Lindstedt series for perturbations of isochronous systems. A review of the general theory

Michele Bartuccelli * and Guido Gentile †

* Department of Mathematics and Statistics, University of Surrey, Guildford, GU2 5XH, UK

† Dipartimento di Matematica, Università di Roma Tre, Roma, I-00146, Italy

ABSTRACT. *We give a proof of the persistence of invariant tori for analytic perturbations of isochronous systems by using the Lindstedt series expansion for the solutions of the equations of motion. With respect to the case of anisochronous systems, there is the additional problem of finding the set of allowed rotation vectors, because they cannot be given a priori simply by looking at the unperturbed system. By considering the involved parameters (size of the perturbation, rotation vector and average action of a persisting invariant torus) as independent parameters we can introduce a function which is analytic in such parameters and only when the latter satisfy some constraint it becomes a solution: this can be regarded as a sort of singular implicit function problem. Therefore, although the dependence of the parameters, hence of the solution, upon the size of the perturbation is not smooth, in this way we construct explicitly the solution by using an absolutely convergent power series.*

1. Introduction

1.1. *Lindstedt series and KAM theorem.* The KAM theorem assures the persistence of a large number of invariant tori under perturbations of integrable systems. For analytic Hamiltonians *a posteriori* one can consider the equations of motion and look directly for analytic quasi-periodic solutions, by writing them as formal power series, *Lindstedt series*, in the size of the perturbation: when such solutions exist, the series representing them must converge. This is a quite natural approach, and in fact it was the first to be attempted, for instance by Poincaré, [P], who, however, doubted that, in general, the series could converge. The problem was then solved by Kolmogorov, [K], and it gave rise to a large amount of literature about what has became known as KAM theory. Exponential bounds on the coefficients of the Lindstedt series are obtained in the proof from the analysis of an implicit function problem, but the mechanism which remained unclear was how the single terms arising in the perturbative expansion of the Lindstedt series and separately growing much faster than exponentially still admit an exponential bound when summed together, [M], and the problem was referred to (improperly) as the "problem of a direct proof of KAM theorem". This was the state of the art until very recently, when the problem was solved by Eliasson, [E], for the anisochronous case. In fact in such a case the non-degeneracy condition for the free Hamiltonian allows us to construct for the perturbed system an invariant torus run with a rotation vector chosen among those of the free system: more precisely, if, with obvious notations, $\{\boldsymbol{\alpha}(t) = \boldsymbol{\omega} t, \mathbf{A}(t) = \mathbf{A}_0\}$ is an orbit on an invariant torus for the integrable anisochronous Hamiltonian $\mathcal{H}_0(\mathbf{A})$, one considers the perturbed system $\mathcal{H}(\boldsymbol{\alpha}, \mathbf{A}) = \mathcal{H}_0(\mathbf{A}) + f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon)$, with $f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon) = O(\varepsilon)$, and, for $|\varepsilon| < \varepsilon_0$, with ε_0 small enough, and $\boldsymbol{\omega}$ satisfying a Diophantine condition, one looks for a quasi-periodic solution of the equations of motion which has the same rotation vector $\boldsymbol{\omega}$.

In the isochronous case an approach of this kind is not so straightforward. In fact for $\mathcal{H}_0(\mathbf{A}) = \boldsymbol{\omega} \cdot \mathbf{A}$ all tori have the same rotation vector $\boldsymbol{\omega}$, while for the perturbed system $\mathcal{H}(\boldsymbol{\alpha}, \mathbf{A}) = \boldsymbol{\omega} \cdot \mathbf{A} +$

* E-mail address: m.bartuccelli@eim.surrey.ac.uk

† E-mail address: gentile@mat.uniroma3.it

$f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon)$, with $f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon) = O(\varepsilon)$, the KAM theorem states the existence of invariant tori for $|\varepsilon| < \varepsilon_0$, with ε_0 small enough, but with rotation vectors different from that of the unperturbed system: in addition the latter depend (in general) on ε and the dependence is not a smooth one. So it is not clear at all how to extend the Lindstedt series, mostly because there is no hope to obtain an analytic dependence on ε.

Here we address the just described question. We prefer to consider a particular case, rather simplified with respect to the most general one can conceive, but still retaining the most important features of the general case. We consider a two-dimensional case, in which the frequency of one harmonic oscillator is fixed once and for all (*i.e.* it is simply a clock). We follow the spirit of the approach by Gallavotti, [G2], where Eliasson's work was revisited by considering a simplified model (*Thirring model*), in order to separate the general strategy of the proof from the technical intricacies which would make it less terse. Moreover such a case is already interesting for physical applications, [BGG], for example in the study of the stability of the upside-down pendulum whose point of support is subjected to a fast oscillation in the vertical direction. Extensions to more general systems will be discussed at the end (see also §1.10 below).

1.2. *The model.* For $m \in \mathbb{N}$, for $\overline{\mathbf{x}} \in \mathbb{R}^m$ and for $\overline{A}_1 \in \mathbb{C}$ define the following domains:

$$\begin{aligned} \Sigma_\kappa &= \{\boldsymbol{\alpha} \in \mathbb{C}^2 \,:\, \operatorname{Re}\alpha_j \in \mathbb{T}, \quad |\operatorname{Im}\alpha_j| < \kappa, j = 1,2\}, \\ \mathcal{B}_r(\overline{\mathbf{x}}) &= \{\mathbf{x} \in \mathbb{C}^m \,:\, |\mathbf{x} - \overline{\mathbf{x}}| < r\}, \\ \mathcal{A}_\rho(\overline{A}_1) &= \{\mathbf{A} \in \mathbb{C}^2 \,:\, |A_1 - \overline{A}_1| < \rho\}. \end{aligned} \tag{1.1}$$

Consider the system described by the Hamiltonian

$$\mathcal{H} = \boldsymbol{\omega} \cdot \mathbf{A} + f(\boldsymbol{\alpha}, A_1, \varepsilon), \tag{1.2}$$

where $\mathbf{A} = (A_1, A_2) \in \mathbb{R}^2$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{T}^2$ are conjugate variables, $\boldsymbol{\omega} = (\mu, 1)$, with $\mu \in (0,1)$, $\cdot$ denotes the inner product in $\mathbb{R}^2$ and f is a function real analytic in the variables $\boldsymbol{\alpha}$, A_1 and ε, and holomorphic in a complex domain

$$\mathcal{D} = \Sigma_\kappa \times D \times \mathcal{B}_{\varepsilon_1}(0), \tag{1.3}$$

with $D \subset \mathbb{C}$ an open subset, and such that $f = 0$ for $\varepsilon = 0$; so one can write in (1.3)

$$f(\boldsymbol{\alpha}, A_1, \varepsilon) = \sum_{k=1}^{\infty} \varepsilon^k f^{(k)}(\boldsymbol{\alpha}, A_1). \tag{1.4}$$

The system (1.2) represents two harmonic oscillators interacting through a potential depending only on the angles and on the action variable A_1. The corresponding equations of motion are

$$\begin{cases} \dot{\alpha}_1 = \mu + \partial_{A_1} f, \\ \dot{\alpha}_2 = 1, \\ \dot{A}_1 = -\partial_{\alpha_1} f, \\ \dot{A}_2 = -\partial_{\alpha_2} f, \end{cases} \tag{1.5}$$

so that the angle α_2 rotates with constant angular velocity, *i.e.* $\alpha_2(t) = t$. Here and henceforth ∂_x denotes the partial derivative with respect to x: if $\mathbf{x} = (x_1, x_2)$ then $\partial_{\mathbf{x}} = (\partial_{x_1}, \partial_{x_2})$.

1.3. *Results.* Assume that $\boldsymbol{\omega}$ in (1.2) satisfies the *Diophantine condition*

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > C|\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{1.6}$$

with Diophantine constants $C > 0$ and $\tau > 1$. We shall prove the following results:
(1) if $\boldsymbol{\omega}_0$ is a rotation vector close enough to $\boldsymbol{\omega}$ and with comparable Diophantine properties (*i.e.* with Diophantine constants $C_0 = bC$ and τ, for some constant $b \in (0,1)$), then for all $\mathbf{A}_0$

close enough to a prefixed $\overline{\mathbf{A}}$, one can fix a value ε for which there exists an invariant torus with rotation vector $\boldsymbol{\omega}_0$ and average action $\mathbf{A}_0$ for the Hamiltonian with perturbative parameter *suitably fixed* (close to 0) *and depending analytically on* $\mathbf{A}_0$;
(2) instead of fixing the average action, we can fix the value of the perturbative parameter ε and look for the invariant tori persisting under perturbation: we have that infinitely many of them persist, with rotation vectors $\boldsymbol{\omega}_0$ close enough to $\boldsymbol{\omega}$ and with average actions depending on $\boldsymbol{\omega}_0$.

The first result is peculiar to isochronous systems, while the second one is quite analogous to the anischronous KAM theorem and it is the usual form in which it is stated in the literature, up to a quantitative characterization of the "infinitely many tori" and an estimate of the relative measure of the points in phase space lying on invariant tori. If we allow a much weaker Diophantine condition, that is we let b to be a power of ε, such a measure can be shown to tend to 1 for $\varepsilon \to 0$; we shall can back to such a problem in §6 (see also §1.10 below).

We can state more formally the above results as follows.

1.4. Theorem. *Fix* $\overline{\mathbf{A}} = (\overline{A}_1, \overline{A}_2)$*, with* $\overline{A}_1 \in D$*, and* $\rho > 0$ *such that* $\mathcal{B}_\rho(\overline{A}_1) \subset D$*. Consider the equations of motions (1.5), corresponding to the Hamiltonian (1.2), with* $\boldsymbol{\omega} = (\mu, 1)$ *satisfying the Diophantine condition (1.6), and suppose that*

$$\int_{\mathbb{T}^2} \mathrm{d}\boldsymbol{\alpha}\, \partial_{A_1} f^{(1)}(\boldsymbol{\alpha}, A_1) \neq 0 \qquad \forall A_1 \in \mathcal{B}_\rho(\overline{A}_1). \tag{1.7}$$

There is a universal constant $b \in (0,1)$ *and three* b*-dependent constants* $a > 0$, $\rho' \in (0,\rho)$ *and* $\kappa' \in (0,\kappa)$ *such that for all* $\mu_0 \in (\mu - aC, \mu + aC)$*, with* $\boldsymbol{\omega}_0 = (\mu_0, 1)$ *satisfying the Diophantine condition*

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \qquad C_0 = bC, \tag{1.8}$$

and for all $\mathbf{A}_0 \in \mathcal{A}_{\rho'}(\overline{A}_1)$*, there is a value* $\varepsilon = \varepsilon(\mu_0, \mathbf{A}_0) \in \mathcal{B}_{\varepsilon_1}(0)$*, depending analytically on* $\mathbf{A}_0 \in \mathcal{A}_{\rho'}(\overline{A}_1)$*, and two functions* $\overline{\mathbf{h}}(\boldsymbol{\psi}, \mathbf{A}_0, \mu_0)$ *and* $\overline{\mathbf{H}}(\boldsymbol{\psi}, \mathbf{A}_0, \mu_0)$*, analytic in* $(\boldsymbol{\psi}, \mathbf{A}_0) \in \Sigma_{\kappa'} \times \mathcal{A}_{\rho'}(\overline{A}_1)$ *and with zero* $\boldsymbol{\psi}$*-average, such that*

$$\begin{cases} \boldsymbol{\alpha}(t) = \boldsymbol{\omega}_0 t + \overline{\mathbf{h}}(\boldsymbol{\omega}_0 t, \mathbf{A}_0, \mu_0), \\ \mathbf{A}(t) = \mathbf{A}_0 + \overline{\mathbf{H}}(\boldsymbol{\omega}_0 t, \mathbf{A}_0, \mu_0) \end{cases} \tag{1.9}$$

is a solution of (1.5). The constant a *depends on* b*, but it is independent of* C*.*

1.5. *Remarks.* (1) The condition (1.7) is not really needed in order to prove the theorem: it is imposed just for simplicity, but it could be considerably weakened. See also the remark 2.13 in sect. 2.

(2) Since the function $\mathbf{H}$ has zero average, the vector $\mathbf{A}_0$ represents the average (over time) of the action variable for the quasi-periodic motion with rotation vector $\boldsymbol{\omega}_0$: this means that we are looking for an invariant torus whose average action equals that of an unperturbed Diophantine one.

1.6. Theorem. *Fix* $\overline{\mathbf{A}} = (\overline{A}_1, \overline{A}_2)$*, with* $\overline{A}_1 \in D$*. Consider the equations of motions (1.5) corresponding to the Hamiltonian (1.2), with* $\boldsymbol{\omega} = (\mu, 1)$ *satisfying the Diophantine condition*

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > C |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{1.10}$$

with Diophantine constants $C > 0$ *and* $\tau > 1$*, and suppose that*

$$M \equiv \int_{\mathbb{T}^2} \mathrm{d}\boldsymbol{\alpha}\, \partial^2_{A_1} f^{(1)}(\boldsymbol{\alpha}, \overline{A}_1) \neq 0. \tag{1.11}$$

There is a universal constant $b \in (0,1)$ *and two* b*-dependent constants* $\overline{\varepsilon} \in (0, \varepsilon_1)$ *and* $\kappa' \in (0,\kappa)$ *such that for all* $\varepsilon \in \mathcal{B}_{\overline{\varepsilon}}(0) \setminus \{0\}$ *there are a constant* $a > 0$*, infinitely many* $\mu_0 \in (\mu - a\varepsilon, \mu + a\varepsilon)$ *with* $\boldsymbol{\omega}_0 = (\mu_0, 1)$ *satisfying the Diophantine condition*

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \qquad C_0 = bC, \tag{1.12}$$

and pairs of functions $\mathbf{h}^*(\boldsymbol{\psi},\varepsilon,\mu_0)$ *and* $\mathbf{H}^*(\boldsymbol{\psi},\varepsilon,\mu_0)$*, analytic for* $\boldsymbol{\psi}\in\Sigma_{\kappa'}$ *and with zero* $\boldsymbol{\psi}$*-average, and a vector* $\mathbf{A}^*(\varepsilon,\mu_0)$ *such that*

$$\begin{cases}\boldsymbol{\alpha}(t)=\boldsymbol{\omega}_0 t+\mathbf{h}^*(\boldsymbol{\omega}_0 t,\varepsilon,\mu_0),\\ \mathbf{A}(t)=\mathbf{A}^*(\varepsilon,\mu_0)+\mathbf{H}^*(\boldsymbol{\omega}_0 t,\varepsilon,\mu_0)\end{cases} \tag{1.13}$$

is a solution of (1.5). One has $A_2^*(\varepsilon,\mu_0)=\overline{A}_2$ *and* $A_1^*(\varepsilon,\mu_0)\in D$*; the constant* a *depends on* b*, but it is independent of* ε*.*

1.7. *Remark.* (1) The condition (1.11) could be considerably weakened. See also the remark 1.5, (1), and the remark 2.15 in sect. 2.

(2) We are fixing ε and $\boldsymbol{\omega}_0$ and we want to detect an invariant torus with rotation vector $\boldsymbol{\omega}_0$: to achieve such a goal we are forced to change (with respect to the unperturbed system) the average action into a value $\mathbf{A}^*(\varepsilon,\mu_0)$: this is the meaning of the vector $\mathbf{A}^*(\varepsilon,\mu_0)$ appearing in the statement of the theorem 1.6.

1.8. *Idea of the proof.* For $\varepsilon=0$ and for all $\mathbf{A}_0\in\mathbb{R}^2$ there is a solution

$$\{\boldsymbol{\alpha}(t)=\boldsymbol{\omega} t,\mathbf{A}(t)=\mathbf{A}_0\}, \tag{1.14}$$

lying on an invariant torus.

Because of the isochrony of the unperturbed system we are not able to fix *a priori* the rotation vectors $\boldsymbol{\omega}_0$ of the quasi-periodic solutions for $\varepsilon\neq 0$. So we proceed by adopting the splitting

$$\mu_0+\eta(\mathbf{A}_0,\varepsilon,\mu_0)=\mu, \tag{1.15}$$

where $\mathbf{A}_0$ is the same as in (1.14) and, to begin, μ_0 is fixed in a rather arbitrary way. The above prescription leads to the system of equations

$$\begin{cases}\dot\alpha_1=\mu_0+\partial_{A_1}f+\eta(\mathbf{A}_0,\varepsilon,\mu_0),\\ \dot\alpha_2=1,\\ \dot A_1=-\partial_{\alpha_1}f,\\ \dot A_2=-\partial_{\alpha_2}f,\end{cases} \tag{1.16}$$

hence, by ignoring the constraint given by equation (1.15), we look for a quasi-periodic solution of the modified system (1.16), where $\eta(\mathbf{A}_0,\varepsilon,\mu_0)$ is to be determined, with rotation vector $\boldsymbol{\omega}_0$, *i.e.* for a solution of the form

$$\begin{cases}\boldsymbol{\alpha}(t)=\boldsymbol{\omega}_0 t+\mathbf{h}(\boldsymbol{\omega}_0 t,\mathbf{A}_0,\varepsilon,\mu_0),\\ \mathbf{A}(t)=\mathbf{A}_0+\mathbf{H}(\boldsymbol{\omega}_0 t,\mathbf{A}_0,\varepsilon,\mu_0),\end{cases} \tag{1.17}$$

where $\boldsymbol{\omega}_0=(\mu_0,1)$. The functions $\mathbf{h}$ and $\mathbf{H}$ are called the *conjugating functions*, while η is called the *counterterm.*

If $\eta(\mathbf{A}_0,\varepsilon,\mu_0)$ is replaced with 0 in (1.15), so that $\boldsymbol{\omega}_0$ becomes $\boldsymbol{\omega}$, it is well known from KAM theory that there are not (in general) quasi-periodic solutions with rotation vector $\boldsymbol{\omega}$; on the other hand, as it is defined, $\eta(\mathbf{A}_0,\varepsilon,\mu_0)$ depends only on μ_0 and it is not obvious how μ_0 has to be chosen nor if it can be chosen at all so that (1.15) is satisfied. Note that, if for some μ_0 (1.15) is verified, then a solution of the true equations of motions (1.5), quasiperiodic with rotation vector $\boldsymbol{\omega}_0=(\mu_0,1)$, will have been found.

We shall show that, by neglecting the constraint (1.15), for fixed μ_0 it will be possible to choose in a unique way $\eta(\mathbf{A}_0,\varepsilon,\mu_0)$ as an analytic function of ε and of $\mathbf{A}_0$ in such a way that there exists a solution of (1.16) the form (1.17) with zero average: this will be the content of the lemma 2.2 in sect. 2. We note that as long as the constraint (1.15) is neglected the solutions (1.17) of the equation of motion (and the corresponding counterterm) are analytic in ε as well in $\mathbf{A}_0$: so we shall use the powerful machinery of the Lindstedt series for the analytic KAM theory also in a case in which the solution of the original equations of motion cannot be expected to be analytic.

The idea of introducing suitable counterterms in the equations of motion in order to make them soluble is not new, and dates back to [M], where it was successfully used in order to give (among other things) a proof of the KAM theorem. Also in [Ge] (see also [GGM]) the existence of whiskers for hyperbolic tori (of codimension one) was proved for a class of almost integrable systems by studying a modified Hamiltonian obtained by introducing suitable counterterms. Of course one had to prove eventually that the original Hamiltonian could be recovered: in [Ge] the proof of such an assertion was a simple application of the implicit function theorem.

Also in the present case, because of the introduction of the counterterm, we study a different system of equations, and, since the counterterm is uniquely determined, in general there is no hope to obtain, for ε, $\mathbf{A}_0$ and μ_0 arbitrarily chosen, that such a counterterm satisfies (1.15). In our case (1.15) is not quite an implicit function problem, because the dependence of $\eta(\mathbf{A}_0, \varepsilon, \mu_0)$ on μ_0 is not even smooth (the counterterm is defined on a Cantor-like set, as far as the dependence on μ_0 is concerned). Nevertheless (1.15), seen either as an equation for ε for fixed $\mathbf{A}_0$ and μ_0 or as an equation for $\mathbf{A}_0$ for fixed ε and μ_0, can be solved: this will lead, respectively, to the theorem 1.4 and to the theorem 1.6, as it will be shown in sect. 2.

This means that we shall look for solutions of the form (1.17) where ε, $\mathbf{A}_0$ and μ_0 are not independent of each other: we can choose ε as a function of μ_0 and $\mathbf{A}_0$, in such a way that the constraint (1.15) is satisfied, so that we can write $\mathbf{h} = \overline{\mathbf{h}}(\boldsymbol{\psi}, \mathbf{A}_0, \mu_0) = \mathbf{h}(\boldsymbol{\psi}, \mathbf{A}_0, \varepsilon(\mu_0, \mathbf{A}_0), \mu_0)$, and analogous expressions hold for the other quantities $\mathbf{H}$ and η. Then, fixed μ_0 and $\mathbf{A}_0$, ε is not a free parameter, and in principle we could not study the dependence of $\mathbf{h}$ in ε by varying ε without changing μ_0 and $\mathbf{A}_0$. Nevertheless, considering ε and $(\mathbf{A}_0, \mu_0)$ as independent parameters, then $\mathbf{h}(\boldsymbol{\psi}, \mathbf{A}_0, \varepsilon, \mu_0)$ turns out to be analytic in ε, and we shall see that we can use the analyticity of such a dependence in order to write the solution as a power series in ε.

Analogously we can fix ε and μ_0 (in an appropriate way) and choose $\mathbf{A}_0 = \mathbf{A}^*(\varepsilon, \mu_0)$ as a function of both parameters, again in such a way that the constraint (1.15) is satisfied, so that we can write the solution as $\mathbf{h} = \mathbf{h}^*(\boldsymbol{\psi}, \varepsilon, \mu_0) = \mathbf{h}(\boldsymbol{\psi}, \mathbf{A}^*(\varepsilon, \mu_0), \varepsilon, \mu_0)$, and the same can be done for $\mathbf{H}$ and η. Note that in such a case even if $\mathbf{h}(\boldsymbol{\psi}, \mathbf{A}_0, \varepsilon, \mu)$ is analytic in ε, the solution $\mathbf{h}^*(\boldsymbol{\psi}, \varepsilon, \mu_0)$ is not, as $\mathbf{A}^*(\varepsilon, \mu_0)$ does not depend analytically on ε.

1.9. *Comments about the statement of the theorems.* The theorem 1.4 deals with the problem of fixing a rotation vector $\boldsymbol{\omega}_0 = (\mu_0, 1)$, with μ_0 close to μ, and looking for a value $\varepsilon = \varepsilon(\mu_0, \mathbf{A}_0)$ such that the Hamiltonian (1.2) admits an invariant torus run with rotation vector $\boldsymbol{\omega}_0$.

Such a problem is of physical relevance, in studying the stability and the persistence of KAM tori near elliptic equilibrium points; applications are discussed in [BGG].

On the other hand one could also consider the following (different) problem: set $\mathcal{B}_\rho(\mathbf{A}_0) = \{\mathbf{A} : |\mathbf{A} - \mathbf{A}_0| < \rho\} \subset D \times \mathbb{C}^2$, fix ε small enough and look for values μ_0 close to μ and, correspondingly, values $\mathbf{A}^* \in \mathcal{B}_\rho(\mathbf{A}_0)$ such that, for that value of ε, the Hamiltonian (1.2) admits a solution parameterized by the action $\mathbf{A}^*$ (*i.e.* such that the average of the action variables is $\mathbf{A}^*$) and run with rotation vector $\boldsymbol{\omega}_0 = (\mu_0, 1)$. The theorem 1.6 deals exactly with such a problem.

1.10. *Invariant tori for fixed* ε. One can also ask how many tori persist under perturbations, that is, with the same notations as in §1.9, which fraction of phase space in $\mathcal{B}_\rho(\mathbf{A}_0) \times \mathbb{T}^2$, once ε has been fixed to some value (small enough), correspond to invariant tori (run with some rotation vector) persisting under perturbation for that value of ε: the answer is that, if the condition (1.11) of the theorem 1.6 is replaced with a non-degeneracy condition like

$$\inf_{A_1 \in D} \left| \int_{\mathbb{T}^2} \mathrm{d}\boldsymbol{\alpha}\, \partial^2_{A_1} f^{(1)}(\boldsymbol{\alpha}, A_1) \right| > 0, \tag{1.18}$$

then the set of initial data $(\mathbf{A}, \boldsymbol{\alpha}) \in \mathcal{B}_\rho(\mathbf{A}_0) \times \mathbb{T}^2$ for trajectories which lie on invariant tori persisting under perturbation form a set of relative measure tending to 1 as $\varepsilon \to 0$. This can be proved with standard arguments of KAM theory (see for instance [CG] and [Pö1]): it could also be studied by using the same techniques introduced in the present paper; we shall briefly (and informally) discuss such an aspect in sect. 6 below.

1.11. *Comments about the proof of the theorem.* In studying the functions $\mathbf{h}, \mathbf{H}, \eta$ in which ε and μ_0 are seen as independent parameters neither the special form of the interaction nor the fact that the dimension is $d = 2$ play any rôle. So the lemma 2.2 below can be extended (essentially with no change) to any perturbations of Hamiltonian isochronous systems in any dimensions. The notations we shall use will make such an extension trivial: simply interpret the vectors as vectors in $\mathbb{R}^d$, and note (while reading the proof) that the special form of the interaction is not really used; for this reason we shall write $f(\boldsymbol{\alpha}, \mathbf{A}) = f(\boldsymbol{\alpha}, A_1)$, even if in our case the perturbation does not depend explicitly on A_2; ee also the remark 3.5.

On the contrary in order to solve the compatibility condition (1.15), the discussion in sect. 2 applies only to the Hamiltonians of the special form (1.2). The methods extend to the general situation, but some further arguments become necessary; in sect. 6 we briefly discuss how such an extension can be carried out.

1.12. *Contents of the paper.* In sect. 2 we state the main technical result of the paper (lemma 2.2), which deals with the modified system (1.16), and we show how it implies the theorems 1.4 and 1.6. The immediately following sections are devoted to the proof of the lemma 2.2: in sect. 3 we introduce the tree formalism which will be used in sect. 4 to prove that the equations (1.16) are formally soluble, for a suitable choice of the counterterm, and in sect. 5 to prove that the formal series defining the conjugating functions and the counterterm are converging: in particular this will implies that all quantities are analytic in the perturbative parameter ε (the more technical aspects of the proofs will be relegated into the Appendices). Finally in sect. 6 we discuss possible extensions and generalizations of the results, particularly those concerning the problem of studying the measure of the persisting invariant tori in phase space.

For the proof of the lemma 2.2 we could have relied on the existing literature, as [BGGM1] and [GM], and simply outlined the main differences with respect to it. However we preferred to give the proof in full details both because the paper is meant as a review of the techniques (hence a selfcontained discussion makes it more readable) and becuase in this way we profit to try to clarify the graphic construction through a number of examples and pictures; furthermore some technical improvements are presented with respect to the quoted papers.

2. Persistence of invariant tori

2.1. *The modified model.* As outlined in 1.8, for the moment, we study the equations (1.16), where μ_0 is fixed, and $\eta(\mathbf{A}_0, \varepsilon, \mu_0)$ is a function to be determined. Of course this is not the original model, so at the end we shall have the problem to show that the results we find can be fruitfully used in order to draw conclusions also for the model in which the constraint (1.15) is taken into account. We shall prove the following result for the *modified model*, given by (1.16) without the constraint (1.15).

2.2. Lemma. *Fix* $\overline{\mathbf{A}} = (\overline{A}_1, \overline{A}_2)$, *with* $\overline{A}_1 \in D$, *and* $\rho > 0$ *such that* $\mathcal{B}_\rho(\overline{A}_1) \subset D$. *Given the equations of motions*

$$\begin{cases} \dot\alpha_1 = \mu_0 + \partial_{A_1} f, \\ \dot\alpha_2 = 1, \\ \dot A_1 = -\partial_{\alpha_1} f, \\ \dot A_2 = -\partial_{\alpha_2} f, \end{cases} \tag{2.1}$$

with $\boldsymbol{\omega}_0 = (\mu_0, 1)$ *satisfying the Diophantine condition*

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.2}$$

with Diophantine constants $C_0 > 0$ *and* $\tau > 1$, *there exist three constants* $\varepsilon_0 \in (0, \varepsilon_1)$, $\rho' \in (0, \rho)$ *and* $\kappa' \in (0, \kappa)$, *two functions* $\mathbf{h}(\boldsymbol{\psi}, \mathbf{A}, \varepsilon, \mu_0)$ *and* $\mathbf{H}(\boldsymbol{\psi}, \mathbf{A}, \varepsilon, \mu_0)$, *analytic in* $(\boldsymbol{\psi}, \mathbf{A}, \varepsilon)$ *for* $\boldsymbol{\psi} \in \Sigma_{\kappa'}$, $\mathbf{A} \in \mathcal{A}_{\rho'}(A_{01})$ *and* $|\varepsilon| < \varepsilon_0$ *and with vanishing* $\boldsymbol{\psi}$*-average on* $\mathbb{T}^2$, *and a unique function* $\eta(\mathbf{A}, \varepsilon, \mu_0)$, *analytic in* ε *for* $|\varepsilon| < \varepsilon_0$ *and in* $\mathbf{A}$ *for* $|A_1 - \overline{A}_1| < \rho'$, *such that all three functions are vanishing for* $\varepsilon = 0$, *and, for all* $\mathbf{A}_0 \in \mathcal{A}_{\rho'}(\overline{A}_1)$ *and* $|\varepsilon| < \varepsilon_0$,

$$\begin{cases} \boldsymbol{\alpha}(t) = \boldsymbol{\omega}_0 t + \mathbf{h}(\boldsymbol{\omega}_0 t, \mathbf{A}_0, \varepsilon, \mu_0), \\ \mathbf{A}(t) = \mathbf{A}_0 + \mathbf{H}(\boldsymbol{\omega}_0 t, \mathbf{A}_0, \varepsilon, \mu_0) \end{cases} \tag{2.3}$$

is a solution of

$$\begin{cases} \dot\alpha_1 = \mu_0 + \partial_{A_1} f + \eta(\mathbf{A}_0, \varepsilon, \mu_0), \\ \dot\alpha_2 = 1, \\ \dot A_1 = -\partial_{\alpha_1} f, \\ \dot A_2 = -\partial_{\alpha_2} f. \end{cases} \tag{2.4}$$

Moreover one has $\varepsilon_0 = \min\{E_0 C_0, \varepsilon_1\}$ for some constant E_0 independent of μ_0.

2.3. *Proof of the theorem 1.4.* The proof of the lemma 2.2 will be performed in the following sections 3÷5 (and in the Appendices).

Now we come back to the original problem (1.5), and we show how the above lemma can be used in order to prove the theorem 1.4.

The lemma 2.2 shows that, for $\boldsymbol{\omega}_0$ satisfying the Diophantine condition

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.5}$$

the series defining the conjugating functions and the counterterm converge and have a radius of convergence in ε bounded from below by $\varepsilon_0 = E_0 C_0$, for some constant E_0 (assume for simplicity $E_0 C_0 \le \varepsilon_1$). Now we want to use such a property to prove the following result, which immediately yields the theorem 1.4.

2.4. Proposition. *Let $\eta(\mathbf{A}_0, \varepsilon, \mu_0)$ be given by the lemma 2.2, such that the functions (1.17) solve (1.16), and suppose (see (1.7)) that*

$$\partial_{A_1} f^{(1)}_{\mathbf{0}}(\mathbf{A}_0) \equiv \int_{\mathbb{T}^2} \mathrm{d}\boldsymbol{\alpha} \partial_{A_1} f^{(1)}(\boldsymbol{\alpha}, A_{01}) \neq 0. \tag{2.6}$$

Then given $\mu \in (0,1)$ such that $\boldsymbol{\omega} = (\mu, 1)$ satisfies the Diophantine condition

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > C |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.7}$$

with Diophantine constants $C > 0$ and $\tau > 1$, there exists $a > 0$ such that it is possible to fix $\mu_0 \in \mathcal{B}_{aC}(\mu)$ such that $\boldsymbol{\omega}_0 = (\mu_0, 1)$ satisfies the Diophantine condition

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.8}$$

with $C_0 = bC$, for some positive constant b, and to fix $\varepsilon \equiv \varepsilon(\mu_0, \mathbf{A}_0)$, with $|\varepsilon| < \varepsilon_0$, such that

$$\mu = \mu_0 + \eta(\mathbf{A}_0, \varepsilon, \mu_0), \tag{2.9}$$

holds for $\varepsilon = \varepsilon(\mu_0)$.

2.5. *Continued fractions and approximants.* We shall prove the proposition 2.4 through a series of (elementary) lemmata. We need some preliminary notations.

Given $\mu \in (0,1)$ denote by $[a_0, a_1, a_2, \ldots]$ its *continued fraction expansion* and by $\{p_k/q_k\}$ its *best approximants*. Then if $\boldsymbol{\omega} = (\mu, 1)$ and $\boldsymbol{\nu}_k = (q_k, p_k)$ one has, [S],

$$\frac{1}{q_{k+1}} > |\boldsymbol{\omega} \cdot \boldsymbol{\nu}_k| > \frac{1}{2 q_{k+1}}, \tag{2.10}$$

and

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > |\boldsymbol{\omega} \cdot \boldsymbol{\nu}_k| \qquad \forall \boldsymbol{\nu} = (q,p) \text{ such that } q_k < q < q_{k+1}. \tag{2.11}$$

Note also that

$$q_k < |\boldsymbol{\nu}_k| < 2 q_k, \tag{2.12}$$

for all $k \in \mathbb{N}$.

2.6. Lemma. *If $\boldsymbol{\omega} = (\mu, 1)$ satisfies the Diophantine condition*

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > \mathcal{C}_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.13}$$

then

$$\frac{q_{k+1}}{q_k^\tau} < \frac{2^\tau}{\mathcal{C}_0}, \tag{2.14}$$

for any $k \in \mathbb{N}$.

2.7. *Proof of the lemma 2.6.* For any k one has by (2.10) and (2.13)

$$\frac{1}{q_{k+1}} > |\boldsymbol{\omega} \cdot \boldsymbol{\nu}_k| > \mathcal{C}_0 |\boldsymbol{\nu}_k|^{-\tau}, \tag{2.15}$$

so that by (2.12)

$$(2q_k)^\tau > |\boldsymbol{\nu}_k|^\tau > \mathcal{C}_0 q_{k+1}, \tag{2.16}$$

and the assertion follows. ■

2.8. Lemma. *If one has*

$$\frac{q_{k+1}}{q_k^\tau} < \frac{1}{2\mathcal{C}_0}, \tag{2.17}$$

for any $k \in \mathbb{N}$, *then* $\boldsymbol{\omega} = (\mu, 1)$ *satisfies the Diophantine condition (2.13).*

2.9. *Proof of the lemma 2.8.* For $\boldsymbol{\nu} = \boldsymbol{\nu}_k$ one has by (2.10), (2.12) and (2.17)

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}_k| > \frac{1}{2q_{k+1}} > \frac{\mathcal{C}_0}{q_k^\tau} > \frac{\mathcal{C}_0}{|\boldsymbol{\nu}_k|^\tau}, \tag{2.18}$$

while for $\boldsymbol{\nu} \neq \boldsymbol{\nu}_k$ one can reason as follows. If $\boldsymbol{\nu} = (q_k, p)$, with $p \neq p_k$, then $|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > 1/2$ and (2.17) is trivially satisfied. If $\boldsymbol{\nu} = (q, p)$, with $q_{k-1} < q < q_k$, then by (2.10) and (2.11)

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > |\boldsymbol{\omega} \cdot \boldsymbol{\nu}_{k-1}| > \frac{1}{2q_k} > \frac{\mathcal{C}_0}{q_{k-1}^\tau} > \frac{\mathcal{C}_0}{q^\tau} > \frac{\mathcal{C}_0}{|\boldsymbol{\nu}|^\tau}, \tag{2.19}$$

so that (2.13) follows. ■

2.10. Lemma. *Given a rotation vector* $\boldsymbol{\omega} = (\mu, 1)$, *with* $\mu \in (0,1)$, *satisfying the Diophantine condition*

$$|\boldsymbol{\omega} \cdot \boldsymbol{\nu}| > C |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.20}$$

and fixed any interval $I \subset \mathbb{R}$ *with center in* μ, *there exist infinitely many* $\mu_0 \in I \cap (0,1)$ *such that* $\boldsymbol{\omega}_0 = (\mu_0, 1)$ *satisfies the Diophantine condition*

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.21}$$

with $C_0 = bC$, *for some constant positive* b.

2.11. *Proof of the lemma 2.10.* Given $\mu = [a_0, a_1, a_2, \ldots]$ and any interval I with center in μ define $\mu' = [a'_0, a'_1, a'_2, , \ldots]$ in the following way:

$$\begin{cases} a'_k = a_k, & \text{if } k \leq k_0, \\ a'_k \leq N, & \text{if } k > k_0, \end{cases} \tag{2.22}$$

where k_0 is so large that $\mu' \in I$ and N is an integer. Then $q'_k = q_k$ for all $k \leq k_0$, so that

$$\frac{q'_{k+1}}{{q'_k}^\tau} < \frac{2^\tau}{C}, \tag{2.23}$$

by the lemma 2.6, while

$$q_{k'+1} = a'_k q'_k + q'_{k-1} \leq N q'_k + q'_{k-1} \leq 2N q'_k < 2N {q'_k}^\tau \tag{2.24}$$

for all $k > k_0$. Then, by the lemma 2.8, $\boldsymbol{\omega}' = (\mu', 1)$ satisfies the Diophantine condition

$$|\boldsymbol{\omega}' \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \tag{2.25}$$

with $C_0 = \min\{C/2^{\tau+1}, 1/4N\}$. If we set $a'_k \in \{1, \ldots, N\}$ for all $k > k_0$, we obtain an infinite set of values satisfying the Diophantine condition (2.25). So, as one has $C \le \mu + 1 \le 2$, the lemma is proved, with $b = \min\{1/2^{\tau+1}, 1/16N\}$. ■

2.12. *Proof of the proposition 2.4.* The lemma 2.2 shows that the function $\eta(\mathbf{A}, \varepsilon, \mu_0)$ is analytic in ε, with radius of convergence $\varepsilon_0 = E_0 C_0$, so that one has $\eta(\mathbf{A}_0, \varepsilon, \mu_0) = \varepsilon \eta^{(1)}(\mathbf{A}_0, \mu_0) + O(\varepsilon^2)$, for $|\varepsilon| < \varepsilon_0$ small enough. The condition (1.7) assures that one has $\eta^{(1)}(\mathbf{A}_0, \mu_0) = -\partial_{A_1} f^{(1)}_{\mathbf{0}}(\mathbf{A}_0) \neq 0$ (see remark 4.4 below): then there exists a positive constant η_1 such that $|\eta^{(1)}(\mathbf{A}_0, \mu_0)| > \eta_1$, for any $\mu_0 \in \mathcal{B}_{aC}(\mu)$ – in fact $\eta^{(1)}(\mathbf{A}_0, \mu_0) \equiv \eta^{(1)}(\mathbf{A}_0)$ does not depend on μ_0. Therefore, varying ε in $(-\varepsilon_0, \varepsilon_0)$, with $\varepsilon_0 = E_0 C_0 = E_0 bC$, the function $\eta(\mathbf{A}_0, \varepsilon, \mu_0)$ covers an interval $\mathcal{B}_{dC}(0)$, for some positive constant d (depending on b). As $|\mu - \mu_0|$ is bounded by aC, one can choose $a < d$ so that by moving ε in $(-\varepsilon_0, \varepsilon_0)$ there is at least one value $\varepsilon = \varepsilon(\mu_0, \mathbf{A}_0)$ such that $\eta(\mathbf{A}_0, \varepsilon(\mu_0, \mathbf{A}_0), \mu_0) = \mu - \mu_0$. ■

2.13. *Remark.* The condition $\partial_{A_1} f^{(1)}_{\mathbf{0}} \neq 0$ imposed on f is not really necessary, and it is simply aimed to assures that the counterterm is not identically vanishing. In fact under such a weaker condition, if $\eta^{(k_0)}$ is the first nonvanishing coefficient in the expansion (3.1) for $\eta(\mathbf{A}_0, \varepsilon, \mu_0)$, then, at worst, when ε is varied in $(-\varepsilon_0, \varepsilon_0)$, the counterterm $\eta(\mathbf{A}_0, \varepsilon, \mu_0)$ covers an interval of width at least $2dC^{k_0}$, for some $d > 0$, so that a result analogous to the proposition 2.4 follows, provided the width of the interval I is chosen $|I| = 2aC^{k_0}$, for some $a < d$. Note that the condition that the counterterm is not identically vanishing to first order amounts to a genericity condition on the perturbation f.

2.14. *Proof of the theorem 1.6.* Now suppose that, instead of the condition (1.7), one requires

$$M \equiv \partial^2_{A_1} f^{(1)}_{\mathbf{0}}(\overline{\mathbf{A}}) \equiv \int_{\mathbb{T}^2} \mathrm{d}\boldsymbol{\alpha}\, \partial^2_{A_1} f^{(1)}(\boldsymbol{\alpha}, \overline{A}_1) \neq 0. \tag{2.26}$$

Then, instead of fixing the action variables and moving ε until the compatibility condition (1.15) is satisfied (as in §2.12), we can fix ε small enough (say smaller than a value $\overline{\varepsilon}$ to be determined) and slightly change $\overline{A}_1$ into a nearby value A'_1 such that one still has

$$\mu_0 + \eta(\mathbf{A}', \varepsilon, \mu_0) = \mu, \qquad \mathbf{A}' = (A'_1, \overline{A}_2), \tag{2.27}$$

for some μ_0 such that $\boldsymbol{\omega}_0 = (\mu_0, 1)$ satisfies the Diophantine condition

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \qquad \boldsymbol{\omega}_0 = (\mu_0, 1), \tag{2.28}$$

with $C_0 = bC$, for some constant $b \in (0, 1)$. This can be easily seen by reasoning as follows. Henceforth we denote $\mathbf{A} = (A_1, \overline{A}_2)$, where $\overline{A}_2$ is fixed once and for all.

Fix $\overline{A}_1 \in D$ and ρ such that $\mathcal{B}_\rho(\overline{A}_1) \subset D$; let us fix $0 < \delta < \rho'$, where ρ' is given by the lemma 2.2, such that for all $\mathbf{A} \in \mathcal{A}_\delta(\overline{A}_1)$ one has $|\partial^2_{A_1} f^{(1)}_{\mathbf{0}}(\mathbf{A})| > M/2$: therefore for $\mathbf{A}$ varying in $\mathcal{A}_{\delta/2}(\overline{A}_1)$ the quantity $\varepsilon \partial_{A_1} f^{(1)}_{\mathbf{0}}(\mathbf{A})$ covers an interval $J(\varepsilon)$ such that $|J(\varepsilon)| = O(\varepsilon)$; more precisely one has $|J(\varepsilon)| > M\varepsilon\delta/4$.

Let $\overline{\varepsilon} \le \varepsilon_0$ be such that for any $|\varepsilon| < \overline{\varepsilon}$ the interval $\mu + J(\varepsilon)$ contains at least one value μ', such that $\boldsymbol{\omega}' = (\mu', 1)$ verifies the the Diophantine condition

$$|\boldsymbol{\omega}' \cdot \boldsymbol{\nu}| > C_1 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \qquad \boldsymbol{\omega}' = (\mu', 1), \tag{2.29}$$

with $C_1 = b_1 C$, for some constant $b_1 \in (0, 1)$; by reasoning as for proving the lemma 2.10 it is easy to realize that this is possible.

By applying once more the lemma 2.10 we can conclude that for all $a' > 0$ the interval $\mathcal{B}_{a'C}(\mu')$ contains values μ_0 such that $\boldsymbol{\omega}_0 = (\mu_0, 1)$ satisfies (2.28), for some constant b such that $C_0 = b_2 C_1 = b_2 b_1 C \equiv bC$: in particular this yields that $J(\varepsilon)$ contains infinitely many such μ_0.

Fix ε such that $|\varepsilon| \le \overline{\varepsilon}$ and choose $\mathbf{A}'' \in \mathcal{A}_{\delta/2}(\overline{A}_1)$ such that

$$\mu_0 = \mu - \varepsilon \eta^{(1)}(\mathbf{A}''), \qquad \eta^{(1)}(\mathbf{A}'') = -\partial_{A_1} f^{(1)}(\mathbf{A}''); \tag{2.30}$$

we can suppose that $\mathbf{A}''$ is such that $\eta^{(1)}(\mathbf{A}'') \neq 0$ (if not simply choose a nearby value $\mu_0 \neq 0$ and use the property (2.26)).

Then write, for any $\mathbf{A} \in \mathcal{A}_\delta(\overline{A}_1)$,

$$\eta(\mathbf{A}, \varepsilon, \mu_0) = \varepsilon\, \eta^{(1)}(\mathbf{A}) + \varepsilon^2 \xi(\mathbf{A}, \varepsilon, \mu_0), \qquad \sup_{\mathbf{A} \in \mathcal{A}_\delta(\overline{A}_1)} |\xi(\mathbf{A}, \varepsilon, \mu_0)| \le \Xi, \tag{2.31}$$

with Ξ a suitable constant: this follows again from the lemma 2.2, by taking into account that $\delta < \rho'$.

Then define $\mathbf{A}'$ implicitly as the solution of the equation

$$\mu = \mu_0 - \varepsilon \partial_{A_1} f^{(1)}(\mathbf{A}'') = \mu_0 - \varepsilon \partial_{A_1} f^{(1)}(\mathbf{A}') + \varepsilon^2 \xi(\mathbf{A}', \varepsilon, \mu_0) \equiv \mu_0 + \eta(\mathbf{A}', \varepsilon, \mu_0) : \tag{2.32}$$

if such a solution exists with A_1' in $\mathcal{B}_{\rho'}(\overline{A}_1)$, then we have proved (2.27). The solution $\mathbf{A}'$ of (2.32) can be found as a simple consequence of the implicit function theorem. In fact the function

$$F(A_1', \varepsilon) \equiv \partial_{A_1} f^{(1)}(\mathbf{A}'') - \partial_{A_1} f^{(1)}(\mathbf{A}') + \varepsilon\, \xi(\mathbf{A}', \varepsilon, \mu_0) \tag{2.33}$$

is analytic both in A_1' and in ε (for $A_1' \in \mathcal{B}_{\rho'}(\overline{A}_1)$ and $\varepsilon \in \mathcal{B}_{\varepsilon_0}(0)$). As

$$F(A_1'', 0) = 0, \qquad \partial_{A_1'} F(A_1'', 0) = -\partial^2_{A_1} f^{(1)}(\mathbf{A}'') \neq 0, \tag{2.34}$$

there exists a value $A_1' = A_1'(\varepsilon)$ such that

$$F(A_1'(\varepsilon), \varepsilon) = 0; \tag{2.35}$$

moreover one has

$$\left| \partial_{A_1} f^{(1)}(\mathbf{A}') - \partial_{A_1} f^{(1)}(\mathbf{A}'') \right| > \frac{M}{2} |\mathbf{A}' - \mathbf{A}''| , \tag{2.36}$$

so that there exists $A_1' \in \mathcal{B}_\delta(\overline{A}_1) \subset \mathcal{B}_{\rho'}(\overline{A}_1)$, provided that $\overline{\varepsilon}$ is so small that one has $4\Xi\overline{\varepsilon} < M\delta$, and the assertion is proved. ∎

2.15. *Remark.* As already noted in the remark 1.7, (1), the condition (1.11) is not really necessary in order to prove the theorem 1.6: what one has to require is that the counterterm is not constant in the action variable A_1, for $A_1 \in D$.

3. Perturbation theory

3.1. *Lindstedt series.* In the following we assume that $\mathbf{A}_0$ and μ_0 are fixed once and for all, and we shall not write explicitly the dependence on $\mathbf{A}_0$ and μ_0: so we shall write $\eta(\mathbf{A}_0, \varepsilon, \mu_0) = \eta(\varepsilon)$, and so on.

We look for a solution of the form (2.3), with

$$\begin{aligned} \mathbf{h}(\boldsymbol{\psi}, \varepsilon) &= \sum_{k=1}^{\infty} \varepsilon^k \sum_{\boldsymbol{\nu} \in \mathbb{Z}^2} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}}\, \mathbf{h}^{(k)}_{\boldsymbol{\nu}}, \\ \mathbf{H}(\boldsymbol{\psi}, \varepsilon) &= \sum_{k=1}^{\infty} \varepsilon^k \sum_{\boldsymbol{\nu} \in \mathbb{Z}^2} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}}\, \mathbf{H}^{(k)}_{\boldsymbol{\nu}}, \\ \eta(\varepsilon) &= \sum_{k=1}^{\infty} \varepsilon^k \eta^{(k)}. \end{aligned} \tag{3.1}$$

The formal series (3.1) are called the *Lindstedt series.*

Note that writing $\mathbf{h} = (h_1, h_2)$ one has $h_2 = 0$ identically as $\alpha_2(t) = t$ for any ε.

More generally, for any function $F = F(\boldsymbol{\psi}, \mathbf{A}_0, \varepsilon)$ analytic in its arguments and 2π-periodic in $\boldsymbol{\psi}$, we denote by $[F]_{\boldsymbol{\nu}}^{(k)}$ the coefficient $F_{\boldsymbol{\nu}}^{(k)}$ with Fourier label $\boldsymbol{\nu}$ and Taylor label k in its expansion

$$F(\boldsymbol{\psi}, \mathbf{A}_0, \varepsilon) = \sum_{k=1}^{\infty} \varepsilon^k \sum_{\boldsymbol{\nu} \in \mathbb{Z}^2} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}} F_{\boldsymbol{\nu}}^{(k)}(\mathbf{A}_0) \equiv \sum_{k=1}^{\infty} \varepsilon^k \sum_{\boldsymbol{\nu} \in \mathbb{Z}^2} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\psi}} F_{\boldsymbol{\nu}}^{(k)}. \tag{3.2}$$

If we put (2.3) into (2.4) by using the expansions (3.1) we obtain, for $\boldsymbol{\nu} \neq \mathbf{0}$,

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\nu}}^{(k)} &= g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}) \, [\partial_{\mathbf{A}} f]_{\boldsymbol{\nu}}^{(k)} \, , \\ \mathbf{H}_{\boldsymbol{\nu}}^{(k)} &= -g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}) \, [\partial_{\boldsymbol{\alpha}} f]_{\boldsymbol{\nu}}^{(k)} \, , \end{aligned} \tag{3.3}$$

with

$$g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}) = \frac{1}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}}, \tag{3.4}$$

provided that, for $\boldsymbol{\nu} = \mathbf{0}$, one has

$$\begin{aligned} \eta^{(k)} + [\partial_{A_1} f]_{\mathbf{0}}^{(k)} &= 0, \\ [\partial_{\boldsymbol{\alpha}} f]_{\mathbf{0}}^{(k)} &= \mathbf{0}. \end{aligned} \tag{3.5}$$

We can write (3.3) as

$$\begin{aligned} \mathbf{h}_{\nu}^{(k)} &= g(\boldsymbol{\nu}) {\sum}^* \frac{1}{p!} \frac{1}{q!} \left(\prod_{p'=1}^{p} i\boldsymbol{\nu}_0 \cdot \mathbf{h}_{\boldsymbol{\nu}_{p'}}^{(k_{p'})} \right) \left(\prod_{q'=p+1}^{p+q} \mathbf{H}_{\boldsymbol{\nu}_{q'}}^{(k_{q'})} \cdot \partial_{\mathbf{A}} \right) \partial_{\mathbf{A}} f_{\boldsymbol{\nu}_0}^{(k)}(\mathbf{A}_0), \\ \mathbf{H}_{\nu}^{(k)} &= -g(\boldsymbol{\nu}) {\sum}^* \frac{1}{p!} \frac{1}{q!} \left(\prod_{p'=1}^{p} i\boldsymbol{\nu}_0 \cdot \mathbf{h}_{\boldsymbol{\nu}_{p'}}^{(k_{p'})} \right) \left(\prod_{q'=p+1}^{p+q} \mathbf{H}_{\boldsymbol{\nu}_{q'}}^{(k_{q'})} \cdot \partial_{\mathbf{A}} \right) (i\boldsymbol{\nu}_0) \, f_{\boldsymbol{\nu}_0}^{(k)}(\mathbf{A}_0), \end{aligned} \tag{3.6}$$

where ${\sum}^*$ is a shorthand for

$${\sum}^* = \sum_{k_0=1}^{\infty} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \sum_{\substack{k_1 \geq 1, \ldots, k_{p+q} \geq 1 \\ k_0 + k_1 + \ldots + k_{p+q} = k}} \sum_{\substack{\nu_1 \in \mathbb{Z}, \ldots, \nu_{p+q} \in \mathbb{Z} \\ \nu_0 + \nu_1 + \ldots + \nu_{p+q} = \nu}} ; \tag{3.7}$$

for $k = 1$ the formulae in (3.6) have to be interpreted in the appropriate (obvious) way. Note in (3.6) that $\mathbf{H}_{\boldsymbol{\nu}}^{(k)}$ is given by a sum of contributions which have always (at least) one derivative with respect to $\boldsymbol{\alpha}$, whereas $\mathbf{h}_{\boldsymbol{\nu}}^{(k)}$ is given by the sum of contributions which have always (at least) one derivative with respect to $\mathbf{A}$. Then we can introduce the following notation: $\mathbf{H}_{\boldsymbol{\nu}}^{(k)}$ is given by a sum of terms which are of the form $H \leftarrow h$, where H denotes that they contribute to $\mathbf{H}_{\boldsymbol{\nu}}^{(k)}$ and h that there is always a derivative with respect to the angle variables; in the same way, $\mathbf{h}_{\boldsymbol{\nu}}^{(k)}$ is given by a sum of terms which are of the form $h \leftarrow H$, where h denotes that they contribute to $\mathbf{h}_{\boldsymbol{\nu}}^{(k)}$ and H that there is always a derivative with respect to the action variables.

So we have two kinds of problems: first to show the formal solubility of equations (2.4), *i.e.* to show that to each perturbative order (3.5) are satisfied so that no division by zero is performed; then to show that the formal series (3.1) defining $\mathbf{h}$, $\mathbf{H}$ and η converge.

3.2. *Notations.* We introduce some notations which will be used in the following.

Given a set of elements S we denote by $|S|$ the number of elements in S.

Recall that ∂_x denotes the partial derivative with respect to x. If a function F depends only on one argument, $F = F(x)$, then we shall write sometimes $\partial F(x)$ for $\partial_x F(x)$, as no ambiguity can arise in such a case.

Given a vector $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ we set $|\mathbf{v}| = |v_1| + |v_2|$. If $\mathbf{v} \in \mathbb{R}^2$ then for any $p \in \mathbb{N}$ the quantity $\mathbf{v}^p$ denotes the tensor with entries $v_{i_1} \dots v_{i_p}$, where $i_j \in \{1,2\}$ for all $j = 1, \dots, p$. Likewise $\partial^q_{\mathbf{A}}$ will denotes the tensor with entries $\partial_{A_{i_1}} \dots \partial_{A_{i_q}}$ where $i_j \in \{1,2\}$ for all $j = 1, \dots, q$.

3.3. *Tree expansion.* An *unlabeled tree* θ is a partially ordered set of points and lines connecting the points. The partial ordering relation between the nodes is from right to left and it denoted by $\preceq$. The leftmost point r is called the *root* of the tree; all the other points are called *nodes* and are denoted by v. The lines are denoted by ℓ; they carry an arrow oriented towards the root. If a line ℓ connects a node v_2 to a node $v_1 \succ v_2$, we shall say that the line is *attached* to the nodes v_1 and v_2, and write $\ell = \ell_{v_2}$ and $v_2' = v_1$: we say also that the line enters v_1 and exits from v_2, and that v_1 is the node immediately following v_2. The line ℓ_0 entering the root is the *root line*. The node v_0 which it exits from is called the *last node* of the tree: one has $\ell_0 = \ell_{v_0}$. See Fig. 1.

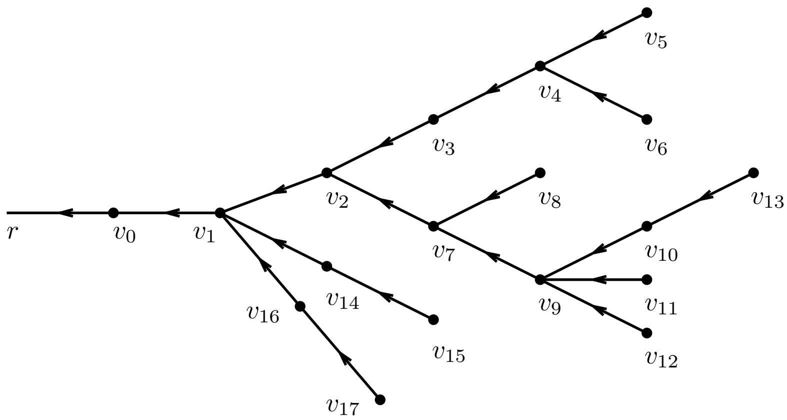

FIG. 1. An unlabeled tree θ with 18 nodes.

We shall call $V(\theta)$ the set of nodes in θ and $\Lambda(\theta)$ the set of lines in θ; one has $|V(\theta)| = |\Lambda(\theta)|$.

Two trees are said to be equivalent if they are obtained from each other by continuously deforming the lines in such a way that the latter do not cross each other: in the following we shall always identify equivalent trees.

Given a tree θ and any node $v \prec v_0$, the set θ' of nodes $w \preceq v$ and of lines connecting them form with the line ℓ_v a tree with root v' and root line ℓ_v: we say that θ' is a subtree of θ. We denote by $\theta \setminus \theta'$ the set of nodes in $V(\theta) \setminus V(\theta')$ and of lines connecting them; we shall write also $V(\theta) \setminus V(\theta') = V(\theta \setminus \theta')$.

We shall write the perturbation f as $f(\boldsymbol{\alpha}, A_1, \varepsilon) = f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon)$, even if the dependence is only through the first component A_1 of $\mathbf{A}$; see the remark 3.5 below. Then f can be expanded as

$$f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon) = \sum_{k=1}^{\infty} \varepsilon^k f^{(k)}(\boldsymbol{\alpha}, \mathbf{A}) = \sum_{k=1}^{\infty} \varepsilon^k \sum_{\boldsymbol{\nu} \in \mathbb{Z}^2} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\alpha}}\, f^{(k)}_{\boldsymbol{\nu}}(\mathbf{A}); \tag{3.8}$$

note that, by the analyticity assumptions on f, one has

$$|f^{(k)}_{\boldsymbol{\nu}}(\mathbf{A})| < F_{01} F_{02}^k e^{-\kappa|\boldsymbol{\nu}|}, \qquad F_{01} = \max_{\substack{|\varepsilon|=\varepsilon' \\ |A_1 - A_{01}|=\rho'}} |f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon)|, \qquad F_{02} = {\varepsilon'}^{-1}, \tag{3.9}$$

for any $0 < \varepsilon' < \varepsilon_1$ and $0 < \rho' < \rho$.

Then to each node v we associate a *mode label* $\boldsymbol{\nu}_v \in \mathbb{Z}^2$ and an *order label* $k_v \in \mathbb{N}$. We define the *order* k of a tree as the sum of the values of the order labels of the nodes:

$$k = \sum_{v \in V(\theta)} k_v. \tag{3.10}$$

Note that if $k_v = 1$ for all $v \in V(\theta)$ then $k = |V(\theta)|$.

Define the *momentum* flowing through a line ℓ_v as

$$\boldsymbol{\nu}_{\ell_v} = \sum_{w \preceq v} \boldsymbol{\nu}_w. \tag{3.11}$$

To each node v we associate a *node factor* F_v, which is a function of $\boldsymbol{\nu}_v$, while to each line ℓ we associate a *propagator* G_ℓ, which is a function of $\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell$. We distinguish between two kinds of lines, lines $h \leftarrow H$ and lines $H \leftarrow h$, and we assign to them and to the nodes they exit from different values for G_ℓ and F_v, if $\ell = \ell_v$, in the following way.

We associate to each node v three non-negative integer labels p_v, q_v and m_v with the constraint $p_v + q_v = m_v$: m_v is the number of lines entering v, while the labels p_v and q_v denote, respectively, the number of lines $h \leftarrow H$ and $H \leftarrow h$ entering it.

Given a node v let us denote by $\ell = \ell_v$ the line exiting from it. Then the *node factor* F_v and the propagator G_ℓ are defined as

$$\begin{array}{lll} \ell & h \leftarrow H & H \leftarrow h \\ F_v & \dfrac{1}{p_v!}\dfrac{1}{q_v!}(i\boldsymbol{\nu}_v)^{p_v}\partial_{\mathbf{A}}^{q_v+1} f_{\boldsymbol{\nu}_v}^{(k_v)}(\mathbf{A}_0) & \dfrac{1}{p_v!}\dfrac{1}{q_v!}(i\boldsymbol{\nu}_v)^{p_v+1}\partial_{\mathbf{A}}^{q_v} f_{\boldsymbol{\nu}_v}^{(k_v)}(\mathbf{A}_0), \\ G_\ell & g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell) & -\, g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell). \end{array} \tag{3.12}$$

Note that for $f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon) = f(\boldsymbol{\alpha}, A_1, \varepsilon)$ the only nonvanishing entry of the tensor $\partial_{\mathbf{A}}^q f_{\boldsymbol{\nu}}^{(k)}(\mathbf{A}_0)$ is $\partial_{A_1}^q f_{\boldsymbol{\nu}}^{(k)}(\mathbf{A}_0)$.

If we introduce a label δ_v such that $\delta_v = 1$ if ℓ_v is a line $h \leftarrow H$ and $\delta_v = 0$ if ℓ_v is a line $H \leftarrow h$, then F_v can be written as

$$F_v = \frac{1}{p_v!}\frac{1}{q_v!}(i\boldsymbol{\nu}_v)^{p_v+(1-\delta_v)}\partial_{\mathbf{A}}^{q_v+\delta_v} f_{\boldsymbol{\nu}_v}^{(k_v)}(\mathbf{A}_0) \tag{3.13}$$

in both cases. Note that one has

$$\begin{aligned} & \frac{1}{q!}|\partial_{\mathbf{A}}^q f_{\boldsymbol{\nu}}^{(k)}(\mathbf{A})| < F_1 F_2^k e^{-\kappa|\boldsymbol{\nu}|} F_3^q, \\ & F_1 = \max_{\substack{|\varepsilon|=\varepsilon' \\ |A_1 - A_{01}|=\rho'}} |f(\boldsymbol{\alpha}, \mathbf{A}, \varepsilon)|, \qquad F_2 = \varepsilon'^{-1}, \qquad F_3 = \rho'^{-1}, \end{aligned} \tag{3.14}$$

by (3.9) and by the assumptions on the dependence on $\mathbf{A}$.

We have that $\mathbf{X} \in \{\mathbf{h}, \mathbf{H}\}$ can be written as

$$\begin{aligned} \mathbf{X}_{\boldsymbol{\nu}}^{(k)} &= \sum_{\theta \in \mathcal{T}_{k,\nu}(\mathbf{X})} \mathrm{Val}(\theta), \\ \mathrm{Val}(\theta) &= \left(\prod_{v \in V(\theta)} F_v \right) \left(\prod_{\ell \in \Lambda(\theta)} G_\ell \right), \end{aligned} \tag{3.15}$$

where $\mathcal{T}_{k,\boldsymbol{\nu}}(\mathbf{X})$ is the set of all labeled trees of order k with momentum $\boldsymbol{\nu}$ flowing through the root line and such that if $\mathbf{X} = \mathbf{h}$ then the root line is a line $h \leftarrow H$, while if $\mathbf{X} = \mathbf{H}$ then the root line is a line $H \leftarrow h$. The proof of (3.15) can be performed by induction on the order k, by using (3.3) and expanding the functions in the square brackets.

Define also

$$\mathrm{Val}'(\theta) = \left(\prod_{v \in V(\theta)} F_v \right) \left(\prod_{\ell \in \Lambda(\theta) \setminus \ell_0} G_\ell \right), \tag{3.16}$$

where ℓ_0 is the root line. If we introduce the vector $\boldsymbol{\eta} = (\eta, 0)$, so that $\boldsymbol{\eta}^{(k)} = (\eta^{(k)}, 0)$, then also $\boldsymbol{\eta}$ admits a representation

$$\boldsymbol{\eta}^{(k)} = -\sum_{\theta \in \mathcal{T}_{k,0}(\mathbf{h})} \mathrm{Val}'(\theta), \tag{3.17}$$

where $\mathcal{T}_{k,\mathbf{0}}(\mathbf{h})$ means that the trees over which the sum is performed have the constraint that the root line is a line $h \leftarrow H$; see (3.5).

We denote also by $\mathcal{T}_{k,\boldsymbol{\nu}}$ the set of all labeled trees of order k and with momentum $\boldsymbol{\nu}$ flowing through the root line, with no condition on the kind of root line.

3.4. *Formal solubility and solubility.* We shall prove first of all that the coefficients defining the formal series (3.1) are finite to each perturbative order, *i.e.* that one has $|\mathbf{h}_{\boldsymbol{\nu}}^{(k)}| < \infty$ for all $k \in \mathbb{N}$ and for all $\boldsymbol{\nu} \in \mathbb{Z}^2$. Hence we shall deal with the problem to prove the convergence of the series: this will require a more careful analysis of the tree values, inspired to the renormalization group approach in quantum field theory.

3.5. *Remark.* The tree representation given in this section can be carried out, essentially unchanged, for any analytic perturbation of isochronous systems of any dimension d. This explains why we used the notation (3.8) for the perturbation and the vectorial notation (3.17) for the counterterms: of course in general all the components of the counterterms are not vanishing.

4. Proof of the formal solubility of the equations of motion

4.1. *Formal solubility.* To show that there exists a formal solution (2.3) of the equations of motion (2.4) we have to show that for any θ no division by zero occurs in $\mathrm{Val}(\theta)$ and in $\mathrm{Val}'(\theta)$. Recall that, given any tree θ, any line $\ell \in \Lambda(\theta)$ can be considered as the root line of the subtree formed by the nodes and lines preceding ℓ. So we have to show that the sum of all trees contributing to $[\partial_{\boldsymbol{\alpha}} f]_{\mathbf{0}}^{(k)}$ is vanishing, provided that $\eta^{(k)}$ is chosen in such a way that $\boldsymbol{\eta}^{(k)} + [\partial_{\mathbf{A}} f]_{\mathbf{0}}^{(k)} = \mathbf{0}$. Then the formal solubility is implied from the following result.

4.2. LEMMA. *There is a unique choice for the coefficients $\eta^{(k)}$ such that, for all $k \geq 1$, one has $[\partial_{\boldsymbol{\alpha}} f]_{\mathbf{0}}^{(k)} = \mathbf{0}$ and $\boldsymbol{\eta}^{(k)} + [\partial_{\mathbf{A}} f]_{\mathbf{0}}^{(k)} = \mathbf{0}$. Such a choice is given by (3.17).*

4.3. *Proof of the lemma 4.2.* The proof can be done by induction. For $k = 1$ the assertion is trivially satisfied, as

$$[\partial_{\boldsymbol{\alpha}} f]_{\boldsymbol{\nu}}^{(1)} = i\boldsymbol{\nu} f_{\boldsymbol{\nu}}^{(1)}(\mathbf{A}_0), \tag{4.1}$$

which is vanishing for $\boldsymbol{\nu} = \mathbf{0}$, while imposing

$$[\partial_{A_1} f]_{\mathbf{0}}^{(1)} + \eta^{(1)} = 0 \tag{4.2}$$

fixes $\eta^{(1)}$ as in (3.17).

If the assertion holds for all $k' < k$ then we can show that it holds also for k. By the inductive hypothesis all lines in θ which are not the root line have a nonvanishing momentum (as they are the root lines of subtrees of order strictly less than k), so that $\mathrm{Val}'(\theta)$ is well defined.

Then consider all contributions arising from the trees $\theta \in \mathcal{T}_{k,\mathbf{0}}(\mathbf{H})$, hence having as root line a line $H \leftarrow h$: we group together all trees obtained from each other by shifting the root line, *i.e.* by changing the node which the root line exits and orienting the arrows in such a way that they still point towards the root. We call $\mathcal{F}(\theta)$ such a class of trees (here θ is any element inside the class). See Fig. 2.

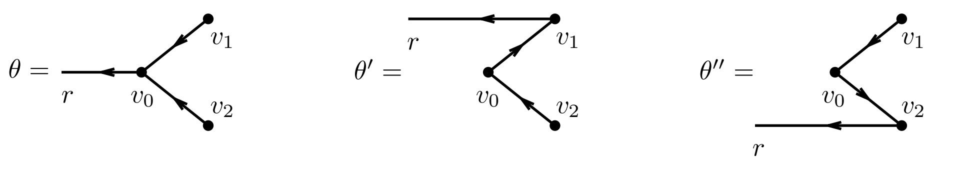

FIG. 2. The family $\mathcal{F}(\theta)=\{\theta,\theta',\theta''\}$ for a tree $\theta\in\mathcal{T}_{3,0}$. The labels are not explicitly shown.

The values $\mathrm{Val}'(\theta')$ of such trees $\theta' \in \mathcal{F}(\theta)$ differ as (1) there is a factor $i\boldsymbol{\nu}_v$ depending on the node v which the root line is attached to (see the definition (3.10) of F_v for the lines $H \leftarrow h$), and (2) some arrows change their directions. More precisely, when the root line is detached from the node v_0 and reattached to the node v, if $\mathcal{P}(v_0, v) = \{w \in V(\theta) \ : \ v_0 \succeq w \succeq v\}$ denotes the path joining the node v_0 to the node v, all the lines $h \leftarrow H$ along the path $\mathcal{P}(v_0, v)$ become lines $H \leftarrow h$ and *vice versa.* As a consequence the signs of the momenta flowing through them and the factorials of the node factors corresponding to the nodes joined by them can change.

The change of the signs of the momenta simply follows from the fact that

$$\sum_{v \in V(\theta)} \boldsymbol{\nu}_v = \mathbf{0}, \tag{4.3}$$

as $\theta \in \mathcal{T}_{k,\mathbf{0}}$: this means that some propagators G_ℓ change from $g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell)$ to $-g(-\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell)$ (compare the definitions of the propagators for lines $h \leftarrow H$ and $H \leftarrow h$ in (3.10)), but, by the definition of propagator (3.2), one has $g(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell) = -g(-\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell)$.

The change of the node factors is due to the fact that for the nodes along the path $\mathcal{P}(v_0, v)$, an entering line can become an exiting line and *vice versa*, so that the labels p_v and q_v can be transformed into $p_v \pm 1$ and $q_v \pm 1$, respectively: this does not modify the factor $(i\boldsymbol{\nu}_v)^{p_v + (1-\delta_v)} \partial_{\mathbf{A}}^{q_v + \delta_v} f_{\boldsymbol{\nu}_v}^{(k_v)}(\mathbf{A}_0)$ in (3.13), as one immediately checks, but it can produce a change of the factorials.

If we neglect the change of the factorials, *i.e.* if we assume that all combinatorial factors are the same, by summing over all possible trees inside the class $\mathcal{F}(\theta)$ we obtain a common value times i times (4.3), and the sum gives zero. One can easily show that a correct counting of the trees implies that all factorials are in fact equal: simply reason as in [BG], sect. 3, by using topological trees. Therefore the above argument proves the second equation in (3.5).

In order to make soluble the equation for $\mathbf{h}$ to order k one has to impose that $\eta^{(k)}$ deletes the Fourier component with label $\boldsymbol{\nu} = \mathbf{0}$ arising from $[\partial_{A_1} f]^{(k)}$ (the one arising from $[\partial_{A_2} f]^{(k)}$ is automatically vanishing as f does not depend on A_2): this gives the condition (3.17).

The summation on the trees can be easily performed, as the summability over the Fourier labels is assured by the Diophantine condition (which is not the optimal condition under which the formal solubility can be proved; see also [BG] for the case of the maps on the cylinder), while all the other labels can assume only a finite number of values. So the proof of the lemma is concluded. ■

4.4. *Remark.* The assumption (1.7) on the perturbation f implies, by (4.2), that one has $\eta^{(1)} \neq 0$.

5. Proof of convergence of the perturbative expansion

5.1. *Bound on the node factors.* To prove the solubility of the equations of motions, *i.e.* the summability of the Lindstedt series (3.1), we shall prove the bounds

$$|\mathbf{h}_{\boldsymbol{\nu}}^{(k)}| \le B^k \, e^{-\kappa' |\boldsymbol{\nu}|}, \qquad |\mathbf{H}_{\boldsymbol{\nu}}^{(k)}| \le B^k \, e^{-\kappa' |\boldsymbol{\nu}|}, \qquad |\eta^{(k)}| \le B^k, \tag{5.1}$$

for suitable constants $B > 0$ and $\kappa' \in (0, \kappa)$.

The sum over the labeled trees in $\mathcal{T}_{\kappa,\boldsymbol{\nu}}$ can be written as the sum over all the unlabeled trees and over all the ways to assign the mode and order labels to the nodes of the unlabeled trees with the constraints

$$\sum_{v \in V(\theta)} k_v = k, \qquad \sum_{v \in V(\theta)} \boldsymbol{\nu}_v = \boldsymbol{\nu}. \tag{5.2}$$

Define

$$M(\theta) = \sum_{v \in V(\theta)} |\boldsymbol{\nu}_v|. \tag{5.3}$$

Of course $M(\theta) \ge |\boldsymbol{\nu}|$ for $\theta \in \mathcal{T}_{k,\boldsymbol{\nu}}$.

The number of unlabeled trees with V nodes is bounded by 2^{2V}.

We can see that, if it was possible to neglect the propagators in the definition of $\mathrm{Val}(\theta)$ and $\mathrm{Val}'(\theta)$, a bound like (5.1) would immediately follow.

In fact suppose for the time being to neglect the propagators, *i.e.* to replace G_ℓ with 1 in the definition of Val(θ) and Val(θ') in §3.3. Then the sum over the labels can be performed as follows: the sum over the mode labels is controlled through

$$\begin{aligned}\sum_{\{\boldsymbol{\nu}_v\}_{v\in V(\theta)}\,:\,\sum_{v\in V(\theta)}\boldsymbol{\nu}_v=\boldsymbol{\nu}}\prod_{v\in V(\theta)}\frac{1}{q_v!}\frac{1}{p_v!}\left|\boldsymbol{\nu}_v\right|^{p_v+(1-\delta_v)}\left|\partial_{\mathbf{A}}^{q_v+\delta_v}f_{\boldsymbol{\nu}}^{(k_v)}(\mathbf{A}_0)\right| \\ \le e^{-\kappa M(\theta)}F_1^{|V(\theta)|}F_2^kF_3^{2k}\frac{M(\theta)^{2|V(\theta)|}}{(2|V(\theta)|)!} \\ \le e^{-\kappa M(\theta)}F_1^{|V(\theta)|}\left(\frac{F_2F_3^2}{\kappa_1^2}\right)^k e^{\kappa_1 M(\theta)},\end{aligned} \tag{5.4}$$

for any $\kappa_1\in(0,\kappa)$, where we used that

$$\sum_{v\in V(\theta)}m_v=|V(\theta)|-1,\qquad \sum_{v\in V(\theta)}p_v\le|V(\theta)|-1,\qquad \sum_{v\in V(\theta)}q_v\le|V(\theta)|-1,\qquad |V(\theta)|\le k, \tag{5.5}$$

while the sum over the unlabeled trees and over the contraction of the indices of the node factors gives a constant to the power V, say F_4^V.

So we are left with the sum over the order labels, which is controlled through

$$\sum_{V=1}^{k}\sum_{k_1+\ldots+k_V=k}2^{2V}F_1^VF_4^V\le 2^{2k}F_1^kF_4^k\sum_{V=1}^{k}\frac{k^V}{k!}\le\left(2^2eF_1F_4\right)^k. \tag{5.6}$$

Then the bound (5.1) follows with $\kappa'=\kappa-\kappa_1$.

So we have to handle the propagators. We shall see that not all propagators can give problems; more exactly only the accumulation of propagators with the same momenta can be source of problems, as the lemma 5.4 below shows.

5.2. *Multi-scale decomposition and clusters.* We introduce a partition of unity through characteristic functions

$$1=\sum_{n=-\infty}^{1}\chi_n(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}), \tag{5.7}$$

where $\chi_n(x)$ has support on $|x|\in[C_02^{n-1},C_02^n)$ for $n\le0$, while $\chi_1(x)$ has support on $|x|\in[C_0,\infty)$; note that $\chi_n(x)=\chi(2^{-n}x)$ if $\chi(x)$ is the characteristic function of the interval $[C_0/2,C_0)$. For each propagator we write

$$G_\ell=\sum_{n_\ell=-\infty}^{1}\chi_{n_\ell}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})G_\ell\equiv\sum_{n_\ell=-\infty}^{1}G_\ell^{(n_\ell)}. \tag{5.8}$$

We say that n_ℓ is the *scale label* of the line ℓ and $G_\ell^{(n_\ell)}$ is a propagator on scale n_ℓ: note that, given the momentum $\boldsymbol{\nu}_\ell$ flowing through the line ℓ, there is only one scale n such that either

$$C_02^{n-1}\le|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell|<C_02^n,\qquad n\le0, \tag{5.9}$$

or $|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell|\ge C_0$, so that, even if (5.8) is written as an infinite series, in fact only one term is really nonvanishing. We shall say that the scale n for which (5.9) holds is the scale *compatible* with the line ℓ.

Once the scale labels have been assigned to the lines one has a natural decomposition of the tree into clusters. A *cluster* T on scale n is a maximal set of nodes and lines connecting them such that all the lines have scales $n'\ge n$ and there is at least one line on scale n. The $m_T\ge0$ lines entering the cluster T and the (only one or zero) exiting line are called the *external lines* of the cluster T. Given a cluster T on scale n, we shall denote by $n_T=n$ the scale of the cluster. We call $T(\theta)$ the set of all clusters in a tree θ; given a cluster $T \in T(\theta)$ call $V(T)$ and $\Lambda(T)$ the set of nodes and the set of lines of T, respectively.

5.3. *Resonances.* We call *resonance* a cluster T with only one entering line ℓ_T^2 such that

$$\sum_{v\in V(T)} \boldsymbol{\nu}_v = \mathbf{0}, \qquad \sum_{v\in V(T)} |\boldsymbol{\nu}_v| \le (2\,2^{(n+3)/\tau})^{-1}, \tag{5.10}$$

if n is the scale of the exiting line ℓ_T^1. Note that the entering line ℓ_T^2 must have, by the first condition in (5.10), the same momentum of the exiting line ℓ_T^1 and, by construction, a scale $n_{\ell_T^2} = n_{\ell_T^1} = n$.

We say that the line ℓ_T^1 exiting a resonance T is a *resonant line.* We call *nonresonant* line a line which is not a resonant line.

For any resonance T and any line $\ell \in \Lambda(T)$ one can write, by setting $\ell = \ell_v$,

$$\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_\ell^0 + \sigma_\ell \boldsymbol{\nu}, \tag{5.11}$$

where

$$\boldsymbol{\nu}_\ell^0 = \sum_{\substack{w\in V(T)\\ w\preceq v}} \boldsymbol{\nu}_w, \tag{5.12}$$

$\boldsymbol{\nu} \equiv \boldsymbol{\nu}_{\ell_T^2}$ is the momentum flowing through the line ℓ_T^2 entering T, and σ_ℓ is defined as follows: writing $\ell \equiv \ell_v$ then $\sigma_\ell = 1$ if ℓ_T^2 enters a node $w \preceq v$ and $\sigma_\ell = 0$ otherwise.

Given a resonance T, define the *resonance value* as

$$\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) = \left(\prod_{v\in V(T)} F_v\right)\left(\prod_{\ell\in\Lambda(T)} G_\ell^{(n_\ell)}\right), \tag{5.13}$$

seen as a function of $\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}$, if $\boldsymbol{\nu} \equiv \boldsymbol{\nu}_{\ell_T^2} = \boldsymbol{\nu}_{\ell_T^1}$ is the momentum flowing through the external lines of the resonance T.

We can have four types of resonances:

$$\begin{array}{lll} & \ell_T^1 & \ell_T^2 \\ 1. & H \leftarrow h & h \leftarrow H, \\ 2. & H \leftarrow h & H \leftarrow h, \\ 3. & h \leftarrow H & h \leftarrow H, \\ 4. & h \leftarrow H & H \leftarrow h. \end{array} \tag{5.14}$$

Given a tree θ, define

$$\begin{aligned} N_n(\theta) &= \{\ell \in \Lambda(\theta) \;:\; n_\ell = n\}, \\ p_n(\theta) &= \{T \subset T(\theta) \;:\; n_T = n\}. \end{aligned} \tag{5.15}$$

Call $N_n^*(\theta)$ the number of nonresonant lines on scale $\le n$ and call $R_n^j(\theta)$ the number of resonant lines on scale $\le n$ exiting from resonances of type j. Of course

$$N_n(\theta) = N_n^*(\theta) + \sum_{j=1}^{4} R_n^j(\theta). \tag{5.16}$$

In Appendix A1 we prove the following result (note that the bound (5.17) is a version of Siegel-Bryuno's lemma).

5.4. Lemma. *For any tree $\theta \in \mathcal{T}_{\kappa,\boldsymbol{\nu}}$ one has*

$$N_n^*(\theta) + p_n(\theta) \le c\,M(\theta)\,2^{n/\tau}, \tag{5.17}$$

and

$$R_n^4(\theta) \le c\,M(\theta)\,2^{n/\tau} + R_n^1(\theta), \tag{5.18}$$

for some constant c.

5.5. *Bound on the nonresonant lines.* Define $\Lambda^*(\theta)$ the set of nonresonant lines in $\Lambda(\theta)$. Then one has

$$\prod_{\ell\in\Lambda^*(\theta)} \left|G_\ell^{(n_\ell)}\right| \le \left(2C_0^{-1}\right)^{|\Lambda^*(\theta)|} \prod_{n=-\infty}^{1} 2^{-nN_n^*(\theta)}. \tag{5.19}$$

Let $n_0 = n_0(\kappa)$ be a negative integer to be fixed later (see (5.21) below). One has in (5.19)

$$\prod_{n=-\infty}^{1} 2^{-nN_n^*(\theta)} \le 2^{-2n_0k} \prod_{n=-\infty}^{n_0} 2^{-nN_n^*(\theta)} \le 2^{-2n_0k} \prod_{n=-\infty}^{n_0} 2^{-cnM(\theta)2^{n/\tau}}, \tag{5.20}$$

where (5.17) has been used for the lines on scale $\le n_0$. Choose n_0 so that

$$c\log 2 \sum_{p=|n_0|}^{\infty} p 2^{-p/\tau} \le \kappa_2, \tag{5.21}$$

for some $\kappa_2 \in (0, \kappa-\kappa_1)$; then (5.19) and (5.20) give

$$\prod_{\ell\in\Lambda^*(\theta)} \left|G_\ell^{(n_\ell)}\right| \le \left(2C_0^{-1}\right)^{|\Lambda^*(\theta)|} 2^{-2n_0k} e^{\kappa_2 M(\theta)}. \tag{5.22}$$

Together with the bounds (5.4) and (5.6), taking into account the product of the node factors, this shows that, by neglecting the resonances, a bound like (5.1) still holds with $\kappa' = \kappa - \kappa_1 - \kappa_2$. So we have to prove that the presence of the resonances does not destroy the results implied by the above discussion.

5.6. *Localization and renormalization operators.* For any resonance T we define

$$\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) = \mathcal{L}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) + \mathcal{R}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}), \tag{5.23}$$

where for resonances of type either 2 or 3 one has

$$\begin{aligned} \mathcal{L}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) &\equiv \mathcal{V}_T(0)\ , \\ \mathcal{R}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) &\equiv (\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) \int_0^1 \mathrm{d}t_T\, \partial\mathcal{V}_T(t_T\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}), \end{aligned} \tag{5.24}$$

while for resonances of type 1 one has

$$\begin{aligned} \mathcal{L}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) &\equiv \mathcal{V}_T(0) + (\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})\,\partial\mathcal{V}_T(0)\ , \\ \mathcal{R}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) &\equiv (\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})^2 \int_0^1 \mathrm{d}t_T\,(1-t_T)\,\partial^2\mathcal{V}_T(t_T\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}), \end{aligned} \tag{5.25}$$

and for resonances of type 4 one has simply

$$\begin{aligned} \mathcal{L}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) &\equiv 0\ , \\ \mathcal{R}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}) &= \mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}). \end{aligned} \tag{5.26}$$

Here $\partial\mathcal{V}_T$ and $\partial^2\mathcal{V}_T$ denote the first and the second derivatives of $\mathcal{V}_T$ with respect to its argument (see §3.2). We shall call $\mathcal{L}$ the *localization operator* and $\mathcal{R}$ the *renormalization operator*: correspondingly $\mathcal{L}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$ and $\mathcal{R}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$ are called the *localized part* and the *renormalized part* of the resonance value.

The quantity $\mathcal{V}_T(0)$ is obtained from $\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$ by replacing $\boldsymbol{\nu}_\ell$ with $\boldsymbol{\nu}_\ell^0$ in the argument of each propagator $G_\ell^{(n_\ell)}$, while $\partial\mathcal{V}_T(0)$ is obtained from $\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$ by deriving it with respect to $x = \boldsymbol{\omega}_0\cdot\boldsymbol{\nu}$, hence replacing $\boldsymbol{\nu}_\ell$ with $\boldsymbol{\nu}_\ell^0$ in the argument of each propagator $G_\ell^{(n_\ell)}$. Analogously $\partial\mathcal{V}_T(t_T\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$

and $\partial^2\mathcal{V}_T(t_T\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$ are obtained from $\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu})$ by deriving it with respect to $x=\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}$, once and twice, respectively, hence replacing $\boldsymbol{\nu}_\ell$ with $\boldsymbol{\nu}_\ell^0+t_T\sigma_\ell\boldsymbol{\nu}$ in the argument of each propagator $G_\ell^{(n_\ell)}$: in such a case we shall write

$$\boldsymbol{\nu}_\ell(t_T)=\nu_\ell^0+t_T\sigma_\ell\boldsymbol{\nu}, \tag{5.27}$$

where, as usual, we denote $\boldsymbol{\nu}=\boldsymbol{\nu}_{\ell_T^2}$. The expressions for the renormalized parts of the resonance values are explicitly written in (5.35) and (5.36) below.

Note that, given a resonance T, even if the renormalization procedure can change the compatible scales of its external lines, nevertheless the two scales have to remain equal to each other: in fact the momenta flowing through the external lines are still equal as their difference is left zero.

5.7. *Remark.* Because of the renormalization procedure it is no more true that there can be only one compatible scale per line (such that the corresponding propagator is not vanishing).

For instance, if T is a resonance and $\boldsymbol{\nu}_\ell$ is the momentum flowing through a line $\ell\in\Lambda(T)$, let n_ℓ be the scale compatible with ℓ before renormalizing the resonance, *i.e.* the scale n_ℓ such that $\chi_{n_\ell}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell)\neq 0$. In the localized part of the resonance value the momentum $\boldsymbol{\nu}_\ell$ has to be replaced with $\boldsymbol{\nu}_\ell^0$, which is in general different from $\boldsymbol{\nu}_\ell$. So it can happen that $\chi_{n_\ell}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell^0)=0$, while $\chi_{n'}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell^0)=0$ for some scale $n'\neq n_\ell$: in such a case the scale label compatible with the line is no more n_ℓ but n'.

Moreover in the renormalized part of the resonance value, even if the mode labels are fixed, the arguments of the propagators can change: they can assume any value reachable by varying $t_T\in[0,1]$, *i.e.* one has

$$\big||\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell^0|-|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}|\big|\le|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell|\le|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell^0|+|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}|\,, \tag{5.28}$$

so that (5.8) are (in principle) really infinite sums. The second condition in (5.10) has been introduced exactly with the aim of preventing the number of compatible scales from being too large: see §5.10 and the lemma 5.11 below.

5.8. *Resummation families.* The reason why to split the resonance values as in (5.21) is given by the fact that the contributions arising from the localized parts of the resonance values, when summed over all trees, give a vanishing contributions.

In order to prove such a (remarkable) property, we need to introduce suitable *resummation families.* Given a tree θ containing a resonance T, we can consider all trees obtained by changing the location of the nodes internal to T which the external lines of T are attached to: we denote by $\mathcal{F}_T(\theta)$ the set of trees so obtained, and call it the resummation family associated to the resonance T. And we shall refer to the operation of detaching and reattaching the external lines, by saying that we are *shifting* such lines. See Fig. 3.

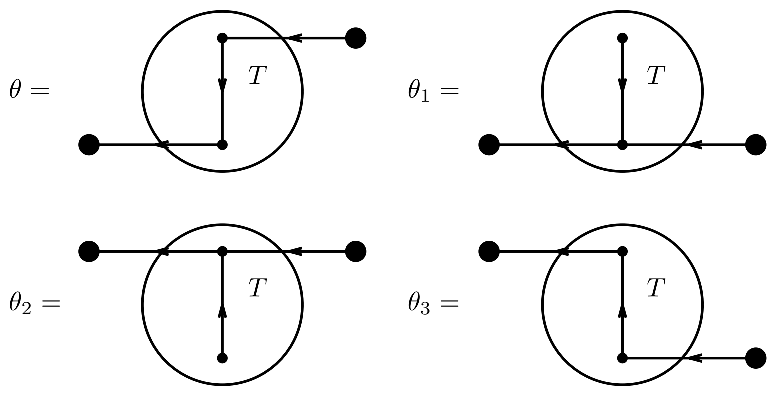

FIG. 3. The resummation family $\mathcal{F}_T(\theta)=\{\theta,\theta_1,\theta_2,\theta_3\}$ obtained by shifting the external lines of the resonance T. The black balls represent the remaining parts of the trees. The labels are not explicitly shown.

Of course shifting the external lines of a resonance produces a change of the propagators of the trees. In particular as all arrows have to point towards the root, some lines can revert their arrows: correspondingly some lines $h \leftarrow H$ become lines $H \leftarrow h$ and *vice versa*.

Moreover the momentum can change, as a reversal of the arrow implies a change of the partial ordering of the nodes inside the resonance and a shifting of the entering line can add or subtract the contribution of the momentum flowing through it. More precisely, if the external lines of a resonance T are detached then reattached to some other nodes in $V(T)$, the momentum flowing through the line $\ell \in \Lambda(T)$ can be changed into $\pm\boldsymbol{\nu}_\ell^0 + \sigma\boldsymbol{\nu}$, with $\sigma \in \{0,1\}$: if we call V_1 and V_2 the two disjoint sets into which ℓ divides T, such that the arrow superposed on ℓ is directed from V_2 to V_1 (before detaching the external lines), then the sign is $+$ if the exiting line is reattached to a node inside V_1 and it is $-$ otherwise, while $\sigma = 1$ if the entering line is reattached to a node inside V_2 when the sign is $+$ and to a node inside V_1 when the sign is $-$, and $\sigma = 0$ otherwise. See Fig. 4.

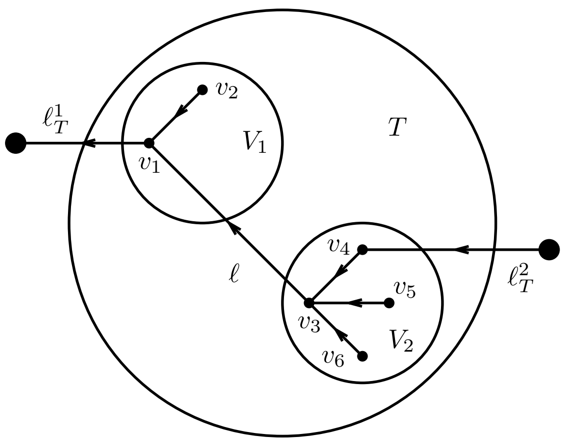

FIG. 4. The sets V_1 and V_2 in a resonance T; note that, even if they are drawn like circles, he sets V_1 and V_2 are not clusters. One has $\boldsymbol{\nu}_\ell^0 = \boldsymbol{\nu}_{v_3} + \boldsymbol{\nu}_{v_4} + \boldsymbol{\nu}_{v_5} + \boldsymbol{\nu}_{v_6}$ and $\boldsymbol{\nu} = \boldsymbol{\nu}_{\ell_T^2}$; of course $\boldsymbol{\nu}_{\ell_T^1} = \boldsymbol{\nu}_{\ell_T^2}$ and $\boldsymbol{\nu}_\ell^0 = -(\boldsymbol{\nu}_{v_1} + \boldsymbol{\nu}_{v_2})$ by definition of resonance. The black balls represent the remaining parts of the trees. The labels are not explicitly shown.

Then the following result follows. The proof is in Appendix A2.

5.9. LEMMA. *For any resonance $T \in T(\theta)$ one has*

$$\sum_{\theta' \in \mathcal{F}_T(\theta)} \mathcal{L}\mathcal{V}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}) = 0, \tag{5.29}$$

where $\boldsymbol{\nu} = \boldsymbol{\nu}_{\ell_T^2}$ and the sum is over the resummation family associated to T.

5.10. *Changing the scales.* When considering separately the localized parts and the renormalized parts of the resonance values, as we said before in the remark 5.7, the scales are no more uniquely fixed by assigning the mode labels.

This means that the sum over the scale labels is no more a fictitious sum as in the case in which no renormalization is performed (in such a case the scale associated to each line is simply determined by the momentum flowing through it by (5.9), so that no sum has to be really done).

Anyway the number of scale labels compatible with a line is not arbitrarily large, as the following result shows (the proof is in Appendix A3).

5.11. LEMMA. *When shifting the lines external to the resonances of a tree θ, for any line $\ell \in \Lambda(\theta)$*

the scale compatible with ℓ can change at most by one unit.

5.12. *Renormalization of the maximal resonances.* If we have several resonances which are external to each other then we apply the lemma 5.9 to each of them: so for all of them we can replace the resonance values with their renormalized parts.

The situation is a little more involved when one has to consider a tree θ in which some resonances are contained inside some other resonances. In such a case we define the *depth* $D(T)$ of a resonance T recursively as follows: given a resonance T, we set $D(T) = 1$ if there is no resonance containing T, and set $D(T) = D(T') + 1$ if T is contained inside a resonance T' and all the other resonances inside T' (if there are any) do not contain T.

Then consider the maximal resonances, *i.e.* the resonances $T \in T(\theta)$ with depth $D(T) = 1$. We call $T_1(\theta)$ the set of such resonances; likewise we call $T_D(\theta)$ the set of the resonances with depth D.

For each $T \in T_1(\theta)$ let $\mathcal{V}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_T^2})$ be its resonance value. By the lemma 5.9 we can neglect the localized part $\mathcal{LV}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_T^2})$, as it will give a vanishing contributions when the values of all trees are summed together, so that we have to consider only the renormalized value $\mathcal{RV}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_T^2})$: we say then that the resonance is a *renormalized resonance.*

Note that the scale compatible with the line ℓ_T^2 has not changed by the renormalization procedure (as the momentum flowing through the line entering the resonance remains the same), so that

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_2^T}| \le C_0 2^n, \tag{5.30}$$

if $n = n_{\ell_T^2}$, while, by shifting ℓ_T^2, the scales compatible with the lines internal to T can change at most by one unit (by the lemma 5.11).

Then for any line $\ell \in \Lambda(T)$ one has, setting $n = n_{\ell_T^2} = n_{\ell_T^1}$, and, by the second condition in (5.10),

$$|\boldsymbol{\nu}_\ell^0| \le \sum_{v \in V(T)} |\boldsymbol{\nu}_v| \le (2\, 2^{(n+3)/\tau})^{-1}, \tag{5.31}$$

so that

$$\left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| > C_0 |\boldsymbol{\nu}_\ell^0|^{-\tau} \ge C_0 2^\tau 2^{n+3}, \tag{5.32}$$

hence, with the notation (5.27),

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t_T)| > C_0 2^\tau 2^{n+3} - C_0 2^n > C_0 2^{n+3}, \tag{5.33}$$

so that the scales compatible with any line $\ell \in \Lambda(T)$ have to be strictly larger than n: the cluster structure imposed by the presence of the resonance is preserved by the renormalization procedure (recall that in order to define the resonances, even before stating the conditions (5.10), we have required them to be clusters!).

As all the lines internal to the resonances in $T_1(\theta)$ have a scale larger than the scales of the lines external to the resonances themselves, we can reason as in [BGGM1] (or [BGGM2]) and consider all the trees having the same structure as the just considered tree θ, but with different scale labels associated to the lines internal to the resonances in $T_1(\theta)$, *i.e.* the trees obtained by assigning to the lines in $\Lambda(T)$, for $T \in T_1(\theta)$, scale labels $n' \ge n + 1$, if $n = n_{\ell_T^2}$. In particular this means that all the considered tree θ' have the same sets of maximal resonances $T_1(\theta')$.

So we have the sum over the scale labels compatible with the resonance structure (see comments about (5.37) below) of renormalized resonance values which (setting $\boldsymbol{\nu}_{\ell_T}^2 = \boldsymbol{\nu}$) are given by

$$\mathcal{RV}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}) = \mathcal{V}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}), \tag{5.34}$$

for resonances of type 4, by

$$\begin{aligned} \mathcal{RV}_T(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}) = \sum_{\ell \in \Lambda(T)} \int_0^1 \mathrm{d}t_T \left(\prod_{v \in V(T)} F_v \right) \\ \left(\partial G_\ell^{(n_\ell)}(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t_T)) \prod_{\ell' \in \Lambda(T) \setminus \ell} G_{\ell'}^{(n_{\ell'})}(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell'}(t_T)) \right), \end{aligned} \tag{5.35}$$

for resonances of type 2 and 3, and by

$$
\begin{aligned}
\mathcal{R}\mathcal{V}_T(\boldsymbol{\omega}_0\cdot\ell) = & \sum_{\ell\in\Lambda(T)} \int_0^1 \mathrm{d}t_T\,(1-t_T) \left(\prod_{v\in V(T)} F_v\right) \\
& \left(\partial^2 G_\ell^{(n_\ell)}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell(t_T))\right) \left(\prod_{\ell'\in\Lambda(T)\setminus\ell} G_{\ell'}^{(n_{\ell'})}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_{\ell'}(t_T))\right) \\
& + \sum_{\ell\neq\ell'\in\Lambda(T)} \int_0^1 \mathrm{d}t_T\,(1-t_T) \left(\prod_{v\in V(T)} F_v\right) \\
& \left(\partial G_\ell^{(n_\ell)}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell(t_T)) \partial G_{\ell'}^{(n_{\ell'})}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_{\ell'}(t_T))\right) \\
& \left(\prod_{\ell''\in\Lambda(T)\setminus\{\ell,\ell'\}} G_{\ell'}^{(n_{\ell'})}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_{\ell''}(t_T))\right),
\end{aligned}
\tag{5.36}
$$

for resonances of type 1. Here we have explicitly written the argument of the propagators and the symbol ∂ is meant as the partial derivative with respect the argument (see §3.2). Note that we can write $\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell(t_T) = \boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell^0 + t_T\sigma_\ell\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}$, by (5.27).

The sum over the scale labels is such that for each line the characteristic functions χ_n reconstruct a step function: each line $\ell\in\Lambda(T)$ can have any scale $n'\geq n_T\geq n_{\ell_T^2}+1$, and

$$
\sum_{n'=n+1}^{1} \chi_{n'}(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell(t_T)) = \vartheta(|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}(t_T)| - C_0 2^n), \qquad n = n_{\ell_T^2}, \tag{5.37}
$$

where ϑ denotes the step function, *i.e.* $\vartheta(x)=1$ if $x>0$ and $\vartheta(x)=0$ otherwise.

For each summand contributing to the renormalized part of the resonance value there are either zero (see (5.34)) or one (see (5.35)) or two (see (5.36)) derived propagators.

5.13. *Remarks.* (1) First of all note that for any value of the interpolation parameter t_T the arguments of the step functions lay inside the region in which their arguments are positive, so that no contribution of the derivatives arise from them. In fact for any line $\ell\in\Lambda(T)$ one has (5.33), while the discontinuity of the theta function is at $C_0 2^n$ by (5.37).

(2) For any line $\ell\in\Lambda(T)$, if n is the scale compatible with it before renormalizing, *i.e.*

$$
C_0 2^{n-1} < |\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell| \leq C_0 2^n, \tag{5.38}
$$

then one has

$$
C_0 2^{n-2} < |\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell(t_T)| \leq C_0 2^{n+1}, \tag{5.39}
$$

by the lemma 5.11.

5.14. *Iterative renormalization of the resonances.* By using the remark 5.13, (1), we redecompose the step functions, so obtaining again characteristic functions. This means that we have to study expressions like (5.35) and (5.36), in which no derivative can acts on the characteristic functions (by the just given argument).

Consider explicitly the case in which only one propagator is derived, *i.e.* the case (5.35) of resonances of type 2 and 3. Then the derived propagator in (5.35) can correspond to a line ℓ contained inside some resonance $T'\subset T$ with depth $D(T')=2$. If this is the case we do not split the resonance value of T' into the sum of a localized and a renormalized part: we say then that such a resonance is not renormalized. Let T'' be the resonance with higher depth containing ℓ. Then we do not renormalize any resonances containing T'' and contained inside T. Then we pass to consider T'' and we repeat the above analysis, *i.e.* we apply once more the lemma 5.9, hence we study the renormalized value as in §5.12. The reason how we proceed in such a way will become clear later (see the remark 5.16 below).

Otherwise, if the line ℓ corresponding to the derived propagator is external to any resonances $T' \in T_2(\theta)$ we pass to consider such resonances and we iterate the above argument, *i.e.* we apply again the lemma 5.9, so getting rid of the localized parts of the resonances values for all resonances $T' \in T_2(\theta)$ such that $T' \subset T$ (which then become renormalized resonances), then, by summing over the scale labels, we group together all trees having the same cluster structure imposed by the set $T_2(\theta)$, and proceed as above.

The only differences with respect with the previous case are that now also the momentum flowing through the line $\ell^2_{T'}$ entering T' can have been changed into $\boldsymbol{\nu}_{\ell^2_{T'}}(t_T)$, and the momenta flowing through the lines internal to T' will depend in general on two interpolation parameters t_T and $t_{T'}$, one for each renormalized resonance.

By the remark 5.13, (2) one has that (5.30) has to be replaced with

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell^2_{T'}}(t_T)| \le C_0 2^{n+1}, \tag{5.40}$$

if $n = n_{\ell^2_{T'}} = n_{\ell^1_{T'}}$ (of course here$n$ has a different value with respect to n in (5.30), where one had $n = n_{\ell^2_T}$).

Moreover for any line $\ell \in \Lambda(T')$ the momentum flowing through it, setting $t \equiv \{t_T, t_{T'}\}$ can be written as

$$\boldsymbol{\nu}_\ell(t) = \boldsymbol{\nu}^0_\ell + t_{T'}\sigma_\ell \left(\boldsymbol{\nu}^0_{\ell^2_{T'}} + t_T \sigma_{\ell^2_{T'}} \boldsymbol{\nu}_{\ell^2_T} \right). \tag{5.41}$$

Then for any line $\ell \in \Lambda(T')$ one has, setting again $n = n_{\ell^2_{T'}}$ and using the second condition in (5.10),

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t_{T'})| > C_0 2^\tau 2^{n+3} - C_0 2^{n+1} > C_0 2^{n+3}, \tag{5.42}$$

so that the same conclusions as before can be drawn.

Note that with respect to the original tree it can happen that the renormalization procedure, by changing the scales compatible with the lines, make some resonances to disappear, while some new resonances can appear: recall that the definition of resonance depends on the scale of the resonant line. But this is not a problem at all: the bound on the number of nonresonant lines (which is of course equal to the number of resonances) is derived in Appendix A1 by using that if a line has a momentum $\boldsymbol{\nu}_\ell(t)$ and a scale n, then

$$C_0 2^{n-2} < |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)| \le C_0 2^{n+1}, \tag{5.43}$$

and such a bound is satisfied also for the trees with renormalized resonances by the lemma 5.11.

One deals in a similar way also with the resonances of type 1: in this case if both lines corresponding to the derived propagators are inside some resonance $T' \in T_2(\theta)$, then the resonance T' is not renormalized, if only one line is inside T' then T' is renormalized to one order less (*i.e.* to zeroth order if of type 2 or 3 and to first order if of type 1), while if both lines are external to T' then T' is renormalized to its proper order (*i.e.* to first order if of type 2 or 3 and to second order if of type 1).

5.15. *Final result of the renormalization procedure.* In this way we have obtained a sum of contributions such that for each of them the following situation arises for a given tree θ. We do not give the full details, as the analysis can be performed as in [BGGM1].

All localized parts of the resonance values cancel out, so that only renormalized parts have to be considered for the resonance values (by the lemma 5.9). As some resonances contained in the tree θ are renormalized (not all, as the discussion in §5.14 shows), we say then that the tree is a *renormalized tree*.

For each renormalized resonance T we have an interpolation parameter t_T. Defining the set interpolation parameters as

$$t = \{t_T \,:\, T \text{ is a renormalized resonance in } \theta\}, \tag{5.44}$$

the momenta of the lines $\ell \in \Lambda(\theta)$ become functions of t, $\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_\ell(t)$. The explicit dependence on such parameters is obtained as follows: if a line $\ell = \ell_v$ is contained inside the renormalized resonances $T_1 \subset T_2 \subset \ldots \subset T_p$ and $t_{T_1}, \ldots, t_{T_p}$ are the corresponding interpolation parameters, then

$$\boldsymbol{\nu}_\ell(t) = \sum_{w \preceq v} \boldsymbol{\nu}_w \prod_{T \in \{T_1, \ldots, T_p\} \,:\, \ell_w \in \Lambda(T)} t_T. \tag{5.45}$$

Of course (5.45) generalizes (5.41) to the case of more than two resonances contained inside each other. See Fig. 5 for an example.

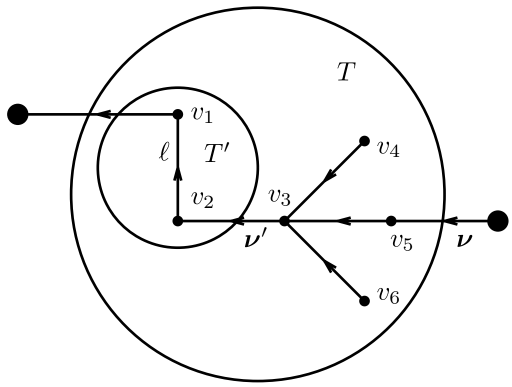

FIG. 5. The momenta as functions of the interpolation parameters in the case of two resonances $T' \subset T$. One has $\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_\ell^0 + t_{T'} \boldsymbol{\nu}'^0 + t_{T'} t_T \boldsymbol{\nu}'$, where $\boldsymbol{\nu}_\ell^0 = \boldsymbol{\nu}_{v_1} + \boldsymbol{\nu}_{v_2}$ and $\boldsymbol{\nu}'^0 = \boldsymbol{\nu}_{v_3} + \boldsymbol{\nu}_{v_4} + \boldsymbol{\nu}_{v_5} + \boldsymbol{\nu}_{v_6}$. The black balls represent the remaining parts of the trees. The labels are not explicitly shown.

Furthermore, by construction, each propagator is derived at most twice, and one has

$$\begin{aligned} \partial^p G_\ell^{(n_\ell)}(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)) &\sim (-1)^p \frac{p!}{(i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t))^{p+1}} \chi_{n_\ell}(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)) \\ &= (-1)^p \frac{p!}{(i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t))^p} G_\ell^{(n_\ell)}(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)), \end{aligned} \tag{5.46}$$

where $\sim$ means that the two quantities differ by the derivatives of the characteristic functions, which, however, have to be discarded in the valuation of the value of the renormalized tree (see the remark 5.13, (1)).

Note that

$$\left| \frac{1}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)} \right| \le C_0^{-1} 2^{n_T - 2}, \tag{5.47}$$

if T is the resonance with highest depth such that $\ell \in \Lambda(T)$, as the compatible scales of ℓ could have been changed at most by one unit with respect to that associated to ℓ before the renormalization procedure was applied, again by the lemma 5.11.

Then for each renormalized resonance, if $\boldsymbol{\nu}(t)$ denotes the momentum flowing through the entering line, we have an extra factor

$$\frac{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t)}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)} \, \frac{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t)}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}'_\ell(t)}, \tag{5.48}$$

(possibly times 2, when arising from a propagator derived twice; see (5.46) with $p = 2$) if T is of type 1, and an extra factor

$$\frac{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t)}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)}, \tag{5.49}$$

if T is of type 2 or 3. Here ℓ and ℓ' denote the lines corresponding to the propagators which have been derived by renormalizing T.

This means that for each resonance of type 2 and 3 one has a factor $i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t)$, which deletes the propagator of the corresponding resonant line. This does not happens for resonances of type 4; on the other hand one has a factor $(i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t))^2$ for resonances of type 1, and we can take advantage of such a fact through (5.18) (see (5.54) below).

As for any resonance T and for any line $\ell \in \Lambda(T)$ one has $|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)| > C_0 2^{n_T - 2}$ (see (5.47)) and $|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t)| \le C_0 2^{n+1}$, if $\boldsymbol{\nu}(t) = \boldsymbol{\nu}_{\ell_T^2}(t)$ and $n = n_{\ell_T^2}$, then (5.48) and (5.49) give a 'factor gain' Γ_T^2 or Γ_T, where

$$\Gamma_T = O(2^{n-n_T}). \tag{5.50}$$

This is evident for resonances which are renormalized, but it is not difficult to realize that it holds also for nonrenormalized resonances.

For simplicity (and for expository clarity) we explicitly discuss only the case in which only resonances of type 2 or 3 are involved, but the argument can be easily extended to cover also the case in which some resonances are of type 1.

Suppose that a resonance $T_1 \in \mathcal{T}_D(\theta)$, for some $D \ge 1$, is renormalized and that the derived propagator corresponds to a line $\ell \in \Lambda(T_p)$, with $T_p \in \mathcal{T}_{D+p}(\theta)$ such that T_p is the resonance with highest depth containing ℓ. Call $T_2, \ldots, T_{p-1}$ the (not renormalized) resonances such that $T_p \subset T_{p-1} \subset \ldots \subset T_2 \subset T_1$. If $\ell_{T_j}^2$ denotes the line entering the resonance T_j, for $j = 1, \ldots, p$, one has $\boldsymbol{\nu}(t) = \boldsymbol{\nu}_{\ell_{T_1}^2}(t)$ in (5.49). Then

$$\frac{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(t)}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)} = \left(\prod_{j=1}^{p-1} \frac{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_{T_j}^2}(t)}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_{T_{j+1}}^2}(t)} \right) \frac{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_{T_p}^2}(t)}{i\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell(t)}, \tag{5.51}$$

and, as each line $\ell_{T_{j+1}}^2$ is a line internal to the cluster T_j, hence on scale $\ge n_{T_j}$ we can bound (5.51) by

$$\prod_{j=1}^{p} O(2^{n_j - n_{T_j}}) = \prod_{j=1}^{p} \Gamma_{T_j}, \qquad n_j = n_{\ell_{T_j}^2}, \tag{5.52}$$

which proves the assertion.

5.16. *Remark.* Finally we have that, by construction, there are at most two derived propagators per cluster. It is exactly with this aim that no remormalization (or a renormalization to one order less) is performed on a cluster $T' \subset T$ when a derived line obtained by renormalizing T is contained inside T'. In fact if all resonances were renormalized we could have some lines derived arbitrarily many times, and this would give dangerous factorials (see (5.46)).

5.17. *Bound on the propagators.* In conclusion, as the effect of the renormalization procedure, we obtain a sum of terms each of which can be bounded as follows. If $\Lambda_1(\theta)$ and $\Lambda_2(\theta)$ denote the sets of lines corresponding to the propagators which are derived once and twice, respectively, we have that each term is bounded by a constant to the power k times

$$\begin{aligned} & C_0^{-k} \left(\prod_{\ell \in \Lambda_1(\theta)} 2^{-n_\ell} \right) \left(\prod_{\ell \in \Lambda_2(\theta)} 2^{-2n_\ell} \right) \\ & \qquad \left(\prod_{n=-\infty}^{1} 2^{-nN_n^*(\theta)} \right) \left(\prod_{n=-\infty}^{1} 2^{-nR_n^4(\theta)} \right) \left(\prod_{n=-\infty}^{1} 2^{nR_n^1(\theta)} \right), \end{aligned} \tag{5.53}$$

where the last two products can be bounded by using (5.18), *i.e.*

$$\left(\prod_{n=-\infty}^{1} 2^{-nR_n^4(\theta)} \right) \left(\prod_{n=-\infty}^{1} 2^{nR_n^1(\theta)} \right) \le \left(\prod_{n=-\infty}^{1} 2^{-cnM(\theta)\, 2^{(n+3)/\tau}} \right), \tag{5.54}$$

while the product taking into account the nonresonant lines can be bounded as in §5.5 (by using of course also the lemma 5.11)

Finally the first two products can be bounded by

$$\left(\prod_{\ell\in\Lambda_1(\theta)} 2^{-n_\ell}\right)\left(\prod_{\ell\in\Lambda_2(\theta)} 2^{-2n_\ell}\right) \le \prod_{n=-\infty}^{1} 2^{-2np_n(\theta)}, \tag{5.55}$$

as, as we noted, there are at most two derivatives per cluster.

So we are left with the problems of counting how many terms we have to sum over. More precisely we have to sum over the possible choices of the propagators to be derived times a sum over all the possible scale labels compatible with the lines.

The first sum is bounded by a constant to the power k as there can be at most two derived propagators per cluster (as already remarked above), while the second sum is bounded by $3^{|V(\theta)|} \le 3^k$: once the mode labels have been fixed for each line the scale label can assume only 3 values, by the lemma 5.11.

Then the bounds (5.1) are proven for the first two quantities $\mathbf{h}_{\boldsymbol{\nu}}^{(k)}$ and $\mathbf{H}_{\boldsymbol{\nu}}^{(k)}$.

5.18. *Remark.* As a matter of fact there are only 2 compatible scales per line. In fact the change of the momentum $\boldsymbol{\nu}(t)$ flowing through a line $\ell\in\Lambda(T)$ is such that at most $|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}|\le C_0 2^{n+1}$, if $n=n_{\ell_T^2}$, while the momentum was originally contained in an interval of width at least $C_0 2^{n+3}$: this means that $\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}(t)$ can fall in one of the two contiguous intervals $[C_0 2^{n'-1}, C_0 2^{n'})$, with $|n'-n|=1$, but not in both of them.

5.19. *Bounds for the counterterms.* Repeating the discussion of the previous sections for the counterterms we obtain a bound of the same kind, the only difference being that there is no factor C_0^{-1} associated to the root line (as there is no propagator corresponding to the root line): such a property will be used in §2.12 below.

So the proof of the lemma 2.2 is complete.

6. Conclusions and extensions of the results

6.1. *General perturbations and higher dimensions.* In this section we briefly review the possible generalizations and extensions of the results discussed in the previous sections. We confine ourselves to give some ideas how the proofs could be carried out, as we think that the main interest of the present paper relies on the techniques described in the previous sections rather than on the results (which are essentially well known from the standard KAM theory).

One can consider more general systems described by Hamiltonians of the form

$$\mathbf{H} = \boldsymbol{\omega}\cdot\mathbf{A} + f(\boldsymbol{\alpha},\mathbf{A},\varepsilon), \tag{6.1}$$

where $(\boldsymbol{\alpha},\mathbf{A})\in\mathbb{T}^d\times\mathbb{R}^d$, with $d\ge 2$, are conjugate variables, the rotation vector $\boldsymbol{\omega}$ satisfies the Diophantine condition

$$|\boldsymbol{\omega}\cdot\boldsymbol{\nu}| > C|\boldsymbol{\nu}|^{-\tau} \qquad \forall\boldsymbol{\nu}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}, \tag{6.2}$$

with Diophantine constants $C>0$ and $\tau>d-1$, and f is a real analytic function, holomorphic ic in a domain $\mathcal{D}=\Sigma_\kappa^{(d)}\times D\times B_{\varepsilon_1}(0)$, where $\Sigma_\kappa^{(d)}=\{\boldsymbol{\alpha}\in\mathbb{C}^d : \operatorname{Re}\alpha_j\in\mathbb{T},\ |\operatorname{Im}\alpha_j|<\kappa,\ j=1,\ldots,d\}$ and $D\subset\mathbb{C}^d$ is an open subset. With respect to the Hamiltonian (1.2) we allow the perturbation to depend on all the action variables and no restriction is made on the dimension d. Then one can ask if a result analogous to the theorem 1.6 holds for the Hamiltonian (6.1).

The reason why we considered the condition (2.26) for the Hamiltonian (1.2) is that the argument given in §2.14 can be repeated also for the Hamiltonian (6.1), with some minor (obvious) changes: this represents the first step in order to study the problem outlined in §1.10.

So, for the Hamiltonian (6.1), given $\overline{\mathbf{A}}\in D$, let ρ be such that $\mathcal{B}_\rho(\overline{\mathbf{A}})\subset D$. By setting

$$M \equiv \det\partial_{\mathbf{A}}^2 f^{(1)}(\overline{\mathbf{A}}) \equiv \det\int_{\mathbb{T}^2} \mathrm{d}\boldsymbol{\alpha}\,\partial_{\mathbf{A}}^2 f^{(1)}(\boldsymbol{\alpha},\overline{\mathbf{A}}) \ne 0, \tag{6.3}$$

then, if one fixes ε small enough (and nonvanishing), one can find infinitely many rotation vectors $\boldsymbol{\omega}_0$ close enough to $\boldsymbol{\omega}$ and satisfying the Diophantine condition

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| > C_0 |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}, \tag{6.4}$$

with $C_0 = bC$, for some constant b, and, for each of them, a value $\mathbf{A}' = \mathbf{A}'(\varepsilon) \in \mathcal{B}_\delta(\overline{\mathbf{A}})$, for a suitable $\delta < \rho' < \rho$, such that one has solutions of the form

$$\begin{cases} \boldsymbol{\alpha}(t) = \boldsymbol{\omega}_0 t + \mathbf{h}(\boldsymbol{\omega}_0 t, \mathbf{A}'(\varepsilon), \varepsilon, \boldsymbol{\omega}_0), \\ \mathbf{A}(t) = \mathbf{A}'(\varepsilon) + \mathbf{H}(\boldsymbol{\omega}_0 t, \mathbf{A}'(\varepsilon), \varepsilon, \boldsymbol{\omega}_0) \end{cases} \tag{6.5}$$

and $\mathbf{A}'(\varepsilon)$ verifies the equation

$$\boldsymbol{\omega}_0 + \boldsymbol{\eta}(\mathbf{A}'(\varepsilon), \varepsilon, \boldsymbol{\omega}_0) = \boldsymbol{\omega}, \tag{6.6}$$

and the functions $\mathbf{h}(\boldsymbol{\psi}, \mathbf{A}, \varepsilon, \boldsymbol{\omega}_0)$ and $\mathbf{H}(\boldsymbol{\psi}, \mathbf{A}, \varepsilon, \boldsymbol{\omega}_0)$ are analytic for $(\boldsymbol{\psi}, \mathbf{A}, \varepsilon) \in \Sigma_{\kappa'}^{(d)} \times \mathcal{B}_{\rho'}(\overline{\mathbf{A}}) \times \mathcal{B}_{\varepsilon_0}(0)$, with suitbale $\kappa' < \kappa$, $\rho' < \rho$ and $\varepsilon_0 < \varepsilon_1$. [Of course $\mathbf{A}'(\varepsilon)$ will depend also on $\boldsymbol{\omega}_0$ (even if we are not explicitly writing such a dependence).] Here $\boldsymbol{\eta}(\mathbf{A}'(\varepsilon), \varepsilon, \boldsymbol{\omega}_0)$ is the counterterm naturally arising when trying to look for solutions of the equations of motion with rotation vector $\boldsymbol{\omega}_0$ (see the lemma 2.2 and the remark 3.5).

The only difference with respect to the simplified problem studied in §2.14 is that now one has to prove that, given $\boldsymbol{\omega}$ verifying the Diophantine condition (6.2) then there exists $\overline{\varepsilon} \in \mathcal{B}_{\varepsilon_1}(0)$ such that, for all $|\varepsilon| < \overline{\varepsilon}$, any neighborhood of radius $O(\varepsilon)$ close enough to $\boldsymbol{\omega}$ contains infinitely many $\boldsymbol{\omega}_0$ satisfying the Diophantine condition (6.4): this can also be easily proved. Then one can reason as in §2.14, and a result analogous to the theorem 1.6 holds for the Hamiltonian (6.1).

6.2. *Measure of the persisting tori in phase space.* Under the condition

$$M \equiv \inf_{\mathbf{A} \in D} \left| \det \partial_{\mathbf{A}}^2 f_{\mathbf{0}}^{(1)}(\mathbf{A}) \right| > 0, \tag{6.7}$$

one can ask how many tori persist for perturbed Hamiltonian systems of the form (6.1).

If one wants to use the Lindstedt series instead of the usual KAM techniques (see [CG] and [Pö1]), one can reason in the following way.

For simplicity let us assume $\varepsilon > 0$ (small enough).

First of all note that it is straightforward to prove that if $\boldsymbol{\omega}$ satisfies the Diophantine condition (6.2), then, for any constant $a > 0$, the Lebesgue measure of the set $\mathcal{I}(\boldsymbol{\omega})$ of vectors $\boldsymbol{\omega}' \in \mathcal{B}_{a\varepsilon}(\boldsymbol{\omega})$ satisfying the Diophantine condition

$$|\boldsymbol{\omega}' \cdot \boldsymbol{\nu}| > C\varepsilon^\beta |\boldsymbol{\nu}|^{-\tau} \qquad \forall \boldsymbol{\nu} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}, \tag{6.8}$$

for some constant $\beta \in (0, 1)$, is given by

$$m(\mathcal{I}(\boldsymbol{\omega})) \geq m(\mathcal{B}_{a\varepsilon}(\boldsymbol{\omega}))\left(1 - c\varepsilon^\alpha\right), \qquad \alpha = \frac{\tau - d + 1}{\tau + 1} + \beta - 1, \tag{6.9}$$

for some constant c, so that $m(\mathcal{I}(\boldsymbol{\omega}))/m(\mathcal{B}_{a\varepsilon}(\boldsymbol{\omega})) \to 0$ for $\varepsilon \to 0$, provided that one has $\beta > d/(\tau + 1)$. The condition $\beta > d/(\tau + 1)$ assures that α in (6.9) is strictly positive, while the condition $\beta < 1$ is required as a consequence of the bound on the radius of convergence (see below); we shall see in §6.4 that in fact such a condition is highly improvable.

Then one can prove that, fixed ε small enough, most of invariant tori persist under perturbation, in the sense that the fraction of initial data in phase space for trajectories lying on invariant tori tends to 1 for $\varepsilon \to 0$. The discussion proceeds as follows.

Fix $\overline{\mathbf{A}} \in D$ and ρ such that $\mathcal{B}_\rho(\overline{\mathbf{A}}) \subset D$. Consider $\boldsymbol{\omega}' \in \mathcal{B}_{a\varepsilon}(\boldsymbol{\omega})$, for some constant a independent of ε (to be fixed), such that (6.8) is satisfied for $\boldsymbol{\omega}'$. Then the radius of convergence (in ε) of the series defining the functions $\mathbf{h}, \mathbf{H}, \eta$ is of the form $\varepsilon_0 = E_0 C_0 \varepsilon^\beta$ (simply use the extension of the lemma 2.2 discussed in §6.1, with C_0 replaced with $C_0\varepsilon^\beta$). As $\beta < 1$, if ε is small enough, one has

$\varepsilon < \varepsilon_0$, so that the series for $\mathbf{h}, \mathbf{H}, \eta$ converge for that value of ε and for $\mathbf{A} \in \mathcal{B}_{\rho'}(\overline{\mathbf{A}})$, for some $\rho' \in (0, \rho)$; in particular $\eta(\mathbf{A}, \varepsilon, \boldsymbol{\omega}')$ depends weakly on $\boldsymbol{\omega}'$, provided $\boldsymbol{\omega}' \in \mathcal{I}(\boldsymbol{\omega})$.

More precisely one can choose ε small enough and $\delta < \rho'$ so that, for all $\mathbf{A} \in \mathcal{B}_\delta(\overline{\mathbf{A}})$,

$$\boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}') = \varepsilon\, \boldsymbol{\eta}^{(1)}(\mathbf{A}) + \varepsilon^{1+\gamma} \boldsymbol{\xi}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}'), \qquad \sup_{\boldsymbol{\omega}' \in \mathcal{I}} \sup_{\mathbf{A} \in \mathcal{B}_\delta(\overline{\mathbf{A}})} |\boldsymbol{\xi}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}')| \le \Xi, \tag{6.10}$$

with $\gamma = 1 - \beta > 0$ and Ξ a suitable constant (we are using that the radius of convergence in ε for the counterterm is $O(\varepsilon^\beta)$).

By the analyticity in $\mathbf{A}$ of the counterterm, by (6.10) and by the condition (6.7) one has that, for ε small enough, for $\mathbf{A}, \mathbf{A}' \in \mathcal{B}_\delta(\mathbf{A})$ and for all $\boldsymbol{\omega}' \in \mathcal{I}$,

$$|\boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') - \boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}')| > \frac{M}{2}\, \varepsilon\, |\mathbf{A}' - \mathbf{A}|\,. \tag{6.11}$$

For $\mathbf{A} \in \mathcal{B}_{\delta/2}(\overline{\mathbf{A}})$ define $\boldsymbol{\omega}' \equiv \boldsymbol{\omega}'(\mathbf{A}) \equiv \boldsymbol{\omega} - \varepsilon\, \boldsymbol{\eta}^{(1)}(\mathbf{A})$. By varying $\mathbf{A} \in \mathcal{B}_{\delta/2}(\overline{\mathbf{A}})$, the corresponding values $\boldsymbol{\omega}'$ cover a set Ω, whose distance from $\boldsymbol{\omega}$ is of order ε: this follows from (6.7), as $\boldsymbol{\eta}^{(1)}(\mathbf{A}) = -\partial_{\mathbf{A}} f^{(1)}_{\mathbf{0}}(\mathbf{A})$. then we can suppose that a is such that $\Omega \subset \mathcal{B}_{a\varepsilon}(\boldsymbol{\omega})$.

Call $\Omega' \subset \Omega$ the set of vectors in Ω satisfying (6.8). By the condition (6.7) and by the bound (6.9) one has that the set of values $\mathbf{A} \in \mathcal{B}_{\delta/2}(\overline{\mathbf{A}})$ such that $\boldsymbol{\omega}'(\mathbf{A}) \in \Omega'$ has complement with measure bounded by $m(\mathcal{B}_{\delta/2}(\overline{\mathbf{A}}))\, d\varepsilon^\alpha$, for some constant d. Then one has, for $\mathbf{A} \in \mathcal{B}_{\delta/2}(\overline{\mathbf{A}})$ and $\mathbf{A}' \in \mathcal{B}_\delta(\overline{\mathbf{A}})$,

$$\begin{aligned} \boldsymbol{\omega}' + \boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') &= \boldsymbol{\omega} - \varepsilon\, \boldsymbol{\eta}^{(1)}(\mathbf{A}) + \boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') \\ &= \boldsymbol{\omega} - \varepsilon\, \boldsymbol{\eta}^{(1)}(\mathbf{A}) + \boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}') + \boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') - \boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}') \\ &= \boldsymbol{\omega} + \varepsilon^{1+\gamma} \boldsymbol{\xi}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}') + \boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') - \boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}'), \end{aligned} \tag{6.12}$$

where $\boldsymbol{\xi}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}')$ admits the bound in (6.10), while $\boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') - \boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}')$ is bounded from below through (6.11).

Choose β so that $\alpha > \gamma$ and a as said above: for ε small enough one has $\delta > 4\Xi\, \varepsilon^\gamma / M$, so that we conclude that, when varying $\mathbf{A}'$ in $\mathcal{B}_{\delta/2}(\mathbf{A}) \subset \mathcal{B}_\delta(\overline{\mathbf{A}})$, one reaches a value $\mathbf{A}'$ such that

$$\varepsilon^{1+\gamma}\, \boldsymbol{\xi}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}') + \boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') - \boldsymbol{\eta}(\mathbf{A}, \varepsilon, \boldsymbol{\omega}') = \varepsilon \left(\eta^{(1)}(\mathbf{A}') - \boldsymbol{\eta}^{(1)}(\mathbf{A}) \right) + \varepsilon^{1+\gamma} \boldsymbol{\xi}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') = 0 \tag{6.13}$$

so that (6.12) gives $\boldsymbol{\omega}' + \boldsymbol{\eta}(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') = \boldsymbol{\omega}$.

Of course the value $\mathbf{A}'$ depends on $\mathbf{A}$: we want to prove that if we take two different values $\mathbf{A}$ and $\tilde{\mathbf{A}}$ and we denote by $\mathbf{A}'$ and $\tilde{\mathbf{A}}'$, respectively, the corresponding values that one finds by following the above procedure, then one has

$$\mathbf{A}' - \tilde{\mathbf{A}}' = (1 + O(\varepsilon^\gamma)) \big(\mathbf{A} - \tilde{\mathbf{A}} \big) + O\big((\mathbf{A} - \tilde{\mathbf{A}})^2 \big). \tag{6.14}$$

To prove the last assertion we shall use that, to each perturbative order k and for all $\boldsymbol{\omega}', \boldsymbol{\omega}'' \in \mathcal{I}$, if $\partial_{\boldsymbol{\omega}} \boldsymbol{\eta}^{(k)}(\mathbf{A}, \boldsymbol{\omega}'')$ denotes the formal derivative of $\boldsymbol{\eta}^{(k)}(\mathbf{A}, \boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ computed in $\boldsymbol{\omega}''$, then one has

$$\boldsymbol{\eta}^{(k)}(\mathbf{A}, \boldsymbol{\omega}') - \boldsymbol{\eta}^{(k)}(\mathbf{A}, \boldsymbol{\omega}'') = \partial_{\boldsymbol{\omega}} \boldsymbol{\eta}^{(k)}(\mathbf{A}, \boldsymbol{\omega}'') \big(\boldsymbol{\omega}' - \boldsymbol{\omega}'' \big) + O\big(\boldsymbol{\omega}' - \boldsymbol{\omega}'' \big)^2, \tag{6.15}$$

uniformly in $\mathbf{A} \in \mathcal{B}_{\rho'}(\overline{\mathbf{A}})$. This follows from the analysis in [CG] (where in fact a much stronger result is proved, *i.e.* C^∞ differentiability in the sense of Whitney, [W]), and it can likely be proved also with the techniques used in the present paper; see §6.3 below.

Then choose two values $\mathbf{A}$ and $\tilde{\mathbf{A}}$, and call $\boldsymbol{\omega}'$ and $\tilde{\boldsymbol{\omega}}'$ the corresponding values $\boldsymbol{\omega}' = \boldsymbol{\omega} - \varepsilon\, \boldsymbol{\eta}^{(1)}(\mathbf{A})$ and $\tilde{\boldsymbol{\omega}}' = \boldsymbol{\omega} - \varepsilon\, \boldsymbol{\eta}^{(1)}(\tilde{\mathbf{A}})$: suppose that $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are both such that $\boldsymbol{\omega}', \tilde{\boldsymbol{\omega}}' \in \Omega'$. The values $\mathbf{A}'$ and $\tilde{\mathbf{A}}'$ are defined so that

$$\begin{aligned} &\Big(\boldsymbol{\eta}^{(1)}(\mathbf{A}') - \boldsymbol{\eta}^{(1)}(\mathbf{A}) \Big) + \varepsilon^\gamma \xi(\mathbf{A}', \varepsilon, \boldsymbol{\omega}') = 0, \\ &\Big(\boldsymbol{\eta}^{(1)}(\tilde{\mathbf{A}}') - \boldsymbol{\eta}^{(1)}(\tilde{\mathbf{A}}) \Big) + \varepsilon^\gamma \xi(\tilde{\mathbf{A}}', \varepsilon, \tilde{\boldsymbol{\omega}}') = 0, \end{aligned} \tag{6.16}$$

(see (6.13)). The difference between the two equations gives

$$\Big(\boldsymbol{\eta}^{(1)}(\mathbf{A}')-\boldsymbol{\eta}^{(1)}(\tilde{\mathbf{A}}')\Big)+\varepsilon^{\gamma}\Big(\boldsymbol{\xi}(\mathbf{A}',\varepsilon,\boldsymbol{\omega}')-\xi(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')\Big)=\Big(\boldsymbol{\eta}^{(1)}(\mathbf{A})-\boldsymbol{\eta}^{(1)}(\tilde{\mathbf{A}})\Big), \tag{6.17}$$

where one can write

$$\begin{aligned}\eta^{(1)}(\mathbf{A}')-\eta^{(1)}(\tilde{\mathbf{A}}')&=\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}'_*)\big(\mathbf{A}'-\tilde{\mathbf{A}}'\big),\\ \eta^{(1)}(\mathbf{A})-\eta^{(1)}(\tilde{\mathbf{A}})&=\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}_*)\big(\mathbf{A}-\tilde{\mathbf{A}}\big),\end{aligned} \tag{6.18}$$

and

$$\begin{aligned}\boldsymbol{\xi}(\mathbf{A}',\varepsilon,\boldsymbol{\omega}')-\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')&=\boldsymbol{\xi}(\mathbf{A}',\varepsilon,\boldsymbol{\omega}')-\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\boldsymbol{\omega}')+\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\boldsymbol{\omega}')-\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')\\ &=\partial_{\mathbf{A}}\boldsymbol{\xi}(\mathbf{A}'_\bullet,\varepsilon,\boldsymbol{\omega}')\big(\mathbf{A}'-\tilde{\mathbf{A}}'\big)+\partial_{\boldsymbol{\omega}}\,\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')\,(\boldsymbol{\omega}'-\tilde{\boldsymbol{\omega}}')+O\big((\boldsymbol{\omega}'-\tilde{\boldsymbol{\omega}}')^2\big),\end{aligned} \tag{6.19}$$

for suitable $\mathbf{A}'_*,\mathbf{A}_*,\mathbf{A}'_\bullet$ (for instance $\mathbf{A}'_*$ is a value between $\mathbf{A}'$ and $\tilde{\mathbf{A}}'$, and so on); by construction one has

$$\boldsymbol{\omega}'-\tilde{\boldsymbol{\omega}}'=\varepsilon\eta^{(1)}(\tilde{\mathbf{A}})-\varepsilon\eta^{(1)}(\mathbf{A})=\varepsilon\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}_*)\big(\tilde{\mathbf{A}}-\mathbf{A}\big), \tag{6.20}$$

and $\partial_{\boldsymbol{\omega}}\,\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')$ is the derivative (in the sense of Whitney) of $\boldsymbol{\xi}$ with respect to the third argument, according to (6.15).

Therefore (6.17) reads as

$$\begin{aligned}&\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}'_*)\big(\mathbf{A}'-\tilde{\mathbf{A}}'\big)+\varepsilon^{\gamma}\partial_{\mathbf{A}}\boldsymbol{\xi}(\mathbf{A}'_\bullet,\varepsilon,\boldsymbol{\omega}')\big(\mathbf{A}'-\tilde{\mathbf{A}}'\big)\\ &\quad+\varepsilon^{1+\gamma}\partial_{\boldsymbol{\omega}}\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}_*)\big(\tilde{\mathbf{A}}-\mathbf{A}\big)+O\big((\tilde{\mathbf{A}}-\mathbf{A})^2\big)=\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}_*)\big(\mathbf{A}-\tilde{\mathbf{A}}\big),\end{aligned} \tag{6.21}$$

so that we obtain (up to corrections $O((\tilde{\mathbf{A}}-\mathbf{A})^2)$)

$$\mathbf{A}'-\tilde{\mathbf{A}}'=\big(\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}'_*)+\varepsilon^{\gamma}\partial_{\mathbf{A}}\boldsymbol{\xi}(\mathbf{A}'_\bullet,\varepsilon,\boldsymbol{\omega}')\big)^{-1}\big(1+\varepsilon^{1+\gamma}\partial_{\boldsymbol{\omega}}\boldsymbol{\xi}(\tilde{\mathbf{A}}',\varepsilon,\tilde{\boldsymbol{\omega}}')\big)\partial_{\mathbf{A}}\eta^{(1)}(\mathbf{A}_*)\big(\mathbf{A}-\tilde{\mathbf{A}}\big), \tag{6.22}$$

which implies (6.14): hence the map $\mathbf{A}\to\mathbf{A}'$ is differentiable where it is defined (a property which indeed follows by the smoothness in the sense of Whitney).

As we have seen before that the set of values $\mathbf{A}\in\mathcal{B}_{\delta/2}(\overline{\mathbf{A}})$ for which $\boldsymbol{\omega}'=\boldsymbol{\omega}'(\mathbf{A})$ belongs to Ω' has measure bounded from below by $m(\mathcal{B}_\delta(\overline{\mathbf{A}}))\,(1-d\varepsilon^\alpha)$, the condition (6.14) implies that also the set of values $\mathbf{A}'$ representing the average values of the action variables along the tori with rotation vector in Ω' has measure bounded from below by $m(\mathbf{B}_\delta(\overline{\mathbf{A}}))\,(1-d'\varepsilon^\alpha)$, for some constant d'.

Therefore we can reason as follows. Fix $\delta_0=\overline{\delta}_0\varepsilon^\alpha$, for some constant $\overline{\delta}_0$ independent of ε, and define $\mathcal{A}$ as the set of values $\overline{\mathbf{A}}\in D$ such that $\mathcal{B}_{\delta+\delta_0}(\overline{\mathbf{A}})\subset D$; of course the complement of the set $\mathcal{A}$ has measure bounded by $m(D)\,c\varepsilon^\alpha$. For each of such $\overline{\mathbf{A}}$ we can repeat the above discussion, by choosing $\mathbf{A}=\overline{\mathbf{A}}$, so that we obtain the following result: the measure of the values $\mathbf{A}'\equiv\overline{\mathbf{A}}'$ which we find by solving the sequence of equations (6.13) – with $\mathbf{A},\mathbf{A}'$ replaced with $\overline{\mathbf{A}},\overline{\mathbf{A}}'$ – has complement (in D) with measure bounded by $m(D)\,c'\varepsilon^\alpha$, for some constant c' independent of ε.

By using the fact that the solution can be written in the form (6.5) (and recalling that $\mathbf{A}'(\varepsilon)$ depends also on $\boldsymbol{\omega}_0$) we can reasoning as above to prove that, for fixed ε, the actions $\mathbf{A}(t)$ are differentiable in $\mathbf{A}'=\mathbf{A}'(\varepsilon)$ (in the sense of Whitney) and with derivative close to 1, so that also the action variables corresponding to motions on invariant tori fulfill the set D up to a set of measure of order $O(\varepsilon^\alpha)$. This answers the question we asked above about the measure in phase space of the persisting invariant tori.

6.3. *Differentiability in the sense of Whitney.* The functions $\mathbf{h},\mathbf{H},\boldsymbol{\eta}$ are well defined for $\boldsymbol{\omega}'\in\mathcal{I}$ and admit bounds which are uniform for such $\boldsymbol{\omega}'$; they are known to be extendible to functions which are C^∞ in the sense of Whitney.

In particular they can be extended to continuous functions: this simply means that (6.15) as well as the equivalent expressions for the functions $\mathbf{h}$ and $\mathbf{H}$ hold for any $\boldsymbol{\omega}',\boldsymbol{\omega}''\in\mathcal{I}$. To prove this one has to compare the Lindstedt series expansions for $\boldsymbol{\eta}^{(k)}(\mathbf{A},\boldsymbol{\omega}')$ and $\boldsymbol{\eta}^{(k)}(\mathbf{A},\boldsymbol{\omega}'')$. Here we simply sketch how the analysis could be performed. One can write both $\boldsymbol{\eta}^{(k)}(\mathbf{A},\boldsymbol{\omega}')$ and $\boldsymbol{\eta}^{(k)}(\mathbf{A},\boldsymbol{\omega}'')$ as sums over trees (see (3.17)), so that the difference $\boldsymbol{\eta}^{(k)}(\mathbf{A},\boldsymbol{\omega}')-\boldsymbol{\eta}^{(k)}(\mathbf{A},\boldsymbol{\omega}'')$ can be written as sum of differences of pairs of tree values, computed the first with the rotation vector $\boldsymbol{\omega}'$ and the second with the rotation vector $\boldsymbol{\omega}''$. For the trees of each pair there is a difference between the propagators appearing in the corresponding values, while the remaining factors of the tree values are equal to each other. Therefore each difference can be decomposed as sum of values corresponding to trees whose lines ℓ have all propagators of the form either

$$\frac{1}{\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_\ell}, \tag{6.23}$$

or

$$\frac{1}{\boldsymbol{\omega}'' \cdot \boldsymbol{\nu}_\ell}, \tag{6.24}$$

up to one, say $\tilde{\ell}$, which has a new propagator given by the difference

$$\frac{1}{\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} - \frac{1}{\boldsymbol{\omega}'' \cdot \boldsymbol{\nu}_{\tilde{\ell}}}. \tag{6.25}$$

Furthermore one can arrange the decomposition in such a way that the set Λ' of the lines with propagator (6.23) is connected and contains the root line (except when $\tilde{\ell}$ is the root line: in such case there is no line with propagator of the form (6.23)), while the set Λ'' of the lines with propagator (6.24) is formed by disjoint sets $\Lambda''_1, \ldots, \Lambda''_N$, each of which is connected. In particular this means that, if for $j = 1, \ldots, N$ we call θ''_j the set of lines in Λ''_j and of the nodes connected by such lines, and, analogously, θ' the set of lines in Λ' and of the nodes connected by such lines, then each set θ''_j is a subtree which has as root either a node of θ' or the node which $\tilde{\ell}$ exit from; for each node v in θ' call $\theta_{j_{v1}}, \ldots, \theta_{j_{vN_v}}$ the subtrees which have v as root.

For the time being let us neglect the line $\tilde{\ell}$. If we denote by $\boldsymbol{\nu}_j$ the momentum flowing through the root line of θ''_j and by k_j its order, we see that when we sum over all the possible subtrees in $\mathcal{T}_{k_j,\boldsymbol{\nu}_j}$ we obtain a quantity which admits the bound (2.5), with $k = k_j$ and $\boldsymbol{\nu}_j$: in fact all the lines of such subtrees have a propagator of the form (6.24), so that the analysis of the previous sections apply. Then we can consider the set θ', which can be considered as a tree, provided that to each node v we replace the mode label $\boldsymbol{\nu}_v$ with $\boldsymbol{\nu}_v + \boldsymbol{\nu}_{j_{v1}} + \ldots + \boldsymbol{\nu}_{j_{vN_v}}$. Again if we sum over all the possible trees θ', we obtain a bound like (2.5).

The newly introduced propagator (6.25) can be written as

$$\frac{1}{\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} - \frac{1}{\boldsymbol{\omega}'' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} = \frac{(\boldsymbol{\omega}' - \boldsymbol{\omega}'')\,\boldsymbol{\nu}_{\tilde{\ell}}}{(\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_{\tilde{\ell}})(\boldsymbol{\omega}'' \cdot \boldsymbol{\nu}_{\tilde{\ell}})}, \tag{6.26}$$

when

$$|\boldsymbol{\nu}_{\tilde{\ell}}| < \left(\frac{C}{2|\boldsymbol{\omega}' - \boldsymbol{\omega}''|} \right)^{-1/(\tau+1)}, \tag{6.27}$$

while it can be only bounded with

$$\left| \frac{1}{\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} - \frac{1}{\boldsymbol{\omega}'' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} \right| \le \left| \frac{1}{\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} \right| + \left| \frac{1}{\boldsymbol{\omega}'' \cdot \boldsymbol{\nu}_{\tilde{\ell}}} \right|, \tag{6.28}$$

when $\boldsymbol{\nu}_{\tilde{\ell}}$ does not satisfy (6.27). In the first case we can add a line $\tilde{\ell}$ both to the set Λ' and to the set Λ'', and reason as above: simply the bound (5.17) has to be multiplied times a factor 2 as the line $\tilde{\ell}$ has to be counted twice. In the latter case one can think to use the exponential decay of the Fourier coefficients in order to obtain a bound small in $|\boldsymbol{\omega}' - \boldsymbol{\omega}''|$; note that there is only one line with propagator of the form (6.25), so that we can use, say, the square root of the product of the Fourier coefficients of all the nodes preceding the line $\tilde{\ell}$ in order to create a quantity bounded by $\exp[-\kappa|\boldsymbol{\nu}_{\tilde{\ell}}|/2]$.

Of course the above analysis is only heuristic: but we think that it can be made easily rigorous by using the analysis of sect. 5, with some minor changes.

With similar arguments we can prove differentiability to all orders, *i.e.* that the functions $\mathbf{h}, \mathbf{H}, \boldsymbol{\eta}$ are C^∞ in the sense of Whitney.

6.4. *Improving the results.* In all the paper we assumed $\boldsymbol{\omega}$ to be Diophantine. In that case the measure of the complement of the set of vectors $\boldsymbol{\omega}'$ whose invariant tori persist under the perturbation, for fixed ε is known to be exponentially small. This is obtained in the usual KAM theory by performing suitably many Birkhoff transformations before applying the KAM theorem. To see such a feature with the Lindstedt series one can reason in a way similar to what was done in [GGM] about the problem of the persistence of KAM tori for three time scale systems.

Again we shall proceed at a heuristic level. Suppose for simplicity f to be a trigonometric polynomial in $\boldsymbol{\alpha}$ of order N, so that to order k one has $|\boldsymbol{\nu}_\ell| \le kN$ for all $\ell \in \Lambda(\theta)$ and for all trees θ of order k. By setting $\boldsymbol{\omega}' = \boldsymbol{\omega} + \tilde{\boldsymbol{\omega}}$, with $|\tilde{\boldsymbol{\omega}}| < a\varepsilon$, one has

$$|\boldsymbol{\omega}' \cdot \boldsymbol{\nu}| \ge |\boldsymbol{\omega} \cdot \boldsymbol{\nu}| - |\tilde{\boldsymbol{\omega}} \cdot \boldsymbol{\nu}| \ge \frac{C}{2} |\boldsymbol{\nu}|^{-\tau} , \tag{6.29}$$

for all $\boldsymbol{\nu}$ such that

$$|\boldsymbol{\nu}| \le N_0 \equiv \left(\frac{C}{2a\varepsilon} \right)^{1/(\tau+1)} , \tag{6.30}$$

So for all k such that $k < k_0 \equiv N_0/N$ one has that (6.27) applies in order to bound all propagators, and only to order $k = k_0$ a line ℓ such that one has to use the bound

$$|\boldsymbol{\omega}' \cdot \boldsymbol{\nu}_\ell| > C\varepsilon^\beta |\boldsymbol{\nu}_\ell|^{-\tau}, \tag{6.31}$$

can really appear. To such an order k_0, by summing over all trees $\theta \in \mathcal{T}_{k_0,\boldsymbol{\nu}}$, one obtains the bound

$$B^{k_0} C^{-k_0} \varepsilon^{-\beta}, \tag{6.32}$$

for some constant B, and this suggests that to any order k one has the bound

$$\left(BC^{-1} \varepsilon^{-\beta N/N_0} \right)^k , \tag{6.33}$$

which requires

$$B\, C^{-1} \varepsilon^{-\beta N/N_0} \varepsilon < 1, \tag{6.34}$$

that is $\beta N/N_0 < 1$. The proof of (6.33) can be performed by reasoning as in [GGM]. Therefore we can take

$$\beta \le \beta_0 = \frac{N_0}{2N} = c \left(\frac{1}{\varepsilon} \right)^{1/(\tau+1)} , \tag{6.35}$$

for some constant c. Then, by using the analysis of sect. 5, the discussion could be extended to analytic perturbations which are not necessarily trigonometric polynomials: it would be interesting to do so, and to compare the results arising by using the Lindstedt series (in particular the exponent $1/(\tau+1)$) with the ones known from the usual KAM theory.

Acknowledgments. We are indebted to L. Chierchia and especially to G. Gallavotti for many discussions.

Appendix A1. Proof of the lemma 5.4

.

A1.1. *Remark.* In the following we shall use that if a line ℓ is on a scale n then

$$C_0 2^{n-2} \le |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| \le C_0 2^{n+1}. \tag{A1.1}$$

Without considering the renormalization procedure then (5.9) holds, and it implies (A1.1). We shall use (A1.1) instead of (5.9), as we have seen that the renormalization procedure can make the number of scales compatible with some lines to increase (see §5.10), but, by the lemma 5.11, the change of the compatible scales with respect to the scale originally (*i.e.* before renormalizing) associated to the line, can be at most by one unit, *i.e.* n_ℓ can be changed into $n'_\ell = n_\ell \pm 1$, so that (A1.1) will continue to hold also when the renormalization procedure is applied. This means that all the following analysis will be still valid for renormalized trees (as it has to be).

A1.2. *Proof of the lemma 5.4.* We prove inductively on the number of nodes of the trees the bounds

$$\begin{aligned} N_n^*(\theta) &\le \max\{0, 2\,M(\theta)\,2^{(n+3)/\tau} - 1\}, \\ p_n(\theta) &\le \max\{0, 2\,M(\theta)\,2^{(n+3)/\tau} - 1\}, \\ R_n^4(\theta) &\le \max\{0, 2\,M(\theta)\,2^{(n+3)/\tau} - 1 + R_n^1(\theta)\}, \end{aligned} \tag{A1.2}$$

where $M(\theta)$ is defined in (5.3). The proof is inspired by [Pö2] (see also [G2]), and gives (5.14) and (5.15) with $c = 2^{2+3/\tau}$.

First of all note that if $M(\theta) < 2^{-(n+3)/\tau}$ then $N_n(\theta) = 0$ as in such a case for any line $\ell \in \Lambda(\theta)$ one has

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| > C_0 |\boldsymbol{\nu}_\ell|^{-\tau} > C_0 M(\theta)^{-\tau} > C_0 2^{n+3}, \tag{A1.3}$$

by the Diophantine hypothesis (2.2).

A1.3. *Bound on $N_n^*(\theta)$.* If θ has only one node the bound is trivially satisfied as, if v is the only node in $V(\theta)$, one must have $M(\theta) = |\boldsymbol{\nu}_v| \ge 2^{-(n+3)/\tau}$ in order that the line exiting from v be on scale $\le n$: then $2\,M(\theta)\,2^{(n+3)/\tau} \ge 2$.

If θ is a tree with $V > 1$ nodes, we assume that the bound holds for all trees having $V' < V$ nodes. Define $E_n = (2\,2^{(n+3)/\tau})^{-1}$: so we have to prove that $N_n^*(\theta) \le \max\{0, M(\theta)\,E_n^{-1} - 1\}$.

If the root line ℓ of θ is either on scale $> n$ or resonant and on scale $\le n$, call $\theta_1, \ldots, \theta_m$ the $m \ge 1$ subtrees entering the last node of θ and with $M(\theta_i) \ge 2^{-(n+3)/\tau}$; see Fig. 6. Then

$$N_n^*(\theta) = \sum_{i=1}^m N_n^*(\theta_i), \tag{A1.4}$$

hence the bound follows by the inductive hypothesis.

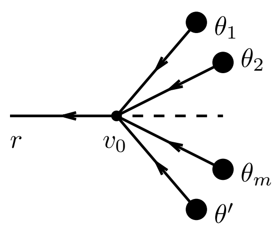

FIG. 6. A tree θ consisting of a node v_0 and m subtrees $\theta_1,\ldots,\theta_m$ entering v_0. The subtrees are represented by black balls. The subtree θ' has $M(\theta') < 2^{-(n+3)/\tau}$. The labels are not explicitly shown.

If the root line ℓ is nonresonant and on scale $\le n$, call $\ell_1, \ldots, \ell_m$ the $m \ge 0$ lines on scale $\le n$ which are the nearest to ℓ (this means that no other line along the paths connecting the lines $\ell_1, \ldots, \ell_m$ to the root line is on scale $\le n$). Note that in such a case $\ell_1, \ldots, \ell_m$ are the entering line of a cluster T on scale $> n$; see Fig. 7.

One has

$$N_n^*(\theta) = 1 + \sum_{i=1}^m N_n^*(\theta_i), \tag{A1.5}$$

so that the bound becomes trivial if either $m = 0$ or $m \ge 2$.

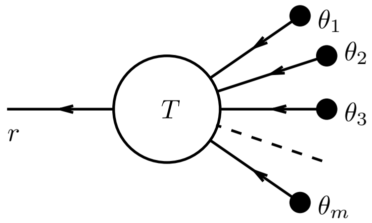

FIG. 7. A tree θ consisting of a line (root line) exiting from a cluster T with m entering lines ℓ_1,...,ℓ_m. Each of such lines is the root line of a subtree which is represented as a black ball. The labels are not explicitly shown.

If $m = 1$ then $T = \theta \setminus \theta_1$, ℓ and ℓ_1 are both on scales $\le n$ and ℓ_1 is not entering a resonance, so that

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| \le C_0 2^{n+1}, \qquad |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_{\ell_1}| \le C_0 2^{n+1}, \tag{A1.6}$$

and either $\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_{\ell_1}$ and one has

$$\sum_{v \in V(T)} |\boldsymbol{\nu}_v| > (2\, 2^{(n+3)/\tau})^{-1} = E_n, \tag{A1.7}$$

or $\boldsymbol{\nu}_\ell \ne \boldsymbol{\nu}_{\ell_1}$, otherwise T would be a resonance. If $\boldsymbol{\nu}_\ell \ne \boldsymbol{\nu}_{\ell_1}$, then, by (A1.6) one has $|\boldsymbol{\omega}_0 \cdot (\boldsymbol{\nu}_\ell - \boldsymbol{\nu}_{\ell_1})| \le C_0 2^{n+2}$, which, by the Diophantine condition (2.2), implies $|\boldsymbol{\nu}_\ell - \boldsymbol{\nu}_{\ell_1}| > 2^{-(n+2)/\tau}$, so that again

$$\sum_{v \in V(T)} |\boldsymbol{\nu}_v| \ge |\boldsymbol{\nu}_\ell - \boldsymbol{\nu}_{\ell_1}| > 2^{-(n+2)/\tau} > (2\, 2^{(n+3)/\tau})^{-1} = E_n, \tag{A1.8}$$

as in (A1.7). Therefore in both cases we get

$$M(\theta) - M(\theta_1) = \sum_{v \in T} |\boldsymbol{\nu}_v| > E_n, \tag{A1.9}$$

which, inserted into (A1.5) with $m = 1$, gives, by using the inductive hypothesis,

$$\begin{aligned} N_n^*(\theta) &= 1 + N_n^*(\theta_1) \le 1 + M(\theta_1)\, E_n^{-1} - 1 \\ &\le 1 + (M(\theta) - E_n)\, E_n^{-1} - 1 \le M(\theta)\, E_n^{-1} - 1, \end{aligned} \tag{A1.10}$$

hence the bound is proved also if the root line is nonresonant and on scale $\le n$.

A1.4. *Bound on $p_n(\theta)$.* The bound is trivial if $M(\theta) < 2^{(n+3)/\tau}$, as $N_n(\theta) = 0$ in such a case. Otherwise one can reason as follows.

If the last node v_0 of θ is not in a cluster on scale n one has

$$p_n(\theta) = \sum_{i=1}^{m} p_n(\theta_i), \tag{A1.11}$$

if $\theta_1, \ldots, \theta_m$ are the subtrees entering v_0 and with $M(\theta_i) \ge 2^{-(n+3)/\tau}$, so that the bound follows from the inductive hypothesis.

If the last node v_0 of θ is inside a cluster T on scale n one has

$$p_n(\theta) = 1 + \sum_{i=1}^{m_T} p_n(\theta_i), \tag{A1.12}$$

where now $\theta_1, \ldots, \theta_{m_T}$ are the subtrees entering the cluster T. The only nontrivial case if the one with $m_T = 1$: in such a case there is only one line ℓ_1 entering the cluster T, and its scale is strictly smaller than n, *i.e.* $n_{\ell_1} \le n - 1$. But then one has

$$\sum_{v \in V(T)} |\boldsymbol{\nu}_v| > 2^{-(n+3)/\tau}, \tag{A1.13}$$

as we are going to show.

First note that T has to contain at least one line on scale n. Call ℓ such a line: then

$$C_0 2^{n-2} \left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell\right| \le C_0 2^{n+1}, \tag{A1.14}$$

by (A1.1). We can write $\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_\ell^0 + \sigma_\ell \boldsymbol{\nu}_1$, where $\boldsymbol{\nu}_\ell^0$ is the sum of the mode labels of the nodes w inside the resonance such that $w \preceq v$, if $\ell = \ell_v$, and $\boldsymbol{\nu}_1$ is the momentum flowing through ℓ_1 (see (5.11) and (5.12)). Therefore if (A1.13) did not hold one would have

$$\left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| \ge C_0 |\boldsymbol{\nu}_\ell^0|^{-\tau} \ge C_0 \Big(\sum_{v \in V(T)} |\boldsymbol{\nu}_v| \Big)^{-\tau} > C_0 2^{n+3}, \tag{A1.15}$$

and, as

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_1| \le C_0 2^{n_{\ell_1}+1} \le C_0 2^n, \tag{A1.16}$$

one would have

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| \ge \left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| - |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_1| \ge C_0 2^{n+3} - C_0 2^n \ge C_0 2^{n+2}, \tag{A1.17}$$

which is not consistent with (A1.14).

Then note that (A1.13) implies $M(\theta_1) \le M(\theta) - E_n$, so that the bound follows.

A1.5. *Bound on* $R_n^4(\theta)$. If θ has only one node the bound is trivially satisfied.

Let θ be a tree with $V > 1$ nodes and suppose that the bound on $R_n^4(\theta)$ holds for all trees with $V' < k$ nodes.

If the root line ℓ of θ is not the exiting line of a resonance or it is the exiting line of a resonance of type either 1 or 2 or 3, then call $\theta_1, \ldots, \theta_m$ the $m \ge 1$ subtrees entering the last node of θ and with $M(\theta_i) \ge 2^{-(n+3)/\tau}$. By the inductive hypothesis one has

$$R_n^4(\theta_i) \le 2\, M(\theta_i)\, 2^{(n+3)/\tau} - 1 + R_n^1(\theta_i), \tag{A1.18}$$

for all $i = 1, \ldots, m$. Moreover

$$R_n^1(\theta) = \sum_{i=1}^m R_n^1(\theta_i), \tag{A1.19}$$

if the line ℓ does not exit from a resonance of type 1, while

$$R_n^1(\theta) = 1 + \sum_{i=1}^m R_n^1(\theta_i) \tag{A1.20}$$

otherwise, so that

$$\begin{aligned} R_n^4(\theta) &= \sum_{i=1}^m R_n^4(\theta_i) \le 2 \Big(\sum_{i=1}^m M(\theta_i) \Big) 2^{(n+3)/\tau} - m + R_n^1(\theta) \\ &\le 2\, M(\theta) 2^{(n+3)/\tau} - 1 + R_n^1(\theta), \end{aligned} \tag{A1.21}$$

as $m \ge 1$.

If the root line ℓ is the exiting line of a resonance T of type 4, then ℓ is a line $h \leftarrow H$ and there is only one subtree θ_0 entering T. One has

$$\begin{aligned} R_n^1(\theta) &= R_n^1(\theta_0), \\ R_n^4(\theta) &= 1 + R_n^4(\theta_0), \end{aligned} \tag{A1.22}$$

and the root line ℓ_0 of θ_0 is a line $H \leftarrow h$ with momentum $\boldsymbol{\nu}_{\ell_0}$ equal to $\boldsymbol{\nu}_\ell$ and on scale $\le n$ (as T is a resonance of type 4).

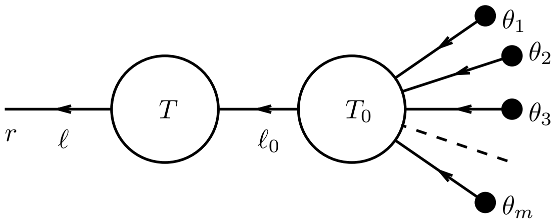

FIG. 8. A tree θ consisting of a line (root line) exiting from a resonance T of type 4 whose entering line ℓ_0 is the exiting line of a cluster T_0 with m entering lines ℓ_1,...,ℓ_m. Each of such lines is the root line of a subtree which is represented as a black ball. The labels are not explicitly shown.

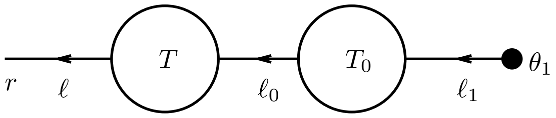

FIG. 9. A tree θ consisting of a line (root line) exiting from a resonance T of type 4 whose entering line ℓ_0 is the exiting line of a cluster T_0 with only one entering line ℓ_1. The subtree θ_1 having ℓ_1 as root line is represented as a black ball. The labels are not explicitly shown.

Call $\ell_1, \ldots, \ell_m$ the $m \ge 0$ lines on scale $\le n$ preceding ℓ_0 and which are the nearest to ℓ_0: by construction they enter a cluster T_0 having ℓ_0 as exiting line; see Fig. 8.

If either $m = 0$ or $m \ge 2$ the bound follows easily from the inductive hypothesis. If $m = 1$, denote by ℓ_1 the line entering T_0; see Fig. 9.

If $\boldsymbol{\nu}_{\ell_1} = \boldsymbol{\nu}_{\ell_0}$ and

$$\sum_{v \in V(T_0)} |\boldsymbol{\nu}_v| < E_n, \tag{A1.23}$$

then T_0 is a resonance, and in such a case it is a resonance of type 1 if ℓ_1 is a line $h \leftarrow H$ and of type 2 if ℓ_1 is a line $H \leftarrow h$; otherwise T_0 is not a resonance.

In the latter case (*i.e.* if T_0 is not a resonance) one has

$$\sum_{v \in V(T_0)} |\boldsymbol{\nu}_v| > E_n, \tag{A1.24}$$

which is obvious if $\boldsymbol{\nu}_{\ell_1} = \boldsymbol{\nu}_{\ell_0}$ (by definition of resonance) and follows from the Diophantine condition if $\boldsymbol{\nu}_{\ell_1} \neq \boldsymbol{\nu}_{\ell_0}$ (see (A1.8); then if T_0 is not a resonance one has

$$M(\theta) - M(\theta_1) \ge M(\theta_0) - M(\theta_1) > E_n. \tag{A1.25}$$

Note also that if T_0 is not a resonance then

$$\begin{aligned} R^1_n(\theta) &= R^1_n(\theta_0) = R^1_n(\theta_1), \\ R^4_n(\theta) &= 1 + R^4_n(\theta_0) = 1 + R^4_n(\theta_1), \end{aligned} \tag{A1.26}$$

so that, by the inductive hypothesis (applied to θ_1) and by (A1.25), one has

$$\begin{aligned} R^4_n(\theta) &\le 1 + M(\theta_1)\, E_n^{-1} - 1 + R^1_n(\theta_1) \\ &\le 1 + (M(\theta) - E_n)\, E_n^{-1} - 1 + R^1_n(\theta) \le M(\theta)\, E_n^{-1} - 1 + R^1_n(\theta), \end{aligned} \tag{A1.27}$$

so that the assertion is proved.

If T_0 is a resonance of type 1, then

$$\begin{aligned} R^1_n(\theta) &= R^1_n(\theta_0) = 1 + R^1_n(\theta_1), \\ R^4_n(\theta) &= 1 + R^4_n(\theta_0) = 1 + R^4_n(\theta_1), \end{aligned} \tag{A1.28}$$

so that, by the inductive hypothesis (applied to θ_1) one has

$$\begin{aligned} R^4_n(\theta) = 1 + R^4_n(\theta_1) &\le 1 + \Big(2\, M(\theta_1) 2^{(n+3)/\tau} - 1 + R^1_n(\theta_1) \Big) \\ &\le 2\, M(\theta_1) 2^{(n+3)/\tau} - 1 + R^1_n(\theta) \le 2\, M(\theta) 2^{(n+3)/\tau} - 1 + R^1_n(\theta), \end{aligned} \tag{A1.29}$$

and the bound follows.

Finally, if T_0 is a resonance of type 2, then

$$\begin{aligned} R^1_n(\theta) &= R^1_n(\theta_0) = R^1_n(\theta_1), \\ R^4_n(\theta) &= 1 + R^4_n(\theta_0) = 1 + R^4_n(\theta_1), \end{aligned} \tag{A1.30}$$

and the line ℓ_1 entering T_0 is again a line $H \leftarrow h$ on scale $\le n$, so that one can repeat for θ_1 the above argument (simply ℓ_1 plays the rôle of ℓ_0) and, if needed, we iterate further the analysis until one of the two following possibilities arise: either one obtains a case for which the bound follows or one reaches a tree $\tilde\theta$ having only a node, so not containing any resonances at all (in particular no other resonances of type 1 and 4). Note that if the first possibility never arise, then the second one sooner or later is unavoidable as at each iterative step the number of the nodes is decreased. In such a case $R^4_n(\theta) = R^4_n(\tilde\theta) + 1 = 1$ and $R^1_n(\theta) = R^1_n(\tilde\theta) = 0$, and the bound holds as $M(\theta)\, 2^{(n+3)/\tau} \ge 1$ in order that there be at least one line on scale $\le n$ in θ. So at the end the assertion follows in all cases.

Appendix A2. Proof of the lemma 5.9

A2.1. *Introduction.* We consider separately the first three types of resonances (for the type 4 there is nothing to prove). As in §3 we ignore the problem of the factorials, which, again, is solved by reasoning as in [BG].

A2.2. *Resonance of type 1.* First we prove that $\mathcal{V}_T(0) = 0$. Given a tree θ consider all trees which can be obtained by shifting the entering line ℓ^2_T; see Fig. 10. Note that the trees so obtained are contained in the resummation family $\mathcal{F}_T(\theta)$ introduced in §5.8.

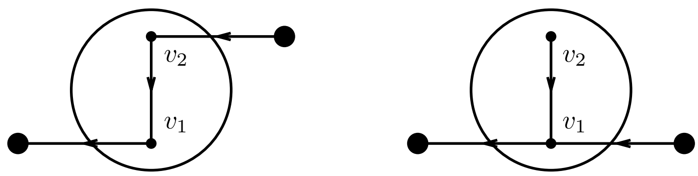

FIG. 10. The trees obtained by shifting the line entering a two-node resonance. The black balls represent the remaining parts of the trees. The labels are not explicitly shown.

Corresponding to such an operation $\mathcal{V}_T(0)$ changes by a factor $i\boldsymbol{\nu}_v$ if v is the node which the entering line is attached to, as all node factors and propagators do not change. By (5.10) the sum of all such values is zero.

Then consider $\partial\mathcal{V}_T(0)$. By construction

$$\partial\mathcal{V}_T(0) = \sum_{\ell\in\Lambda(T)} \left(\prod_{v\in V(T)} F_v \right) \left(\partial G^{(n_\ell)}_\ell \prod_{\ell'\in\Lambda(T)\setminus\ell} G^{(n_{\ell'})}_{\ell'} \right), \tag{A2.1}$$

where all propagators have to be computed for $\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu} = 0$, and

$$\partial G_\ell^{(n_\ell)} = \frac{\mathrm{d}}{\mathrm{d}x} G_\ell^{(n_\ell)}(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0 + x)\bigg|_{x=0}, \qquad x = \boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}. \tag{A2.2}$$

The line ℓ divides $V(T)$ into two disjoint set of nodes V_1 and V_2, such that ℓ_T^1 exits from a node inside V_1 and ℓ_T^2 enters a node inside V_2: if $\ell = \ell_v$ one has $V_2 = \{w \in (T) : w \preceq v\}$ and $V_1 = V(T) \setminus V_2$; see Fig. 4. By (5.10) if

$$\boldsymbol{\nu}_1 = \sum_{v \in V_1} \boldsymbol{\nu}_v, \qquad \boldsymbol{\nu}_2 = \sum_{v \in V_2} \boldsymbol{\nu}_v, \tag{A2.3}$$

one has $\boldsymbol{\nu}_1 + \boldsymbol{\nu}_2 = \mathbf{0}$. Then consider the families $\mathcal{F}_1(\theta)$ and $\mathcal{F}_2(\theta)$ of trees obtained as follows: $\mathcal{F}_1(\theta)$ is obtained from θ by detaching ℓ_T^1 then reattaching to all the nodes $w \in V_1$ and by detaching ℓ_T^2 then reattaching to all the nodes $w \in V_2$, while $\mathcal{F}_2(\theta)$ is obtained from θ by reattaching the line ℓ_T^1 to all the nodes $w \in V_2$ and by reattaching the line ℓ_T^2 to all the nodes $w \in V_2$, and, simultaneously, by replacing all lines $h \leftarrow H$ with lines $H \leftarrow h$ and *vice versa*; note that $\mathcal{F}_1(\theta) \cup \mathcal{F}_2(\theta) \subset \mathcal{F}_T(\theta)$. See Fig. 11 for a simple example.

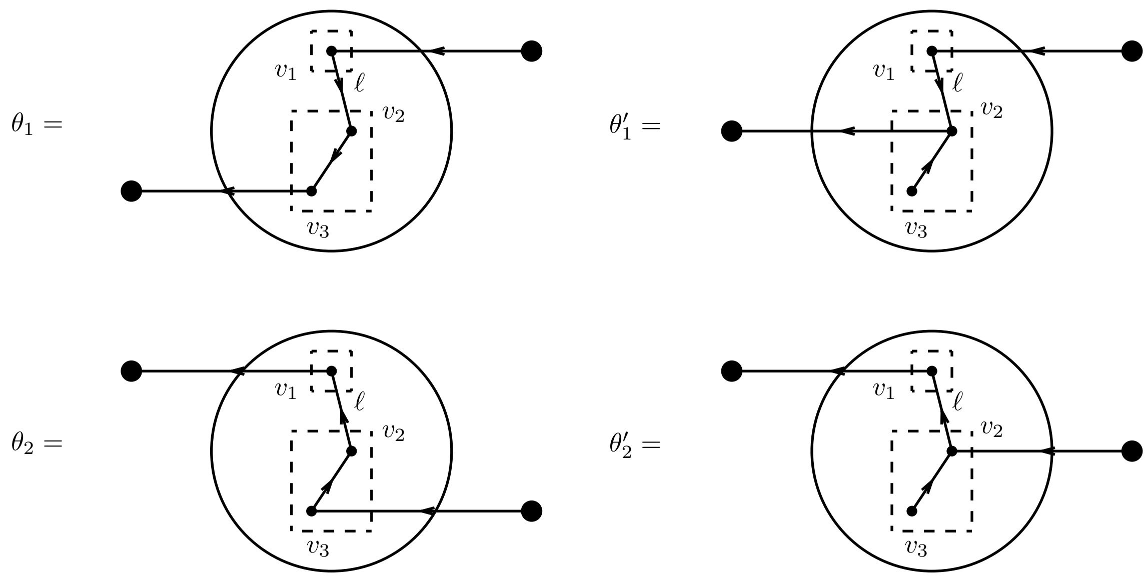

FIG. 11. The two families $\mathcal{F}_1(\theta)=\{\theta_1,\theta_1'\}$ and $\mathcal{F}_2(\theta)=\{\theta_2,\theta_2'\}$ for a three-node resonance. Here $V_1=\{v_2,v_3\}$ and $V_2=\{v_1\}$. The black balls represent the remaining parts of the trees. The labels are not explicitly shown.

As a consequence of such an operation the arrows of some lines $\ell' \in \Lambda(\theta) \setminus \ell_0$ change their directions: this means that some line $h \leftarrow H$ becomes $H \leftarrow h$ and *vive versa*, and, correspondingly, some propagator $g(\boldsymbol{\nu})$ becomes $-g(-\boldsymbol{\nu})$, but $g(-\boldsymbol{\nu}) = -g(\boldsymbol{\nu})$, so that no overall change is produced by such factors. On the other hand one has a derived propagator $\pm g'(\boldsymbol{\nu}_\ell)$ for the trees in $\mathcal{F}_1(\theta)$ and a derived propagator $\mp g'(-\boldsymbol{\nu}_\ell)$ for the trees in $\mathcal{F}_2(\theta)$, and $g'(\boldsymbol{\nu}) = g'(-\boldsymbol{\nu})$. Then by summing over all the possible trees in $\mathcal{F}_1(\theta)$ we obtain a value $i^2\boldsymbol{\nu}_1\boldsymbol{\nu}_2$ times a common factor, while by summing over all the possible trees in $\mathcal{F}_2(\theta)$ we obtain $-i^2\boldsymbol{\nu}_1\boldsymbol{\nu}_2$ times the same common factor, so that the sum of two sums gives zero.

A2.3. *Resonance of type 3.* To prove that $\mathcal{V}_T(0) = 0$ simply reason as for $\mathcal{V}_T(0)$ in the previous case, by using that the entering line is a line $h \leftarrow H$.

A2.4. *Resonance of type 2.* We want to show that also in such a case $\mathcal{V}_T(0) = 0$. Given a tree θ consider all trees obtained by detaching the exiting line, then reattaching to all the nodes $v \in V(T)$; see Fig. 12. Note again that the trees so obtained are contained in the resummation family $\mathcal{F}_T(\theta)$.

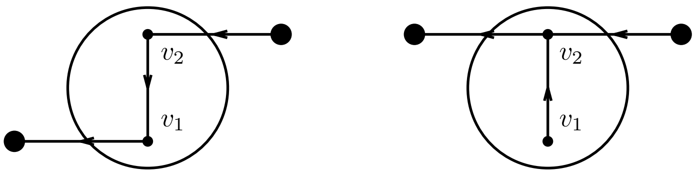

FIG. 12. The trees obtained by shifting the line entering a two-node resonance. The black balls represent the remaining parts of the trees. The labels are not explicitly shown.

In such a case again some propagators can change, *i.e.* a line $h \leftarrow H$ can become $H \leftarrow h$ and *vice versa* but the corresponding propagator does not change (reason as above for resonances of type 1). So at the end we obtain a common factor times $i\boldsymbol{\nu}_v$, where v is the node which the exiting line is attached to. By (5.10) again we obtain $\mathcal{V}_T(0) = 0$.

So the lemma is proved.

Appendix A3. Proof of the lemma 5.11

A3.1. *Notations.* As in §5.12 we define the depth $D(T)$ of a resonance T by setting $D(T) = 1$ if there is no resonance containing T, and setting $D(T) = D(T') + 1$ if T is contained inside a resonance T' and all the other resonances inside T' (if there are any) do not contain T.

Given a resonance T, we denote here by T_0 the set of nodes and lines internal to T_0 not contained in any resonance inside T.

Given a resonance T and a line $\ell \in T$ we write $\boldsymbol{\nu}_\ell = \boldsymbol{\nu}_\ell^0 + \sigma_\ell \boldsymbol{\nu}$ as in (5.27), with $\boldsymbol{\nu} = \boldsymbol{\nu}_{\ell_T^2}$. By shifting the lines external to the resonances a momentum $\boldsymbol{\nu}_\ell$ can be changed into $\pm\boldsymbol{\nu}_\ell^0 + \sigma\boldsymbol{\nu}$ (as noted in §5.8).

A3.2. *Proof of the lemma 5.11.* The proof is by induction of the depth D of the resonances.

If T is a resonance with depth $D(T) = 1$ and $\ell \in T_0$, denoting by n the scale of the line ℓ_T^2, then one must have $n_T \ge n + 3$ by definition of resonance, as it can easily proved. Moreover $n_\ell \ge n_T$ for all $\ell \in T$, by definition of cluster).

Then, by denoting by $\boldsymbol{\nu}$ the momentum of the line entering T, one has

$$C_0 2^{n_\ell} \ge |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| > C_0 2^{n_\ell - 1}, \tag{A3.1}$$

while

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \le C_0 2^{n} \le C_0 2^{n_T - 3} \le C_0 2^{n_\ell - 3}, \tag{A3.2}$$

so that

$$\begin{aligned} \left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| &\ge |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| - |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \ge C_0 2^{n_\ell - 1} - C_0 2^{n_\ell - 3} \ge C_0 2^{n_\ell - 2}, \\ \left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| &\le |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| + |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \le C_0 2^{n_\ell} + C_0 2^{n_\ell - 3} \le C_0 2^{n_\ell + 1}, \end{aligned} \tag{A3.3}$$

which proves the assertion for lines in T_0 with $D(T) = 1$.

Fix $D > 1$. Then suppose that the assertion holds for all resonances of depth $D' < D$: we show that then it holds also for resonances of depth D.

Let ℓ be a line in T_0, for some resonance $T \in \mathcal{T}(\theta)$ of depth $D(T) = D$. By the inductive hypothesis ℓ_T^2 is contained inside a resonance of depth $D - 1$, so that its scale can be changed at most by one unit, *i.e.* , if one had $n_{\ell_T^2} = n$ before shifting the lines, the scale $n_{\ell_T^2}$ has become $n'_{\ell_T^2}$ with $n_{\ell_T^2} - 1 \le n'_{\ell_T^2} \le n_{\ell_T^2} + 1$: then $n'_{\ell_T^2} \le n_T - 2$.

Therefore one has

$$C_0 2^{n_\ell} \ge |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| > C_0 2^{n_\ell - 1}, \tag{A3.4}$$

while

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \le C_0 2^{n+1} \le C_0 2^{n_T - 2} \le C_0 2^{n_\ell - 2}, \tag{A3.5}$$

so that again

$$\begin{aligned} \left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| &\ge |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| - |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \ge C_0 2^{n_\ell - 1} - C_0 2^{n_\ell - 2} \ge C_0 2^{n_\ell - 2}, \\ \left|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell^0\right| &\le |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}_\ell| + |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \le C_0 2^{n_\ell} + C_0 2^{n_\ell - 2} \le C_0 2^{n_\ell + 1}, \end{aligned} \tag{A3.6}$$

which proves the assertion for the lines in T_0 with $D(T) = 1$.

References

[BGG] M.V. Bartuccelli, K.V. Georgiou, G. Gentile: KAM theorem and stability of the upside-down pendulum, Preprint (2000).

[BG] A. Berretti, G. Gentile: Scaling properties for the radius of convergence of Lindstedt series: generalized standard maps, *J. Math. Pures Appl. (9)* **79** (2000), no. 7, 691–713.

[BGGM1] F. Bonetto, G. Gallavotti, G. Gentile, V. Mastropietro: Lindstedt series, ultraviolet divergences and Moser's theorem, *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **26** (1998), no. 3, 545–593.

[BGGM2] F. Bonetto, G. Gallavotti, G. Gentile, V. Mastropietro: Quasi linear flows on tori: regularity of their linearization, *Comm. Math. Phys.* **192** (1998), no. 3, 707–736.

[CG] L. Chierchia, G. Gallavotti: Smooth prime integrals for quasi-integrable Hamiltonian systems, *Nuovo Cimento B* **67** (1982), no. 2, 277–295.

[E] L.H. Eliasson: Absolutely convergent series expansions for quasi-periodic motions, *Math. Phys. Electron. J.* **2** (1996), paper 4, 1–33.

[G1] G. Gallavotti: Quasi-integrable mechanical systems, in *Critical Phenomena, Random Systems, Gauge Theories*, Les Houches, Session XLIII (1984), Vol. II, 539–624, Ed. K. Osterwalder & R. Stora, North Holland, Amsterdam, 1986.

[G2] G. Gallavotti: Twistless KAM tori, *Comm. Math. Phys.* **164** (1994), no. 1, 145–156.

[GGM] G. Gallavotti, G. Gentile, V. Mastropietro: A field theory approach to Lindstedt series for hyperbolic tori in three time scales problems, *J. Math. Phys.* **40** (1999), no. 12, 6430–6472.

[Ge] G. Gentile: Whiskered tori with prefixed frequencies and Lyapunov spectrum, *Dynamics and Stability of Systems* **10** (1995), no. 3, 269–308.

[GM] G. Gentile, V. Mastropietro: Methods for the analysis of the Lindstedt series for KAM tori and renormalizability in Classical Mechanics. A review with some applications, *Rev. Math. Phys.* **8** (1996), no. 3, 393–444.

[K] A.N. Kolmogorov: On the preservation of conditionally periodic motions, *Dokl. Akad. Nauk* **96**, 527–530 (1954); English translation in G. Casati, J. Ford: *Stochastic behavior in classical and quantum Hamiltonians*, Lecture Notes in Physics **93**, Springer, Berlin, 1979.

[M] J. Moser: Convergent series expansions for quasi–periodic motions, *Math. Ann.* **169** (1967), 136–176.

[P] H. Poincaré: *Les méthodes nouvelles de la mécanique céleste*, Gauthier-Villars, Paris, Vol. I, 1892, Vol. II, 1893, Vol. III, 1899.

[Pö1] J. Pöschel: Über invariante Tori in differenzierbaren Hamiltonschen Systemen, Bonner Mathematische Schriften **120**, Universität Bonn, Mathematisches Institut, Bonn, 1980. Integrability of Hamiltonian systems on Cantor sets, *Comm. Pure Appl. Math.* **35** (1982), no. 5, 653–696.

[Pö2] J. Pöschel: Invariant manifolds of complex analytic mappings near fixed points, in *Critical Phenomena, Random Systems, Gauge Theories*, Les Houches, Session XLIII (1984), Vol. II, 949–964, Ed. K. Osterwalder & R. Stora, North Holland, Amsterdam, 1986.

[S] W.M. Schmidt: *Diophantine approximation*, Lecture Notes in Mathematics **785**, Springer, Berlin, 1980.

[W] H. Whitney: Analytic extensions of differentiable functions defined in closed sets, *Trans. Amer. Math. Soc.* **36** (1934), no. 1, 63–89.