M P E J

Mathematical Physics Electronic Journal

ISSN 1086-6655
Volume 3, 1997

Paper 5
Received: Mar 18, 1997, Revised: Jul 28, 1997, Accepted: Oct 1, 1997
Editor: G. Gallavotti

ON CLASSICAL SERIES EXPANSIONS FOR QUASI–PERIODIC MOTIONS

ANTONIO GIORGILLI
Dipartimento di Matematica dell' Università di Milano
Via Saldini 50, 20133 Milano, Italy.

UGO LOCATELLI
Observatoire de la Côte d'Azur,
BP 4229, 06304 – NICE Cedex 4, France.

Abstract. We reconsider the problem of convergence of classical expansions in a parameter ε for quasiperiodic motions on invariant tori in nearly integrable Hamiltonian systems. Using a reformulation of the algorithm proposed by Kolmogorov, we show that if the frequencies satisfy the nonresonance condition proposed by Bruno, then one can construct a normal form such that the coefficient of ε^s is a sum of $O(C^s)$ terms each of which is bounded by $O(C^s)$. This allows us to produce a direct proof of the classical ε expansions. We also discuss some relations between our expansions and the Lindstedt's ones.

Work 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).

1. Introduction and statement of the result

We reconsider the classical problem of constructing quasiperiodic solutions for a near to integrable Hamiltonian system. The classical approach to this problem, going back to Lindstedt, Gyldén and Poincaré (see [1]), is to look for series expansions of the solution. However, the method of Lindstedt encounters difficulties both concerning the formal consistence of the construction and regarding the question of convergence. As is well known, in chapt. IX of *Méthodes Nouvelles* Poincaré solved the formal problem of consistency, but his long discussion about the convergence in chapt. XIII left the latter problem still unsolved. The existence of quasiperiodic solutions was established in the year 1954 with the work of Kolmogorov[2]. The more delicate matter of the convergence of Lindstedt's series was clarified with the recent works of Eliasson (see report [3], now published as [4], and [5][6]) and the more recent ones of Gallavotti and coworkers[7][8][9][10][11][12][13] and of Chierchia and Falcolini[14][15].

In the paper [16] we tackled the problem by introducing a constructive algorithm based on a sequence of canonical transformations which gives the Hamiltonian a suitable normal form in a neighborhood of a properly chosen unperturbed torus; the ε–expansion of the solutions are then given as a byproduct of the sequence of transformations.

In the present paper we discuss the problem of the convergence of our expansions, making use of a tree structure inspired by, but not identical to, the representation introduced in the papers quoted above. The use of a tree representation is the main technical difference with respect to our previous paper. Such a procedure allows us to achieve a great simplification of the proof, and also to improve the nonresonance condition required on the frequencies. Indeed, we replace the Diophantine condition used in all papers concerned with Lindstedt's series with the weaker one proposed by Bruno. It must be noted that a proof of KAM theorem with a condition similar to Bruno's one was given by Rüssmann[17][18].

We also try a comparison between the series produced by Lindstedt's algorithm and the ones resulting from our procedure. The use of powers of ε in our expansions makes the comparison direct, at least in principle. However, we must admit that we are still unable to give a complete explanation of the algebraic relations between the two expansions. We can only say that we obtain different expansions of one and the same solution, the relevant differences being due to different mechanisms of accumulation of small divisors.

We conclude this section with a formal statement of the result. In order to simplify the discussion we consider the case of Thirring's model, namely a system of weakly coupled rotators described by the Hamiltonian

$$H(p,q,\varepsilon) = \omega\cdot p + \frac{1}{2}J^{-1}p\cdot p + \varepsilon f(q)\ , \tag{1}$$

where $(p,q) \in \mathbf{R}^n \times \mathbf{T}^n$ are action–angle variables, J is the (diagonal) matrix of the moments of inertia, and $f(q)$ is a trigonometric polynomial of degree K in the angles q. For a discussion of the general case see [13], [15] and [16].

We consider an unperturbed torus carrying conditionally periodic motions with frequencies satisfying the nonresonance condition proposed by Bruno[19]. Precisely, in terms of the real sequence $\{\alpha_r\}_{r\geq 1}$ defined by

$$\alpha_r = \min_{0<|k|\leq rK} |k\cdot\omega| \ , \tag{2}$$

we state Bruno's condition as follows: there is an increasing sequence of positive integers $\{r_j\}_{j\geq 0}$ such that

$$-\sum_{j\geq 0} \frac{1}{r_j} \log \alpha_{r_{j+1}} < \infty \ .$$

In particular, such a condition was proved by Bruno (see [19], pp. 222–224) to be equivalent to the condition

$$-\sum_{j\geq 0} \frac{1}{2^{j-1}} \log_2 \alpha_{2^j} = \text{Б} < \infty \ , \tag{3}$$

which is the one we will use here.

Theorem: *If the matrix J satisfies $\|J^{-1}v\| \geq m\|v\|$ for some positive m and for all $v \in \mathbf{R}^n$, and ε is small enough, then in a neighborhood of a nonresonant unperturbed torus with frequencies ω satisfying Bruno's condition (3) one can construct an ε-analytic canonical transformation of the form*

$$q = q' + \varepsilon\varphi^{(1)}(q') + \varepsilon^2\varphi^{(2)}(q') + \dots \ , \quad p = p' + \varepsilon\psi^{(1)}(p',q') + \varepsilon^2\psi^{(2)}(p',q') + \dots$$

which gives the Hamiltonian the Kolmogorov's normal form

$$H(p,q,\varepsilon) = \omega\cdot p + R(p,q,\varepsilon) \ , \quad R(p,q,\varepsilon) = O(p^2) \ . \tag{4}$$

Remarks.

1. The condition $\|J^{-1}v\| \geq m\|v\|$ is not necessary if the Hamiltonian has the form (1). This "twistless property" has been pointed out, e.g., in [7]. We keep this hypothesis so that the statement and the scheme of proof of the theorem can easily be adapted to the slightly more general model discussed in Thirring's book [20], §3.6, where a linear p–dependence is allowed in the perturbation f. In sect. 3.3 we discuss when and why this hypothesis can be removed.
2. In the analytic case, $K \to \infty$, the condition of Bruno, as stated above, looks meaningless. However, in such a case one can follow the suggestion of Poincaré of splitting the Hamiltonian in trigonometric polynomials of degree $K, 2K, \dots$, with some finite K. This has been done, e.g., in [16].

The paper is organized as follows. Our formulation of Kolmogorov's algorithm is given in sect. 2.2. A discussion concerning the relations among our method and previous ones is given in sect. 3. The technical proofs have been moved to sect. 4, so that a reader not particularly interested in technicalities can easily skip them.

Acknowledgements. This paper was prompted by long and constructive discussions with G. Gallavotti, L. Galgani, G. Benettin, D. Bambusi, A. Carati, G. Gentile and V. Mastropietro. We are indebted with all of them. We are also grateful to E. Lega for useful discussions.

2. Algebraic setting and formal algorithm

We start here with a list representation of functions which can be used (e.g., with an algebraic manipulator) in order to implement both Lindstedt's algorithm and the algorithm of Kolmogorov in our scheme. The aim is to make clear in which sense we take into account every term generated by the algorithms, with no simplifications. The reader should figure out a possible representation of functions on a computer, with commonly used tools such as lists and pointers to list elements. The second section below is devoted to our algorithm. The next sections contain the tree representation of our construction, which is useful in order to control the accumulation of small divisors and to obtain the necessary estimates for convergence.

2.1 Representation of functions

We consider functions $g(p,q,\varepsilon)$ which can be represented as powers series in ε the coefficients of which are trigonometric polynomials in the angles q and polynomials of finite degree in the actions p. More precisely, we consider the following structure, illustrated in fig. 1. A function $g(p,q,\varepsilon) = g_0(p,q) + \varepsilon g_1(p,q) + \ldots$ is uniquely associated to a list of coefficients of different powers of ε. An entry in the list of powers of ε points to a list of Fourier modes. That is, we write the coefficient of ε^s as a Fourier expansion $g_s(p,q) = \sum_k c_k(p) \exp(ik\cdot q)$, where $k \in \mathbf{Z}^n$, and represent such an expansion by the list of Fourier modes. An entry in the list of Fourier modes points to a further list of monomials. That is, we write $c_k(p) = \sum_j a_{k,j} p^j$, where j is an n–array of nonnegative integers labeling the elements of the list of monomials. An entry in the list of monomials points to a further list of summands, and an entry in the list of summands points to a numerical coefficient with a list of divisors of the form $\nu\cdot\omega$, with $\nu \in \mathbf{Z}^n$. That is, we represent by the list of summands the expression $a_{k,j} = \sum \frac{\beta}{\prod \nu\cdot\omega}$. To summarize, a generic function g takes the form

$$g(p,q,\varepsilon) = \sum_{s\geq 0} \varepsilon^s g_s(p,q) \ , \quad g_s(p,q) = \sum_k \exp(ik\cdot q) \sum_j p^j \sum^{*} \frac{\beta}{\prod \nu\cdot\omega} \ . \tag{5}$$

The symbol $\sum^*$ refers to the list of summands, and the product is over the list of divisors.

Algebraic operations like sum, product, differentiation and Poisson brackets can be performed by manipulation of the lists above, either modifying some existing list entry or creating a new list structure representing the result of an operation. Sum acts on the lists of summands, being just a concatenation of two lists. Differentiation with respect to one angle acts on the list of numerical coefficients, being just a multiplication by an exponent taken from the list of Fourier modes. Differentiation with respect to one action acts on the list of monomials, moving a pointer to an entry of lower order, and on the numerical coefficients. Product between two functions acts on all lists. The Poisson bracket is just a combination of differentiation, product and sum. Similar considerations apply to the inverse of the operator $\partial_\omega := \omega\cdot\partial_q$, that acts on the numerical coefficient and adds one divisor to the list of divisors. We stress in particular that the contents of the list of summands depend on the operations performed on the list.

In all the rest, we intend that all algebraic operations are performed keeping the lists as they are generated. This point will be crucial in comparing the series of Lindstedt and the series generated by our algorithm.

In order to investigate the convergence of the series we proceed as follows: let $g_s(p,q)$ be the coefficient of some power of ε according to (5), and consider the list associated to it. Then we define

$$\|g_s\| = \sum_k \sum_j \sum^{*} \frac{|\beta|}{|\prod \nu\cdot\omega|} \ . \tag{6}$$

where the sums and the product are to be performed over all elements of the corresponding lists, as in (5). Since the norm involves all elements of the sublist describing the coefficient of a given power of ε we never need to count the elements in a given list: in order to prove the convergence of a power series $g = g_0 + \varepsilon g_1 + \ldots$ it will be sufficient to prove that $\sum_{s\geq 0} \varepsilon^s \|g_s\|$ converges for some positive ε.

2.2 Formal algorithm for Kolmogorov's normal form

We perform a sequence of canonical transformations which bring the Hamiltonian to Kolmogorov's normal form order by order in ε. The transformations are performed using the algorithm of Lie series. For an account on this method see, e.g., [21] or [22]. As it will be evident from the algorithm, the property of the Hamiltonian (1) of being at most quadratic in the action p is preserved by the transformation.

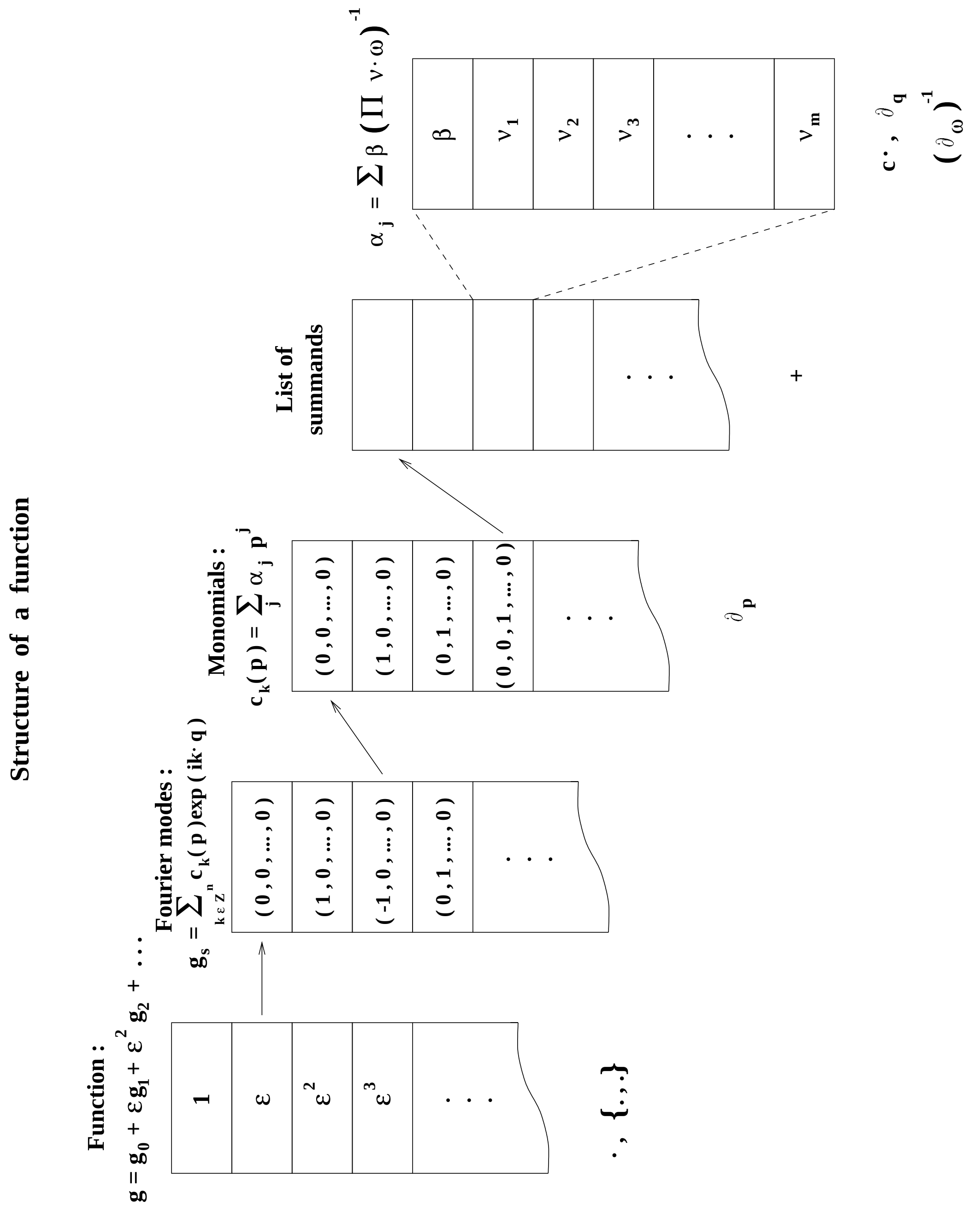


Figure 1. Illustrating a possible computer representation of the list structure associated to a function. At the first three levels an entry of the list consists of a label (e.g., the integer vector of exponents for the list of monomials) and a pointer to an unique list of the next level. The list of summands has no labels, and contains only pointers to the coefficients; it acts just as a set of indexes in a sum.

Let us write the Hamiltonian after r normalization steps in the form

$$H^{(r)} = \sum_i \omega_i p_i + \sum_{s=0}^{r} \varepsilon^s h^{(s)} + \sum_{s>r} \varepsilon^s \left(f_0^{(r,s)} + f_1^{(r,s)} + f_2^{(r,s)} \right) , \tag{7}$$

where $h^{(s)}$ is a homogeneous polynomial of degree 2 in the actions p and a trigonometric polynomial of degree sK in the angles q, and the functions $f_l^{(r,s)}$ are homogeneous polynomials of degree l in the actions and trigonometric polynomials of degree sK in the angles. We can always assume that $\langle f_0^{(r,s)} \rangle = 0$, where $\langle \cdot \rangle$ denotes averaging over all angles. The upper index r refers to the current iteration of the normalization algorithm, and is missing in $h^{(s)}$ because this is the part of the Hamiltonian already in normal form. The original Hamiltonian (1) has clearly this form: just put $r = 0$, and identify $h^{(0)} = J^{-1} p \cdot p/2$, $f_0^{(0,1)} = f$. All the rest of the expansion is initially zero.

Assuming that we have performed $r-1$ normalization steps, thus obtaining $H^{(r-1)}$ as in (7), we perform two canonical transformations with generating functions $\varepsilon^r \chi_1^{(r)}(q) = \varepsilon^r (X^{(r)}(q) + \xi^{(r)} \cdot q)$ and $\varepsilon^r \chi_2^{(r)}(p,q)$ respectively, where $X^{(r)}$ does not depend on p, $\xi^{(r)}$ is a real vector, and $\chi_2^{(r)}$ is linear in p. Performing the transformation in two separate steps allows us to easily control the degree in the actions of every term in the expansion.

With the first transformation we compute an intermediate Hamiltonian $\hat{H}^{(r)} = \exp(\varepsilon^r L_{\chi_1^{(r)}}) H^{(r-1)}$ of the form

$$\hat{H}^{(r)} = \sum_i \omega_i p_i + \sum_{s=0}^{r-1} \varepsilon^s h^{(s)} + \sum_{s \geq r} \varepsilon^s \left(\hat{f}_0^{(r,s)} + \hat{f}_1^{(r,s)} + \hat{f}_2^{(r,s)} \right) , \tag{8}$$

where $\chi_1^{(r)} = X^{(r)} + \xi^{(r)} \cdot q$ has to be determined so that $\hat{f}_0^{(r,r)} = 0$ and $\hat{f}_1^{(r,r)}$ has zero average over the angles. Then we perform the second transformation, computing $H^{(r)} = \exp(\varepsilon^r L_{\chi_2^{(r)}}) \hat{H}^{(r)}$ of the form (7), where $\chi_2^{(r)}$ has to be determined with the condition that the linear term $\hat{f}_1^{(r,r)}$ is removed.

The generating functions are determined by the equations

$$\partial_\omega X^{(r)} + f_0^{(r-1,r)} = 0 , \tag{9}$$

$$J^{-1} \xi^{(r)} \cdot p + \langle f_1^{(r-1,r)} \rangle = 0 , \tag{10}$$

$$\partial_\omega \chi_2^{(r)} + \{X^{(r)}, h^{(0)}\} + \left(f_1^{(r-1,r)} - \langle f_1^{(r-1,r)} \rangle \right) = 0 , \tag{11}$$

where $\partial_\omega \cdot = \{\cdot, \omega \cdot p\}$ and $\langle \cdot \rangle$ denotes averaging over the angles q. Moreover, the normalized term $h^{(r)}$ in the Hamiltonian is given by

$$h^{(r)} = L_{\chi_2^{(r)}} h^{(0)} + f_2^{(r-1,r)} .$$

Looking at the explicit expressions of the exponential operators $\exp(\varepsilon^r L_{\chi_j^{(r)}})$ one immediately sees that terms of the same degree in ε and in the actions p are easily isolated, so that one can obtain explicit recursive formulæ for all functions $f_l^{(r,s)}$ in the expression (7). The explicit recursive formulæ for the transformations are reported in [16].

2.3 Representation of the algorithm by trees

The following construction keeps track of the algebraic operations performed in constructing all functions generated by the algorithm of the previous section.

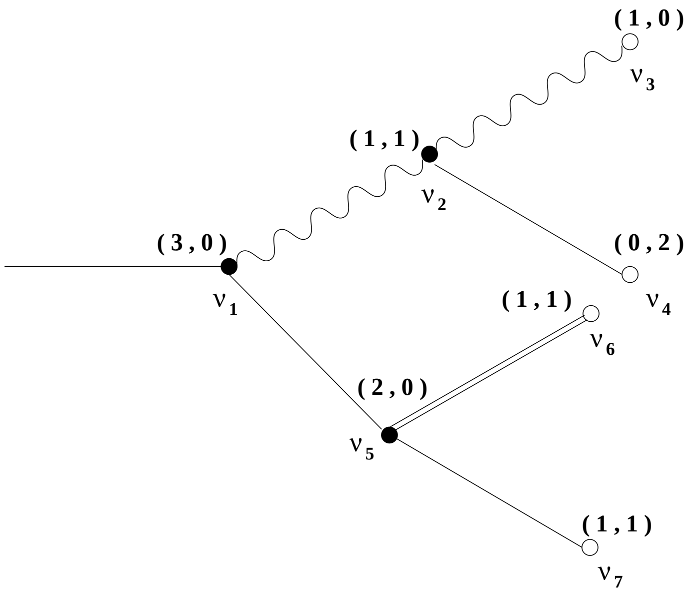
Let us represent a function as a union of rooted trees. A tree $\mathcal{A}$ is composed by a root branch uniquely associated to a node; the node in turn can either be an end node or be expanded in a set of trees with their root attached to that node; we shall represent an end node by a circle, and a node to be expanded in trees as a box; the latter node will be also called an internal node. We consider a node as associated to its branch; the notation $\#\mathcal{A}$ will denote the number of nodes in the tree $\mathcal{A}$, and so also the number of branches. The set of internal nodes and of the end nodes will be denoted by $\mathcal{A}_i$ and $\mathcal{A}_e$, respectively. A tree has a natural partial ordering: a node a precedes the node $b \neq a$ if the path from b to the root branch goes through the node a. In that case we shall write $a \prec b$. In particular, if the branch of the node b is attached to the node a we shall say that a is the left node of b and that b is a right node of a. A node has exactly one left node, unless it is attached to the root branch. An internal node has one or more right nodes.

At the lowest level we represent a function by a single node with a root branch. When determining the generating functions (i.e., when solving the equation $\partial_\omega S = f$) we represent the solution S by the same tree as f, but with the root branch replaced by a wavy line. That is, the wavy line means division of all summands by $-ik\cdot\omega$, where k is the Fourier mode corresponding to the summand. The subset of all nodes with a wavy line will be denoted by $\mathcal{B}_\sim$.

A special role is played by the solution of eq. (10), since the function $\xi\cdot q$ can not be represented with the structure of sect. 2.1. However, this expression enters only in Poisson brackets with functions, and the result is actually a function. Therefore, it acts as a temporarily created expression that can be represented by just keeping track of the vector ξ (actually n functions with zero Fourier mode and zero monomial exponents). We associate a tree to $\xi\cdot q$ by replacing the root branch by a double line in the tree corresponding to the known function in (10). That is, the double line means extracting the part of a list corresponding to the Fourier mode $k=0$, and solving eq. (10). The subset of all nodes with a double line will be denoted by $\mathcal{B}_=$.

New trees are generated via Poisson brackets. The multiple Poisson bracket $\frac{1}{s!}L_\chi^s f$ will be represented by creating a new root branch with a node, and attaching s trees representing generating functions χ and a tree representing the function f to that node, in descending order.

To every node we attach two labels, the first one representing the exponent of ε, and the second one representing the polynomial degree on the actions. In the following we call these two quantities ε–weight and polynomial weight, respectively; moreover, for the ε–weight of a node ν we use the notation $w(\nu)$. An internal node has labels determined according to the rule that the exponent of ε is the sum of the ε–weights of all its right nodes, and the polynomial exponent is the sum of the polynomial weights of its right nodes decreased by the number of wavy lines and twice the number of double lines attached to that node. For instance, in the representation



the weights of the node ν_2 are computed from those of its right nodes ν_3 and ν_4; similarly, the weights of ν_6 and ν_7 determine the weights of ν_5, and the weights of ν_2 and ν_5 determine the weights of ν_1. We emphasize that the weights of all internal nodes are completely determined by the weights of the end nodes and by the wavy lines and the double lines inside the tree. For this reason we represent the internal nodes by a black circle, meaning that there is no arbitrary information associated to it. These rules are obvious in view of the mechanism above of construction of trees via Poisson brackets.

Remark. One should not try to identify our tree structures with the trees of [3]–[15]. In these papers a tree represents one coefficient of a given Fourier mode, and so it must be identified with an element of the list of summands in the list representation of sect. 2.1 (recall that many trees can represent different coefficients of the same Fourier mode). In contrast, our trees are meant to represent a set of rules that allow one to construct a new function from existing ones, through the operations of solving the equations (9), (10) and (11) for the generating functions and calculating Lie derivatives

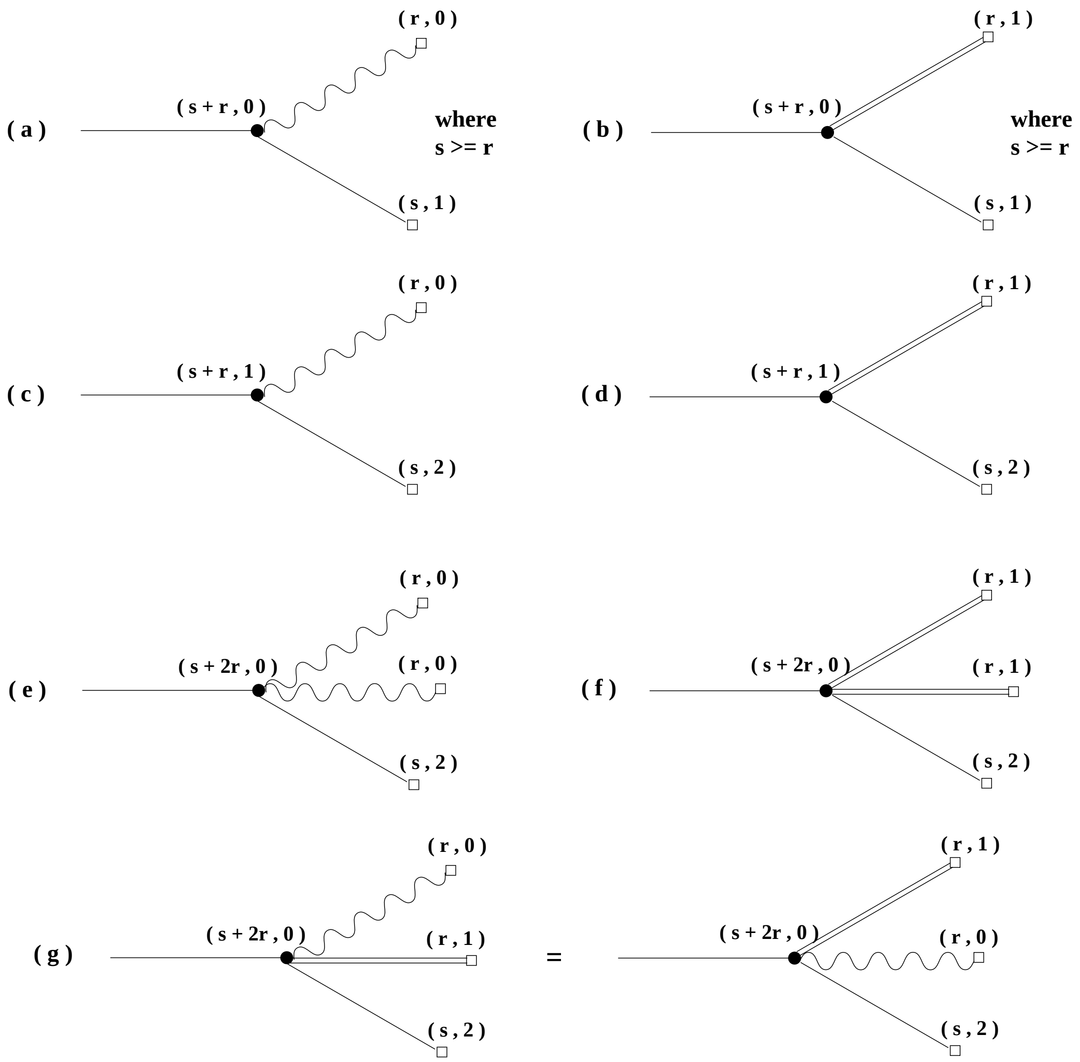
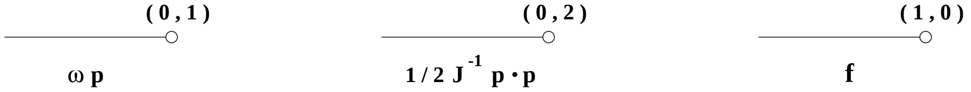


Figure 2. Tree structures generated by the transformation with the generating function $\chi_1^{(r)}$.

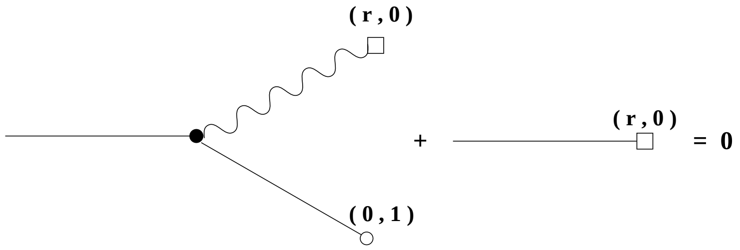
of functions. A function defined as the sum of many expressions involving the operations above is represented as a union of trees. The use of a single tree with a square box as representing a function reflects the recursive nature of our algorithm: once a function has been computed, its actual representation as a union of trees is no more useful.

We are now ready to describe the algorithm for Kolmogorov's normal form in terms of trees. The Hamiltonian $H^{(0)}$, i.e., the Hamiltonian (1), is represented by simple trees with a root and one end node, namely

(12) 

Similarly, the Hamiltonian $H^{(r)}$ is represented by two trees as above for $\omega\cdot p$ and $h^{(0)}$, and a set of trees that can be generically represented by a root branch with a box, and with labels $(s,2)$ for $h^{(s)}$ and (s,l) for $f_l^{(r,s)}$.

Assume that all trees which represent $H^{(0)},\ldots,H^{(r-1)}$ have been constructed. The algorithm of sect. 2.2 tells us how to construct the trees of $H^{(r)}$. Starting with the trees of $H^{(r-1)}$, that we represent just by a root branch with a box, we construct the generating function $X^{(r)}$ by solving (9). To this end, let us represent symbolically the equation as

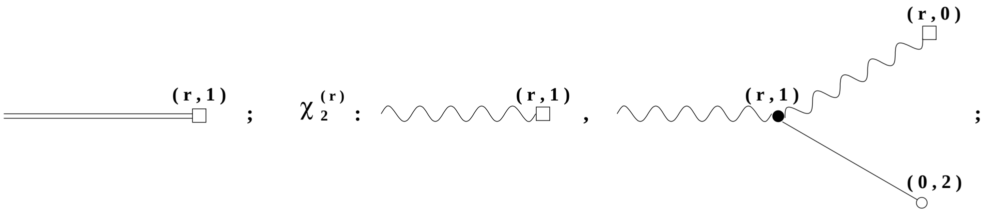
(13) $\partial_\omega \mathbf{X}^{(r)} + \mathbf{f}_0^{(r-1,r)} = 0$: 

Here, the unknown generating function $X^{(r)}$ is represented by the tree

(14) 

namely, the node with a wavy line. Changing the root branch of $f_0^{(r-1,r)}$ to a wavy line represents the operation of solving eq. (13).

We also construct $\xi^{(r)}\cdot q$ and $\chi_2^{(r)}$ by solving (10) and (11), and represent symbolically the result as

(15) $\xi^{(r)}\cdot\mathbf{q}$: 

The two different trees for $\chi_2^{(r)}$ are obtained by the tree expansion of the two terms on the right of (11).

Now we construct the intermediate Hamiltonian $\hat{H}^{(r)}$. Applying the generating function $\chi_1^{(r)}$ to every term in $H^{(r-1)}$ produces the trees represented in fig. 2. The union of all trees with the same root's weights represents the functions $\hat{f}$ in (8). The

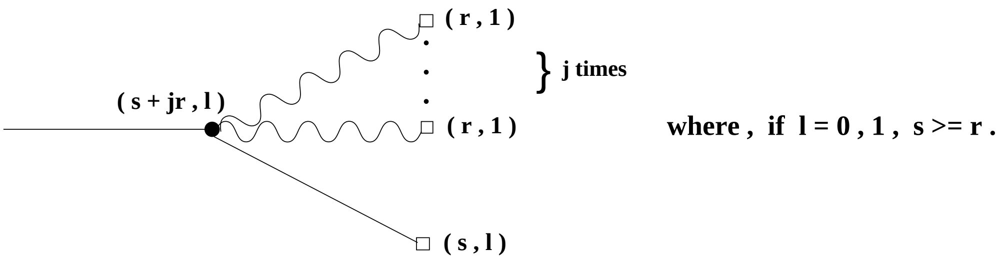


Figure 3. Tree structures generated by the transformation with the generating function $\chi_2^{(r)}$.

tree representation of $\hat{H}^{(r)}$ is obtained by making the union of the trees in fig. 2 with the trees in the representation of $H^{(r-1)}$.

In order to construct $H^{(r)}$ we apply the transformation with $\chi_2^{(r)}$ to $\hat{H}^{(r)}$, and represent again the trees of $\hat{H}^{(r)}$ by a box with a root branch. We obtain the trees represented in fig. 3. A complete representation of the trees could be done by recursively expanding the boxes in their actual representation.

2.4 Some properties of the trees

The trees generated by our algorithm have some particular properties which will be useful in order to produce convergence estimates. We collect here all these properties.

We start with some definitions. Let $\mathcal{A}$ be a tree. We denote by $\mathcal{Q}(\nu)$ the path connecting the node ν with the root, including the node ν itself. If ν is an end node, then the path $\mathcal{Q}(\nu)$ will be called a primary path. The set of all right nodes of an internal node ν will be denoted by $\mathcal{R}(\nu)$. By construction, the set $\mathcal{R}(\nu)$ of an internal node $\nu \in \mathcal{A}_i$ is the union $\mathcal{G}(\nu) \cup \nu'$, where ν' is a node with a line, namely represents a function, and $\mathcal{G}(\nu)$ is a set of nodes with a wavy line or a double line, namely represents generating functions. All nodes $\nu'' \in \mathcal{G}(\nu)$ have the same ε–weight, that we denote by $v(\nu)$, i.e., $v(\nu) = w(\nu'')$ where $\nu'' \in \mathcal{G}(\nu)$.

A subset $\mathcal{I} \subset \mathcal{A}$ is said to be a set of independent nodes in case for any $\nu, \nu' \in \mathcal{I}$ there is no primary path in $\mathcal{A}$ containing both ν and ν'.

To any non negative integer q we associate the set $\mathcal{N}^{(q)} \subset \mathcal{A}$ defined as

$$\mathcal{N}^{(q)} = \left\{ \nu \in \mathcal{B}_{\sim} \ : \ 2^{q-1} < w(\nu) \le 2^q \right\} \ ;$$

we also define $\mathcal{N}_j^{(q)}$ for a positive integer j as the set

$$\mathcal{N}_j^{(q)} = \left\{ \nu \in \mathcal{N}^{(q)} \ : \ \#\left(\mathcal{Q}(\nu) \cap \mathcal{N}^{(q)}\right) = j \right\} \ .$$

For any positive integer r we define the set

$$\mathcal{M}^{(r)} = \{ \nu \in \mathcal{A}_i \ : \ v(\nu) = r \} \ ;$$

moreover for any positive integer j we define

$$\mathcal{M}_j^{(r)} = \left\{ \nu \in \mathcal{M}^{(r)} \ :\ \#(\mathcal{Q}(\nu) \cap \mathcal{M}^{(r)}) = j \right\} \ .$$

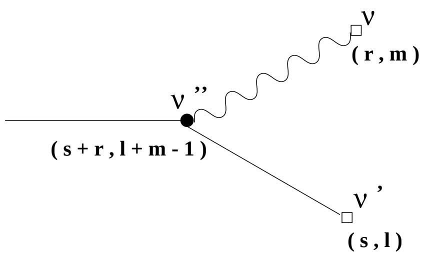
We claim that for the trees generated by our algorithm the following statements are true:

(a) *If $\mathcal{I}$ is a set of independent nodes, then* $\sum_{\nu \in \mathcal{I}} w(\nu) \le w(\mathrm{root})$.

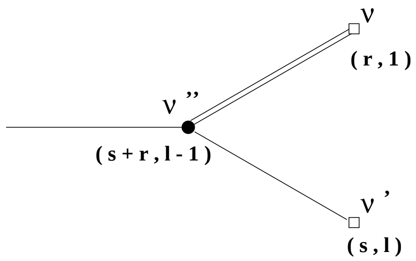
(b) *If $\nu \prec \nu'$ then $w(\nu) \ge w(\nu')$ and, if $\nu' \in \mathcal{A}_i$, then $v(\nu) \ge v(\nu')$.*

(c) *With reference to the trees*

(a)



(b)



if the polynomial weight l of ν' equals 0 or 1, then $w(\nu) \le w(\nu')$.

(d) $\mathcal{N}_j^{(q)} = \mathcal{M}_j^{(r)} = \emptyset$ *for $j > 2$.*

(e) $\mathcal{N}_1^{(q)}$, $\mathcal{N}_2^{(q)}$, $\mathcal{M}_1^{(r)}$ *and* $\mathcal{M}_2^{(r)}$ *are sets of independent nodes, and moreover one has* $\mathcal{N}_1^{(q)} \cap \mathcal{N}_2^{(q)} = \emptyset$, $\mathcal{N}_1^{(q)} \cup \mathcal{N}_2^{(q)} = \mathcal{N}^{(q)}$, $\mathcal{M}_1^{(r)} \cap \mathcal{M}_2^{(r)} = \emptyset$ *and* $\mathcal{M}_1^{(r)} \cup \mathcal{M}_2^{(r)} = \mathcal{M}^{(r)}$.

(f) *for any integer $q \ge 0$ one has*

$$\#\mathcal{N}^{(0)} \le 2w(\mathrm{root}) \ , \quad \#\mathcal{N}^{(q)} \le 2 \left\lfloor \frac{w(\mathrm{root})}{2^{q-1}+1} \right\rfloor \ .$$

(g) *The maximal number of elements of a generic tree, according to whether it represents a function or a generating function, is given by the following table, which holds for $r, s > 0$ and for $l = 0, 1, 2$:*

	$\#\mathcal{A}$	$\#\mathcal{A}_e$	$\#\mathcal{B}_\sim$	$\#\mathcal{B}_=$
$X^{(r)}$	$6r-5$	$3r-2$	$2r-1$	$r-1$
$\xi^{(r)}$	$6r-3$	$3r-1$	$2r-1$	r
$\chi_2^{(r)}$	$6r-3$	$3r-1$	$2r$	$r-1$
$h^{(s)}$	$6s-1$	$3s$	$2s$	$s-1$
$f_l^{(r,s)}$	$6s+2l-5$	$3s+l-2$	$2s-2+l$	$s-1$

(h) *The total number of trees representing any of the functions $h^{(s)}$, $\chi_1^{(s)}$, $\chi_2^{(s)}$, $f_l^{(r,s)}$ does not exceed 84^s, independent of r and l.*

We postpone the proof of all these statements to sect. 4.

The control of the accumulation of small divisors strongly depends on the property *(f)*. With reference to the list structure of sect. 2.1, we can rephrase the property as follows: *for the coefficient of order s in ε of any function, the list of divisors $k \cdot \omega$ contains at most $2s$ elements with $|k| \leq K$ and at most $2 \cdot \lfloor s/2^{q-1} \rfloor$ elements with $2^{q-1} < |k|/K \leq 2^q$ for $q \geq 1$.* In view of this elementary property the accumulation of small divisors in the coefficient of order ε^s is bounded by C^s, with some constant C, provided the frequencies ω satisfy the condition of Bruno.

3. Some relations with previous methods

The aim of this section is to discuss the relations of our algorithm with the original Kolmogorov's algorithm and with the recent works on Lindstedt series.

3.1 Relations with Kolmogorov's algorithm

Our method is in fact an algorithmic reformulation of the original method proposed by Kolmogorov. A short and elegant formulation of Kolmogorov's method can be found in Thirring's book [20], where the model considered here was proposed. Our approach differs from that of Thirring in two main points: (i) we perform the canonical transformations with the method of Lie series, as suggested in [23], and (ii) we use traditional expansions in the parameter ε instead of using the so called quadratic method suggested by Kolmogorov and almost universally used in KAM theory.

The main benefit offered by the use of Lie series is that no inversion is required in order to determine the explicit form of the canonical transformation, and moreover the construction of the transformed Hamiltonian requires only the application of the exponential operator $\exp(L_\chi)$, where $\chi = \xi \cdot q + X(q) + Y(q) \cdot p$ is the generating function. This presents no particular advantage if one is interested only in quantitative estimates leading to the proof of the theorem, as was the purpose of Thirring. From the algorithmic point of view instead the advantage is very relevant, because applying the exponential operator is substantially faster than doing inversions and substitutions of functions. For instance, substituting the expression $p = p' + \varepsilon\psi_1(p', q') + \ldots + \varepsilon^r \psi_r(p', q')$ in p^2 requires $O(r^2)$ products, while doing the same operation with the Lie series method requires only $O(r)$ differentiations and products. Remark that for the functions considered here differentiation is a trivial operation with little computational cost.

The choice of making a step by step procedure in powers of ε allows us to find series that are easily comparable with Lindstedt's ones. Indeed, the analyticity in ε of the original Hamiltonian is preserved by the whole process of transformation to Kolmogorov's normal form, but extra work is necessary in order to make this dependence explicit. Let the Hamiltonian H to be explicitly expanded in powers of ε, so that applying the first normalization step one determines the generating function (in the Lie series formalism) $\chi = \varepsilon\chi_1 + \varepsilon^2\chi_2 + \ldots$. Then one has to compute the transformed Hamiltonian H' as

$$H' = \exp\bigl(L_{\varepsilon\chi_1+\varepsilon^2\chi_2+\ldots}\bigr)H \ .$$

Writing the ε expansion of the latter expression in explicit form shows immediately that one actually has to perform at least the same number of Poisson brackets that are required by our algorithm, so that at least the same computational time is required. This situation does not change if one decides, e.g., to proceed by quadratic steps, as usual, by killing at step r all terms in the perturbation of order $\varepsilon^{2^{r-1}}, \ldots, \varepsilon^{2^r-1}$. Therefore, our choice of proceeding by powers of ε is fully justified from the algorithmic viewpoint.

The usefulness of the quadratic method has often been emphasized in connection with the problem of controlling the small divisors. As remarked, e.g., by Moser[24], proving the convergence of the ε expansions with traditional methods based on Cauchy's majorant's turns out usually to be unsuccessful. This is related to the compensations among resonant trees discussed in the recent works on Lindstedt's series. In fact, in the present paper we show that the rules for the accumulations of small divisors stated by property (f) implies that the size of the coefficient of ε^s is $O(C^s)$. This immediately leads to a convergence proof essentially equivalent to Cauchy's majorants method. Moreover, we show that our analysis introduces in a very natural way the nonresonance condition of Bruno instead of the usual diophantine one. Since this argument is not common, in the technical section 4 we include a complete proof.

3.2 Comparison with Lindstedt's series

Having determined the sequence of generating functions $\chi_1^{(1)}\,,\,\chi_2^{(1)}\,,\,\chi_1^{(2)}\,,\,\chi_2^{(2)}\,,\ldots$ the canonical transformation leading to Kolmogorov's normal form is

$$\begin{aligned} p &= \ \ldots \circ \exp\bigl(L_{\chi_2^{(2)}}\bigr) \circ \exp\bigl(L_{\chi_1^{(2)}}\bigr) \circ \exp\bigl(L_{\chi_2^{(1)}}\bigr) \circ \exp\bigl(L_{\chi_1^{(1)}}\bigr)p' =: \psi(p',q') \\ q &= \ \ldots \circ \exp\bigl(L_{\chi_2^{(2)}}\bigr) \circ \exp\bigl(L_{\chi_2^{(1)}}\bigr)q' =: \varphi(q') \ , \end{aligned} \tag{16}$$

where the generating functions are considered as functions of p', q'. Remark that the transformation with the generating functions $\chi_1^{(r)}$ leaves the angles unchanged; for this reason in the second formula only $\chi_2^{(r)}$ appears. In view of Kolmogorov's normal form, the quasiperiodic solution on the torus is $p' = 0\,,\ q' = \omega t + q_0'$, where q_0' is the initial phase. Thus, in the original variables the parametric equation of the torus and the corresponding quasiperiodic motion are

$$p = \psi(0, q') \ , \quad q = \varphi(\omega t + q_0') \ , \tag{17}$$

respectively, where ψ and φ are the functions defined in (16). The second expression must be compared with Lindstedt's solutions.

By the way, a canonical transformation leading to Kolmogorov's normal form could be constructed starting from Lindstedt's series. Indeed, let us write the invariant torus as $q = \varphi(q')$, $p = \psi(q')$, as given by Lindstedt's algorithm. Looking at the first relations as a point transformation, by the known method of the extended point transformation we can construct a generating function $S(p, q') = p \cdot \varphi(q') + w(q')$, with an arbitrary function $w(q')$. The latter can be determined by requiring that the equation of the invariant torus in the new variables be $p' = 0$. This can be done because the invariant torus is a Lagrangian manifold, so that the form $p \cdot dq$ is exact on it. Therefore, the two methods are fully equivalent.

Coming to the comparison, let us first remark that the expansions generated by the algorithms coincide after resummation of all coefficients of the same power of ε and of the same Fourier mode. This appears as evident, since both expansions are solutions of the same Hamiltonian equations. A more careful analysis reveals that in both constructions there are some arbitrary choices which could result in a change of the initial phases. However, if one assumes, as in [7], that the perturbation $f(q)$ in the original Hamiltonian (1) is even in the angles then the initial phases turn out to be the same in view of the fact that the function $\varphi(q')$ in (16) is odd in the angles.

On the other hand, looking at the algorithm in [7] one immediately realizes that the coefficient of ε^s will contain a term with denominator $(k_1 \cdot \omega)^2 (k_2 \cdot \omega)^2 \cdot \ldots \cdot (k_s \cdot \omega)^2$, with $|k_j| = jK$. Such a term can not appear in our expansions because it violates the property (f) of sect. 2.4. We conclude that the expansions must contain the same Fourier modes, but the expressions of the coefficients of a given Fourier mode do not coincide term by term.

Such a state of affairs is better illustrated by an explicit example. We implemented both our algorithm and Lindstedt's one using an algebraic manipulator (Mathematica). We performed the construction of Lindstedt's series using the algorithm given by formula (8) in [7]. §

We consider the Hamiltonian (even in the angles)

$$H(p, q) = \omega_1 p_1 + \omega_2 p_2 + \frac{1}{2}(p_1^2 + p_2^2) + \varepsilon\bigl(\cos q_1 + \cos(q_1 - q_2)\bigr) \ ,$$

§ The list representation of sect. 2.1 can be used in this case, too. Indeed, using the notations of [7], the order k of the tree corresponds to the entry ε^k in the list of powers of ε; the "momentum" $\vec{\nu} = \vec{\nu}(v_0)$ of the tree corresponds to an entry in the list of Fourier modes; in turn, this entry points to a list of summands which is the list of all trees in the sum $\sum^*$; each element of the latter list contains a numerical coefficient and a list of divisors $\vec{\nu}(v)$. No reference is made here to the monomials list, because there is no dependence on p in Lindstedt's series.

so that $K = 2$. We write the transformation of the angle variables as a power series in ε

$$q_j(q'_1, q'_2) = q'_j + \varepsilon\varphi_j^{(1)} + \varepsilon^2\varphi_j^{(2)} + \ldots \ ,$$

with $j = 1, 2$. We computed the equations of the invariant torus up to terms of order ε^5. The comparison could be made by just looking at the lists produced by the program. However, for reader's convenience, we report the results after translating the list representation in more readable analytic expressions.

At order ε both algorithm give exactly the same expansion. The first differences appear at order ε^2. Indeed, reporting for brevity only the expressions for $\varphi_1^{(2)}$, with our algorithm we find (primes are omitted)

$$\begin{aligned}\varphi_1^{(2)} = {} & \frac{\sin 2q_1}{8\omega_1^4} + \frac{\sin(2q_1 - 2q_2)}{4(\omega_1-\omega_2)^4} + \frac{\sin(2q_1 - q_2)}{2\omega_1^2(\omega_1-\omega_2)^2} - \frac{\sin(2q_1 - q_2)}{\omega_1(\omega_1-\omega_2)(2\omega_1-\omega_2)^2} \\ & - \frac{\sin q_2}{2\omega_1\omega_2(\omega_1-\omega_2)^2} - \frac{\sin q_2}{2\omega_1^2(\omega_1-\omega_2)\omega_2} \ .\end{aligned}$$

On the other hand, with Lindstedt's algorithm we obtain

$$\begin{aligned}\varphi_1^{(2)} = {} & \frac{\sin 2q_1}{8\omega_1^4} + \frac{\sin(2q_1 - 2q_2)}{4(\omega_1-\omega_2)^4} + \frac{\sin(2q_1 - q_2)}{2\omega_1^2(2\omega_1-\omega_2)^2} \\ & + \frac{\sin(2q_1 - q_2)}{2(2\omega_1-\omega_2)^2(\omega_1-\omega_2)^2} + \frac{\sin q_2}{2\omega_1^2\omega_2^2} - \frac{\sin q_2}{2\omega_2^2(\omega_1-\omega_2)^2} \ .\end{aligned}$$

The expressions look different, but using some simple algebraic relations one sees that they are actually the same. For instance, the coefficients of $\sin q_2$ are identified by applying twice the elementary algebraic identity

$$\frac{1}{ab} = \frac{1}{b(a-b)} - \frac{1}{a(a-b)} \ . \tag{18}$$

The same situation, namely that there are different expressions that are algebraically equivalent, occurs for all coefficients that we computed explicitly. However, we could not discover any simple algebraic relations that allow to immediately deduce one of the expansions from the other. So, let us just illustrate with some specific examples that Lindstedt's series contain terms that violate the property (f) of sect. 2.4. This can not occur at order ε^3, so let us consider the order ε^4. The complete expression contains more than hundred terms; therefore we report here only the expansion of the coefficient of the Fourier mode $(4, -3)$. With our algorithm we have

$$\begin{aligned}\frac{1}{2\pi^2}\left\langle \varphi_1^{(4)} \, , \sin(4q_1 - 3q_2)\right\rangle = {} & \frac{19}{96(\omega_1)^2(\omega_1-\omega_2)^6} + \frac{3}{32(\omega_1)^2(\omega_1-\omega_2)^5(2\omega_1-\omega_2)} \\ & + \frac{7}{24(\omega_1)^2(\omega_1-\omega_2)^5(3\omega_1-2\omega_2)} + \frac{9}{16(\omega_1)^2(\omega_1-\omega_2)^5(4\omega_1-3\omega_2)}\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{16(\omega_1)^2(\omega_1-\omega_2)^4(2\omega_1-\omega_2)(4\omega_1-3\omega_2)}-\frac{5}{4(\omega_1)^2(\omega_1-\omega_2)^4(3\omega_1-2\omega_2)^2}\\
&+\frac{5}{2(\omega_1)^2(\omega_1-\omega_2)^4(3\omega_1-2\omega_2)(4\omega_1-3\omega_2)}-\frac{21}{4(\omega_1)^2(\omega_1-\omega_2)^4(4\omega_1-3\omega_2)^2}\\
&-\frac{3}{32\,\omega_1(\omega_1-\omega_2)^6(2\omega_1-\omega_2)}-\frac{1}{24\,\omega_1(\omega_1-\omega_2)^6(3\omega_1-2\omega_2)}+\frac{1}{16\,\omega_1(\omega_1-\omega_2)^6(4\omega_1-3\omega_2)}\\
&-\frac{3}{8\,\omega_1(\omega_1-\omega_2)^5(2\omega_1-\omega_2)^2}+\frac{3}{2\,\omega_1(\omega_1-\omega_2)^5(2\omega_1-\omega_2)(3\omega_1-2\omega_2)}\\
&-\frac{11}{16\,\omega_1(\omega_1-\omega_2)^5(2\omega_1-\omega_2)(4\omega_1-3\omega_2)}-\frac{15}{8\,\omega_1(\omega_1-\omega_2)^5(3\omega_1-2\omega_2)^2}\\
&+\frac{15}{4\,\omega_1(\omega_1-\omega_2)^5(3\omega_1-2\omega_2)(4\omega_1-3\omega_2)}-\frac{4}{\omega_1(\omega_1-\omega_2)^5(4\omega_1-3\omega_2)^2}\\
&-\frac{3}{2\,\omega_1(\omega_1-\omega_2)^4(2\omega_1-\omega_2)(4\omega_1-3\omega_2)^2}
\end{aligned}
$$

On the other hand, with Lindstedt's algorithm we find

$$
\begin{aligned}
\frac{1}{2\pi^2}\left\langle\varphi_1^{(4)}\,,\,\sin(4q_1-3q_2)\right\rangle &= \frac{3}{8(\omega_1)^2(\omega_1-\omega_2)^4(4\omega_1-3\omega_2)^2}\\
&+\frac{1}{2(\omega_1)^2(\omega_1-\omega_2)^2(2\omega_1-\omega_2)^2(4\omega_1-3\omega_2)^2}+\frac{1}{2(\omega_1)^2(\omega_1-\omega_2)^2(3\omega_1-2\omega_2)^2(4\omega_1-3\omega_2)^2}\\
&+\frac{1}{2(\omega_1)^2(2\omega_1-\omega_2)^2(3\omega_1-2\omega_2)^2(4\omega_1-3\omega_2)^2}+\frac{1}{8(\omega_1-\omega_2)^6(4\omega_1-3\omega_2)^2}\\
&+\frac{1}{4(\omega_1-\omega_2)^4(2\omega_1-\omega_2)^2(4\omega_1-3\omega_2)^2}+\frac{1}{8(\omega_1-\omega_2)^4(3\omega_1-2\omega_2)^2(4\omega_1-3\omega_2)^2}\\
&+\frac{1}{4(\omega_1-\omega_2)^2(2\omega_1-\omega_2)^2(3\omega_1-2\omega_2)^2(4\omega_1-3\omega_2)^2}
\end{aligned}
$$

According to our rule *(f)* of sect. 2.4 every denominator can contain at most 8, 4 and 2 expressions of the form $k\cdot\omega$ with $|k|\le 2$, $3\le|k|\le 4$ and $5\le|k|\le 8$, respectively. The reader can easily check that this rule is violated by four terms in Lindstedt's expression. In recent works on Lindstedt's series these terms are dominated using the Siegel–Bruno lemma.

Resonant terms (in Eliasson's language) that violate the properties of our trees do appear only starting at order ε^5, but in this case we do not report any expansion for brevity reason (e.g., the complete expansion of $\varphi_1^{(5)}$ given by our algorithm contains more than 2500 terms). We just remark that in the expression of the coefficient of the Fourier mode $(3,-2)$ generated by the Lindstedt algorithm we found some resonant terms, for instance, one of them has the denominator:

$$
(\omega_1-\omega_2)^2(2\omega_1-\omega_2)^2(3\omega_1-2\omega_2)^4(4\omega_1-3\omega_2)^2\ .
$$

Once again, it is easy to check that such a term can not appear in our developments since it violates the rule *(f)* of sect. 2.4.

3.3 On the twistless property

We discuss here the twist condition $\|J^{-1}v\| \geq m\|v\|$ in our statement of Kolmogorov's theorem. This condition is used here in order to assure that eq. (10) for the translation of the torus can be solved. However, in the case of the Hamiltonian (1) the translation of the torus is actually zero, as shown, e.g., in [7]. We show here that such a result can be obtained also with our construction.

We recall that eq. (10) contains the average of the function $f_1^{(r-1,r)}$ with respect to the angles. We show that such an average is always of the form $J^{-1}\eta \cdot p$, with some real vector η, so that eq. (10) can be solved without any assumption on J^{-1}. Our claim follows from the simple remark that all functions which are linear in p in our construction have the form $J^{-1}p \cdot w(q)$ with some vector function $w(q)$, and that all functions which are quadratic in p, with the only exception of $h^{(0)} = J^{-1}p \cdot p$, have the form $(J^{-1}p \cdot v(q))(J^{-1}p \cdot v'(q))$ with some vector functions $v(q)$ and $v'(q)$. This is easily checked by looking at the trees of figs. 2 and 3. Indeed, terms which are linear in p are produced either by the trees (c) and (d) of fig. 2, namely by Poisson brackets of a p–independent function with a function which is quadratic in p or by the tree of fig. 3 when $l = 1$, namely by Poisson brackets between functions which are linear in p. It is an easy matter to check that such operations preserve the form above for functions which are linear in p, even when the quadratic function is $h^{(0)}$. Similarly, quadratic functions are produced by the tree of fig. 3, namely by Poisson brackets between functions linear and quadratic in p, respectively. Again, just check that the form above of the quadratic functions is preserved by the Poisson bracket.

This argument applies only when the Hamiltonian is the sum of a kinetic term quadratic in p and of a potential energy depending only on the angles.

4. Proofs

We collect here all the technicalities that are needed in order to prove the theorem.

4.1 Proof of the properties of the trees

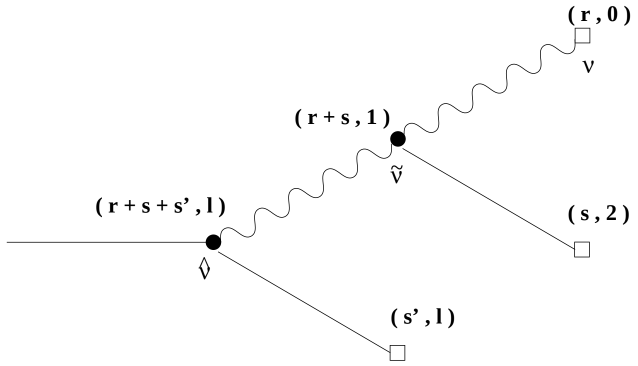
We refer to the properties *(a)–(h)* stated in sect. 2.4.

Proof of *(a)*. Recall that for every internal node ν one has $w(\nu) = \sum_{\nu' \in \mathcal{R}(\nu)} w(\nu')$, where $\mathcal{R}(\nu)$ is the set of right nodes of ν; from this one easily deduces that $w(\nu)$ is the sum of the ε–weights of all end nodes connected to ν through a primary path, i.e., all $\nu' \in \mathcal{A}_e$ such that $\nu \in \mathcal{Q}(\nu')$. Then the statement follows from the trivial remark that the sets of end nodes connected to two independent nodes are disjoint.

Proof of *(b)*. For the first property, just use the fact that $w(\nu)$ is the sum of the ε–weights of all right nodes of ν. For the second property, recall that the transformations with the generating functions are applied in increasing weight in ε.

Proof of *(c)*. This is an elementary consequence of the fact that the generating functions $\chi_1^{(r)}$ and $\chi_2^{(r)}$ are applied to the Hamiltonian $H^{(r-1)}$, which is in Kolmogorov's normal form up to order r.

Proof of *(d)*. For the second statement, recall that any internal node has right nodes representing the same generating function, either χ_1 or χ_2. Consider now any primary path $\mathcal{P}$, and remark that, by construction, two different nodes of $\mathcal{P}$ can not have the same generating function as right node. Since there are only two generating functions with the same ε–weight, our claim follows. The statement that $\mathcal{N}_j^{(r)} = \emptyset$ for $j > 2$ requires some attention. Let $\overline{\nu} \in \mathcal{N}^{(q)}$, and consider any primary path $\mathcal{P}$ through $\overline{\nu}$. Let us follow that path from its end point to the root. By *(b)*, the nodes with wavy lines have non decreasing ε–weight w. Let ν be the first node satisfying $\nu \in \mathcal{N}^{(q)}$; by construction of $\mathcal{P}$, such a node does exist. If ν is the root node, then the claim is trivially true. If not, then ν must be a node with wavy line of one of the structures of figs. 2 or 3. Let $\tilde{\nu}$ be the left node of ν, i.e., the root of any of these structures. The weights of ν are either $(r,0)$ or $(r,1)$, with $r = w(\nu)$. By *(c)*, in the cases of figs. 2.a and 3 for $l = 0, 1$ one has $s \geq r$, and so $w(\tilde{\nu}) \geq 2w(\nu) > 2^q$. Thus, by *(b)*, $\mathcal{N}^{(q)} \cap \mathcal{P} = \{\nu\}$, i.e., $\nu \in \mathcal{N}_1^{(q)}$. In the cases of figs. 2.e, 2.g and 3 with $j > 1$ it is evident that $w(\tilde{\nu}) \geq 2w(\nu)$, so that $\nu \in \mathcal{N}_1^{(q)}$, as above. In case of fig. 3 with $j = 1$ and $l = 2$ the node $\tilde{\nu}$ has polynomial weight 2, and so its branch can not be changed to a wavy line. On the other hand, the left node $\hat{\nu}$ of $\tilde{\nu}$ (if any) has $v(\hat{\nu}) \geq v(\tilde{\nu}) = w(\nu)$. Thus, $w(\hat{\nu}) \geq 2w(\nu)$, and $\nu \in \mathcal{N}_1^{(q)}$, as above. The last case is that of fig. 2.c. If $\tilde{\nu} \notin \mathcal{N}^{(q)}$, then the latter argument applies, and one concludes again $\nu \in \mathcal{N}_1^{(q)}$. If $\tilde{\nu} \in \mathcal{N}^{(q)}$, then the polynomial weight of $\tilde{\nu}$ is 1, and one has the structure



which takes into account the worst case. Thus, $\tilde{\nu}$ belongs to the structure of fig. 3, and so, by the argument above, $\tilde{\nu} \in \mathcal{N}_1^{(q)}$. We conclude $\nu \in \mathcal{N}_2^{(q)}$. This exhausts all possible cases, so that our claim follows.

Proof of *(e)*. By contradiction, let $\mathcal{P}$ be a primary path, and let $\nu, \nu' \in \mathcal{N}_j^{(q)} \cap \mathcal{P}$, with $\nu \prec \nu'$. Then, $\#(\mathcal{Q}(\nu') \cap \mathcal{N}^{(q)}) > \#(\mathcal{Q}(\nu) \cap \mathcal{N}^{(q)})$, which contradicts the assumption.

The same argument, mutatis mutandis, applies to $\mathcal{M}_j^{(q)}$. The rest of the statement is a trivial consequence of the definition of $\mathcal{N}_j^{(q)}$ and of *(d)*.

Proof of *(f)*. Let $q > 0$. By definition, $\sum_{\nu\in\mathcal{N}_j^{(q)}} w(\nu) \geq \#\mathcal{N}_j^{(q)} \cdot (2^{q-1}+1)$. On the other hand, by *(e)* and *(a)*, $\sum_{\nu\in\mathcal{N}_j^{(q)}} w(\nu) \leq w(\mathrm{root})$. The conclusion immediately follows in view of *(e)*. The case $q=0$ is now trivial.

Proof of *(g)*. The columns 2,3 and 4 are easily proved by induction, looking at (12), (13), (14), (15) and figures 2 and 3. We prove the first column. Let σ be the total number of nodes and σ_0 be the number of end nodes of a tree $\mathcal{A}$, and let $\eta(\nu) = \#\mathcal{R}(\nu)$, namely the number of right nodes of any node ν of the tree; then one has $\sum_{\nu\in\mathcal{A}} \eta(\nu) = \sigma - 1$. By construction, every node in our trees either is an end node, or has at least two right nodes, so that we have $\sum_{\nu\in\mathcal{A}} \eta(\nu) \geq 2(\sigma-\sigma_0)$. Thus, we have $\sigma - 1 \geq 2(\sigma-\sigma_0)$, i.e., $\sigma \leq 2\sigma_0 - 1$. The values of column 1 follow, using σ_0 as given by the second column.

Proof of *(h)*. Denote by $\mathcal{F}(g)$ the set of all trees in the representation of a generic function g. By the construction of sect. 2.2 it is evident that

$$
\begin{aligned}
\#\mathcal{F}(X^{(r)}) &= \#\mathcal{F}(f_0^{(r-1,r)}) \\
\#\mathcal{F}(\xi^{(r)}\cdot q) &\leq \#\mathcal{F}(f_1^{(r-1,r)}) \\
\#\mathcal{F}(\chi_2^{(r)}) &\leq \#\mathcal{F}(X^{(r)}) + \#\mathcal{F}(f_1^{(r-1,r)}) \\
\#\mathcal{F}(h^{(r)}) &= \#\mathcal{F}(f_2^{(r-1,r)}) + \#\mathcal{F}(\chi_2^{(r)}) \ .
\end{aligned}
$$

Thus, it is enough to find an upper bound on $\max_{l=0,1,2} \#\mathcal{F}(f_l^{(r-1,r)})$. Looking at figs. 2 and 3 it is easily seen that $\max_l \#\mathcal{F}(f_l^{(r,s)}) \leq \mu_{r,s}$ and $\max_l \#\mathcal{F}(\hat{f}_l^{(r,s)}) \leq \hat{\mu}_{r,s}$, where the sequences $\{\mu_{r,s}\}_{r\geq 0\,,\,s\geq 0}$ and $\{\hat{\mu}_{r,s}\}_{r\geq 1\,,\,s\geq 0}$ are recursively defined as

$$
\begin{aligned}
\mu_{0,s} &= 1 \\
\hat{\mu}_{r,s} &= \sum_{j=0}^{\lfloor s/r\rfloor} (2\mu_{r-1,r})^j \mu_{r-1,s-jr} \ , \\
\mu_{r,s} &= \sum_{j=0}^{\lfloor s/r\rfloor} (2\mu_{r-1,r})^j \hat{\mu}_{r,s-jr} \ .
\end{aligned}
\tag{19}
$$

In view of the remark above we get $\#\mathcal{F}(X^{(r)}) \leq \mu_{r-1,r}$, $\#\mathcal{F}(\xi^{(r)}\cdot q) \leq \mu_{r-1,r}$, $\#\mathcal{F}(\chi_2^{(r)}) \leq 2\mu_{r-1,r}$ and $\#\mathcal{F}(h^{(r)}) \leq \mu_{r,r}$. Proceeding as in the proof of lemma 6 in [16] we find $\mu_{r,s} \leq 84^s$.

4.2 Quantitative estimates

We now come to the investigation of the convergence of our algorithm. To this end, we make use of the norm (6) introduced in sect. 2.1.

Our goal is to estimate the norm of every function, in particular the generating functions, produced by the algorithm. Referring to the Hamiltonian (1), we denote $E = \max(\|h^{(0)}\|\,, \|f\|\,, 1)$, with $h^{(0)} = J^{-1}p\cdot p/2$; moreover we recall $\|J^{-1}v\| \geq m\|v\|$, and that K is the trigonometric degree of f. We claim that:

(i) The norm of the function represented by any tree does not exceed

$$\left(2^{2Б+16}e^{\pi^2/3}\frac{72K^2}{m}E^3\right)^{w(\mathrm{root})} . \tag{20}$$

(ii) The norms of the generating functions are bounded by

$$\left\|X^{(r)}\right\|, |\xi^{(r)}|, \left\|\chi_2^{(r)}\right\| \leq \left(2^{2Б+21}e^{\pi^2/3}\frac{189K^2}{m}E^3\right)^r . \tag{21}$$

In view of the latter estimate, it is standard matter to prove that the sequence of canonical transformations with generating functions $\chi_1^{(r)}$ and $\chi_2^{(r)}$ converges to an analytic canonical mapping provided ε is small enough, for instance if

$$2^{2Б+21}e^{\pi^2/3}\frac{189K^2}{m}E^3\varepsilon < \frac{1}{2} . \tag{22}$$

This also implies the convergence of the sequence $H^{(r)}$ of Hamiltonians of the form (7) to an analytic Hamiltonian, $H^{(\infty)}$ say, which by construction is in Kolmogorov's normal form, as stated in the main theorem. Thus, in order to prove the theorem it is sufficient to verify estimate (ii).

The estimate (ii) trivially follows from (i) and from the estimate *(h)* of sect. 2.4. The rest of this section is devoted to the proof of (i).

The norm of a tree is estimated as follows: every end node contributes with its norm; every internal node contributes with a factor (depending on the structure of its right nodes) due to the Poisson bracket(s) represented by it; every wavy line contributes with a small divisor depending on its ε–weight; every double line contributes with a factor m^{-1}.

The contribution of Poisson brackets is estimated as follows: let f be a homogeneous polynomial of degree l in p and a trigonometric polynomial of degree sK in q; then we have

$$\begin{aligned}
\left\|\frac{1}{j!}L^j_{X^{(r)}}f\right\| &\leq \left(2lKr^2\right)^j e^{(jr+s)/r^2}\left\|X^{(r)}\right\|^j \|f\| \ ,\\
\left\|L^j_{\xi^{(r)}\cdot q}f\right\| &\leq l^j\left|\xi^{(r)}\right|^j \|f\| \ ,\\
\left\|\frac{1}{j!}L^j_{\chi_2^{(r)}}f\right\| &\leq \left(2(l+1)Kr^2\right)^j e^{(jr+s)/r^2}\left\|\chi_2^{(r)}\right\|^j \|f\| \ .
\end{aligned} \tag{23}$$

These estimates are a direct consequence of the definitions of Poisson bracket and of our norm, recalling also that $X^{(r)}$ and $\chi_2^{(r)}$ are trigonometric polynomials of degree rK in q. Only the factor $2^j r^{2j} e^{(jr+s)/r^2}$ requires some justification. Considering for instance the estimate for $\chi_2^{(r)}$, first obtain

$$\left\| \frac{1}{j!} L^j_{\chi_2^{(r)}} f \right\| \le (l+1)^j \frac{(s+r)\cdot\ldots\cdot(s+jr)}{j!} K^j \bigl\|\chi_2^{(r)}\bigr\|^j \, \|f\| \ ;$$

then use the elementary inequality

$$\frac{(s+r)\cdot\ldots\cdot(s+jr)}{j!} \le 2^j r^{2j} \left(1+\frac{1}{r}\right)^{\lfloor s/r\rfloor + j} \le 2^j r^{2j} e^{(jr+s)/r^2} \ .$$

Recalling that any function of order ε^r is a trigonometric polynomial of degree rK, and in view of (2), (9) and (11), the contribution of a wavy line is at most $1/\alpha_r$, where r is the ε–weight of its node. The contribution m^{-1} due to a double line is a trivial consequence of the solution of eq. (10).

We come now to the proof of (20). Denote by $b_\sim = \#\mathcal{B}_\sim$ and $b_= = \#\mathcal{B}_=$. The norm of the tree $\mathcal{A}$ is bounded by collecting the contributions of every node, according to (23); this gives

$$\left(\frac{2}{m}\right)^{b_=} (6K)^{b_\sim} \prod_{\nu\in\mathcal{B}_\sim} \frac{w(\nu)^2}{\alpha_{w(\nu)}} \prod_{\nu\in\mathcal{A}_i} e^{w(\nu)/v(\nu)^2} \prod_{\nu\in\mathcal{A}_e} \|g\| \ . \tag{24}$$

The factor $(2/m)^{b_=}$ is the contribution of double lines and of the factor $l\le 2$ in the estimate (23) of the Poisson bracket; the factor $(6K)^{b_\sim}$ takes into account the factor independent from r in the estimate (23) of the Poisson bracket. In view of property *(g)* of sect. 2.4 we estimate $b_= \le w(\mathrm{root})$ and $b_\sim \le 2w(\mathrm{root})$.

The contribution of small divisors and of the factor r^2 in the estimate of the Poisson bracket are taken into account by $\prod_{\nu\in\mathcal{B}_\sim} w(\nu)^2/\alpha_{w(\nu)}$. Recalling that for an internal node ν one has $\mathcal{R}(\nu) = \mathcal{G}(\nu)\cup\nu'$ (see sect. 2.4), and that all nodes in $\mathcal{G}(\nu)$ have the same ε–weight $v(\nu)$, the factor $\prod_{\nu\in\mathcal{A}_i} e^{w(\nu)/v(\nu)^2}$ comes from the contributions of all internal nodes. Finally, recalling that every end node ν represents a function $h^{(0)}$ or $f_l^{(0,s)}$ coming from the original Hamiltonian, the factor $\prod_{\nu\in\mathcal{A}_e}\|g\|$ takes into account all these contributions, where g represents any of the functions above, as appropriate; this is estimated using the definition of E and that the number of end nodes does not exceed $3w(\mathrm{root})$.

In order to conclude the proof of (20) it is enough to prove that

$$\prod_{\nu\in\mathcal{B}_\sim} \frac{w(\nu)^2}{\alpha_{w(\nu)}} \le 2^{(2\text{Б}+16)w(\mathrm{root})} \ , \tag{25}$$

$$\prod_{\nu\in\mathcal{A}_i} e^{w(\nu)/v(\nu)^2} \le e^{\pi^2 w(\mathrm{root})/3} \ . \tag{26}$$

Using the definition of $\mathcal{N}^{(q)}$ and $\mathcal{N}_j^{(q)}$ in sect. 2.4 we immediately get

$$\prod_{\nu\in\mathcal{B}_\sim} \frac{w(\nu)^2}{\alpha_{w(\nu)}} \le \prod_{q=0}^{\lceil \log_2 w(\mathrm{root})\rceil} \prod_{\nu\in\mathcal{N}^{(q)}} \frac{2^{2q}}{\alpha_{2^q}} \ .$$

on the other hand, by property *(e)* of sect. 2.4 we have $\mathcal{N}^{(q)} = \mathcal{N}_1^{(q)} \cup \mathcal{N}_2^{(q)}$, and by property *(f)* we have $\#\mathcal{N}^{(q)} < 2w(\mathrm{root})/2^{q-1}$. Thus, we get

$$\log_2 \prod_{\nu\in\mathcal{B}_\sim} \frac{w(\nu)^2}{\alpha_{w(\nu)}} \le -\sum_{q\ge 0} \frac{2w(\mathrm{root})}{2^{q-1}} \log_2 \alpha_{2^q} + \sum_{q\ge 0} \frac{4qw(\mathrm{root})}{2^{q-1}} \ ,$$

and (25) follows from Bruno's condition (3) and from the elementary equality $\sum_{q\ge 0} q/2^{q-1} = 4$.

Concerning (26), recalling the definition of $\mathcal{M}^{(r)}$ in sect. 2.4 we have

$$\prod_{\nu\in\mathcal{A}_i} e^{w(\nu)/v(\nu)^2} \le \prod_{1\le r\le w(\mathrm{root})} \prod_{\nu\in\mathcal{M}^{(r)}} e^{w(\nu)/r^2} \ .$$

Using the properties *(e)* and *(a)* of sect. 2.4 we bound the latter quantity by

$$\prod_{r\ge 1} \exp\left[\frac{1}{r^2} \sum_{\nu\in\mathcal{M}^{(r)}} w(\nu)\right] \le \exp\left[\sum_{r\ge 1} \frac{2w(\mathrm{root})}{r^2}\right] = \exp\left(\frac{\pi^2 w(\mathrm{root})}{3}\right) \ .$$

This proves (26).

References

[1] Poincaré, H.: *Les méthodes nouvelles de la mécanique céleste*, Gauthier–Villars, Paris (1892).

[2] Kolmogorov, A. N.: *Preservation of conditionally periodic movements with small change in the Hamilton function*, Dokl. Akad. Nauk SSSR, **98**, 527 (1954).

[3] Eliasson, L. H.: *Absolutely convergent series expansion for quasi–periodic motions*, report 2–88, Dept. of Math., Univ. of Stockolm (1988).

[4] Eliasson, L. H.: *Absolutely convergent series expansion for quasi–periodic motions*, MPEJ **2**, 1–33 (1996).

[5] Eliasson, L. H.: *Hamiltonian system with normal form near an invariant torus*, in *Nonlinear dynamics*, G. Turchetti ed., World scientific (1989).

[6] Eliasson, L. H.: *Generalization of an estimate of small divisors by Siegel*, in *Analysis et cætera*, P. Rabinowitz and E. Zehnder eds., Academic Press (1990).

[7] Gallavotti, G.: *Twistless KAM tori*, Comm. Math. Phys. **164**, 145–156 (1994).

[8] Gallavotti, G.: *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**, 343–411 (1994).

[9] Gallavotti, G. and Gentile, G.: *Majorant series convergence for twistless KAM tori*, Ergodic Theory and Dynamical Systems **15**, 857–869 (1995).

[10] Gentile, G. and Mastropietro, V.: *Tree expansion and multiscale analysis for KAM tori*, Nonlinearity **8**, 1159–1178 (1995).

[11] Gallavotti, G.: *Lindstedt series and Kolmogorov theorem*, talk given at the "Congress on dynamical systems", Montevideo, Uruguay, (1995).

[12] Gentile, G. and Mastropietro, V.: *KAM theorem revisited*, Physica D **90**, 225–234 (1996).

[13] Gentile, G. and Mastropietro, V.: *Methods of analysis of the Lindstedt series for KAM tori and renormalizability in classical mechanics. A review with some applications*, Reviews in Math. Phys **8**, 393–444 (1996).

[14] Chierchia, L. and Falcolini, C.: *A direct proof of a theorem by Kolmogorov in Hamiltonian systems*, Ann. Sc. Norm. Sup. Pisa, Serie IV. **XXI** 541-593 (1994).

[15] Chierchia, L. and Falcolini, C.: *Compensations in small divisor problems*, Comm. Math. Phys. **175**, 135-160, (1996).

[16] Giorgilli, A. and Locatelli, U.: *Kolmogorov theorem and classical perturbation theory*, ZAMP **48**, 220–261 (1997).

[17] Rüssmann, H.: *On the one–dimensional Schrödinger equation with a quasi–periodic potential*, Ann. New York Acad. Sci. **357**, 90–107 (1980).

[18] Rüssmann, H.: *On the frequencies of quasi periodic solutions of analytic nearly integrable Hamiltonian systems*, in *Seminar on Dynamical Systems*, S. Kuksin, V. Lazutkin and J. Pöschel eds., PNDLE 12, 160–183, Birkaüser Verlag (1994).

[19] Bruno, A.D.: *Analytic form of differential equations*, Trans. Moscow Math. Soc. **25**, 131–288 (1971).

[20] Thirring, W.: *Classical dynamical systems*, Springer Verlag, New York (1978).

[21] Gröbner, W.: *Die Lie–Reihen und Ihre Anwendungen*, Springer Verlag, Berlin (1960); it. transl.: *Le serie di Lie e le loro applicazioni*, Cremonese, Roma (1973).

[22] A. Giorgilli: *Quantitative methods in classical perturbation theory*, in*From Newton to chaos*, A.E. Roy and B.D. Steves eds., Plenum Press, New York (1995).

[23] Benettin, G., Galgani, L., Giorgilli, A. and Strelcyn, J. M.: *A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the Lie method.* Il Nuovo Cimento, **79**, 201 (1984).

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