METHODS OF COMPLEX ANALYSIS IN CLASSICAL PERTURBATION THEORY

by

Antonio Giorgilli

Dipartimento di Matematica dell'Università di Milano
Via Saldini 50, 20133 MILANO, Italy

D.Benest, C.Froeschlé (eds.), Chaos and Diffusion (Cham94), 1-33

1. Overview

At the beginning of the second volume of his *Méthodes nouvelles de la Mécanique Céleste* Poincaré devoted the chapter VIII to the problem of the reliability of the formal expansions of perturbation theory. In particular, he pointed out that these series could have the same character of Stirling's series. Recent work in perturbation theory has enlighten this conjecture of Poincaré, bringing into evidence that the series of perturbation theory, although non convergent in general, furnish nevertheless valuable approximations to the true orbits for a very large time, which in some practical cases could be comparable with the age of the universe.

The aim of my lectures is to introduce the quantitative methods of perturbation theory which allow to obtain such powerful results. In particular, I will discuss here the theorem of Nekhoroshev on stability over exponentially large times.

I will proceed in two steps. First of all, I will pay particular attention to the use of Lie transforms as an useful tool for constructing canonical transformations, and to the use of a few ideas from the theory of analytic functions in order to produce quantitative estimates on series arising from perturbation theory. These methods are the common ingredient of all perturbation schemes. I will try to present all these technical tools at an elementary level, using only a few informations coming from the theory of analytic functions. It could be said that almost everything is based only on Cauchy's estimates for the derivatives of an analytic function and on Weierstrass theorem on analytic functions defined as sums of infinite series. My hope is that the present text will be an useful introduction for people that are perhaps familiar with the formal expansions of Celestial Mechanics, but find a bit hard the step of making quantitative estimates.

The second step is a discussion of the theorem of Nekhoroshev in a context general enough. This involves on the one hand the proof of convergence of the normal form, which is treated with methods of complex analysis, and, on the other hand, the so called geometric part of the theorem, namely, a clever analysis of the geometry of resonances in phase space. As a matter of fact, the latter makes very little use of the theory of analytic functions (which should be the main argument of my lectures, according to the title). Indeed, it is rather based on the one hand on geometry and number theory, and, on the other hand, on a process of optimization of parameters which is commonly used in dealing with asymptotic expansions. The combination of geometric results and optimization leads to a stability estimate over exponentially large times. This result is in fact a full confirmation of Poincaré's suggestion that the series of Celestial Mechanics have an asymptotic character. This justifies my choice of including also an almost complete scheme of proof of Nekhoroshev's theorem.

The text is organised in three sections. Sect. 2 introduces general tools (Lie transforms, norms, generalized Cauchy's estimates, convergence of canonical transformation defined through Lie transforms) that can be used in a wide number of problems. Sect. 3 is devoted to normal form theory, with quantitative estimates. The conclusion is a local stability result in an apparently useless form; however, it is the basis for the global estimate of Nekhoroshev. Sect. 4 is devoted to the so called geometric part of Nekhoroshev's theorem and to the optimization process referred to above. The conclusion is the statement of Nekhoroshev's theorem for an apparently particular class of Hamiltonian systems. However, I emphasize that extending the result to a completely general class is just a technical matter: the case discussed here contains all the difficulties of the general problem.

Sections 2 and 3 are developed in full detail, and should be self contained. Sect. 4 instead contains only a formal statement of the main results and a short discussion of the trace of proof.

2. Lie series and Lie transforms

The use of Lie series and Lie transforms is quite recent in Celestial Mechanics, and is generally known at a formal level. I give here a quantitative formulation of the theory.

2.1. *Formal definitions*

On a $2n$–dimensional phase space endowed with canonical coordinates p, q, consider an analytic function $\chi(p, q)$, which will be called a *generating function.* The *Lie derivative* L_χ is a linear differential operator acting on functions on the phase space, defined as

$$L_\chi \cdot = \{\chi, \cdot\} \ . \tag{1}$$

This is nothing but the derivative along the Hamiltonian field generated by χ. The *Lie series* operator is then defined as the exponential of L_χ, namely

$$\exp(L_\chi) = \sum_{s\geq 0} \frac{1}{s!} L_\chi^s \tag{2}$$

(see Gröbner, 1960). This operator represents the time one evolution of the canonical flow generated by the autonomous Hamiltonian χ.

As an example, consider the case of action–angle variables $p \in \mathbf{R}^n$ and $q \in \mathbf{T}^n$. The function $\chi = \sum_l \xi_l q_l$, with $\xi \in \mathbf{R}^n$ generates through the Lie series operator the canonical transformation $p' = \exp(L_\chi)p = p + \xi$, $q' = \exp(L_\chi)q = q$, namely a translation in the action space. Similarly, a generating function $\chi = \chi(q)$ (independent of p) generates the transformation $p' = p + \frac{\partial\chi}{\partial q}$, $q' = q$, namely a deformation of the action variables.

The *Lie transform* differs from the Lie series in that it is connected with the flow of a *nonautonomous* (i.e., time dependent) Hamiltonian. In such a form, it has been introduced by Hori (1966) and Deprit (1969). In fact, several algorithms have been devised in order to give the Lie transform an algorithmic recursive form: everybody, of course, has his favourite one. To make a definite choice, I will make use here of my favourite one, which is related to the "algorithm of the inverse", introduced by Henrard. Consider a *generating sequence* $\chi = \{\chi_s\}_{s\geq 1}$ of analytic functions on the phase space. The Lie transform operator T_χ is defined as

$$T_\chi = \sum_{s\geq 0} E_s \ , \tag{3}$$

where the sequence $\{E_s\}_{s\geq 0}$ of operators is recursively defined as

$$E_0 = \mathrm{Id} \ , \quad E_s = \sum_{j=1}^{s} \frac{j}{s} L_{\chi_j} E_{s-j} \ . \tag{4}$$

As a simple example, consider the case $\chi = \{\chi_1, 0, 0, \ldots\}$, namely a generating sequence containing only the first term. Then the Lie transform generated by χ coincides with the Lie series generated by χ_1, i.e.,

$$T_\chi = \exp(L_{\chi_1}) \ .$$

From now on, it will be useful to look at the algebraic aspect of the definitions of Lie series and Lie transform, rather than to their property of being related to the canonical flow of some Hamiltonian. In particular, all the properties stated here can be proved on a purely algebraic basis, without any reference to the canonical flow (see for instance Giorgilli and Galgani, 1978). The relevant properties are the following: both the Lie series and the Lie transform operators are linear, and preserve products and Poisson brackets, namely

$$T_\chi(fg) = (T_\chi f)(T_\chi g)\ , \quad T_\chi\{f,g\} = \{T_\chi f, T_\chi g\}\ ,$$

with analogous formulæ for the Lie series. Moreover, both the operators are invertible. Finding the inverse of the Lie series is an easy matter: recalling that it represents the flow of an autonomous canonical system, one immediately concludes that the inverse of $\exp(L_\chi)$ is $\exp(L_{-\chi}) = \exp(-L_\chi)$, i.e., the generating function of the inverse is $-\chi$. Concerning the Lie transform, a careful analysis of the algorithm(3)–(4) shows that finding the inverse transformation is an easy matter, although finding the generating sequence of the inverse is not as easy. Indeed, assume we are given a function $f = f_0 + f_1 + \ldots$, and denote by $g = g_0 + g_1 + \ldots$ its transformed function $g = T_\chi f$; using the linearity of the Lie transform, we can apply T_χ separately to every term of f. It is useful to rearrange terms according to the triangular diagram

$$\begin{array}{llllll}
g_0 & f_0 & & & & \\
 & \downarrow & & & & \\
g_1 & E_1 f_0 & f_1 & & & \\
 & \downarrow & \downarrow & & & \\
g_2 & E_2 f_0 & E_1 f_1 & f_2 & & \\
 & \downarrow & \downarrow & \downarrow & & \\
g_3 & E_3 f_0 & E_2 f_1 & E_1 f_2 & f_3 & \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array} \tag{5}$$

where terms of the same order appear on the same line. Remark that the operator T_χ acts *by columns*, as indicated by the arrows: the knowledge of f_j and of the generating sequence allows one to construct the whole column below f_j. Thus, the first line gives $g_0 = f_0$, the second line gives $g_1 = E_1 f_0 + f_1$, and so on. This shows how to practically perform the transformation. Concerning the inverse, assume that g is given, and f is unknown. Then, the first line gives $f_0 = g_0$; having determined f_0, all the column below f_0 can be constructed, and the second line gives immediately $f_1 = g_1 - E_1 f_0$; having determined f_1, all the corresponding column can be constructed, so that f_2 can be determined from the third line as $f_2 = g_2 - E_2 f_0 - E_1 f_1$, and so on. There is also an explicit formula for the inverse, namely

$$T_\chi^{-1} = \sum_{s\geq 0} D_s \ ,$$

where

$$D_0 = \mathrm{Id} \ , \quad D_s = -\sum_{j=1}^{s} \frac{j}{s} D_{s-j} L_{\chi_j} \ .$$

However, this expression is actually useless for a practical computation: the algorithm described above is much more efficient. Nevertheless, the explicit recursive formula is useful for quantitative estimates.

The property of preserving the Poisson brackets is relevant, as it means that the coordinate transformation

$$p = T_\chi p' \ , \quad q = T_\chi q' \ ,$$

where the generating sequence χ depends on the "new" variables p', q', is canonical (I omit here the analogous statement concerning the Lie series). Now, if we consider a function $f(p,q)$ on the phase space, we should in principle compute the transformed function, $f'(p',q')$ say, as

$$f'(p',q') = f(p,q)\big|_{p=T_\chi p',q=T_\chi q'} \ ,$$

or, equivalently,

$$f'(p',q') = f(T_\chi p', T_\chi q') \ ,$$

namely by substitution of the coordinate transformation. Now, the properties of being linear and of preserving products can be used to prove that in fact one has

$$f(T_\chi p', T_\chi q') = (T_\chi f)\,(p',q') \ . \tag{6}$$

Briefly, "don't try to make substitutions: just transform the function, and change the name of the variables".

2.2. Cauchy estimates

Consider an open disk $\Delta_\varrho(0)$ centered at the origin of the complex plane $\mathbf{C}$. Consider a function f analytic and bounded on the disk $\Delta_\varrho(0)$. The *supremum norm* $|f|_\varrho$ of f in the domain $\Delta_\varrho(0)$ is defined as

$$|f|_\varrho = \sup_{z\in\Delta_\varrho(0)} |f(z)| \ . \tag{7}$$

Cauchy's estimate for the derivative f' of f at the origin states that

$$|f'(0)| \le \frac{1}{\varrho}|f|_\varrho \ .$$

More generally, for the s–th derivative $f^{(s)}$ one has the estimate

$$|f^{(s)}(0)| \le \frac{s!}{\varrho^s}|f|_\varrho \ .$$

For instance, let $\varrho = 1$, and consider the function $f(z) = z^s$; it is an easy matter to check that $|f|_1 = 1$, so that Cauchy estimate gives $|f^{(s)}(0)| \le s!$. This shows that the estimate cannot be improved in general. The proof of the inequalities above is an easy consequence of Cauchy formula

$$f^{(n)}(z) = \frac{n!}{2\pi i}\oint \frac{f(\zeta)}{(\zeta - z)^{n+1}}\mathrm{d}z \ .$$

In the case of n variables the extension is quite straightforward. Let the domain $\Delta_\varrho(0)$ be the polydisk of radius ϱ centered at the origin of $\mathbf{C}^n$, namely

$$\Delta_\varrho(0) = \{z \in \mathbf{C}^n \ : \ |z| < \varrho\} \ , \tag{8}$$

where $|z| = \max_j |z_j|$ is the l_∞ norm on $\mathbf{C}^n$. This is nothing but the cartesian product of complex disks of radius ϱ in the complex plane. Define the supremum norm of an analytic function f as above. Then Cauchy estimate reads

$$\left|\frac{\partial f}{\partial z_j}(0)\right| \le \frac{1}{\varrho}|f|_\varrho \ , \quad 1 \le j \le n \ .$$

I omit here the estimates for higher order derivatives, since we shall not need them.

2.3. Analytic framework

Consider the common case of a phase space $\mathcal{G}\times\mathbf{T}^n$, where $\mathcal{G}\subset\mathbf{R}^n$, endowed with action–angle variables $q\in\mathbf{T}^n$ and $p\in\mathcal{G}$.

In order to use Cauchy estimates, it is necessary to introduce a *complexification* of the domains. For $p\in\mathcal{G}$ consider the complex polydisk $\Delta_\varrho(p)$ of radius $\varrho>0$ with center p defined as in (8). The complexification $\mathcal{G}_\varrho$ of the real domain $\mathcal{G}$ is defined as

$$\mathcal{G}_\varrho=\bigcup_{p\in\mathcal{G}}\Delta_\varrho(p)\ . \tag{9}$$

Similarly, the complexification $\mathbf{T}^n_\sigma$ for $\sigma>0$ of the n–torus is defined as

$$\mathbf{T}^n_\sigma=\{q\in\mathbf{C}^n\ :\ |\Im q|<\sigma\}\ , \tag{10}$$

namely the cartesian product of strips of width σ around the real axis on the complex plane. From now on, the phase space will be the domain

$$\mathcal{D}_{(\varrho,\sigma)}=\mathcal{G}_\varrho\times\mathbf{T}^n_\sigma\ .$$

Introducing a norm on the space of analytic functions on the domain $\mathcal{D}_{(\varrho,\sigma)}$ is now straightforward, in view of the previous section. For instance, for a function f analytic on $\mathcal{G}_\varrho$ the supremum norm is

$$|f|_\varrho=\sup_{p\in\mathcal{G}_\varrho}|f(p)|\ . \tag{11}$$

Similarly, for a function g analytic on $\mathcal{D}_{(\varrho,\sigma)}$ the supremum norm $|g|_{(\varrho,\sigma)}$ is

$$|g|_{(\varrho,\sigma)}=\sup_{(p,q)\in\mathcal{D}_{(\varrho,\sigma)}}|g(p,q)|\ . \tag{12}$$

The use of Cauchy estimates goes as follows. Consider for simplicity a function f analytic and bounded on the complex domain $\mathcal{G}_\varrho$. Consider now a point $p\in\mathcal{G}_{\varrho-\delta}$ for $0<\delta\le\varrho$ (that is, the union of polydisks of radius $\varrho-\delta$ centered at every point of $\mathcal{G}$). Remark that the polydisk $\Delta_\delta(p)$ is a subset of $\mathcal{G}_\varrho$, so that f is analytic and bounded on $\Delta_\delta(p)$, and moreover we have the estimate $|f(p')|\le|f|_\varrho$ for all $p'\in\Delta_\delta(p)$. By Cauchy estimate, we immediately get

$$\left|\frac{\partial f}{\partial p_j}(p)\right|\le\frac{1}{\delta}|f|_\varrho\ ,\quad 1\le j\le n\ .$$

Since this is true for every point $p \in \mathcal{G}_{\varrho-\delta}$, we conclude

$$\left|\frac{\partial f}{\partial p_j}\right|_{\varrho-\delta} \le \frac{1}{\delta}|f|_{\varrho}\ , \quad 1 \le j \le n\ .$$

The relevant fact here is that we have an explicit bound on the derivative of a function, but we must pay this information with a restriction of the complex domain. Analogous formulæ could be obtained for the derivatives with respect to q of a function $f(p,q)$, but we shall not need them here.

2.4. *Weighted Fourier norms*

As a matter of fact, the supremum norm is not the best one, up to my knowledge, when dealing with small denominators. A more useful norm is the so called *weighted Fourier norm*, which can be introduced on the basis of the following considerations. It is known that an analytic function f on the domain $\mathcal{D}_{(\varrho,\sigma')}$ admits the Fourier expansion

$$f(p,q) = \sum_{k\in\mathbf{Z}^n} f_k(p) e^{ik\cdot q}\ .$$

It is easy to show that the coefficients $f_k(p)$ of the Fourier expansion decay exponentially with k. More precisely, one has

$$|f_k|_{\varrho} \le |f|_{(\varrho,\sigma')} e^{-|k|\sigma'}\ ,$$

where $|k| = |k_1| + \cdots + |k_n|$ (the l_1 norm in $\mathbf{Z}^n$). To prove this, consider the simple case of a function $f(q)$ of one variable $q \in \mathbf{T}$, and recall that the coefficients are computed as

$$f_k = \frac{1}{2\pi}\int_0^{2\pi} f(q) e^{-ikq} \mathrm{d}q\ .$$

If k is positive, move the contour of integration by $i\sigma$ and get

$$f_k = \frac{1}{2\pi}\int_0^{2\pi} f(q+i\sigma) e^{-ikq} e^{-k\sigma} \mathrm{d}q\ .$$

The inequality then follows in view of $|f(q+i\sigma)| \le |f|_{\sigma}$ (by definition of supremum norm) and $|e^{-ikq}| = 1$. For negative k, just move the contour by $-i\sigma$. In the case $q \in \mathbf{T}^n$ the coefficients f_k, $k \in \mathbf{Z}^n$ are defined as

$$f_k = \frac{1}{(2\pi)^n}\int_0^{2\pi} \cdots \int_0^{2\pi} f(q) e^{-ik\cdot q} \mathrm{d}q_1\, \mathrm{d}q_n\ ,$$

and a similar argument applies. In the general case $f(p,q)$ just repeat the same argument for every $p \in \mathcal{G}_\varrho$, and the claim is proved.

The exponential decay of coefficients suggests to define the weighted Fourier norm $\|f\|_{(\varrho,\sigma)}$ for $\sigma < \sigma'$ as

$$\|f\|_{(\varrho,\sigma)} = \sum_{k\in\mathbf{Z}^n} |f_k|_\varrho e^{|k|\sigma} \ . \tag{13}$$

The condition $\sigma < \sigma'$ ensures that the series defining the norm converges. It is also an easy matter to check that the supremum norm is bounded by the weighted Fourier norm, i.e.,

$$|f|_{(\varrho,\sigma)} \le \|f\|_{(\varrho,\sigma)} \ .$$

2.5. *Generalized Cauchy estimates*

The next step is concerned with the generalization of Cauchy estimates, taking into account that we are interested in Lie derivatives (or Poisson brackets), and that we are going to use the weighted Fourier norm instead of the supremum one.

A first result concerns the Poisson bracket between two functions, and the derivative of a function with respect to one of the canonical coordinates (i.e., the Poisson bracket with the conjugate one).

Lemma : *Let f be analytic on the domain $\mathcal{D}_{(\varrho,\sigma)}$, and g be analytic in $\mathcal{D}_{(1-d')(\varrho,\sigma)}$ for some $0 \le d' < 1$; assume moreover that $\|f\|_{(\varrho,\sigma)}$ and $\|g\|_{(1-d')(\varrho,\sigma)}$ are finite. Then:*

i. for $0 < d < 1$ and for $1 \le j \le n$ one has

$$\left\|\frac{\partial f}{\partial p_j}\right\|_{(1-d)(\varrho,\sigma)} \le \frac{1}{d\varrho}\|f\|_{(\varrho,\sigma)} \ , \qquad \left\|\frac{\partial f}{\partial q_j}\right\|_{(1-d)(\varrho,\sigma)} \le \frac{1}{ed\sigma}\|f\|_{(\varrho,\sigma)} \ ; \tag{14}$$

ii. for $0 < d < 1 - d'$ one has

$$\|\{f,g\}\|_{(1-d'-d)(\varrho,\sigma)} \le \frac{2}{ed(d+d')\varrho\sigma}\|f\|_{(\varrho,\sigma)}\|g\|_{(1-d')(\varrho,\sigma)} \ . \tag{15}$$

Proof. Using the Fourier expansion, compute

$$\frac{\partial f}{\partial p_j} = \sum_k \frac{\partial f_k}{\partial p_j}(p)e^{ik\cdot q} \ .$$

According to the definition of the norm, we have

$$\left\| \frac{\partial f}{\partial p_j} \right\|_{(1-d)(\varrho,\sigma)} = \sum_k \left| \frac{\partial f_k}{\partial p_j} \right|_{(1-d)\varrho} e^{(1-d)|k|\sigma} \ ;$$

by Cauchy estimate, and forgetting d in the exponent, the right hand side of the latter expression is bounded by

$$\frac{1}{d\varrho} \sum_k |f_k|_\varrho e^{|k|\sigma} \ ,$$

and the first of (14) follows in view of the definition of the norm. Coming to the second one, compute

$$\frac{\partial f}{\partial q_j} = i \sum_k k_j f_k(p) e^{ik\cdot q} \ ,$$

and use again the definition of the norm to estimate

$$\begin{aligned} \left\| \frac{\partial f}{\partial q_j} \right\|_{(1-d)(\varrho,\sigma)} &= \sum_k |k_j|\, |f_k|_{(1-d)\varrho} e^{(1-d)|k|\sigma} \\ &\le \sum_k |f_k|_\varrho e^{|k|\sigma} |k| e^{-d|k|\sigma} \ . \end{aligned}$$

By the general inequality

$$x^\alpha e^{-\delta x} \le \left(\frac{\alpha}{e\delta}\right)^\alpha \qquad \text{for positive } \alpha,\ x,\ \delta \tag{16}$$

(just compute the supremum over x of $x^\alpha e^{-\delta x}$), with $\alpha = 1$ and $|k|$ and $d\sigma$ in place of x and δ respectively, one gets $|k|e^{-d|k|\sigma} \le 1/(ed\sigma)$; the second of (14) then follows in view of the definition of the norm. The proof of (15) requires some more work. Compute

$$\{f, g\} = i \sum_{k,k'} \left[\sum_{l=1}^n \left(k_l \frac{\partial g_{k'}}{\partial p_l} f_k + k'_l \frac{\partial f_k}{\partial p_l} g_{k'} \right) \right] e^{ik\cdot q} e^{ik'\cdot q} \ ,$$

and use the definition of the norm to estimate

$$\|\{f,g\}\|_{(1-d'-d)(\varrho,\sigma)} < \sum_{k,k'}\Biggl[\sum_{l=1}^{n}\Bigl(|k_l|\left|\frac{\partial g_{k'}}{\partial p_l}\right|_{(1-d'-d)\varrho}|f_k|_\varrho + |k'_l|\left|\frac{\partial f_{k'}}{\partial p_l}\right|_{(1-d'-d)\varrho}|g_k|_{(1-d')\varrho}\Bigr)\Biggr]e^{(1-d'-d)|k+k'|\sigma}\ .$$

By Cauchy estimates, the right hand side of the latter expression is found to be smaller than

$$\frac{1}{d\varrho}\sum_{k,k'}|g'_k|_{(1-d')\varrho}e^{(1-d')|k'|\sigma}|f_k|_\varrho e^{|k|\sigma}\sum_{l=1}^{n}|k_l|e^{-(d'+d)|k|\sigma} + \frac{1}{(d'+d)\varrho}\sum_{k,k'}|g_{k'}|_{(1-d')\varrho}e^{(1-d')|k'|\sigma}|f_k|_\varrho e^{|k|\sigma}\sum_{l=1}^{n}|k'_l|e^{-d|k'|\sigma}\ .$$

Remark now that the only terms depending on the index l are $\sum_l |k_l| = |k|$ and $\sum_l |k'_l| = |k'|$; moreover, using the inequality (16) get

$$|k|e^{-(d'+d)|k|\sigma} \le \frac{1}{(d'+d)e\sigma}\ ,\quad |k'|e^{-d|k'|\sigma} \le \frac{1}{de\sigma}\ .$$

Reordering terms, the expression above is found to be smaller than

$$\frac{2}{ed(d'+d)\varrho\sigma}\sum_{k'}|g'_k|_{(1-d')\varrho}e^{(1-d')|k'|\sigma}\cdot\sum_{k}|f_k|_\varrho e^{|k|\sigma}\ ,$$

and the conclusion follows in view of the definition of the norms $\|f\|_{(\varrho,\sigma)}$ and $\|g\|_{(1-d')(\varrho,\sigma)}$. *Q.E.D.*

The next step concerns the estimate of multiple Poisson brackets. Consider an expression of the form $L_{g_s}\circ\ldots\circ L_{g_1}f$. We can estimate it by iterating s times the estimate of lemma 2.1. However, we must take into account the restriction of the domains. To this end, having fixed the final restriction d, we choose s positive quantities $\delta_1,\ldots,\delta_s$ the sum of which is d, and estimate $L^j_\chi f$ on the domain restricted by the factor $1-\delta_1-\ldots-\delta_j$. The simplest choice is $\delta_1=\ldots=\delta_s=d/s$. The final result is given by

Lemma 1: *Let $g_1,\ldots,g_s$ and f be analytic on the domain $\mathcal{D}_{(\varrho,\sigma)}$, with finite norms and s positive. Then for every positive $d<1$ one has*

$$\|L_{g_s}\circ\ldots\circ L_{g_1}f\|_{(1-d)(\varrho,\sigma)} \le \frac{s!}{e^2}\left(\frac{2e}{d^2\varrho\sigma}\right)^s\|g_1\|_{(\varrho,\sigma)}\cdot\ldots\cdot\|g_s\|_{(\varrho,\sigma)}\|f\|_{(\varrho,\sigma)}\ . \tag{17}$$

Remark. The operators L_{g_j} do not generally commute. However, the estimate (17) actually does not depend on the order: with a little abuse, we could say that the estimates do commute. Furthermore, the coefficient in front of the estimate depends only on how many operators L_{g_j} are applied to f. Thus, the lemma could be reformulated in a recursive form as follows: let $\vartheta_0 = f$, and ϑ_s, for $s > 1$, be the result of the application of the s operators $L_{g_1}, \ldots, L_{g_s}$ to f in any order (e.g., $\vartheta_s = L_{g_1} \circ \ldots \circ L_{g_s} f$ or $\vartheta_s = L_{g_s} \circ \ldots \circ L_{g_1} f$). Then

$$\|\vartheta_s\|_{(1-d)(\varrho,\sigma)} \le \frac{2se}{d^2 \varrho\sigma} \|g_s\|_{(1-d)(\varrho,\sigma)} \|\vartheta_{s-1}\|_{(1-d)(\varrho,\sigma)} \ .$$

Remark also that both ϑ_s and ϑ_{s-1} are estimated on the same domain $(1-d)(\varrho,\sigma)$: no further restriction is needed.

Proof of lemma 2. For $s = 1$ this is just lemma 1. For $s > 1$ let $\delta = d/s$. A straightforward application of lemma 1 gives

$$\|L_{g_1} f\|_{(1-\delta)(\varrho,\sigma)} \le \frac{2}{e\delta^2 \varrho\sigma} \|g_1\|_{(\varrho,\sigma)} \|f\|_{(\varrho,\sigma)} \ .$$

For $j = 2, \ldots, s$, recalling that $L_{g_j} \circ \ldots \circ L_{g_1} f$ can be estimated on the domain restricted by the factor $1 - j\delta$, use again lemma 1 with δ in place of d and $(j-1)\delta$ in place of d', and get the recursive estimate

$$\begin{aligned} &\bigl\|L_{g_j} \circ \ldots \circ L_{g_1} f\bigr\|_{(1-j\delta)(\varrho,\sigma)} \\ &\qquad \le \frac{2}{ej\delta^2 \varrho\sigma} \|g_j\|_{(\varrho,\sigma)} \|L_{g_{j-1}} \circ \ldots \circ L_{g_1} f\|_{(1-(j-1)\delta)(\varrho,\sigma)} \ . \end{aligned}$$

By recursive application of this formula up to s, and recalling $\delta = d/s$ we get

$$\begin{aligned} &\|L_{g_s} \circ \ldots \circ L_{g_1} f\|_{(1-d)(\varrho,\sigma)} \\ &\qquad \le \left(\frac{2}{ed^2 \varrho\sigma}\right)^s \frac{s^{2s}}{s!} \|g_1\|_{(\varrho,\sigma)} \cdot \ldots \cdot \|g_s\|_{(\varrho,\sigma)} \|f\|_{(\varrho,\sigma)} \ . \end{aligned}$$

Then, (17) follows in view of the trivial inequality $s^s \le e^{s-1} s!$ for $s \ge 1$ (prove it by induction). *Q.E.D.*

2.6. Analyticity of Lie series and of Lie transform

Lie series and Lie transform have been defined as series of analytic functions. It is natural to ask whether or not the sums of these series are analytic functions on some domain. Sufficient conditions can be found on the basis of the results of the previous section.

Proposition 1: *Let f and the generating sequence $\chi = \{\chi_s\}_{s\geq 1}$ be analytic on the domain $\mathcal{D}_{(\varrho,\sigma)}$, and assume that $\|f\|_{(\varrho,\sigma)}$ is finite, and that*

$$\|\chi_s\|_{(\varrho,\sigma)} \leq \frac{b^{s-1}}{s} G \tag{18}$$

for some $b \geq 0$ and $G > 0$. Then for every positive $d < 1/2$ the following statement holds true: if the condition

$$\frac{2eG}{d^2\varrho\sigma} + b \leq \frac{1}{2} \tag{19}$$

is satisfied, then the operator T_χ and its inverse T_χ^{-1} define an analytic canonical transformation on the domain $\mathcal{D}_{(1-d)(\varrho,\sigma)}$ with the properties

$$\begin{aligned} \mathcal{D}_{(1-2d)(\varrho,\sigma)} &\subset T_\chi \mathcal{D}_{(1-d)(\varrho,\sigma)} \subset \mathcal{D}_{(\varrho,\sigma)} \ , \\ \mathcal{D}_{(1-2d)(\varrho,\sigma)} &\subset T_\chi^{-1} \mathcal{D}_{(1-d)(\varrho,\sigma)} \subset \mathcal{D}_{(\varrho,\sigma)} \ . \end{aligned}$$

The proof of the proposition is based on the following

Lemma : *With the hypotheses of proposition 1 the series $T_\chi f$, $T_\chi^{-1} f$, $T_\chi p$, $T_\chi^{-1} p$, $T_\chi q$ and $T_\chi^{-1} q$ are absolutely convergent on $\mathcal{D}_{(1-d)(\varrho,\sigma)}$, and for integer $r > 0$ one has the estimates*

$$\begin{aligned} \|T_\chi p - p\|_{(1-d)(\varrho,\sigma)} &\leq d\varrho \ , \\ \|T_\chi q - q\|_{(1-d)(\varrho,\sigma)} &\leq d\sigma \ , \\ \left\|T_\chi^{-1} p - p\right\|_{(1-d)(\varrho,\sigma)} &\leq d\varrho \ , \\ \left\|T_\chi^{-1} q - q\right\|_{(1-d)(\varrho,\sigma)} &\leq d\sigma \ , \\ \left\|T_\chi f - \sum_{s=0}^{r} E_s f\right\|_{(1-d)(\varrho,\sigma)} &\leq \frac{2}{e^2}\left(\frac{2eG}{d^2\varrho\sigma} + b\right)^{r+1} \|f\|_{(\varrho,\sigma)} \\ \left\|T_\chi^{-1} f - \sum_{s=0}^{r} D_s f\right\|_{(1-d)(\varrho,\sigma)} &\leq \frac{2}{e^2}\left(\frac{2eG}{d^2\varrho\sigma} + b\right)^{r+1} \|f\|_{(\varrho,\sigma)} \ . \end{aligned} \tag{20}$$

Proof. First we prove that

$$\begin{aligned} \|E_s f\|_{(1-d)(\varrho,\sigma)} &\leq B_s \|f\|_{(\varrho,\sigma)} \ , \\ \|D_s f\|_{(1-d)(\varrho,\sigma)} &\leq B_s \|f\|_{(\varrho,\sigma)} \ , \quad s \geq 1 \ , \end{aligned} \tag{21}$$

where the real sequence $\{B_s\}_{s\geq 1}$ is defined as

$$(22)\qquad \begin{aligned} B_1 &= \frac{2}{ed^2\varrho\sigma}G\ ,\\ B_s &= \frac{2e}{d^2\varrho\sigma}\sum_{j=1}^{s-1}\frac{s-j+1}{s}b^{j-1}GB_{s-j} + \frac{2}{sed^2\varrho\sigma}b^{s-1}G\ . \end{aligned}$$

For $s=1$ it is enough to apply the estimate (15) of lemma 1 with $d'=0$. For $s>1$ we look for a recursive estimate. We shall do it for the operators E_s, but the same argument applies word by word also to the operators D_s. The whole argument is based on the remark that E_s can be written as $E_s=\sum_{\alpha\in\mathcal{J}_s}c_\alpha F_\alpha$, where $\mathcal{J}_s$ is a set of indexes, $\{c_\alpha\}_{\alpha\in\mathcal{J}_s}$ a set of real coefficients, and $\{F_\alpha\}_{\alpha\in\mathcal{J}_s}$ a set of linear operators, each F_α being a composition of at most s operators L_{χ_j}. This is evident from the recursive definition (4) of E_s. By lemma 2 we have $\|F_\alpha f\|_{(1-d)(\varrho,\sigma)}\leq A_\alpha\|f\|_{(\varrho,\sigma)}$, with some set of positive constants $\{A_\alpha\}_{\alpha\in\mathcal{J}_s}$, and so also $\|E_s f\|_{(1-d)(\varrho,\sigma)}\leq\sum_{\alpha\in\mathcal{J}_s}|c_\alpha|A_\alpha\|f\|_{(\varrho,\sigma)}$. This information suffices in order to establish a recursive estimate. To this end, assume that we know $\sum_{\alpha\in\mathcal{J}_r}|c_\alpha|A_\alpha\leq B_r$ for $1\leq r<s$; this is true for $s=1$. Recalling the recursive definition (4) of E_s, look for an estimate of $L_{\chi_j}E_{s-j}=\sum_{\alpha\in\mathcal{J}_{s-j}}c_\alpha L_{\chi_j}F_\alpha$. By lemma 2 we get

$$\left\|L_{\chi_j}F_\alpha f\right\|_{(1-d)\varrho,\sigma}\leq\frac{2e(s-j+1)}{d^2\varrho\sigma}\|\chi_j\|_{\varrho,\sigma}\,\|F_\alpha f\|_{(1-d)(\varrho,\sigma)}$$

(see the remark after the statement of lemma 2, recalling also that F_α contains at most $s-j$ operators L_{χ_j}; also remark that this argument applies also to D_s because commuting operators does not affect the estimate). Thus, we also get

$$\begin{aligned}\left\|L_{\chi_j}E_{s-j}f\right\|_{(1-d)(\varrho,\sigma)} &\leq \frac{2e(s-j+1)}{d^2\varrho\sigma}\|\chi_j\|_{(\varrho,\sigma)}\sum_{\alpha\in\mathcal{J}_{s-j}}|c_\alpha|A_\alpha\|f\|_{(\varrho,\sigma)}\\ &\leq\frac{2e(s-j+1)}{d^2\varrho\sigma}\|\chi_j\|_{(\varrho,\sigma)}B_{s-j}\|f\|_{(\varrho,\sigma)}\ ,\end{aligned}$$

where the inductive hypothesis has been used. Using (4) and hypothesis (18) we immediately get (22).

We prove now that

$$(23)\qquad B_s\leq\left(\frac{2eG}{d^2\varrho\sigma}+b\right)^{s-1}\frac{2}{sed^2\varrho\sigma}G\ .$$

For $s = 1, 2$ this is directly checked. For $s > 2$ isolate the term $j = 1$ in the sum in (22), and change the summation index j to $j-1$, thus getting

$$\begin{aligned} B_s &= \frac{2eG}{d^2\varrho\sigma} B_{s-1} + b \sum_{j=1}^{s-2} \frac{(s-1)-j+1}{s} b^{j-1} G B_{s-1-j} + \frac{2}{sed^2\varrho\sigma} b^{s-1} G \\ &\le \frac{2eG}{d^2\varrho\sigma} B_{s-1} + \frac{s-1}{s} b B_{s-1} \ ; \end{aligned}$$

here, the definition of B_{s-1} in (22) has been used. Thus we get

$$B_s < \left(\frac{2eG}{d^2\varrho\sigma} + b \right) B_{s-1} \ ;$$

and (23) immediately follows.

By (21) and (23) we have

$$\sum_{s>0} \|E_s f\|_{(1-d)(\varrho,\sigma)} \le \frac{2G}{ed^2\varrho\sigma} \sum_{s>1} \left(\frac{2eG}{d^2\varrho\sigma} + b \right)^{s-1} \|f\|_{(\varrho,\sigma)} \ ;$$

in view of condition (19) the series in the right hand side of the latter inequality is a convergent geometric series. Since the supremum norm is bounded by the weighted Fourier norm, this implies that the series $T_\chi f$ is absolutely convergent on the domain $\mathcal{D}_{(1-d)(\varrho,\sigma)}$, as claimed. The fifth of (20) follows by simply computing the sum of the geometric series above starting from $s = r+1$ and taking into account the condition (19). The same argument applies to $T_\chi^{-1} f$. Concerning the estimate on the transformation of the coordinates p, q, a similar argument applies, with a minor difference: the terms $L_{\chi_j} p$ and $L_{\chi_j} q$ must be estimated using (14) of lemma 1. Proceeding as above, we get

$$\begin{aligned} \|E_1 p\|_{(1-d)(\varrho,\sigma)} &\le \frac{G}{ed\sigma} \ , \\ \|E_s p\|_{(1-d)(\varrho,\sigma)} &\le \left(\frac{2eG}{d^2\varrho\sigma} + b \right)^{s-1} \frac{G}{ed\sigma} \ , \\ \|E_1 q\|_{(1-d)(\varrho,\sigma)} &\le \frac{G}{d\sigma} \ , \\ \|E_s q\|_{(1-d)(\varrho,\sigma)} &\le \left(\frac{2eG}{d^2\varrho\sigma} + b \right)^{s-1} \frac{G}{d\sigma} \ , \end{aligned}$$

and similar estimates for the operators D_s. Then the same convergence argument is used, giving

$$\begin{aligned} \|T_\chi p - p\|_{(1-d)(\varrho,\sigma)} &\le \frac{G}{ed\sigma} \sum_{s>0} \left(\frac{2eG}{d^2\varrho\sigma} + b \right)^{s-1} \\ &\le \frac{2G}{ed\sigma} < \left(\frac{2eG}{d^2\varrho\sigma} + b \right) d\varrho \ , \end{aligned}$$

and the first of (20) immediately follows by condition (19). The same argument applies to the remaining estimates. *Q.E.D.*

Proof of proposition 1. By lemma 3, the series of analytic functions defining the canonical transformation are absolutely convergent in $\mathcal{D}_{(1-d)(\varrho,\sigma)}$. Thus, the same series are uniformly convergent on every compact subset of $\mathcal{D}_{(1-d)(\varrho,\sigma)}$. By Weierstrass theorem, this implies that the sums of the series are analytic functions on $\mathcal{D}_{(1-d)(\varrho,\sigma)}$, as claimed. The statement concerning the inclusions of the domains follows from the estimates (20). *Q.E.D.*

A straightforward consequence of proposition 1 is the following: if $f(p,q)$ is analytic in $\mathcal{D}_{(\varrho,\sigma)}$, then the transformed function $f\bigl(T_\chi p, T_\chi q\bigr)$ is analytic in $\mathcal{D}_{(1-d)(\varrho,\sigma)}$, being a composition of analytic functions. However, as has been remarked at the end of sect. 2.1, the transformed function is nothing but $T_\chi f$. The last estimate (20) gives in fact a direct proof of the analyticity of $T_\chi f$. Moreover, taking into account that an explicit computation must be truncated at some order r, it gives an estimate of the error.

A similar result holds for the transformation defined via Lie series: just reformulate the statement of proposition 1 by replacing T_χ and T_χ^{-1} with $\exp(L_\chi)$ and $\exp(-L_\chi)$ respectively, replacing (18) with $\|\chi\|_{(\varrho,\sigma)} = G$, and putting $b = 0$ in (19). Indeed, the generating sequence $\chi = \{\chi_1, 0, 0, \ldots\}$ generates the transformation $T_\chi = \exp L_{\chi_1}$, and (19) clearly holds with $G = \|\chi_1\|_{(\varrho,\sigma)}$ and $b = 0$.

3. Normal form theory

This section is devoted to the so called analytic part of the theorem of Nekhoroshev. As usual, I will consider a canonical system of differential equations with Hamiltonian

$$H(p,q) = h(p) + f(p,q) \ , \tag{24}$$

where p, q are action angle variables in a domain $\mathcal{G} \times \mathbf{T}^n$, with $\mathcal{G} \subset \mathbf{R}^n$ open. The Hamiltonian will be characterized by four parameters ϱ, σ (analyticity parameters), ε (perturbation parameter) and K (Fourier cutoff), in the following sense:

i. both h and f are assumed to be analytic bounded functions on the complex extension $\mathcal{D}_{(\varrho,\sigma)}$ of the real domain $\mathcal{G} \times \mathbf{T}^n$;

ii. the perturbation f is assumed to be bounded by some (small) positive quantity ε, i.e.,

$$\|f\|_{(\varrho,\sigma)} \le \varepsilon \; ; \tag{25}$$

iii. the perturbation f is assumed to have a finite Fourier representation of the form

$$f(p,q) = \sum_{|k|\le K} f_k(p) e^{ik\cdot q} \; , \tag{26}$$

K being a positive integer.

At first sight, the latter condition might seem quite strong, but in fact it is nothing but a technical simplification. The general case is discussed, for instance, in (Giorgilli and Zehnder, 1992). Indeed, in the case of an infinite Fourier expansion it is nevertheless necessary to consider a suitable Fourier cutoff in order to perform even a single perturbation step, unless the unperturbed part h of the Hamiltonian is strongly degenerate. Remark also that in view of the assumption (iii) the parameter σ is in fact arbitrary. For a short note on the possible optimization of the choice of σ see sect. 4.3.

3.1. Formal scheme

In the following, a special role will be played by those functions which have a finite Fourier representation. Thus, let me state some elementary but useful properties. For a non–negative integer K, denote by $\mathcal{F}_K$ the distinguished class of functions which can be represented as in (26), i.e., do not contain harmonics with $|k| > K$ (so that the assumption (iii) above could be be replaced by "f is of class $\mathcal{F}_K$"). The following properties are immediately checked: if $f \in \mathcal{F}_K$ and $g \in \mathcal{F}_{K'}$ then $f+g \in \mathcal{F}_{\max(K,K')}$, $fg \in \mathcal{F}_{K+K'}$ and $\{f,g\} \in \mathcal{F}_{K+K'}$. Moreover, looking at the triangular diagram (5) one checks by induction that if $f_s \in \mathcal{F}_{sK}$ and $\chi_s \in \mathcal{F}_{sK}$, then one has $E_j f_s \in \mathcal{F}_{(j+s)K}$, and so also $g_s \in \mathcal{F}_{sK}$. Briefly, the algebraic scheme of Lie transform is consistent with the request that at every order s all the functions are of class $\mathcal{F}_{sK}$.

Let me now come to the normal form, using the formalism of Lie transform. Starting with the Hamiltonian (24), one looks for a truncated generating sequence $\chi^{(r)} = \{\chi_1, \ldots, \chi_r\}$, where r is an arbitrary but fixed positive integer, such that the transformed Hamiltonian $H^{(r)} = T_{\chi^{(r)}} H$ takes the form

$$(27) \quad T_{\chi^{(r)}} H(p,q) = h(p) + Z_1(p,q) + \ldots + Z_r(p,q) + \mathcal{R}^{(r+1)}(p,q) \ ,$$

where $Z_1, \ldots, Z_r$ are "in normal form", that is, they have some nice property that we are going to identify, while $\mathcal{R}^{(r+1)}$ is a still unnormalized remainder. To this end, we consider (27), as an equation for the unknowns $\chi_1, \ldots, \chi_r$ and $Z_1, \ldots, Z_r$, to be solved with some conditions on the Z's (e.g., we require $Z = Z(p)$).

Look now at the triangular diagram (5), with $H^{(r)}$ in place of g, h in place of f_0 and f in place of f_1 (so that all columns corresponding to $f_2, \ldots$ are actually empty). It is immediately seen that the first $r+1$ rows give

$$(28) \qquad E_s h + E_{s-1} f = Z_s \ , \quad s = 1, \ldots, r \ ,$$

while the rest of the diagram defines the remainder $\mathcal{R}^{(r+1)}$ (the explicit form of the remainder is not necessary here). Using definition (4) of E_s in the form

$$E_s h = L_{\chi_s} h + \sum_{j=1}^{s-1} \frac{j}{s} L_{\chi_j} E_{s-j} h$$

and in view of $L_{\chi_s} h = -L_h \chi_s$, one immediately obtains the equations

$$(29) \qquad L_h \chi_s + Z_s = \Psi_s \ , \quad s = 1, \ldots, r \ ,$$

where Ψ_s is known, being

$$(30) \qquad \Psi_1 = f \ , \quad \Psi_s = E_{s-1} f + \sum_{j=1}^{s-1} \frac{j}{s} L_{\chi_j} E_{s-j} h \quad \text{for } s \geq 2 \ .$$

A more convenient expression is

$$(31) \qquad \Psi_s = \frac{1}{s} E_{s-1} f + \sum_{j=1}^{s-1} \frac{j}{s} L_{\chi_j} Z_{s-j} \quad \text{for } s \geq 2 \ .$$

In order to obtain this formula, transform the expression (30) of Ψ_s by replacing the recursive definition (4) of E_{s-1}, and using (28) for $E_{s-j} h$;

this gives

$$\begin{aligned}\Psi_s &= \sum_{j=1}^{s-1} \frac{j}{s(s-1)} L_{\chi_j} \left(sE_{s-1-j} f + (s-1) E_{s-j} h\right) \\ &= \sum_{j=1}^{s-1} \frac{j}{s(s-1)} L_{\chi_j} \left(E_{s-1-j} f + (s-1) Z_{s-j}\right) \\ &= \frac{1}{s} \sum_{j=1}^{s-1} \frac{j}{s-1} L_{\chi_j} E_{s-j-1} f + \sum_{j=1}^{s-1} \frac{j}{s} L_{\chi_j} Z_{s-j} \ ;\end{aligned}$$

then remark that the first sum is nothing but $E_{s-1} f$, and (31) is recovered.

Equation (29) is usually solved as follows. Since the q variables are angles, expand Ψ_s in Fourier series as

$$\Psi_s(p,q) = \sum_{k\in\mathbf{Z}^n} \psi_k(p) \exp(ik\cdot q) \ ,$$

with known coefficients $\psi_k(p)$, and consider the same expansion for χ_s and Z_s, with unknown coefficients $c_k(p)$ and $z_k(p)$ respectively. Then the equation above splits into the system of equations for the coefficients

$$i(k\cdot\omega(p))\, c_k(p) + z_k(p) = \psi_k(p) \ , \quad \omega(p) = \frac{\partial h}{\partial p} \ , \tag{32}$$

where $\omega(p)$ are the frequencies of the unperturbed system. The key point here is that the coefficient $c_k(p)$ should be determined as

$$c_k(p) = -i \frac{f_k(p)}{k\cdot\omega(p)} \ ;$$

this is formally consistent only if the denominator $k\cdot\omega(p)$ does not vanish. Thus, the simplest rule seems to be the following: if $k\cdot\omega = 0$, then put $z_k = f_k$ and $c_k = 0$; else put $z_k = 0$ and determine c_k as above. However, the set $\omega \in \mathbf{R}^n$ such that $k\cdot\omega = 0$ for some nonzero $k \in \mathbf{Z}^n$ is clearly dense in $\mathbf{R}^n$; thus, unless some strong restriction is imposed on h and/or on f, one has to expect that arbitrarily close to every value p of the actions there is some value, p' say, where at least one denominator $k\cdot\omega(p')$ does vanish. This is the well known problem of the "small denominators", which lies at the basis of Poincaré's theorem on nonexistence of uniform integrals. The way out of these troubles consists in making some truncation of the Fourier expansion of the Hamiltonian (in fact, it is a common practice to consider only a *finite* set of Fourier components), and constructing a *local* theory (i.e., one restricts the action variables to some suitable domain). The simplest way, that I choose here, is to start from the beginning with a truncated Fourier expansion. The relevant fact is that at every order s the equation (29) involves only functions of class $\mathcal{F}_{sK}$. Thus, due to the finite number of small denominators, the generating function can be determined on open domains.

3.2. *Quantitative estimates on the generating sequence*

A more precise formulation of the remarks at the end of the previous section is based on the following technical definitions.

(i) A *resonance module* is defined as a subgroup $\mathcal{M} \in \mathbf{Z}^n$ satisfying $\mathrm{span}(\mathcal{M}) \cap \mathbf{Z}^n = \mathcal{M}$; here, both $\mathcal{M}$ and $\mathbf{Z}^n$ are considered as subsets of $\mathbf{R}^n$, and $\mathrm{span}(\mathcal{M})$ is the linear subspace in $\mathbf{R}^n$ spanned by $\mathcal{M}$.

(ii) A function $Z(p,q)$ is said to be in *normal form with respect to the resonance module* $\mathcal{M}$ in case its Fourier expansion has the form

$$Z(p,q) = \sum_{k\in\mathcal{M}} z_k(p)\exp(ik\cdot q) \ ,$$

namely, if it contains only harmonics belonging to $\mathcal{M}$ (resonant harmonics).

(iii) A set $\mathcal{V} \subset \mathbf{R}^n$ is said to be a *nonresonance domain of type* $(\mathcal{M},\alpha,\delta,N)$ in case one has

$$|k\cdot\omega| > \alpha \text{ for all } p \in \mathcal{V}_\delta \ , \quad k \in \mathbf{Z}^n \setminus \mathcal{M} \text{ and } |k| \le N \ .$$

Here, α and δ are positive parameters, N is a positive integer, $\mathcal{M}$ is a resonance module, and $\mathcal{V}_\delta$ is the complex extension of the real domain $\mathcal{V}$, as described in sect. 2.3.

The nonresonance domain is the subset of the action space where the Hamiltonian can be given a normal form with respect to the given resonance module. Indeed, still referring to the equation (32) for the coefficients of the Fourier expansion, the equation (29) can be consistently solved as follows: if $k \in \mathcal{M}$, then put $z_k = \psi_k$ and $c_k = 0$; else put $z_k = 0$ and $c_k = -i\psi_k/(k\cdot\omega)$.

Let me now come to quantitative estimates. First of all, remark that if $\|\Psi_s\|_{(1-d')(\varrho,\sigma)}$ is known for some $d' < 1$, then the definition of the norm together with the form above of the solution of (29) and the definition of nonresonance domain immediately gives

$$\begin{aligned} \|\chi_s\|_{(1-d')(\varrho,\sigma)} &\le \frac{1}{\alpha}\,\|\Psi_s\|_{(1-d')(\varrho,\sigma)}\ , \\ \|Z_s\|_{(1-d')(\varrho,\sigma)} &\le \|\Psi_s\|_{(1-d')(\varrho,\sigma)}\ . \end{aligned} \tag{33}$$

Remarking that the recursive definition (31) of Ψ_s involves Poisson brackets, and so a restriction of the domains in the estimate, fix the final restriction $d < 1$ of the domain, and define

$$d_s = \frac{s-1}{r-1} d\ , \quad 1 \le s \le r\ ; \tag{34}$$

look now for an estimate of $\|\Psi_s\|_{(1-d_s)(\varrho,\sigma)}$, from which the estimate for $\|\chi_s\|_{(1-d_s)(\varrho,\sigma)}$ is obtained in view of (33). To this end, look for two sequences $\{\eta_s\}_{1\le s\le r}$ and $\{\vartheta_s\}_{1\le s\le r}$ such that

$$\|E_{s-1} f\|_{(1-d_s)(\varrho,\sigma)} \le \eta_s \|f\|_{(\varrho,\sigma)}\ , \quad \|\Psi_s\|_{(1-d_s)(\varrho,\sigma)} \le \vartheta_s \|f\|_{(\varrho,\sigma)}\ . \tag{35}$$

Clearly, in view of $E_0 f = f$ and of $\Psi_1 = f$, we can choose $\eta_1 = \vartheta_1 = 1$. Moreover, the recursive definition (4) of E_{s-1} and the expression (31) of Ψ_s give the estimates

$$\|E_{s-1} f\|_{(1-d_s)(\varrho,\sigma)} \le \frac{C}{s-1} \sum_{j=1}^{s-1} \frac{1}{s-j} \vartheta_j \eta_{s-j} \|f\|_{(\varrho,\sigma)}\ ,$$

$$\|\Psi_s\|_{(1-d_s)(\varrho,\sigma)} \le \frac{1}{s}\eta_{s-1}\|f\|_{(\varrho,\sigma)} + \frac{C}{s} \sum_{j=1}^{s-1} \frac{1}{s-j} \vartheta_j \vartheta_{s-j} \|f\|_{(\varrho,\sigma)}\ ,$$

where

$$C = \frac{2(r-1)^2}{e\alpha d^2 \varrho\sigma}\,\|f\|_{(\varrho,\sigma)}\ . \tag{36}$$

Thus, the sequences $\{\eta_s\}$ and $\{\vartheta_s\}$ can be recursively defined as

$$\begin{aligned} \eta_s &= \frac{C}{s-1} \sum_{j=1}^{s-1} \frac{1}{s-j}\,\vartheta_j \eta_{s-j} \\ \vartheta_s &= \frac{1}{s}\eta_s + \frac{C}{s} \sum_{j=1}^{s-1} \frac{1}{s-j}\,\vartheta_j \vartheta_{s-j}\ . \end{aligned} \tag{37}$$

Remark now that for our purposes it is enough to find an estimate of the sequence $\{\vartheta_s\}$. It is not difficult to see that one has

$$\vartheta_s \le \frac{1}{s}(4C)^{s-1} \ . \tag{38}$$

To this end, let us first prove that $\vartheta_s \le C^{s-1}\lambda_s$, where the sequence $\{\lambda_s\}_{s\ge1}$ is recursively defined as

$$\lambda_1 = 1 \ , \quad \lambda_s = \sum_{j=1}^{s-1} \lambda_j \lambda_{s-j} \ . \tag{39}$$

This is clearly true for $s = 1$. For $s > 1$ we proceed by induction. In view of the second of (37) we have $\eta_j < j\vartheta_j$ for all $j > 1$, and so, by the inductive hypothesis, also $\eta_j \le jC^{j-1}\lambda_j$ for $1 \le j < s$. Replacing this in the first of (37) we get the stronger inequality

$$\eta_s \le \frac{C^{s-1}}{s-1} \sum_{j=1}^{s-1} \lambda_j \lambda_{s-j} = \frac{C^{s-1}}{s-1}\lambda_s \ .$$

Using this in the second of (37) we get

$$\vartheta_s < \frac{C^{s-1}}{s(s-1)}\lambda_s + \frac{C^{s-1}}{s} \sum_{j=1}^{s-1} \lambda_j \lambda_{s-j} = \frac{C^{s-1}}{s-1}\lambda_s \ ,$$

which completes induction.

Thus, in order to prove (38) it is enough to prove that the sequence (39) is bounded by $\lambda_s \le 4^{s-1}/s$. To this end, let the function $g(z)$ be defined as $g(z) = \sum_{s\ge1} \lambda_s z^s$, so that $\lambda_s = g^{(s)}(0)/s!$. Then it is immediate to check that the recursive definition (39) is equivalent to the equation $g = z + g^2$. By repeated differentiation of the latter equation one readily finds (check by induction)

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

where the standard notation $(2n+1)!! = 1 \cdot 3 \cdot \ldots \cdot (2n+1)$ has been used. From this we get

$$\lambda_s = \frac{2^{s-1}(2s-3)!!}{s!} \le 4^{s-1}/s \ ,$$

as claimed.

Using (33) and (35) and recalling the hypothesis $\|f\|_{(\varrho,\sigma)} \le \varepsilon$, we readily get the estimate for the generating sequence

$$\|\chi_s\|_{1-d} \le \frac{b^{s-1}}{s} G \tag{40}$$

with

$$b = \frac{8(r-1)^2\varepsilon}{e\alpha d^2 \varrho\sigma} \ , \quad G = \frac{\varepsilon}{\alpha} \ . \tag{41}$$

Finally, we apply lemma 3 and proposition 1. With the choice $d = 1/4$, and using $\mathcal{R}^{(r+1)} = \sum_{s>r}(E_s h + E_{s-1} f)$ it is an easy matter to prove the following

Proposition 2: *C*onsider the Hamiltonian (24); let $\mathcal{V}$ be a nonresonance domain of type $(\mathcal{M}, \alpha, \delta, rK)$, and assume

$$\mu := \frac{2^7}{\alpha\varrho\sigma} r^2 \varepsilon \le \frac{1}{2} \ . \tag{42}$$

Then:

(i) there exists a finite generating sequence $\chi^{(r)} = \{\chi_1, \ldots, \chi_r\}$, with $\chi_s \in \mathcal{F}_{sK}$ which transforms the Hamiltonian to

$$H^{(r)}(p,q) = h(p) + Z_1(p,q) + \ldots + Z_r(p,q) + \mathcal{R}^{(r+1)}(p,q) \ ,$$

where $Z_1, \ldots, Z_r$ are in normal form with respect to the resonance module $\mathcal{M}$;

(ii) for every $p \in \mathcal{V}_{\delta/2}$ one has the estimate

$$\left|p - T_{\chi^{(r)}} p\right| \le \frac{\delta}{4} \ , \quad \left|p - T^{-1}_{\chi^{(r)}} p\right| \le \frac{\delta}{4} \ ;$$

(iii) the remainder $\mathcal{R}^{(r+1)}$ is estimated by

$$\left\|\mathcal{R}^{(r+1)}\right\|_{1-2d} \le B\mu^r \ , \quad B = \frac{1}{e^2}\bigl(\|h\|_{(\varrho,\sigma)} + 2\|f\|_{(\varrho,\sigma)}\bigr) \ .$$

3.3. Local estimates

Consider now the case of a nondegenerate unperturbed Hamiltonian, namely assume

$$\det\left(\frac{\partial^2 h}{\partial p_i \partial p_j}\right) \neq 0 \ .$$

Suppose that we have determined a nonresonance domain $\mathcal{V}$ of type $(\mathcal{M}, \alpha, \delta, rK)$ for some nonresonance module $\mathcal{M}$, some integer r and some positive α and $\delta \leq \varrho$. Thus, according to proposition 2, the Hamiltonian can be given a normal form with respect to $\mathcal{M}$ in $\mathcal{V}_{\delta/2}$. Let me denote by p', q' the new canonical variables. As a first approximation, ignore the remainder $\mathcal{R}^{(r+1)}$. It is an easy matter to check that there are $n - \dim \mathcal{M}$ independent first integrals of the form

$$\Phi_\lambda(p') = \sum_{j=1}^{n} \lambda_j p'_j \ ,$$

where $\lambda \in \mathbf{R}^n$ must satisfy the condition $\lambda \perp \mathcal{M}$, i.e., $\sum_j \lambda_j k_j = 0$ for all $k \in \mathcal{M}$. Given now an initial point $p'(0) \in \mathcal{V}$ (the nonresonance domain, endowed with the new coordinates p'), the orbit issuing from $p'(0)$ lies on the intersection of the surfaces $\Phi_\lambda(p') = \Phi_\lambda\bigl(p'(0)\bigr)$, which is a plane, $\Pi_{\mathcal{M}}\bigl(p'(0)\bigr)$ say, parallel to the subspace in $\mathbf{R}^n$ generated by $\mathcal{M}$; this will be referred to as the *plane of fast drift*. The reason for this name is that the velocity of the motion along the plane $\Pi_{\mathcal{M}}\bigl(p'(0)\bigr)$ is of the order of Z_1, which in turn is of the order of the perturbation f.

Turning back to the original action variables p, q, one concludes that the orbit lies on an invariant surface of dimension $n - \dim \mathcal{M}$ passing through the initial point $p(0)$, and close to a plane. More precisely, consider the plane $\Pi_{\mathcal{M}}\bigl(p(0)\bigr)$ in the original action variables; then by statement (ii) of proposition 2 the invariant surface is contained in a strip of width $\delta/4$ around this plane, i.e., the distance of the orbit from the plane of fast drift does not exceed $\delta/4$. The effect of change from new to original coordinates is called *deformation*.

Let us now take into account the effect of the remainder. In the new coordinates, this could cause a slow motion transversal to the plane of fast drift, named *noise*. However, the velocity of the motion is bounded by the size of the remainder, i.e., by statement (iii) of proposition 2. Thus, it takes a time of the order of μ^{-r} for the distance from the plane of fast drift to reach the size, say, $\delta/4$. In the original variables, the effect of the noise must be added to that of the deformation; this means that the orbit is still contained in a strip of width $\delta/2$ around the plane of fast drift.

The last point is that the description above of the motion is true provided the orbit does not exit from the nonresonance domain $\mathcal{V}$. Thus, one concludes that *the distance of the orbit from the plane of fast drift* $\Pi_{\mathcal{M}}\bigl(p(0)\bigr)$ *does not exceed* $\delta/2$ *for all times* $D\mu^{-r}$*, unless the orbit leaves the nonresonance domain* $\mathcal{G}$*. The constant* D *is estimated as* $D = 8\delta\sigma/B$*, with* B *as in proposition 2.* The formal proof of this statement can be found in (Giorgilli and Zehnder, 1992).

Thus, we know a lot about the behaviour of the orbit as far as it remains in the nonresonance domain. The natural question is: what happens if the motion along the plane of fast drift moves the orbit out of the nonresonance domain? The answer to this question is the achievement of the so called geometric part of Nekhoroshev's theorem.

4. Exponential stability

In order to answer the question at the end of the previous section we must come back to considering the whole domain $\mathcal{G}$ of the original Hamiltonian. Following Nekhoroshev, I will show in this section how the domain $\mathcal{G}$ can be covered with nonresonance domains. From this construction a global stability result over an exponentially large time will be deduced.

A relevant role will be played by the condition of *convexity* of the unperturbed Hamiltonian $h(p)$. Precisely, referring to the Hamiltonian (24), and still denoting by $\omega(p) = \frac{\partial h}{\partial p}$ the unperturbed frequencies, we assume that there are positive constants m and M such that

$$(43) \qquad |A(p)v \cdot v| \ge m\|v\|^2 \ , \quad \|A(p)v\| \le M\|v\| \quad \text{for all } v \in \mathbf{R}^n \ ,$$

where $A(p) = \frac{\partial \omega}{\partial p}$. In the original proof of Nekhoroshev the condition of *steepness* was assumed, which is weaker than convexity, but more difficult to check. The relevant consequence of the convexity is the so called *nonoverlapping of resonances*. Stating such a property requires the technical discussion which follows, but the main consequence can be described in rough terms: if the motion along the plane of fast drift causes an orbit to leave (to enter) a nonresonance domain in a short time, then the orbit goes to (comes from) a region characterized by less resonances. This prevents the formation of chains of resonances, which could cause the orbit to go far from the initial point.

4.1. Geography of resonances

The input to the geometric part of Nekhoroshev's theorem is represented by the following quantities:

(i) the domain $\mathcal{G}$ and the positive parameter δ defining the complex domain $\mathcal{G}_\delta$ of the action variables;

(ii) a positive integer N, to be identified with the final truncation order rK of the analytic part.

(iii) the mapping $\omega(p)$ satisfying the conditions (43).

The following construction returns a covering of the domain $\mathcal{G}$ with non-resonance domains characterized by parameters $(\mathcal{M}, \beta_s/2, \delta_s, N)$, where β_s and δ_s are determined as functions of δ, N, $s = \dim \mathcal{M}$ and the number n of degrees of freedom. Explicit values are

$$
\begin{aligned}
\beta_0 &= \left[\left(\frac{2M}{m} \right)^{n+1} (n+1)!\, N^{(n^2+n-2)/2} \right]^{-1} \delta \ , \\
\beta_s &= \left(\frac{2M}{m} \right)^{s} s!\, N^{[s(s-1)]/2} \beta_0 \quad \text{for } 1 \le s \le n \ , \\
\delta_s &= \frac{\beta_s}{2N} \ , \quad 0 \le s \le n \ .
\end{aligned}
\tag{44}
$$

Let $\mathcal{M}$ be a resonance module with dimension s, and consider the set $\mathcal{M}_N = \{k \in \mathcal{M} \,:\, |k| \le N\}$; we call $\mathcal{M}$ an N–module in case $\mathcal{M}_N$ contains s independent vectors k. In view of the results of the analytic part, we shall not need to consider resonance moduli which are not N–moduli.

1. The *resonant zone* $\mathcal{Z}_\mathcal{M}$ associated to a N–module $\mathcal{M}$ of dimension s is defined as the set of points $p \in \mathcal{G}$ satisfying $|k \cdot \omega(p)| < \beta_s$ for s independent vectors of $\mathcal{M}_N$, namely, the points close "within β_s" to resonance. The dimension s of $\mathcal{M}$ will be called the multiplicity of the resonance, or also the multiplicity of the zone. The union of all resonant zones with the same multiplicity s will be denoted by $\mathcal{Z}^*_s$, and will be called the *resonant region of order* s; we also define $\mathcal{Z}^*_{n+1} = \emptyset$. In rough terms, $\mathcal{Z}^*_s$ is the set of points which admit at least s resonances within β_s; further resonances are not excluded till now. In particular, resonant zones

$\mathcal{Z}_{\mathcal{M}}$ corresponding to different $\mathcal{M}$'s can intersect, giving rise to regions characterized by resonances of higher multiplicity.

2. A *resonant block* $\mathcal{B}_{\mathcal{M}}$ associated to $\mathcal{M}$ of dimension s is defined as the set $\mathcal{B}_{\mathcal{M}} = \mathcal{Z}_{\mathcal{M}} \setminus \mathcal{Z}^*_{s+1}$. In view of this definition the block is the set of points admitting s resonances within β_s, but no further resonances; it is a good basis for the construction of nonresonant domains.

3. To every point p of a given resonant block $\mathcal{B}_{\mathcal{M}}$ we associate the plane in $\mathbf{R}^n$ passing through p and parallel to span$(\mathcal{M})$ (i.e., the plane of fast drift), and define the *extended plane* $P_{\mathcal{M},\delta_s}(y)$ as the set of points of $\mathbf{R}^n$ the distance of which from the plane above is less than δ_s. The *cylinder* $\mathcal{C}_{\mathcal{M},\delta_s}$ is defined as the intersection $\mathcal{P}_{\mathcal{M},\delta_s} \cap \mathcal{Z}_{\mathcal{M}}$ of the extended plane with the corresponding resonant zone, and the bases of the cylinder are defined as the intersection $\mathcal{P}_{\mathcal{M},\delta_s} \cap \partial\mathcal{Z}_{\mathcal{M}}$ of the extended plane with the border of the zone. Remark that the bases of the cylinder do not belong to the zone, because the latter is defined as an open set.

4. Finally, the extended resonant block $\mathcal{B}_{\mathcal{M},\delta_s}$ is defined as the union of all cylinders corresponding to all points of the block $\mathcal{B}_{\mathcal{M}}$.

Having completed this construction, one proves the following

Proposition 3: *With the hypotheses (43), and with the constants $\beta_0, \ldots, \beta_n$ and $\delta_0, \ldots, \delta_n$ defined by (44) one has the following properties:*

(i) The blocks are a covering of the domain $\mathcal{G}$; that is, every point in $\mathcal{G}$ belongs to at least one block.

(ii) The extended block $\mathcal{B}_{\mathcal{M},\delta_s}$ is a nonresonance domain of type $(\mathcal{M}, \beta_s/2, \delta_s, N)$ (remark that $\delta_s < \varrho$).

(iii) The complement of a resonant region $\mathcal{Z}^*_s$ of multiplicity s is covered by the union of all block of multiplicity less than s; that is, if a point does not belong to any zone of multiplicity s, then it belongs to at least one block of multiplicity less than s.

(iv) The closure of an extended block $\mathcal{B}_{\mathcal{M},\delta_s}$ and the resonant zone $\mathcal{Z}_{\mathcal{M}'}$ corresponding to a different N–module with the same dimension as $\mathcal{M}$ are disjoint; that is, the basis of a cylinder $\mathcal{C}_{\mathcal{M},\delta_s}(p)$ corresponding to a point $p \in \mathcal{B}_{\mathcal{M}}$ does not belong to any resonant

zone of the same multiplicity $\dim(\mathcal{M})$.

(v) *The diameter of a cylinder does not exceed* δ.

The proof of these properties contains only one delicate point concerning the estimate of the diameter of the cylinders. It is precisely at this point that we use the hypothesis of convexity (43). The interested reader will find all the details in (Benettin, Galgani and Giorgilli, 1985).

4.2. Global estimates

Having fixed $\delta \le \varrho$, r and K we apply proposition 3, thus constructing the geometric structure in $\mathcal{G}_\delta$. Let us now exploit the consequences of the properties (i)–(v) above. Property (ii) means that the extended blocks are the domains where the analytic theory of the first part can be consistently applied; in particular, setting $N = rK$ with a given K allows us to perform r normalization steps in every block. It should be remarked that in the statement of proposition 2 we must use $\beta_s/2$ and δ_s in place of α and ϱ, respectively, with $s = \dim \mathcal{M}$. Thus, to a block of multiplicity s we should associate a stability time $T_s = D\mu_s^{-r}$. However, the argument below will be valid for the shortest time

$$T_0 = \frac{8\delta_0\sigma}{B}\,\mu_0^{-r}\ , \tag{45}$$

with

$$\mu_0 = \frac{2^8}{\beta_0\delta_0\sigma}\,r^2\varepsilon \le \frac{1}{2}\ . \tag{46}$$

The picture of the motion is as follows: take an initial value p_0 for the actions; find a resonant block $\mathcal{B}_{\mathcal{M}}$ to which p_0 belongs (may be it is not unique, but this is not relevant), and consider the corresponding cylinder $\mathcal{C}_{\mathcal{M},\delta_s}(p_0)$. Then the orbit either is confined inside the cylinder for all t in the interval $(-T_0, T_0)$, or it leaves (enters) the cylinder through a basis. In the latter case, by property (iv) the intersection of the orbit with the basis of the cylinder is a point not belonging to the resonant region of multiplicity s; moreover, by property (iii), this point belongs to a block of multiplicity less than s.

It is now easy to conclude that there is a cylinder containing the orbit for a time interval of length $T_0 \simeq \mu_0^{-r}$. For, assume that during the time interval $(-T_0, T_0)$ the orbit visits several blocks, and determine an instant, t_0 say, at which the point $p(t_0)$ on the orbit belongs to a block of minimal multiplicity. Consider the cylinder corresponding to $p(t_0)$: this is the wanted cylinder. Indeed, during the time interval $(t_0 - T_0, t_0 + T_0) \cap (-T_0, T_0)$ the orbit could leave the cylinder only by entering a block of lower multiplicity, thus contradicting the assumption that the multiplicity of the block was minimal.

Property (v) gives then a bound on the time evolution of the actions p. A formal statement is the following.

Proposition 4: *Assume that the hypotheses of proposition 2 are satisfied; assume also the hypotheses (43), and let* μ_0 *be as in (46). Let also* $\delta \le \varrho$ *and* r *be given. Then for every solution with* $p(0) \in \mathcal{G}$ *one has*

$$\mathrm{dist}(p(t), p(0)) < \delta \quad \text{for} \quad |t| \le T_0 = \frac{8\delta\sigma}{B}\mu_0^{-r} \ ,$$

with B *as in proposition 2.*

4.3. The exponential estimate

Proposition 4 gives a stability result depending on several parameters, namely: the parameters ε, K and ϱ depending on the analyticity properties of the Hamiltonian and on the choice of the domain of initial data, the parameter σ in the norm, and the arbitrary parameters $r > 1$ and $\delta \le \varrho$. The constants β_s and δ_s are instead given by the geometric part. Now we discuss how to make a choice for the parameters not intrinsically related to the Hamiltonian, namely σ, r and δ.

Concerning the choice of σ, the natural remark is that it actually influences ε, since $\|f\|_{(\varrho,\sigma)}$ is expected to grow roughly as $e^{K\sigma}$. On the other hand, by (46) we have $\mu_0 \simeq e^{K\sigma}/\sigma$. Since we are interested in minimizing μ_0, it is reasonable to put $\sigma \simeq 1/K$.

We now look for an optimal choice of r and δ, i.e., we try to obtain the best estimate for the stability time. To this end, recalling (46) and (44), let us write

$$\mu_0 = \frac{\mu_*}{e} \cdot \frac{r^a \varepsilon}{\delta^2} \ , \quad a = n^2 + n + 1 \ ,$$

with the constant μ_* depending on M, m, σ and the number n of degrees of freedom. Impose the condition $\mu_0 \le 1/e$, which is stronger than (46), and set

$$r = \left[\left(\frac{\delta^2}{\mu_* \varepsilon} \right)^{1/a} \right] \ , \quad \delta = (\mu_* \varepsilon)^{1/4} \ ,$$

where $[\cdot]$ denotes the integer part. A straightforward substitution in proposition 4 leads to the following

Theorem: *Let the Hamiltonian $H = h(p) + f(p,q)$ be an analytic function of p in the complex domain $\mathcal{G}_\varrho$, with $\mathcal{G}$ open and convex and $\varrho > 0$, and let f be of class $\mathcal{F}_K$ for some positive integer K. Assume that for some positive constants ε, M and M one has:*

(i) $\|f\|_{(\varrho\sigma)} \leq \varepsilon$ *for some* $\sigma > 0$*;*

(ii) $\|A(p)v\| \leq M\|v\|$ *and* $A(p)v \cdot v \geq m\|v\|$ *for all* $v \in \mathbf{R}^n$ *and all* $p \in \mathcal{G}_\varrho \cap \mathbf{R}^n$*, with* $A(p) = \frac{\partial h}{\partial p \partial p}$.

Then there exist positive constants μ_* and T_* such that the following holds true: if

$$\mu_* \varepsilon < \min(1, \varrho^4) \ ,$$

then for every orbit $p(t), q(t)$ satisfying $p(0) \in \mathcal{G}$ one has

$$\mathrm{dist}(p(t), p(0)) < (\mu_* \varepsilon)^{1/4}$$

for all times t satisfying

$$|t| \leq T_* \varepsilon^{1/4} \exp \left[\left(\frac{1}{\mu_* \varepsilon} \right)^{1/(2a)} \right] \ ,$$

with $a = n^2 + n + 1$.

REFERENCES

Benettin, G., Galgani, L. and Giorgilli, A. (1985): "A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems", *Cel. Mech.* **37**, 1–25.

Deprit, A. (1969): "Canonical transformations depending on a small parameter", *Cel. Mech.* **1**, 12–30.

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

Giorgilli, A. and Galgani, L. (1978): "Formal integrals for an autonomous Hamiltonian system near an equilibrium point", *Cel. Mech.* **17**, 267–180.

Giorgilli, A. and Galgani, L. (1985): "Rigorous estimates for the series expansions of Hamiltonian perturbation theory", *Cel. Mech.* **37**, 95–112.

Giorgilli, A. and Zehnder, E. (1992): "Exponential stability for time dependent potentials", *ZAMP* **43**, 827–855.

Hori, G. (1966):: "Theory of general perturbations with unspecified canonical variables", *Publ. Astron. Soc. Japan*, **18**, 287–296.

Littlewood, J.E. (1959): "On the equilateral configuration in the restricted problem of three bodies", *Proc. London Math. Soc.* **9**, 343–372;

Littlewood, J. E. (1959): "The Lagrange configuration in celestial mechanics", *Proc. London Math. Soc.* **9**, 525–543.

Moser, J., (1955): "Stabilitätsverhalten kanonisher differentialgleichungssysteme", *Nachr. Akad. Wiss. Göttingen, Math. Phys.* **K1 IIa**, nr.6, 87–120.

Nekhoroshev, N., N., (1977): "Exponential estimates of the stability time of near–integrable Hamiltonian systems.", *Russ. Math. Surveys*, **32**, 1-65.

Nekhoroshev, N., N., (1979): "Exponential estimates of the stability time of near–integrable Hamiltonian systems, 2.", *Trudy Sem. Petrovs.*, **5**, 5-50.

Poincaré, H. (1892): *Le méthodes nouvelles de la mécanique célèste*, Gauthier–Villard, Paris.