

Kolmogorov theorem and classical perturbation theory

Antonio Giorgilli and Ugo Locatelli

Abstract. We reconsider the original proof of Kolmogorov's theorem in the light of classical perturbation methods based on expansions in some parameter. With a careful analysis of the accumulation of small divisors we prove that their effect is bounded by a geometrically increasing numerical sequence. This allows us to achieve the proof without using the so called quadratic method.

Mathematics Subject Classification (1991). 58F05, 58F30, 34C27.

Keywords. KAM-theory, quasiperiodic motion, Hamiltonian perturbations.

1. Introduction

The aim of this paper is to reconsider the proof of Kolmogorov's theorem (usually referred to as KAM theorem) in the light of classical perturbation theory. By this we mean that we give a proof based on classical expansions on some parameter. As in Kolmogorov's original proof, we proceed performing an infinite sequence of canonical transformations bringing the Hamiltonian in a suitable normal form. This presents some strong connections with the method of Lindstedt's series: one of the main differences is that we find absolutely convergent series (see point 4 below).

We recall that Kolmogorov's theorem concerns the dynamical problem of a canonical system of differential equations with Hamiltonian

$$H(p, q) = h(p) + \varepsilon f(p, q, \varepsilon), \quad p \in \mathcal{G} \subset \mathbf{R}^n, \quad q \in \mathbf{T}^n, \quad (1)$$

where \mathcal{G} an open set and ε is a small parameter (see [1]). The problem is to prove the existence of quasi periodic solutions for the system above, with given non resonant frequencies (see [1]). As is well known, the main technical difficulty in

Work partially supported by grant CE n. CHRX-CT93-0330/DG, CE n. ERB-CHRX-CT94-0460 and CE n. ERB-FMBI-CT95-0316 (the last one in TMR programme of the European Community).

studying such a system with the methods of classical perturbation theory is due to the so called *small denominators*.

From a formal viewpoint, this problem has been faced with two different methods, namely (i) direct construction of solutions in the form of series expansions, and (ii) reduction of the Hamiltonian to a suitable normal form, from which the existence of the wanted solution turns out to be evident. The first method goes back to Lindstedt and Poincaré; the second one has been used by Poincaré, but the crucial contribution is due to Kolmogorov.

There is a strict relation between these two methods, which has been pointed out by Poincaré in chapt. IX of his *Méthodes Nouvelles*^[2]. Indeed, the method of Lindstedt consists in looking for solutions $p(t), q(t)$ which are quasiperiodic functions of time, with fixed (nonresonant) frequencies ω . The functions $p(t), q(t)$ are constructed as power series expansions in the parameter ε . However, as remarked by Poincaré, this method presents some problems of formal consistency. Poincaré used the method of normal form to overcome this difficulty. Precisely, his suggestion was to look for a near to identity canonical transformation of the form $p = p' + \psi(p', q')$, $q = q' + \varphi(p', q')$ such that the transformed Hamiltonian depends only on the action variables p' . Here too, the functions ψ, φ are constructed as power series expansions in ε . In this case no consistency problems show up, so that the procedure can be formally performed. On the other hand, the solutions for the Hamiltonian in normal form are $p' = p'_0, q' = \omega t + q'_0$, where p'_0, q'_0 are integration constants. Thus, the transformation of variables gives also the solutions looked for by Lindstedt.

Having settled the formal consistency, and the equivalence, of the two methods, one is naturally lead to consider the problem of convergence. This has been indeed discussed by Poincaré (see chapt. XIII, N. 149): it is known that he attempted to prove the divergence of Lindstedt's expansions, without success. His conclusion was that convergence can not be excluded in case the frequencies satisfy some suitable nonresonance conditions, but "il est fort invraisemblable".

Our paper is concerned precisely with the convergence of the series expansions generated by the method of construction of the normal form, but taking into account also the ideas introduced by Siegel and Kolmogorov. In order to explain our view point, it is useful to recall some historical aspects of the development of this problem. We do not attempt to give a complete report. We just limit our discussion to a few points that lie at the basis of our idea.

1. A first significant result concerning the convergence of series with small divisors was obtained by Siegel studying two different problems, namely: (i) the problem of conjugating an analytic mapping of the complex plane into itself, with the origin as a fixed point, to the linear part of the mapping (see [3]), and (ii) the normal form of a system of differential equations in the neighbourhood of an equilibrium point (see [4]). The first problem is known as "Siegel center problem", and the formal solution is known as Schröder series; the corresponding series for the Hamiltonian case are the series of Lindstedt. In both cases the formal solution is given as a series

containing small divisors. Siegel proved that in both cases the series is convergent, and so the problem has an analytic solution, provided the rotation number of the mapping, or the eigenvalues of the linear part of the differential equations, satisfy a strong irrationality condition of diophantine type. The proof is based on a clever analysis of the accumulation of small divisors. An interesting aspect is that one and the same idea allows one to achieve the proof of convergence both for the series of Schröder and for the normal form.

2. The idea of Siegel does not apply to the Hamiltonian case. In rough terms, the reason is the following: in Siegel's case the small denominators do not repeat; in the Hamiltonian case they do repeat. To make things simple, let us discuss the problem in terms of an expansion in powers of a small parameter ε (actually, in the case of Siegel the small parameter is the radius of a neighbourhood of the origin). Consider a real nonresonant vector ω , and define two sequences $\{\alpha_r\}_{r>0}$ and $\{\beta_r\}_{r>0}$ as

$$\alpha_r = \min_{0 < |k| \leq r} |k \cdot \omega|, \quad \beta_r = \min_{|k|=r} |k \cdot \omega|, \quad k \in \mathbf{Z}^n. \quad (2)$$

The nonresonance condition implies that all α 's and β 's are different from zero. Remark that the sequence α_r is monotonically decreasing to zero, while the sequence β_r is quite random, although its \liminf is zero. In the cases considered by Siegel the worst possible denominator in the coefficient of ε^r is the product $\beta_1 \cdot \dots \cdot \beta_r$. In the series of Lindstedt and in the normal form of Poincaré one finds instead a denominator $\alpha_1 \cdot \dots \cdot \alpha_r$. Therefore, one is lead to investigate the convergence of a series with a general term of either form

$$\frac{\varepsilon^r}{\beta_1 \cdot \dots \cdot \beta_r}, \quad \frac{\varepsilon^r}{\alpha_1 \cdot \dots \cdot \alpha_r}.$$

The first case applies to Siegel's problems; the latter case to the Hamiltonian problem. Assume now that the vector ω satisfies the diophantine condition $\alpha_r \geq \gamma r^{-\tau}$ for some $\gamma > 0$ and $\tau \geq n - 1$. A naive use of this condition would give $\beta_1 \cdot \dots \cdot \beta_r \geq \gamma^r (r!)^{-\tau}$ (remark that $\beta_r \geq \alpha_r$). Siegel's argument is that one has instead $\beta_1 \cdot \dots \cdot \beta_r \geq C^r$ for some constant $C < 1$; this exactly because the sequence β_r is not monotonic: very small values are very rare. In the words of Siegel, "this simple remark is the main argument of the whole proof". However, this argument does not apply to the product $\alpha_1 \cdot \dots \cdot \alpha_r$, that can be estimated from below only by $\gamma^r (r!)^{-\tau}$. Therefore, the series above turn out to be convergent for the sequence β_r , namely in Siegel's cases, and divergent for the sequence α_r , namely in the Hamiltonian case.

3. The proof of existence of quasiperiodic solution was first given by Kolmogorov in the short note [1]. He solved the problem by introducing two novel ideas: (i) looking for a special set of initial conditions leading to conditionally periodic motions on an invariant torus, and (ii) replacing the classical expansion in the small parameter ε by a superconvergent iteration procedure using the generalized Newton's method. His method is reminiscent of Poincaré's construction of the normal form,

with two crucial differences. Firstly, instead of eliminating all dependencies on the angles, Kolmogorov looks for an Hamiltonian of the form (hereafter referred to as Kolmogorov's normal form)

$$H(p, q) = \omega \cdot p + R(p, q), \quad R = O(p^2),$$

with frequencies ω satisfying a diophantine irrationality condition. This makes evident the existence of a single invariant torus $p = 0$, carrying quasiperiodic motions with frequencies ω . The second difference is that no use is made of power expansions in a small parameter during the construction of the normal form. This results in a superconvergent procedure known as "quadratic method". It is a common opinion that the achievement of the proof of Kolmogorov's theorem is strongly based on the use of this fast convergence (in addition to the original Kolmogorov's paper [1], see also the subsequent articles by Moser [5][6] and Arnold[7][8]).

4. The convergence of Lindstedt's series in view of Kolmogorov's theorem has been discussed by Moser in [6] (see also [9]). His conclusion is that the series of Lindstedt are actually convergent, but a proof of their convergence can not be achieved via the usual method of majorants going back to Cauchy. However, Moser shows that an *indirect* proof of convergence can be given using the method of Kolmogorov. This complicated state of affairs has been clarified in the recent works of Eliasson [10][11][12], Gallavotti[13][14], and Chierchia and Falcolini[15] (see also [16], [17], [18]). Indeed, Eliasson proved that in Lindstedt's expansions there are cancellations of critical terms. In slightly more precise terms the situation is the following. The method of Lindstedt produces the coefficients of the power expansion in ε as Fourier expansions in the angles q . The coefficient of a single Fourier mode is actually the sum of several terms, with different combinations of small divisors. Formally, the solution has the form $\sum_s \varepsilon^s a_s(q)$, with $a_s(q) = \sum_{j \in \mathcal{J}} b_j \exp(k_j \cdot q)$, where \mathcal{J} is a set of indexes, and b_j and k_j are a numerical coefficient and a Fourier mode, respectively. The set \mathcal{J} of indexes labels the set of terms produced by Lindstedt's algorithm, and takes into account that the same Fourier mode is generated many times, with different coefficients. Now, some divisors cause the corresponding coefficient b_j to grow very fast, so that $\sum_s \varepsilon^s \sum_j |b_j|$ diverges (for $\varepsilon \neq 0$). That is: the series generated by Lindstedt's algorithm is not absolutely convergent. Now, the coefficients of the critical terms do compensate each other, leaving a remainder which grows not faster than geometrically. Eliasson did not give an *explicit* mechanism of cancellation. Such an explicit mechanism has been provided in the works of Gallavotti and, more recently, of Chierchia and Falcolini. As a matter of fact, they give an algorithm of summation of Lindstedt's series from which convergence follows. Therefore, the problem of convergence for Lindstedt's series can be considered as definitely settled.

In view of the discussion above, we think it is natural to ask what happens if one tries to implement the construction of Kolmogorov's normal form via classical expansions in powers of ε (as was suggested by Poincaré), instead of using the quadratic method introduced by Kolmogorov. This paper is concerned with ex-

actly this point. One is tempted to conjecture that one will find exactly the same situation as for Lindstedt's series; that is, the algorithm produces series which are not absolutely convergent, but there are cancellations which make them to converge. By the way, this conjecture is in agreement with the common belief that the quadratic method is an essential tool for the proof of Kolmogorov's theorem. One could also conjecture that the algorithm will produce exactly the same expansions as the method of Lindstedt. Now, our point is that both these conjectures are false. We show instead that the construction of the normal form can be implemented in such a way that one produces *absolutely convergent series*.

The basic idea, that we borrow from Siegel, is that one should start with a careful analysis of the mechanism of accumulation of small divisors. Pushing the ideas of Siegel a little further, we discover that the accumulation in the expansions in a parameter is the same that is expected in the quadratic method of Kolmogorov. This fact replaces the "simple remark" of Siegel, and allows us to prove the *absolute* convergence of our expansions. By the way, we point out that our considerations could be easily applied also to the problems investigated by Siegel, thus sharpening his analysis in such a way that it can be applied to the Hamiltonian case.

The non necessity of a mechanism of cancellations makes the conclusion trivial, in some sense: one could say that our proof is just a modification of that of Kolmogorov. However, we emphasize that the algorithm of Kolmogorov has never been compared with the classical methods: that the power expansions in a parameter could be absolutely convergent is quite unexpected. We believe that our scheme brings into light some mechanisms which are present in all problems involving small divisors, and that are not evident in the proofs based on the quadratic scheme. Furthermore, one should notice that in practical applications the Hamiltonian is typically known only as an expansion in ε with coefficients expanded in Fourier series. In such a case one practically works with a finite order expansion in ε and some suitable truncation of the Fourier series. Thus, it is interesting to have an explicit constructive algorithm supported by quantitative estimates. Therefore, we believe that a direct proof of the convergence of the classical scheme of expansion in a parameter is of interest in itself.

The paper is organized as follows. In sect. 2 we recall the statement of Kolmogorov's theorem, and illustrate in an informal way the main idea of our scheme. We also discuss the relations with the algorithms of Lindstedt and of Kolmogorov. The rest of the paper is devoted to the proof. In sect. 3 we establish an appropriate algebraic and analytic framework, and reformulate the theorem in a form adapted to our scheme of proof. In sect. 4 we give the scheme of proof. We try to simplify the scheme omitting many technical details which are the tedious but unavoidable tools in this kind of results. However, in order to make the paper self contained, most technical details are explicitly worked out in sect. 5, which contains only technical matter.

Acknowledgements. During the preparation of this paper we had fruitful discussions with D. Bambusi, G. Benettin, A. Carati, F. Fassò, L. Galgani, G. Gallavotti, U. Kirchgraber, A. Morbidelli and D. Stoffer (in alphabetical order), whom we thank for their interest and for useful suggestions. Part of this work has been done while one of the authors (A. Giorgilli) was guest of the Forschungsinstitut für Mathematik, ETH, Zürich. The hospitality of ETH is greatly appreciated.

2. Formulation of the theorem and discussion

We recall the formal statement of the theorem, essentially in the form given by Kolmogorov in his original paper [1]. The Hamiltonian (1) will be assumed to be real analytic for sufficiently small values of the parameter ε , and to be a real holomorphic function of p, q in the complex domain

$$\mathcal{D}_{\delta, \xi} = \mathcal{G}_\delta \times \mathbf{T}_\xi^n \quad (3)$$

where δ and ξ are positive parameters, $\mathcal{G} \subset \mathbf{R}^n$, and

$$\begin{aligned} \mathbf{T}_\xi^n &= \{q \in \mathbf{C}^n : |\operatorname{Im}(q)| < \xi\} \\ \mathcal{G}_\delta &= \bigcup_{p \in \mathcal{G}} B_\delta(p), \end{aligned} \quad (4)$$

$B_\delta(p)$ being the open ball of radius δ and center p in \mathbf{C}^n .

Following the original idea of Kolmogorov, we consider a point $p^* \in \mathcal{G}$ such that the frequencies $\omega := \frac{\partial h}{\partial p}(p^*)$ satisfy a diophantine condition

$$|k \cdot \omega| \geq \gamma |k|^{-\tau} \quad (5)$$

for some positive γ and some $\tau \geq n - 1$. Furthermore, we denote

$$C_{ij} = \frac{\partial^2 h}{\partial p_i \partial p_j}(p^*). \quad (6)$$

Kolmogorov's theorem can now be stated as follows.

Theorem. *Consider the Hamiltonian (1); assume that $h(p)$ and $f(p, q, \varepsilon)$ for ε small enough are uniformly bounded in ε and real analytic functions in the domain $B_\delta(p^*) \times \mathbf{T}_\xi^n$, where $p^* \in \mathcal{G}$ is such that the corresponding unperturbed frequencies ω satisfy the diophantine condition (5). Suppose moreover that there exists a positive constant m such that for every $v \in \mathbf{R}^n$ one has*

$$m \sum_i |v_i| \leq \sum_i \left| \sum_j C_{ij} v_j \right|, \quad (7)$$

where C_{ij} is the matrix defined by (6). Then there exists a positive ε^* such that for $|\varepsilon| < \varepsilon^*$ the following statement holds true: there exists a real analytic canonical

transformation

$$p = p^* + p' + \varphi(p', q', \varepsilon) , \quad q = q' + \psi(p', q', \varepsilon) , \tag{8}$$

mapping the domain $B_\varrho(0) \times \mathbf{T}_\sigma^n$ to $B_\delta(p^*) \times \mathbf{T}_\xi^n$ for some positive $\varrho < \delta$ and $\sigma < \xi$, which gives the Hamiltonian the Kolmogorov's normal form

$$H'(p', q', \varepsilon) = \sum_{j=1}^n \omega_j p'_j + R(p', q', \varepsilon) , \tag{9}$$

where $R(p', q', \varepsilon)$ is at least quadratic in p' . The change of variables is near the identity, in the sense that one has $\varphi, \psi = O(\varepsilon^{1/2})$.

In view of the form of the transformed Hamiltonian it is evident that the torus $p' = 0$ is invariant under the canonical flow, and that the flow on that torus is given by $q'(t) = \omega t + q'_0$, where q'_0 is the initial phase. Therefore, in view of (8), the solutions of Lindstedt should be compared with

$$p = p^* + \varphi(0, \omega t + q'_0, \varepsilon) , \quad q = \omega t + q'_0 + \psi(0, \omega t + q'_0, \varepsilon) . \tag{10}$$

As a byproduct of the proof we obtain explicit estimates that we report here. Let E_0 and F_0 be constants such that

$$\sup_p |h(p)| \leq E_0 , \quad \sup_{p,q,\varepsilon} |f(p, q, \varepsilon)| \leq F_0 , \tag{11}$$

the supremum being taken over $p \in B_\delta(p^*)$, $q \in \mathbf{T}_\xi^n$ and ε in some interval around zero. Then one has the following quantitative estimates:

$$\varepsilon^* = \frac{\bar{\varepsilon}}{A} \tag{12}$$

$$\varrho = \frac{\delta}{2} , \quad \sigma = \frac{\xi}{2} \tag{13}$$

$$|\varphi| < \frac{1}{8} |\varepsilon|^{1/2} A^{1/2} \varrho , \quad |\psi| \leq \frac{1}{8} |\varepsilon|^{1/2} A^{1/2} \sigma \tag{14}$$

$$\bar{\varepsilon} = \min \left\{ 1, \left[\frac{2^{n+2} E_0}{\delta} \left(\frac{2K^\tau}{\gamma\xi} + \frac{4e}{m\delta} \right) \right]^{-2}, \left[\frac{2^{n+2} E_0 K^\tau}{\gamma\delta\xi} \left(1 + \frac{2^{n+3} E_0 K^\tau}{e\gamma\delta\xi} \right) \right]^{-2} \right\} \tag{15}$$

$$K = \left\lceil -\frac{4}{\xi} \ln \bar{\mu} \right\rceil \tag{16}$$

$$\bar{\mu} = \frac{1}{3 \cdot 2^{2\tau+21} (\zeta(3/2))^4} \tag{17}$$

$$A = e^{K\xi/4} \left(\frac{1 + e^{-\xi/4}}{1 - e^{-\xi/4}} \right)^n \frac{F_0}{E_0} ; \tag{18}$$

here, $\zeta(s) = \sum_{j \geq 1} 1/j^s$ denotes the Riemann zeta function, and $\lceil \cdot \rceil$ denotes the smallest integer larger than the argument.

We come now to illustrating in an informal way the idea which lies at the basis of our proof of Kolmogorov's theorem. We include this part with the purpose of

making transparent the technical mechanism of control of the accumulation of small divisors that will be used in the next sections.

Let us briefly recall the scheme of proof proposed by Kolmogorov in his original memoir [1]. Consider the Hamiltonian

$$H(p, q) = \omega \cdot p + A(q) + B(q) \cdot p + \frac{1}{2}C(q)p \cdot p + O(p^3) \quad (19)$$

where the function $A(q)$ and the vector valued function $B(q)$ are assumed to be small. As usual, all functions are assumed to be analytic. The goal is to kill the unwanted terms independent of p and linear in p , namely $A(q) + B(q) \cdot p$. To this end, Kolmogorov's suggestion is to look for a canonical transformation with generating function

$$S(\hat{p}, q) = \hat{p} \cdot q + X(q) + \xi \cdot q + Y(q) \cdot \hat{p} , \quad (20)$$

where $X(q)$ is a function, $Y(q)$ is a vector function, and ξ is a real vector. The latter three quantities are so determined that the Hamiltonian (6) is transformed to a new Hamiltonian $\hat{H}(\hat{p}, \hat{q})$ of the same form, with new functions $\hat{A}(\hat{q})$ and $\hat{B}(\hat{q})$, which, however, have smaller size than $A(q)$ and $B(q)$. By iterating this procedure one constructs an infinite sequence of canonical transformations which is proven to converge to an analytic canonical transformation, say, $p = \psi(p', q')$, $q = \varphi(p', q')$ which brings the Hamiltonian in Kolmogorov's normal form (9).

A technical remark is that no use is made here of power expansions in the small parameter ε . Indeed, at every step all terms independent of p and linear in p are collected together, and used in order to determine the generating function for the next step. This results in a fast convergence, usually called quadratic. However, this also makes impossible a direct comparison of the equations representing the flow on the torus with the solutions of Lindstedt. Indeed, looking at Kolmogorov's proof one immediately realizes that if the Hamiltonian (1) is analytic in ε then the functions ψ and φ are analytic in ε , too. This means that the solutions can be expanded in power series of ε . The resulting series must coincide with a *resummation* of Lindstedt's series. Therefore, the convergence of Kolmogorov's construction implies the convergence, but not necessarily the *absolute* convergence, of Lindstedt's series. A direct comparison can not be made, because the explicit expansion in ε is not given by Kolmogorov's method.

We now modify the algorithm of Kolmogorov by using explicitly the expansion in ε . In order to simplify the discussion, let us consider here the simple case of an Hamiltonian of the form

$$H(p, q) = \omega \cdot p + \frac{p^2}{2} + \varepsilon [A(q) + B(p, q) + C(p, q)] , \quad (22)$$

where $A(q)$ is independent of p , and $B(p, q)$ and $C(p, q)$ are assumed to be linear and quadratic in p , respectively; moreover, let us assume that A , B and C are trigonometric polynomials of degree $K > 1$. By the way, we can always assume that $A(q)$ has zero average over the angles, because a constant in the Hamiltonian

is irrelevant. The present case already contains all the essential difficulties. The general case will be considered in the following sections.

We use the formalism of Lie series in order to perform canonical transformations. In brief, having given a generating function $\varepsilon\chi$ the transformation of the Hamiltonian is written as $H' = \exp(L_{\varepsilon\chi})H = H + \varepsilon L_\chi H + \frac{1}{2}\varepsilon^2 L_\chi^2 H + \dots$, where $L_\chi \cdot = \{\chi, \cdot\}$. A similar formula is easily written for a generating function $\varepsilon^s \chi$, with $s > 1$. The advantage is that reordering all expansions in powers of ε is very easy, because the algorithm of Lie series gives exactly that power expansion. For more details see for instance [19].

Let us illustrate the first step. In order to keep trace also of the powers of p , we perform the transformation of Kolmogorov via composition of two canonical transformations with generating functions, respectively, $\chi_1 = X(q) + \xi \cdot q$ and $\chi_2 = Y(q) \cdot p$. With the first transformation we get

$$\hat{H} = \exp(L_{\varepsilon\chi_1})H = \omega \cdot p + \frac{p^2}{2} + \varepsilon \left[A + \{\chi_1, \omega \cdot p\} + B + \{\chi_1, \frac{p^2}{2}\} + C \right] + \varepsilon^2 \dots ,$$

and we determine the function $X(q)$ and the real vector ξ so that $A + \{X, \omega \cdot p\} = 0$ and the average of $B + \{\xi \cdot q, p^2/2\}$ is zero. The constant term $\{\xi \cdot q, \omega \cdot p\} = \xi \cdot \omega$ is constant, and so it can be neglected. This kills the unwanted term $A(q)$. With the second transformation we get

$$H' = \exp(L_{\varepsilon\chi_2})\hat{H} = \omega \cdot p + \frac{p^2}{2} + \varepsilon \left[\{\chi_2, \omega \cdot p\} + \hat{B} + \{\chi_2, \frac{p^2}{2}\} + C \right] + \varepsilon^2 \dots ,$$

where $\hat{B} = \{\chi_1, p^2/2\} + B$ has vanishing average, so that we can determine χ_2 by solving the equation $\{\chi_2, \omega \cdot p\} + \hat{B} = 0$.

After this step we find an Hamiltonian of the form

$$H'(p, q) = \omega \cdot p + \frac{p^2}{2} + \varepsilon C'_1(p, q) + \varepsilon^2 [A'_2(q) + B'_2(p, q) + C'_2(p, q)] + \varepsilon^3 \dots , \quad (22)$$

namely, in Kolmogorov's normal form up to terms of order ε^2 . The functions A'_s , B'_s and C'_s are determined as the part of the coefficient of ε^s which is independent of p , linear in p and quadratic in p , respectively. Since the generating functions are at most linear in p , no terms of degree higher than 2 in p are generated. Moreover, the A'_s , B'_s and C'_s turn out to be trigonometric polynomials of degree sK in the angles q .

Concerning the small divisors, from the equations determining χ_1 and χ_2 one immediately sees that each term in the Fourier expansion of $X(q)$ contains exactly one denominator $k \cdot \omega$ with $k \in \mathbf{Z}^n$ and $0 < |k| \leq K$, and that each term in the Fourier expansion of χ_2 contains at most two such denominators. Changing a little the definition of the sequence $\{\alpha_r\}_{r \geq 1}$ in (2) to

$$\alpha_r = \min_{0 < |k| \leq rK} |k \cdot \omega| , \quad (23)$$

we see that the worst possible denominator in χ_1 and χ_2 is α_1 and α_1^2 , respectively. Furthermore, the worst possible denominator in the coefficient of ε^s in transformed Hamiltonian is α_1^{2s-2} for terms independent of p , α_1^{2s-1} for terms linear in p and α_1^{2s} for terms quadratic in p . This is easily checked.

We proceed now by successively eliminating the part independent of p and linear in p from the coefficients of $\varepsilon^2, \varepsilon^3, \dots$, without adding them up. That is, we proceed as almost everybody did before Kolmogorov. This way we always work with powers of ε . Following the idea of Siegel, we look carefully at the accumulation of small divisors. A naive consideration would be that the transformation with the generating functions of order s adds two denominators α_s ; thus, after r steps of normalization one expects to find a denominator $(\alpha_1 \cdot \dots \cdot \alpha_r)^2$. This would bring us back to the considerations of point 2 in the introduction, thus supporting the conjecture that there are cancellations. But it is not so. The key remark is the following. A term which appears in the generating function χ_1 at order ε^s contributes to χ_2 at orders $\varepsilon^s, \dots, \varepsilon^{2s-1}$, but can contribute again to χ_1 only at order ε^{2s} or higher. On the other hand, a term which appears in the generating function χ_2 at order ε^s does not contribute to the generating functions of order $\varepsilon^{s+1}, \dots, \varepsilon^{2s-1}$, and can contribute again to the generating functions only at order ε^{2s} or higher. Checking this fact requires a careful analysis of the algorithm, but does not involve any essential difficulty; a detailed exposition can be found in [19]. In *rough* terms, let the Hamiltonian be in Kolmogorov's normal form up to order ε^{s-1} , and let the term of order ε^s contain a product of small divisors that we denote with d_s . For instance, from the discussion above we get $d_1 = 1$ and $d_2 = \alpha_1^2$. Performing the normalization at order ε^s we introduce a new pair of small divisors α_s , so that the generating functions will contain a denominator $d_s \alpha_s^2$. On the other hand, the Poisson bracket between a generating function and a term of order ε^s in the Hamiltonian generates a term of order ε^{2s} containing a denominator $d_s^2 \alpha_s^2$. By the argument above, no small divisor will be added until we determine the generating function of order ε^{2s} . Thus, we get the recursive relation $d_{2s} = d_s^2 \alpha_s^2$. If we consider only orders which are powers of 2, this leads to the conclusion that the worst possible denominator at order ε^{2^r} is

$$d_{2^r} = \left(\alpha_1^{2^{r-1}} \alpha_2^{2^{r-2}} \cdot \dots \cdot \alpha_{2^{r-1}} \right)^2 .$$

Consequently, the worst possible denominator in the generating function χ_2 at order ε^{2^r} is $d_{2^r} \alpha_{2^r}^2$. This is exactly the form expected in the quadratic procedure of Kolmogorov (see, e.g., [20]).

Now, the problem is that there are terms of order ε^s , where s is not a power of two. Thus, we need some conjecture on the worst possible denominator at every order. In fact, we prove that:

The worst denominator that can appear at order ε^r is a product of the form $\alpha_{j_1} \cdot \dots \cdot \alpha_{j_s}$ satisfying the following rules:

- (i) *the indexes j_k do not exceed r ;*
- (ii) *the number of factors α_{j_k} is at most $2r$;*

(iii) *the indexes obey the selection rule*

$$\sum_k \log_2 j_k \leq 2(r-1) .$$

It is an easy matter to check that the conjecture fits perfectly the general form $d_{2^s} \alpha_{2^s}^2$ of the worst possible denominator in χ_2 at order $r = 2^s$ (use induction). Proving that it works fine at every order is the main part of our scheme of proof of Kolmogorov's theorem.

Thus, we are naturally led to the problem of investigating the convergence of a series with general term

$$\frac{\varepsilon^r}{\alpha_{j_1} \cdot \dots \cdot \alpha_{j_{2r}}} ,$$

and with the denominators obeying the rules above. It is now an easy matter to check that if the denominators are bounded from below by the usual diophantine condition (5), then the expression above grows not faster than geometrically with r . Indeed, according to (5) and (23) one has $\alpha_s \geq \gamma(sK)^{-\tau}$, and so

$$\frac{\varepsilon^r}{\alpha_{j_1} \cdot \dots \cdot \alpha_{j_{2r}}} \leq \varepsilon^r \prod_{s=1}^{2r} \frac{K^\tau}{\gamma} j_s^\tau .$$

Now, in view of (ii) the number of factors K^τ/γ is at most $2r$, and in view of (iii) one has

$$\log_2 \prod_s j_s^\tau = \tau \sum_s \log_2 j_s \leq 2(r-1)\tau ;$$

We conclude

$$\frac{\varepsilon^r}{\alpha_{j_1} \cdot \dots \cdot \alpha_{j_{2r}}} \leq \frac{1}{2^\tau} \left(\frac{2^\tau K^\tau \varepsilon}{\gamma} \right)^{2r} ,$$

as claimed. This elementary remark is the key of our result.

In order to compare our expansions with the series generated by Lindstedt's algorithm, let us remark that the restrictions above on the accumulation of small divisors are actually very strong. For instance, a denominator of the form $\alpha_1 \cdot \dots \cdot \alpha_r$ is forbidden by our rules. As a matter of fact, a denominator of this form *does* appear in the series generated by Lindstedt's algorithm (see for instance the construction given in [14]). This immediately leads to the conclusion that our series and the series of Lindstedt are actually different, and that they coincide only after resummation of all coefficients of the same Fourier modes. The advantage of our algorithm with respect to Lindstedt is that it produces *absolutely convergent* expansions, while Lindstedt's series are not absolutely convergent. The disadvantage is that it is less efficient, because it requires the construction of a lot of auxiliary function, while the algorithm of Lindstedt produces immediately the wanted solutions.

In the rest of the paper we consider the case of a general Hamiltonian. This requires in particular considering functions which are not trigonometric polynomials in the angles. This seems to invalidate the mechanism above for the control of the accumulation of small divisors. However, this is just a technical matter, and there are many ways to get rid of this problem. Following Poincaré, we use the exponential decay of the coefficients in the Fourier expansion of an analytic function (see [2], chapt. XIII, n. 146, p. 98–99). In view of this fact, we can split the Hamiltonian in a sum of trigonometric polynomials and introduce an artificial small parameter (that will be denoted by μ) which plays the role of the expansion parameter. This seems to be a rather artificial procedure, but is fully equivalent to considering the case of an Hamiltonian which is a trigonometric polynomial. Moreover, we emphasize that in a practical computation one is always restricted to considering truncated Fourier expansions. This is reflected in our scheme.

3. General setting

Our aim in this section is to introduce an appropriate algebraic and analytic framework, also fixing the notations. In particular, in sect. 3.3 we introduce the technical tools which allow us to dominate the accumulation of small divisors in a direct manner.

3.1 Algebraic and analytic framework

For fixed non negative integers N and l we shall denote by $\mathcal{P}_{l,N}$ the distinguished class of functions which are homogeneous polynomials of degree l in $p \in \mathbf{R}^n$ and trigonometric (nonhomogeneous) polynomials of degree N in $q \in \mathbf{T}^n$. In particular, $\mathcal{P}_{0,N}$ is the class of trigonometric polynomials of degree N in q independent of p , $\mathcal{P}_{l,0}$ is the class of homogeneous polynomials of degree l in p independent of q , and $\mathcal{P}_{0,0}$ turns out to be the class of constants. It is convenient to let $0 \in \mathcal{P}_{l,N}$ for all l, N , so that the class $\mathcal{P}_{l,N}$ has the structure of a linear space. A function $f \in \mathcal{P}_{l,N}$ can be written as

$$f(p, q) = \sum_{|j|=l} \sum_{|k| \leq N} f_{j,k} p^j \exp(ik \cdot q). \quad (24)$$

Here, $j \in \mathbf{Z}^n$ is an integer vector with non negative components; $k \in \mathbf{Z}^n$; $|k| = |k_1| + \dots + |k_n|$, and similarly for $|j|$; $f_{j,k} \in \mathbf{C}$ are complex coefficients; moreover we use the multi index notation $p^j = p_1^{j_1} \dots p_n^{j_n}$ and $k \cdot q = k_1 q_1 + \dots + k_n q_n$. We shall also use the notation $\langle \cdot \rangle$ for the average of a function over the angles q .

The following algebraic properties are immediate: if $f \in \mathcal{P}_{l,N}$ and $f' \in \mathcal{P}_{l',N'}$ then:

- (i) if $l = l'$ then $f + f' \in \mathcal{P}_{l, \max(N, N')}$.
- (ii) for $1 \leq j \leq n$ one has $\frac{\partial f}{\partial p_j} \in \mathcal{P}_{l-1, N}$ and $\frac{\partial f}{\partial q_j} \in \mathcal{P}_{l, N}$; in particular for $l = 0$ one has $\frac{\partial f}{\partial p_j} = 0$, and for $N = 0$ one has $\frac{\partial f}{\partial q_j} = 0$.
- (iii) $ff' \in \mathcal{P}_{l+l', N+N'}$.

- (iv) $\{f, f'\} \in \mathcal{P}_{l+l'-1, N+N'}$; in particular, $\{f, f'\} = 0$ in either case $l = l' = 0$ or $N = N' = 0$. Here, $\{\cdot, \cdot\}$ denotes the Poisson bracket.

We introduce norms on the spaces $\mathcal{P}_{l,N}$ as follows. Recall that a function $f \in \mathcal{P}_{l,N}$ can be written as in (24); having fixed a positive parameter σ we define the norm of f as

$$\|f\|_\sigma = \sum_{|j|=l} \sum_{|k| \leq N} |f_{j,k}| \exp(|k|\sigma) . \tag{25}$$

The norm is well defined, due to the finite number of terms in the sum. By the way, the definition is correct also in the case of an infinite Fourier expansion, provided f is analytic in q and σ is small enough. A useful property of this norm is that one has

$$|f(p, q)| \leq \|f\|_\sigma \varrho^l \quad \text{for all } (p, q) \in B_\varrho \times \mathbf{T}_\sigma^n . \tag{26}$$

3.2 Reformulation of the theorem

Here we reformulate the main theorem in a form adapted to our algebraic scheme.

Proposition 1. *Consider a real analytic Hamiltonian*

$$H(p, q) = \sum_{j=1}^n \omega_j p_j + \sum_{l \geq 2} h_l^{(0)}(p) + \sum_{s > 0} \sum_{l \geq 0} f_l^{(0,s)}(p, q) , \tag{27}$$

with $h_l^{(0)} \in \mathcal{P}_{l,0}$ and $f_l^{(0,s)} \in \mathcal{P}_{l,sK}$ for some positive integer K . Assume that (5) and (7) are satisfied with positive γ, τ and m . Assume moreover

$$\|h_l^{(0)}\|_\sigma \leq E \eta_0^l , \quad \|f_l^{(0,s)}\|_\sigma \leq \varepsilon_0 E \mu^s \eta_0^l \tag{28}$$

with positive constants $E, \varepsilon_0, \mu, \eta_0$ and σ . Then there exist positive constants $\varrho, \bar{\varepsilon}$ and $\bar{\mu}$ such that the following statement holds true: if $\varepsilon_0 < \bar{\varepsilon}$ and $\mu < \bar{\mu}$ then there exists a real analytic canonical transformation \mathcal{C} satisfying

$$B_{\varrho/4}(0) \times \mathbf{T}_{\sigma/4}^n \subset \mathcal{C}(B_{\varrho/2}(0) \times \mathbf{T}_{\sigma/2}^n) \subset B_{3\varrho/4}(0) \times \mathbf{T}_{3\sigma/4}^n \tag{29}$$

which changes the Hamiltonian to Kolmogorov’s normal form

$$H^{(\infty)}(p, q) = \sum_j \omega_j p_j + h^{(\infty)}(p, q) , \tag{30}$$

$h^{(\infty)}$ being a power series in p starting with terms of degree 2. The canonical transformation is close to the identity, in the sense that one has $|p - p'| = O(\varepsilon_0^{1/2})$, $|q - q'| = O(\varepsilon_0^{1/2})$, where $(p', q') = \mathcal{C}(p, q)$.

Estimated values are

$$\varrho = \frac{1}{\eta_0} \tag{31}$$

$$\bar{\mu} = \frac{1}{3 \cdot 2^{2\tau+21} (\zeta(3/2))^4}, \quad \zeta(s) = \sum_{j \geq 1} 1/j^s \tag{32}$$

$$\bar{\varepsilon} = \min \left\{ 1, \left[4E\eta_0 \left(\frac{K^\tau}{\gamma\sigma} + \frac{2e\eta_0}{m} \right) \right]^{-2}, \left[\frac{2E\eta_0 K^\tau}{\gamma\sigma} \left(1 + \frac{4E\eta_0 K^\tau}{e\gamma\sigma} \right) \right]^{-2} \right\} \tag{33}$$

$$|p - p'| \leq \frac{1}{8}\varepsilon_0^{1/2} \varrho, \quad |q - q'| \leq \frac{1}{8}\varepsilon_0^{1/2} \sigma. \tag{34}$$

The main theorem will be obtained from this proposition in sect. 4.4.

3.3 Small divisors

Instead of using the explicit diophantine condition (5) we may fix a positive constant Ω and a sequence $\{\alpha_r\}_{r \geq 1}$ of positive real numbers such that

$$|k \cdot \omega| \geq \Omega \alpha_r \quad \text{for } 0 < |k| \leq rK, \quad k \in \mathbf{Z}^n, \tag{35}$$

and

$$\alpha_r \leq 1 \quad \text{for all } r \geq 1. \tag{36}$$

The latter condition can be satisfied, of course, via an appropriate choice of Ω .

Recalling that the small divisors are expected to accumulate according to the rules stated at the end of sect. 2, we introduce a special function defined as follows. Let $\{\vartheta_r\}_{r \geq 1}$ be a sequence of positive real numbers, with $\vartheta_r \leq 1$ for all r . Let $r \geq 0$ and $s \geq 0$ be integers, and $q \geq 0$ be real. Let also $\mathcal{J}_{r,s,q}$ for $r \geq 1$ and $s \geq 1$ be defined as the set of integers arrays the elements of which are positive and not exceeding r , the array containing at most s elements, and satisfying the selection rule that the sum of the base 2 logarithms of the elements does not exceed q . Formally:

$$\mathcal{J}_{r,s,q} = \left\{ j = \{j_1, \dots, j_k\} : j_m \in \{1, \dots, r\}, 1 \leq k \leq s, \sum_{m=1}^k \log_2 j_m \leq q \right\}. \tag{37}$$

Then we define the function $T(\vartheta, r, s, q)$ as

$$\begin{aligned} T(\vartheta, 0, s, q) &= T(\vartheta, r, 0, q) = 1, \\ T(\vartheta, r, s, q) &= \max_{j \in \mathcal{J}_{r,s,q}} \prod_m \vartheta_{j_m}^{-1}. \end{aligned} \tag{38}$$

In the following, we shall refer to the argument s as the *number of divisors*, and to the argument q as the *selection rule*. The following inequalities are immediate:

$$\begin{aligned} \text{if } r' \leq r & \quad \text{then } T(\vartheta, r', s, q) \leq T(\vartheta, r, s, q), \\ \text{if } s' \leq s & \quad \text{then } T(\vartheta, r, s', q) \leq T(\vartheta, r, s, q), \\ \text{if } q' \leq q & \quad \text{then } T(\vartheta, r, s, q') \leq T(\vartheta, r, s, q). \end{aligned} \tag{39}$$

The condition $\vartheta_r \leq 1$ is required here for the cases $r' = 0$ or $s' = 0$. It is also an easy matter to prove that the following inequalities hold:

$$T(\vartheta, r, s, q) T(\vartheta, r', s', q') \leq T(\vartheta, \max(r, r'), s + s', q + q') , \tag{40}$$

$$T(\vartheta, r, s, q) \leq \vartheta_{r'} T(\vartheta, \max(r, r'), s + 1, q + \log_2 r') . \tag{41}$$

In order to prove (40) just note that in view of definition (37) the union of two arrays $j \in \mathcal{J}_{r,s,q}$ and $j' \in \mathcal{J}_{r',s',q'}$ does belong to $\mathcal{J}_{\max(r,r'),s+s',q+q'}$; then the proof is trivial. In order to prove (41), note that by definition (38) one has

$$\vartheta_{r'}^{-1} \leq T(\vartheta, r', 1, \log_2 r') ;$$

then use (40).

4. Scheme of proof

In the spirit of Kolmogorov’s scheme, we construct an infinite sequence $\{H^{(r)}\}_{r \geq 0}$ of Hamiltonians, with the condition that $H^{(r)}$ is in Kolmogorov’s normal form up to order r . To this end, we perform an infinite sequence of perturbation steps, each consisting of a near to identity canonical transformation which changes $H^{(r-1)}$ to $H^{(r)}$. The transformation removes all terms of order r which are not yet in normal form. The canonical transformation at step r is generated via composition of two Lie series of the form

$$\exp(L_{\chi_2^{(r)}}) \circ \exp(L_{\chi_1^{(r)}}) \tag{42}$$

where

$$\chi_1^{(r)}(q) = X^{(r)}(q) + \sum_i \xi_i^{(r)} q_i , \quad \chi_2^{(r)}(p, q) = \sum_i Y_i^{(r)}(q) p_i \tag{43}$$

with $\xi^{(r)} \in \mathbf{R}^n$, $X^{(r)} \in \mathcal{P}_{0,rK}$ and $\chi_2^{(r)} \in \mathcal{P}_{1,rK}$. Here, as usual, we denote $L_g \cdot = \{g, \cdot\}$. The quantities $X^{(r)}(q)$, $Y_i^{(r)}(q)$ and $\xi^{(r)}$ are unknowns to be determined so that $H^{(r)}$ is in Kolmogorov’s normal form up to order r .

4.1 Formal algorithm

After r steps the Hamiltonian is written as

$$H^{(r)} = \sum_i \omega_i p_i + \sum_{s=0}^r \sum_{l \geq 2} h_l^{(s)} + \sum_{s > r} \sum_{l \geq 0} f_l^{(r,s)} , \tag{44}$$

where $h_l^{(s)}$ is of class $\mathcal{P}_{l,sK}$ and $f_l^{(r,s)}$ is of class $\mathcal{P}_{l,sK}$. We can always assume that $\langle f_0^{(r-1,r)} \rangle = 0$, because a constant in the Hamiltonian can be neglected. The functions are recursively defined as follows. The terms of order r of the Hamiltonian are

$$h_l^{(r)} = L_{\chi_2^{(r)}} h_l^{(0)} + \hat{f}_l^{(r,r)} \quad \text{for } l \geq 2 . \tag{45}$$

The perturbation is

$$\begin{aligned}
 f_0^{(r,kr+m)} &= \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2^{(r)}}^j \hat{f}_0^{(r,(k-j)r+m)} \quad \text{for } k \geq 1, 0 \leq m < r, kr+m > r; \\
 f_1^{(r,kr)} &= \frac{k-1}{k!} L_{\chi_2^{(r)}}^{k-1} \hat{f}_1^{(r,r)} + \sum_{j=0}^{k-2} \frac{1}{j!} L_{\chi_2^{(r)}}^j \hat{f}_1^{(r,(k-j)r)} \quad \text{for } k \geq 2; \\
 f_1^{(r,kr+m)} &= \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2^{(r)}}^j \hat{f}_1^{(r,(k-j)r+m)} \quad \text{for } k \geq 1, 0 < m < r; \\
 f_l^{(r,kr+m)} &= \frac{1}{k!} L_{\chi_2^{(r)}}^k h_l^{(m)} + \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2^{(r)}}^j \hat{f}_l^{(r,(k-j)r+m)} \\
 &\quad \text{for } l \geq 2, k \geq 1, 0 \leq m < r, kr+m > r.
 \end{aligned} \tag{46}$$

The functions \hat{f} are given by

$$\begin{aligned}
 \hat{f}_0^{(r,r)} &= 0, \\
 \hat{f}_0^{(r,r+m)} &= f_0^{(r-1,r+m)} \quad \text{for } 0 < m < r; \\
 \hat{f}_l^{(r,kr+m)} &= \frac{1}{k!} L_{\chi_1^{(r)}}^k h_{l+k}^{(m)} + \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_1^{(r)}}^j f_{l+j}^{(r-1,(k-j)r+m)} \\
 &\quad \text{for } l \geq 0, k \geq 1, l+k > 1, 0 \leq m < r.
 \end{aligned} \tag{47}$$

The generating functions $\chi_1^{(r)} = X^{(r)} + \sum_i \xi_i^{(r)} q_i$ and $\chi_2^{(r)}$ are determined by the equations

$$\sum_i \omega_i \frac{\partial X^{(r)}}{\partial q_i} + f_0^{(r-1,r)} = 0, \tag{48}$$

$$\sum_j C_{ij} \xi_j^{(r)} + b_i^{(r)} = 0, \tag{49}$$

$$\sum_i \omega_i \frac{\partial \chi_2^{(r)}}{\partial q_i} + \{X^{(r)}, h_2^{(0)}\} + \tilde{f}_1^{(r-1,r)} = 0. \tag{50}$$

where

$$\tilde{f}_1^{(r-1,r)} = f_1^{(r-1,r)} - \langle f_1^{(r-1,r)} \rangle,$$

and the real matrix C and the real vector $b^{(r)}$ are defined as

$$C_{ij} = \frac{\partial^2 h_2^{(0)}}{\partial p_i \partial p_j}, \quad b_i^{(r)} = \frac{\partial \langle f_1^{(r-1,r)} \rangle}{\partial p_i}. \tag{51}$$

This algorithm is obtained by writing in explicit form the exponential operators in (42). The details are deferred to the technical section 5.1

4.2 Estimates depending on parameters

We now translate our formal algorithm, as expressed in section 4.1, into a recursive scheme of estimates on the norms of the various functions. To this end, having fixed $d \in \mathbf{R}$, $0 < d \leq 1/4$, we consider a sequence $\{\delta_r\}_{r \geq 1}$ of positive real numbers satisfying

$$\delta_{r+1} \leq \delta_r, \quad \sum_{r \geq 1} \delta_r \leq \frac{d}{2}. \tag{52}$$

Moreover, we introduce a further sequence $\{d_r\}_{r \geq 0}$ of positive real numbers recursively defined as

$$d_0 = 0, \quad d_r = d_{r-1} + 2\delta_r. \tag{53}$$

With this setting we look for estimates on the Hamiltonian and on the generating functions at every normalization step r . However, we need first two elementary technical estimates concerning on the one hand the generalization of Cauchy’s estimates in view of our choice of the norm, and, on the other hand, the estimate of the generating functions as determined by equations (48)–(51).

Lemma 1. *Let $\chi_1^{(r)}$ and $\chi_2^{(r)}$ be as in (43), and let $f \in \mathcal{P}_{l,N}$. Then for every positive $d < 1$ one has*

$$\left\| \frac{\partial \chi_1^{(r)}}{\partial q} \right\|_{(1-d)\sigma} \leq \frac{\|X^{(r)}\|_\sigma}{ed\sigma} + |\xi^{(r)}|, \quad \text{for} \tag{54}$$

$$\left\| \frac{\partial \chi_2^{(r)}}{\partial q} \right\|_{(1-d)\sigma} \leq \frac{\|\chi_2^{(r)}\|_\sigma}{ed\sigma}, \quad \text{for} \tag{55}$$

$$\left\| \frac{\partial \chi_2^{(r)}}{\partial p} \right\|_{(1-d)\sigma} \leq \|\chi_2^{(r)}\|_\sigma, \quad \text{for} \tag{56}$$

$$\left\| L_{\chi_1^{(r)}}^s f \right\|_{(1-d)\sigma} \leq s! \binom{l}{s} \left(\frac{\|X^{(r)}\|_\sigma}{d\sigma} + e|\xi^{(r)}| \right)^s \|f\|_\sigma \quad \text{for } 0 \leq s \leq l, \tag{57}$$

$$\left\| L_{\chi_2^{(r)}}^s f \right\|_{(1-d)\sigma} \leq s! \binom{l+s}{s} \left(\frac{\|\chi_2^{(r)}\|_\sigma}{d\sigma} \right)^s \|f\|_\sigma \quad \text{for } s \geq 0. \tag{58}$$

Note that in view of $f \in \mathcal{P}_{l,N}$ one has $L_{\chi_1^{(r)}}^s f = 0$ for $s > l$. The proof is deferred to the technical section 5.2. We emphasize that the estimates do not depend on the order of truncation of the Fourier expansion, namely rK and N . In fact the estimates hold true also in the case of infinite Fourier series, provided all the functions are analytic.

Lemma 2. *The equations (48)–(50) admit solutions $X^{(r)} \in \mathcal{P}_{0,rK}$, $\xi^{(r)} \in \mathbf{R}^n$ and*

$\chi_2^{(r)} \in \mathcal{P}_1, rK$ satisfying

$$\|X^{(r)}\|_{(1-d_{r-1})\sigma} \leq \frac{\|f_0^{(r-1,r)}\|_{(1-d_{r-1})\sigma}}{\alpha_r \Omega}, \tag{59}$$

$$|\xi^{(r)}| \leq \frac{1}{m} \|f_1^{(r-1,r)}\|_{(1-d_{r-1})\sigma}, \tag{60}$$

$$\|\chi_2^{(r)}\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \frac{1}{\alpha_r \Omega} \left(\frac{2\|f_0^{(r-1,r)}\|_{(1-d_{r-1})\sigma} \|h_2^{(0)}\|_{\sigma}}{\alpha_r \delta_r e \Omega \sigma} + \|f_1^{(r-1,r)}\|_{(1-d_{r-1})\sigma} \right) \tag{61}$$

Here, the fact that at order r we work with Fourier expansions truncated at order rK is crucial, of course. The proof is deferred to the technical section 5.2

Lemma 3. *Let $H(p, q)$ be as in proposition 1 and satisfy the nondegeneracy condition (7). Let the sequences $\{\alpha_r\}_{r \geq 1}$, $\{\delta_r\}_{r \geq 1}$ and $\{d_r\}_{r \geq 0}$ be determined according to (35), (52) and (53) respectively. Given η_0 as in proposition 1, define the sequence $\{\eta_r\}_{r \geq 1}$ as*

$$\eta_r = (1 + \delta_r)(1 + \alpha_r \delta_r^2) \eta_{r-1}. \tag{62}$$

Define also the sequences $\{a_l\}_{l \geq 0}$ and $\{b_l\}_{l \geq 0}$ as

$$a_0 = 2, \quad b_0 = 0, \quad a_1 = 1, \quad b_1 = 1, \quad \text{and} \quad a_l = 0, \quad b_l = 2 \quad \text{for } l \geq 2. \tag{63}$$

Assume

$$\varepsilon_0 \leq 1, \tag{64}$$

$$4\varepsilon_0^{1/2} E \eta_0 \left(\frac{1}{\Omega \sigma} + \frac{2e\eta_0}{m} \right) \leq 1, \tag{65}$$

$$2\varepsilon_0^{1/2} E \eta_0 \frac{1}{\Omega \sigma} \left(1 + \frac{4E\eta_0}{e\Omega \sigma} \right) \leq 1. \tag{66}$$

Then at every step $r > 0$ in the normalization procedure the following estimates hold: for the generating functions we have

$$\frac{\|X^{(r)}\|_{(1-d_{r-1})\sigma}}{e\delta_r \sigma} + |\xi^{(r)}| \leq \frac{\varepsilon_0^{1/2}}{2e\eta_0} (24\mu)^r T(\alpha\delta^2, r, 2r - 1, 2r - 2 - \log_2 r), \tag{67}$$

$$\|\chi_2^{(r)}\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \varepsilon_0^{1/2} \delta_r^2 \sigma (24\mu)^r T(\alpha\delta^2, r, 2r, 2r - 2); \tag{68}$$

for the functions $h^{(r)}$ and $f_l^{(r,s)}$ we have

$$\left\| h_l^{(r)} \right\|_{(1-d_r)\sigma} \leq \frac{1}{8} \varepsilon_0^{1/2} E \eta_r^l (24\mu)^r T(\alpha\delta^2, r, 2r, 2r - 2), \tag{69}$$

$$\left\| f_l^{(r,s)} \right\|_{(1-d_r)\sigma} \leq \frac{1}{8} \varepsilon_0 E \eta_r^l (24\mu)^s T(\alpha\delta^2, r, 2s - a_l, 2s - 2 + b_l \log_2 r - 2 \log_2 s) \tag{70}$$

where $l \geq 2$ in (69), and $l \geq 0, s \geq r$ in (70).

Here, $\alpha\delta^2$ denotes the sequence $\{\alpha_r \delta_r^2\}$. The proof is deferred to sect. 5.2.

4.3 Choice of parameters and proof of proposition 1

In this section we make a choice for the sequences $\{\alpha_r\}_{r \geq 1}$ and $\{\delta_r\}_{r \geq 1}$, and for the parameter d . According to the diophantine condition (5), which was assumed by proposition 1, we set

$$\alpha_r = r^{-\tau} \ , \quad \Omega = \gamma K^{-\tau} \ , \tag{71}$$

with $\gamma > 0$ depending on the frequencies ω , and with $\tau > n - 1$. We also set

$$\delta_r = \frac{d}{Br^{3/2}} \ , \quad B = 2\zeta(3/2) \ , \tag{72}$$

where $\zeta(s)$ is the Riemann zeta function; with this choice, condition (52) is fulfilled. Finally we set

$$d = \frac{1}{4} \ . \tag{73}$$

Substituting Ω as given by (71) in the hypotheses (64), (65) and (66) of lemma 3 we readily recover the condition (33) of proposition 1. Thus, lemma 3 applies. We now specialize the estimates with our choice of the constants. We get

$$\frac{\|X^{(r)}\|_{(1-d_{r-1})\sigma}}{e\delta_r\sigma} + |\xi^{(r)}| \leq \frac{\varepsilon_0^{1/2}\delta_r^2}{2^{2\tau+7}e\eta_0r^\tau} \left(\frac{\mu}{\bar{\mu}}\right)^r \tag{74}$$

$$\|\chi_2^{(r)}\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \frac{\varepsilon_0^{1/2}\delta_r^2\sigma}{2^{2\tau+6}} \left(\frac{\mu}{\bar{\mu}}\right)^r \ , \tag{75}$$

with $\bar{\mu}$ given by

$$\bar{\mu} = \frac{d^4}{3 \cdot 2^{2\tau+9}B^4} \ ;$$

by the way, (32) is computed from this by substituting (72) and (73). In order to prove (74) and (75) we proceed as follows. Using the definition (38) of the function T , with ϑ_j replaced by $\alpha_j\delta_j^2$, we get

$$T(\alpha\delta^2, r, s, q) = \max_{j \in \mathcal{J}_{r,s,q}} \prod_m \frac{1}{\alpha_{j_m}\delta_{j_m}^2} = \left(\frac{B}{d}\right)^{2s} \max_{j \in \mathcal{J}_{r,s,q}} \prod_m j_m^{\tau+3}$$

(recall that the number of divisors is at most s , and that $B/d > 1$). Furthermore, in view of the selection rule we have

$$\log_2 \prod_m j_m^{\tau+3} = (\tau + 3) \sum_m \log_2 j_m \leq (\tau + 3)q \ .$$

This holds for every $j \in \mathcal{J}_{r,s,q}$. We conclude

$$T(\alpha\delta^2, r, s, q) \leq \left(\frac{B}{d}\right)^{2s} 2^{(\tau+3)q} \ . \tag{76}$$

Substituting this inequality in (68) we readily get (75). In order to obtain (74) from (67) we use also $T(\alpha\delta^2, r, 2r - 1, 2r - 2 - \log_2 r) \leq \alpha_r\delta_r^2 T(\alpha\delta^2, r, 2r, 2r - 2)$ and $\alpha_r = r^{-\tau}$.

We use now (74) and (75) in order to estimate the canonical transformation. We denote by $(p^{(0)}, q^{(0)})$ the original coordinates, and by $(p^{(r)}, q^{(r)})$ the coordinates at step r . We also denote by $\hat{\mathcal{C}}^{(r)}$ the canonical transformation mapping $(p^{(r)}, q^{(r)})$ to $(p^{(r-1)}, q^{(r-1)})$. The transformation is written explicitly as

$$\begin{aligned}
 p^{(r-1)} &= \exp(L_{\chi_1^{(r)}}) \hat{p}^{(r)} = \hat{p}^{(r)} + \frac{\partial \chi_1^{(r)}}{\partial q^{(r-1)}} \\
 \hat{p}^{(r)} &= \exp(L_{\chi_2^{(r)}}) p^{(r)} = p^{(r)} + \sum_{s \geq 1} \frac{1}{s!} L_{\chi_2^{(r)}}^{s-1} \frac{\partial \chi_2^{(r)}}{\partial q^{(r)}} \\
 q^{(r-1)} &= \exp(L_{\chi_2^{(r)}}) q^{(r)} = q^{(r)} - \sum_{s \geq 1} \frac{1}{s!} L_{\chi_2^{(r)}}^{s-1} \frac{\partial \chi_2^{(r)}}{\partial p^{(r)}}.
 \end{aligned} \tag{77}$$

Note that the generating function $\chi_1^{(r)}$ does not change the angles. Consider now a sequence of domains $B_{(3d-d_r)\varrho}(0) \times \mathbf{T}_{(3d-d_r)\sigma}^n$, with d_r and ϱ given by (53) and by (31) respectively. Using lemma 1, (31), (73) and (74), and recalling (26), we get the estimate

$$\begin{aligned}
 |p^{(r-1)} - \hat{p}^{(r)}| &< \frac{\varepsilon_0^{1/2} \delta_r^2 \varrho}{2^{2\tau+7} e^{r\tau}} \left(\frac{\mu}{\bar{\mu}} \right)^r \\
 |\hat{p}^{(r)} - p^{(r)}| &< \frac{\varepsilon_0^{1/2} \delta_r \varrho}{2^{2\tau+6}} \left(\frac{\mu}{\bar{\mu}} \right)^r \sum_{s \geq 1} \left[\frac{\varepsilon_0^{1/2} \delta_r}{2^{2\tau+6}} \left(\frac{\mu}{\bar{\mu}} \right)^r \right]^{s-1} \\
 |q^{(r-1)} - q^{(r)}| &< \frac{\varepsilon_0^{1/2} \delta_r^2 \sigma}{2^{2\tau+6}} \left(\frac{\mu}{\bar{\mu}} \right)^r \sum_{s \geq 1} \left[\frac{\varepsilon_0^{1/2} \delta_r}{2^{2\tau+6}} \left(\frac{\mu}{\bar{\mu}} \right)^r \right]^{s-1}.
 \end{aligned}$$

In view of the condition $\mu < \bar{\mu}$ the series in the latter estimates converges, with

$$\sum_{s \geq 1} \left[\frac{\varepsilon_0^{1/2} \delta_r}{2^{2\tau+6}} \left(\frac{\mu}{\bar{\mu}} \right)^r \right]^{s-1} < 2.$$

Thus, the series (77) defining the canonical transformation is absolutely convergent in the domain $B_{(3d-d_{r-1}-\delta_r)\varrho}(0) \times \mathbf{T}_{(3d-d_{r-1}-\delta_r)\sigma}^n$, and so, by Weierstrass theorem, is analytic. Furthermore, one has the estimates

$$|p^{(r-1)} - p^{(r)}| < \varepsilon_0^{1/2} \delta_r \varrho, \quad |q^{(r-1)} - q^{(r)}| < \varepsilon_0^{1/2} \delta_r \sigma.$$

A similar argument applies to the inverse of $\hat{\mathcal{C}}^{(r)}$, which is defined as a composition of Lie series generated by $-\chi_2^{(r)}$ and $-\chi_1^{(r)}$. We conclude that

$$B_{(3d-d_r)\varrho} \times \mathbf{T}_{(3d-d_r)\sigma}^n \subset \hat{\mathcal{C}}^{(r)}(B_{(3d-d_{r-1}-\delta_r)\varrho} \times \mathbf{T}_{(3d-d_{r-1}-\delta_r)\sigma}^n) \subset B_{(3d-d_{r-1})\varrho} \times \mathbf{T}_{(3d-d_{r-1})\sigma}^n.$$

Consider now the sequence of transformations $\mathcal{C}^{(r)} = \hat{\mathcal{C}}^{(1)} \circ \dots \circ \hat{\mathcal{C}}^{(r)}$. For $(p^{(r-1)}, q^{(r-1)}) \in B_{(3d-d_{r-1})\delta_r} \times \mathbf{T}_{(3d-d_{r-1})\sigma}^n$ the transformation is clearly analytic,

and one has

$$|p^{(0)} - p^{(r)}| < \varepsilon_0^{1/2} \varrho \sum_{j=1}^r \delta_j, \quad |q^{(0)} - q^{(r)}| < \varepsilon_0^{1/2} \sigma \sum_{j=1}^r \delta_j.$$

On the other hand, by (72) and (73) one has $\sum_{j \geq 1} \delta_j \leq d/2 = 1/8$, and so, by Weierstrass theorem, the sequence $\mathcal{C}^{(r)}$ converges to an analytic canonical transformation \mathcal{C} satisfying (34) and (29).

It remains to prove that the transformed Hamiltonian $H \circ \mathcal{C}$ is given the form (30), where $h^{(\infty)}$, according to (44), is in Kolmogorov's normal form, being

$$h^{(\infty)} = \sum_{s \geq 0} \sum_{l \geq 2} h_l^{(s)}.$$

To this end, we first prove that the sequence of Hamiltonians $\{H^{(r)}\}_{r \geq 0}$ given by (44) actually converges to an analytic Hamiltonian, $H^{(\infty)}$ say. Using (69), (70) and (76) we get

$$\|h_l^{(r)}\|_{(1-d_r)\sigma} \leq \frac{\varepsilon_0^{1/2} E}{2^{2\tau+9}} \left(\frac{\mu}{\bar{\mu}}\right)^r \eta_r^l, \quad \|f_l^{(r,s)}\|_{(1-d_r)\sigma} \leq \frac{\varepsilon_0 E}{2^{2\tau+9}} \left(\frac{\mu}{\bar{\mu}}\right)^s \left(\frac{r}{s}\right)^{2\tau+6} \eta_r^l. \tag{78}$$

Recalling now that we are considering $H^{(r)}$ as defined on the domain $B_{(3d-d_{r-1}-\delta_r)\varrho}(0) \times \mathbf{T}_{(3d-d_{r-1}-\delta_r)\sigma}^n$, where the canonical transformation $\mathcal{C}^{(r)}$ is defined, for all (p, q) in that domain we estimate

$$|h_l^{(r)}(p, q)| < \|h_l^{(r)}\|_{(1-d_r)\sigma} (3\varrho/4)^l, \quad |f_l^{(r,s)}(p, q)| < \|f_l^{(r,s)}\|_{(1-d_r)\sigma} (3\varrho/4)^l. \tag{79}$$

We need now the elementary inequality

$$\ln \prod_{r \geq 1} (1 + \delta_r)(1 + \alpha_r \delta_r^2) < \sum_{r \geq 1} \delta_r (1 + \alpha_r \delta_r) < (1 + \frac{d}{2}) \sum_{r \geq 1} \delta_r < (1 + \frac{d}{2}) \frac{d}{2} = \frac{9}{64}$$

(use $\alpha_r \leq 1$, $\sum_r \delta_r \leq d/2$ and $d = 1/4$). From this we get $\eta_r < e^{9/64} \eta_0$ and, using $\varrho = 1/\eta_0$, also $3\eta_r \varrho/4 < 1$. Putting this inequality in (79) and (78) and adding up we get

$$\sum_{l \geq 2} |h_l^{(r)}(p, q)| < C_1 \left(\frac{\mu}{\bar{\mu}}\right)^r, \quad \sum_{s > r} \sum_{l \geq 0} |f_l^{(r,s)}(p, q)| < C_2 \left(\frac{\mu}{\bar{\mu}}\right)^{r+1},$$

where C_1 and C_2 are finite constants, the actual value of which is not relevant. Thus, recalling

$$H^{(r)} - H^{(r-1)} = \sum_{l \geq 2} h_l^{(r)} + \sum_{s > r} \sum_{l \geq 0} f_l^{(r,s)} - \sum_{s \geq r} \sum_{l \geq 0} f_l^{(r-1,s)},$$

in the domain $B_{(3d-d_{r-1}-\delta_r)\varrho}(0) \times \mathbf{T}_{(3d-d_{r-1}-\delta_r)\sigma}^n$ we have the uniform estimate

$$|H^{(r)}(p, q) - H^{(r-1)}(p, q)| < C_3 \left(\frac{\mu}{\bar{\mu}}\right)^r$$

with some positive constant C_3 . In view of $\mu < \bar{\mu}$, this proves that the sequence of functions $H^{(r)}$ is uniformly convergent in the domain $B_{\rho/2} \times \mathbf{T}_{\sigma/2}^n$, and so, by Weierstrass theorem, it defines an analytic Hamiltonian $H^{(\infty)}$.

The proof that $H^{(\infty)} = H \circ \mathcal{C}$ is a straightforward consequence of a well known property of Lie series. Indeed, given a canonical mapping $\tilde{\mathcal{C}}^{(r)}$ of the form (77) and any analytic function $g(p, q)$ on the domain where the mapping is defined, then one has $g \circ \tilde{\mathcal{C}}^{(r)} = \exp(L_{\chi_2^{(r)}}) \circ \exp(L_{\chi_1^{(r)}})g$. A repeated use of this property allows us to establish that $H^{(r)} = H \circ \mathcal{C}^{(r)}$ for every r , and the conclusion $H^{(\infty)} = H \circ \mathcal{C}$ easily follows. This concludes the proof of proposition 1.

4.4 Proof of the main theorem

We give the Hamiltonian (1) the form (27), also producing explicit estimates for the constants $\varepsilon_0, \eta_0, E, \mu$ and σ . Having fixed the point p^* we translate the origin of the action variables in p^* via the canonical transformation $p' = p - p^*, q' = q$. Then we expand both $h(p)$ and $f(p, q, \varepsilon)$ in power series of p' around the origin, and get (primes are omitted, and an unessential constant is ignored)

$$h(p) = \sum_{j=1}^n \omega_j p_j + \sum_{l \geq 2} h_l(p), \quad f(p, q, \varepsilon) = \sum_{l \geq 0} f_l(p, q, \varepsilon),$$

$h_l(p)$ and $f_l(p, q, \varepsilon)$ denoting homogeneous polynomials of degree l in p . Finally, we pick an arbitrary positive integer K and split the expansion of $f_l(p, q, \varepsilon)$ in Fourier series of the angles as

$$f_l^{(1)} = \sum_{0 \leq |k| \leq K} c_{l,k}(p, \varepsilon) \exp(ik \cdot q),$$

$$f_l^{(s)} = \sum_{(s-1)K < |k| \leq sK} c_{l,k}(p, \varepsilon) \exp(ik \cdot q) \quad \text{for } s > 1.$$

Identifying h_l and $f_l^{(s)}$ with $h_l^{(0)}$ and $f_l^{(0,s)}$ respectively, the form (27) is achieved.

By the analyticity assumptions there are constants $\varepsilon_0, E, \eta_0, \mu$ and σ such that (28) is satisfied. Estimated values are

$$\eta_0 = \frac{2}{\delta}, \quad E = 2^{n-1}E_0, \quad \mu = e^{-K\xi/4}, \quad \varepsilon_0 = \varepsilon e^{K\xi/4} \left(\frac{1 + e^{-\xi/4}}{1 - e^{-\xi/4}} \right)^n \frac{F_0}{E_0}, \quad \sigma = \frac{\xi}{2}, \tag{80}$$

E_0 and F_0 being given by (11). The estimate of the constants η_0 and E is a straightforward application of Cauchy’s estimate to Taylor’s expansion of an analytic function. The estimate of μ, σ and ε_0 uses also the exponential decay of the coefficients in Fourier expansion of a periodic analytic function, and can be found in [21], lemma 8.

Thus, we have given the Hamiltonian (1) a form to which proposition 1 can be applied, provided the conditions $\varepsilon_0 \leq \bar{\varepsilon}$ and $\mu \leq \bar{\mu}$ are satisfied. We show that this is possible by just computing the explicit estimates for the main theorem.

The condition on μ can be easily satisfied by suitably determining the parameter K . For, in view of (80), we ask $e^{-K\xi/4} \leq \bar{\mu}$; on the other hand, in view of (33) we are interested in keeping K as small as possible; this gives (16). Thus, we are left only with the condition $\varepsilon_0 \leq \bar{\varepsilon}$; in view of (80) this gives (15) and (12). Finally, ϱ is determined by (31), and this, using again (80), gives (13). This concludes the proof.

5. Proofs and technical lemmas

In this section we work out in detail the proof of all lemmas of sect. 4. The formal part of sect. 5.1 is nothing but an explicit formulation of the normalization step with use of the Lie series algorithm. It reduces in fact to a suitable reordering of terms. Some estimates of sect. 5.2 (proofs of lemmas 1 and 2) are just an adaptation of known estimates to our norm. The main part, which includes the mechanism of control of the small divisors, is the proof of lemma 3 in sect. 5.3.

5.1 Justification of the formal algorithm

We use induction. For $r = 0$ the form (44) of the Hamiltonian coincides with (27). Assuming that the Hamiltonian has the form (44) with $r - 1$ in place of r , we perform the canonical transformation (42), thus transforming the Hamiltonian $H^{(r-1)}$ to $H^{(r)}$. The proof of the lemma is nothing but an algebraic calculation which leads to the recursive transformation formulæ stated in the lemma.

It is convenient to reorder all terms in the Hamiltonian $H^{(r-1)}$, splitting it into blocks of size r . To this end we replace the index $s \geq r$ in the perturbation with the index $kr + m$ with $k \geq 1$ and $0 \leq m < r$, and write

$$H^{(r-1)} = \sum_i \omega_i p_i + \sum_{s=0}^{r-1} \sum_{l \geq 2} h_l^{(s)} + \sum_{k \geq 1} \sum_{m=0}^{r-1} \sum_{l \geq 0} f_l^{(r-1,kr+m)}. \tag{81}$$

We first compute $\exp(L_{\chi_1^{(r)}})H^{(r-1)}$. To this end we notice that by (48) one has $X^{(r)} \in \mathcal{P}_{0,rK}$. A straightforward application of the properties (ii) and (iv) of sect. 3.1 gives $L_{\chi_1^{(r)}}(\mathcal{P}_{l,sK}) \subset \mathcal{P}_{l-1,(s+r)K}$ for $l \geq 1$, and $L_{\chi_1^{(r)}}(\mathcal{P}_{0,sK}) = \{0\}$.

Thus in the expansion of $\exp(L_{\chi_1^{(r)}})H^{(r-1)}$ one has

$$\begin{aligned} \sum_i \omega_i p_i &\in \mathcal{P}_{1,0} \\ L_{\chi_1^{(r)}}\left(\sum_i \omega_i p_i\right) &\in \mathcal{P}_{0,rK} ; \\ L_{\chi_1^{(r)}}^k h_{l+k}^{(m)} &\in \mathcal{P}_{l,(kr+m)K} \quad \text{for } l \geq 0, k \geq 1, l+k > 1, 0 \leq m < r ; \\ L_{\chi_1^{(r)}}^j f_{l+j}^{(r-1,(k-j)r+m)} &\in \mathcal{P}_{l,(kr+m)K} \quad \text{for } 0 \leq j < k, l \geq 0, k \geq 1, 0 \leq m < r ; \end{aligned} \tag{82}$$

this takes into account all terms. Now we define new functions $\hat{f}_l^{(r,kr+m)}$ by collecting all terms of class $\mathcal{P}_{l,kr+m}$, with $l \geq 0, k \geq 1$ and $0 \leq m < r$. For $l = 0, k = 1$, forgetting the constant term $\sum_i \omega_i \xi_i^{(r)}$, we define

$$\hat{f}_0^{(r,r)} = \sum_i \omega_i \frac{\partial X^{(r)}}{\partial q_i} + f_0^{(r-1,r)} .$$

For all other values of the indexes l, k, m we define $\hat{f}_l^{(r,kr+m)}$ as in the second and third of (47); we must split the definition because there is no contribution to the class $\mathcal{P}_{0,r+m}$ coming from the third of (82). We determine $X^{(r)}$ by the condition $\hat{f}_0^{(r,r)} = 0$, which is nothing but equation (48). By the way, this also justifies the first of (47). Thus, all terms independent of p at order r have been removed. We conclude that the Hamiltonian transformed with $\chi_1^{(r)}$ has the form

$$\hat{H}^{(r)} := \exp(L_{\chi_1^{(r)}}) H^{(r-1)} = \sum_i \omega_i p_i + \sum_{s=0}^{r-1} \sum_{l \geq 2} h_l^{(s)} + \sum_{k \geq 1} \sum_{m=0}^{r-1} \sum_{l \geq 0} \hat{f}_l^{(r,kr+m)} . \quad (83)$$

Remark that this is similar to the form (81), with however $\hat{f}^{(r,kr+m)}$ in place of $f^{(r-1,kr+m)}$ and with $\hat{f}_0^{(r,r)} = 0$.

We compute now $\exp(L_{\chi_2^{(r)}}) \hat{H}^{(r)}$. To this end we notice that by (50) one has $\chi_2^{(r)} \in \mathcal{P}_{1,rK}$. A straightforward application of the properties (ii) and (iv) of sect. 3.1 gives $L_{\chi_2^{(r)}}(\mathcal{P}_{l,sK}) \subset \mathcal{P}_{l,(s+r)K}$ for all s, l . Thus in the expansion of $\exp(L_{\chi_2^{(r)}}) \hat{H}^{(r)}$ one has

$$\begin{aligned} L_{\chi_2^{(r)}}^j \left(\sum_i \omega_i \cdot p_i \right) &\in \mathcal{P}_{1,jrK} && \text{for } j \geq 0; \\ L_{\chi_2^{(r)}}^j h_l^{(s)} &\in \mathcal{P}_{l,(jr+s)K} && \text{for } j \geq 0, l \geq 2, 0 \leq s < r; \\ L_{\chi_2^{(r)}}^j \hat{f}_l^{(r,(k-j)r+m)} &\in \mathcal{P}_{l,(kr+m)K} && \text{for } 0 \leq j < k, l \geq 0, k \geq 1, 0 \leq m < r; \end{aligned}$$

this takes into account all terms. Proceeding as above, we define new functions $\hat{f}_l^{(r,kr+m)}$ with $l \geq 0, k \geq 1$ and $0 \leq m < r$ by collecting all terms belonging to

the same class $\mathcal{P}_{l,(kr+m)K}$. Thus we define

$$\begin{aligned}
 f_0^{(r,kr+m)} &= \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2}^j \hat{f}_0^{(r,(k-j)r+m)} \\
 &\qquad\qquad\qquad \text{for } k \geq 1, 0 \leq m < r; \\
 f_1^{(r,kr)} &= \frac{1}{k!} L_{\chi_2}^k \left(\sum_i \omega_i \cdot p_i \right) + \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2}^j \hat{f}_1^{(r,(k-j)r)} \\
 &\qquad\qquad\qquad \text{for } k \geq 1; \\
 f_1^{(r,kr+m)} &= \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2}^j \hat{f}_1^{(r,(k-j)r+m)} \\
 &\qquad\qquad\qquad \text{for } k \geq 1, 0 < m < r; \\
 f_l^{(r,kr+m)} &= \frac{1}{k!} L_{\chi_2}^k h_l^{(m)} + \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2}^j \hat{f}_l^{(r,(k-j)r+m)} \\
 &\qquad\qquad\qquad \text{for } l \geq 2, k \geq 1, 0 \leq m < r.
 \end{aligned}
 \tag{84}$$

Remark that for $k = 1, m = 0$ the first equation gives $f_0^{(r,r)} = \hat{f}_0^{(r,r)} = 0$. Now we also impose the condition $f_1^{(r,r)} = 0$, so that all terms which are linear in p are removed at order r . To this end we replace (47) with $k = 1, m = 0$ and $l = 1$ in the second of (84) with $k = 1$, thus getting the condition

$$\sum_i \omega_i \frac{\partial \chi_2^{(r)}}{\partial q_i} + L_{\chi_1} h_2^{(0)} + f_1^{(r-1,r)} = 0.
 \tag{85}$$

Recalling now that $h_2^{(0)} \in \mathcal{P}_{2,0}$, write it as

$$h_2^{(0)} = \frac{1}{2} \sum_{i,j} C_{ij} p_i p_j,$$

with C_{ij} defined by (51). This gives

$$L_{\chi_1} h_2^{(0)} = \{X^{(r)}, h_2^{(0)}\} + \sum_{i,j} C_{ij} p_i \xi_j^{(r)}.
 \tag{86}$$

Furthermore, separate out the average part of $f_1^{(r-1,r)}$, writing

$$f_1^{(r-1,r)} = \langle f_1^{(r-1,r)} \rangle + \tilde{f}_1^{(r-1,r)},
 \tag{87}$$

and remark that, in view of $\langle f_1^{(r-1,r)} \rangle \in \mathcal{P}_{1,0}$, we have

$$\langle f_1^{(r-1,r)} \rangle = \sum_i b_i^{(r)} p_i,
 \tag{88}$$

with $b_i^{(r)}$ given by (51). Substituting (86) and (87) in condition (85), and using (88) we readily get the equations (49) and (50), thus defining $\xi^{(r)}$ and $\chi_2^{(r)}$. We use now equation (50) in order to transform the second of (84) with $k > 1$ as follows. We separate the term $j = k - 1$ in the sum over j , thus writing

$$f_1^{(r,kr)} = \frac{1}{k!} L_{\chi_2^{(r)}}^k \left(\sum_i \omega_i \cdot p_i \right) + \frac{1}{(k-1)!} L_{\chi_2^{(r)}}^{k-1} \hat{f}_1^{(r,r)} + \sum_{j=0}^{k-2} \frac{1}{j!} L_{\chi_2^{(r)}}^j \hat{f}_1^{(r,(k-j)r)} .$$

Then we remark that the first two terms, in view of (50), give

$$\frac{1}{k!} L_{\chi_2^{(r)}}^{k-1} \left[L_{\chi_2^{(r)}} \left(\sum_i \omega_i \cdot p_i \right) + k \hat{f}_1^{(r,r)} \right] = \frac{k-1}{k!} L_{\chi_2^{(r)}}^{k-1} \hat{f}_1^{(r,r)} .$$

Thus we get

$$f_1^{(r,kr)} = \frac{k-1}{k!} L_{\chi_2^{(r)}}^{k-1} \hat{f}_1^{(r,r)} + \sum_{j=0}^{k-2} \frac{1}{j!} L_{\chi_2^{(r)}}^j \hat{f}_1^{(r,(k-j)r)} , \tag{89}$$

to be used in place of the second of (84). We conclude that the transformed Hamiltonian $H^{(r)} := \exp(L_{\chi_2^{(r)}}) \circ \exp(L_{\chi_1^{(r)}}) H^{(r-1)}$ takes the form

$$H^{(r)} = \sum_i \omega_i p_i + \sum_{s=0}^{r-1} \sum_{l \geq 2} h_l^{(s)} + \sum_{k \geq 1} \sum_{m=0}^{r-1} \sum_{l \geq 0} f_l^{(r,kr+m)} , \tag{90}$$

with $f_0^{(r,r)} = f_1^{(r,r)} = 0$, i.e., is in Kolmogorov’s normal form up to order r , as required. Writing it in the form (44) given in the statement of the lemma is just a matter of reordering all terms in a convenient fashion, as follows. First, put $h_l^{(r)} = f_l^{(r,r)}$ for $l \geq 2$, and this gives (45). Notice that this takes into account the case $k = 1, m = 0$ in (84). Then rewrite the equations (84) by excluding the case above, which is made through the condition $kr + m > r$, and by replacing the second equation by (89). Finally, identify the double sum over $k \geq 1$ and $0 \leq m < r$ in (90), excluding the case $k = 1, m = 0$, with the sum over $s > r$ in (44).

5.2 Generalized Cauchy estimates

We start with the

Proof of lemma 1. For $s = 0$ the statement is trivial. For $s > 0$ we proceed by first proving the inequalities

$$\begin{aligned} \|L_{\chi_1^{(r)}} f\|_{(1-d'-d)\sigma} &\leq l \left(\frac{\|X^{(r)}\|_\sigma}{e(d+d')\sigma} + |\xi^{(r)}| \right) \|f\|_{(1-d')\sigma} , \\ \|L_{\chi_2^{(r)}} f\|_{(1-d'-d)\sigma} &\leq \left(\frac{l}{e(d+d')\sigma} + \frac{1}{ed\sigma} \right) \|\chi_2^{(r)}\|_\sigma \|f\|_{(1-d')\sigma} , \end{aligned} \tag{91}$$

which hold for $0 \leq d' < 1$ and $0 < d < 1 - d'$. To this end, recall that $f \in \mathcal{P}_{l,N}$ can be given the form (24), and write $X^{(r)}$ in the similar form

$$X^{(r)} = \sum_{k'} x_{k'} \exp(ik' \cdot q) .$$

Then compute

$$L_{X^{(r)}} f = i \sum_{k',j,k} \sum_m \frac{k'_m j_m}{p_m} x_{k'} f_{jk} p^j \exp(i(k+k') \cdot q) ,$$

$$L_{\xi^{(r)} \cdot q} f = i \sum_{j,k} \sum_m \frac{j_m \xi_m^{(r)}}{p_m} f_{jk} p^j \exp(ik \cdot q) .$$

By the definition of the norm we get the estimate

$$\begin{aligned} \|L_{X^{(r)}} f\|_{(1-d'-d)\sigma} &\leq \sum_{k',j,k} \left| \sum_m k'_m j_m \right| |x_{k'}| |f_{jk}| e^{(1-d'-d)|k+k'|\sigma} \\ &\leq l \sum_{k'} |k'| e^{-(d+d')|k'|\sigma} |x_{k'}| e^{|k'|\sigma} \sum_{j,k} |f_{jk}| e^{(1-d')|k|\sigma} \quad (92) \\ &\leq \frac{l}{e(d+d')\sigma} \|X^{(r)}\|_\sigma \|f\|_{(1-d')\sigma} . \end{aligned}$$

Here, we used the elementary inequality

$$\left| \sum_m k'_m j_m \right| \leq \sum_m |k'_m| |j_m| \leq |k'| \sum_m |j_m| = l|k'| ,$$

and the general inequality

$$x^\alpha e^{-\delta x} \leq \left(\frac{\alpha}{e\delta}\right)^\alpha \quad \text{for positive } \alpha, x, \delta \quad (93)$$

with $1, (d+d')\sigma$ and $|k'|$ in place of α, δ and x respectively. Similarly, using the inequality

$$\left| \sum_m j_m \xi_m^{(r)} \right| \leq \sum_m |j_m| |\xi_m^{(r)}| \leq |\xi^{(r)}| \sum_m |j_m| = l|\xi^{(r)}| ,$$

we get

$$\|L_{\xi^{(r)} \cdot q} f\|_{(1-d'-d)\sigma} \leq l|\xi^{(r)}| \|f\|_{(1-d')\sigma} .$$

Recalling that $\chi_1^{(r)} = X^{(r)} + \xi^{(r)} \cdot q$, the first of (91) follows from the latter inequality together with (92). The proof of the second of (91) is similar. Write

$$\chi_2^{(r)} = \sum_{j',k'} y_{j'k'} p^{j'} \exp(ik' \cdot q) , \quad (94)$$

and compute

$$L_{\chi_2^{(r)}} f = i \sum_{j,k,j',k'} \sum_m \frac{k'_m j_m - k_m j'_m}{p_m} y_{j'k'} f_{jk} p^{j+j'} \exp(i(k+k') \cdot q) .$$

By the definition of the norm we estimate

$$\begin{aligned} \|L_{\chi_2^{(r)}} f\|_{(1-d'-d)\sigma} &\leq \sum_{j,k,j',k'} |y_{j'k'}| |f_{jk}| e^{(1-d'-d)|k+k'|\sigma} \left| \sum_m k'_m j_m - k_m j'_m \right| \\ &\leq \sum_{j,k} |f_{jk}| e^{(1-d')|k|\sigma} \sum_{j',k'} e^{-(d'+d)|k'|\sigma} |y_{j'k'}| e^{|k'|\sigma} \left| \sum_m k'_m j_m \right| \\ &\quad + \sum_{j,k} e^{-d|k|\sigma} |f_{jk}| e^{(1-d')|k|\sigma} \sum_{j',k'} |y_{j'k'}| e^{|k'|\sigma} \left| \sum_m k_m j'_m \right| , \end{aligned}$$

and the second of (91) follows from $|\sum_m k'_m j_m| \leq l|k'|$ and $|\sum_m k_m j'_m| \leq |k|$ (use $|j'| = 1$ in view of $\chi_2^{(r)} \in \mathcal{P}_{1,rK}$). From (91) one readily gets (54) and (55) just putting $f = p$, i.e., $l = 1$ and $\|f\|_\sigma = 1$. Concerning (56), just differentiate (94) and use the definition of the norm. Coming to (57) and (58), for $s = 0$ they are trivial, and for $s = 1$ are nothing but (91) with an extra factor e . For $s > 1$, let $\delta = d/s$. Note also that from $f \in \mathcal{P}_{l,N}$ one gets $L_{\chi_1^{(r)}}^j f \in \mathcal{P}_{l-j, jrK+N}$. Thus, using the first of (91) with δ in place of d , $(j-1)\delta$ in place of d' and $L_{\chi_1^{(r)}}^{j-1} f$ in place of f we get

$$\|L_{\chi_1^{(r)}}^j f\|_{(1-j\delta)\sigma} \leq (l-j+1) \left(\frac{\|X^{(r)}\|_\sigma}{ej\delta\sigma} + |\xi^{(r)}| \right) \|L_{\chi_1^{(r)}}^{j-1} f\|_{(1-(j-1)\delta)\sigma} .$$

Applying s times this inequality we get

$$\begin{aligned} \|L_{\chi_1^{(r)}}^s f\|_{(1-d)\sigma} &\leq \|f\|_\sigma \prod_{j=1}^s \left[(l-j+1) \left(\frac{\|X^{(r)}\|_\sigma}{ej\delta\sigma} + |\xi^{(r)}| \right) \right] \\ &\leq \|f\|_\sigma \prod_{j=1}^s \left[(l-j+1) \left(\frac{s\|X^{(r)}\|_\sigma}{ejd\sigma} + \frac{s|\xi^{(r)}|}{j} \right) \right] \\ &\leq \left(\frac{s}{e} \right)^s \left(\frac{\|X^{(r)}\|_\sigma}{d\sigma} + e|\xi^{(r)}| \right)^s \|f\|_\sigma \prod_{j=1}^s \frac{(l-j+1)}{j} , \end{aligned}$$

and in view of the trivial inequality $(s/e)^s < s!$ the estimate (57) readily follows. With a similar procedure, from the second of (91) we get the recursive estimate

$$\begin{aligned} \|L_{\chi_2^{(r)}}^j f\|_{(1-j\delta)\sigma} &\leq \left(\frac{l}{ej\delta\sigma} + \frac{1}{e\delta\sigma} \right) \|\chi_2^{(r)}\|_\sigma \|L_{\chi_2^{(r)}}^{j-1} f\|_{(1-(j-1)\delta)\sigma} \\ &= \frac{s(l+j)}{ejd\sigma} \|\chi_2^{(r)}\|_\sigma \|L_{\chi_2^{(r)}}^{j-1} f\|_{(1-(j-1)\delta)\sigma} . \end{aligned}$$

Iterating s times this estimate, and proceeding as above, one readily gets (58).
Q.E.D.

Lemma 4. *Let $f \in \mathcal{P}_{l,rK}$ and $\langle f \rangle = 0$. Then the equation*

$$\sum_j \omega_j \frac{\partial \chi}{\partial q_j} = f \tag{95}$$

admits a solution $\chi \in \mathcal{P}_{l,rK}$ with

$$\|\chi\|_\sigma \leq \frac{1}{\alpha_r \Omega} \|f\|_\sigma \tag{96}$$

The procedure for solving (95) is well known. Write

$$\chi = \sum_{j,k} x_{jk} p^j \exp(ik \cdot q), \quad f = \sum_{j,k} f_{jk} p^j \exp(ik \cdot q),$$

with known coefficients f_{jk} and unknowns x_{jk} . Then one gets $i(k \cdot \omega)x_{jk} = f_{jk}$. In view of (35) we get the estimate $|x_{jk}| \leq |f_{jk}|/(\alpha_r \Omega)$, and (96) follows from the definition of the norm.

Proof of lemma 2. The estimate (59) is a straightforward application of lemma 4 to eq. (48). By the property of the matrix C_{ij} of being nondegenerate there exists a solution $\xi^{(r)}$ of (49), which by (7) satisfies

$$|b^{(r)}| = \left| \sum_j C_{ij} \xi_j^{(r)} \right| \geq m |\xi^{(r)}|.$$

On the other hand, by the definition of the norm one has

$$|b^{(r)}| = \|\langle f_1^{(r-1,r)} \rangle\|_{(1-d_{r-1})\sigma} \leq \|f_1^{(r-1,r)}\|_{(1-d_{r-1})\sigma}.$$

Combining the latter two inequalities one gets (60). Proceeding as in the proof of lemma 1, eq. (92), and using (28) and (59), we get

$$\begin{aligned} \|L_{X^{(r)}} h_2^{(0)}\|_{(1-d_{r-1}-\delta_r)\sigma} &\leq \frac{2}{\delta_r e \sigma} \|X^{(r)}\|_{(1-d_{r-1})\sigma} \|h_2^{(0)}\|_\sigma \\ &\leq \frac{2}{\alpha_r \delta_r e \Omega \sigma} \|f_0^{(r-1,r)}\|_{(1-d_{r-1})\sigma} \|h_2^{(0)}\|_\sigma. \end{aligned}$$

Then, (61) is a straightforward application of lemma 4 to eq. (50). *Q.E.D.*

5.3 Proof of lemma 3

The proof is essentially a translation of the recursive algorithm of section 4.1 into an algebraic scheme of recursive estimates. The underlying idea of the proof is that the norm of every function is estimated taking into account three contributions: (i) accumulation of constants and factors coming from the estimate of

Poisson brackets (lemma 1) and of the solution of the equations for the generating functions (lemma 2); (ii) accumulation of small divisors; (iii) counting the number of Poisson brackets generated by the algorithm. The constants of the point (i) are controlled using hypotheses (64)–(66). The meaning is essentially that we get rid of all constants by just spending a factor $\varepsilon^{1/2}$. The accumulation of small divisors is controlled using the function $T(\dots)$ introduced in section 3.3. Remark that also the estimate of Poisson brackets generates small denominators due to the restrictions δ of the analyticity domain. Actually, these denominators obey the same accumulation mechanism of the small denominators due to integer combinations of the frequencies. This control is the most unpleasant part of the proof, because it requires checking several cases, and repeating the same calculations in different situations. The number of Poisson brackets is estimated via a recursive sequence $\{\nu_{r,s}\}$ introduced in the next lemma. An arithmetic argument is needed in order to bound geometrically that sequence. All proofs are made by induction. The main part of the proof is the following

Lemma 5. *Let the sequences $\{\nu_{r,s}\}_{r \geq 0, s \geq r}$ and $\{\hat{\nu}_{r,s}\}_{r \geq 1, s \geq r}$ be recursively defined as*

$$\begin{aligned} \nu_{0,s} &= 1, \\ \hat{\nu}_{r,kr+m} &= \nu_{r-1,r}^k \nu_{m,m} + \sum_{j=0}^{k-1} \nu_{r-1,r}^j \nu_{r-1,(k-j)r+m}, \\ \nu_{r,kr+m} &= \nu_{r-1,r}^k \nu_{m,m} + \sum_{j=0}^{k-1} \nu_{r-1,r}^j \hat{\nu}_{r,(k-j)r+m}, \end{aligned} \tag{97}$$

where $s \geq 0, k \geq 1, 0 \leq m < r$. Then for $r \geq 1$ one has the following estimates.

(i) *The generating functions are estimated by*

$$\left\| \frac{X^{(r)}}{\delta_r e \sigma} \right\|_{(1-d_{r-1})\sigma} + |\xi^{(r)}| \leq \frac{\varepsilon_0 E}{e} \nu_{r-1,r} \mu^r \zeta_1 T(\alpha \delta^2, r, 2r-1, 2r-2 - \log_2 r), \tag{98}$$

$$\|\chi_2^{(r)}\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \varepsilon_0 E \nu_{r-1,r} \eta_{r-1} \mu^r \zeta_2 \delta_r^2 \sigma T(\alpha \delta^2, r, 2r, 2r-2), \tag{99}$$

where

$$\zeta_1 = \frac{1}{\Omega \sigma} + \frac{2e\eta_0}{m}, \quad \zeta_2 = \frac{1}{\Omega \sigma} \left(1 + \frac{4E\eta_0}{e\Omega \sigma} \right). \tag{100}$$

(ii) *The functions $\hat{f}_l^{(r,s)}$ are estimated by*

$$\begin{aligned} \|\hat{f}_l^{(r,r)}\|_{(1-d_{r-1}-\delta_r)\sigma} &\leq \varepsilon_0^{1/2} E \hat{\nu}_{r,r} (1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^r \\ &\quad \times T(\alpha \delta^2, r, 2r - a_l, 2r - 2 + (b_l - 2) \log_2 r) \quad \text{for } l \geq 1, \end{aligned} \tag{101}$$

$$\begin{aligned} \|\hat{f}_l^{(r,s)}\|_{(1-d_{r-1}-\delta_r)\sigma} &\leq \varepsilon_0 E \hat{\nu}_{r,s} (1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^s \\ &\quad \times T(\alpha \delta^2, r, 2s - a_l, 2s - 2 + b_l \log_2 r - 2 \log_2 s) \quad \text{for } s > r, l \geq 0. \end{aligned}$$

Here, a_l and b_l are defined by (63).

(iii) The functions $h_l^{(r)}$ are estimated by

$$\|h_l^{(r)}\|_{(1-d_r)\sigma} \leq \varepsilon_0^{1/2} E\nu_{r,r}\eta_r^l \mu^r T(\alpha\delta^2, r, 2r, 2r-2) \quad \text{for } l \geq 2. \quad (103)$$

(iv) The functions $f_l^{(r,s)}$ are estimated by

$$\begin{aligned} \|f_l^{(r,s)}\|_{(1-d_r)\sigma} &\leq \varepsilon_0 E\nu_{r,s}\eta_r^l \mu^s \quad (104) \\ &\times T(\alpha\delta^2, r, 2s - a_l, 2s - 2 + b_l \log_2 r - 2 \log_2 s) \quad \text{for } s > r, l \geq 0. \end{aligned}$$

We prove all estimates by induction. This requires a direct check for $r = 1$, and an induction step from $r - 1$ to r . Throughout the proof we shall repeatedly use some elementary estimates that we collect here.

In view of definition (62) of the sequence $\{\eta_r\}$ and of the properties (36) of $\{\alpha_r\}$ and (52) of $\{\delta_r\}$ one readily proves the chain of inequalities

$$\eta_0 < \dots < \eta_{r-1} < (1 + \alpha_r \delta_r^2) \eta_{r-1} < \eta_r < \dots < 2\eta_0. \quad (105)$$

In view of hypotheses (64)–(66) and of (100) and (105) we have

$$2\varepsilon_0 E\eta_r \zeta_1 < \varepsilon_0^{1/2}, \quad \varepsilon_0 E\eta_r \zeta_2 < \varepsilon_0^{1/2}. \quad (106)$$

The binomial coefficients satisfy the following estimate

$$\binom{r+s}{s} < \frac{(1+d)^{r+s}}{d^s} \quad \text{for } r \geq 0, s \geq 0, d > 0; \quad (107)$$

we can always assume $s \leq r$, of course. Finally, for $r \geq 1$ and $k \geq 1$ the following inequalities hold true:

$$\begin{aligned} k + \log_2 r &\geq \log_2(kr + m) && \text{for } 0 \leq m \leq r, \\ j + \log_2((k-j)r + m) &\geq \log_2(kr + m) && \text{for } 0 \leq m \leq r, 0 \leq j < k. \end{aligned} \quad (108)$$

(i) *Estimates on the generating functions.* We use lemma 2, and produce estimates for the expressions in lemma 1. Considering first the case $r = 1$, write the second of (28) as

$$\|f_l^{(0,s)}\|_\sigma \leq \varepsilon_0 E\nu_{0,s}\eta_0^l \mu^s T(\alpha\delta^2, 0, 0, 0),$$

which is true in view of $\nu_{0,s} = T(\alpha\delta^2, 0, 0, 0) = 1$. Substituting this in (59), (60) and (61) we get

$$\begin{aligned} \frac{\|X^{(1)}\|_\sigma}{\delta_1 e \sigma} + |\zeta^{(1)}| &\leq \varepsilon_0 E\nu_{0,1}\mu \left(\frac{T(\alpha\delta^2, 0, 0, 0)}{\alpha_1 \delta_1 e \Omega \sigma} + \frac{\eta_0 T(\alpha\delta^2, 0, 0, 0)}{m} \right), \\ \|\chi_2^{(1)}\|_{(1-\delta_1)\sigma} &\leq \frac{\varepsilon_0 E\nu_{0,1}\eta_0 \mu}{\alpha_1 \Omega} \left(\frac{2E\eta_0 T(\alpha\delta^2, 0, 0, 0)}{\alpha_1 \delta_1 e \Omega \sigma} + T(\alpha\delta^2, 0, 0, 0) \right). \end{aligned}$$

Then (98) and (99) follow from (100) and the inequalities (see (41))

$$T(\alpha\delta^2, 0, 0, 0) < \frac{1}{\alpha_1\delta_1}T(\alpha\delta^2, 0, 0, 0) < T(\alpha\delta^2, 1, 1, 0) ,$$

$$T(\alpha\delta^2, 1, 1, 0) \leq \alpha_1\delta_1^2T(\alpha\delta^2, 1, 2, 0) .$$

For $r > 1$ we substitute (104) in (59), (60) and (61), and using (63) and (105) we get

$$\begin{aligned} & \frac{1}{\delta_r e^\sigma} \|X^{(r)}\|_{(1-d_{r-1})\sigma} + |\xi^{(r)}| \\ & \leq \varepsilon_0 E\nu_{r-1,r}\mu^r \left(\frac{T(\alpha\delta^2, r-1, 2r-2, 2r-2-2\log_2 r)}{\alpha_r \delta_r e^\sigma \Omega} \right. \\ & \qquad \qquad \qquad \left. + \frac{\eta_{r-1}T(\alpha\delta^2, r-1, 2r-1, 2r-2-\log_2 r)}{m} \right) , \\ & \|X_2^{(r)}\|_{(1-d_{r-1}-\delta_r)\sigma} \\ & \leq \frac{\varepsilon_0 E\nu_{r-1,r}\eta_{r-1}\mu^r}{\alpha_r \Omega} \left(\frac{2E\eta_{r-1}T(\alpha\delta^2, r-1, 2r-2, 2r-2-2\log_2 r)}{\alpha_r \delta_r e^\sigma \Omega} \right. \\ & \qquad \qquad \qquad \left. + T(\alpha\delta^2, r-1, 2r-1, 2r-2-\log_2 r) \right) . \end{aligned}$$

Then (98) and (99) follow from (100), (105) and from the inequalities (see (41))

$$\frac{1}{\alpha_r \delta_r} T(\alpha\delta^2, r-1, 2r-2, 2r-2-2\log_2 r) < T(\alpha\delta^2, r, 2r-1, 2r-2-\log_2 r) ,$$

$$T(\alpha\delta^2, r, 2r-1, 2r-2-\log_2 r) \leq \alpha_r \delta_r^2 T(\alpha\delta^2, r, 2r, 2r-2) .$$

This concludes the estimates for the generating functions.

(ii) *Estimates for the functions $\hat{f}_l^{(r,s)}$.* We apply lemma 1 to equations (47), using (28), (98), (103) and (104) where appropriate. With separate estimates for the cases $m = 0$ and $m > 0$, for $l \geq 0$, $k \geq 1$, $l+k > 1$ we get

$$\begin{aligned} & \frac{1}{k!} \|L_{\chi_1^{(r)}}^k h_{l+k}^{(0)}\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \varepsilon_0^{k/2} E\nu_{r-1,r}^k \nu_{0,0} (1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^{kr} \\ & \qquad \qquad \qquad \times T(\alpha\delta^2, r, 2kr - a_l, 2kr - 2 + b_l \log_2 r - 2\log_2(kr)) , \\ & \frac{1}{k!} \|L_{\chi_1^{(r)}}^k h_{l+k}^{(m)}\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \varepsilon_0 E\nu_{r-1,r}^k \nu_{m,m} (1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^{kr+m} \\ & \qquad \qquad \qquad \times T(\alpha\delta^2, r, 2(kr+m) - a_l, 2(kr+m) - 2 + b_l \log_2 r - 2\log_2(kr+m)) \end{aligned} \tag{109}$$

for $0 < m < r$.

To this end, recalling that $r = 1$ implies $m = 0$, from (57) we get

$$\begin{aligned} \frac{1}{k!} \|L_{\chi_1}^k h_{l+k}^{(0)}\|_{(1-d_{r-1}-\delta_r)\sigma} &\leq E\nu_{0,0}\eta_0^{l+k} (\varepsilon_0 E\nu_{r-1,r}\mu^r \zeta_1)^k \\ &\times \binom{l+k}{k} \left(T(\alpha\delta^2, r, 2r-1, 2r-2-\log_2 r)\right)^k \quad \text{for } r \geq 1, m = 0; \\ \frac{1}{k!} \|L_{\chi_1}^k h_{l+k}^{(m)}\|_{(1-d_{r-1}-\delta_r)\sigma} &\leq \varepsilon_0^{1/2} E\nu_{m,m}\eta_m^{l+k}\mu^m (\varepsilon_0 E\nu_{r-1,r}\mu^r \zeta_1)^k \\ &\times \binom{l+k}{k} \left(T(\alpha\delta^2, r, 2r-1, 2r-2-\log_2 r)\right)^k T(\alpha\delta^2, m, 2m, 2m-2) \\ &\quad \text{for } r \geq 2, 0 < m < r. \end{aligned} \tag{110}$$

In order to estimate the latter expressions we need the further estimates

$$\begin{aligned} &\binom{l+k}{k} \left(T(\alpha\delta^2, r, 2r-1, 2r-2-\log_2 r)\right)^k \\ &\leq 2(1+\alpha_r\delta_r^2)^{l+k} T(\alpha\delta^2, r, 2kr-a_l, 2kr-2+b_l\log_2 r-2\log_2(kr)), \\ &\binom{l+k}{k} \left(T(\alpha\delta^2, r, 2r-1, 2r-2-\log_2 r)\right)^k T(\alpha\delta^2, m, 2m, 2m-2) \\ &\leq 2(1+\alpha_r\delta_r^2)^{l+k} T(\alpha\delta^2, r, 2(kr+m)-a_l, 2(kr+m)-2+b_l\log_2 r-2\log_2(kr+m)). \end{aligned} \tag{111}$$

The first estimate is to be used for $r \geq 1, m = 0$, while the second one applies to the case $r \geq 2, 0 < m < r$. The proof requires some care, since several cases must be considered. For $l = 0$ the binomial coefficient is 1 (recall also that $l = 0$ implies $k \geq 2$); for $l = k = 1$ the binomial coefficient is 2, which justifies the factor 2. For all other values of l and k the binomial coefficient can be estimated by (107) with $\alpha_r\delta_r^2$ in place of d ; in the latter case the numerator justifies the factor $(1+\alpha_r\delta_r^2)^{l+k}$ in the estimate, while the denominator, is included in the estimate of the functions $T(\alpha\delta^2, \dots)$ according to (41); notice that we use (41) with denominator $(\alpha_r\delta_r^2)^{\min(l,k)}$. The number of divisors is estimated according to the following table:

$(2r-1)k$	for $r \geq 1, l = 0, k \geq 2, m = 0$,
$2r-1$	for $r \geq 1, l = 1, k = 1, m = 0$,
$(2r-1)k+1$	for $r \geq 1, l = 1, k \geq 2, m = 0$,
$2kr$	for $r \geq 1, l \geq 2, k \geq 1, m = 0$,
$2(kr+m)-k$	for $r > 1, l = 0, k \geq 2, 0 < m < r$,
$2(r+m)-1$	for $r > 1, l = 1, k = 1, 0 < m < r$,
$2(kr+m)+1-k$	for $r > 1, l = 1, k \geq 2, 0 < m < r$,
$2(kr+m)$	for $r > 1, l \geq 2, k \geq 1, 0 < m < r$.

The selection rule is 0 for $r = 1$, while for $r > 1$ it is evaluated according to the table

$$\begin{array}{ll}
 2kr - 2k - k \log_2 r & \text{for } l = 0, k \geq 2, m = 0, \\
 2r - 2 - \log_2 r & \text{for } l = 1, k = 1, m = 0, \\
 2kr - 2k + \log_2 r - k \log_2 r & \text{for } l = 1, k \geq 2, m = 0, \\
 2kr - 2k & \text{for } l \geq 2, k \geq 1, m = 0, \\
 2(kr + m) - 2 - 2k - k \log_2 r & \text{for } l = 0, k \geq 2, 0 < m < r, \\
 2(r + m) - 4 - \log_2 r & \text{for } l = 1, k = 1, 0 < m < r, \\
 2(kr + m) - 2 - 2k + \log_2 r - k \log_2 r & \text{for } l = 1, k \geq 2, 0 < m < r, \\
 2(kr + m) - 2 - 2k & \text{for } l \geq 2, k \geq 1, 0 < m < r.
 \end{array}$$

In view of (63) and (108) it is now an easy matter to check that the number of divisors never exceeds $2(kr + m) - a_l$ and that the selection rule never exceeds $2(kr + m) - 2 + b_l \log_2 r - 2 \log_2(kr + m)$, as stated in (111). Substituting (111) in (110), and using also (105) and (106) one readily gets (109).

We obtain now the estimate

$$\begin{aligned}
 & \left\| \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_1}^j f_{l+j}^{(r-1, (k-j)r+m)} \right\|_{(1-d_{r-1}-\delta_r)\sigma} \leq \varepsilon_0 E(1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^{kr+m} \quad (112) \\
 & \times T(\alpha \delta^2, r, 2(kr + m) - a_l, 2(kr + m) - 2 + b_l \log_2 r - 2 \log_2(kr + m)) \\
 & \times \sum_{j=0}^{k-1} \nu_{r-1, r}^j \nu_{r-1, (k-j)r+m}
 \end{aligned}$$

for $l \geq 0, k \geq 1, 0 \leq m < r, l + k + m > 1$;

this estimate applies to both the second and third equation (47), which justifies the condition $l + k + m > 1$ instead of $l + k > 1$. Recalling that $r = 1$ implies $m = 0$, we get

$$\begin{aligned}
 & \frac{1}{j!} \left\| L_{\chi_1}^j f_{l+j}^{(0, k-j)} \right\|_{(1-\delta_1)\sigma} \leq \varepsilon_0 E \nu_{0, k-j} \eta_0^{l+j} \mu^{k-j} (\varepsilon_0 E \nu_{0, 1} \mu \zeta_1)^j \\
 & \times \binom{l+j}{j} \left(T(\alpha \delta^2, 1, 1, 0) \right)^j \quad \text{for } l \geq 0, k \geq 1, l + k > 1, 0 \leq j < k; \\
 & \frac{1}{j!} \left\| L_{\chi_1}^j f_{l+j}^{(r-1, (k-j)r+m)} \right\|_{(1-d_{r-1}-\delta_r)\sigma} \\
 & \leq \varepsilon_0 E \nu_{r-1, (k-j)r+m} \eta_{r-1}^{l+j} \mu^{(k-j)r+m} (\varepsilon_0 E \nu_{r-1, r} \mu^r \zeta_1)^j \quad (113) \\
 & \times \binom{l+j}{j} \left(T(\alpha \delta^2, r, 2r - 1, 2r - 2 - \log_2 r) \right)^j \\
 & \times T(\alpha \delta^2, r - 1, 2((k - j)r + m) - a_{l+j}, \\
 & \quad 2((k - j)r + m) - 2 + b_{l+j} \log_2(r - 1) - 2 \log_2((k - j)r + m)) \\
 & \quad \text{for } r > 1, l \geq 0, k \geq 1, 0 \leq m < r, l + k + m > 1, 0 \leq j < k.
 \end{aligned}$$

In order to estimate the latter expressions we use

$$\begin{aligned}
 & \binom{l+j}{j} \left(T(\alpha\delta^2, 1, 1, 0) \right)^j \leq (1 + \alpha_1\delta_1^2)^{l+j} T(\alpha\delta^2, 1, 2k - a_l, 2k - 2 - 2\log_2 k), \\
 & \binom{l+j}{j} \left(T(\alpha\delta^2, r, 2r - 1, 2r - 2 - \log_2 r) \right)^j \\
 & \quad \times T(\alpha\delta^2, r - 1, 2((k - j)r + m) - a_{l+j}, \\
 & \quad \quad 2((k - j)r + m) - 2 + b_{l+j} \log_2(r - 1) - 2\log_2((k - j)r + m)) \\
 & \leq c_j (1 + \alpha_r \delta_r^2)^{l+j} T(\alpha\delta^2, r, 2(kr+m) - a_l, 2(kr+m) - 2 + b_l \log_2 r - 2\log_2(kr+m))
 \end{aligned} \tag{114}$$

with $c_0 = 1$ and $c_j = 2$ for $j > 0$. The first inequality must be used for $r = 1$, and the second one for $r > 1$. In order to check the estimates we proceed as for (111), bounding the binomial coefficients the same way. We evaluate the number of divisors according to the table

j	for $r = 1, l = 0, k \geq 2, m = 0,$
$j + 1$	for $r = 1, l = 1, k \geq 1, m = 0,$
$2j$	for $r = 1, l \geq 2, k \geq 1, m = 0,$
$2(kr + m) - j - a_j$	for $r > 1, l = 0, k \geq 1, j \geq 0, 0 \leq m < r$
$2(kr + m) - 1$	for $r > 1, l = 1, k \geq 1, j = 0, 0 \leq m < r,$
$2(kr + m) - 1$	for $r > 1, l = 1, k \geq 2, j = 1, 0 \leq m < r,$
$2(kr + m) + 1 - j$	for $r > 1, l = 1, k \geq 3, j \geq 2, 0 \leq m < r,$
$2(kr + m)$	for $r > 1, l \geq 2, k \geq 1, j \geq 0, 0 \leq m < r.$

On the other hand, the selection rule is 0 for $r = 1$, while for $r > 1, k \geq 1, 0 \leq m < r$ and $0 \leq j < k$ it is evaluated according to the table

$2(kr + m) - 2 - 2j - j \log_2 r + b_j \log_2(r - 1) - 2\log_2((k - j)r + m)$	for $l = 0, j \geq 0,$
$2(kr + m) - 2 + \log_2(r - 1) - 2\log_2(kr + m)$	for $l = 1, j = 0,$
$2(kr + m) - 4 - \log_2 r + 2\log_2(r - 1) - 2\log_2((k - 1)r + m)$	for $l = 1, j = 1,$
$2(kr + m) - 2 + \log_2 r - 2j - j \log_2 r + 2\log_2(r - 1) - 2\log_2((k - j)r + m)$	for $l = 1, j \geq 2,$
$2(kr + m) - 2 - 2j + 2\log_2(r - 1) - 2\log_2((k - j)r + m)$	for $l \geq 2, j \geq 0.$

In view of (63) and (108) it is now an easy matter to check that the number of divisors never exceeds $2(kr + m) - a_l$ (recall that $j < k$) and that the selection rule never exceeds $2(kr + m) - 2 + b_l \log_2 r - 2\log_2(kr + m)$, as stated in (114). Substituting (114) in (113), and using also (105) and (106) one readily gets (112). Finally, using (112) together with (109) in order to estimate (47) we get (101) and (102). This concludes the estimates for the functions $\hat{f}_l^{(r,s)}$.

(iii) *Estimates for the functions $h_l^{(r)}$ and $f_l^{(r,s)}$.* The proof is similar to the previous one for the functions $\hat{f}_l^{(r,s)}$. Some simplification is due to the extra factor δ_r in the estimate (99) for the generating function $\chi_2^{(r)}$. We apply lemma 1 to equations (46),

using (28), (101), (102) and (103) where appropriate. Separating the cases $m = 0$ and $m > 0$, for $l \geq 2, k \geq 1$ we estimate

$$\begin{aligned} \frac{1}{k!} \left\| L_{\chi_2}^k h_l^{(0)} \right\|_{(1-d_r)\sigma} &\leq \varepsilon_0^{k/2} E\nu_{r-1,r}^k \nu_{0,0} \eta_r^l \mu^{kr} \\ &\quad \times T(\alpha\delta^2, r, 2kr, 2kr - 2 + 2 \log_2 r - 2 \log_2(kr)) , \\ \frac{1}{k!} \left\| L_{\chi_2}^k h_l^{(m)} \right\|_{(1-d_r)\sigma} &\leq \varepsilon_0 E\nu_{r-1,r}^k \nu_{m,m} \eta_r^l \mu^{kr+m} \\ &\quad \times T(\alpha\delta^2, r, 2(kr+m), 2(kr+m) - 2 + 2 \log_2 r - 2 \log_2(kr+m)) \\ &\quad \text{for } 0 < m < r . \end{aligned} \tag{115}$$

The first estimate applies to (45) and to the fourth of (46). In order to prove (115), recalling that $r = 1$ implies $m = 0$ we get

$$\begin{aligned} \frac{1}{k!} \left\| L_{\chi_2}^k h_l^{(0)} \right\|_{(1-d_r)\sigma} &\leq E\nu_{0,0} \eta_0^l (\varepsilon_0 E\nu_{r-1,r} \eta_{r-1} \mu^r \zeta_2)^k \\ &\quad \times \binom{l+k}{k} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^k \quad \text{for } r \geq 1, m = 0 ; \\ \frac{1}{k!} \left\| L_{\chi_2}^k h_l^{(m)} \right\|_{(1-d_r)\sigma} &\leq \varepsilon_0^{1/2} E\nu_{m,m} \eta_m^l \mu^m (\varepsilon_0 E\nu_{r-1,r} \eta_{r-1} \mu^r \zeta_2)^k \\ &\quad \times \binom{l+k}{k} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^k T(\alpha\delta^2, m, 2m, 2m-2) \\ &\quad \text{for } r \geq 2, 0 < m < r . \end{aligned} \tag{116}$$

In order to estimate the latter expressions we use

$$\begin{aligned} &\binom{l+k}{k} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^k \\ &\leq (1 + \delta_r)^{l+k} T(\alpha\delta^2, r, 2kr, 2kr - 2 + 2 \log_2 r - 2 \log_2(kr)) , \\ &\binom{l+k}{k} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^k T(\alpha\delta^2, m, 2m, 2m-2) \\ &\leq (1 + \delta_r)^{l+k} T(\alpha\delta^2, r, 2(kr+m), 2(kr+m) - 2 + 2 \log_2 r - 2 \log_2(kr+m)) . \end{aligned} \tag{117}$$

This is seen as follows. In all cases we may use the estimate (107) for the binomial coefficient, with δ_r in place of d ; for, the denominator δ_r is compensated by the same factor in front of $T(\alpha\delta^2, \dots)$. Recalling that $l \geq 2$, the number of divisors is $2kr + m$ in all cases. The selection rule gives 0 for $r = 1$, while for $r > 1$ it is evaluated as

$$\begin{aligned} &2kr - 2k \quad \text{for } k \geq 1, m = 0 , \\ &2(kr+m) - 2 - 2k \quad \text{for } k \geq 1, 0 < m < r . \end{aligned}$$

In view of (108) it is now an easy matter to check that the selection rule never exceeds $2(kr+m) - 2 + 2 \log_2 r - 2 \log_2(kr+m)$, as stated in (117). Substituting (117) in (116), and using also (106) one readily gets (115).

We obtain now the estimate

$$\begin{aligned} & \left\| \sum_{j=0}^{k-1} \frac{1}{j!} L_{\chi_2}^j \hat{f}_l^{(r, (k-j)r+m)} \right\|_{(1-d_r)\sigma} \leq \varepsilon_0 E \eta_r^l \mu^{kr+m} \tag{118} \\ & \times T(\alpha\delta^2, r, 2(kr+m) - a_l, 2(kr+m) - 2 + b_l \log_2 r - 2 \log_2(kr+m)) \\ & \times \sum_{j=0}^{k-1} \nu_{r-1,r}^j \hat{\nu}_{r, (k-j)r+m} \end{aligned}$$

which holds for $l \geq 0, k \geq 1, 0 \leq m < r, kr+m > r$. This estimate can be used for all equations (46), including the second one. For (45), which corresponds to the case $k = 1, m = 0$, we just use (101). Remark also that a separate check for $r = 1$ is not necessary here, because all functions involved have already been estimated. Proceeding as in the previous cases we get

$$\begin{aligned} & \frac{1}{j!} \left\| L_{\chi_2}^j \hat{f}_l^{(r,r)} \right\|_{(1-d_r)\sigma} \\ & \leq \varepsilon_0^{1/2} E \hat{\nu}_{r,r} (1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^r (\varepsilon_0 E \nu_{r-1,r} \eta_{r-1} \mu^r \zeta_2)^j \\ & \quad \times \binom{l+j}{j} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^j T(\alpha\delta^2, r, 2r - a_l, 2r - 2 + (b_l - 2) \log_2 r) \\ & \hspace{15em} \text{for } l \geq 1, j > 0, \\ & \frac{1}{j!} \left\| L_{\chi_2}^j \hat{f}_l^{(r, (k-j)r+m)} \right\|_{(1-d_r)\sigma} \tag{119} \\ & \leq \varepsilon_0 E \hat{\nu}_{r, (k-j)r+m} (1 + \alpha_r \delta_r^2)^l \eta_{r-1}^l \mu^{(k-j)r+m} (\varepsilon_0 E \nu_{r-1,r} \eta_{r-1} \mu^r \zeta_2)^j \\ & \quad \times \binom{l+j}{j} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^j \\ & \quad \times T(\alpha\delta^2, r, 2((k-j)r+m) - a_l, \\ & \hspace{10em} 2((k-j)r+m) - 2 + b_l \log_2 r - 2 \log_2((k-j)r+m)) \\ & \hspace{15em} \text{for } l \geq 0, 0 \leq m < r, (k-j)r+m > r. \end{aligned}$$

In order to estimate the latter expressions we use

$$\begin{aligned} & \binom{l+j}{j} \left(\delta_r T(\alpha\delta^2, r, 2r, 2r-2) \right)^j \\ & \quad \times T(\alpha\delta^2, r, 2((k-j)r+m) - a_l, \tag{120} \\ & \hspace{10em} 2((k-j)r+m) - 2 + b_l \log_2 r - 2 \log_2((k-j)r+m)) \\ & \leq (1+\delta_r)^{l+j} T(\alpha\delta^2, r, 2(kr+m) - a_l, 2(kr+m) - 2 + b_l \log_2 r - 2 \log_2(kr+m)), \end{aligned}$$

which holds for both cases (119). Indeed, using again (107) with δ_r in place of d the number of divisors is evaluated as $2(kr+m) - a_l$; on the other hand, the selection rule is evaluated as $2(kr+m) - 2 + b_l \log_2 r - 2j - 2 \log_2((k-j)r+m)$, which in view of (108) does not exceed $2(kr+m) - 2 + b_l \log_2 r - 2 \log_2(kr+m)$. Substituting the

latter inequality in (119), and using also (106), one readily gets (118). Using (118) together with (115) in order to estimate all equations (46) we get (104). Finally, using (101) and the first of (115) we get (103). This concludes the estimates for the functions $h_l^{(r)}$ and $f_l^{(r,s)}$.

In order to complete the proof of lemma 3 we need only an estimate of the sequences (97).

Lemma 6. *The sequence $\{\nu_{r,s}\}$ is bounded by*

$$\nu_{r,s} \leq \frac{24^s}{8} \quad \text{for } r \geq 1, s \geq 1. \quad (121)$$

We start by replacing (97) by

$$\nu_{0,s} = 1, \quad \nu_{r,kr+m} = \sum_{j=0}^k (j+1) \nu_{r-1,r}^j \nu_{r-1,(k-j)r+m} \quad (122)$$

for $s \geq 0, k \geq 0, 0 \leq m < r$. This sequence is equivalent to (97) in the sense that for $s \geq r$ the values $\nu_{r,s}$ computed via (97) and via (122) actually coincide. Let us prove this claim. We extend the sequences (97) to $\{\nu_{r,s}\}_{r \geq 0, s \geq 0}$ and $\{\hat{\nu}_{r,s}\}_{r \geq 1, s \geq 0}$ (i.e., we add terms with $s < r$) via the definition

$$\begin{aligned} \nu_{0,s} &= 1 \\ \hat{\nu}_{r,kr+m} &= \sum_{j=0}^k \nu_{r-1,r}^j \nu_{r-1,(k-j)r+m} \\ \nu_{r,kr+m} &= \sum_{j=0}^k \nu_{r-1,r}^j \hat{\nu}_{r,(k-j)r+m} \end{aligned} \quad (123)$$

for $s \geq 0, k \geq 0, 0 \leq m < r$. For $s \geq r$ the values $\nu_{r,s}$ and $\hat{\nu}_{r,s}$ computed by the sequences (97) and (123) coincide. Indeed, just remark that for $k = 0$ one has $\nu_{r,m} = \hat{\nu}_{r,m} = \nu_{r-1,m} = \dots = \nu_{m,m}$, so that the term $j = k$ in the sums in (123) coincides with the term out of sum in (97). Obtaining (122) from (123) requires a few calculations:

$$\begin{aligned} \nu_{r,kr+m} &= \sum_{j=0}^k \nu_{r-1,r}^j \sum_{i=0}^{k-j} \nu_{r-1,r}^i \nu_{r-1,(k-j-i)r+m} \\ &= \sum_{j=0}^k \nu_{r-1,r}^j \sum_{l=j}^k \nu_{r-1,r}^{l-j} \nu_{r-1,(k-l)r+m} \\ &= \sum_{l=0}^k \sum_{j=0}^l \nu_{r-1,r}^l \nu_{r-1,(k-l)r+m} \end{aligned}$$

$$= \sum_{l=0}^k (l+1) \nu_{r-1,r}^l \nu_{r-1,(k-l)r+m} ;$$

here, the only hidden point is that in the second line we introduced the index $l = i + j$. The last line is nothing but (122).

Coming to the study of the sequence (122), we observe that the following properties hold true:

$$\nu_{0,s} < \dots < \nu_{s,s} = \nu_{s+1,s} = \dots \quad \text{for } s \geq 0 , \tag{124}$$

$$\nu_{r,r} = 3\nu_{r-1,r} \quad \text{for } r \geq 1 , \tag{125}$$

$$\nu_{1,s} \leq \frac{2}{3}\nu_{s-1,s-1} \quad \text{for } s > 2 , \tag{126}$$

$$\nu_{r,s} \leq \nu_{r-1,s} + \frac{2}{3}\nu_{r,r}\nu_{s-r,s-r} \quad \text{for } s > r > 1 . \tag{127}$$

The relevant property here is the first one. It means that we can estimate any element in the (infinite) matrix $\nu_{r,s}$ by the diagonal one $\nu_{s,s}$. That is, it is enough to estimate the diagonal elements. Let us prove these properties. In order to prove (124) isolate the first term in the sum (122), thus writing

$$\nu_{r,s} = \nu_{r-1,s} + \sum_{j=1}^k \dots ,$$

and remark that the sum does not exist for $k = 0$ (i.e. for $s < r$), while it is positive for $k \geq 1$; then (124) follows. Using $\nu_{r-1,0} = \nu_{0,0} = 1$, by direct check one readily obtains (125). The proof of (127) requires some extra calculation. Still writing $s = kr + m$, which implies here $k \geq 1$, compute

$$\begin{aligned} \nu_{r,kr+m} &= \nu_{r-1,kr+m} + \sum_{j=1}^k (j+1) \nu_{r-1,r}^j \nu_{r-1,(k-j)r+m} \\ &= \nu_{r-1,kr+m} + \nu_{r-1,r} \sum_{l=0}^{k-1} (l+2) \nu_{r-1,r}^l \nu_{r-1,(k-1-l)r+m} \\ &\leq \nu_{r-1,kr+m} + 2\nu_{r-1,r} \sum_{l=0}^{k-1} (l+1) \nu_{r-1,r}^l \nu_{r-1,(k-1-l)r+m} \\ &= \nu_{r-1,kr+m} + 2\nu_{r-1,r} \nu_{r,(k-1)r+m} ; \end{aligned}$$

then (127) follows in view of $(k-1)r + m = s - r$ and of (124) and (125). The proof of (126) is a small modification of the latter calculation: observing that $r = 1$

implies $m = 0$, and so $k = s$, write

$$\begin{aligned} \nu_{1,s} &= \nu_{0,s} + \nu_{0,1} \sum_{l=0}^{s-1} (l+2) \nu_{0,1}^l \nu_{0,s-1-l} \\ &\leq \nu_{0,s} - 1 + 2\nu_{0,1} \sum_{l=0}^{s-1} (l+1) \nu_{0,1}^l \nu_{0,s-1-l} \\ &= 2\nu_{1,s-1} \end{aligned}$$

(the point here is that for $l \geq 2$ one has $l+2 \leq 2(l+1) - 1$; recall also $s > 2$) then use again (124).

We come now to the estimate of the diagonal elements $\nu_{r,r}$ of (122). We claim:

$$\nu_{r,r} \leq \frac{6^r}{2} \lambda_r \quad \text{for } r \geq 1, \tag{128}$$

where $\{\lambda_r\}_{r \geq 1}$ is a sequence recursively defined as

$$\lambda_1 = 1, \quad \lambda_r = \sum_{j=1}^{r-1} \lambda_j \lambda_{r-j}. \tag{129}$$

Indeed, for $r = 1, 2$ just compute $\nu_{1,1} = 3$, $\nu_{2,2} = 18$ and $\lambda_2 = 1$, so that equality holds. For $r > 2$ start with (125), then repeatedly use (127) and (124), and at the end use (126) in order to calculate

$$\begin{aligned} \nu_{r,r} &= 3\nu_{r-1,r} \\ &\leq 3\nu_{r-2,r} + 2\nu_{r-1,r-1}\nu_{1,1} \\ &\quad \dots \\ &\leq 3\nu_{1,r} + 2(\nu_{2,2}\nu_{r-2,r-2} + \dots + \nu_{r-1,r-1}\nu_{1,1}) \\ &< 2 \sum_{j=1}^{r-1} \nu_{j,j} \nu_{r-j,r-j}. \end{aligned}$$

From this, (128) is readily proved by induction (recall that it is true for $r = 1, 2$).

The final step is the estimate of the sequence (129). We claim

$$\lambda_r = \frac{2^{r-1}(2r-3)!!}{r!} \leq 4^{r-1}, \tag{130}$$

where the standard notation $(2n+1)!! = 1 \cdot 3 \cdot \dots \cdot (2n+1)$ has been used. In order to prove this claim, let the function $g(z)$ be defined as $g(z) = \sum_{r \geq 1} \lambda_r z^r$, so that $\lambda_r = g^{(r)}(0)/r!$. Then it is immediate to check that the recursive definition (129) is equivalent to the equation $g = z + g^2$. By repeated differentiation of the latter equation one readily finds

$$g' = \frac{1}{1-2g}, \quad \dots, \quad g^{(r)} = \frac{2^{r-1}(2r-3)!!}{(1-2g)^{2r-1}}$$

(check by induction). From this, (130) follows.

Collecting now (130), (128) and (124) one gets (121).

Substituting (121) in (98), (99), (103) and (104) and using (106) one gets the estimates (67), (68), (69) and (70). This concludes the proof of lemma 3.

References

- [1] A. N. Kolmogorov, Preservation of conditionally periodic movements with small change in the Hamilton function, *Dokl. Akad. Nauk SSSR*, **98** (1954), 527.
- [2] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Gauthier-Villars, Paris 1892.
- [3] C. L. Siegel, Iterations of analytic functions, *Annals of Math.* **43** (1942), 607–612.
- [4] C. L. Siegel, Über die Normalformen analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung, *Nachr. Acad. Wiss. Göttingen, Math. Phys. Kl*, **IIA** (1952) 21–30.
- [5] J. Moser, On invariant curves of area-preserving mappings of an annulus, *Nachr. Akad. Wiss. Gött., II Math. Phys. Kl* 1962 (1962), 1–20.
- [6] J. Moser, Convergent series expansions for quasi-periodic motions, *Math. Ann.* **169** (1967), 136–176.
- [7] V. I. Arnold, Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian, *Usp. Mat. Nauk* **18** (1963), 13; *Russ. Math. Surv.* **18** (1963), 9.
- [8] V. I. Arnold, Small denominators and problems of stability of motion in classical and celestial mechanics, *Usp. Math. Nauk* **18** N.6 (1963), 91; *Russ. Math. Surv.* **18** N.6 (1963), 85.
- [9] J. Moser, Old and new applications of KAM theory, Proceedings Nato ASI School *Hamiltonian Systems with Three or More Degrees of Freedom*, S'Agaro (Spain), June 19–30, 1995 (to appear).
- [10] L. H. Eliasson, Absolutely convergent series expansion for quasi-periodic motions, report 2–88, Dept. of Math., Univ. of Stockholm (1988), now published in *MPEJ* **2**, paper 4 (1996).
- [11] L. H. Eliasson, *Hamiltonian system with normal form near an invariant torus*, in *Nonlinear dynamics*, G. Turchetti ed., World Scientific, Singapore 1989.
- [12] L. H. Eliasson, *Generalization of an estimate of small divisors by Siegel*, in P. Rabinowitz and E. Zehnder eds., book in honor of J. Moser, Academic Press, New York 1990.
- [13] G. Gallavotti, Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review, *Reviews in Math. Phys.* **6** (1994), 343–411.
- [14] G. Gallavotti, Twistless KAM tori, *Comm. Math. Phys.* **164** (1994), 145–156.
- [15] L. Chierchia and C. Falcolini, *A direct proof of a theorem by Kolmogorov in Hamiltonian systems*, preprint (1993).
- [16] M. Vittot, *Lindstedt perturbation series in Hamiltonian mechanics: explicit formulation via a multidimensional Burmann–Lagrange formula*, preprint CNRS–Luminy, Marseille (1992).
- [17] G. Gallavotti and G. Gentile, *Nonrecursive proof of the KAM theorem*, preprint (1994).
- [18] G. Gentile and V. Mastropietro, *Tree expansion and multiscale analysis for KAM tori*, preprint (1994).
- [19] A. Giorgilli and U. Locatelli, A classical self-consistent proof of Kolmogorov's theorem

on invariant tori, Proceedings Nato ASI School *Hamiltonian Systems with Three or More Degrees of Freedom*, S'Agaro (Spain), June 19–30, 1995 (to appear).

- [20] H. Rüssmann, On the one-dimensional Schrödinger equation with a quasi-periodic potential, *Ann. New York Acad. Sci.* **357** (1980), 90–107.
- [21] A. Giorgilli and E. Zehnder, Exponential stability for time dependent potentials, *ZAMP* **43** (1992), 827–855.

Antonio Giorgilli
Dipartimento di Matematica
dell' Università di Milano
Via Saldini 50
20133 Milano, Italy
Fax: +39 2 2392480

Ugo Locatelli
Observatoire de la Côte d'Azur
BP 229, 06304
NICE Cedex 4, France

(Received: April 4, 1995; revised June 7, 1996)