j(x) = \pm(-1)^j M^-_j(x^{-1})$, where the sign is $+$ if M is even, and $-$ if it is odd. This means that $\sigma^j M_j$ have the same parity for $x \longleftrightarrow x^{-1}$ as M for $t \longleftrightarrow -t$. Interesting examples are the functions in (3.3).
It will be checked, by induction, that the functions X^k and F^k are in $\mathcal{M}$ and their representation (3.1) is such that the sum over j runs up to $2k$. We say that $F^k, X^k \in \mathcal{M}^{2k}$, see definition 1. More precisely:

$$X^k \in \mathcal{M}^{2k-1}, \qquad F^k \in \mathcal{M}^{2(k-1)}, \qquad \vec{F}^k_\uparrow \in \mathcal{M}^{2(k-1)}_0 \tag{3.4}$$

and, furthermore, the singularity cone consists of just the imaginary axis (*i.e.* the singularities of the functions defining X^k, F^k are on the segments on the imaginary axis $(-i\infty, -i]$ and $[+i, +i\infty)$). See below for details.
On the class $\hat{\mathcal{M}}$ we can define the following operation. If $M \in \hat{\mathcal{M}}$, and $t = \tau + i\vartheta$, with τ, ϑ real, *and* $\tau = \mathrm{Re}\,t \neq 0$, $\sigma = \mathrm{sign}\,\mathrm{Re}\,t$, the function:

$$\mathcal{I}_R M(t) \equiv \int_{\sigma\infty + i\vartheta}^{t} e^{-Rg\sigma z} M^\sigma(z)\, dz \tag{3.5}$$

is defined for $\mathrm{Re}\, R > 0$ and large enough, the integral being on an axis parallel to the real axis.

If $M \in \hat{\mathcal{M}}$ then the function of R in (3.5) admits an analytic continuation to $\mathrm{Re}\, R < 0$ with possible poles at the integer values of R and at the values $i\vec{\omega}\cdot\vec{\nu}$ with $|\vec{\nu}| <$ (trigonometric degree of M in the angles $\vec{\psi}$); and we can then set:

$$\mathcal{I}M(t) \equiv \oint \frac{dR}{2\pi i R}\,\mathcal{I}_R M(t) \tag{3.6}$$

where the integral is over a small circle of radius $r < 1$ and $r < \min|\vec{\omega}\cdot\vec{\nu}|$, with the minimum being taken over the $\vec{\nu} \neq \vec{0}$ which appear in the Fourier expansion of M (which is finite by definition 1).

¿From the above definition it follows immediately that if $M(t) = t^j$ then:

$$\mathcal{I}M(t) = \tfrac{1}{j+1}\, t^{j+1} \tag{3.7}$$

and more generally, if j, h are integers and $\chi = 0, 1$, the $\mathcal{I}$ acts on the monomial (3.2) as:

$$\mathcal{I}M(t) = \begin{cases} -g^{-1}\sigma^{\chi+1}x^h e^{i\vec{\omega}\cdot\vec{\nu}t}\sum_{p=0}^{j}\frac{(g\sigma t)^{j-p}}{(j-p)!}\frac{1}{(h-i\sigma g^{-1}\vec{\omega}\cdot\vec{\nu})^{p+1}} & \text{if } |h|+|\vec{\nu}|>0 \\ g^{-1}\sigma^{\chi+1}\frac{(\sigma g t)^{j+1}}{(j+1)!} & \text{otherwise} \end{cases} \tag{3.8}$$

showing, in particular, that the radius of convergence in x of $\mathcal{I}M$, for a general M, is the same of that of M. But in general the singularities at $\pm i$ will no longer be polar, even if those of the M_j's were such.

In fact one might want to check that $\mathcal{I}M \in \hat{\mathcal{M}}$: this certainly happens if no singularity can appear in the $(\mathcal{I}M)_j$'s outside the singularity cone of M. Note that $M \in \hat{\mathcal{M}}$ can be written, for fixed $\sigma = \mathrm{sign}\,\mathrm{Re}\,t$, as a sum of finitely many monomials $t^j x^{-r} e^{i\Omega t}$ with $r \geq 0$ plus a function $M'(t)$ of the form (3.1) with $M_j(0,\vec{\psi}) = 0$.

If the $\mathcal{I}$ operates on the monomials with $r \geq 0$ (*i.e.* on monomials not vanishing at $x = 0$ or with a polar singulariy at the origin) then it can be explicitly computed by using (3.8), and the check of the claimed property is immediate.

The $\mathcal{I}M$ for M such that $M_j(0,\vec{\psi}) \equiv 0$ can be computed as an ordinary integral: it is obvious from (3.5) that in such case the $\mathcal{I}_R M(t)$ is holomorphic in R for $R = 0$ (as the integral in (3.5) is convergent for $R = 0$). Consider first the case in which $M(t) = e^{i\vec{\omega}t}M^\sigma(x)$ with $M^\sigma(x)$ in $\hat{\mathcal{M}}$ and $M^\sigma(0) = 0$, one gets for all Ω real:

$$\textstyle\int_{\sigma\infty}^{t} e^{i\Omega\tau}M(e^{-\sigma g\tau})\,d\tau = \int_0^x e^{-i\Omega\sigma g^{-1}\log y}M(y)\frac{dy}{y} = e^{i\Omega t}\int_0^1 e^{-i\Omega\sigma g^{-1}\log y}M(yx)\frac{dy}{y} \tag{3.9}$$

which is clearly holomorphic in x in the above considered x plane deprived, outside the unit disk, of the singularity cone of M. The more general case, in which we consider $t^j e^{i\Omega t}M(x)$ with $M(0) = 0$, is derived from (3.9) by differentiation with respect to Ω. Any other case is a finite linear combination of the considered cases.

Furthermore, if $M(t) \in \hat{\mathcal{M}}$ and if $\mathrm{sign}\,\mathrm{Re}\,t = \mathrm{sign}\,\mathrm{Re}\,t_0$, it follows from (3.8) (as declared at the beginning of §3) that the $\mathcal{I}M$ function is a primitive of M:

$$\int_{t_0}^{t} M(\tau)d\tau = \mathcal{I}M(t) - \mathcal{I}M(t_0) \tag{3.10}$$

In general, $\mathcal{I} : \hat{\mathcal{M}}^k \to \hat{\mathcal{M}}^{k+1}$, because of (3.7). Furthermore, because of the similarities of the $\mathcal{I}$ operation with a definite integral, we shall often use the notation:

$$\fint_{(\sigma)}^{t} M(\tau)d\tau \equiv \mathcal{I}M(t)\ , \qquad M \in \hat{\mathcal{M}},\ \sigma = \mathrm{sign}\ \mathrm{Re}\,t \tag{3.11}$$

In fact many standard properties of integration are, in such a way, extended to the space $\hat{\mathcal{M}}$; for instance:

$$\fint_{(\sigma)}^{t} = \fint_{(\sigma)}^{0^\sigma} + \int_0^t \quad , \qquad \mathcal{I}^2 \equiv \fint_{(\sigma)}^{t}\fint_{(\sigma)}^{\tau} = \fint_{(\sigma)}^{t}(t-\tau) \tag{3.12}$$

where $\sigma = \mathrm{sign}\,t$ and $\fint_{(\sigma)}^{0^\sigma}$ means $\lim_{\varepsilon\to 0^\sigma}\fint_{(\sigma)}^{\varepsilon}$, of course.

This leads to a few more natural definitions and properties.

If $\sigma_t = \text{sign Re}\, t$, if $M \in \hat{\mathcal{M}}$ and if R is large enough, we define $\mathcal{I}_{\pm,R}M$ via:

$$\mathcal{I}_{\pm,R}M(t) \equiv \int_{\pm\infty}^{t} M(\tau)e^{-R\sigma_\tau \tau g}d\tau = \mathcal{I}_R M(0^\pm) + \int_0^t M(\tau)e^{-R\sigma_t \tau g}d\tau \tag{3.13}$$

where we use the definition (3.5), and in the r.h.s. the integral is on the straight line joining 0 to t. This allows us to consider the analytic continuation in R of $R^{-1}\mathcal{I}_{\pm,R}M(t)$ and its residue $\mathcal{I}_\pm M(t)$ at $R = 0$. The latter is linked to the operation $\mathcal{I}$ already defined in (3.6) so that the following two definitions (of $\mathcal{I}_\pm M(t)$ and the consequent one of $⨍_{-\infty}^{+\infty} M(\tau)d\tau$) are natural:

$$\begin{aligned}\mathcal{I}_\pm M(t) \equiv ⨍_{\pm\infty}^{t} M(\tau)\,d\tau &= \mathcal{I}M(0^\pm) + \int_0^t M(\tau)d\tau \\ ⨍_{-\infty}^{+\infty} M(\tau)d\tau &= \mathcal{I}_- M(t) - \mathcal{I}_+ M(t)\end{aligned} \tag{3.14}$$

where the integral from 0 to t is over a straight path joining 0 with t. The (3.14) will be a quite useful extension of the operation $\mathcal{I}$ introduced in (3.11).

Given the definition (3.14), a natural question arises at this point: is there a class of functions $M \in \mathcal{M}$ such that the following *shift of contour* formula:

$$⨍_{-\infty}^{+\infty} M(\tau)d\tau = ⨍_{-\infty}^{+\infty} M(\tau + ig^{-1}\xi)d\tau \tag{3.15}$$

holds for all ξ smaller than the complement to $\frac{\pi}{2}$ of the half opening of the singularity cone of M (*i.e.* for all ξ for which (3.15) makes sense)? A very simple answer follows immediately from the definitions: (3.15) holds if M is holomorphic as a function of t in a strip around the real axis, wider than $g^{-1}\xi$. Note that this might be at first surprising as the operation $⨍$ is an improper integral operating on generally non summable functions.

The (3.15) can be proved by remarking that for R large (and positive):

$$\mathcal{I}_{-,R}M(0^-) - \mathcal{I}_{+,R}M(0^+) = \int_{-\infty}^{0} e^{R\tau g}M(\tau)d\tau + \int_0^{+\infty} e^{-R\tau g}M(\tau)d\tau \tag{3.16}$$

which, by the assumed analyticity of M, differs from:

$$\int_{-\infty}^{0} e^{R\tau g}M(\tau + ig^{-1}\xi)d\tau + \int_0^{+\infty} e^{-R\tau g}M(\tau + ig^{-1}\xi)d\tau \tag{3.17}$$

precisely by:

$$\begin{aligned}&(e^{iR\xi} - 1)\int_{-\infty}^{0} d\tau e^{R\tau g}M(\tau + ig^{-1}\xi) + (e^{-iR\xi} - 1)\int_0^{+\infty} d\tau e^{-R\tau g}M(\tau + ig^{-1}\xi) + \\ &-i\int_0^{g^{-1}\xi}(e^{-Ri\tau g} - e^{+Ri\tau g})M(i\tau)\,d\tau\end{aligned} \tag{3.18}$$

i.e. (3.16) is the sum of (3.17) and (3.18), by the analyticity properties of M. This implies (3.15) by taking the residues at $R = 0$, as the analytic continuation of (3.18) vanishes if $R \to 0$, when $M \in \hat{\mathcal{M}}_0$, *i.e.* it has no polynomial component; while if M is a polynomial in t one can check by direct calculation (using (3.7)) that both sides of (3.15) give 0.

It is also useful for the purposes of a better understanding, to realize that, if M is holomorphic as a function of t in a strip wider than $g^{-1}\xi$, then for t real it is:

$$\frac{1}{2}\sum_\rho ⨍_{\rho\infty}^{t+ig^{-1}\xi} M(\tau)d\tau = \oint \frac{dR}{2\pi i R}\frac{1}{2}\sum_\rho \int_{\rho\infty}^{t} e^{-R\sigma_\tau g(\tau+ig^{-1}\xi)}M(\tau + ig^{-1}\xi)d\tau \tag{3.19}$$

where the integrals in the r.h.s. have to be considered to be the analytic continuation on R from $R > 0$ and large. This is so because the two sides differ by the residue at $R = 0$ of $-iR^{-1}\int_0^{g^{-1}\xi} d\tau(e^{-Ri\xi} - e^{Ri\xi})M(i\tau)$ which vanishes.

Note that the r.h.s. is *different* from $\frac{1}{2}\sum_\rho \int\!\!\!\!-_{\rho\infty}^{t} M(\tau + ig^{-1}\xi)d\tau$, which would be the residue ar $R = 0$ of $R^{-1}\sum_\rho \int_{\rho\infty}^{t} e^{-R\sigma_\tau g\tau}\ M(\tau + ig^{-1}\xi)\ d\tau$. The two quantities coincide under the mentioned analyticity assumption, however, if M is in $\mathcal{M}_0$; and one could verify that this remains true if M has no polynomial component, *i.e.* $M \in \hat{\mathcal{M}}_0$.

This completes the discussion of the operations $\mathcal{I}$.

§4 Analytic expressions of the expansion coefficients for the whiskers and the KAM tori. Parity properties.

A): *Whiskers.*

We shall show that X^k admits rather simple expressions in terms of the operation $\mathcal{I}$,(and other related operations introduced below). Recall that in §2 we have fixed $\vec{\alpha} \in T^{l-1}$ and $\varphi = \pi$, and we are looking for the motions, on the stable ($\sigma = +$) or unstable ($\sigma = -$) whisker, which start with the given $\vec{\alpha}$ and $\varphi = \pi$ at $t = 0$; in the following $\vec{\alpha}$ is kept constant and usually notationally omitted.

We suppose inductively that $F^h \in \mathcal{M}^{2(h-1)}$, $F^h_\uparrow \in \mathcal{M}_0^{2(h-1)}$ for $h \le k$, and that $X^{h\sigma} \in \mathcal{M}^{2h-1}$, for $h < k$: see definition 1, §3. This means, in particular, that $F^{h\sigma}, X^{h\sigma}$ can be represented as:

$$
\begin{aligned}
F^{h\sigma}(x,\vec{\psi},t) &= \sum_{j=0}^{2(h-1)} \frac{t^j}{j!} F_j^{h\sigma}(x,\vec{\psi}), \qquad h = 1,\ldots,k \\
X^{h\sigma}(x,\vec{\psi},t) &= \sum_{j=0}^{2h-1} \frac{t^j}{j!} X_j^{h\sigma}(x,\vec{\psi}), \qquad h = 1,\ldots,k-1
\end{aligned}
\tag{4.1}
$$

by setting $\vec{\psi} = \vec{\omega}t$, $\sigma = \operatorname{sign} t$, $x = e^{-g\sigma t}$; with $F_j^{k\sigma}, X_j^{k\sigma}$ holomorphic at $x = 0$ and vanishing at $x = 0$ if $j > 0$. Hnece if $x = e^{-g\sigma t}$ and $\vec{\psi}$ is kept fixed the F_j^h, X_j^h tend exponentially to zero as $t \to \infty$, if $j > 0$; while if $j = 0$ they tend exponentially fast to a limit as $t \to \sigma\infty$ (*i.e.* as $x \to 0$), which we denote $F^h(\vec{\psi},\sigma\infty)$ dropping the subscript 0 as there is no ambiguity.

Furthermore the inductive hypothesis is enriched by:

$$
\vec{F}^{k\sigma}_{\uparrow\vec{0}}(\sigma\infty) = \vec{0}, \qquad \text{for all } h \le k \tag{4.2}
$$

recalling that, in general, a subscript $\vec{\nu}$ affixed to a function denotes the Fourier component of order $\vec{\nu} \in Z^{l-1}$ of the considered function.

We denote $X^{h\sigma}_{j\vec{\nu}}(t)$ and $F^{h\sigma}_{j\vec{\nu}}(t)$ the Fourier transforms in $\vec{\psi}$ of $X_j^{h\sigma}(t,\vec{\psi})$ and $F_j^{h\sigma}(t,\vec{\psi})$. *It follows from the KAM theory mentioned in (2.6), that $X^{h\sigma}(t)$ and, from (2.14), hence also $F^{h\sigma}(t)$ are bounded as $t \to \sigma\infty$ for all h*, so that $X_j^{h\sigma}(0,\vec{\psi}) = 0$ if $j \ge 1$. We show that the latter information is very strong and permits us to determine X^k.

We note that, since $F^{k\sigma} \in \mathcal{M}^{2(k-1)}$ and $\vec{F}^{k\sigma}_{\uparrow\vec{0}} = \vec{0}$ hold, the function $\vec{X}^{k\sigma}_\uparrow(t)$, given by the first of (2.18), is in fact in $\mathcal{M}^{2(k-1)}$ (by integration). But of course we do not know (yet) the initial data $X^{k\sigma}(0)$.

To find expressions for $X^k_\uparrow$ we start from the equations (2.13) with initial time at some instant T. And we use that $\mathcal{I}F(t)$ is a primitive of the function $F(t)$, see (3.10), so that:

$$
\vec{X}^{k\sigma}_\uparrow(t) = \vec{X}^{k\sigma}_\uparrow(T) + \mathcal{I}F^{k\sigma}(t) - \mathcal{I}F^{k\sigma}(T) \tag{4.3}
$$

where $\sigma = \operatorname{sign} t$, and T has the same sign of t.

The function $\vec{X}^{k\sigma}_\uparrow(T)$ tends to become quasi periodic with exponential speed as $T \to \sigma\infty$: in fact it becomes asymptotic to the $j = 0$ component, see (4.1), at $x = 0$: $X^{k\sigma}_{0\uparrow}(0,\vec{\omega}T)$, (in the sense that the difference tends to 0, bounded by $(g|T|)^{2k-1}e^{-g|T|})$. The function $\mathcal{I}F^{k\sigma}_{\uparrow}(T)$ also becomes asymptotically quasi periodic with exponential speed *and* $\vec 0$ *average*, because $\vec F^{k\sigma}_{\uparrow}\in\mathcal{M}^{2(k-1)}_0$ and by the definition of $\mathcal{I}$: therefore the two quasi periodic functions of T must cancel modulo a constant equal to $\langle\vec X^{k\sigma}_{0\uparrow}(0,\cdot)\rangle\equiv\vec X^{k\sigma}_{\uparrow\vec 0}(\sigma\infty)$.

Hence it follows that:

$$\vec X^{k\sigma}_{\uparrow}(t)=\vec X^{k\sigma}_{\uparrow\vec 0}(\sigma\infty)+\mathcal{I}\vec F^{k\sigma}_{\uparrow}(t) \tag{4.4}$$

and, by inserting (4.4) into the second of(2.18), (considering also that the time average of $\mathcal{I}\vec F^{k\sigma}_{\uparrow}$ vanishes, and therefore $\int_0^t\tau\vec F^{k\sigma}_{\uparrow}(\tau)\,d\tau=t\mathcal{I}\vec F^{k\sigma}_{\uparrow}(t)+$ a t-bounded function), we see that the $\vec X^{k\sigma}_{\downarrow}(t)$ can be bounded only if:

$$\vec X^{k\sigma}_{\uparrow\vec 0}(\sigma\infty)=\vec 0,\qquad\text{hence:}\qquad\vec X^{k\sigma}_{\uparrow}(t)=\mathcal{I}\vec F^{k\sigma}_{\uparrow}(t) \tag{4.5}$$

yielding, setting $t=0^\sigma$, the initial values of $X^k_{\uparrow}$ *and* a simple form for its time evolution. Analogously, recalling that $\vec X^{k\sigma}_{\downarrow}(0)=\vec 0$, essentially by definition, one finds:

$$\vec X^{k\sigma}_{\downarrow}(t)=J^{-1}\big(\mathcal{I}^2\vec F^{k\sigma}_{\uparrow}(t)-\mathcal{I}^2\vec F^{k\sigma}_{\uparrow}(0^\sigma)\big)\equiv J^{-1}\bar{\mathcal{I}}^2\vec F^{k\sigma}_{\uparrow}(t) \tag{4.6}$$

which gives a simple form to the time evolution of the $\vec\alpha$ (*i.e.* $\downarrow$) component of X^k in terms of the operator $\overline{\mathcal{I}}^2$ defined by the r.h.s. of (4.6).

Likewise considering the (2.17) and the behaviour at $\sigma\infty$ of $\overline{W}$ in (2.15) and *recalling that $X^{k\sigma}(t)$ has to be bounded at $\sigma\infty$ by (2.6)*, we see from the second of (2.17) that:

$$X^{k\sigma}_{+}(0)=-\int_0^{\sigma\infty}w_{00}(\tau)F^{k\sigma}_{+}(\tau)\;d\tau \tag{4.7}$$

Thus we get (defining at the same time also $\mathcal{O}$ and $\mathcal{O}_+$):

$$\begin{aligned}X^{k\sigma}_{+}(t)&=w_{ll}(t)\,⨍_{(\sigma)}^{t}w_{00}(\tau)F^{k\sigma}_{+}(\tau)d\tau-w_{l0}(t)\int_0^t w_{0l}(\tau)F^{k\sigma}_{+}(\tau)d\tau\equiv\mathcal{O}_+F^{k\sigma}_{+}(t)\\X^{k\sigma}_{-}(t)&=w_{0l}(t)\,⨍_{(\sigma)}^{t}w_{00}(\tau)F^{k\sigma}_{+}(\tau)d\tau-w_{00}(t)\int_0^t w_{0l}(\tau)F^{k\sigma}_{+}(\tau)d\tau\equiv\mathcal{O}F^{k\sigma}_{+}(t)\end{aligned} \tag{4.8}$$

The (4.5),(4.6),(4.8), and (2.5) imply (2.2) for $h=k+1$. As already remarked before (4.3) we note again that, since $F^{h\sigma}_{\uparrow\vec 0}(\sigma\infty)=\vec 0$ for $h\le k$, the $\vec X^{k+1}_{\uparrow},\vec X^{k+1}_{\downarrow}$ functions are in fact in $\mathcal{M}^{2(k-1)}$, (as the $\mathcal{I}$ operation, on such $\vec F^{k}_{\uparrow}$ functions does not increase the degree). Also, if one looks carefully at the $X^{h\sigma}_{\pm}$–evaluation in terms of $F^{h\sigma}_{+}$, one realizes that the $\mathcal{O},\mathcal{O}_+$ operations may increase the degree but by at most 1. Thus the inductive hypothesis made in connection with (4.1) is proved for X^{k+1}, and it remains to check it for F^{k+1}.

The latter check follows from the expression of F^{k+1}, see (2.14), in terms of the X^h with $h\le k$: see (2.14). One treats separately the sums in (2.14) with $|\vec m|\ge 2$ and $|\vec m|\ge 0$. One just has to consider that in the first case, which might look dangerous for the inductive hypothesis, the products of X's contains *at least two factors* (which therefore have order labels smaller than k and verify the inductive hypothesis); and, furthermore, the coefficients $(\partial_{\vec\alpha}f)_{\vec m}(\varphi_0,\vec\omega t)$ or $g^2J_0\sin\varphi_0$ or $g^2J_0\cos\varphi_0$, by (3.3), do not contain any terms that can possibly increase the degree. Hence $F^{k+1}\in\mathcal{M}^{2k}$.

To see that $\vec F^{(k+1)\sigma}_{\uparrow}\in\mathcal{M}^{2k}_0$, *i.e.* $\vec F^{(k+1)\sigma}_{\uparrow\vec 0}=\vec 0$, we simply remark that otherwise the second of (2.18) could not be bounded in t as $t\to\infty$; but we know that it is bounded by (2.5).

Remark: The use of (2.5) is clearly a spurious element: it should not be necessary to invoke a rather involved analytic discussion (*e.g.* the "KAM" theorem yielding the (2.5)) to prove an algebraic fact, namely $\vec F^{(k+1)\sigma}_{\uparrow\vec 0}=\vec 0$, that if not valid would prove the claims of the theorem leading to (2.5) false. At least it is unpleasant to do so (although logically consistent) and a direct check of the algebraic property is highly desirable: it would show that (2.14), and the equations $\vec F^{k\sigma}_{\uparrow\vec 0}=\vec 0$, and (4.4),(4.6) and (4.8) yield recursively a formal power series expression for the whiskers.

Such a check is possible, see [CG] appendix 14. This means that the property $\vec F^{k\sigma}_{\uparrow\vec 0}=\vec 0$ and the (4.4),(4.6), and (4.8), coupled with the initial condition (2.9), *always* have a solution verifying (3.4). It gives us a formal power series solution to the problem of finding the whiskers equations (and those of their tori as well, which

are easily related to the $t \to \pm\infty$ behaviour of the X^k functions).
The convergence of the series, however, does not follow from what said so far. In the case $l = 2$ it will be checked in §8, so that the $l = 2$ case can be made fully independent on the KAM–type results expressed by the (2.5) (and imply them, of course). If $l > 2$ we still rely on the KAM results for the convergence, although it will be clear that with some extra work it should be possible to obtain convergence estimates along the same lines as in the $l = 2$ case.

We can summarize the above analytic considerations as:

$$\vec F^{k\sigma}_{\uparrow\vec 0}(\sigma\infty) \equiv \int_{T^{l-1}} \vec F^{k\sigma}_{\uparrow}(\vec\psi,\sigma\infty)\frac{d\vec\psi}{(2\pi)^{l-1}} \equiv \langle \vec F^{k\sigma}_{\uparrow}(\cdot,\sigma\infty)\rangle = \vec 0 \tag{4.9}$$

for all $k \ge 1$, and, still for all $k \ge 1$, by:

$$\begin{aligned}
X^h_-(t) =& w_{0l}(t)\mathcal I(w_{00}F^h_+)(t) - w_{00}(t)\big(\mathcal I(w_{0l}F^h_+)(t) - \mathcal I(w_{0l}F^h_+)(0^\sigma)\big) \equiv \mathcal O(F^h_+)(t)\\
\vec X^h_\downarrow(t) =& J^{-1}\left(\mathcal I^2(\vec F^h_\uparrow)(t) - \mathcal I^2(\vec F^h_\uparrow)(0^\sigma)\right) \equiv J^{-1}\overline{\mathcal I}(\vec F^h_\uparrow(t)\\
X^h_+(t) =& w_{ll}(t)\mathcal I(w_{00}F^h_+)(t) - w_{l0}(t)\big(\mathcal I(w_{0l}F^h_+)(t) - \mathcal I(w_{0l}F^h_+)(0^\sigma)\big) \equiv \mathcal O_+(F^h_+)(t)\\
\vec X^h_\uparrow(t) =& \mathcal I(\vec F^h_\uparrow)(t)
\end{aligned} \tag{4.10}$$

where $\mathcal O, \mathcal O_+, \overline{\mathcal I}^2, \mathcal I$ are defined here (and in §3); and $X^h \equiv (X_-, \vec X_\downarrow, X_+, \vec X_\uparrow) = (X^h_j)$, $j = 0, \ldots 2l-1$, $F^h = (0, \vec 0, F^h_+, \vec F^h_\uparrow)$ so that $-,+$ are synonims of $0, l$ respectively and $\downarrow, \uparrow$ denote collectively the labels $j = 1, \ldots, l-1$ and $l+1, \ldots, 2l-1$ respectively (see also §3). Note that while X^h has non zero components over both the *angle* ($j = 0, \ldots, l-1$) components and over the *action* ($j = l, \ldots, 2l-1$) the F^h has only the action compnents non zero. *Furthermore if $\sigma t > 0$ the above functions describe a motion on the whisker W^σ with initial data at some $\vec\alpha$ and $\varphi = \pi$.*

B): *Tori:*

A case of special interest is the case in which f in (1.1) is φ–independent. In such case the pendulum and the rotators decouple and we are really studying the perturbation theory of a completely integrable $(l-1)$–dimensional system, of rather special form, namely that of model 1) in (1.1).
The whiskers will, in this case, be degenerate and, at $\varphi = \pi$, have the form:

$$\vec\alpha \to (\vec X_\uparrow(\vec 0; \vec\alpha), I_0(\pi)) \tag{4.11}$$

and we can deduce the geometric locus of the torus, by letting $\vec\psi$ vary in T^{l-1}, via:

$$\vec A = \vec X_\uparrow(\vec\psi, \sigma\infty; \vec 0) = \textstyle\sum_{k=1}^\infty \mu^k \vec X^{k\sigma}_\uparrow(\vec\psi,\sigma\infty;\vec 0), \qquad \vec\alpha = \vec\psi + \sum_{k=1}^\infty \mu^k \vec X^{k\sigma}_\downarrow(\vec\psi,\sigma\infty;\vec 0) \tag{4.12}$$

and, in fact, in this case $X^\sigma(\vec\psi, t; \vec 0)$ is t–independent, and also σ–independent.
If $0 < J_0 < J < +\infty$ we can use the general KAM theory to say that the above invariant torus does exist for $|\mu| < b^{-1}$ small enough. The analysis in A) of the whiskered tori then yields, in this particular case, that the motion $t \to X(t) = (\vec X_\downarrow(t), \vec X_\uparrow(t))$, given by the following (4.13),(4.14), is a quasi periodic motion on the invariant torus:

$$\vec X^h_\downarrow(t) = J^{-1}\big(\mathcal I^2 \vec F^h_\uparrow(t) - \mathcal I^2 \vec F^h_\uparrow(0)\big), \qquad \vec X^h_\uparrow(t) = \mathcal I \vec F^h_\uparrow(t) \tag{4.13}$$

where we have dropped the σ labels as the functions $\vec F, \vec X$ no longer depend on them. And, by (2.14), if $\vec m = (m_1, \ldots, m_{l-1})$, $k^i_j \ge 1$:

$$\vec F^h_\uparrow = -\sum_{|\vec m|>0} (\partial_{\vec\alpha} f)_{\vec m}(\vec\omega t) \sum_{(k^i_j)_{\vec m, k-1}} \prod_{i=1}^{l-1}\prod_{j=1}^{m_i} X_i^{k^i_j\sigma} \tag{4.14}$$

because the pendulum and the whiskers really disappear as a consequence of the decoupling.
The analysis in A) of the whiskered tori can be repeated to show that $\vec X^h_i, \vec F^h_i$, for $i = 1, \ldots, l-1$ ($\downarrow$ components) and for $i = l+1, \ldots, 2l-1$ ($\uparrow$ components), are in $\mathcal M^0$ for all $h \ge 1$. This is a stronger

statement than (3.3), valid only for the $\vec{X}^k$ components and in the present special case: note that in our notations $\mathcal{M}^0$, with the label 0 as superscript, means that not only the functions in $\mathcal{M}^0$ are bounded (*i.e.* in $\mathcal{M}$), but also that they have t–degree 0: see definition 1, §3.
Furthermore:

$$\vec{F}^{h\sigma}_{\uparrow\vec{0}} = \vec{0} \tag{4.15}$$

As in A) the property that $\vec{F}^{h}_{\uparrow\vec{0}} = \vec{0}$ is derived from the KAM theorem, (as done in (4.9)). But also in this case one can check it directly. In this way we free ourselves from the KAM theorem and we are in a position to study it independently. This means that the property $\vec{F}^{k\sigma}_{\uparrow\vec{0}} = \vec{0}$ and the (4.13),(4.14), coupled with the initial condition $\vec{F}^1_{\uparrow}(t) = \partial_{\vec{\alpha}} f(\vec{\omega}t)$, (see (2.9)), *always* have a solution with $\vec{X}^h$ quasi periodic (with spectrum $\vec{\omega}$). Such solution gives us a formal power series solution to the problem of finding the invariant tori equations. In this case we shall discuss and prove the convergence of such series. The first proof ("non KAM") of (4.15) is due to [CZ]: in §6 we shall find a more direct proof of (4.15), see remark 3) after (6.18): which is very simple although far less general than [CZ]. But for clarity of exposition we prefer to appeal to [CZ] and continue, without stopping to prove (4.15).

The above statements do dot require that J be a constant: it can be, for instance, a diagonal matrix (of dimension $l-1$), provided we interpret conveniently the multiplications by J^{-1}. It is then interesting to note that all the above statements remain true if some, or all, the elements of the diagonal matrix J become infinite. This is not difficult to check by going through the classical proof of KAM theorem, (it is, in fact, not easy to find it explicitly stated except in the case in which either only one of the J_i is infinite ("periodically forced systems") or all of them are infinite). However one finds a result with a analyticity radius (in μ) which, as mentioned in the introduction, *is not bounded away from* 0 *uniformly in the size of the largest among the non infinite* J_i*'s.* Although this is nevertheless sufficient to prove that $\vec{F}^{k\sigma}_{\uparrow\vec{0}} = \vec{0}$, it is reassuring that the algebraic check, that we refer to above, is independently possible.
If we let J be a diagonal matrix and we allow for some, or all, its elements to be $+\infty$ and if we set $\eta \equiv 1$ in (1.2), and ask whether the torus $\vec{A} = \vec{0}$ can be analytically continued for μ complex, $|\mu| < b^{-1}$, with b being J–independent, then we say that we are considering a "twistless KAM" problem.
The convergence of the formal series for the tori equations studied in §7 will yield a radius of convergence, b^{-1}, independent on the J_j (a result stronger than the usual KAM theorem relying on the twist property). Hence we show, by direct bounds, that the just posed twistless KAM problem has a solution (a fact that could be checked also by a careful exam of some of the classical proof of KAM theore, as mentioned in the introduction).
This is less surprising if one studies the $J = +\infty$ case, see [G1]: problems 1,16,17 §5.10, showing that the system is completely integrable. In the language of the present work the quoted reference [G1] can be easily worked out: one finds that $\vec{X}^h_{\downarrow} \equiv \vec{0}$ if $h \geq 1$ and $\vec{F}^h_{\uparrow} \equiv \vec{0}, \vec{X}^h_{\uparrow} \equiv \vec{0}$ if $h \geq 2$, while $\vec{X}^1_{\uparrow}(t) = \mathcal{I}(\partial_{\vec{\alpha}} f)(t)$, so that:

$$\vec{A} = \vec{A}(\vec{\alpha}) \equiv -\mu \sum_{\vec{\nu}\neq\vec{0}} i\vec{\nu} f_{\vec{\nu}} \frac{e^{i\vec{\omega}\cdot\vec{\nu}t} e^{i\vec{\alpha}\cdot\vec{\nu}}}{i\vec{\omega}\cdot\vec{\nu}}, \qquad \vec{\alpha} \equiv \vec{\alpha} \in T^{l-1} \tag{4.16}$$

are the equations of the invariant, twistless, torus arising from the unperturbed $\vec{A} = \vec{0}$ torus.

C) *Parity properties. (I).*

We close this section by pointing out a few *parity properties* of the operator $\mathcal{I}$, which will be useful below (see [CG], §10).
Suppose that M depends on other $l-1$ dimensional angles $\vec{\alpha}$ as a linear combination (with σ–independent coefficients) of monomials:

$$\frac{(g\sigma t)^p}{p!} x^k \sigma^\chi \cos_{\chi'}(\vec{\omega}\cdot\vec{\nu}t + \vec{\alpha}\cdot\vec{\mu}) \equiv \frac{(g\sigma t)^p}{p!} x^k \sigma^\chi \frac{1}{2i^{\chi'}} \sum_{\rho=\pm1} \rho^{\chi'} e^{i\rho(\vec{\omega}\cdot\vec{\nu}t + \vec{\mu}\cdot\vec{\alpha})} \tag{4.17}$$

with $\chi, \chi' = 0, 1$ and $\cos_{\chi'} y = \cos y$ if $\chi' = 0$ and $\cos_{\chi'} y = \sin y$ if $\chi' = 1$, then $\mathcal{I}M$ has the same form.
We shall say that M is *time-angle even* if $\chi + \chi' =$ even for all monomials of M. If, instead, $\chi + \chi' =$ odd for all monomials we say that M is *time-angle odd.* It then follows that the time angle parities of M and $\mathcal{I}M$ are opposite (when either is well defined).

After the above remarks we make the inductive assumption that $F^{h\sigma}(t;\vec\alpha)$ has action components $(+,\uparrow)$, denoted symbolically d, of odd time-angle parity in the above sense and angle components $(-,\downarrow)$, denoted p, of even time angle parity. Opposite parity assumptions will be made for $X^{h\sigma}(t;\vec\alpha)$. We shall write:

$$F^h = \begin{pmatrix} p \\ d \end{pmatrix}, \quad X^h = \begin{pmatrix} d \\ p \end{pmatrix} \tag{4.18}$$

dropping the label σ from F and X. In fact the main goal of the above formalism is to treat simultaneously the stable and the unstable whiskers: for $t>0$ it is $\sigma=1$ and F^h, X^h represent F^{h+}, X^{h+} while for $t<0$, $\sigma=-1$ and F^h, X^h represent F^{h-}, X^{h-}. Hence we can symbolically write:

$$\begin{aligned} F^h &= \sum \delta\, x^k\, (g\sigma t)^{k'} \sigma^\chi \cos_{\chi'}(\vec\omega\cdot\vec\nu t + \vec\alpha\cdot\vec\mu) \\ X^h &= \sum \xi\, x^k\, (g\sigma t)^{k'} \sigma^\chi \cos_{\chi'}(\vec\omega\cdot\vec\nu t + \vec\alpha\cdot\vec\mu) \end{aligned} \tag{4.19}$$

with suitable σ-independent coefficients δ, ξ and $\chi+\chi' =$ even for the $(+,\uparrow)$ components and odd for the $(-,\downarrow)$ components in the case of X, and with reversed parities in the case of F (the symbol $\cos_\chi y$ being defined after (4.17)).

By remarking that $X(t)$ can be expressed via the wronskian:

$$X(t) = W(t)\Big(X(0) + \int_0^t W(\tau)^{-1} F(\tau) d\tau\Big) \tag{4.20}$$

where the wronskian matrices $W(t)$ and $W(t)^{-1}$ are respectively:

$$(w_{jq}) \equiv \begin{pmatrix} w_{00}(t) & 0 & w_{0l}(t) & 0 \\ 0 & 1 & 0 & J^{-1}t \\ w_{l0}(t) & 0 & w_{ll}(t) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (w^{-1}_{jq}) \equiv \begin{pmatrix} w_{ll}(t) & 0 & -w_{0l}(t) & 0 \\ 0 & 1 & 0 & -J^{-1}t \\ -w_{l0}(t) & 0 & w_{00}(t) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.21}$$

(w_{jq}, $j,q=0,l$, being the matrix in (2.15), with w_{jj} even in t and w_{l0}, w_{0l} odd in t), one deduces immediately from the above property of $\mathcal{I}$ that X^h will have the *opposite* structure to F^h (*i.e.* if $F^h = \begin{pmatrix} p \\ d \end{pmatrix}$ then $X^h = \begin{pmatrix} d \\ p \end{pmatrix}$).

The above remark and (2.11) imply that if $X^{h'}$ has the structure $\begin{pmatrix} d \\ p \end{pmatrix}$ for $h'<h$ then F^h has $\begin{pmatrix} p \\ d \end{pmatrix}$ structure. And since it is obvious that F^1 has $\begin{pmatrix} p \\ d \end{pmatrix}$ structure, the (4.18) follows by induction.

An important consequence of the parity of $X^k_+, X^k_\uparrow$ is that if $\vec\alpha=\vec 0$ they are even functions of t so that, see (4.17), if $\chi=0$ it is $\chi'=1$ and viceversa. Hence $X^{k+}_j(0,\vec 0) = X^{k-}_j(0,\vec 0)$ and we see that $\vec\alpha=\vec 0, \varphi=\pi$ is a homoclinic point (and more precisely $\alpha_j=0,\pi$, and $\varphi=\pi$ ar 2^{l-1} homoclinic points. In the following we study the homoclinic point $\vec\alpha=\vec 0, \varphi=\pi$.

D) *Parity properties. (II).*

It is important to study the following operators:

$$\begin{aligned} U\, f(t) &= \frac12 \sum_{\rho=\pm} \int_{\rho\infty}^t J^{-1}(t-\tau) f(\tau)\, d\tau \\ V\, f(t) &= \frac12 \sum_{\rho=\pm} \int_{\rho\infty}^t (w_{0l}(t) w_{00}(\tau) - w_{0l}(\tau) w_{00}(t)) f(\tau)\, d\tau \end{aligned} \tag{4.22}$$

Then one checks that U, V *preserve the time angle parity of f as well as the holomorphy.* This is not the case of the operation $\mathcal{I}$: the latter has *floating* integration axtremes (depending on the siign of t): therefore it maps analytic functions of t into functions with a possible non analyticity at $\mathrm{Re}\, t=0$, unless the function is odd in t.

Hence if f is analytic in t in a strip $\operatorname{Im} t < \xi g^{-1}$ and time angle even it will be:

$$U\, f(0^\sigma) = \sigma \sum_{\vec\nu} u^1_{\vec\nu} \cos a\, \vec\alpha \cdot \vec\nu + \sum_{\vec\nu} u_{\vec\nu}\, \sin \vec\alpha \cdot \vec\nu \tag{4.23}$$

for suitable coefficents u^1, u; and similarly for V. But (4.23) has to be analytic; hence:

$$U\, f(0) = \sum_{\vec\nu} u_{\vec\nu} \sin \vec\alpha \cdot \vec\nu, \qquad V\, f(0) = \sum_{\vec\nu} v_{\vec\nu} \sin \vec\alpha \cdot \vec\nu \tag{4.24}$$

for suitable coefficients $u_{\vec\nu}, v_{\vec\nu}$. If f is analytic and odd we exchange the role of sines and cosines.

Note, as implicit in the discussion of C) and D) above, that if f time angle even and analytic then $\mathcal{I} f(0^\sigma)$ has the form $\sum_{\vec\nu} (f^1_{\vec\nu} \sin \vec\alpha \cdot \vec\nu + \, f^2_{\vec\nu} \sigma \cos \vec\alpha \cdot \vec\nu)$, *i.e.* the generally present (in the non analytic cases) part $\sum_{\vec\nu}\, f^2_{\vec\nu} \sigma \cos \vec\alpha \cdot \vec\nu$ is not necessarily 0; and if f is time–angle odd and analytic then $\mathcal{I} f(0^\sigma) = \sum_{\vec\nu} (f^1_{\vec\nu}\, \sigma\, \sin \vec\alpha \cdot \vec\nu + f^2 \vec\nu \cos \vec\alpha)$: *i.e.* "no simplification" occurs because of analyticity.

We see that the parity properties, although very simple, can become quite intricate to visualize.

§5 Trees, roots, nodes, branches and fruits: the formalism.

We develop a graphical formalism to represent, via (4.10) and (2.14), the generic h-th order contribution to various quantities related to the invariant tori and their whiskers.

We shall consider, for instance, the whiskers splitting in model 2), (1.1), at the point with coordinates $\varphi = 0$ and $\vec\alpha$, which is $\Delta^k_j(\vec\alpha) = X^{k+}_j(0;\vec\alpha) - X^{k-}_j(0;\vec\alpha)$ if j is an action component subscript (hence a homoclinic point (at $\varphi = 0$) corresponds to the values $\vec A$ such that $\Delta^k_j(\vec\alpha) = 0$). We shall also consider, as a second example, the *homoclinic scattering phase shifts* (a notion introduced in [CG]), and as a third example the parametric equations of the invariant tori in model 1).

The case of model 1) can be regarded as a special case of model 2), with f independent on the pendulum position φ, hence the two problems can be treated with the same formalism: although the latter is simpler and, if independently formulated, would require a slightly easier analysis.

Recall our label convention, in (2.7), for the action angle variables: we label with $j = 0$ the φ angle, with $j = 1, \ldots, l-1$ the angles $\alpha_1, \ldots, \alpha_{l-1}$ and we label with $j = l$ the pendulum action I and with $j = l+1, \ldots, 2l-1$ the rotators actions $\vec A$. The F^k are given by (2.14) and the X^k are related to the F^k by (4.10).

Note that the action angle labels are enumerated, between 0 and $2l-1$, in the order in which they appear in (4.10): because in (2.14) *a subscript $-$ or $+$ is synonimous of a subscript 0 or l; and $\downarrow$ or $\uparrow$ denote, collectively, the subscripts $j = 1, \ldots, l-1$ or $j > l$.* Furthermore F^k depends only on X^h_j, $h < k$, with $j = 0, \ldots, l-1$, (*i.e.* it depends only on the the lower order $-,\downarrow$ components of X^k; this is a consequence of the f in (1.1) depending only on the angles $\varphi, \vec\alpha$).

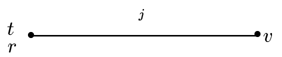
We imagine to use (2.14) recursively to express everything in terms of F^1 only. The structure of (2.14), (or, more generally, (2.11); see also [G1] Ch. 5, §11), leads us naturally to a graphical representation, quite familiar in perturbation theories, see [G2]. This fact was noted explicitly in the context of KAM theory by [E], (see also [V]).

We can see a tree grow out, of (2.14), (4.10) as follows.

First we represent $F^h_j(\tau)$ as a fat point v:

$$F^h_j(\tau) = \quad {}_v\bullet^{\tau,h,j} \tag{5.1}$$

Then we consider the improper integration operations with upper limit t, denoted $\mathcal{O}$, $J^{-1}\overline{\mathcal{I}}^2$, $\mathcal{O}_+$, $\mathcal{I}$ in (4.10). We represent them with a line segment of unit length, called *branch*, joining two points r, v, and calling r the *branch root*, and v the *branch node*. The branch will bear a *branch label* $j = 0$ when representing $\mathcal{O}$, or a label $j = 1, \ldots, l-1$ for $J^{-1}\overline{\mathcal{I}}^2$, or $j = l$ for $\mathcal{O}_+$, or $j = l+1, \ldots, 2l-1$ for $\mathcal{I}$, see (4.10). A label t will be attached to the root r:



(5.2)

Then we can represent the whole formula (4.10) as:

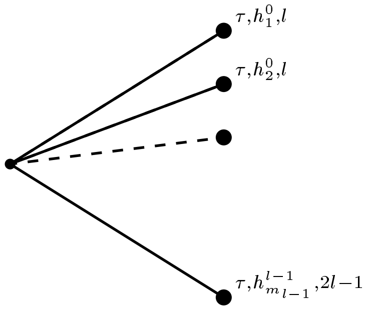
$$X_j^h(t) = {}^{t}_{r}\bullet\!\!\xrightarrow{\;\;\;\;\;\;\;\;\;j\;\;\;\;\;\;\;\;\;}\!\!\bullet^{\tau,h,\tilde{j}}_{v} \tag{5.3}$$

and there is an implicit constraint as $j, \tilde{j}$ cannot be independent, as one sees from (4.10): in fact $j = 0 \Rightarrow \tilde{j} = l$ (first or third line in (4.10)); and $j = 1, \ldots, l-1 \Rightarrow \tilde{j} = j + l$ (second line in (4.10)); and $j \geq l \Rightarrow \tilde{j} = j$ (fourth line in (4.10).

Going back to the representation of $F_j^h(\tau)$ as a fat point, (5.1), we can use (2.14) to express it in terms of $X_i^{h_p^i}(\tau)$ with $\sum_{i=0}^{l}\sum_{p=1}^{m_i} h_p^i = h$ or $h-1$.

Using the representation (5.1) for $X_s^{h_p^s}(\tau)$ we see that (2.14) can be written, $(j \geq l)$:

$$F_j^h(\tau) = \;{}_v\bullet^{\tau,h,j} \;=\; \sum \frac{1}{\prod_{i=0}^{l-1} m_i!} \quad {}_v \tag{5.4}$$



where the first m_0 branches are labeled 0, the next m_1 are labeled 1, etc.; $\delta = 0,1$ and $\sum h_p^i = h - \delta$, because of the meaning of the symbols $(k_p^i)_{\vec{m},k}, (k_p^i)_{\vec{m},k-1}$ in (2.14).

The vertex v, that we call a *node*, corresponds to the factors $\vec{m}!\,(g^2 J_0 \sin\varphi)_{\vec{m}}(\varphi^0)$ and $\vec{m}!\,(\partial f)_{\vec{m}}$ appearing in (2.14) and deprived of the combinatorial factor $\vec{m}! = \prod_{i=0}^{l-1} m_i!$, see (2.14), (2.12). And the label ν is introduced to split such factors as sums of their Fourier components. Namely, see (1.1), let:

$$\begin{aligned}
&f^\delta(\vec{\alpha},\varphi) \equiv \sum_{\nu=(n,\vec{\nu})} f_\nu^\delta\, e^{i(n\varphi+\vec{\nu}\cdot\vec{\alpha})}, \qquad \delta = 0,1 \\
&f^0(\varphi,\vec{\alpha}) \equiv J_0 g^2 \cos\varphi = \sum_{\substack{\nu,\ \vec{\nu}=\vec{0} \\ n=\pm 1}} \frac{f_\nu^0}{2}\, e^{in\varphi}, \qquad f^1(\varphi,\vec{\alpha}) \equiv \sum_{\nu} \frac{f_\nu^1}{2}\, e^{i(n\varphi+\vec{\nu}\cdot\vec{\alpha})}
\end{aligned} \tag{5.5}$$

then the factor represented by the node v, with its labels τ, ν, δ, j is:

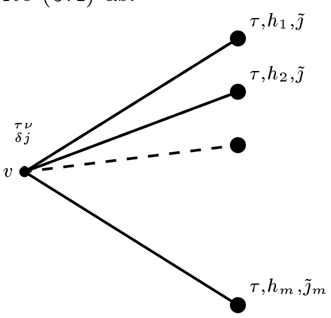
$$\frac{i\nu_{j-l}}{2} \cdot \Big(\prod_{s=1}^{l-1} (i\nu_s)^{m_s}\Big)\, f_\nu^\delta e^{i(n\varphi^0(\tau)+\vec{\nu}\cdot(\vec{\alpha}+\vec{\omega}\tau))} \tag{5.6}$$

where $j - l$ is the branch label of the branch leading to v, and the introduction of the above Fourier representation is convenient as it eliminates the derivatives with respect to $\varphi, \vec{\alpha}$ in the coefficients of (2.14).

We call τ a *time label*, ν a mode label, δ an *order* label and $\tilde{j}$ an *action* label associated with the node v.

One can further simplify the (5.4) by imagining that the $m = \sum_{i=0}^{l-1} m_i$ branches coming out of v are distinguished by a number counting them from 1 to m and appended to them, which however is not explicitly written and is thought as associated with the branch. Then we drop the condition that the branch labels $j = 0, \ldots, l-1$ are in non decreasing order and rewrite (5.4) as:

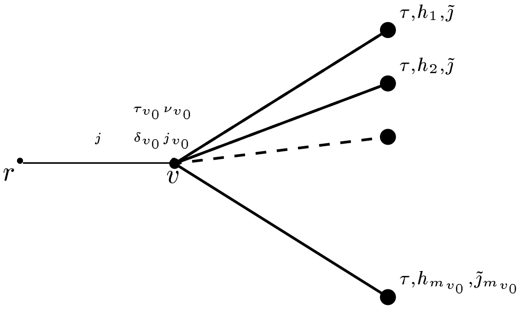
$$F_j^h(\tau) = \;{}_v\bullet^{\tau,h,j} \;=\; \sum \frac{1}{m!} \quad {}_v^{\tau\nu}_{\delta j} \tag{5.7}$$



so that we can put freely the labels without worrying about the order (if we did not replace the combinatorial factor in (5.4) with that of (5.7) we would count the same term $m!\prod_{i=0}^{l-1} m_i!^{-1}$ times and this explains the change in the combinatorial factor).

It is now easy to write (4.10) with the F^h in it expressed in terms of lower order $X^{h'}$, $h' < h$, via (2.14): thinking X^h as written in (5.3) and (2.14) as written in (5.7), we get:

$$X_j^h(t) = \sum \frac{1}{m_{v_0}!} \tag{5.8}$$



where the sum is over all the possible choices of the labels, subject to the conditions described in their definitions ($e.g.$ $j_{v_0} = \tilde{\jmath}$).

We can now iterate the representation (5.8) by replacing the fat points by nodes out of which new branches emerge, until we reach fat points with order labels $h = 1$, where the expansion ends and we call the last fat points *top nodes*, drawing them as slim as the other nodes. The root r of the first branch drawn will not be regarded as a node, but it will be called *root*.

We see that the above expansion leads to a sum over *trees* bearing "decorations": such trees have branches emerging from the same node which are regarded as distinguished by a number, never marked in the drawings, see fig. 1 below. On such trees a natural group of transformations acts: it is generated by the following operations; fix a node v and permute the subtrees emerging from it. Two trees that can be transformed into each other by a transformation of the group just defined will be regarded as identical.

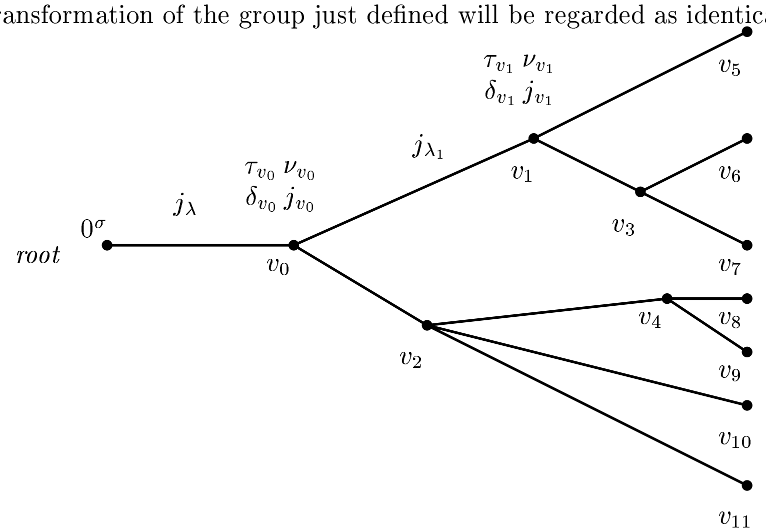


fig. 1: A tree ϑ with $m_{v_0} = 2, m_{v_1} = 2, m_{v_2} = 3, m_{v_3} = 2, m_{v_4} = 2$ and $m = 12$, $n(\vartheta)$=$n(\vartheta_1)n(\vartheta_2)$ $= (1\cdot 1^2)(2!2!) = 4$, $\prod m_v! = 2^4\cdot 6$, and some decorations.

We call the above trees *semitopological trees*: we call *topological trees* the trees in which the branches emerging from the same node are not numbered, and also such trees will be identified when superposable modulo the action of the same group of transformations.

A third type of trees, that we call "numbered trees" or simply *trees*, are obtained by imagining to have a deposit of m branches numbered from 1 to m and depositing them on the branches of a topological tree with m branches: numbered trees that can be superposed by transforming them with the above group of tranfsormations will still be regarded as identical. The number of such trees asociated with a given semitopological tree is $m!\prod_v m_v!^{-1}$ and using such trees, which will be the only ones considered from now on, unless otherwise stated, the (5.8) becomes:

$$X_j^h(t) = \sum_{\vartheta\in trees} \frac{1}{m(\vartheta)!} \sum_{labels;\ \sum_v \delta_v = h} \Big[\text{Fig. 1}\Big] \tag{5.9}$$

where $m(\vartheta)$ = number of branches of ϑ; and the drawing of Fig. 1 symbolizes, now, a well defined hierarchical chain of operations of improper integrations.[4]

The expression (5.9) is very convenient, as we shall see, from a combinatorial point of view. It is however clear tha the summation in (5.9) over the labels produces a number $w(\vartheta)$ depending *only* on the topological tree associated with ϑ, *i.e.* it does not depend on the branch identification numbers (hidden in the pictures) of the branches. Therefore in real calculations the above sum over trees can be replaced by a single sum over semitopological trees, or by an even simpler sum over topological trees.[5]

Let us fix some terminology and conventions: given a tree ϑ and a family of labels for it we denote Θ their pair and call it a *labeled tree.*

(a) the trees are dawn from left to right and are regarded as partially ordered sets of nodes in the obvious sense (see footnote [4] above), with the higher nodes to the right (*i.e.* we draw them as "fallen trees", unfortunately: the vertical notation would have required too much space). The branches are naturally ordered as well; all of them have two nodes at their extremes (possibly one of them is a top node) except the lowest or *first* branch which has only one node, the first node v_0 of the tree. The other extreme r of the first branch, (which is the root of the branch), will be called the *root* of the tree and it *will not* be regarded as a node. If v_1 and v_2 are two nodes we say that $v_1 < v_2$ if v_2 follows v_1 in the order established by the tree: *i.e.* if on has to pass v_1 before reaching v_2, while climbing the tree.

(b) each node carries a *time label* τ_v, a *mode label* ν_v, and *order label* δ_v and an *action label* $\tilde{\jmath}_v = l, \ldots, 2l-1$. Each branch carries and *angle label,* $j_\lambda = 0, \ldots, l-1$; but if λ is the root branch its label can also be an *action label* $j_\lambda \ge l$. If λ leads to v then $j_\lambda = \tilde{\jmath}_v + l$ if $j_\lambda = 0, \ldots, l-1$, while $\tilde{\jmath}_v = j_v$ if $j_\lambda = l, \ldots, 2l-1$ (which is allowed only if λ is the root branch).

(c) the *order* $h(\Theta)$ of the labeled tree Θ is $h = \sum_v \delta_v$, *i.e.* the sum of the order labels of the nodes. The number of branches emerging from the node v is $m = 1 + \sum_v m_v$ (as one has to count also the root branch). Of course, as the order label $\delta_v = 0, 1$ and as each node v with $\delta_v = 0$ *must* have $m_v \ge 2$, (see above and (2.14)), it is $h \le m < 2h$.

---

[4] More formally a topological tree is a partially ordered set with the property that any two elements follow some common predecessor (hence there is a minimum element, or *root*) and have no common follower if they are not comparable. This is visualized (see fig. 1 above) by representing its elements by oriented unit segments. We draw the first segment (in the partial order) and attach to its endpoint the $m_{v_0} \ge 0$ segments representing its immediate followers, and so on. The endpoints of the segments become *nodes* of the tree: to each node v is associated a *branching number* $m_v \ge 0$ as the segments will be called *branches.* The first branch will be called the *root branch* and its first point will not be a node, but it will, nevertheless, be called the *root.* The endpoint v_0 of the root branch will be called the *first node.* Two topological trees giving rise to the same partially ordered set will be considered identical. Each node v with $m_v > 0$ can be regarded as the root of a *subtree*: and the operation of graphical permutation of two subtrees emerging from the same node establishes an equivalence relation between trees which is a notion coinciding with that of giving rise to the same partially ordered set.
It is clear that the set of topological trees with m branches contains at most as many elements as the random walks on the lattice of the integers ≥ 1 with $2m-2$ steps: hence the number of topological trees with m branches is bounded by 2^{2m-2}.

[5] For this purpose the following identity is used. A natural combinatorial factor $n(\vartheta)$ that can be associated with a topological tree ϑ that bifurcates at the first node v_0 into m_{v_0} subtrees among which there are only q topologically different subtrees $\vartheta_1, \ldots, \vartheta_q$ each of which is repeated $p_1, \ldots, p_q$ times, is:

$$n(\vartheta) = \textstyle\prod_{i=1}^q p_i!\, n(\vartheta_i)^{p_i} \tag{5.10}$$

Such combinatorial factors are useful when one has to consider sums over trees ϑ with m branches (root branch included) of functions $F(\vartheta)$ whose values depend only on the topological tree. In such cases one finds:

$$\tfrac{1}{m!} \textstyle\sum_{\vartheta\in\, trees} F(\vartheta) \equiv \sum_{\vartheta\in\frac{1}{2}topological} (\prod_{v\in\vartheta} \frac{1}{m_v!})\, F(\vartheta) \equiv \sum_{\vartheta\in topological} \frac{1}{n(\vartheta)}\, F(\vartheta) \tag{5.11}$$

The above identity is closely related to Cayley's formulae, see also [G2], (6.1), (6.2) and (5.13)), and [FG]. It simply reflects that the number $N(m; \{m_v\})$ of trees with m branches and branching numbers m_v, hence generating the same semitopological tree, is $m! \prod_v m_v!^{-1}$ (which is (an adaptation of) Cayley's formula, (see [H], §1.7)), while the number of trees generating the same topological tree ϑ is $m!\, n(\vartheta)^{-1}$.

(d) the *momentum* of a node v or of the branch λ_v leading to v is $\vec\nu(v)=\sum_{w\ge v}\vec\nu_w$, if $\nu_v=(n_v,\vec\nu_v)$ is the *mode label* of v. The *total momentum* is $\vec\nu(v_0)\equiv\sum_{v\ge v_0}\vec\nu_v$.

(e) Given all the above decorations on a labeled tree Θ *we define its value* $V_j(t;\Theta)$ via the following operations: (1) We first lay down a set of parentheses () ordered hierarchically and reproducing the tree structure (in fact any ordered (topological) tree can be represented as a set of matching parentheses representing the tree nodes). Matching parentheses corresponding to a node v will be made easy to see by appending to them a label v. The root will not be represented by a (unnecessary) parenthesis.
(2) Inside the parenthesis $(_v$ and next to it we write, setting $\tilde j_v\equiv j_v-l$, the *node function*, see (5.6):

$$\Big(_{v}\; -\tfrac12 i(\nu_v)_{\tilde j_v}\, f^{\delta_v}_{\nu_v} e^{i(n_v\varphi(\tau_v)+(\vec\alpha+\vec\omega\tau_v)\cdot\vec\nu)}\,\prod_{s=0}^{l-1}(i\nu_{vs})^{m^s_v} \tag{5.12}$$

here j_v is the branch label of the branch leading to v. The last product is missing if no nodes follow v. (3) Furthermore out of $(_v$ and next to it we write a symbol $\mathcal{E}^T_v$ which we interpret differently, depending on the label $j'=j_{\lambda_v}$ on λ_v and on the action label $j''=\tilde j_v$ on v:

$$\mathcal{E}^T_v\Big(_v\,\cdot\,\Big)_v\equiv\begin{cases}\mathcal{O}\Big(_v\,\cdot\,\Big)_v(\tau_{v'}), & \text{if } v>v_0\ ,\quad j_{\lambda_v}=0;\\ J^{-1}\overline{\mathcal{I}}^2\Big(_v\,\cdot\,\Big)_v(\tau_{v'})\ , & \text{if } v>v_0\ ,\quad 1\le j_{\lambda_v}\le l-1;\end{cases} \tag{5.13}$$

for $v>v_0$, otherwise:

$$\mathcal{E}^T_v\Big(_v\,\cdot\,\Big)_v\equiv\begin{cases}\mathcal{O}\Big(_v\,\cdot\,\Big)_v(t^\sigma)\ , & \text{if } v=v_0\ ,\quad j_{\lambda_v}=0\ ,\\ J^{-1}\overline{\mathcal{I}}^2\Big(_v\,\cdot\,\Big)_v(t^\sigma) & \text{if } v=v_0\ ,\quad 1\le j_{\lambda_v}\le l-1\ ,\\ \mathcal{O}_+\Big(_v\,\cdot\,\Big)_v(t^\sigma)\ , & \text{if } v=v_0\ ,\quad j_{\lambda_v}=l;\\ \mathcal{I}\Big(_v\,\cdot\,\Big)_v(t^\sigma)\ , & \text{if } v=v_0\ ,\quad l+1\le j_{\lambda_v}\le 2l-1\end{cases} \tag{5.14}$$

t being the root time label of the tree and the superscript σ attached to t is important only if $t=0$: in such case (5.14) has to be interpreted as the limit as $t\to0^\sigma$.

Remarks:

1) One realizes that the giant symbol thus constructed has a perfectly defined meaning as a hierarchically ordered chain of operations $\mathcal{I}$: it gives a "single contribution" to the value $X^h_{j\vec\nu}(t;\vec\alpha)\equiv X^h_{j\vec\nu}(t)e^{i\vec\alpha\cdot\vec\nu}$. For instance: given $\sigma=\pm$ multiply σ times all the above values of the trees Θ with order $h(\Theta)=h$, momentum $\vec\nu$, and $j>l,t=0^\sigma$. *Such sum will give the h-th order contribution* $\Delta^h_j=X^h_j(0^+;\vec\alpha)-X^h_j(0^-;\vec\alpha)$ *to the homoclinic splitting.*

2): If we do not perform the operation $\mathcal{E}^T$ relative to the time τ_{v_0} of the first node v_0 and set it to be equal to t, setting also $j\equiv j_{v_0}$, we see that the result is a representation of $F^h_{j_{v_0}}(t)$, if j_{v_0} is the label of the node v_0.

3): Note that if $\vec\alpha=\vec0$ then we are at a homoclinic point, because the hamiltonian (1.1) is even: so that the sum in remark 1) above yields a value 0 for all components $j=l,\ldots 2l-1$, see the final comments in §4, C), and all fixed tree shapes.

4): Suppose that one whishes to study model 1) in (1.1). For instance suppose one wants to determine the parametric equations for the invariant torus run quasi periodically with angular velocity $\vec\omega$, analytically continuing the unperturbed torus $\vec A=\vec0$. The (4.10) must be replaced by (4.13) and (2.14) by (4.14), while (4.9) remains the same. But *very little* has to be changed in the above graphical representations.
In fact it is sufficient to consider, in the sum $X^h_{j,\vec\nu}(t)=\sum_{\Theta:\,h(\Theta)=h}V_{j\vec\nu}(t;\Theta)e^{i\vec\nu\cdot\vec\alpha}$, only trees Θ with *no node or branch labels* j equal to $0,l$ (*i.e.* all the labels have to be rotators labels); and $\delta_v\equiv1$, so that $m(\Theta)=h(\Theta)=h$. This is obvious because we can think that the pendulum is present but that it is completely decoupled from the rotators.
The functions $\vec X^h_\downarrow(t),\vec X^h_\uparrow(t)$ thus defined are, of course, the solutions of the equations of motion of a point starting at $t=0$ on the above invariant torus $\mathcal{T}$ at a angular position $\vec\alpha$ and at action position $\vec A_\uparrow(0)=\sum_{h=1}^\infty\mu^h\vec X^h_\uparrow(0)$, for μ small enough (this is a convergent series for μ small, at fixed J_j, by KAM, as already mentioned): if, in view of the aim of *proving* the theorem by directly showing the convergence of the series, one does not want to use the KAM theorem then the latter series will only be a formal power series solution

to the problem of finding a motion on the invariant torus $\mathcal{T}$.
Therefore setting:

$$\vec{\alpha}(t)=\vec{\alpha}+\vec{\omega}t+\sum_{h=1}^{\infty}\mu^h\vec{X}_{\downarrow}(t),\qquad \vec{A}(t)=\sum_{h=1}^{\infty}\mu^h\vec{X}_{\uparrow}(t) \tag{5.15}$$

and letting t vary in $[0,+\infty)$, or $(-\infty,0]$ (or, for that matter, in any infinite connected subinterval of the above two) we describe (at least formally) a dense set on the torus; *i.e.* the torus itself.
The formalism necessary to see the cancellations, and to make use of them, is completed.

§6 Cancellation mechanisms.

A): *Homoclinic cancellations.*
1) (representation of $\mathcal{E}^T$):
To study cancellation mechanisms for the whiskers splitting we consider trees with root time $t=0$, (*i.e.* $t=0^\sigma$), and we introduce a useful *representation of the $\mathcal{E}$ operations.* Given a tree ϑ contributing to order h and a total momentum $\vec{\nu}$, let v be a node. We fix our attention on one such node v and call v' its predecessor: the case in which v is a top node and the predecessor v' of v is the root is the simple first order case, so that this case will not be considered as it can be studied easily by direct evaluation and no cancellations occurr in it, in general, (hence v' will not be the root and $v>v_0$).
Supposing that v' is not the root, we realize that the "integral" describing the $\mathcal{I}$ operation (associated with the vertex v) can be written, if $j=j_{\lambda_v}$ is a $\downarrow$ component (*i.e.* if $0<j\le l-1$, see (4.6)):

$$J^{-1}\Big(\fint_{(\sigma)}^{\tau_{v'}}(\tau_{v'}-\tau_v)S_{v,\vartheta}(\tau_v)d\tau_v-\fint_{(\sigma)}^{0}(-\tau_v)S_{v,\vartheta}(\tau_v)d\tau_v\Big) \tag{6.1}$$

with $S_{v,\vartheta}(\tau_v)$ being the result of the operations performed to evaluate the integrals and sums inside the parentheses $({}_v)_v$. Note that the $S_{v,\vartheta}(\tau)$ depends only on the subtree $\vartheta_v\subset\vartheta$ rooted at v' and consisting of the nodes following v in ϑ (and bearing all the decorations of ϑ).
Calling $t\equiv\tau_{v'}$, $\tau\equiv\tau_v$, using (3.13),(3.14), we can replace the above expression by:

$$J^{-1}\Big(\chi(\sigma<0)t\fint_{-\infty}^{+\infty}S_{v,\vartheta}(\tau)d\tau+\fint_{+\infty}^{t}(t-\tau)S_{v,\vartheta}(\tau)\,d\tau-\fint_{+\infty}^{0}(-\tau)S_{v,\vartheta}(\tau)\,d\tau\Big) \tag{6.2}$$

Note that a rearrangement has been necessary to simplify the term with the characteristic function as well as the other; here we define $\fint_{-\infty}^{+\infty}$ to be identical to $\fint_{-\infty}^{0^-}-\fint_{+\infty}^{0^+}$, see (3.14).
A more symmetric representation is obtained by noting that a "mirror" formula must hold with $-\infty$ playing the role of $+\infty$. Averaging over the two formulae for the same quantity (6.1) we get:

$$\begin{aligned}&J^{-1}\Big(-\frac{1}{2}\sigma t\fint_{-\infty}^{+\infty}S_{v,\vartheta}(\tau)\,d\tau+\frac{1}{2}\Big[\fint_{+\infty}^{t}(t-\tau)S_{v,\vartheta}(\tau)\,d\tau+\fint_{-\infty}^{t}(t-\tau)S_{v,\vartheta}(\tau)\,d\tau\Big]+\\&+\frac{1}{2}\big(\fint_{+\infty}^{0}\tau S_{v,\vartheta}(\tau)d\tau+\fint_{-\infty}^{0}\tau S_{v,\vartheta}(\tau)d\tau\big)\Big)\end{aligned} \tag{6.3}$$

such symmetrization is not really needed, but it has some aestethic value.
A similar representation can be achieved also for the case $j=j_{\lambda_v}=0$:

$$\begin{aligned}&\chi(\sigma<0)w_{0l}(t)\int_{-\infty}^{+\infty}w_{00}(\tau)S_{v,\vartheta}(\tau)d\tau+\Big(w_{0l}(t)\int_{+\infty}^{t}w_{00}(\tau)S_{v,\vartheta}(\tau)d\tau-\\&-w_{00}(t)\fint_{+\infty}^{t}w_{0l}(\tau)S_{v,\vartheta}(\tau)d\tau\Big)+w_{00}(t)\fint_{+\infty}^{0}w_{0l}(\tau)S_{v,\vartheta}(\tau)d\tau\end{aligned} \tag{6.4}$$

see (2.17) and symmetrization over the choice of $\pm\infty$ yields the more symmetric formula:

$$\begin{aligned}&-\frac{1}{2}\sigma w_{0l}(t)\fint_{-\infty}^{+\infty}w_{00}(\tau)S_{v,\vartheta}(\tau)d\tau+\\&+\frac{1}{2}\Big[\fint_{+\infty}^{t}\Big(w_{0l}(t)w_{00}(\tau)-w_{00}(t)w_{0l}(\tau)\Big)S_{v,\vartheta}(\tau)d\tau+\fint_{-\infty}^{t}(\text{same})d\tau\Big]+\\&+\frac{1}{2}w_{00}(t)\Big(\fint_{+\infty}^{0}w_{0l}(\tau)S_{v,\vartheta}(\tau)d\tau+\fint_{-\infty}^{0}w_{0l}(\tau)S_{v,\vartheta}(\tau)d\tau\Big)\end{aligned} \tag{6.5}$$

We describe the above representation of the contribution of the considered tree branch by affixing a "bubble" around the node v and enclosing all the branches following it: the bubble will have a wiggly boundary or a smooth boundary to distinguish between the first and the third terms in the above sums. If a node is not enclosed in a bubble that cuts the branch linking it to the previous node then this means that in the above sum we selected instead the intermediate term: in this case we mark the node v by a label $\mathcal{R}$ (in analogy with renormalization theory, see [G2]; here $\mathcal{R}$ has nothing to do with the parameter R appearing in the definition of $\mathcal{I}$).

We can repeat the same representation operation on the (improper) integrations pertaining to all the other nodes, starting from the highest on the tree and going down towards the root: we stop at the first nodes following v_0 (the latter node is somewhat different from the others as the corresponding $\mathcal{E}^T$ integration has a un upper limit t fixed).

Thus the most general contribution to $X_j^h(t)$ by trees with order h will be represented by a tree with h nodes with index $\delta = 1$ (and up to $2h-1$ branches, see c) preceding (5.12)) and all the labels introduced so far ($\mathcal{R}$ labels included) plus an arbitrary number of bubbles drawn around tree nodes above v_0, and drawn as described (avoiding overlappings). Of course we could replace the bubbles with labels affixed on the nodes following the root: but we prefer the "bubbles notation" as it reminds us of the analogous bubbles that can be used to describe the renormalization cancellations in quantum field theory, (see [G2]).

We call *free* the branches that are not enclosed inside bubbles of any type. The total *free momentum* of the tree will be the sum of the node modes associated with the free nodes: *note that it is quite different from the total tree momentum previously defined (see (d)) preceding (5.12).*

Looking at (6.3),(6.5) we see that the first and the last terms contain constant factors which can be taken out of the integration operations associated with the other tree nodes. Thus the bubbles involve simpler integrals. Furthermore the three terms in the (6.3),(6.5) individually preserve the time–angle parities, as all the operations in §4 did: this is essential as it will permit us to use the parity arguments derived at the end of §4 (see C) and D) in §4).

2): (Resummations):

Consider a bubble containing a subtree of order h linked to its root node $v' \ge v_0$ by a branch $v'v$ carrying a momentum $\vec{\nu} = \sum_{w \ge v} \vec{\nu}(w)$ and a label j $(0 \le j \le l-1)$. Fixed $\vec{\nu}, j, h$ we can sum over the possible choices (consistent with the labels $\vec{\nu}, j, h$) of the subtree and its decorations:

$$\beta^{h,1}_{\vec{\nu}j} e^{i\vec{\nu}\cdot\vec{\alpha}} = \begin{cases} (Jg^2)^{-1} \sum \frac{1}{2} \fint_{-\infty}^{\infty} F^h_{\vec{\nu}j}(\tau)\, dg\tau \ , & j = 1, ..., l-1 \\ (J_0 g^2)^{-1} \sum \frac{1}{2} \fint_{-\infty}^{\infty} w_{00}(\tau) F^h_{\vec{\nu}j}(\tau)\, dg\tau \ , & j = 0 \end{cases} \tag{6.6}$$

$$\beta^{h,2}_{\vec{\nu}j} e^{i\vec{\nu}\cdot\vec{\alpha}} = \begin{cases} -(Jg^2)^{-1} \sum \frac{1}{2} \fint_{-\infty}^{+\infty} \sigma\, g\tau F^h_{\vec{\nu}j}(\tau)\, dg\tau \ , & j = 1, ..., l-1 \\ -(J_0 g^2)^{-1} \sum \frac{1}{2} \fint_{-\infty}^{+\infty} \sigma\, g\bar{w}_{0l}(\tau) F^h_{\vec{\nu}j}(\tau)\, dg\tau \ , & j = 0 \end{cases}$$

where $\bar{w}_{0l} \equiv (J_0 g)^{-1} w_{0l} = \frac{1}{4}\bar{w}$, see (2.15), is a convenient adimensional matrix element, and $F^h_{\vec{\nu}j}(\tau)$ is the function resulting from the resummation. The terms $\beta^{h,1}_{\vec{\nu}j}$ correspond to wiggly bubbles, while the $\beta^{h,2}_{\vec{\nu}j}$ correspond to smooth bubbles. The "value" of the subtrees on which we are summing is given according to the rules described in 1)÷ 4) above. The $\beta^{h,1}_{\vec{\nu}j}$ have the interpretation, which we leave to the reader to verify, of being proportional to the j-th component of the Fourier transform of mode $\vec{\nu}$ of the homoclinic splitting $\Delta^h(\vec{\alpha})$ (hint: just note that $-\fint_{-\infty}^{\infty} = \fint_{+\infty}^{0} - \fint_{-\infty}^{0}$ and recall the expression of $X^{h+}(0^+)$ and $X^{h-}(0^-)$ in terms of trees and the remark 2) following (5.14) which implies that $F^h_{\vec{\nu}j}$ in (6.6) is really the same function F^h considered in the previous sections, see (2.14)).

Taking care also of the constants of proportionality it is:

$$\Delta^h_{\vec{\nu},l} = J_0 g\, \beta^{h,1}_{\vec{\nu},0} \ , \qquad \Delta^h_{\vec{\nu},l+j} = Jg\, \beta^{h,1}_{\vec{\nu},j} \ , \quad (1 \le j \le l-1) \tag{6.7}$$

(note that the bubbles values are dimensionless).

The $\beta^{h,2}_{\vec{\nu}j}$ are new objects, *while the quantities $\sigma^h_{\vec{\nu}j}$ that one would obtain by eliminating the sign function σ from the integral $\fint$ in the definition of $\beta^{h,2}$ would be the j-th component of the scattering phase shift* (see [CG] for a definition of the notion of phase shift): however such interpretation will not play a role in our discussion; hence we do not pursue the enterprise of checking the latter statement.

The parity properties, discussed at the end of §4, allow us to conclude immediately that:

$$\Delta^h_{\vec{\nu}j} \equiv -\Delta^h_{-\vec{\nu}j}, \quad \beta^{h,2}_{\vec{\nu}j} \equiv -\beta^{h,2}_{-\vec{\nu}j}, \quad \sigma^h_{\vec{\nu}j} \equiv \sigma^h_{-\vec{\nu}j} \tag{6.8}$$

Hence: $\Delta^h_{\vec 0j}\equiv 0, \beta^{h,2}_{\vec 0j}\equiv 0$.

3): (Definition of resummation trees and of dry and ripe fruits):

The above remark suggests introducing a new type of trees, that we shall call "resummation trees", or $\mathcal{R}$-trees. A resummation tree, with all its decorations ($\mathcal{R}$ and bubbles included), is defined by drawing a tree with all its decorations and by *deleting* the contents of the bubbles, leaving only the branches connecting the top free nodes to the nodes v inside the outermost bubbles (if any). We shall leave the bubble around each v and on the branch leading to v we write the label j_{λ_v} and the total momentum $\vec\nu_{\lambda_v}$ of the deleted subtree together with the total order h_{λ_v} of the deleted subtree.

By construction the resummation trees have bubbles which can contain only one node v. It is natural to call such resummation bubbles *"fruits"* ("of the resummations" or "of the trees", as preferred). If the bubble is wiggly we call it a "dry fruit" while if it is smooth we call it a "ripe fruit"; the node of the fruit will be called the *seed* of the fruit. As we shall see the dry fruits values will be estimated, in our inductive construction of bounds on $\Delta^h_{\vec\nu}$, easily in terms of the inductive hypothesis, while the ripe fruits will, at each inductive step, be just "ripe to be bounded".

It is convenient to define the "value" of a resummation tree by the same prescription we used for the previously introduced trees, but changing the symbol corresponding to the seed v (preceded by v') from (5.12) to (see (6.3), (6.5), (6.6)):

$$x^1_j(\tau_{v'})\beta^{h,1}_{\vec\nu j}e^{i\vec\nu\cdot\vec\alpha}, \qquad \text{with} \qquad x^1_j(t)=\begin{cases}\sigma g t, & \text{if } j=1,\ldots,l-1\\ \sigma\bar w_{0l}(t), & \text{if } j=0\end{cases} \tag{6.9}$$

if the fruit surrounding the node is dry (the dimensionless matrix element $\bar w_{0l}$ is defined after (6.6)); or:

$$x^2_j(\tau_{v'})\beta^{h,2}_{\vec\nu j}e^{i\vec\nu\cdot\vec\alpha}, \qquad \text{with} \qquad x^2_j(t)=\begin{cases}1, & \text{if } j=1,\ldots,l-1\\ w_{00}(t), & \text{if } j=0\end{cases} \tag{6.10}$$

if the fruit is ripe.

The name *seed* is fairly appropriate as it really represnts a sum of appropiately evaluated tree values of trees having the seed has a root. "A seed can be magnified to reveal its content in trees" whose values add up to the value, (6.9),(6.10) of the fruit containing the seed (as it should be).

Finally we *renumber* such trees (as described at the beginning of the section) according to their topological structure (*i.e.* resummation trees do not inherit the numbering from the trees that generated them).

This is a good definition as we can now say that the sum of the values of the decorated resummation trees of given order is equal to the sum of the values of the previously considered trees of the same order. The non trivial part of such claim is dealing with the combinatorial factors; the correctness of the combinatorics is easily verified if the trees are regarded (as we are doing) as having all the branches distinct.

A simple cancellation can already be seen: consider the contribution to $F^h_{\vec\nu,j}(\tau)$ coming from trees *with* $\vec 0$ *total free momentum and even fruits number*: then given a tree we can consider another "conjugated" tree with the node momenta of the free nodes simultaneously changed in sign; summing their two contributions we see that the integrals to be performed are integrals of a time odd analytic function over the whole line: such integral will vanish (by the parity considerations of §4). Hence we have the mechanism:

I) Cancellation of the contribution to $\beta^{h,1}_{j\vec\nu}$ from the $\mathcal{R}$–trees with an even number of fruits, summing over the choices of a global sign multiplying the free node modes. *Hence trees without fruits and with* $\vec 0$ *free momentum do not contribute to the value of* $\beta^{h,1}_{j,\vec\nu}$, *i.e. to the Fourier transform (in the* $\vec\alpha$ *angles) of the splitting (note that* $\vec\nu$ *is the total momentum, hence in general it is* $\vec\nu\ne\vec 0$*).* We call this a "parity cancellation".

A deeper cancellation mechanism is:

II) Cancellation of the contribution to the splitting $b^{h,1}_{j,\vec\nu}$ from $\mathcal{R}$–trees carrying at least one fruit with label $j>0$, summing over all the possible ways of attaching the fruits to the tree. We call the latter the "KAM cancellations" as they extend the cancellations discussed below, leading to I), II) at the end of §6, and are sufficient to carry to KAM tori the Siegel–Brjiuno method (§7). The (easy) proof is exactly the same. And it is not a parity cancellation: it will be called a "KAM cancellation".

We see that there are still quite a few cases of trees with $\vec 0$ free momentum that can contribute to the homoclinic splitting. A few more cancellations can be spotted.[6]

We shall see that we do not have to worry about trees containing at least one dry fruit; but the *basic problem remains for the trees with only ripe fruits.* The problem is solved, for the trees with a root label

[6] for instance:

$j > l$ by the following really remarkable mechanism, which is a non trivial extension of the above mechanism II).

IV) Cancellation of the $\mathcal{R}$–tree values contributions to the splitting from $\vec 0$–free momentum trees with root carrying a label $j > l$, by summing over all the possible ways of shifting the root branch location to other free nodes of a given *rootless* tree.

This is another cancellation which is not a parity cancellation: it implies as a particular case the KAM cancellation above, but the latter is very elementary (see below), while the more general cancellation IV) is much deeper, and it is essential for the homoclinic splitting theory (for which the KAM cancellation is not sufficient). With some thought one realizes that without cancellation II) for the KAM theory, and IV) for the homoclinic splitting the above "nice" formalism would be essentially useless. The proof of IV) is based on the "tree root identity", due to L. Chierchia.

4): (Tree root identity), (Chierchia):

Consider a resummation tree ϑ without fruits and consider its R–value. Suppose that the total mode of ϑ is $\vec\nu(v_0) = \vec 0$, and let j be the root label, $\bar h$ the order of ϑ. Fix $\vec\nu_1, \ldots, \vec\nu_s$ and $j_1, \ldots, j_s$, where $s \ge 0$ and $j_i \in (0, \ldots, l-1)$. Consider all trees obtained by adding to ϑ, in all possible ways, s branches with a ripe fruit carrying the labels $h_i, \vec\nu_i, j_i$ with $i = 1, \ldots, s$, or no fruit at all ($s = 0$). Suppose that the root label is $j > l$.

We want to prove that such trees (with vanishing free mode) contribute 0 to the splitting $\Delta^h_{\vec\nu j}$, by a cancellation mechanism based on the symmetry property of the following quadratic form, setting $F_R(t) = F(t)e^{-R\sigma_t tg}$:

$$QG(t) = \sum_{\rho=\pm} \int_{\rho\infty}^{t} w(t,\tau) G(\tau)\, d\tau \tag{6.11}$$

$$(F_{R_1}, QG_{R_2}) \equiv \int_{-\infty}^{+\infty} e^{-R_1\sigma_t tg} dt \sum_{\rho=\pm} \int_{\rho\infty}^{t} w(t,\tau) F(t) e^{-R_2\sigma_\tau\tau g} G(\tau)\, d\tau = (QF_{R_1}, G_{R_2})$$

where $w(t,\tau)$ is either $t-\tau$ or $w_{00}(t)w_{0l}(\tau) - w_{00}(\tau)w_{0l}(t)$, see (6.3), (6.5), and F, G are *arbitrary* elements of $\hat{\mathcal{M}}$. The (6.11) is immediately checked, for any R_1, R_2 large enough, from the definitions and uses only the antisymmetry of $w(t,\tau)$.

Integrating (6.11) over R_1, R_2 on the appropriately small contours, to get the residues at $R_1, R_2 = 0$, see appendix A2:

$$\fint_{-\infty}^{+\infty} dt \sum_\rho \fint_{-\rho\infty}^{t} w(t,\tau) F(t) G(\tau) d\tau = \fint_{-\infty}^{+\infty} dt \sum_\rho \fint_{-\rho\infty}^{t} w(t,\tau) G(t) F(\tau) d\tau \tag{6.12}$$

This identity, and the remark that it is relevant for the cancellations, is a key property and is due to L. Chierchia (private communication).

If we consider a tree ϑ with root label $j > l$ and we take $t = \tau_{v_0}$, $\tau = \tau_{v_1}$, v_1 being any of the nodes immediately following v_0, the above identity has a simple graphical interpretation (formally discussed in appendix A2). Let $\vartheta_0 = \vartheta, \vartheta_1, \ldots, \vartheta_n$ be all the trees that can be obtained from ϑ by detaching the root branch from the node v_0 and attaching it to another node v_j with $\delta_{v_j} = 1$ (*i.e.* to a node to which the above operation gives rise to another of our trees). Then the symmetry (6.12) implies that the $\mathcal{R}$ value of any of the trees thus obtained has the form $V \cdot i(\vec\nu_{v_p})_j$ with V *independent* of ϑ_p

Hence we see the following basic cancellation mechanism: if the root free mode is $\vec 0$ then $\sum_{v,\,\delta_v=1} \nu_{vj} = 0$ and the sum of the contributions to the homoclinic splitting coming from the above family of trees vanishes.

We conclude, in particular, that the $\mathcal{R}$–tree value of the contribution to the action splitting of the resummation $\mathcal{R}$–tree values ϑ with ripe fruits only and $\vec 0$ free mode is always 0*, if the root label is* $j > l$. Hence we do not consider them in the analysis of the splitting.

5): (The energy conservation cancellation):

III) Cancellation of the contributions to $\beta^{h,1}_{j,\vec\nu}$ from $\mathcal{R}$–trees with $\vec 0$ free momentum, ripe fruits only, and with ripe fruits carrying only the label $j = 0$, *and* with all the free branches carrying a $j > 0$ label, by summing over all the possible ways of attaching the fruits to the tree. This cancellation is also remarkable but its role is not very clear; its proof, not immediate, will not be given here.

We cannot, however, conclude that also the general contribution to the splitting Δ_l^h of the resummation $\mathcal{R}$–tree values from trees with $\vec 0$ free mode is always 0.

This will not be so bad, after all, as the energy of the two whiskers is the same so that (see (1.1), (2.2)) $H_\mu(X^\sigma_\mu(t;\vec\alpha)\equiv E_\mu\equiv\sum_{h\ge0}E_h\mu^h$ where, by the KAM results reported in §2, E_μ is an analytic function of μ near $\mu=0$ and is independent of σ (and of course of t and $\vec\alpha$). Recalling that $\vec X_{\mu\downarrow}(0;\vec\alpha)\equiv\vec\alpha$, $X_{\mu-}(0;\vec\alpha)\equiv\pi$ and that $I^0(0)\equiv X^0_+(0)=2J_0g$ we find for $h\ge1$:

$$
\begin{aligned}
&\vec\omega\cdot\vec X^{h\sigma}_\uparrow(0)+2gX^{h\sigma}_+(0)+\frac12\sum_{\substack{h_1+h_2=h\\1\le h_i\le h-1}}\Big(J^{-1}\vec X^{h_1\sigma}_\uparrow(0)\cdot\vec X^{h_2\sigma}_\uparrow(0)+\\
&+\,J_0^{-1}X^{h_1\sigma}_+(0)X^{h_2\sigma}_+(0)\Big)+\delta_{h1}\sum_{\substack{|\nu|\le N\\ \vec\nu\ne\vec0}}f_\nu\cos(\vec\alpha\cdot\vec\nu+n\pi)=E_h
\end{aligned}
\tag{6.13}
$$

where $X^{h\sigma}(0)$ is short for $X^{h\sigma}(0;\vec\alpha)$ and δ_{1h} is the Kronecker symbol (and for $h=1$ the sum over h_i is absent); recall that a label l has been equivalently denoted $+$ and that the labels $j>l$ are collectively denoted $\uparrow$. Taking the difference when $\sigma=+$ and $\sigma=-$ one obtains:

$$
\begin{aligned}
\Delta^h_+(\vec\alpha)=&(-2g)^{-1}\Big(\vec\omega\cdot\vec\Delta^h_\uparrow(\vec\alpha)+\frac12\sum_{k=1}^{h-1}J^{-1}\vec\Delta^{h-k}_\uparrow(\vec\alpha)\cdot[\vec X^{k+}_\uparrow(0;\vec\alpha)+\vec X^{k-}_\uparrow(0;\vec\alpha)]+\\
&+\,J_0^{-1}\Delta^{h-k}_+(\vec\alpha)[X^{k+}_+(0;\vec\alpha)+X^{k-}_+(0;\vec\alpha)]\Big)
\end{aligned}
\tag{6.14}
$$

Hence in some sense the Δ_l^h components of the splitting may be "less important" to control. One may also think that the energy conservation allows us to "transfer" the cancellations that we have shown to exist in the expressions for the rotator components of the homoclinic splitting to the pendulum components.

B): *KAM cancellations.*

The above analysis can be applied also to the case in which the pendulum in (1.1) is decoupled from the rotators, to provide a graphical representation of the invariant torus which continues the unperturbed torus $\vec A=\vec0$. The simplest representation of the torus is:

$$
\vec\alpha=\vec\alpha\in T^{l-1},\qquad \vec A=\vec X_\uparrow(0;\vec\alpha)
\tag{6.15}
$$

which is obtained by simply considering the trees ϑ with only "rotators labels", *i.e.* with node and branch, angle or action, labels j taking values $j=1,\ldots,l-1$ and $j=l+1,\ldots,2l-1$ (hence excluding the uninteresting $0,l$). The root time label is $t=0^+$ (or $t=0^-$) as we are computing $X_\uparrow$ at time 0).

The root time t is again taken to be $t=0$, and:

$$
\vec X_{\uparrow\vec\nu}(\vec0;\vec\alpha)=\int_{+\infty}^0F^h_{\uparrow\vec\nu}(\tau)\,d\tau,\qquad \vec X_{\downarrow\vec\nu}(0;\vec\alpha)\equiv\vec0
\tag{6.16}
$$

where $F^h_{\vec\nu}$ is the same function in (6.6), resulting from the resummations (hence, by the remark following (6.6) it is the $\vec\nu$–Fourier component, in the $\vec\alpha$–angles, contribution to (4.14)). This time we only integrate up to 0 because we are not looking at the splitting (which, see below, vanishes) but at the value of $\vec X_\uparrow$ at $t=0$.

However the (6.16) does not seem to be the most convenient representation, not only from a technical viewpoint but also because it does not lead to a simple representation of the motion on the torus.

A better representation is obtained by remarking that $F^h(\tau)$ corresponds to a solution a solution $t\to X(t)$, on the torus and starting at $\vec\alpha$, and has the form $F^h(t;\vec\alpha)\in\mathcal{M}$, see (3.4). We shall denote $\Phi^h_i(\vec\psi,t;\vec\alpha)$ the functions in the representation (3.1) of $F^h(t;\vec\alpha)=\sum_{i\ge0}\frac{t^i}{i!}\Phi_i(\vec\omega t,t;\vec\alpha)$. The notation is slightly different from that of §2,§3,§4 where argument t in $\Phi_i(\vec\psi,t;\vec\alpha)$ appears in the form $x=e^{-\sigma gt}$; (this is irrelevant and we shall see that $\Phi_i(\vec\psi,t;\vec\alpha)$ is in fact t–independent).

The absence of coupling between rotators and pendulum produces the following great simplifications, with respect to the case of the homiclinic splitting.

0) The $+\infty$ in (6.16) can be replaced with $-\infty$; *i.e.* the splitting vanishes.

1) The functions Φ_i have rotator components vanishing if $i>0$, so that the rotator components of $F(t;\vec\alpha)$ can be simply written $\Phi_0(\vec\omega t,t;\vec\alpha)$ (*i.e.* no secular terms $\frac{t^i}{i!}$ are really present);

2) The function $\Phi_0(\vec\psi,t;\vec\alpha)$ *is t–independent and* $\vec\Phi_{0\uparrow}(\vec\psi,t;\vec\alpha)$ *has* $\vec 0$ *average over* $\vec\psi$.

The above properties are an immediate inductive consequence of the (6.15), and of (4.9),(4.10), see below, (or, *alternatively*, of the KAM theorem).

One sees, in fact, that the splitting is zero, to all orders of perturbation theory, simply by remarking that from (4.14),(4.13) we see that there are no traces, in the present case, of the pendulum separatrix motion. Therefore F^1 is a analytic quasi periodic function with $\vec 0$ average. Hence X^1 is also quasi periodic and analytic because the operators $\mathcal I$ and $J^{-1}\overline{\mathcal I}^2$ do not generate any non quasi periodic nor non analytic, nor secular terms *when applied to* $\vec 0$ *average quasi periodic functions of* t. Therefore $F^2_\uparrow$, see (4.14), is still analytic, and also quasi periodic in t with $\vec 0$ average over t, by the cancellation (4.15), *etc.* Therefore we never generate non analytic non quasi periodic terms. In particular $X^h(0)$ can be equivalently given by the (6.16) with $-\infty$ replacing ∞: hence the splitting at 0 vanishes (being $X^h(0^+)-X^h(0^-)$. This simply means that if there is no interaction between the pendulum and the rotators, the whiskers remain degenerate to all orders of perturbation theory.

The above implies that we can write:

$$\vec A(t)=\mathcal I\vec\Phi_\uparrow(\vec\omega t;\vec\alpha),\qquad \vec\alpha(t)=\vec\alpha+\vec\omega t+J^{-1}\overline{\mathcal I}^2\vec\Phi_\uparrow(\vec\omega t;\vec\alpha)\tag{6.17}$$

where $\Phi(\vec\psi;\vec\alpha)$ is the function $F(\vec\psi;\vec\alpha)$ regarded at fixed $\vec\alpha$ and as a function of $\vec\psi$.

Remarks:

1): We *have not* invoked here the KAM theorem in describing the above algorithm to build recursively the $\Phi^h(\omega t;\vec\alpha)$, *i.e.* to construct a formal power series describing via (6.17) the motion on the invariant torus (which, unless we use the KAM theorem we still have to prove to exist by proving the convergence of the series). The (6.17), which coincides with (5.15), is here a well defined formal power series with the coefficients defined by the recursive algorithm discussed in connection with (5.15) and recalled above.

2): in the present case, in the notations of §4, see comments to (4.1), it is $F^{h\sigma}(\vec\psi,\sigma\infty)\equiv\Phi^h(\vec\psi;\vec\alpha)$ so that (4.9) implies $\Phi^h_{\vec 0}=\vec 0$.

3): Since the $\vec\omega$ is a diophantine vector the above points $(\vec A(t),\vec\alpha(t))$ cover densely the torus as t varies; hence calling $\vec\omega t=\vec\psi$ and setting $\vec\alpha=\vec 0$ and $\Phi(\vec\psi)\equiv\Phi(\vec\psi;\vec 0)$, we realize that a natural parametrization of the torus is:

$$\begin{aligned}&\vec A=\sum_{k=1}^\infty\mu^k\,\vec a^k(\vec\psi),\qquad \vec\alpha=\vec\psi+\sum_{k=1}^\infty\mu^k\vec b^k(\vec\psi),\qquad \vec\psi\in T^{l-1}\\ &\vec a^h(\vec\psi)=\sum_{\vec\nu\ne\vec 0}\vec\Phi^h_{\uparrow\vec\nu}\frac{e^{i\vec\nu\cdot\vec\psi}}{i\vec\omega\cdot\vec\nu},\qquad \vec b^h=\vec\psi+J^{-1}\sum_{\vec\nu\ne\vec 0}\vec\Phi^h_{\uparrow\vec\nu}\frac{e^{i\vec\nu\cdot\vec\psi}-1}{(i\vec\omega\cdot\vec\nu)^2}\end{aligned}\tag{6.18}$$

and in such representation $\vec\psi$ has the interpretation of "average anomaly", *i.e.* the evolution is simply $\vec\psi\to\vec\psi+\vec\omega t$. By (4.9) $\vec\Phi^h_{\vec 0}=\vec 0$, and the $\vec\nu=\vec 0$ terms are therefore absent (care has to be exercised here in not confusing the Fourier transorm with respect to $\vec\psi$ and that with respect to $\vec\alpha$: we are not talking about the latter as $\vec\alpha$ is set $\equiv\vec 0$). Since we set $\vec\alpha=\vec 0$ we could use the parity properties of $\vec\Phi^h$ which, if $\vec\alpha$ were not $\vec 0$, would be jointly odd in $\vec\alpha,\vec\psi$, reflecting that $F^h_\uparrow$ is time angle odd and implying that if $\vec\alpha=\vec 0$ then it is odd in $\vec\psi$ and hence $\vec\Phi_{\vec 0}=\vec 0$. But $\vec\Phi^h_{\vec 0}$ is equal to $\vec 0$ also if $\vec\alpha\ne\vec 0$, by the cancellation remarked in (4.2) (see (4.9); see also remark 2) above), which is not a parity cancellation.

4): To calculate $\vec F^h_\uparrow$ at $\vec\alpha=\vec 0$, hence to calculate $\vec\Phi$ at $\vec\alpha=\vec 0$, we can use the tree expansion. The $\vec\Phi^h_{j\vec\nu}$ can be identified to be given by the trees of order h with time label τ_{v_0} of the first node fixed equal to the root time label t, and with $j_{v_0}=j$ and total free momentum $\vec\nu$. We use the remark 2) after (5.14) to establish the connection between F^h and the tree expansion.

In fact, if we remark that:

$$\sum_\rho\int_{\rho\infty}^\tau e^{i\vec\omega\cdot\vec\nu\tau'}\,d\tau'\equiv\frac{e^{i\vec\omega\cdot\vec\nu\tau}}{i\vec\omega\cdot\vec\nu}\tag{6.19}$$

we see that a resummation tree with free momentum $\vec\nu$ has a t_{v_0}–dependence proportional to $e^{i\vec\omega\cdot\vec\nu\,\tau_{v_0}}$. Since the trees are evaluated stopping short of computing the "last integral" over the τ_{v_0} which is, instead, set equal to t we see that the tree value for the evaluation of $e^{i\vec\nu\cdot\vec\psi}\,\Phi^h_{\vec\nu}$ is then obtained by replacing such exponentials in t by $e^{i\vec\nu\cdot\vec\psi}$: summing over the trees yields $e^{i\vec\nu\cdot\vec\psi}\,\vec\Phi^h_{\vec\nu}$. In the above calculations $\vec\alpha$ is set $\equiv\vec 0$.

5): since the splitting is identically zero to all orders there will be no resummation trees with dry fruits.

6): thus we se that we can use the formalism developed in the case of the theory of the splitting, to study X^h for $\vec\alpha$ arbitrary.

But to take advantage of the easy evaluation of the integrals and of the fact that we only want $\vec\alpha=\vec 0$ in the present case, it is useful to quickly go through an essentially independent and more detailed analysis. We can start by remarking that the integrals associated with the tree vertices are straightforward because:

$$\int_{(\sigma)}^t (t-\tau)^\zeta\, e^{i\vec\omega\cdot\vec\nu\tau}\,d\tau-\int_{(\sigma)}^0 (t-\tau)^\zeta\, e^{i\vec\omega\cdot\vec\nu\tau}\,d\tau=\frac{e^{i\vec\omega\cdot\vec\nu t}}{(i\vec\omega\cdot\vec\nu)^{1+\zeta}}-\zeta\,\frac{1}{(i\vec\omega\cdot\vec\nu)^{1+\zeta}},\qquad \zeta=0,1 \tag{6.20}$$

which takes the place of (6.3) and (6.5).

We can mimick, therefore, the procedure followed in the resummation scheme developed for the homoclinic splitting cancellations.

Given a tree ϑ contributing to order h we represent the $\mathcal{E}^T$ operations associated with a node v by affixing a "bubble" around the node v and enclosing all the branches following it, or by writing a label $\mathcal{R}$ on the node v. The bubble will mean that after evaluating the $J^{-1}\overline{\mathcal{I}}^2$ or $\mathcal{I}$ operation via (6.20) we chose the second term in the r.h.s. of (6.20) (with $\zeta=1$). If the node is not enclosed in a bubble that cuts the branch linking it to the previous node, then this means that it is marked by $\mathcal{R}$ and that in the sum in (6.20) we selected instead the first term (like in the previous case).

We can repeat the same representation operation on the (improper) integrations pertaining to all the other nodes, starting from the highest on the tree and going down towards the root: we stop at the first nodes following v_0, (the latter node is somewhat different from the others as the corresponding $\mathcal{I}$ operation will not be performed, as we evaluate $\vec F^h$ and not X^h).

Thus the most general contribution by trees of order h to $\vec\Phi^h_\uparrow$ will be represented by a tree with h nodes with index $\delta=1$ (and h branches, because now there are no nodes of order $\delta=0$) and all the labels introduced so far ($\mathcal{R}$ labels included) plus an arbitrary number of bubbles drawn around tree nodes above v_0, and drawn as described (avoiding overlappings).

We call *free* the branches that are not enclosed inside bubbles of any type. The total *free momentum* of the tree will be the sum of the node modes associated with the free nodes. By construction the t dependence of the tree value appears *only* through $e^{i\vec\omega\cdot\vec\nu t}$, if $\vec\nu$ is the total free momentum of the tree. This is special for the present case, and it does not hold for the homoclinic case.

Hence the $\vec\nu$–th Fourier coefficients, in $\vec\psi$, of $\vec\Phi^h_\uparrow(\vec\psi)$ are simply equal to the sum of the tree values of the trees with order k and total free momentum $\vec\nu$, evaluated at $t=0$. They will be denoted, as usual, $\vec\Phi^h_{\uparrow\vec\nu}$. This also provides an interesting interpretation of the free momenta. Note that in this way the concept of total momentum (as opposed to that of total free momentum) does not even arise.

Looking at (6.20) we see that the second term is a constant factor which can be taken out of the integration operations $J^{-1}\overline{\mathcal{I}}^2$ associated with the other tree nodes. Thus the trees with bubbles involve somewhat simpler integrals (although not dramatically simpler as in the case of the homoclinic splitting, because in this case all the integrals are essentially trivial).

Consider a bubble containing a subtree of order h linked to its root node $v'\ge v_0$ by a branch $v'v$ carrying a total free momentum $\vec\nu\equiv\vec\nu_f(v)=\sum^*_{w\ge v}\vec\nu_w$, where the $*$ means that the sum is over the nodes which are inside the bubble but not inside inner bubbles, *i.e.* which are "relatively free"), and carrying a label j ($1\le j\le l-1$), and a label h equal to the order. Fixed $\vec\nu,j,h$ we can sum over the possible choices (consistent with the labels $\vec\nu,j,h$) of the subtree and its decorations obtaining the "resummed bubble values":

$$\beta^h_{\vec\nu j}=-\frac{1}{(i\vec\omega\cdot\vec\nu)^2}\Phi^h_{\vec\nu j},\qquad \vec\nu\ne\vec 0 \tag{6.21}$$

The cancellation remarked after (4.14), see also (6.18), implies that $\beta^h_{\vec 0j}\equiv 0$ and (6.21) is interpreted as $\vec 0$ for $\vec\nu=\vec 0$. In fact it could be easily checked by parity considerations the more general relation:

$$\beta^h_{\vec\nu j}\equiv-\beta^h_{-\vec\nu j},\qquad j=1,\ldots,l-1 \tag{6.22}$$

Therefore the above resummation suggests introducing again "resummation trees". A resummation tree, with all its decorations (bubbles included), is defined by drawing a tree with all its decorations and by *deleting* the contents of the bubbles, leaving only the branches connecting the top free nodes to the nodes v inside the outermost bubbles if any: we call $\tilde V_f$ the set of such "non free" nodes. We shall leave the bubble around each v, and on the branch leading to v we write the label j_{λ_v}, together with the total free momentum and the order h_{λ_v} of the deleted subtree.

By construction the resummation trees have bubbles which can contain only one node. It is natural to call such resummation bubbles *"fruits"*, as in the previous case. And the nodes $v\in\tilde V_f$ we again shall rightly call *seeds*, see remark following (6.10). Note that also from the new viewpoint there is only one type of fruits.

Setting $f_{\vec\nu,0}\equiv f_{\vec\nu}$ and calling $\vec\nu_f(v)=\sum_{v\le w<\tilde V_f}\vec\nu_w$ the free momentum of the node v, the "value" of a resummation tree contribution to $\Phi^h_{\uparrow\vec\nu}$ will be simply:

$$\Big(-\tfrac12 (i\nu_{v_0})_{j_{v_0}} f_{\vec\nu_{v_0}}\prod_{j=1}^{l-1}(i\nu_{v_0})_j)^{m^j_{v_0}}\Big)\cdot\prod_{v_0<v\notin\tilde V_f}\frac{-\frac12(i\nu_v)_{j_v}f_{\nu_v}\prod_{j=1}^{l-1}(i\nu_v)_j^{m^j_v}}{J\,(i\vec\omega\cdot\vec\nu_f(v))^2}\prod_{v\in\tilde V_f}\frac{\beta^{h_v}_{\vec\nu(v)j_v}}{J\,(i\vec\omega\cdot\vec\nu f(v))^2}\tag{6.23}$$

by (5.12), as in this case $n_v\equiv0$, $\delta_v\equiv1$. The value of $\vec F^h_{\uparrow\vec\nu}$ is obtained by summing over all trees ϑ with order h and total free momentum $\vec\nu$: naturally the order h is the number of free nodes plus the orders of the fruits, but the total momentum is the sum of the node modes *excluding* the modes of the fruits.

Finally we *renumber* such trees (as described at the beginning of the section) according to their topological structure (*i.e.* resummation trees do not inherit the numbering from the trees that generated them).

This is a good definition as we can now say that the sum of the values of the decorated resummation trees of given order is equal to the sum of the values of the previously considered trees of the same order. As in the homoclinic splitting case the correctness of the combinatorics is not difficult to verify if the trees are regarded (as we are doing) as having all the branches distinct.

Remarks:

i): The collection of all the trees of a given order and total free momentum $\vec 0$ must give a zero contribution to $\vec\Phi^h$ because $\vec F^h_\uparrow(\vec\psi\,;\vec 0)$ has zero average by (4.9), or (6.22). Hence $\vec\nu_f(v)\ne\vec0$ can, *and will*, be supposed without fear of errors and without ever having to divide by zero in the trees evaluations (recall the diophantine property (1.3), showing that zeros can only occur if $\vec\nu_f(v)=\vec 0$): the terms in (6.23) with some of the denominators equal to 0 have to be regarded as 0.

ii) Another essential cancellation is the following. Given a tree ϑ suppose it to be such that $\vec\nu_f(v)=\vec\nu_f(v')$, $(\ne\vec0)$, with v',v being a pair of consecutive nodes in ϑ, for some $v>v_0$. Then the v' must be a node with a bifurcation with at least two branches (as, of course, $\vec\nu_v\ne\vec0$ for all v). We can then consider the subtree $\vartheta_2\subset\vartheta$, with the same root as ϑ, obtained by deleting the branch $v'v$ and the following ones. Since $\vec\nu_f(v')=\vec\nu_f(v)$ by assumption, such tree will be such that $\sum_{w\ge v'}\vec\nu_w=\vec0$.
with root v': imagine to attach it to the remaining tree by pinning it to the nodes $w>v'$. From (6.23) it is clear that in so doing we get terms equal to some w–independent constant times $i\nu_{wj'}$ (if j' is the branch label of $v'v$) provided we could neglect that, in so doing, some of the denominators of a few of the branches above v' change value, The change in value is simply the addition or subtraction of $\varepsilon=\vec\omega\dot{\vec\nu}(v)$. Hence this can be hoped to be a good approximation if ε is very small. In fact summing also over a simoultanous change in sign of all the modes $\vec\nu_w$ to which the pinning can be done it is clear that the sum of the values of the considered family of trees vanishes *to second order in* ε.

iii): The same argument holds if v and v' are comparable in the partial order established by the tree, *i.e.* if $v'<v$ but v' is not necessarily the immediate predecessor of v.

Therefore we can approximately compute $\vec X^h_\uparrow(\vec0;\vec\alpha)$ by summing only trees such that:

I) $\vec\nu_f(v)\ne\vec0$ *if v is any node (seeds included).*

II) $\vec\nu_f(v)\ne\vec\nu_f(v')$ *for all pairs of comparable nodes v',v, (not necessarily next to each other in the tree order, however), with $v'\ge v_0$.*

The consequences of the above proved cancellations are analyzed in the next sections (and in the relative appendices).

§7 Twistless tori: Siegel–Brjiuno–Eliasson method for KAM tori.

This section has heuristic nature; we shall suppose that $\vec{X}_\uparrow^{(h)}(\vec{0};\vec{\alpha})$ can be computed by summing over the trees verifying I),II) at the end of §6. The corrextions, *i.e.* the contribution from the other trees are accounted for in the Appendix A3, because the natural continuation of the heuristic argument is §8, while the corrections to the approximation II) require some new ideas (constituting Eliasson's main contribution). It will be clear that the approximation II) is very good, particularly if one applies it to momenta $\vec{\nu}(v)$ which are very large.

To simplify further the proof we shall make a further assumption on $\vec{\omega}_0$ besides (1.3); namely we shall suppose that:

$$\min_{0\ge p\ge n} \big|C_0|\vec{\omega}_0\cdot\vec{\nu}| - \gamma^p\big| > \gamma^{n+1} \qquad \text{if } n\le 0,\ 0<|\vec{\nu}|\le (\gamma^{n+c})^{-\tau^{-1}} \tag{7.1}$$

and it is easy to see that the *strongly diophantine vectors*, as we shall call the $\vec{\omega}_0$ verifying (1.3) and (7.1), have full measure in R^l if $\gamma>1$ and c are fixed and if τ is fixed $\tau>l-1$: we take $\gamma=2, c=3$ for simplicity; note that (7.1) is empty if $n>-3$ or $p<n+3$.

We proceed to prove theorem 1, §1: *i.e.* the persistence of the torus which for $\mu=0$ is $\vec{A}=\vec{0}$. We recall the hypothesis H: hence, in this section, the parameter $\eta\equiv 1$ and therefore the system time scales are set by $|\vec{\omega}|, C_0^{-1}$ and by the perturbation; and J is an arbitrary diagonal matrix with matrix elements $J_j\ge J_0$ where $J_0>0$ is a positive constant.

The basic idea of the following proof goes back to Siegel, [S]: his somewhat difficult proof has been greatly clarified by Brjuno and a very clear exposition and generalization of Brjiuno's work by Pöschel is in [P]. But the connection with the KAM theorem and the tree cancellations, exposed in §6, are due to Eliasson, [E].

Let $\mathcal{N}$ be the set of the harmonics $\vec{\nu}$ for which $f_{\vec{\nu}}\equiv f_{0,\vec{\nu}}$ in (1.1) does not vanish: the number of such "perturbation harmonics" does not exceed $(2N+1)^{l-1}$ and the length $|\vec{\nu}|=\sum|\nu_j|$ is $\le N$. But the long sought connection with the KAM theorem is due entirely to Eliasson, [E], who also realized the role of the tree expansions in establishing the connection just mentioned.

Consider the functions $\vec{\Phi}^h_{\uparrow\vec{\nu}}$ defined in §6, B), see (6.18). We shall prove that:

$$b^h_{j\vec{\nu}} \equiv \frac{1}{J_0|\vec{\omega}\cdot\vec{\nu}|^2}|\Phi^h_{j\vec{\nu}}| \le DB^{h-1}, \qquad j=1,\ldots,l-1 \tag{7.2}$$

and D,B will be determined below, and depend only on $J_0,N,l,C_0,\tau,|\vec{\omega}|$. *Hence they do not depend on a lower bound on* $\det J^{-1}$. Furthermore $\vec{F}^h_{\uparrow\vec{\nu}}\equiv\vec{0}$ if $|\vec{\nu}|>Nh$, because the initial perturbation has only harmonics with $|\vec{\nu}|\le N$.

The discussion of §6, B), shows that (7.2) implies theorem 1, if valid under the just stated conditions, and the function $\vec{x}$ can be taken $\vec{x}_{\vec{\nu}}=\frac{i}{E}\vec{\Phi}_{\vec{\nu}}$.

Consider a tree Θ with labels and fruits and regard it as representing a contribution to the l.h.s. in (7.2). We imagine to express each fruit value as a sum over the trees that can be grown on its seeds, and so on. We can think that "each seed is magnified to show its content in subtrees", see remark following (6.10). In other words we undo, for the purposes of the estimates, the collection of contributions to $\Phi^h_{\vec{\nu}j}$ which gave rise to the resummations and to the notion of fruits.

But we take into account the assumed cancellations by constraining, from now on, the labeled trees appearing inside each of the bubbles (representing a magnified seed) to verify the properties I) (which is exact), and II) (which is approximated) at the end of §6.

Therefore the contribution from the considered Θ, and as a consequence the full value of $\Phi^h_{\vec{\nu}j}$, is split as a sum of contributions coming from trees ϑ carrying an arbitrary number of bubbles, hierarchically ordered to avoid overlappings and each of which encloses a node of ϑ as well as all the subsequent ones. The number of nodes is of course h.

Having undone all our painful collection of the trees contributing to $\Phi^h_{\vec{\nu}j}$ it is convenient to regard again all the branches of the trees as distinct and therefore the correct combinatorial factor with which to multiply the tree contribution to the $\Phi^h_{\vec{\nu}j}$ is again $h!^{-1}$, see (5.11).

Fixed h and an unlabeled tree ϑ the number of ways one can put bubbles around the nodes is bounded by 2^h (as there can be at most one bubble per node by our decoration rules); the number of node modes $\vec{\nu}_v$ is bounded by $(2N+1)^{(l-1)h}$; the number of branch labels is $(l-1)^h$ and the number of trees is, by Cayley's count, see §5, footnote [5]. Therefore a bound on $b^h_{\vec{\nu}j}$ is simply:

$$2^h(2N+1)^{(l-1)h}(l-1)^h 2^{2h}\, M_h \tag{7.3}$$

where M_h is just the maximum possible value that a single fully decorated tree can contribute to $b^h_{\vec{\nu}j}$.
Suppose that in one of the possible trees ϑ there are n bubbles, counting among them also an extra bubble (drawn for convenience) and enclosing the whole tree except the root, and suppose that the i-th of them encloses h^*_i branches. Then $h = n + \sum_{i=1}^n h^*_i$: because, (having added one bubble encircling the whole tree but the root), for every bubble there is one branch not entirely contained inside a bubble.
Suppose that we can prove that for a fruitless tree the contribution to $b^k_{\vec{\nu}j}$ is bounded by $D_0 B_0^{k-1}$, for all $k \ge 1$. Then it clear from (6.23) that one can take:

$$M_h = \max \prod_{i=1}^n D_0 B_0^{h^*_i} \tag{7.4}$$

where the maximum is over n, h^*_i such that $h = n + \sum_{i=1}^n h^*_i$. Supposing that $B_0 \ge D_0$ it follows immediately that (7.2) holds and that one can take (see (7.3)):

$$D = (2^3(2N+1)^{l-1}(l-1)) \cdot D_0, \qquad B = B_0 \tag{7.5}$$

Therefore the problem is reduced to the analysis of a fruitless tree. Hence we only must study:

$$\delta_{h,\vec{\nu}} = \max_{\vartheta}{}^* \prod_{v \ge v_0} \frac{1}{|C_0\vec{\omega} \cdot \vec{\nu}_f(v)|^2} \tag{7.6}$$

where the maximum is considered over all fruitless trees with total free mode $\vec{\nu}$ and the $*$ reminds us that only trees with the considered constraints are admmissible to compete for the maximum. The constant C_0 is introduced to make $\delta_{h,\vec{\nu}}$ adimensional.
Proceeding as in [P], we fix, given $\vec{\nu}, h$, one tree ϑ on which the maximum is attained and we consider its mode labels and the subtrees ϑ_v with first node at v and root at v', for any $v > v_0$. It is necessarily true that the subtree ϑ_v with the given mode labels is also a maximizer for $\delta_{h_v, \vec{\nu}(v)}$.
A first remark, that should be at this point really obvious, is that we can suppose that $0 < |\vec{\nu}_f(v)| \le h_v N$ if h_v is the order of the subtree with root at v' and first node at v. And therefore we can always suppose:

$$|C_0\vec{\omega} \cdot \vec{\nu}_f(v)| \ge \frac{1}{N^\tau h_v^\tau} \tag{7.7}$$

by the diophantine assumption on $\vec{\omega}$.
Given, arbitrarily, $q \ge 1$ integer we say that a harmonic $\vec{\nu} \ne \vec{0}$ *is q–resonant* if:

$$C_0|\vec{\omega} \cdot \vec{\nu}| < \frac{1}{3}\frac{1}{N^\tau q^\tau} \equiv \zeta(q) \tag{7.8}$$

where $\zeta(q)$ is defined here.
Then we define $N(\vec{\nu}, h; q)$ to be the number of q–resonant harmonics among the harmonics found in the tree where the maximum in (7.6) is reached.
The following (extension) of Brjuno's lemma (see [P]) holds:

$$N(\vec{\nu}, h; q) \le 2\frac{k}{q} \tag{7.9}$$

for the trees verifying I),II) at the end of §6. We do not repeat the proof as the present analysis has only heuristic value.
Assuming (7.7) we can easily conclude the discussion; in fact choosing $q = 2^p$:

$$\prod_{v \ge v_0} \frac{1}{(C_0|\vec{\omega} \cdot \vec{\nu}_f(v)|)^2} \le \prod_{p=0}^{\infty} (3N^\tau 2^{\tau p})^{4h2^{-p}} \le B_4^h \tag{7.10}$$

with $B_4 = \exp(\sum_{p=0}^{\infty} \frac{4}{2^p} \log(3N^\tau 2^{\tau p})$.
Therefore we can bound (7.6) by B_4^h. Recalling also that in the notations of §1 $J\vec{\omega}^2$ is also called E and it is supposed to be an upper bound on $|f_\nu|$, see comment after (1.1) and (1.4), and if $B_5 \equiv \left(\frac{J_0\vec{\omega}^2}{JC_0^{-2}}\right)$, this means that D_0, B_0 in (7.4) holds with:

$$D_0 = (C_0^2\vec{\omega}^2)(B_4 B_5), \qquad B_0 = B_4 B_5 \tag{7.11}$$

And therefore (7.2) holds with:

$$B_0 = b_l \frac{J_0}{J_{min}} \vec{\omega}^2 C_0^2 N^{8\tau+l-1} 2^{8\tau}, \qquad D_0 = b_l' B_0 \tag{7.12}$$

where b_l, b_l' are suitable constants depending only on l and b_l' can be supposed without loss of generality ≥ 1. Hence we see that B in (7.2) does not depend on $J_{\max}$ *but* on $J_{\min}$.

We see that if $J_{\min} = +\infty$ then $B = 0$: and the system is in fact completely integrable as already remarked in §4, around (4.16), (see also [G1], probles 1,16,17 of §5.11).

Clearly the bound (7.2) implies the convergence of our approximation to the series for the invariant tori for $|\mu| < B^{-1}$: this gives an idea of the connection between Siegel's theorem and the proof of the KAM theorem. This was the situation before Eliasson's work. To complete the proof we must turn to the corrections present because the cancellation II) at the end of §6 is only approximate. A complete analysis is in Appendix A3 and A4.

§8 Theory of the homoclinic splitting.

As a consequence of the above analysis we get that, in general, the angles of homoclinic splitting (or δ, see (1.8)) are smaller than any power in η. Let us denote $\Delta_{\vec{\nu}}^h$ the coefficient of order h in the Taylor expansion in powers of μ and of order $\vec{\nu}$ in the Fourier expansion in $\vec{\alpha}$ of the splitting $(\mu, \vec{\alpha}) \to \Delta(\vec{\alpha}) \equiv X_\mu^+(0;\vec{\alpha}) - X_\mu^-(0;\vec{\alpha})$; then the property of smallness is an immediate consequence of the following bounds.

Let $d \in (0, \frac{\pi}{2})$, and let:

$$\varepsilon_h \equiv \varepsilon_h(d) \equiv \sup_{0<|\vec{\nu}_0|\le Nh} e^{-|\vec{\omega}\cdot\vec{\nu}_0| g^{-1}(\frac{\pi}{2}-d)}, \qquad \beta = 4(N_0+1), \qquad p = 4\tau \tag{8.1}$$

where N_0 is the maximal φ–harmonic of the perturbation f in (1.1). Then, for $j = 0, \ldots, l-1$ and for all $J \in [J_0, +\infty)$ and $h \ge 1$:

$$|\Delta_{\vec{\nu},j}^h| \le J_0 g\, D\, (b\eta^Q)^{-h}, \qquad |\Delta_{\vec{\nu},j}^h| \le J_0 g\, D\, d^{-\beta} (Bd^{-\beta})^{h-1} (h-1)!^p \varepsilon_h \tag{8.2}$$

where $b < 1, D, Q, B$ are suitable dimensionless constants depending on the various parameters describing (1.1), *but not on the perturbation parameters* η, μ; the proof in the appendix A1 shows how to construct a bound on B; the latter is the only real problem as the existence of the b, Q, D is well known, (see §1 and [CG], for instance, §5, eq. (5.76), (5.78)). The constant Q has been briefly discussed in §1 and we take it to be the minimal such that the properties of peristence of the tori hold (*e.g.* $Q = 10$ if $l > 2$ and $Q = 1$ if $l = 2$, see §1).

If $l = 2$ both bounds in (8.2) are uniform in $J \ge J_0$ and one can take $J \to +\infty$; furthermore τ can be taken $\tau = 0$. If $l > 2$ the second bound in (8.2) is still uniform in $J \ge J_0$, but the first may fail to be uniform in the values $J_j \ge J_0$. Perturbation theory in μ is well defined to all orders, but it might be non uniformly (as the $J_j \to \infty$, i.e. as the twist rate goes to 0*) convergent even at small μ;* see, however, the conjecture at the end of appendix A1.

The bound (8.2) is of great interest in the case $l = 2$ as it is sufficient to determine the exact asymptotic behaviour as $\eta \to 0$ of the splitting, (hence in the periodically forced pendulum, by taking $J = +\infty$, a permissible choice). *Note that, if $l = 2$, it is $\varepsilon_h \equiv \varepsilon_1$.*

We proceed to explain the strategy of our bounds. The first inequality in (8.2) is just (2.6). Note also that since we always suppose that f is a trigonometric polynomial of degree N, it is actually $\Delta_{\vec{\nu},j}^h = 0$ if $|\vec{\nu}| > Nh$.

To explain the appearance of the "small" factor ε_h in (8.2), let us consider the most general resummation tree, ϑ, of order h with all its labels (see item 3) in part A) of §6), representing one contribution to the homoclinic splitting $\Delta_{\vec{\nu}j}^h$.

Let r be the root of ϑ; let v_0 be the first node after the root; let V_f be the set of top free nodes, *i.e.* the set of top nodes still outside the fruits (if any) of ϑ; and recall that v_0 cannot be surrounded by a fruit. Let $\tilde{V}_f$ be the set of top nodes contained inside a fruit ($\tilde{V}_f$ might be empty): *i.e.* $\tilde{V}_f$ is the set of fruit seeds.

Recall that in general h_{λ_v} denote the order of the subtree following v: the fruits contribute to the order a quantity given by the order label h_{λ_v} leading to them. We shall call $v_1,\ldots,v_{m_v}$ the m_v nodes following v, $v_0\le v\le V_f$, and v' will be the node preceding v; $m_v=0$ if $v\in V_f$.

With the above notations it is easy to "write" the sum over σ of the values of the contributions to $\beta^{h,\gamma}_{\vec\nu,j}$ from the resummation trees ϑ of order h and total mode $\vec\nu$ and root label $j\equiv j_{v_0}\ge l$:

$$\begin{aligned}\beta^h_{\vec\nu,j}(\vartheta)=&\frac{1}{m!}\oint\frac{dR_{v_0}}{2\pi iR_{v_0}}\,⨍_{-\infty}^{\infty}e^{-\sigma_{v_0}gR_{v_0}\tau_{v_0}}w^{\gamma_{v_0}}_{j_{v_0}}(\tau_{v_0})\,d\,g\tau_{v_0}\cdot\\&\cdot\prod_{v_0<v\le V_f}\oint\frac{dR_v}{2\pi iR_v}\Big[\frac12⨍_{+\infty}^{\tau_{v'}}d\,g\tau_v+\frac12⨍_{-\infty}^{\tau_{v'}}d\,g\tau_v\Big]e^{-\sigma_v gR_v\tau_v}\cdot\\&\cdot\prod_{v_0\le v\le V_f}\Big(\big(\frac{-i(\nu_v)_{\tilde j_v}}{2}\big)\;c_{\nu_v}\;\;e^{i(n_v\varphi^0(\tau_v)+\vec\nu_v\cdot\vec\omega\tau_v)}\prod_{s=0}^{l-1}(i\nu_{vs})^{m_v^s}\Big)\cdot\\&\cdot\prod_{v_0<v\le V_f}\Big(w_{j_v}(\tau_{v'},\tau_v)\Big)\cdot\Big(\prod_{v\in\tilde V_f}x^{\gamma_v}_{j_{\lambda_v}}(\tau_{v'})\beta^{h_v,\gamma_v}_{\vec\nu_{\lambda_v}j_{\lambda_v}}\Big)\end{aligned}\tag{8.3}$$

where (see (5.12), (3.4), (5.13), (5.14), (6.6), (6.9), (6.10)), and the dimensionless coefficients c_{ν_v} are given by: $c_{\nu_v}\equiv(Jg^2)^{-1}f_{\nu_v}\delta_v\delta_{j\ne0,l}+(J_0g^2)^{-1}\;f_\nu\delta_v\delta_{j=0,l}-(1-\delta_v)$ where $\delta_{j=0,l}$ is 1 if $j=0,l$ and 0 otherwise and a similar, complementary, meaning is given to $\delta_{j\ne0,l}$ (recall also that if $\delta_v=0$ then $\vec\nu_v=\vec0$) and:

$$\begin{aligned}&w_{j_v}(t,\tau)\equiv\begin{cases}w_{00}(t)\bar w_{0l}(\tau)-\bar w_{0l}(t)w_{00}(\tau), & v>v_0\;,j_v=l\\ g(t-\tau), & v>v_0\;,j_v>l\end{cases}\\&x^\gamma_j(\tau)\equiv\begin{cases}\sigma\bar w_{0l}(\tau), & \text{if } j=l\text{ and }\gamma=1\\ \sigma g\tau, & \text{if } j>l\text{ and }\gamma=1\\ w_{00}(\tau), & \text{if } j=l\text{ and }\gamma=2\\ 1, & \text{if } j>l\text{ and }\gamma=2\end{cases}\qquad w^\gamma_j=\begin{cases}x^1_j & \text{if }\gamma=2\\ x^2_j & \text{if }\gamma=1\end{cases}\end{aligned}\tag{8.4}$$

where the dimensionless matrix element $\bar w_{0l}$ was defined after (6.6); $\gamma=\gamma_v$ is 1 or 2, respectively, if the fruit encircling v is dry or ripe; m is the total number of branches (root included); the integers m_v^s decompose m_v and count the number of branches emerging from v and carrying the labels $s=0,\ldots,l-1$; $\lambda_v,\vec\nu_{\lambda_v},j_{\lambda_v}$ in the product over $\tilde V_f$ are, respectively, the branch entering the fruit around v, the momentum carried by λ_v and its j-label.

Remark: the τ_v integrals in (8.3) are, in general, not convergent for R_v small: therefore the improper integral symbols $⨍$ are used and they have to be thought as the analytic continuation in R_v from $\mathrm{Re}\,R_v>0$ and large.

If we do not perform the R_v integrals in the above (8.3), we see that the result of the improper integrals is a holomorphic function of R_v which can be continued to the region of positive and large $\mathrm{Re}\,R_v$. At such large R_v the integrands are well decaying as $\mathrm{Re}\,\tau_v\to\pm\infty$.

Suppose, furthermore, that the tree contains only ripe fruits and represents a contribution to the splitting, (*i.e.* $\gamma_{v_0}=1$), (the case of no fruits being in our notations a special case of only ripe fruits). Then the integrals are holomorphic functions of the τ_v in the strip $g|\mathrm{Im}\,\tau_v|<\frac{\pi}{2}-d$, except for the factors $e^{-\sigma_v gR_v\tau_v}$ and provided the $(\gamma_v)_{v\in\tilde V_f}$, are all equal to 2, and provided $\gamma_{v_0}=1$.

The residues have to be evaluated starting with those relative to the top free nodes (and in arbitrary order) and continuing hierarchically, (*i.e.* always evaluating the residues relative to the highest remaining nodes, in arbitrary order).

Note that the evaluation of contributions from trees with dry fruits as well as from trees with $\gamma_{v_0}=2$ (*i.e.* to values of ripe fruits) involve, instead, integrands with singularities at $\mathrm{Re}\,t=0$ see (6.9). It is a fact, unfortunately, that the interplay between $\gamma=1$ and $\gamma=2$ might look, at first, confusing.

If the tree has $\gamma_{v_0}=1$ and $\gamma_v=2$ for all $v\in\tilde V_f$, then its value (8.3), *i.e.* its contribution to $\beta^{h,1}_{\vec\nu,j}$, is such that we shift the integrations over the $g\tau_v$ to the branches with fixed imaginary part $\xi=\pm(\frac{\pi}{2}-d)$ choosing the sign as we like, by using the shift of contour formulae (3.15).

This implies that, if $\vec\nu_0$ is the total free momentum of our tree, we can compute (8.3) as:

$$\Delta^h_{\vec\nu j}(\vartheta)=\frac{e^{-\xi g^{-1}|\vec\omega\cdot\vec\nu_0|}}{m!}\oint\frac{dR_{v_0}}{2\pi iR_{v_0}}⨍_{-\infty}^{\infty}d\,g\tau_{v_0}e^{-\sigma_{v_0}R_{v_0}(g\tau_{v_0}+i\xi)}w^1_j(\tau_{v_0}+ig^{-1}\xi)\cdot$$

$$
\begin{aligned}
\cdot \prod_{v_0<v\le V_f} & \oint \frac{dR_v}{2\pi i R_v}\Big[\frac{1}{2}\int_{+\infty}^{\tau_{v'}} d\tau_v + \frac{1}{2}\int_{-\infty}^{\tau_{v'}} d\, g\tau_v\Big] e^{-\sigma_v R_v (g\tau_v + i\xi)} . \\
\cdot \prod_{v_0\le v\le V_f} & \Big((\frac{-i(\nu_v)_{\tilde j_v}}{2}) \;\; c_{\nu_v} \;\; e^{i(n_v\varphi^0(\tau_v+ig^{-1}\xi)+\vec\nu_v\cdot\vec\omega\tau_v)} \prod_{s=0}^{l-1} (i\nu_{vs})^{m_v^s} \Big) \cdot \\
\cdot \prod_{v_0\le v\le V_f} & \Big(w_{j_v}(\tau_{v'}+ig^{-1}\xi, \tau_v + ig^{-1}\xi)\Big) \cdot \Big(\prod_{v\in\tilde V_f} x^2_{j_{\lambda_v}}(\tau_{v'}+ig^{-1}\xi)\beta^{h_v,2}_{\vec\nu_{\lambda_v} j_{\lambda_v}}\Big)
\end{aligned}
\tag{8.5}
$$

if all the γ_v, $v\in V_f$, are equal to 2 and if $\gamma_{v_0}=1$ (otherwise the x^{γ_v} or the $w_j^{\gamma_{v_0}}$ are not holomorphic in the strip and the argument does not apply). The factor $e^{-\xi g^{-1}|\vec\omega\cdot\vec\nu_0|}$ is extracted, after shifting the contour up or down according to the sign of $\vec\omega\cdot\vec\nu_0$, from the factors $e^{i\vec\omega\cdot\vec\nu(\tau_v+ig^{-1}\xi)}$.

The factor $e^{-\xi g^{-1}|\vec\omega\cdot\vec\nu_0|}$ can be bounded by ε_h, provided $\vec\nu_0\ne\vec 0$: the latter case $\vec\nu_0=\vec 0$ can indeed be; but we have proved that the cancellations I) show that it gives in fact exactly 0 contribution to the total splitting, after summing its value for all the fruitless trees. Thus the presence of the factor ε_h is explained, to all orders, in the contributions coming from the (sums) of tree values of trees with only ripe fruits and $\vec\nu_0\ne\vec 0$ or from trees with no fruit at all and any $\vec\nu_0$ (equal to $\vec 0$ or not).

The case of trees with only ripe fruits and $\vec 0$ free momentum is also giving a zero contribution to the splitting if the root label is $j>l$, by the cancellations II) and IV) of §6. It remains the case in which the root label is $j=l$: this case, however, is covered by the energy cancellation of §6, once we understand why the small factor ε_h appears in the bound of the splitting to order h in the cases $j>l$.

But the bound (8.2) holds by direct calculation for $h=1$; therefore one assumes it inductively for $h<k$. It is clear that to order k the trees with at least one dry fruit will have a value containing as a factor the value of the splitting to some lower order, in turn bounded by a quantity containing a small factor $\varepsilon_h\le\varepsilon_k$, by the inductive assumption. Hence we just have to worry about the trees with only ripe fruits (no fruit at all being a special case): but they are the ones for which the above analyticity argument is correct. Hence the factor ε_k is present in the bounds of the contributions to the splitting from the trees of order k and either with total free momentum $\ne\vec 0$ or with no fruits or with at least one dry fruit.

The contribution to the splitting from the trees with ripe fruits only and $\vec 0$ total free momentum also gives, by the above reasons, $\vec 0$ contribution to the splitting for the $\uparrow$ components which are therefore, to order k bounded proportionally to ε_k: hence also the $+$,*i.e.* the $j=1$–component of the splitting is bounded proportionally to ε_k.

It remains to estimate the integrals: but this can be done rather easily from the analysis of integrals like (8.3) or (8.5) performed with the aim of finding bounds which do not exhibit the small factors ε_h (which we already know have to be there). In fact such integrals can be essentially exactly computed.

A full discussion of the straightforward but lenghty analysis is reported, for completeness, in appendix A.

We proceed to show that the bounds (8.2) (*i.e.* ultimately the proved canellations) and the proved cancellations $\vec\Delta^h_{\vec 0\uparrow}=0$ imply that the splitting is smaller than any power.

By (8.2), the splitting can be bounded, for any multiindex $\vec a$, by:

$$
|\partial^{\vec a}_{\vec\alpha}\vec\Delta_\uparrow(\vec 0)| \le DJ_0 g\sum_{h=1}^\infty\sum_{0<|\vec\nu|\le Nh} |\mu|^h|\vec\nu|^{|\vec a|}\min\{(b\eta^Q)^{-h}, B_h\varepsilon_h(d)\} \tag{8.6}
$$

having denoted $B_h=B^{h-1}d^{-\beta h}(h-1)!^p$. Note that, if N is the trigonometric degree of the polynomial f in (1.1), the sums over $\vec\nu$ can be suppressed by multiplying the h-th term by the mode counting factor $(2N+1)^{h(l-1)+|\vec a|}$ (*i.e.* the maximum number of non zero Fourier components times the maximum of $|\vec\nu|^{\vec a}$).

¿From this bound it follows that $|\partial^{\vec a}_{\vec\alpha}\vec\Delta_\uparrow|$ is smaller than any power in η (see (1.2)), a result that, as mentioned above after (1.8), also follows from the theory of normal forms ([Nei]). In fact we can split the sum over h in (8.6) into a finite sum, $\sum_{1\le h\le h_0}(\cdot)$ and a "remainder", $\sum_{h>h_0}(\cdot)$; then, if $|\mu|<\frac12 b\eta^Q$ and η is small we find:

$$
\sum_{h>h_0}(\cdot)\le D\,J_0 g\sum_{h>h_0}(\frac{|\mu|\bar C}{b\eta^Q})^h\le 2D\,J_0 g\,(\frac{\bar C|\mu|}{b\eta^Q})^{h_0} \tag{8.7}
$$

where $\bar C=(2N+1)^{(l-1)+|\vec a|}$ takes into account the summation over the $(2N+1)^{l-1}$ modes for which $\vec\Delta^h_{\vec\nu}$ does not vanish; and:

$$
\sum_{h=1}^{h_0}(\cdot)\le DJ_0 g h_0|\mu|\bar C^{h_0}d^{-\beta h_0}B^{h_0-1}(h_0-1)!^p\varepsilon_{h_0}(d) \tag{8.8}
$$

Thus if $\mu=\eta^{Q+s}$, $d=\sqrt{\eta}$, and $s\ge 1$ we see that fixing $h_0=r/s$, for any $r>1$, the $|\partial^{\vec a}_{\vec\alpha}\vec\Delta_\uparrow|$ is bounded by a (r–dependent) constant times η^r (as in such a case (8.8) is just a remainder, exponentially small in $\eta^{-1/2}$).
Note that when $l=2$, the splitting δ introduced in (1.8) is just $\partial_\alpha\Delta_\uparrow$ and $|\vec a|=1$ in (8.6).
In the case $l=2$ ($\omega_0>0$) we can get more precise asymptotic estimates: namely we can prove the well known exponential decay of the splitting, as $e^{-\pi\omega_0/2g\sqrt{\eta}}$, for $\eta\to 0$ when $|\mu|<O(\eta^Q)$ with Q large enough, see the theorem 2 in §1, and below.
To check the latter claim on the asymptotic value of the splitting we simply remark that (8.2) imply the convergence of the series for the splitting for $|\mu|<B^{-1}\eta^{2(N_0+1)}$ (recall that $d=\sqrt{\eta}$. Thus the second and higher order terms can be simply bounded by:

$$\varepsilon_1\, J_0 g\, D\sum_{h=2}^{\infty}\Big(\frac{B|\mu|}{\eta^{N_0+1}}\Big)^2,\qquad \text{if}\quad |\mu|<\frac{1}{2B}\eta^{-2(N_0+1)} \tag{8.9}$$

because, if $l=2$, it is $\varepsilon_h\equiv\varepsilon_1$.
Hence we compute explicitly the first order term. We find (computing from (4.5)), with the above notations, that the first order term $\delta_1\equiv\partial_\alpha\Delta^1_\uparrow(0)$ is given by:

$$\sum_{\substack{|n|\le N\\ |m|\le N_0}} n^2 f_{mn}\int_{-\infty}^{\infty}[\sin\omega nt\sin m\varphi^0-\cos\omega nt(\cos m\varphi^0-1)]dt \tag{8.10}$$

For all $J\le+\infty$. Note that $J=+\infty$ is, if $l=2$, possible.
A simple calculation shows that the leading order (easily studied in terms of the elementarily expressible auxiliary functions: $\int_{-\infty}^{\infty}e^{i\omega nt}[e^{im\varphi^0(t)}-1]dt$) is given, as $\omega\equiv\omega_0/\sqrt{\eta}\to\infty$, by the terms in (8.10) with $m=\pm1$, $n=\pm N_0$ and it is:

$$\eta^{-N_0+\frac12}A_* e^{-\pi\omega_0/2g\sqrt{\eta}}\ ,\qquad A_*\equiv\ -g^{-1}\big(\frac{\omega_0}{g}\big)^{2N_0-1}(f_{N_0,1}+f_{N_0,-1})\frac{4\pi(-1)^{N_0}}{(2N_0-1)!} \tag{8.11}$$

provided $A_*\ne 0$.
Thus we see that the first order dominates if:

$$\frac{|\mu|}{\eta^{N_0-\frac12}}>2\frac{J_0 g}{A_*}\,D\,\frac{B^2|\mu|^2}{\eta^{4N_0+4}} \tag{8.12}$$

which means $|\mu|<\frac{A_*}{2J_0gD}\eta^{3N_0+\frac92}$ and it is implied by $|\mu|<\eta^Q$ with $Q=3N_0+\frac92$ and η small enough.

Comments:

1) A theory for the latter case, in a somewhat different set up[7], can also be found in [DS], [Ge], [GLT], [HMS], [La], based on arguments not relying directly on the cancellations between the contributions to the coefficients of the perturbation series.
2) Note that a value $p\ne0$ in (8.1) would not spoil the asymptotic formula (1.9), if $l=2$: it is easy to check that it would simply make it valid for $Q>N_0+\frac92+p$.
3)The above estimates can be obviously adapted to prove that the quantities $\sigma^h_{\vec\nu,j}$, obtained by looking at the definition (6.6) of $\beta^{h,1}_{\vec\nu,j}$ but replacing the σ by 1 inside the integrals, also verify the bounds (8.2), with different constants. Note, in fact, that the replacement of σ by 1 makes the integrands, in the contribution to $\vec\sigma^h_{\vec\nu,j}$ coming from trees with only ripe fruits, analytic thereby making it possible to imitate the above shift of contour argument.
In [CG] the notion of homoclinic scattering, and of *scattering phase shift function* $\vec\sigma(\vec\alpha)$, was introduced and it could be checked that the $\sigma^h_{\vec\nu,j}$ have the interpretation of Fourier transform of the scattering phase shift function $\vec\sigma(\vec\alpha)$: we do not give a formal proof of the latter identification.

[7] In [DS], [Ge], [HMS] a forced pendulum $\ddot\varphi+\eta\sin\varphi=\mu\eta\sin t$ is considered. After suitable rescalings, such equation arises as Hamilton equation for the hamiltonian $\frac{A}{\sqrt{\eta}}+\frac{I^2}{2}-\cos\varphi-\mu\varphi\sin\alpha$, to which our techniques seems to be adaptable. The adaptation is needed because the latter hamiltonian is not periodic in φ, a property used explicitly in our analysis. In [HMS], [Ge] the exponential smallness of the splitting is obtained under the assumption that $|\mu|\le$ const. η^q with q given, respectively, by 4 and 5/2; in [DS] a is any positive number. In [La], [GLT] the splitting of separatices for the so–called standard map is considered. Our technique seems to be adaptable to cover also such a case and would confirm the exponentially small value of the separatrix splitting found in [La] and [GLT] (see formula (6) in [GLT]).

Appendix A1: Estimates. A conjecture.

Our purpose is to study the integral (8.3) and (8.5). The estimates are straightforward as we are willing to tolerate for $l > 2$, bounds proportional to a power, (any), of the factorial of the order k and to an inverse power, (any), of d^k. The only really non trivial point being the understanding of the cancellations in §6 which, by the shift of contour argument of §8, provide the small factor ε_k. We provide the estimates only for completeness, as the results that we are deriving have often been presented in an incomplete form.

The reader will realize that the following "estimates" are in fact an essentially exact calculation of the integrals involved. Such a feat is made possible by the simplicity of the model 2) in (1.1).

Let ϑ be a resummation tree with $m = m_f + m_0$ nodes, m_f of which are endnodes v contained in some fruit ($m_f \ge 0$). We first consider the (8.3), (8.5) with $\beta^{h,\gamma}_{\vec{\nu},j} \equiv 1$ and with the $m!$ replaced by 1.

Such case contains the case of fruitless trees: once understood one obtains the general case quite easily; see below.

To simplify the algebra we split (8.3) and (8.5) as a sum of many (c^m for some large c, see below) terms. The splitting is rather trivial and it is performed with the aim of being left only with integrals of functions which are products of single argument functions. Recall that $d = \frac{\pi}{2} \pm \xi$, see (8.5).

1) each $w(\tau,\tau')$ is split into two addends, by using the expression (8.4), *i.e.* $\bar{w}_{0l}(\tau)\, w_{00}(\tau') - w_{00}(\tau)\, \bar{w}_{0l}(\tau')$ is regarded as a difference of two terms, and so is $\tau - \tau'$. This splits each of (8.3), (8.5) into up to 2^m terms (in the estimates we sometimes exceed to get simpler expressions: here we bound m_0 by m, for instance). Hence we produce up to 2^m terms.

2) each $\bar{w}_{0l}(t) = 2(gt)\,(\cosh gt)^{-1} - 2\sinh gt$ is split into its two composing addends. Hence we produce, for each of the preceding ones, up to 2^m more terms.

3) in the case of (8.5) we split the $g\tau_v + i\xi$ coming from the action angle elements of the wronskians, or from $\bar{w}_{0l}$, into the two addends composing them, thereby producing up to 2^m more terms for each (see 2)) preceding ones.

4) we split $2^{-1}\sum_\rho$ into two addends getting a factor 2^{-m_0} and up to $2^{m_0} \le 2^m$ more terms (for each of the preceding), and 2^{-m_0} will compensate in the bounds the 2^{m_0} that we get by extracting the factor 2 appearing in 2) above. Note however that it is not important to keep track of powers of m: we are doing the counting just to help the reader checking what we are doing.

The integral is thus split as a sum of up to 2^{5m} terms each of which has the following form ($\omega_v = \vec{\omega}\cdot\vec{\nu}_v$):

$$\prod_v \oint \frac{dR_v}{2\pi i R_v} \prod_{v_0 \le v \in \vartheta} \int_{\rho_{v'}\infty}^{\tau_{v'}} d\, g\tau_v\ \left[e^{-R_v\sigma_v i\xi}\right]\, e^{-R_v\sigma_v g\tau_v}(g\tau_v)^{n_v} e^{i\omega_v\tau_v} \prod_{j=1}^{m_v+1} y^v_j(\tau_v) \tag{A1.1}$$

where the terms in square brackets are present only if we are condidering (8.5) and the $y^v_j(\tau)$ are elements of a *finite* set of functions: namely, in the case of (8.3), $e^{in\varphi(\tau)}$ with $n = 0, \pm 1, \ldots, \pm N_0$ if N_0 is the trigonometric degree of f in (1.1), and $(\cosh gt)^{-1}$, $\sinh gt$, see (2.15). Here $\rho_v = \pm$ *and it is not the sign of* τ_v, but it is an independent variable. The factor 2^{-m_0}, found in step 5) above, is not included in (A1.1) and it will be brought back later, (see (A1.2) below). The product over v is over the m_0 "free nodes" of ϑ, *i.e.* which correspond to actual integration operations. And $\tau'_{v_0} \equiv 0$.

Likewise, in the case of (8.5), the $y^v_i(\tau)$ are to be found among the $e^{in\varphi(\tau + ig^{-1}\xi)}$ with $n = 0, \pm 1, \ldots, \pm N_0$, and $(\cosh(g\tau + i\xi))^{-1}$, $\sinh(gt + i\xi)$, and, as said above, the terms in square brackets have to considered present.

Furthermore $n_v \le m_v + 1$ is the number of factors τ_v that are collected from the wronskians in performing the above operations. But $\sum_v n_v \le m$, because each node v can contain at most one factor τ_v coming from the wronskians (recall that only $\bar{w}_{0l}(\tau)$ and w_{ij} with $0 < i < l, l < j < 2l-1$ can contain τ explicitly).

The functions y_j are holomorphic for $|\,\mathrm{Im}\, gt| < \frac{\pi}{2} - \xi$ or, if thought as functions of $x = e^{-\sigma gt}$, $\sigma = \mathrm{sign}\, t$, are holomorphic for $|x| < 1$ *and* outside the cone with half opening ξ centered at the origin and symmetric about the imaginary axis (the case $\xi = 0$ refers to (8.3)). Furthermore they are in the class $\hat{\mathcal{M}}$ introduced in §2,§3: see (2.15) and (3.3).

The contour integrals involving the R_v are over a small circle, around the origin, and they really denote the evaluation of the appropriate residues, see §8.

Note, as this will be quite relevant below, that the y functions, as functions of x at fixed σ are holomorphic near $x = \pm 1$, as they only have polar singularities at $x = \pm i$ or at $\pm e^{i\xi}$

We reconstruct (8.3), from the integrals (A1.1), as a linear combination of the above intergals, (up to 2^{4m}, as one can check by keeping track of the above accounted proliferation of the terms), times suitable factors, all bounded by:

$$F^{m_0}\left(\tfrac{\pi}{2}\tfrac{N}{2}\right)^{m_0}N^{m_0} \tag{A1.2}$$

and N is the maximum degree of the trigonometric polynomial f in (1.1) and $F \geq 1$ is a bound on the Fourier components of $(J_0 g^2)^{-1} f$. The factors bounded by powers of ξ ($\leq m_0$) coming from the binomial expansions met in the splittings considered in 3) have been bounded by $\frac{\pi}{2}$ to simplify the notation.

5) We consider one of the nodes $\bar v$ following v_0 and split the integration over $\tau_{\bar v}$, *i.e.* $⨍_{\rho_{\bar v}\infty}^{\tau_{v_0}} d\tau_{\bar v}$, as a sum of $⨍_{\rho_{\bar v}\infty}^{0} - ⨍_{\rho_{v_0}\infty}^{0} + ⨍_{\rho_{v_0}\infty}^{\tau_{v_0}}$, leaving for later consideration the first two choices. And we repeat the procedure, generating up to 3^m terms: all of them left for future consideration except the one which has all the τ_v variables integrated in the interval $[0, \rho_{v_0}\infty)$, *i.e.* all having the same sign. We shall return to the terms left out for future consideration in item 8) below.

We are thus left with an integral like:

$$I_m = \prod_{v \geq v_0} \int_{-\infty}^{\tau_v'} d\tau_v [e^{R_v i\xi}] e^{R_v \tau_v g} e^{i\omega_v \tau_v} (g\tau_v)^{n_v} \prod_{j=1}^{m_v+1} y_j^v(\tau_v) \tag{A1.3}$$

where the terms in square brackets are present only in the case (8.5) and $\rho_{v_0}\infty$ has been supposed, to fix the ideas, to be $-\infty$. As before the term in square brackets is present only in the case of (8.5).

Of course one might think that we are undoing all that was painfully done in §6. And in fact this is essentially the case; the work of §6 was performed only to exhibit the cancellations.

We shall let d be $\frac{\pi}{4}$ in the case (8.3), while in the case (8.5) it will be actually d (to unify the notation). If we now show that (A1.3) can be bounded by:

$$|I_m| \leq I_m^0 \equiv B_1^m [d^{-\beta m}] \max_{0<|\vec\nu|<Nm} \left(g^{-1}|\vec\omega \cdot \vec\nu|\right)^{-2m} \tag{A1.4}$$

we will have shown that all the up to 3^m terms generated by the above decoposition of the above integral are bounded by the same quantity. In fact the terms left above "for later consideration" are manifestly products of quantities bounded by $I_{m_1}^0 \cdot \ldots \cdot I_{m_p}^0$ with $\sum m_i = m$.

6) We write the first integral in (A1.3) as $⨍_{-\infty}^{-g^{-1}} d\tau_{v_0} + \int_{-g^{-1}}^{0} d\tau_{v_0}$: this gives us two terms. The first of which is:

$$\int_{-\infty}^{-g^{-1}} d\,g\tau_{v_0} \prod_{v>v_0} \int_{-\infty}^{\tau_{v'}} \cdot \Big(\prod_{v\geq v_0} [e^{R_v i\xi}] e^{R_v \tau_v g} e^{i\omega_v \tau_v} (g\tau_v)^{n_v} \prod_{j=1}^{m_v+1} y_j^v(\tau_v)\Big) \tag{A1.5}$$

We shall come back to the second integral in item 7) below.

The functions $y(\tau)$ can be expanded into a series:

$$y_j^v(\tau) = \frac{1}{x}\sum_{p=0}^{\infty} y_{j,p}^v x^p, \quad x = e^{-\sigma g\tau} \tag{A1.6}$$

and some of the $y_{j,0}$ may vanish, but not necessarily all (because of the possible choice: $y(\tau) = \sinh g\tau$).

If each of the y_j is expanded as in (A1.6), then each of the (A1.1) is broken into a sum over labels $\{k_j^v\}$, with $v \in \vartheta$, and $j = 1, \ldots, m_v + 1$, with convenient weights, namely:

$$\prod_{v\in\vartheta}\prod_{j=1}^{m_v+1} y_{j,k_j^v}^v \tag{A1.7}$$

of integrals like:

$$\left[\prod_v e^{i\xi R_v}\frac{\partial^{n_v}}{\partial E_v^{n_v}}\right] \prod_{v_0\leq v\in\vartheta} \int_{-\infty}^{\tau_{v'}} e^{+R_v g\tau_v}[e^{iR_v\xi}]\, e^{(i\omega_v+E_v)\tau_v} x_v^{k_v}\, d\,g\tau_v \tag{A1.8}$$

where v' denotes the node immediately preceding v, and the auxiliary paramenters E_v have to be put equal to 0 after differentiation while the R_v will have to be integrated over the above small contour around the origin. The $[e^{i\xi R_v}]$ factors are present only in the case of (8.5).

The integrals (A1.8) are performed hierarchically. This means that we first integrate with respect to the τ_v's with v being a top node, in arbitrary order.

To perform such integration we use (3.8):

$$\int_{-\infty}^{t} x^K e^{(i\Omega+E)\tau}\, d\, g\tau = \frac{x^K e^{(i\Omega+E)t}}{K+g^{-1}(i\Omega+E)} \tag{A1.9}$$

valid if the denominator does not vanish, (needless to say), and if $t<0$.

After integrating over the τ_v variable corresponding to a top free node v, arbitrarily fixed, we compute the E_v derivatives and the residues at $R_v=0$, and then we shall repeat the procedure after deleting the top free node v and the branches outgoing from it.

The E_v derivatives can act either on the numerators or on the denominators of (A1.9). The n_v differentiations with respect to E_v give results, if v' denotes the node immediately preceding v, like:

$$n!\, \frac{x^{k_v} e^{gR_v\tau_{v'}} e^{i\xi R_v} e^{i\omega_v\tau_v}}{(k_v+R_v+ig^{-1}\omega_v)^{n+1}}\, (g\tau_{v'})^{n'} \tag{A1.10}$$

with $n+n'=n_v$ and the terms like (A1.10) add up with coefficients ± 1 to give the result of the derivative of the τ_v integral that we are considering. The number of addends is not greater than 2^{n_v} (because we are taking derivatives of a product of at most two factors, when $\lambda_v=1$). Here $\omega_v=\vec\omega\cdot\vec\nu_v$.

The residue of R_v^{-1} times (A1.10) at $R_v=0$ is simply (A1.10) itself evaluated at $R_v=0$ if k_v or $g^{-1}\omega_v$ are different from 0 (because at $R_v=0$ the denominator in (A1.10) does not vanish being bounded below by either $|k_v|\ge 1$ or by $|g^{-1}\omega_v|$; and otherwise it is:

$$\frac{n!}{(n+1)!}(g\tau_{v'})^{n'}(g\tau_{v'}+i\xi)^{n+1} \tag{A1.11}$$

still with $n+n'=n_v$ and, as above, the terms with $i\xi$ are present only in the bound of (8.5). Developing the binomial in (A1.11) we can say that the result of the derivative and residue evaluation is a sum of at most 2^{n_v+1} terms, per each of the already found $3^m 2^{n_v+1}$, which have the form:

$$G\,\tilde n!\, x_v'^{\tilde k_v} e^{i\tilde\omega_v\tau_{v'}} (g\tau_{v'})^{\bar n}, \qquad x'=e^{-\sigma_{v'}\tau_{v'}g} \tag{A1.12}$$

with $\tilde n+\bar n\le n_v+1$ and $\tilde\omega_v=\vec\omega\cdot\vec\nu_v$; the latter in general will be, when considering the other integrals correponding to inner nodes, the sum of the modes $\vec\nu_w$ with $w\ge v$. The coefficients G are bounded by the maximum between 1 (corresponding to the a bound on the denominator in (A1.10) when $k_v\ne 0$), or $(\min_{0<|\vec\nu|\le N}|\vec\omega\cdot\vec\nu|g^{-1})^{-n_v-1}$ (corresponding to $k_v=0,\omega_v\ne 0$) or $(\pi/2)^{n_v+1}$ (corresponding to $k_v=\omega_v=0$, $n=n_v, n'=0$, using (A1.11) and recalling that $|\xi|<\frac{\pi}{2}$).

Continuing the procedure we integrate successively over the τ variables of all the top free nodes (in arbitrary order), each time deleting the considered node and its outgoing branches.

Therefore after all the integrations have been performed, all derivatives and residues computed, we shall have expressed the result of the evaluation of (A1.3) (taking into account the listed proliferation of terms described after (A1.4), (3^m), after (A1.10), ($\prod_v 2^{n_v}$), and before (A1.12), ($\prod_v 2^{n_v+1}$)), as a sum of up to $3^m 2^{3m}$ terms: *we use here that* $\sum n_v\le m$ *as there can be only up to one factor* τ_v *per node.* Each of which is bounded by:

$$\begin{aligned} &e^{-\sum_{v,j}k_j^k} c_1^m\cdot \max_{0<|\vec\nu|\le Nm}(g^{-1}|\vec\omega_0\cdot\vec\nu|)^{-2m} \quad &&\text{if}\quad l>2\\ &e^{-\sum_{v,j}k_j^k} c_1^m\cdot 1 &&\text{if}\quad l=2\end{aligned} \tag{A1.13}$$

because the factors x_v are $\le e^{-1}$ in the integration interval, and by (A1.9), eventually, they are evaluated at $x_v=e^{-1}$; here c_1 is a suitable constant and in the cases $l>2$ we supposed $g/|\vec\omega_0|>1$ and neglect the (favourable) $\sqrt{\eta}$ in $\vec\omega=\vec\omega_0/\sqrt{\eta}$, to simplify the algebra (which is not restrictive unless one cares about numerically good estimates, which is not our desire here). In the case $l=2$, on the contrary, we use explicitly that $\min_{\vec\nu\ne\vec 0}|\vec\omega\cdot\vec\nu|=\omega_0/\sqrt{\eta}$, very special for this "non resonant" case.

The reason why the minimum has to be considered only for $|\vec\nu|\le Nm$ is that at each integration the oscillating exponent will have the form $\vec\omega\cdot\vec\nu$ with $\vec\nu$ being sum of the, at most m, modes present in f in (1.1) (and in fact of $m_0\le m$).

Collecting all the above terms and estimates and using $\sum_v n_v \le m$ we have a bound on (A1.8):

$$\begin{cases} \bar{\mathcal{N}}_l^m\,(Cm^\tau)^{2m} \le \mathcal{N}_l^m m!^{2\tau} & \text{if } l \ge 2 \\ \mathcal{N}_2^m & \text{if } l = 2 \end{cases} \tag{A1.14}$$

for some suitable constants $\mathcal{N}_l$, and if C, τ denote respectively the diophantine constant and the diophantine exponent of $\vec{\omega}_0$, see (1.3).

The latter bound should be multiplied by the absolute values of the weights (A1.7), by (A1.2) and by the "multiplicity" (produced in the course of the analysis 1)%5)) 2^{5m} in order to produce, after summing over the k labels in (A1.7), a bound on (8.3) and on (8.5) with the $\beta^{\cdot,\cdot}_{\cdot,\cdot} = 1$.

The weights are Taylor coefficients of a few (*i.e.* a finite number) of functions of x, with radius of convergence $|x| = 1$.

Therefore they can be bounded by a common bound M_λ on the maxima of such functions in any disk of radius $\lambda < 1$ times λ^{-k}. The integrals over the τ's run over the interval $(-\infty, -g^{-1})$ so that the x_v are always $\le e^{-1}$. Hence, for $\lambda = 2^{-1}$, we get convergent bounds because of the exponential factors in (A1.13). Clearly this was the reason for the splitting of the integrals over the τ_v into the part with $\tau_v < -g^{-1}$ and the part with $\tau_v \in [-g^{-1}, 0]$.

Note that there are at most $(2N_0+3)$ different functions y and at most $3m$ of them appear: and of these there will be up to m trigonometric functions of $n\varphi_0(\tau)$ (in the case of (8.3)) or of $n\varphi_0(\tau + i\xi)$ (in the case of (8.5)) and up to $2m$ functions coming from the wronskians. The first admit a bound proportional to d^{-2N_0} as the $\cos\varphi_0(\tau)$ and $\sin\varphi_0(\tau)$ carry a polar singularity of second order on the unit circle at distance of order d (*i.e.* ξ) from the real axis. The wronskians are bounded proportionally to d^{-1} as they carry, at worst, a simple pole on the unit circle.

Hence we see that a bound on the the products of y_j that can be met in the integrals (8.5) has the form $M_*^m d^{-2(N_0+1)m}$ for some constant M_*. In the case of (8.3) we can take the bound to have the same form with the same M_* and $d = \frac{\pi}{4}$, say.

7) We can therefore switch to considering the "left out part" $\int_{-g^{-1}}^0 d\tau_{v_0}$. Let v_1 be one of the nodes following v_0. We break the integral $⨍_{-\infty}^{\tau_{v_0}} d\tau_{v_1}$ as $⨍_{-\infty}^{-g^{-1}} d\tau_{v_1} + \int_{-g^{-1}}^0 d\tau_{v_1}$. If we make the first choice and if m_1 is the number of nodes following v_0 in the direction v_1 the first integral can be bounded, by the previous argument, by $I^0_{m_1}$ and we are left with the problem of bounding $I^0_{m_1} \cdot \int_{-g^{-1}}^0 d\tau_{v_0} \prod_{v \ni \vartheta_{v_0 \bar{v}}} ⨍_{-\infty}^{\tau_v} \ldots$.

We repeat the procedure hierarchically and we are eventually left with up to 2^m integrals like $I^0_{m_1} \ldots I^0_{m_p}$ $\prod_{v \in \tilde{\vartheta}} \int_{-g^{-1}}^{\tau_{v'}} d\tau_v \ldots$ to bound, where $\tilde{\vartheta}$ is a subtree of ϑ with $\tilde{m}$ nodes, and $\tilde{m} + \sum m_j = m$.

The last integral is manifestly bounded by the maximum of the integrand, which is $\tilde{B}^m d^{-2(N_0+1)\tilde{m}}$ for a suitable $\tilde{B}$, because $\int_{-g^{-1}}^0 dg\tau = 1$, we see that there is a constant B_2 such that each of the the up to $3^m 2^{3m}$ terms we generated is bounded by $B_2^m [d^{-2(N_0+1)m}]$.

8) If one selects for consideration any other of the $3^m - 1$ integrals left aside in the analysis of item 5), see lines preceding (A1.3), one is left with essentially the same problem discussed in items 5)%7) above. In fact such choices involve factorized integrals, each of which has exactly the same form as the one studied in 5)%7) above.

9) *We conclude that the final bound on the* <u>*sum*</u> *over the mode values* $\vec{\nu}$ *and over all the other mode labels (which is a sum over up up to to* $(2N+1)^{lm}$ *terms) of the absolute values of (8.3) or (8.4) with all the* $\beta^{\cdot,\cdot}_{\cdot,\cdot} = 1$*, is:*

$$\begin{aligned} & c_3^m M_*^m\, m!^{2\tau}, && \text{in the case of (8.3)} \\ e^{-|\xi\vec{\omega}\cdot\vec{\nu}_0|} d^{-2(N_0+1)m} & c_3^m M_*^m\, m!^{2\tau}, && \text{in the case of (8.5)} \end{aligned} \tag{A1.15}$$

with $\xi = \frac{\pi}{2} - d$ and c_3 conveniently fixed; and we shall use the (A1.15) to get a recursive bound on the coefficients $\beta^{\cdot,\cdot}_{\cdot,\cdot}$. *Furthermore if* $l = 2$ *the diophantine exponent* τ *can be taken* $\tau = 0$.

The (A1.15) can be used to play the role plaid by the bound on the fruitless trees in the KAM proof in §7 in terms of D_0, B_0: in our case the corresponding quantities are given by $\varepsilon_1 M_*^2 c_3^2 d^{-4h(N_0+1)} (2h)^{4\tau}$ and $M_*^2 c_3^2 d^{-4h(N_0+1)} (2h)^{4\tau}$ for trees with order up to h. The extra power of 2 is lost here because $m \le 2\tau$ (recall that the order h of a tree is in general different from the number of nodes because some nodes may carry an order label $\delta_v = 0$, and $h \le m < 2h$).

And one has simply to make an argument parallel to that presented in the paragraph containing (7.4) and (7.5). The cancellations discussed in §6 show that, among the subtrees that are generated by "magnifying

the fruit seeds", there has to be at least one which will provide a small factor $e^{-|\xi\vec{\omega}\cdot\vec{\nu}_0|}$ for some $\vec{\nu}$ with $|\vec{\nu}|\le Nh$. We leave the details to the reader, to avoid repetitions of the bounds discussed in §7.
The final estimate of D,B is that $\beta=4(N_0+1)$, $p=4\tau$ for the same reason which shows that in (7.5) one can take D,B proportional to D_0,B_0.

Remark: A more careful analysis of the higher orders would show that this can be improved to $|\mu|<\eta^Q$ with $Q>N_0+\frac{9}{2}$, simply because the origin of the $4N_0h$ was a "poor" bound on the number of factors $f_\nu e^{i(n\varphi+\vec{\nu}\cdot\vec{\alpha})}$ present in the expressions of the splitting to order h and each of which contributes a d^{-2N_0}, (hence eventually a $(\sqrt{\eta})^{-2N_0}$): the number of such factors, in deriving (8.2), by $2h$. But to order h there are, obviously, exactly h such factors, and we can therefore essentially replace $4N_0$ by $2N_0$ (changing the various constants).

If $J=+\infty$ the above analysis can be repeated: we realize that several simplifications take place. For instance $\beta^{h,\gamma}_{\vec{\nu},j}\equiv 0$ if $j=1,\ldots,l-1$ and in fact $\bar{D}$ can be replaced by D_j with $D_j=\bar{D}$ if $j=0$ and $D_j=\frac{J_0}{J}$ if $j>0$. It follows also that the trees with at least one inner branch with $j_\lambda=1$ give a vanishing contribution (corresponding to the fact that $X^h_\downarrow\equiv\vec{0}$ as the angles $\vec{\alpha}$ are isochronous, *i.e.* $\vec{\alpha}=\vec{\omega}$ even if $\mu\neq 0$). The splitting being $Jg\beta^{h,1}_{\vec{\nu},j}$ we see that the second bound in (8.2) holds uniformly in the size of $J\ge J_0$, *i.e.* the constants D,B,β,p do not depend on J. This is again as in the KAM theory of §7.
The first bound in (8.2) is more delicate, if $J=+\infty$, as it is not "purely algebraic", resting on the KAM theory: if $l=2$ one does not really need the KAM theory as mentioned in §1 and also the first bound holds. If $l>2$ the KAM theory does not apply, in general, and only the second estimate (8.2) is proved by the above arguments. This implies in particular the (not surprising) fact that perturbation theory for the whiskers and their splitting is well defined to all orders; but convergence is not guaranteed (*i.e.* the first of (8.2) may fail) and one cannot guarantee even the persistence of the invariant tori.
To summarize (8.2) holds uniformly in $J\ge J_0$ if $l=2$; if $l>2$ only the second of (8.2) holds uniformly in $J\ge J_0$, and it is not even sufficient to yield the convergence of the formal perturbation theory for the tori and the splitting. If $l>2$ the first of (8.2) holds at fixed $J<+\infty$: as needed and claimed in §8.

A conjecture.

Nevertheless we think that even if $l>2$, and for all $J\le+\infty$, but with $J\ge J_0$, the τ can be set equal to 0: in other words *we conjecture that the above theory can be completely freed from its dependence on a KAM type of proof of existence of the invariant tori and whiskers (like the quoted [CG]), so that the two theories are essentially equivalent and independent even if $l>2$.* We believe that the essential has been done already in the present paper. One would have to improve a little the cancellations analysis of §6. And the form of (1.1) will have to play an essential role (as it already did above). The main feature of (1.1) to keep, in order to be able to believe the above conjecture, is the absence of action variables in the perturbation f.

Appendix A2: Symmetry of rootless trees.

We first prove (6.12). Integrating the second of (6.11) over R_1,R_2 on circles with radii r_1,r_2 with $0<r_2<r_1$ and r_1 small enough we get a l.h.s. equal to the l.h.s. of the relation (6.12); while the r.h.s. of (6.12) is obtained by taking $0<r_1<r_2$, and r_2 small enough. Hence we want to deform the radius r_1 to become larger than r_2. This can be done if no singularities in R_1,R_2 are met near the product of two circles with radius r_2. The evaluation of the double integral in (6.11) leads to a function which can have some singular terms when $R_1=\pm R_2$: it is easy to see, representing F,G as a sum of finitely many terms like (3.2) plus very fastly decreasing functions at $\pm\infty$, that the singular terms have the form $(R_1\pm R_2)^{-n}$, $n>0$. Therefore after multiplying them by $R_1^{-1}R_2^{-1}$ we see that they have equal double residue at 0, whether we take first the residue in R_1 or in R_2 (and the result is zero in both cases). So that (6.12) follows.
Given a tree ϑ_0 (with all its labels) we can consider the object obtained by deleting the root branch (with its labels). The object $\bar{\vartheta}$ thus obtained will be a *rootless tree* $\bar{\vartheta}$.
The same rootless tree can be obtained from several trees, as by deleting the root branch we forget the action label it beared. And we also forget which node was the first node.
We can find the trees which generate the same rootless tree $\bar{\vartheta}$ by attaching a root branch to any of the free nodes $v\in\bar{\vartheta}$. Some restrictions may apply to the labels that can appear on the new root branch. In fact from the rules of tree labeling of §5, and fron the (2.14), (4.10) it appears that if the type δ_v of the node is

1 no restrictions apply; but if $\delta_v = 0$ the action label of the new branch has to be 0.
A rootless tree with a distinguished node will be a pair $(\bar\vartheta, v)$ of a rootless tree and one of its nodes.
It is convenient to define the value $V(\bar\vartheta, v)$ of a rootless tree with a distinguished node: we shall use the following prescription.
1) Imagine to attach a root branch at the node v with any possible action label.
2) Evaluate the tree value with the rules of §5, but the node function introduced in (5.12) relative to the distinguished node will be modified by deleting the factor $-\frac{i}{2}(\nu_v)_{j_v}$ and replacing it by 1.
3) The $\mathcal{E}^T_v$ operation is *always* $\mathcal{I}$ (even when the attached root branch action label is $j_v = 0$, which would require, for a normal tree, the interpretation $\mathcal{E}^T_v(\cdot) = \mathcal{I}(w_{00}\cdot)$).
By construction $V(\bar\vartheta, v)$ does not depend on the label j_v that we attach to the new root to perform the above calculation. In principle, however, it does depend on the distinguished node v. The main consequence of the symmetry (6.12) is that *in fact it is v independent.* Thus we can denote it $V(\bar\vartheta)$, calling it the value of the rootless tree $\bar\vartheta$.
To prove the above independence we write the just defined value $V(\bar\vartheta, v)$ as:

$$\int_{-\infty}^{+\infty} d\tau \Big(\prod_{p=1}^r (i\nu_v)_{j_p}\Big) e^{in_v\varphi_0(\tau)+(\vec\alpha+\vec\omega\tau)\cdot\vec\nu_v} \prod_{p=1}^r \tilde X^{h_p}_{j_p}(\tau) \tag{A2.1}$$

if r is the number of branches arriving in v and h_p is the order of the p-th branch (defined as the sum of the type labels of the nodes that can be reached leaving v along the p-th branch; the j_p is the angle label of the p-th branch. The $w(\tau,\tau')$ is either $(\tau-\tau')$ if $j_1 > 0$ or $w_{00}(\tau)w_{0l}(\tau') - w_{0l}(\tau)w_{00}(\tau')$ if $j_1 = 0$ (apart from some proportionality constant fixing the dimensions). The $\tilde X^{h_p}_{j_p}(\tau)$ is the result of the integrations over the time labels of the nodes of $\bar\vartheta$ which follow the p-th branch emerging from v (in the order generated on $\bar\vartheta$ by the insertion of a root branch).
Let v' be one of the r nodes linked by a branch of $\bar\vartheta$ to v. We suppose that the branch is the one corresponding to $p = 1$. Then, by the definition of the $\mathcal{R}$–tree evaluation we see that:

$$X^{h_1}_{j_1}(\tau) = \sum_\rho \frac{1}{2}\int_{\rho\infty}^{\tau} w(\tau,\tau')d\tau' \Big(\prod_{q=1}^s (i\nu_{v'})_{j'_q}\Big) e^{in_v\varphi_0(\tau')+(\vec\alpha+\vec\omega\tau')\cdot\vec\nu_v} \prod_{q=2}^r \tilde X^{h'_q}_{j'_q}(\tau') \tag{A2.2}$$

if s is the number of branches arriving in v' and h_q is the order of the q-th branch (defined as the sum of the type labels of the nodes that can be reached leaving v' along the q-th branch. The j'_q is the angle label of the q-th branch.
Substituting (A2.2) into (A2.1) we get an expression like the l.h.s. of (6.12): then it is immediate to check that the expression to which it is equal, by applying (6.12), is in fact $V(\bar\vartheta, v')$.
Finally one remarks that the value of a tree ϑ_0 with root label $l + j_{v_0}$ with $j = j_{v_0} > 0$ is expressed in terms of the corresponding rootless tree $\bar\vartheta$ value as $-\frac{1}{2}(i\nu_{v_0})_j V(\bar\vartheta)$. This would not be true for $j = 0$ because the evaluation of the tree value would require using the operation $\mathcal{I}(w_{00}\cdot)$ instead of $\mathcal{I}$.
If $\vartheta_0, \vartheta_1, \ldots, \vartheta_u$ are the trees that can be associated with a given rootless tree $\bar\vartheta$ and which have a root branch action label $l + j$ with $j > 0$, we see that the sum of their values is:

$$-\frac{i}{2}\sum_k (\nu_{v_k})_j V(\bar\vartheta) = -\frac{i}{2}(\vec\nu_0)_j V(\bar\vartheta) \tag{A2.3}$$

where $\vec\nu_0$ is the total free mode of the trees (which is the same for all).
Hence we see that if the total free mode vanishes the corresponding trees give a vanishing contribution to the homoclinic splitting. And the cancellation takes place separately among the trees that have the same rootless tree. In particular we can say that the family of all the trees with only ripe fruits and $\vec 0$ total free mode gives a zero contribution to the action splitting for $j > 0$, (of course the same could be said of the family of the trees with only dry fruits, or with a prefixed number of dry and ripe fruits and of free nodes).

Appendix A3: Analysis of the approximate cancellations.

We have seen in §7 the basic mechanism, *given for the purpose of illustration*, showing that, restricting the sum in (6.23) to a sum over the trees such that:

I) $\vec{\nu}_f(v) \neq \vec{0}$ if v is any node (seeds included).

II) $\vec{\nu}_f(v) \neq \vec{\nu}_f(v')$ for all pairs of comparable nodes v', v, (not necessarily next to each other in the tree order, however), with $v' \geq v_0$.

leads to a convergent series. The trees not verifying II) are called *resonant* by Eliasson. And in this section we deal with them. Which is the hardest part of the problem and the most original contribution by Eliasson to the field.

I shall consider only the case of trees without fruits, as the reduction, in §7, to this case was not based on assumption II).

However *there are* resonant trees. The key remark is that they cancel *almost exactly*. The reason is very simple: imagine to detach from a tree ϑ the subtree ϑ_2 with first node v. Then attach it to all the *remaining* nodes $w \geq v', w \in \vartheta/\vartheta_2$. We obtain a family of trees whose contributions to $h^{(k)}$ differ because:

1) some of the branches above v' have changed total momentum by the amount $\vec{\nu}(v)$: this means that some of the denominators $(\vec{\omega} \cdot \vec{\nu}(w))^{-2}$ have become $(\vec{\omega} \cdot \vec{\nu}(v) \pm \varepsilon)^{-2}$ if $\varepsilon \equiv \vec{\omega}_0 \cdot \vec{\nu}(v)$; and:
2) because there is one of the node factors which changes by taking successively the values ν_{wj}, j being the branch label of the branch leading to v, and $w \in \vartheta/\vartheta_2$ is the node to which such branch is reattached.

Hence if $\vec{\omega} \cdot \vec{\nu} = \varepsilon = 0$ we would build in this *resummation* a quantity proportional to: $\sum \vec{\nu}_w = \vec{\nu}(v) - \vec{\nu}(v')$ which is zero, because $\vec{\nu}(v) = \vec{\nu}(v')$ means that the sum of the $\vec{\nu}_w$'s vanishes. Since $\vec{\omega} \cdot \vec{\nu} = \varepsilon \neq 0$ we can expect to see a sum of order ε^2, if we sum as well on a overall change of sign of the $\vec{\nu}_w$ values (which sum up to $\vec{0}$).

But this can be true only if $\varepsilon \ll \vec{\omega} \cdot \vec{\nu}'$, for any branch momentum $\vec{\nu}'$ of a branch in ϑ/ϑ_2. If the latter property is not true this means that $\vec{\omega} \cdot \vec{\nu}'$ is small and that there are many nodes in ϑ/ϑ_2 of order of the amount needed to create a momentum with small divisors of order ε.

Examining carefully the proof of Brjiuno's lemma one sees that such extreme case 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 and are called "overlapping divergences". Their systematic analysis is made through the renormalization group methods. We argue here that Elliasson's method can be interpreted in the same way.

The above introduced trees will play the role of *Feynman diagrams*; and they will be plagued by overlapping divergences. They will therefore be collected into another family of graphs, that we shall call *trees*, on which the bounds are easy. The $(\vec{\omega} \cdot \vec{\nu})^{-2}$ are the *propagators*, in our analogy.

We fix an *scaling* parameter γ, which we take $\gamma = 2$ for consistency with (7.1), and we also define $\vec{\omega} \equiv C_0 \vec{\omega}_0$: it is an adimensional frequency. Then we say that a propagator $(\vec{\omega} \cdot \vec{\nu})^{-2}$ is *on scale* n if $2^{n-1} < |\vec{\omega} \cdot \vec{\nu}| \leq 2^n$, for $n \leq 0$, and we set $n = 1$ if $1 < |\vec{\omega} \cdot \vec{\nu}|$.

Proceeding as in quantum field theory, see [G3], given a tree ϑ we can attach a *scale label* to each branch $v'v$ in (6.23) (with v' being the node preceding v): 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 function of helping to visualize the orders of magnitude of the various tree branches.

Looking at such labels we identify the connected clusters T of nodes that are linked by a continuous path of branches with the same scale label n_T or a higher one. We shall say that *the cluster T has scale n_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. Here we use that the tree branches carry an arrow pointing to the root: this gives a meaning to the words "incoming" and "outgoing".

If V is one such cluster we denote λ_V the incoming branch: the branch scale $n = n_{\lambda_V}$ is smaller than the smallest scale $n' = n_V$ of the branches inside V. We call w_1 the node into which the branch λ_V ends, inside V. *We say that such a V is a resonance if the number of branches contained in V is* $\leq E\,2^{-n\varepsilon}$, where $n = n_{\lambda_V}$, and E, ε are defined by: $E \equiv 2^{-3\varepsilon} N^{-1}$, $\varepsilon = \tau^{-1}$. We shall say that n_{λ_V} is *the resonance scale*.

Let us consider a tree ϑ and its clusters. We wish to estimate the number N_n of branches with scale $n \leq 0$ in it, assuming $N_n > 0$.

Denoting T a cluster of scale n let m_T be the number of resonances of scale n contained in T (*i.e.* with incoming branches of scale n), we have the following inequality, valid for any tree ϑ:

$$N_n \leq \frac{3k}{E\,2^{-\varepsilon n}} + \sum_{T,\, n_T = n} (-1 + m_T) \tag{A3.1}$$

with $E = N^{-1} 2^{-3\varepsilon}, \varepsilon = \tau^{-1}$. This is a version of Brjuno's lemma: a proof is in appendix A4.

Consider a tree ϑ^1 we define the family $\mathcal{F}(\vartheta^1)$ generated by ϑ^1 as follows. Given a resonance V of ϑ^1 we detach the part of ϑ^1 above λ_V and attach it successively to the points $w \in \tilde{V}$, where $\tilde{V}$ is the set of nodes of V (including the endpoint w_1 of λ_V contained in V) outside the resonances contained in V. Note that all the branches λ in $\tilde{V}$ have the same scale $n_\lambda = n_V$.

For each resonance V of ϑ^1 we shall call M_V the number of nodes in $\tilde{V}$. To the just defined set of trees we add the trees obtained by reversing simoultaneously the signs of the node modes $\vec{\nu}_w$, for $w \in \tilde{V}$: the change of sign is performed independently for the various resonant clusters. This defines a family of $\prod 2M_V$ trees that we call $\mathcal{F}(\vartheta_1)$. The number $\prod 2M_V$ will be bounded by $\exp \sum 2M_V \le e^{2k}$.

It is important to note that the definition of resonance is such that the above operation (of shift of the node to which the branch entering V is attached) does not change too much the scales of the tree branches inside the resonances: the reason is simply that inside a resonance of scale n the number of branches is not very large being $\le \overline{N}_n \equiv E\, 2^{-n\varepsilon}$.

Let λ be a branch, in a cluster T, contained inside the resonances $V = V_1 \subset V_2 \subset \ldots$ of scales $n = n_1 > n_2 > \ldots$; then the shifting of the branches λ_{V_i} can cause at most a change in the size of the propagator of λ by at most $2^{n_1} + 2^{n_2} + \ldots < 2^{n+1}$.

Since the number of branches inside V is smaller than $\overline{N}_n$ the quantity $\vec{\omega}\cdot\vec{\nu}_\lambda$ of λ has the form $\vec{\omega}\cdot\vec{\nu}^0_\lambda + \sigma_\lambda \vec{\omega}\cdot\vec{\nu}_{\lambda_V}$ if $\vec{\nu}^0_\lambda$ is the momentum of the branch λ "inside the resonance V", *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 V; and $\sigma_\lambda = 0, \pm 1$.

Therefore not only $|\vec{\omega}\cdot\vec{\nu}^0_\lambda| \ge 2^{n+3}$ (because $\vec{\nu}^0_\lambda$ is a sum of $\le \overline{N}_n$ node modes, so that $|\vec{\nu}^0_\lambda| \le N\overline{N}_n$) but $\vec{\omega}\cdot\vec{\nu}^0_\lambda$ is "in the middle" of the diadic interval containing it and by (7.1) does not get out of it if we add a quantity bounded by 2^{n+1} (like $\sigma_\lambda \vec{\omega}\cdot\vec{\nu}_{\lambda_V}$). Hence no branch changes scale as ϑ varies in $\mathcal{F}(\vartheta^1)$, if $\vec{\omega}_0$ verifies (7.1).

This implies, by the strong diophantine hypothesis on $\vec{\omega}_0$, (7.1), that the resonant clusters of the trees in $\mathcal{F}(\vartheta^1)$ all contain the same sets of branches, and the same branches go in or out of each resonance (although they are attached to generally distinct nodes inside the resonances: the identity of the branches is here defined by the number label that each of them carries in ϑ^1). Furthermore the resonance scales and the scales of the resonant clusters, and of all the branches, do not change.

Let ϑ^2 be a tree not in $\mathcal{F}(\vartheta^1)$ and construct $\mathcal{F}(\vartheta^2)$, *etc.* We define a collection $\{\mathcal{F}(\vartheta^i)\}_{i=1,2,\ldots}$ of pairwise disjoint families of trees. We shall sum all the contributions to $\vec{h}^{(k)}$ coming from the individual members of each family. This is the *Eliasson's resummation.*

We call ε_V the quantity $\vec{\omega}\cdot\vec{\nu}_{\lambda_V}$ associated with the resonance V. If λ is a branch with both extremes in $\tilde{V}$ we can imagine to write the quantity $\vec{\omega}\cdot\vec{\nu}_\lambda$ as $\vec{\omega}\cdot\vec{\nu}^0_\lambda + \sigma_V\varepsilon_V$, with $\sigma_V = 0, \pm 1$. Since $|\vec{\omega}\cdot\vec{\nu}_\lambda| > 2^{n_V-1}$ we see that the product of the propagators is holomorphic in ε_V for $|\varepsilon_V| < 2^{n_V-3}$. While ε_V varies in such complex disk the quantity $|\vec{\omega}\cdot\vec{\nu}_\lambda|$ does not become smaller than $2^{n_V-1} - 2^{n_V-3} \ge 2^{n_V-2}$. Note the main point here: the quantity 2^{n_V-3} will usually be $\gg 2^{n_{\lambda_V}-1}$ which is the value ε_V actually can reach in every tree in $\mathcal{F}(\vartheta^1)$; this can be exploited in applying the maximum priciple, as done below.

It follows that, calling n_λ the scale of the branch λ in ϑ^1, each of the $\prod 2M_V \le e^{2k}$ products of propagators of the members of the family $\mathcal{F}(\vartheta^1)$ can be bounded above by $\prod_\lambda 2^{-2(n_\lambda-2)} = 2^{4k}\prod_\lambda 2^{-2n_\lambda}$, if regarded as a function of the quantities $\varepsilon_V = \vec{\omega}\cdot\vec{\nu}_{\lambda_V}$, for $|\varepsilon_V| \le 2^{n_V-3}$, associated with the resonant clusters V. This even holds if the ε_V are regarded as independent complex parameters.

By construction it is clear that the sum of the $\prod 2M_V \le e^{2k}$ terms, giving the contribution to $\vec{h}^{(k)}$ from the trees in $\mathcal{F}(\vartheta^1)$, vanishes to second order in the ε_V parameters (by the approximate cancellation discussed above). Hence by the maximum principle, and recalling that each of the scalar products in (6.23) can be bounded by N^2, we can bound the contribution from the family $\mathcal{F}(\vartheta^1)$ by:

$$\left[\frac{1}{k!}N\Big(\frac{f_0 C_0^2 N^2}{J_0}\Big)^k 2^{4k} e^{2k} \prod_{n\le 0} 2^{-2nN_n}\right]\left[\prod_{n\le 0}\prod_{T,\,n_T=n}\prod_{i=1}^{m_T} 2^{2(n-n_i+3)}\right] \tag{A3.2}$$

where:

1) N_n is the number of propagators of scale n in ϑ^1 ($n = 1$ does not appear as $|\vec{\omega}\cdot\vec{\nu}| \ge 1$ in such cases),

2) the first square bracket is the bound on the product of individual elements in the family $\mathcal{F}(\vartheta^1)$ times the bound e^{2k} on their number,

3) The second term is the part coming from the maximum principle, applied to bound the resummations, and is explained as follows.

i) the dependence on the variables $\varepsilon_{V_i} \equiv \varepsilon_i$ relative to resonances $V_i \subset T$ with scale $n_{\lambda_V} = n$ is holomorphic for for $|\varepsilon_i| < 2^{n_i-3}$ if $n_i \equiv n_{V_i}$, provided $n_i > n+3$ (see above).

ii) the resummation says that the dependence on the ε_i's has a second order zero in each. Hence the maximum principle tells us that we can improve the bound given by the first factor in (A3.2) by the product of factors $(|\varepsilon_i|\, 2^{-n_i+3})^2$ if $n_i > n+3$. If $n_i \le n+3$ we cannot gain anything: but since the contribution to the bound from such terms in (A3.2) 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 $\ll 1$).

Hence substituting (A3.1) into (A3.2) we see that the m_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 $+m_T$ in (A3.1) taken from the factors with $T = V_i$, (note that there are always enough -1's).

Hence the product (A3.2) is bounded by:

$$\frac{1}{k!} N\,(C_0^2 J_0^{-1} f_0 N^2)^k e^{2k} 2^{4k} 2^{6k} \prod_n 2^{-4nkE^{-1}\,2^{\varepsilon n}} \le \frac{1}{k!} N\, B_0^k \tag{A3.3}$$

with: $B_0 = 2^{10} e^2 C_0^2 f_0 N^2 \exp[N\,(2^{2+2\tau^{-1}} \log 2) \sum_{p=1}^{\infty} p 2^{-p\tau^{-1}}]$.

To sum over the trees we note that, fixed ϑ the collection of clusters is fixed. Therefore we only have to multiply (A3.3) by the number of tree shapes for ϑ, $(\le 2^{2k} k!)$, by the number of ways of attaching mode labels, $(\le (3N)^{lk})$, so that we can bound $|h^{(k)}_{\vec\nu j}|$ by an exponential of k and (1.4) follows.

Remarks.

The strong diophantine condition is quite unpleasant as it seems to put an extra requirement on $\vec\omega_0$: I think that in fact such condition is not necessary: I explain why in [G3]. The basic reason is that one is not forced to introduce the scales 2^n in an exact geometric growth: an approximate one, in which 2^n is replaced by γ_n and $1 \le \gamma_n 2^{-n} \le 2$ would be enough. This gives much more freedom in fulfilling the (7.1), given $\vec\omega_0$ verifying (1.3) only. Such a squence can be shown to exist always if $\vec\omega_0$ verifies (1.3), possibly replacing C_0 by a larger constant (to fulfill the analogue of (7.1)).

Appendix A4: Resonant Siegel-Brjuno bound.

Calling N_n^* the number of non resonant lines carrying a scale label $\le n$. We shall prove first that $N_n^* \le 2k(E2^{-\varepsilon n})^{-1} - 1$ if $N_n^* > 0$.

If ϑ has the root line with scale $> n$ then calling $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ the subdiagrams of ϑ emerging from the first vertex of ϑ and with $k_j > E\,2^{-\varepsilon n}$ lines, it is $N^*(\vartheta) = N^*(\vartheta_1) + \ldots + N^*(\vartheta_m)$ and the statement is inductively implied from its validity for $k' < k$ provided it is true that $N^*(\vartheta) = 0$ if $k < E2^{-\varepsilon n}$, which is is certainly the case if E is chosen as in (A3.1).*

In the other case it is $N_n^* \le 1 + \sum_{i=1}^m N^*(\vartheta_i)$, and if $m = 0$ the statement is trivial, or if $m \ge 2$ the statement is again inductively implied by its validity for $k' < k$.

If $m = 1$ we once more have a trivial case unless the order k_1 of ϑ_1 is $k_1 > k - \frac{1}{2} E\,2^{-n\varepsilon}$. Finally, and this is the real problem as the analysis of a few examples shows, we claim that in the latter case the root line is either a resonance or it has scale $> n$.

Accepting the last statement it will be: $N^*(\vartheta) = 1 + N^*(\vartheta_1) = 1 + N^*(\vartheta'_1) + \ldots + N^*(\vartheta'_{m'})$, with ϑ'_j being the m' subdiagrams emerging from the first node of ϑ'_1 with orders $k'_j > E\,2^{-\varepsilon n}$: this is so because the root line of ϑ_1 will not contribute its unit to $N^*(\vartheta_1)$. Going once more through the analysis the only non trivial case is if $m' = 1$ and in that case $N(\vartheta'_1) = N^*(\vartheta''_1) + \ldots + N(\vartheta''_{m''})$, *etc*, until we reach a trivial case or a diagram of order $\le k - \frac{1}{2} E\,2^{-n\varepsilon}$.

It remains to check that if $k_1 > k - \frac{1}{2} E\,2^{-n\varepsilon}$ then the root line of ϑ_1 has scale $> n$, unless it is entering a resonance.

Suppose that the root line of ϑ_1 is not entering a resonance. Note that $|\vec\omega \cdot \vec\nu(v_0)| \le 2^n, |\vec\omega \cdot \vec\nu(v)| \le 2^n$, if v_0, v_1 are the first vertices of ϑ and ϑ_1 respectively. Hence $\delta \equiv |(\vec\omega \cdot (\vec\nu(v_0) - \vec\nu(v_1))| \le 2\,2^n$ and the diophantine assumption implies that $|\vec\nu(v_0) - \vec\nu(v_1)| > (2\,2^n)^{-\tau^{-1}}$, or $\vec\nu(v_0) = \vec\nu(v_1)$. The latter case being

* Note that if $k \le E\,2^{-n\varepsilon}$ it is, for all momenta $\vec\nu$ of the lines, $|\vec\nu| \le NE\,2^{-n\varepsilon}$, *i.e.* $|\vec\omega \cdot \vec\nu| \ge (NE\,2^{-n\varepsilon})^{-\tau} = 2^3\,2^n$ so that there are *no* clusters T with $n_T = n$ and $N^* = 0$. The choice $E = N^{-1}2^{-3\varepsilon}$ 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 resonance concept.

discarded as $k - k_1 < \frac{1}{2}E\,2^{-n\varepsilon}$ (and we are not considering the resonances), it follows that $k - k_1 < \frac{1}{2}E\,2^{-n\varepsilon}$ is inconsistent: it would in fact imply that $\vec{\nu}(v_0) - \vec{\nu}(v_1)$ is a sum of $k - k_1$ vertex modes and therefore $|\vec{\nu}(v_0) - \vec{\nu}(v_1)| < \frac{1}{2}NE\,2^{-n\varepsilon}$ hence $\delta > 2^3\,2^n$ which is contradictory with the above opposite inequality.

A similar, far easier, induction can be used to prove that if $N_n^* > 0$ then the number p of clusters of scale n verifies the bound $p < 2k\,(E2^{-\varepsilon n})^{-1} - 1$. Thus (11) is proved.

Remark: the above argument is a minor adaptation of Brjiuno's proof of Siegel's theorem, as remarkably exposed by Pöschel, [P].

Acknowledgements: partial support from NSF grant # DMR 89-18903, from Rutgers University and Institut des Hautes Etudes Scientifiques (IHES). The work was also partly supported by Ministero della Ricerca (fondi 40%). I am indebted to L. Chierchia for many suggestions and for critical readings of early versions; but mainly for showing me, prior to publication, his "tree root identity" which is the key idea for the homoclinic splitting theory, and without which the part of the paper on the homoclinic splitting would have been left incomplete, and at best with a strange looking conjecture. It is with gratitude that I thank H. Epstein and S. Miracle–solé for criticism, suggestions and encouragement; I am also indebted to J. Bricmont and M. Vittot formany clarifying discussions.

References

[A] Arnold, V.: *Instability of dynamical systems with several degrees of freedom*, Sov. Mathematical Dokl., 5, 581-585, 1966.

[A2] Arnold, V.: *Proof of a A.N. Kolmogorov theorem on conservation of conditionally periodic motions under small perturbations of the hamiltonian function*, Uspeki Matematicheskii Nauk, 18, 13– 40, 1963.

[ACKR] Amick C., Ching E.S.C., Kadanoff L.P., Rom–Kedar V.: *Beyond All Orders: Singular Perturbations in a Mapping* J. Nonlinear Sci. 2, 9–67, 1992.

[BG] Benettin, G., Gallavotti, G.: *Stability of motions near resonances in quasi-integrable hamiltonian systems*, J. Statistical Physics, 44, 293-338, 1986.

[BfG] Benfatto, G., Gallavotti, G.: *Perturbation theory of the Fermi surface in a quantum liquid. A general quasi particle formalism and one dimensional systems*, Journal of Statistical Physics, 59, 541- 664, 1990.

[CG] Chierchia, L., Gallavotti, G.: *Drift and diffusion in phase space*, in mp_arc, # 92-92. This paper is deposited in the archive mp_arc: to get a TeX version send an empty E-mail message to `mp_arc@math.utexas.edu`: instructions will be sent back. In print in Annales de l'Institut H. Poincarè.

[CZ] Chierchia, L., Zehnder, E.: *Asymptotic expansions of quasi-periodic motions*, Annali della Scuola Normale Superiore di Pisa, Serie IV Vol XVI Fasc.2 (1989).

[DS] Delshams, A., Seara, M.T.: *An asymptotic expression for the splitting of separatrices of rapidly forced pendulum*, preprint 1991.

[E] Eliasson L. H.: *Absolutely convergent series expansions for quasi–periodic motions*, report 2–88, Dept. of Math., University of Stockholm, 1988.

[FG] Felder, G., Gallavotti, G.: *Perturbation theory and non renormalizable scalar fields*, Communications in Mathematical Physics, 102, 549-571, 1986.

[G1] Gallavotti, G.: *The elements of Mechanics*, Springer, 1983.

[G2] Gallavotti, G.: *Renormalization theory and ultraviolet stability for scalar fields via renormalization group methods*, Reviews in Modern Physics, 57, 471- 572, 1985. See also, Gallavotti, G.: *Quasi integrable mechanical systems*, Les Houches, XLIII (1984), vol. II, p. 539– 624, Ed. K. Osterwalder, R. Stora, North Holland, 1986.

[Ge] Gelfreich, V.: *Separatrices splitting for the rapidly forced pendulum* preprint 1992.

[GLT] Gelfreich, V. G., Lazutkin, V.F., Tabanov, M.B.:*Exponentially small splitting in Hamiltonian systems*, Chaos, 1 (2), 1991.

[H] Harary, F., Palmer, E.: *Graphical enumeration*, Academic Press, 1973, New York.

[HMS] Holmes, P., Marsden, J., Scheurle,J: *Exponentially Small Splittings of Separatrices in KAM Theory and Degenerate Bifurcations*, Preprint, 1989.

[K] Kolmogorov, N.: *On the preservation of conditionally periodic motions*, Doklady Akademia Nauk SSSR, 96, 527– 530, 1954. See also: Benettin, G., Galgani, L., Giorgilli, A., Strelcyn, J.M.: *A proof of Kolmogorov theorem on invariant tori using canonical transormations defined by the Lie method*, Nuovo Cimento, 79B, 201– 223, 1984.

[La] Lazutkin, V.F.: *Separatrices splitting for standard and semistandard mappings*, Preprint, 1989.
[M] Moser, J.: *On invariant curves of an area preserving mapping of the annulus*, Nachrichten Akademie Wiss. Göttingen, 11, 1– 20, 1962.
[N] Nekhorossev, N.: An exponential estimate of the time of stability of nearly integrable hamiltonian systems, Russian Mathematical Surveys, 32, 1-65, 1975.
[Nei] Neihstad, A.I.: *The separation of motions in systems with rapidly rotating phase*, Prikladnaja Matematika i Mekhanika **48**, 133–139, 1984 (Translation in Journal of Applied Mathematics and Mechanics).
[P] Pöschel, J.: *Invariant manifolds of complex analytic mappings*, Les Houches, XLIII (1984), vol. II, p. 949–964, Ed. K. Osterwalder, R. Stora, North Holland, 1986.
[S] Siegel, K.: *Iterations of analytic functions*, Annals of Mathematics, **43**, 607– 612, 1943.
[T] Thirring, W.: *Course in Mathematical Physics*, vol. 1, p. 133, Springer, Wien, 1983.
[V] Vittot, M.: *Lindstedt perturbation series in hamiltonian mechanics: explicit formulation via a multidimensional Burmann–Lagrange formula*, Preprint CNRS–Luminy 1992.