A CLASSICAL SELF–CONTAINED PROOF OF KOLMOGOROV'S THEOREM ON INVARIANT TORI

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

UGO LOCATELLI
Observatoire de la Côte d'Azur
BP 229, 06304 – NICE Cedex 4, France.

Abstract. The celebrated theorem of Kolmogorov on persistence of invariant tori of a nearly integrable Hamiltonian system is revisited in the light of classical perturbation algorithm. It is shown that the original Kolmogorov's algorithm can be given the form of a constructive scheme based on expansion in a parameter. A careful analysis of the accumulation of the small divisors shows that it can be controlled geometrically. As a consequence, the proof of convergence is based essentially on Cauchy's majorant's method, with no use of the so called quadratic method. A short comparison with Lindstedt's series is included.

1. Overview

The present lectures are concerned with the celebrated theorem of Kolmogorov on persistence of conditionally periodic motions in nearly integrable Hamiltonian systems. The aim is to extract from the original Kolmogorov's scheme of proof[1] an explicit constructive algorithm based on classical expansions in a parameter. In particular, no use is made here of the so called quadratic scheme first introduced in this framework by Kolmogorov.

This problem has already been discussed during this conference in the lectures of Moser and of Gallavotti. In particular, they discussed the problem of convergence of the classical Lindstedt's series for quasi periodic solutions lying on invariant tori (see also [2]). Direct proofs of the convergence of these series have been produced on the basis of a tree representation of the coefficients and of a suitable grouping of terms (see the works of Eliasson[3][4][5]), Gallavotti[6][7], Chierchia and Falcolini[8], Gallavotti and Gentile[9], Gentile and Mastropietro[10]).

We shall follow here a different procedure, based on the original Kolmogorov's idea of transforming the Hamiltonian in a suitable form –Kolmogorov's normal form– in a neighbourhood of a diophantine unperturbed torus. This is accomplished through a sequence of canonical transformations which is proven to be convergent in a neighbourhood of the wanted torus (see below for more details). The original part of this lecture is the analysis of accumulation of small divisors, which can be estimated geometrically. In order to better illustrate this process we shall discuss in some detail the very simple case of a perturbed system of harmonic oscillators admitting a linear Birkhoff normal form. Although not particularly interesting, this model furnishes the simplest possible example, up to our knowledge, which illustrates the thesis of this paper.

The lectures are organized as follows. Sect. 2 contains some preliminary information concerning the model problem, the characterization of Kolmogorov's normal form, and the method of Lie series for generating canonical transformations. In sect. 3 we shall discuss the case of linear Birkhoff's normal form for a perturbed system of harmonic oscillators. In this section we shall explain in detail the heuristic analysis of the accumulation of small divisors. The case of Kolmogorov's normal form will be discussed in sect. 4. Many details will be omitted, referring to [11]. Finally, a short comparison with the method of Lindstedt series will be given.

2. Preliminaries

For simplicity we shall consider the Thirring model, which has been considered also in Gallavotti's lecture. Precisely, we shall assume that the Hamiltonian has the form

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

where $\omega \in \mathbf{R}^n$ is the vector of frequencies, $p \in \mathbf{R}^n$ are action variables and $q \in \mathbf{T}^n$ are angle variables. The frequencies ω will be assumed to satisfy a diophantine condition

(2) $$|k\cdot\omega| > \gamma|k|^{-\tau} \quad \text{for } 0 \neq k \in \mathbf{Z}^n\ ,$$

where $|k| = |k_1| + \ldots + |k_n|$. Moreover, $f(p,q)$ will be assumed to be a trigonometric polynomial of degree $K > 0$ for some K.

2.1 Kolmogorov's normal form

Following Kolmogorov, we look for a canonical transformation which gives the Hamiltonian (1) the particular form (Kolmorogorov's normal form)

(3) $$H(p,q) = \omega\cdot p + \mathcal{R}(p,q)\ ,$$

where $\mathcal{R}(p,q)$ is at least quadratic in the actions p. For the model (1) $\mathcal{R}(p,q)$ will actually be a homogeneous polynomial of degree 2. It is immediately seen that the torus $p = 0$

is invariant for the flow generated by the Hamiltonian (3), and morever that the orbits on the torus are conditionally periodic with frequencies ω.

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

$$H(p,q) = \omega \cdot p + \frac{p^2}{2} + A(q) + B(q) \cdot p + O(p^2) \tag{4}$$

where the function $A(q)$ and the vector valued function $B(q)$ are assumed to be small. 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

$$S(p', q) = p' \cdot q + X(q) + \xi \cdot q + Y(q) \cdot p' \ , \tag{9}$$

where $X(q)$ and $Y(q)$ are functions, and ξ is a real vector to be determined. Performing such a transformation the Hamiltonian (4) is transformed to a new Hamiltonian $H'(p', q')$ of the same form, with new functions $A'(q)$ and $B'(q)$, which, however, are smaller than $A(q)$ and $B(q)$. By iterating this procedure one ends up with an Hamiltonian in Kolmogorov's normal form. Under analyticity conditions on the Hamiltonian, the transformation turns out to be analytic in a neighbourhood of the invariant torus. Thus, Kolmogorov's normal form contains information not only on the invariant torus, but also on a neighbourhood of it. A technical remark is that no use is made here of power expansions in a parameter. This results in a fast convergence, usually called quadratic. A moment's thought will allow one to realize that the procedure proposed by Kolmogorov is constructive.

2.2 *Canonical transformations via Lie series*

As is well known, the method of Lie series is nothing but an explicit algoritm based on a series expansion of the solution of a system of differential equations (see for instance [12]). However, we shall insist a little on the expansion formulæ, with the aim of making the algorithm clear in all details.

We are particularly interested in canonical transformations. So, let us consider a generating function $\varepsilon\chi(p, q)$ (that will be supposed to be analytic), where ε is a small parameter. The map at time one generated by the canonical flow due to $\varepsilon\chi$ can be represented via the exponential operator

$$\exp(\varepsilon L_\chi) = \sum_{s \geq 0} \frac{\varepsilon^s}{s!} L_\chi^s \ . \tag{6}$$

More explicitly, the map can be written as

$$
\begin{aligned}
p &= \exp(\varepsilon L_\chi)p' = p' + \varepsilon \left.\frac{\partial\chi}{\partial q}\right|_{p',q'} + \frac{\varepsilon^2}{2} L_\chi \left.\frac{\partial\chi}{\partial q}\right|_{p',q'} + \ldots \\
q &= \exp(\varepsilon L_\chi)q' = q' + \varepsilon \left.\frac{\partial\chi}{\partial p}\right|_{p',q'} + \frac{\varepsilon^2}{2} L_\chi \left.\frac{\partial\chi}{\partial p}\right|_{p',q'} + \ldots \;;
\end{aligned}
\tag{7}
$$

The main advantage of using this formalism is that there is no need of inversions, as happens instead in the common method of using a generating function in mixed variables. Moreover, for a function $f(p,q)$ one has the relevant property

$$
f(p,q)\Big|_{p=\exp(\varepsilon L_\chi)p',\,p=\exp(\varepsilon L_\chi)q'} = \exp(\varepsilon L_\chi) f\Big|_{p',q'} \;. \tag{8}
$$

That is, the transformation of the function $f(p,q)$ via the change of variables (7) can be equivalently performed by direct application of the exponential operator (6) to the function f.

The action of the exponential operator is better illustrated as follows: let $f = f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \ldots$ and the generating function $\varepsilon\chi_1$ be known, and look for the series expansion in ε of the transformed function $g := \exp(\varepsilon L_{\chi_1}) f = g_0 + \varepsilon g_1 + \varepsilon^2 g_2 + \ldots$. This operation is illustrated by the triangular diagram

$$
\begin{array}{cccccc}
g_0 & f_0 & & & & \\
 & \downarrow & & & & \\
g_1 & L_{\chi_1} f_0 & f_1 & & & \\
 & \downarrow & \downarrow & & & \\
g_2 & \frac{1}{2} L_{\chi_1}^2 f_0 & L_{\chi_1} f_1 & f_2 & & \\
 & \downarrow & \downarrow & \downarrow & & \\
g_3 & \frac{1}{3!} L_{\chi_1}^3 f_0 & \frac{1}{2} L_{\chi_1}^2 f_1 & L_{\chi_1} f_2 & f_3 & \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
\tag{9}
$$

where terms of the same order appear on the same row. Remark that the operator $\exp L_{\chi_1}$ acts *by columns*, as indicated by the arrows: the knowledge of f_j and of the generating function χ_1 allows one to construct the whole column below f_j. Thus, the first line gives $g_0 = f_0$, the second line gives $g_1 = L_{\chi_1} f_0 + f_1$, and so on. We emphasize that the algorithm furnishes immediately the power expansion in ε of the transformed function.

A similar diagram can be constructed for a generating function $\varepsilon^r \chi_r$, where r is an arbitrary positive integer: there are empty cases, of course. An algebraic expression can also be given. Precisely, the function $g := \exp(\varepsilon^r L_{\chi_r}) f = g_0 + \varepsilon g_1 + \varepsilon^2 g_2 + \ldots$ has coefficients

$$g_s = \sum_{j=0}^{k} \frac{1}{j!} L^j_{\chi_r} f_{s-jr} \ , \quad k = \left\lfloor \frac{s}{r} \right\rfloor \ .$$

In the following we shall also consider the composition of Lie series. Precisely, having given a sequence $\{\varepsilon\chi_1, \varepsilon^2\chi_2 \ldots, \}$ of generating functions one can consider a sequence of operators $\{\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \ldots\}$ recursively defined by $\mathcal{C}^{(0)} = \mathrm{Id}$ and $\mathcal{C}^{(r)} = \exp(\varepsilon^r L_{\chi_r}) \circ \mathcal{C}^{(r-1)}$.

It is natural to look for sufficient conditions for convergence of Lie series and in particular for a composition of an infinite sequence of transformations. Concerning the exponential operator (6), proving the convergence of the series means essentially proving the theorem of existence of solutions for a system of differential equations. This can be done, for instance, using the majorants method of Cauchy. For the composition of Lie series it is not difficult to prove that a sufficient condition is, e.g.,

$$\|\chi_s\| \le b^{s-1} G$$

for some constants b and G, where $\|\cdot\|$ denotes some norm (e.g., the supremum norm). This condition will be used later on.

Let us now consider more closely our model problem. In order to be able to compare the results with the series generated by the method of Lindstedt it is necessary to explain in a much more detailed fashion how the computation should proceed, at least in principle. In the case of Thirring's model we can represent a generic function $f(p, q, \varepsilon)$ as follows.

(i) the function has a power expansion in ε, i.e.

$$f(p, q, \varepsilon) = f_0(p, q) + \varepsilon f_1(p, q) + \varepsilon^2 f_2(p, q) + \ldots \ .$$

(ii) The coefficients f_s admit a Fourier expansion

$$f_s = \sum_{k \in \mathbf{Z}^n} c_k(p) \exp(ik \cdot q) \ ;$$

in our case f_s will be a trigonometric polynomial of some finite degree.

(iii) The coefficient $c_k(p)$ of a Fourier mode is a polynomial in p of the form

$$c_k(p) = \sum_j \gamma_j p^j \ ,$$

j being an array of n nonnegative integers. In our case $c_k(p)$ will actually be a polynomial of degree 2.

(iv) The coefficient γ_j has generically the form

$$\gamma_j = \sum \frac{\beta}{\prod \nu \cdot \omega} \ ,$$

where β is a numerical coefficient, $0 \neq \nu \in \mathbf{Z}^n$ is a Fourier mode, and the sum and the product run over a suitable set of indexes.

The form of the coefficient in (iv) looks natural if one recalls that the perturbation scheme introduces small divisors. This structure is perhaps better understood in terms of lists. The sum of point (iv) is represented as a list of summands. In turn, each element in the latter list contains a numerical coefficient β and a further list of small divisor to be multiplied together. The power of ε, the Fourier mode and the exponents of the polynomial coefficient are just pointers to a list of summands.

Having in mind this representation, the algorithm of Lie series described above can be implemented by simple manipulation of lists. In order to make the algorithm uniquely defined, let us add the rule that sums of functions are performed by simply concatenating the list of summands of the point (iv) above, namely lists corresponding to the same order in ε, Fourier mode and polynomial coefficient: no algebraic operation of partial summing of coefficient is to be performed.

3. Birkhoff's normal form for an isochronous system

As an elementary example, let us consider the case of Birkhoff's normal form. Precisely, we shall consider a system of perturbed harmonic oscillators with Hamiltonian

$$H(p,q) = \omega \cdot p + \varepsilon f_1(p,q) + \varepsilon^2 f_2(p,q) + \ldots \ , \tag{10}$$

where the frequencies ω are assumed to satisfy the diophantine condition (2), and the perturbations $f_s(p,q)$ are assumed to be trigonometric polynomials of degree sK in q, K being a positive integer.

3.1 Formal algorithm

The process of construction of Birkhoff's normal form is well known. Let us recall the first step. One looks for a canonical transformation with generating function $\varepsilon\chi_1$ that kills the dependence on the angles in the first term f_1 of the perturbation. Looking at the triangular diagram (9) it is immediately seen that the term of order ε in the transformed Hamiltonian is $L_{\chi_1} h_0 + f_1$. Thus, one determines the generating function χ_1 and the normal form h_1 via the equations

$$\partial_\omega \chi_1 = f_1 - \langle f_1 \rangle \ , \quad h_1 = \langle f_1 \rangle \ , \tag{11}$$

where $\langle\cdot\rangle$ denotes averaging with respect to the angles. This introduces a small denominator in χ_1. By iterating this procedure, the Hamiltonian after r steps is transformed to Birkhoff's normal form up to order r, say

$$H^{(r)}(p,q) = \omega \cdot p + \varepsilon h_1(p) + \ldots + \varepsilon^r h_r(p) + \varepsilon^{r+1} f^{(r)}_{r+1}(p,q) + \ldots \ ,$$

the superscript r being used to denote the Hamiltonian after r normalization steps. Looking again at the triangular diagram (9) it is easily seen that the perturbation $f_s^{(r)}$, with $s > r$, is a trigonometric polynomial of degree sK in q.

The series so constructed are known to be generally divergent. However they can be proved to be convergent in the very particular case of an isochronous system, i.e., if for some very fortunate event it happens that $h_1 = h_2 = \ldots = 0$. That is, Birkhoff's normal form reduces to the linear part $\omega \cdot p$. In fact, this is a particular case of a more general theorem (see [13]).

Our aim now is to show that an heuristic analysis on the process of accumulation of small divisors shows that the latter case behaves well differently from the general one. Indeed, in the general case the global contribution of the small divisors is expected to be a factorial of r. In contrast, in the completely isochronous case it the global contribution can be controlled geometrically.

3.2 Heuristic analysis of the accumulation of small divisors

We shall first consider the general case. Define a sequence $\{\alpha_r\}_{r>0}$ by

$$\alpha_r = \min_{0<|k|\leq rK} (|k \cdot \omega|) \; ; \tag{12}$$

this is clearly a nonincreasing sequence.

In order to follow the process of accumulation of small divisors, we need the following two informations: (i) solving equation (11) adds a further denominator α_r in the generating function χ_r, besides the ones already present in $f_r^{(r-1)}$; (ii) in computing the Poisson bracket $L_{\chi_r} f$ (where f is any of the functions in the expansion of the Hamiltonian) the denominators are multiplied together. The process is illustrated in table 1. One starts with the original Hamiltonian $H^{(0)}$, which contains terms of order $\varepsilon^0, \varepsilon^1, \ldots$ with no small denominators. By solving equation (11) with $r = 1$ one determines $h_1 \sim \varepsilon$ and $\chi_1 \sim \varepsilon/\alpha_1$. The expansion of $\exp(L_{\chi_1})(\omega \cdot p)$ produces a term of order ε which kills the angle dependent terms in $f_1^{(0)}$. The rest of the expansion gives the same contributions, apart from a numerical factor, as $\exp(L_{\chi_1}) f_1^{(0)}$, and this produces terms $L_{\chi_1} f_1^{(0)} \sim \varepsilon^2/\alpha_1$, $L^2_{\chi_1} f_1^{(0)} \sim \varepsilon^3/\alpha_1^2$, and so on. These terms actually contain the worst possible combination of small divisors. This justifies the line corresponding to $H^{(1)}$. Solving equation (11) for $r = 2$ generates $h_2 \sim \varepsilon^2/\alpha_1$ and $\chi_2 \sim \varepsilon^2/(\alpha_1\alpha_2)$. The transformation of h_1 produces $L_{\chi_2} h_1 \sim \varepsilon^3/(\alpha_1\alpha_2)$, $L^2_{\chi_2} h_1 \sim \varepsilon^5/(\alpha_1^2\alpha_2^2)$, and so on. Similarly, the transformation of $f_2^{(1)}$ produces $L_{\chi_2} f_2^{(1)} \sim \varepsilon^4/(\alpha_1^2\alpha_2)$, and so on. Again, it is an easy matter to check that these terms contain the worst possible combinations of small divisors. The rest of the table is constructed by the same procedure. We emphasize that at every step r the term h_1 of the normal form produces terms of order ε^{r+1}, which contribute to the generating function of the next step. A moment's thought allows us

Table 1. Scheme of accumulation of small divisors throughout the process of Birkhoff normalization.

$H^{(0)}$:	1	ε	ε^2	ε^3	ε^4	ε^5	$\ldots$
$\chi_1 : \frac{\varepsilon}{\alpha_1}$							
$H^{(1)}$:	1	ε	$\frac{\varepsilon^2}{\alpha_1}$	$\frac{\varepsilon^3}{\alpha_1^2}$	$\frac{\varepsilon^4}{\alpha_1^3}$	$\frac{\varepsilon^5}{\alpha_1^4}$	$\ldots$
$\chi_2 : \frac{\varepsilon^2}{\alpha_1\alpha_2}$							
$H^{(2)}$:	1	ε	$\frac{\varepsilon^2}{\alpha_1}$	$\frac{\varepsilon^3}{\alpha_1\alpha_2}$	$\frac{\varepsilon^4}{\alpha_1^2\alpha_2}$	$\frac{\varepsilon^5}{\alpha_1^2\alpha_2^2}$	$\ldots$
$\chi_3 : \frac{\varepsilon^3}{\alpha_1\alpha_2\alpha_3}$							
$H^{(3)}$:	1	ε	$\frac{\varepsilon^2}{\alpha_1}$	$\frac{\varepsilon^3}{\alpha_1\alpha_2}$	$\frac{\varepsilon^4}{\alpha_1\alpha_2\alpha_3}$	$\frac{\varepsilon^5}{\alpha_1^2\alpha_2\alpha_3}$	$\ldots$
$\chi_4 : \frac{\varepsilon^4}{\alpha_1\alpha_2\alpha_3\alpha_4}$							
$H^{(4)}$:	1	ε	$\frac{\varepsilon^2}{\alpha_1}$	$\frac{\varepsilon^3}{\alpha_1\alpha_2}$	$\frac{\varepsilon^4}{\alpha_1\alpha_2\alpha_3}$	$\frac{\varepsilon^5}{\alpha_1\alpha_2\alpha_3\alpha_4}$	$\ldots$
$\chi_5 : \frac{\varepsilon^5}{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5}$							
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

to conclude that the generating function at step r, namely χ_r, is expected to contain terms of order

$$\frac{\varepsilon^r}{\alpha_1\alpha_2\cdot\ldots\cdot\alpha_r}\ .$$

In view of the diophantine condition (1), we have $\alpha_s > \gamma(sK)^{-\tau}$, so that the best we can do is to estimate

$$\chi_r \sim \left(\frac{K^\tau\varepsilon}{\gamma}\right)^r (r!)^\tau\ .$$

Table 2. Scheme of accumulation of small divisors in the case of linear Birkhoff's normal form.

	$H^{(0)}$:	1	ε	ε^2	ε^3	ε^4	ε^5	$\ldots$
χ_1 :	$\frac{\varepsilon}{\alpha_1}$							
	$H^{(1)}$:	1	0	$\frac{\varepsilon^2}{\alpha_1}$	$\frac{\varepsilon^3}{\alpha_1^2}$	$\frac{\varepsilon^4}{\alpha_1^3}$	$\frac{\varepsilon^5}{\alpha_1^4}$	$\ldots$
χ_2 :	$\frac{\varepsilon^2}{\alpha_1\alpha_2}$							
	$H^{(2)}$:	1	0	0	$\frac{\varepsilon^3}{\alpha_1^2}$	$\frac{\varepsilon^4}{\alpha_1^2\alpha_2}$	$\frac{\varepsilon^5}{\alpha_1^3\alpha_2}$	$\ldots$
χ_3 :	$\frac{\varepsilon^3}{\alpha_1^2\alpha_3}$							
	$H^{(3)}$:	1	0	0	0	$\frac{\varepsilon^4}{\alpha_1^2\alpha_2}$	$\frac{\varepsilon^5}{\alpha_1^3\alpha_2}$	$\ldots$
χ_4 :	$\frac{\varepsilon^4}{\alpha_1^2\alpha_2\alpha_4}$							
	$H^{(4)}$:	1	0	0	0	0	$\frac{\varepsilon^5}{\alpha_1^3\alpha_2}$	$\ldots$
χ_5 :	$\frac{\varepsilon^5}{\alpha_1^3\alpha_2\alpha_5}$							
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Thus, we cannot prove the convergence of Birkhoff's normal form via this kind of estimates.

Let us now turn to the case of a completely isochronous system. The main point is that the terms $h_1, h_2, \ldots$ do not appear, and so they do not contribute to the next orders. This has strong consequences on the accumulation of small denominators, as illustrated in table 2. A comparison with table 1 immediately shows that starting with $H^{(2)}$ the estimated denominators behave better than in the general case. The reason is precisely that there are no terms h_1 , h_2 , ... in the normal form. Indeed, the transformation of

$H^{(1)}$ by χ_2 does not change the terms of order ε^3, because the lowest order contribution is $L_{\chi_2} f_2^{(1)}$, which is of order ε^4. Similarly, the transformation of $H^{(2)}$ by χ_3 does not change the terms of order ε^4 and ε^5, the first change occurring at order ε^6. A moment's thought suffices to realize that the transformation from $H^{(r-1)}$ to $H^{(r)}$ generated by χ_r does not change the terms of order $\varepsilon^{r+1}, \ldots, \varepsilon^{2r-1}$, the first change in the transformed Hamiltonian occurring at order ε^{2r}. By the way, this is similar to the mechanism of accumulation of small divisors of the quadratic method. In order to exploit this idea, let us follow only orders which are powers of 2. The relevant information is that the generating function χ_r is determined by $f_r^{(r-1)}$, and determines $f_{2r}^{(r)} = \ldots = f_{2r}^{(2r-1)}$ (that is, $f_{2r}^{(r)}$ remains unchanged until it enters the generating function χ_{2r} at step $2r$). We get the following table:

$$
\begin{array}{lllll}
f_1^{(0)} : \varepsilon & \longrightarrow & \chi_1 : \varepsilon \cdot \dfrac{1}{\alpha_1} & \longrightarrow \\
f_2^{(1)} : \dfrac{\varepsilon^2}{\alpha_1} & \longrightarrow & \chi_2 : \dfrac{\varepsilon^2}{\alpha_1} \cdot \dfrac{1}{\alpha_2} & \longrightarrow \\
f_4^{(2)} : \dfrac{\varepsilon^4}{\alpha_1^2 \alpha_2} & \longrightarrow & \chi_4 : \dfrac{\varepsilon^4}{\alpha_1^2 \alpha_2} \cdot \dfrac{1}{\alpha_4} & \longrightarrow \\
f_8^{(4)} : \dfrac{\varepsilon^8}{\alpha_1^4 \alpha_2^2 \alpha_4} & \longrightarrow & \chi_8 : \dfrac{\varepsilon^8}{\alpha_1^4 \alpha_2^2 \alpha_4} \cdot \dfrac{1}{\alpha_8} & \longrightarrow \\
f_{16}^{(8)} : \dfrac{\varepsilon^{16}}{\alpha_1^8 \alpha_2^4 \alpha_4^2 \alpha_8} & \longrightarrow & \chi_{16} : \dfrac{\varepsilon^{16}}{\alpha_1^8 \alpha_2^4 \alpha_4^2 \alpha_8} \cdot \dfrac{1}{\alpha_{16}} & \longrightarrow \\
\ldots : \ldots & & &
\end{array}
$$

In view of this, it is natural to guess that at order 2^s we shall get

$$
f_{2^s}^{(2^{s-1})} : \frac{\varepsilon^{2^s}}{\alpha_1^{2^{s-1}} \alpha_2^{2^{s-2}} \alpha_4^{2^{s-3}} \cdot \ldots \cdot \alpha_{2^{s-1}}} \longrightarrow \chi_{2^s} : \frac{\varepsilon^{2^s}}{\alpha_1^{2^{s-1}} \alpha_2^{2^{s-2}} \alpha_4^{2^{s-3}} \cdot \ldots \cdot \alpha_{2^{s-1}}} \cdot \frac{1}{\alpha_{2^s}} \ .
$$

Of course, this tells us nothing about the form of the worst possible denominator at orders which are not powers of 2. We need a conjecture to be checked by induction. The conjecture is the following:

The worst possible denominators in the generating function χ_r are products of the form $\alpha_{j_1} \cdot \ldots \cdot \alpha_{j_s}$ satisfying the following rules:

(i) the indexes j_k do not exceed r;

(ii) the number of factors α_{j_k} is at most r;

(iii) the indexes obey the selection rule

$$
\sum_k \log_2 j_k \le r - 1 \ .
$$

It is an easy matter to check that the conjecture fits exactly the general form of the worst possible denominator at order $r = 2^s$ (use induction). Proving that it works fine at every order requires a recursive scheme of estimates on all functions entering the process of construction of the normal form. We omit this part, which is purely technical.

Thus, we are led to the conclusion that the term of order r of the generating function is estimated as

$$\chi_r \simeq \frac{\varepsilon^r}{\alpha_{j_1} \cdot \ldots \cdot \alpha_{j_r}}$$

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 (1), then the expression above grows not faster than geometrically with r. Indeed, according to (1) one has $\alpha_s \geq \gamma(sK)^{-\tau}$, and so

$$\chi_r \simeq \varepsilon^r \prod_{s=1}^{r} \frac{K^\tau}{\gamma} j_s^\tau \ .$$

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

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

We conclude

$$\chi_r \simeq \frac{1}{2} \left(\frac{2^\tau K^\tau \varepsilon}{\gamma} \right)^r \ ,$$

as claimed.

Using this elementary remark, it is not difficult to show that the generating functions χ_r grow not faster than geometrically with r. This kind of estimates is now quite well known, so we do not enter this matter in more detail.

4. Kolmogorov's normal form

My aim is now to show that the same mechanism which shows up in the (very exceptional) case of isochronous systems works, essentially in the same form, in the (much more interesting) case of Kolmogorov's normal form.

4.1 Formal algorithm

Let us reformulate the algorithm of Kolmogorov in terms of Lie series, applying it to the Hamiltonian (1). By Kolmogorov's normal form up to order r we mean an Hamiltonian of the form

$$H^{(r)} = \omega\cdot p + \sum_{s=0}^{r} \varepsilon^s h^{(s)} + \sum_{s\ge r} \varepsilon^s \left[A_s^{(r)} + B_s^{(r)} + C_s^{(r)}\right] \ , \tag{13}$$

where $A_s^{(r)}$ and $B_s^{(r)}$ are independent of p and linear in p, respectively, and $C_s^{(r)}$ and $h^{(s)}$ are quadratic in p. The Hamiltonian (1) is to be identified with $H^{(0)}$, and has already this form with $h^{(0)} = p^2/2$ and $A_1^{(0)} = f$, while all other functions A, B and C are zero.

Explicit expressions for all functions can be calculated by referring again to the triangular diagram (9). We report here the result, i.e., the algorithm for performing a normalization step.

Assume that the Hamiltonian is already in Kolmogorov's normal form up to order $r-1$ (i.e., $H^{(r-1)}$ is known). The generating functions are determined by the equations

$$\begin{aligned} &\partial_\omega X^{(r)} + A_r^{(r-1)} = 0 \ ,\\ &\xi^{(r)}\cdot p + \langle B_r^{(r-1)}\rangle = 0 \ ,\\ &\partial_\omega \chi_2^{(r)} + \left\{X^{(r)}, h^{(0)}\right\} + B_r^{(r-1)} - \langle B_r^{(r-1)}\rangle = 0 \ , \end{aligned} \tag{14}$$

Having determined $\chi_1^{(r)} = X^{(r)} + \xi^{(r)}\cdot q$, we transform

$$\hat H^{(r)} := \exp(\varepsilon^r L_{\chi_1^{(r)}}) H^{(r-1)} = \omega\cdot p + \sum_{s=0}^{r} \varepsilon^s h^{(s)} + \sum_{s\ge r} \varepsilon^s \left[\hat A_s^{(r)} + \hat B_s^{(r)} + C_s^{(r)}\right]$$

where

$$\hat A_r^{(r)} = 0$$

$$\hat A_s^{(r)} = \begin{cases} A_s^{(r-1)} A_s^{(r-1)} & r < s < 2r \\ \frac{1}{2} L^2_{\chi_1^{(r)}} h^{(s-2r)} + L_{\chi_1^{(r)}} B_{s-r}^{(r-1)} + A_s^{(r-1)} & 2r \le s < 3r \\ \frac{1}{2} L^2_{\chi_1^{(r)}} C_{s-2r}^{(r-1)} + L_{\chi_1^{(r)}} B_{s-r}^{(r-1)} + A_s^{(r-1)} & s \ge 3r \end{cases}$$

$$\hat B_s^{(r)} = \begin{cases} L_{\chi_1^{(r)}} h^{(r-1)} + B_s^{(r-1)} & r \le s < 2r \\ L_{\chi_1^{(r)}} C_{s-r}^{(r-1)} + B_s^{(r-1)} & s \ge 2r \end{cases}$$

Next, denoting

$$k = \left\lfloor \frac{s}{r} \right\rfloor \ , \quad m = s \ (\mathrm{mod}\, r) \ , \quad s = kr + m \ ,$$

we perform the transformation

$$H^{(r)} = \exp(\varepsilon^r L_{\chi_2^{(r)}}) \hat H^{(r)}$$

with $H^{(r)}$ in normal form (13), where

$$
\begin{aligned}
h^{(r)} &= L_{\chi_2^{(r)}} h^{(0)} + C_r^{(r)} \\
A_s^{(r)} &= \sum_{j=0}^{k-1} \frac{1}{j!} L^j_{\chi_2^{(r)}} \hat{A}^{(r)}_{s-jr} \\
B_s^{(r)} &= \begin{cases} \dfrac{k-1}{k!} L^{k-1}_{\chi_2^{(r)}} \hat{B}^{(r)}_r + \displaystyle\sum_{j=0}^{k-2} \frac{1}{j!} L^j_{\chi_2^{(r)}} \hat{B}^{(r)}_{s-jr} & k \geq 2\,,\ m = 0 \\ \displaystyle\sum_{j=0}^{k-1} \frac{1}{j!} L^j_{\chi_2^{(r)}} \hat{B}^{(r)}_{s-jr} & k \geq 1\,,\ m \neq 0 \end{cases} \\
C_s^{(r)} &= \frac{1}{k!} L^k_{\chi_2^{(r)}} h^{(m)} + \sum_{j=0}^{k-1} \frac{1}{j!} L^j_{\chi_2^{(r)}} C^{(r)}_{s-jr}
\end{aligned}
$$

Obtaining these formulæ is a bit tedious, but not difficult. There is just one point which could result somehow obscure, in computing the first expression of $B_s^{(r)}$ for $m = 0$. One should take into account that in view of (14) all expressions $L^k_{\chi_2^{(r)}}(\omega \cdot p)$ are partially compensated by $L^{k-1}_{\chi_2^{(r)}} \hat{B}^{(r)}_r$. This explains why the former expression does not appear, and there is a factor $(k-1)/k!$ in front of the latter.

4.2 Equations of the invariant torus

Suppose that we are looking for the equation of the invariant torus up to terms of order ε^r. To this end, we should first construct the generating functions $\chi_1^{(1)}, \chi_2^{(1)}, \ldots, \chi_1^{(r)}, \chi_2^{(r)}$ according to the scheme of the previous section. Then, by composition, we have determined a canonical transformation

$$
\mathcal{C}^{(r)} = \exp(\varepsilon^r L_{\chi_2^{(r)}}) \circ \exp(\varepsilon^r L_{\chi_1^{(r)}}) \circ \cdots \circ \exp(\varepsilon L_{\chi_2^{(1)}}) \circ \exp(\varepsilon L_{\chi_1^{(1)}}) \ .
$$

Thus, denoting by $p^{(r)}, q^{(r)}$ the canonical variables after r steps, we can write the canonical transformation, forgetting terms of order higher than r, in the form

$$
\begin{aligned}
q &= \mathcal{C}^{(r)} q^{(r)} = q^{(r)} + \varepsilon \varphi_1(q^{(r)}) + \varepsilon^2 \varphi_2(q^{(r)}) + \ldots + \varepsilon^r \varphi_r(q^{(r)}) \\
p &= \mathcal{C}^{(r)} p^{(r)} = p^{(r)} + \varepsilon \psi_1(p^{(r)}, q^{(r)}) + \varepsilon^2 \psi_2(p^{(r)}, q^{(r)}) + \ldots + \varepsilon^r \psi_r(p^{(r)}, q^{(r)}) \ ,
\end{aligned}
\tag{15}
$$

where the functions $\varphi_1(q^{(r)}), \ldots \varphi_r(q^{(r)})$ and $\psi_1(p^{(r)}, q^{(r)}), \ldots, \psi_r(p^{(r)}, q^{(r)})$ are completely determined (that is, are not changed when performing transformation of order higher than r).

The corresponding transformation for any function $f(p,q)$ takes the form

$$f^{(r)}(p^{(r)},q^{(r)}) = f(p,q)\Big|_{p=\mathcal{C}^{(r)}p^{(r)},q=\mathcal{C}^{(r)}q^{(r)}} = \mathcal{C}^{(r)}f\Big|_{p^{(r)},q^{(r)}} \ ;$$

in order to check this, use (8).

By construction, the transformed Hamiltonian is in Kolmogorov's normal form (13). Thus, forgetting terms of order higher than r, the torus $p^{(r)} = 0$ is invariant and carries a quasiperiodic motion $q^{(r)}(t) = \omega t + q^{(r)}(0)$. The equation of the torus is obtained by replacing $p^{(r)} = 0$ in the equations (15). This gives the parametric equations

$$\begin{aligned} q &= q^{(r)} + \varepsilon\varphi_1(q^{(r)}) + \varepsilon^2\varphi_2(q^{(r)}) + \ldots + \varepsilon^r\varphi_r(q^{(r)}) \\ p &= \varepsilon\psi_1(0,q^{(r)}) + \varepsilon^2\psi_2(0,q^{(r)}) + \ldots + \varepsilon^r\psi_r(0,q^{(r)}) \ . \end{aligned} \tag{16}$$

The convergence of the transformation to normal form, still to be proved, guarantees that this is the correct approximation of the invariant torus, within an error of order ε^r. The expressions (16) are to be compared with the result of Lindstedt's series.

4.3 Heuristic analysis of the accumulation of small divisors

My aim now is to illustrate the mechanism of accumulation of small divisors, and in particular to show that it is essentially the same as in the case of Birkhoff's normal form for an isochronous system. Here too we shall restrict our attention to the worst possible combination of small divisors. Again, the relevant information is that a new small denominator appears when determining the generating functions $X^{(r)}$ and $\chi_2^{(r)}$ by solving the equations (14), and propagates through the application of the corresponding exponential operators, as illustrated by the triangular diagram (9).

The first step of the procedure is illustrated in table 3. The parameter ϱ takes into account the degree in p. So, for instance, in $H^{(0)}$ the terms $\omega\cdot p$ and $p^2/2$ are represented by ϱ and ϱ^2, respectively; the perturbations $\varepsilon^s A_s^{(0)}(q)$, $\varepsilon^s B_s^{(0)}(p,q)$ and $\varepsilon^s C_s^{(0)}(p,q)$ are represented by ε^s, $\varepsilon^s\varrho$ and $\varepsilon^s\varrho^2$, respectively. The fact that in the Hamiltonian (1) only $A_1^{(0)}$ is non zero is not relevant.

Solving the equation for $X^{(1)}$ introduces in $\chi_1^{(1)}$ a small denominator α_1 in the worst case (recall the definition (12) of the numbers α_r). Applying the exponential operator $\exp(\varepsilon L_{\chi_1^{(1)}})$ to the term ϱ, i.e., $\omega\cdot p$, produces $\varepsilon L_{\chi_1^{(1)}}(\omega\cdot p)$, which kills the term $\sim\varepsilon$ of degree zero in the actions; the rest of the expansion is zero. Applying the exponential operator to $p^2/2$ produces $\varepsilon L_{\chi_1^{(1)}}p^2/2 \sim \varepsilon\varrho/\alpha_1$ and $\varepsilon^2 L^2_{\chi_1^{(1)}}p^2/2 \sim \varepsilon^2/\alpha_1^2$, and nothing else. More generally, applying the exponential operator to a term $\sim\varepsilon^s\varrho^2$ produces $\varepsilon^{s+1}\varrho/\alpha_1$ and $\varepsilon^{s+2}/\alpha_1^2$. All these terms are accounted for in the three lines corresponding to $\hat{H}^{(1)}$. The reader can easily check that these are the paths producing the worst possible accumulation of divisors.

Table 3. Scheme of accumulation of small divisors during the first step of construction of Kolmogorov's normal form.

	ϱ^2	$\varepsilon\varrho^2$	$\varepsilon^2\varrho^2$	$\varepsilon^3\varrho^2$	$\varepsilon^4\varrho^2$	$\varepsilon^5\varrho^2$	$\ldots$
$H^{(0)}$:	ϱ	$\varepsilon\varrho$	$\varepsilon^2\varrho$	$\varepsilon^3\varrho$	$\varepsilon^4\varrho$	$\varepsilon^5\varrho$	$\ldots$
	0	ε	ε^2	ε^3	ε^4	ε^5	$\ldots$
$\varepsilon\chi_1^{(1)}$: $\dfrac{\varepsilon}{\alpha_1}$							
	ϱ^2	$\varepsilon\varrho^2$	$\varepsilon^2\varrho^2$	$\varepsilon^3\varrho^2$	$\varepsilon^4\varrho^2$	$\varepsilon^5\varrho^2$	$\ldots$
$\hat{H}^{(1)}$:	ϱ	$\dfrac{\varepsilon\varrho}{\alpha_1}$	$\dfrac{\varepsilon^2\varrho}{\alpha_1}$	$\dfrac{\varepsilon^3\varrho}{\alpha_1}$	$\dfrac{\varepsilon^4\varrho}{\alpha_1}$	$\dfrac{\varepsilon^5\varrho}{\alpha_1}$	$\ldots$
	0	0	$\dfrac{\varepsilon^2}{\alpha_1^2}$	$\dfrac{\varepsilon^3}{\alpha_1^2}$	$\dfrac{\varepsilon^4}{\alpha_1^2}$	$\dfrac{\varepsilon^5}{\alpha_1^2}$	$\ldots$
$\varepsilon\chi_2^{(1)}$: $\dfrac{\varepsilon\varrho}{\alpha_1^2}$							
	ϱ^2	$\dfrac{\varepsilon\varrho^2}{\alpha_1^2}$	$\dfrac{\varepsilon^2\varrho^2}{\alpha_1^4}$	$\dfrac{\varepsilon^3\varrho^2}{\alpha_1^6}$	$\dfrac{\varepsilon^4\varrho^2}{\alpha_1^8}$	$\dfrac{\varepsilon^5\varrho^2}{\alpha_1^{10}}$	$\ldots$
$H^{(1)}$:	ϱ	0	$\dfrac{\varepsilon^2\varrho}{\alpha_1^3}$	$\dfrac{\varepsilon^3\varrho}{\alpha_1^5}$	$\dfrac{\varepsilon^4\varrho}{\alpha_1^7}$	$\dfrac{\varepsilon^5\varrho}{\alpha_1^9}$	$\ldots$
	0	0	$\dfrac{\varepsilon^2}{\alpha_1^2}$	$\dfrac{\varepsilon^3}{\alpha_1^4}$	$\dfrac{\varepsilon^4}{\alpha_1^6}$	$\dfrac{\varepsilon^5}{\alpha_1^8}$	$\ldots$

Solving the equation for $\chi_2^{(1)}$ introduces a further divisor α_1. Then the application of the exponential operator $\exp(\varepsilon L_{\chi_2^{(1)}})$ to $p^2/2$ produces terms of order $\varepsilon\varrho^2/\alpha_1^2, \varepsilon^2\varrho^2/\alpha_1^4, \ldots$, which are accounted for in the first line referring to $H^{(1)}$. Similarly, the application of the exponential operator to $\omega \cdot p$ kills the term of order $\varepsilon\varrho/\alpha_1$, and generates terms of order $\varepsilon^2\varrho/\alpha_1^3, \varepsilon^3\varrho/\alpha_1^5, \ldots$, as indicated in the second line. Finally, the application of the exponential operator to the term of order ε^2/α_1^2 generates $\varepsilon^3/\alpha_1^4, \ldots$, as indicated in the third line. As above, this produces the worst possible accumulation of small divisors.

The second step is illustrated in table 4. The part referring to $H^{(1)}$ is just a copy of the corresponding part of table 3. Determining the generating function $X^{(2)}$ introduces a small divisor α_2, and kills the term of order ε^2 and of degree zero in the actions. Then the small divisors propagate, e.g., from ϱ^2 to $\varepsilon^2\varrho/(\alpha_1^2\alpha_2)$ and to $\varepsilon^4/(\alpha_1^4\alpha_2^2)$; no further

Table 4. Scheme of accumulation of small divisors during the second step of construction of Kolmogorov's normal form.

	ϱ^2	$\frac{\varepsilon\varrho^2}{\alpha_1^2}$	$\frac{\varepsilon^2\varrho^2}{\alpha_1^4}$	$\frac{\varepsilon^3\varrho^2}{\alpha_1^6}$	$\frac{\varepsilon^4\varrho^2}{\alpha_1^8}$	$\frac{\varepsilon^5\varrho^2}{\alpha_1^{10}}$	$\ldots$
$H^{(1)}$:	ϱ	0	$\frac{\varepsilon^2\varrho}{\alpha_1^3}$	$\frac{\varepsilon^3\varrho}{\alpha_1^5}$	$\frac{\varepsilon^4\varrho}{\alpha_1^7}$	$\frac{\varepsilon^5\varrho}{\alpha_1^9}$	$\ldots$
	0	0	$\frac{\varepsilon^2}{\alpha_1^2}$	$\frac{\varepsilon^3}{\alpha_1^4}$	$\frac{\varepsilon^4}{\alpha_1^6}$	$\frac{\varepsilon^5}{\alpha_1^8}$	$\ldots$
$\varepsilon^2\chi_1^{(2)}$:	$\frac{\varepsilon^2}{\alpha_1^2\alpha_2}$						
	ϱ^2	$\frac{\varepsilon\varrho^2}{\alpha_1^2}$	$\frac{\varepsilon^2\varrho^2}{\alpha_1^4}$	$\frac{\varepsilon^3\varrho^2}{\alpha_1^6}$	$\frac{\varepsilon^4\varrho^2}{\alpha_1^8}$	$\frac{\varepsilon^5\varrho^2}{\alpha_1^{10}}$	$\ldots$
$\hat{H}^{(2)}$:	ϱ	0	$\frac{\varepsilon^2\varrho}{\alpha_1^2\alpha_2}$	$\frac{\varepsilon^3\varrho}{\alpha_1^4\alpha_2}$	$\frac{\varepsilon^4\varrho}{\alpha_1^6\alpha_2}$	$\frac{\varepsilon^5\varrho}{\alpha_1^8\alpha_2}$	$\ldots$
	0	0	0	$\frac{\varepsilon^3}{\alpha_1^4}$	$\frac{\varepsilon^4}{\alpha_1^4\alpha_2^2}$	$\frac{\varepsilon^5}{\alpha_1^6\alpha_2^2}$	$\ldots$
$\varepsilon^2\chi_2^{(2)}$:	$\frac{\varepsilon^2\varrho}{\alpha_1^2\alpha_2^2}$						
	ϱ^2	$\frac{\varepsilon\varrho^2}{\alpha_1^2}$	$\frac{\varepsilon^2\varrho^2}{\alpha_1^2\alpha_2^2}$	$\frac{\varepsilon^3\varrho^2}{\alpha_1^4\alpha_2^2}$	$\frac{\varepsilon^4\varrho^2}{\alpha_1^4\alpha_2^4}$	$\frac{\varepsilon^5\varrho^2}{\alpha_1^6\alpha_2^4}$	$\ldots$
$H^{(2)}$:	ϱ	0	0	$\frac{\varepsilon^3\varrho}{\alpha_1^4\alpha_2}$	$\frac{\varepsilon^4\varrho}{\alpha_1^4\alpha_2^3}$	$\frac{\varepsilon^5\varrho}{\alpha_1^6\alpha_2^3}$	$\ldots$
	0	0	0	$\frac{\varepsilon^3}{\alpha_1^4}$	$\frac{\varepsilon^4}{\alpha_1^4\alpha_2^2}$	$\frac{\varepsilon^5}{\alpha_1^6\alpha_2^2}$	$\ldots$

terms are generated. A similar propagation occurs for all terms which are quadratic in ϱ. This is accounted for in the line corresponding to $\hat{H}^{(2)}$. Determining the generating function $\chi_2^{(2)}$ introduces a further divisor α_2, and kills the term $\varepsilon^2\varrho/(\alpha_1^2\alpha_2)$. Then the application of the exponential operator propagates horizontally on each of the three lines corresponding to $H^{(2)}$, starting from ϱ^2 and $\varepsilon\varrho^2/\alpha_1^2$ for the first line, from $\varepsilon^2\varrho/(\alpha_1^2\alpha_2)$ and $\varepsilon^3\varrho/(\alpha_1^4\alpha_2)$ for the second line, and from ε^3/α_1^4 and $\varepsilon^4/(\alpha_1^4\alpha_2^2)$ for the third line.

The correspondence with the case of Birkhoff's normal form for an isochronous system appears here, although it requires a bit of attention. Indeed, one should notice that the transformation with $\chi_2^{(r)}$ at step r propagates horizontally (i.e., the degree in the actions p is preserved), and leaves unchanged all terms of the second and third line up to ε^{2r-1}. This corresponds exactly to what happens in the Birkhoff's case. However, some doubt can raise because the transformation with $\chi_1^{(r)}$ actually changes the second line, thus affecting $\chi_2^{(r)}$; moreover, the transformation with $\chi_2^{(r)}$ affects everything starting from order ε^r in the first line (i.e., it changes the quadratic terms). We claim that this does not invalidate the scheme. Indeed, the change in the term of order ε^r due to $\chi_1^{(r)}$ just means that two small denominators instead of one are generated at each step. On the other hand, no term of the first line (which are quadratic in p) affects the determination of the generating functions unless it is first composed with a generating function $\chi_1^{(r)}$, which pushes its contribution to higher orders. This remark is the key of our mechanism of control of small denominators.

Of course, the heuristic considerations made here are not sufficient in order to establish the convergence of the procedure. We need a conjecture that can be checked, for instance by induction. Here is a good conjecture. Referring to the representation of function illustrated in sect. 2.2, consider in particular a generic summand in the expression of point (iv). Define three quantities, that will be called *number of factors*, *maximal size* and *selection rule* as follows:

$$\frac{\beta}{\prod \nu\cdot\omega} \quad : \quad \begin{cases} \text{number of factors} & = \quad \#\{\nu\} \\ \text{maximal size} & = \quad \max(|\nu|) \\ \text{selection rule} & = \quad \sum \log_2 \left\lceil \dfrac{|\nu|}{K} \right\rceil \end{cases}$$

That is, recalling that the product on the left runs over a set of values of ν, the number of factors is actually the length of the list of small denominators associated to the product, the maximal size bounds $|\nu|$, and so also the smallest possible denominator, and the selection rule is a quantity that can be used in order to put some restriction on the accumulation of small divisors. The correspondence with the case of isochronous systems is evident.

These quantities can be associated to every function: recalling that the structure related to any function contains several summands of the type above, just take the worst case. Now, we claim that through the process of construction of Kolmogorov's normal form these three quantities are bounded according to the following table:

	size	number	selection
$\chi_1^{(r)}$	r	$2r-1$	$2r-2-\log_2 r$
$\chi_2^{(r)}$	r	$2r$	$2r-2$
$h^{(r)}$	r	$2r$	$2r-2$
$A_s^{(r)}, \hat{A}_s^{(r)}$	r	$2s-2$	$2s-2-2\log_2 s$
$B_s^{(r)}, \hat{B}_s^{(r)}$	r	$2s-1$	$2s-2+\log_2 r-2\log_2 s$
$C_s^{(r)}$	r	$2s$	$2s-2+2\log_2 r-2\log_2 s$

Making the conjecture which leads to this table requires some time. However, checking it is not difficult: just check the effect of solving the equations for the generating functions and of performing a Poisson bracket. This is left to the reader.

Looking in particular at the generating function, it is now immediately realized that the argument at the end of sect. 3.2 applies again. This proves that in view of the diophantine condition the contribution of small divisors to the coefficient of ε^r is bounded by C^r, for some positive C.

4.4 Scheme of the proof of convergence

The discussion of the previous section shows that the presence of small denominators does not affect the convergence of the procedure. It remains to verify that no other obstacles subsist. For instance, we should check that the number of summands generated by the algorithm does not increase too fast. A possibility is to make use of the known technology of perturbation expansions. We illustrate here the main points of the proof.

Considering a function

$$f(p,q) = \sum_{j,k} c_{jk} p^j \exp(ik\cdot q) \ , \quad c_{jk} = \sum \frac{\beta}{\prod \nu\cdot\omega}$$

and recalling the representation of sect. 2.2, we define the norm of f, depending on a weight σ, as

$$\|f\|_\sigma = \sum_{j,k} |c_{jk}| \exp(|k|\sigma) \ , \quad |c_{jk}| = \sum \frac{|\beta|}{\prod |\nu\cdot\omega|} \ .$$

The action of Poisson brackets is estimated by the inequalities

$$\left\| \frac{1}{s!} L^s_{\chi_1^{(r)}} f \right\|_{(1-d)\sigma} \le \left(\frac{2\,\|X^{(r)}\|_\sigma}{d\sigma} + 2e|\xi^{(r)}| \right)^s \|f\|_\sigma$$

$$\left\| \frac{1}{s!} L^s_{\chi_2^{(r)}} f \right\|_{(1-d)\sigma} \le \left(\frac{3\|\chi_2^{(r)}\|_\sigma}{d\sigma} \right)^s \|f\|_\sigma \ ,$$

where $d < 1$ is any positive number.

With these tools we can construct a scheme of recursive estimates. Consider the sequence $\{\alpha_r\}_{r\geq 1}$ defined by (12); let $\{\delta_r\}_{r\geq 1}$, be a nonincreasing sequence of positive numbers satisfying $\sum_{r\geq 1}\delta_r < 1/2$, and let the sequence $\{d_r\}_{r\geq 0}$ be recursively defined as $d_0 = 0$ and $d_r = d_{r-1} + 2\delta_r$. Then look for a recursive scheme of estimates in such a way that all functions in the Hamiltonian $H^{(r)}$ have bounded norms $\|\cdot\|_{(1-d_r)\sigma}$. Such bound can clearly be produced for $H^{(0)}$, so let us proceed recursively, assuming that we have estimates for $H^{(r-1)}$.

In view of the equations (14) the generating functions satisfy the estimates

$$\begin{aligned}
\|X^{(r)}\|_{(1-d_{r-1})\sigma} &\leq \frac{1}{\alpha_r}\|A_r^{(r-1)}\|_{(1-d_{r-1})\sigma} \\
|\xi^{(r)}| &\leq \|B_r^{(r-1)}\|_{(1-d_{r-1})\sigma} \\
\|\chi_2^{(r)}\|_{(1-d_{r-1}-\delta_r)\sigma} &\leq \frac{1}{\alpha_r}\left(\frac{2\|A_r^{(r-1)}\|_{(1-d_{r-1})}\left\|h^{(0)}\right\|_\sigma}{\alpha_r\delta_r e\sigma} + \|B_r^{(r-1)}\|_{(1-d_{r-1})\sigma}\right)
\end{aligned}$$

where $|\xi^{(r)}| = |\xi_1^{(r)}| + \ldots|\xi_n^{(r)}|$, and α_r are the lower bounds to small denominators defined in (12).

Using these estimates it is not difficult to realize that one will find the following estimates for the generating functions

$$\begin{aligned}
\frac{\|X^{(r)}\|_{(1-d_{r-1})\sigma}}{\delta_r\sigma} + e|\xi^{(r)}| &\leq \mu_r\zeta^{2r-1}\,\|f\|_\sigma^r \max\left(\frac{1}{\prod\delta\alpha}\right) \\
\frac{\chi_2^{(r)}}{\delta_r\sigma} &\leq \mu_r\zeta^{2r}\,\|f\|_\sigma^r \max\left(\frac{1}{\prod\delta\alpha}\right)
\end{aligned}$$

where $\{\mu_r\}_{r\geq 1}$ is a sequence of numbers to be determined,

$$\zeta = \max\left\{\frac{2}{\sigma} + 2e, \frac{3}{\sigma}\left(1 + \frac{2\left\|h^{(0)}\right\|_\sigma}{e\sigma}\right)\right\}\ ,$$

and the product is over a set of α's and δ's to be identified.

For $r = 1$ these estimates can be checked immediately, with $\mu_1 = 1$. For $r > 1$ one could proceed by induction. However, the following arguments show that one must obtain an estimate of the type above. The factor $\|f\|_\sigma$ always appears with ε, so it must have the same exponent as ε, namely r. The factor ζ is due to the estimate of Poisson brackets with generating functions. Now, according to the estimates above, every generating function introduces a new divisor α_r. Thus, the number of factors ζ is the same as the number of factors α_r, which is maximized by the table of the previous section. Since, by definition, $\zeta > 1$, we determine the exponent of ζ from that table. The quantity $\max\,(1/\prod\alpha\delta)$ takes into account the contribution of the small divisors and of the numbers δ_r in the estimate of Poisson brackets. The latter divisors seem to introduce another possible source of divergence, analogous to that of the small divisors. However, it is an easy matter to see that the contribution of the δ's can be controlled exactly with the same mechanism which works for the α's. The relevant remark here is that the solution of the equation for a generating function introduces a denominator α_r which estimates the small divisors. On the other hand, the generating functions contributes to the transformed Hamiltonian only via Poisson brackets, the estimate of which introduces a denoninator δ_r. Thus, the denominators α_r and δ_r always appear in pairs $\alpha_r\delta_r$ (this argument seems to fail for $\xi\cdot q$, but remark that the Poisson bracket in this case is just a derivative with respect to p, which does not involve any denominator δ in the estimate). If we choose $\delta_r = B/r^2$, then the condition $\sum_r \delta_r < 1/2$ is easily satisfied. On the other hand, in view of the diophantine condition (2), we have $\alpha_r\delta_r \sim r^{\tau+2}$. Thus, only the exponent of r is changed, which does not invalidate our argument at the end of sect. 3.2 which proves that the contribution of small divisors is controlled geometrically. Finally, the factor μ_r takes into account the number of Poisson brackets generated during the process of construction of normal form, and needs to be estimated.

In order to estimate μ_r let us refer to the explicit formulæ giving the formal algorithm, which are reported in sect. 4.1. It is actually necessary to estimate the number of Poisson bracket in every fuction which appears in the algorithm. We give the result, which requires only elementary considerations. We have $\mu_r = \nu_{r-1,r}$, where the ν's are recursively defined for $s, r \geq 0$ and $k = \lfloor s/r \rfloor$ by

$$\begin{aligned} \nu_{0,s} &= 1 \\ \hat{\nu}_{r,s} &= \sum_{j=0}^{\min(k,2)} \nu_{r-1,r}^{j}\nu_{r-1,s-jr} \\ \nu_{r,s} &= \sum_{j=0}^{k} \nu_{r-1,r}^{j}\hat{\nu}_{r,s-jr}\ ; \end{aligned} \tag{17}$$

here, $\hat{\nu}_{r,s}$ is an auxiliary sequence which estimates the number of Poisson brackets in $\hat{H}^{(r)}$. Studying this sequence one obtains, e.g., $\nu_{r-1,r} \leq \nu_{r,r} \leq 24^r$.

Thus, all contribution to the size of the generating functions are estimated geometrically, i.e., one has estimates of the type

$$\|\chi_r^{(1)}\|\ ,\quad \|\chi_r^{(2)}\| \leq b^{r-1}G$$

with some positive constants b and G. According to the general theory of Lie transforms, this is enough in order to prove the convergence of the sequence of transformations leading to Kolmogorov's normal form.

5. Some comments on the relation with Lindstedt's method

It is now interesting to investigate the relation between the present approach and the similar method of computing Lindstedt's series. we shall make a short comparison concerning two main points: (i) the effectiveness of the algorithms, and (ii) the control of small denominators. In order to establish a well definite basis, we shall refer in particular to the algorithm discussed in Gallavotti's paper [7]. We report the formulæ, using the same notations as the author. The Hamiltonian is written in the form

$$H = \frac{1}{2} J^{-1} A \cdot A + \varepsilon f(\alpha) \ ,$$

where $A_1, \ldots, A_l, \alpha_1, \ldots, \alpha_l$ are action–angle variables, J is the diagonal matrix of the moments of inertia, and f is a trigonometric polynomial of degree N which is assumed to be even in the angles, i.e., $f(\alpha) = f(-\alpha)$. The equations of the torus with frequency ω_0 are

$$A = A_0 + \sum_{k\ge 0} \varepsilon^k H^{(k)}(\omega_0 t) \ , \quad \alpha = \omega_0 t + \sum_{k\ge 0} \varepsilon^k h^{(k)}(\omega_0 t) \ , \quad A_0 = J\omega_0$$

where

$$h^{(1)}(\psi) = -\sum_{\nu\ne 0} \frac{iJ^{-1}\nu}{(i\omega_0\cdot\nu)^2} f_\nu e^{i\nu\cdot\psi} \ , \quad \psi = \omega_0 t$$

$$\omega_0 \cdot \partial H_j^{(k)} = -\sum_{\substack{m_1,\ldots,m_l \\ |m|\ge 0}} \frac{1}{\prod_{s=1}^{l} m_s!} \left[\partial_{\alpha_j} \partial_{\alpha_1\ldots\alpha_l}^{m_1+\ldots+m_l} f(\omega_0 t)\right] \sum{}^{*} \prod_{s=1}^{l} \prod_{j=1}^{m_s} h_s^{(k_j^s)}(\omega_0 t) \tag{18}$$

$$\omega_0 \cdot \partial h_j^{(k)} = J_j^{-1} H_j^{(k)} \ ,$$

where $\sum^*$ is estended to all $k_j^s \ge 1$ so that $\sum_{s=1}^{l}\sum_{j=1}^{m_s} k_j^s = k-1$. From this formula is immediately clear that $h^{(k)}, H^{(k)}$ are trigonometric polynomials of degree kN (this corresponds to rK in the previous sections). On the other hand, it is also evident that every Fourier mode has a coefficient of the form $\sum \frac{\beta}{\prod \nu\cdot\omega}$, similar to the form of the coefficients in the algorithm of the preceeding sections. Thus, the two algorithms produce similar expansions. However, if one keeps all the coefficients as they are generated without performing any algebraic simplification the expansions are actually different. They coincide only after summation of all coefficients of the same Fourier mode.

Concerning the effectiveness of the algorithms, it is immediate to remark that if one is interested only on constructing the equation of the invariant torus then the method of Lindtsedt is much more efficient. Indeed, it does not require determining generating functions and Hamiltonians at every step: the construction is very direct. Moreover, if one performs explicitly the first steps of Kolmogorov's algorithm one sees that the number of coefficients generated is much higher. In this respect, the only advantage of Kolmogorov's algorithm is that it gives information also on the neighbourhood of the invariant torus. But, as far as the equation of the torus is concerned, Lindstedt's algorithm seems to be in a definitely better position.

Concerning the accumulation of small divisors the main difference is that the coefficients produced by Lindstedt's algorithm do not satisfy the selection rule of sect. 4.3, namely that in $h^{(r)}$ the denominator of the expression $\sum \frac{\beta}{\prod \nu\cdot\omega}$ should satisfy $\sum \log_2 \lceil |\nu|/N \rceil \le 2r-2$. Indeed, only the weaker rule $\sum \log_2 \lceil |\nu|/N \rceil \le 2\log_2 r!$ can be proved. Proving the latter inequality is an easy matter (do it by induction). The proof that the former selection rule is violated follows from considering the term $\partial_{\alpha_j} \partial^{m_1+\ldots+m_l}_{\alpha_1\ldots\alpha_l} f(\omega t) h^{(k)}$, which clearly appears in (18). Indeed, assume that in $h^{(r-1)}$ there is a Fourier mode ν with $\lceil |\nu|/N \rceil = r-1$ and with a coefficient satisfying $\sum \log_2 \lceil |\nu|/N \rceil \le 2(r-1)-2$. For $r=2$ such a coefficient exists, and the argument which follows implies that it must exist also for $r>2$. Then the expression above contains all Fourier modes $\nu+\nu'$, where ν' is any Fourier mode of f, with the same coefficient, since derivatives do not modify the denominators. In general, at least one of these modes has $\lceil |\nu+\nu'|/N \rceil = r$, and solving eq. (18) one adds two more denominators with ν with $\lceil |\nu|/N \rceil = r$. This produces a coefficient in $h^{(r)}$ for which $\sum \log_2 \lceil |\nu|/N \rceil = 2(r-1) - 2 + 2\log_2 r > 2r-2$, unless $r=2$. Thus, the selection rule is violated, and the simple argument of sect. 3.2 for the control of small denominators does not apply. Actually, as we learned from Gallavotti's lecture, it is possible to group together some coefficients in Lindstedt's expansion so that the global effect of the small denominators is dominated geometrically. Such a procedure, however, is not necessary in Kolmogorov's algorithm.

References

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

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

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

[5] Eliasson, L. H.: *Generalization of an estimate of small divisors by Siegel*, in P. Rabinowitz and E. Zehnder eds., book in honor of J. Moser, Academic Press (1990).

[6] 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).

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

[8] Chierchia, L. and Falcolini, C.: *A direct proof of a theorem by Kolmogorov in Hamiltonian systems*, preprint (1993).

[9] Gallavotti, G. and Gentile, G.: *Nonrecursive proof of the KAM theorem*, preprint (1994).

[10] Gentile, G. and Mastropietro, V.: *Tree expansion and multiscale analysis for KAM tori*, preprint (1994).

[11] Giorgilli, A. and Locatelli, U.: *Kolmogorov theorem and classical perturbation theory*, preprint (1995).

[12] 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).

[13] Rüssmann, H.: *Über die Normalform analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung*, Math. Annalen, **169**, 55–72 (1967).