Effective stability and KAM theory

AMADEU DELSHAMS[1]
Departament de Matemàtica Aplicada I
Universitat Politècnica de Catalunya
Av. Diagonal 647, 08071 Barcelona
E-mail: amadeu@ma1.upc.es

PERE GUTIÉRREZ[2]
Departament de Matemàtica Aplicada II
Universitat Politècnica de Catalunya
Pau Gargallo 5, 08071 Barcelona
E-mail: gutierrez@ma2.upc.es

[1]Partially supported by the Spanish grant DGCYT PB90–0580.
[2]Supported by the U.P.C. grant PR9409.

Abstract

The two main stability results for nearly-integrable Hamiltonian systems are revisited: Nekhoroshev theorem, concerning exponential lower bounds for the stability time (effective stability), and KAM theorem, concerning the preservation of a majority of the nonresonant invariant tori (perpetual stability). To stress the relationship between both theorems, a common approach is given to their proof, consisting of bringing the system to a normal form constructed through the Lie series method. The estimates obtained for the size of the remainder rely on bounds of the associated vectorfields, allowing to get the "optimal" stability exponent in Nekhoroshev theorem for quasiconvex systems. On the other hand, a direct and complete proof of the isoenergetic KAM theorem is obtained. Moreover, a modification of the proof leads to the notion of nearly-invariant torus, which constitutes a bridge between KAM and Nekhoroshev theorems.

1 Introduction

We consider a *nearly-integrable Hamiltonian* written in *action–angle variables:*

$$H(\phi, I) = h(I) + f(\phi, I), \tag{1}$$

where $\phi = (\phi_1, \dots, \phi_n) \in \mathbf{T}^n$ and $I = (I_1, \dots, I_n) \in \mathcal{G} \subset \mathbf{R}^n$ are, respectively, the angular and action variables, and f is a small perturbation, of size ε, of the integrable Hamiltonian h. It is well-known that the dynamics associated to the unperturbed Hamiltonian h is very simple: the action $I(t)$ remains constant for all motions. Then, all n-dimensional tori $I = \text{const.}$ in phase space $\mathbf{T}^n \times \mathcal{G}$ are invariant. The flow on each torus is linear, with frequency vector $\omega(I) = \operatorname{grad} h(I)$.

In general, for the perturbed system associated to (1), the dynamics can be very complicated. It is thought that there are unstable motions, and that Arnold diffusion takes place. Concerning stability, the main results are provided by Nekhoroshev and KAM (Kolmogorov–Arnold–Moser) theorems.

Nekhoroshev theorem, which was first proved in [24], leads to the concept of *effective stability.* Roughly speaking, it states that an estimate of the type

$$|I(t) - I(0)| < R_0\, \varepsilon^b \qquad \text{for} \quad |t| \le T_0 \cdot \exp\left\{ \left(\frac{\varepsilon_0}{\varepsilon}\right)^a \right\},$$

holds for all initial conditions $(\phi(0), I(0)) \in \mathbf{T}^n \times \mathcal{G}$, provided *steepness* conditions are fulfilled by h. The *stability exponents* a and b are positive constants. For the case of a perturbation of a *quasiconvex* Hamiltonian (the simplest kind of steepness), these exponents have been successively improved along several papers. Thus, the exponent $a = 2/(n^2 + n)$ was found in [2]; and $a < 1/(2n + 1)$ in [15]. Finally, the exponents

$$a = b = \frac{1}{2n},$$

are stated in [17], [27]. It has been conjectured [6] that the exponent $a = 1/2n$ is *optimal.*

Estimates of an analogous type can be obtained for the case of a perturbation of a system of *harmonic oscillators:* $H(\phi, I) = \omega \cdot I + f(\phi, I)$, where ω is now a constant vector satisfying a Diophantine condition:

$$|k \cdot \omega| \ge \frac{\gamma}{|k|_1^\tau} \qquad \forall k \in \mathbf{Z}^n \setminus \{0\}, \tag{2}$$

for some $\tau \ge n - 1$ and $\gamma > 0$ (it is well-known that, if $\tau > n - 1$, the set of vectors ω satisfying this condition for a given $\gamma > 0$ has relative measure $1 - \mathcal{O}(\gamma)$ in $\mathbf{R}^n$). We use the notation $|k|_1 = \sum_{j=1}^n |k_j|$ for $k = (k_1, \dots, k_n)$. We say a vector ω satisfying (2) to be τ, γ-*Diophantine.* In this case, the optimal stability exponent seems to be $a = 1/(\tau + 1)$. This exponent has been obtained in [11], [10] and [7].

KAM theorem states, under a suitable nondegeneracy condition, that most of the n-dimensional invariant tori are preserved with some deformation in the perturbed system (1) if the size ε of the perturbation is small enough. More precisely, this preservation is guaranteed for tori that have frequency vector $\omega(I) = \operatorname{grad} h(I)$ satisfying a Diophantine condition. In this way, one gets *perpetual stability,* but only for initial conditions on a Cantorian set, which does not contain any open set although its measure is large. In fact, it was first stated by Kolmogorov [13], for analytic Hamiltonians, the preservation of one given torus, suitably chosen. Afterwards, Arnold proved in [1] (see also [26]) the existence of a large family of invariant tori and estimated the measure of the complement of the invariant set. An analogous theorem for area-preserving maps of the plane was proved by Moser [20], without the hypothesis of analiticity.

Concerning the nondegeneracy condition required for the validity of KAM theorem, two sorts of conditions are usually imposed on the unperturbed frequency map $\omega = \operatorname{grad} h$, namely the *(standard) nondegeneracy* and the *isoenergetic nondegeneracy* (see definitions (31–32) in section 4.1). There are slight differences between the statements of KAM theorem under both nondegeneracies. Indeed, in the standard case every preserved invariant torus keeps its frequency vector in the perturbation. In the isoenergetic case, the frequency vector is not usually kept but, nevertheless, every invariant torus keeps its frequency ratios and its energy and, moreover, on each fixed energy hypersurface most of the invariant tori are preserved. A well-known consequence is that, for two degrees of freedom ($n = 2$), it follows from the isoenergetic nondegeneracy the stability of the perturbed system.

The usual proofs of Nekhoroshev and KAM theorems do not allow to stress the close relationship existing between the different types of stability provided by these theorems. Actually, no use is made of the existence of the KAM tori in the proof of Nekhoroshev theorem, which gives a uniform stability time for all trajectories in phase space. These trajectories include the ones lying in KAM tori, which are the most numerous, and clearly have an infinite stability time. But one can also expect that, for a trajectory starting near a KAM torus, the stability time is much larger than the one predicted by Nekhoroshev theorem. Results concerning this "stickiness" of KAM tori, with Nekhoroshev-like estimates, have recently been obtained in [25], [19].

In this paper we are concerned about a unified approach to Nekhoroshev and KAM theorems, already announced in [8]. After a preliminary part where the common method is set up, we give quantitative proofs of Nekhoroshev theorem under the assumption of quasiconvexity (theorem D in section 3.5) and the isoenergetic KAM theorem (theorem E in section 4.4). We notice that our approach to the isoenergetic theorem is direct, unlike the usual proofs where it is deduced from the standard KAM theorem (see, for example, [4]) or making use of the associated Poincaré map (see [21]).

Moreover, under the same hypothesis of KAM theorem, we get a Nekhoroshev-like stability result (theorem F in section 4.5) which is slighlty different from the ones of [25] and [19]. The result we prove considers the invariant tori of the unperturbed system such that their associated frequency vector satisfies appoximately, up to a given precision r, a Diophantine condition. In the perturbed system, these tori survive in the form of *nearly-invariant tori,* i.e., the trajectories starting on such a torus remain near to it up to a stability time which is exponentially long in $1/r$. This result is similar to the one of [18] where, however, the estimates are expressed in terms of the stability time, which is previously fixed.

On the other hand, in [25] and [19] the stability estimates are expressed in terms of the distance to a given KAM torus, meaning in this way the "stickiness" of KAM tori. The stability time is then exponentially long in the inverse of this distance (or even "superexponentially" long for quasiconvex Hamiltonians [19]). We do not prove that

KAM tori are "sticky" but we believe that our result is more useful for practical purposes, since our estimates for nearly-invariant tori do not require the existence of a KAM torus nearby.

The method we follow for the proof of Nekhoroshev and KAM theorems is standard in classical perturbation theory. It consists of seeking for a suitable canonical transformation Ψ, bringing our Hamiltonian H into a *normal form* $H^* = H \circ \Psi$, asked to depend on less angles, none if possible. The transformation Ψ is constructed iteratively as a product of successive canonical transformations $\Phi^{(1)}, \Phi^{(2)}, \ldots$, near to the identity, which provide a sequence of Hamiltonians $H^{(1)}, H^{(2)}, \ldots$ coming nearer and nearer to the normal form.

We construct the successive canonical transformations with the help of the well-known *Lie series formalism,* which we describe in section 2.1. This is a very suitable procedure for practical applications, since it allows to carry out explicit computations in concrete examples. The procedure can be directly implemented in computers, since we only use harmonics of finite order.

It is well-known that an obstruction for the construction of the normal form is found on the resonances or near them. A *resonant manifold* is characterized by a given module $\mathcal{M} \subset \mathbf{Z}^n$:

$$S_{\mathcal{M}} := \{ I \in \mathcal{G} \ : \ \ k \cdot \omega(I) = 0 \ \ \forall k \in \mathcal{M} \}.$$

The obstruction comes from the presence of the *small divisors* $k \cdot \omega(I)$, with $k \in \mathcal{M}$, which can be zero or too small. It is because of the presence of the small divisors that one considers, near the resonance $S_{\mathcal{M}}$, a *resonant normal form,* which accepts dependence on combinations of angles $k \cdot \phi$, with $k \in \mathcal{M}$. The union of all resonances is dense in the set of frequencies, but one only needs to consider resonances up to a given suitable finite *order:* $|k|_1 \le K$, since it turns out that the effect of higher-order resonances is exponentially small in K. Thus, we say a function $g(\phi, I)$ to be in *normal form with respect to $\mathcal{M}$ of degree* K if its Fourier series expansion in the angular variables is restricted to the form

$$g(\phi, I) = \sum_{\substack{k \in \mathcal{M} \\ |k|_1 \le K}} g_k(I)\, e^{i\,k\cdot\phi}.$$

We express this by writing $g \in \mathbf{R}(\mathcal{M}, K)$. Note that a fuction is in normal form with respect to the trivial module $\mathcal{M} = 0$ if it does not depend on the angular variables.

In the first part of this paper, we restrict ourselves to a subset $G \subset \mathcal{G}$, where the frequency vector $\omega(I)$ is allowed to satisfy resonance relations corresponding to a fixed module $\mathcal{M}$, but a neighborhood of all other resonances of order less or equal than K are excluded. For such a set we say that $\omega(G)$ is *nonresonant modulo $\mathcal{M}$ up to order K* (see definition (16) in section 2.3). On this set G we make successive reductions to the type of normal form defined above: the harmonics corresponding to integer vectors satisfying $k \notin \mathcal{M}$, $|k|_1 \le K$, become smaller and smaller in the successive Hamiltonians, whereas the harmonics corresponding to $k \in \mathcal{M}$ or $|k|_1 > K$ have to be kept because in the set G the small divisors $k \cdot \omega(I)$ associated to these harmonics are not avoided. In this way the final Hamiltonian $H^* = H \circ \Psi$ can be written as

$$H^*(\phi, I) = h(I) + Z^*(\phi, I) + R^*(\phi, I),$$

where $Z^* \in \mathbf{R}(\mathcal{M}, K)$ and the remainder R^* is exponentially small in K.

As in [24] and [2], the proof of Nekhoroshev theorem is divided two parts, usually named analytic and geometric ones. The analytic part concerns the iterative process and the estimates for the successive remainders. In the geometric part the whole action space $\mathcal{G}$ is covered by a family of sets associated to every module in order to get stability estimates for the trajectories corresponding to all initial conditions. A similar distinction may also be carried out in the proof of KAM theorem. In this case the geometric part concerns the estimates for the measure of the complement of the invariant set.

For given $\mathcal{M}$ and K, we consider a subset $G \subset \mathcal{G}$ where the nonresonance condition quoted above is satisfied. In section 2.1, we show how the Lie series method is applied to the construction of the iterative process, which is finite for the proof of Nekhoroshev theorem and infinite for KAM theorem. In the case of KAM theorem we always take $\mathcal{M} = 0$. The Iterative Lemma (theorem A in section 2.3), which is common to both proofs, provides estimates for one given step of this process. We make use of a *vectorfield norm,* introduced in section 2.2, which allows us to optimize the estimates of the Iterative Lemma with respect to other related papers.

From successive application of the Iterative Lemma, with a fixed K, and carrying out an appropiate number of iterative steps, we get the Normal Form Theorem (theorem B in section 3.1), where the estimate obtained for the remainder is exponentially small in K, and hence H^* is an approximate normal form, specific for the set G. This theorem completes the analytic part of the proof of Nekhoroshev theorem. We point out that this approach is carried out along the lines of Pöschel's proof [27] of Nekhoroshev theorem, but our proof is somewhat simpler because the Iterative Lemma has been optimized.

From the Normal Form Theorem, one can deduce stability estimates for the trajectories starting in $\mathbf{T}^n \times G$, which hold up to an exponentially long time. The estimates for nonresonant ($\mathcal{M} = 0$) and resonant ($\mathcal{M} \neq 0$) regions are given in sections 3.2 and 3.3, respectively. In the resonant case we impose a quasiconvexity condition. Following [27], in the geometric part of the proof, the whole domain $\mathcal{G}$ is covered by a family of sets $G = G_{\mathcal{M}}$ associated to the different modules $\mathcal{M} \subset \mathbf{Z}^n$, with suitably chosen parameters (section 3.4). One sees that it suffices to consider K*-modules* (a module $\mathcal{M} \subset \mathbf{Z}^n$ is said to be a K-module if it is generated by vectors of order less or equal than K). To complete the proof of Nekhoroshev theorem (section 3.5), with optimal exponents, we choose K as a suitable function of ε and apply the stability estimates to each set $G_{\mathcal{M}}$. In this way, we obtain estimates for all trajectories starting in $\mathbf{T}^n \times \mathcal{G}$.

As an additional application of the Normal Form Theorem, we also consider a perturbation of a system of n harmonic oscillators with a Diophantine frequency vector. The nonresonant estimates of the case $\mathcal{M} = 0$ give rise to effective stability in such a system (theorem C in section 3.2).

Our approach to KAM theorem is parallel, in its main lines, to the Arnold's one [1]. We first prove the Inductive Lemma (proposition 11 in section 4.3), which concerns the estimates given by the Iterative Lemma, with $\mathcal{M} = 0$, for one given step of the iterative process. In this case, it does not suffice to bring our Hamiltonian H to an approximate normal form with an exponentially small remainder. It is necessary to perform an infinite number of iterations, with orders $K_1, K_2, \ldots$ increasing to infinity. Then, the resonances up to higher and higher orders are removed from the domain along the successive iterative steps. In this way, the remainders tend quickly to zero and the final Hamiltonian becomes integrable: $H^*(\phi, I) = h^*(I)$. Therefore, the domain where the transformation holds is filled with n-dimensional invariant tori with linear flow, but it shrinks to a Cantorian set corresponding to Diophantine frequencies. To finish the proof of KAM theorem, the measure of the invariant set can be estimated assuming that a suitable nondegeneracy condition is fulfilled by the unperturbed system.

In section 4.4, we give this direct proof of KAM theorem under the hypothesis that the unperturbed frequency map ω is isoenergetically nondegenerate. We point out that the same scheme would be useful for the standard nondegeneracy. An explanation of both nondegeneracy conditions and the technical difficulties arising in the isoenergetic case is given in section 4.1 (quantitative lemmas are provided in section 4.2). This common approach to both nondegeneracy conditions can be seen as a first step towards the proof of KAM theorem under higher-order nondegeneracy conditions. We recall that a very general condition has been announced by Rüssmann [28]. See [29], [5], [30] for very recent results along this line.

Finally, we see in section 4.5 that, inside the same iterative scheme used for KAM theorem but stopping it at an appropiate step, instead of carrying it to the limit, we find that Nekhoroshev-like estimates hold for the trajectories starting in the domain at this step. This domain is then filled with nearly-invariant tori (theorem F). This result is quantitatively very close to KAM theorem. Qualitatively, the perpetual stability of KAM tori is sacrificed but, on the other hand, the domain where the result holds contains inner points, and hence it is not a Cantorian set.

It is worth reminding that KAM theorem is meaningless from a practical point of view despite its theoretical importance. This is due to the fact that, from an approximation of a concrete frequency vector, one cannot decide whether this vector is Diophantine or not. The notion of nearly-invariant torus may be understood as an attempt to compensate this deficiency.

Acknowledgements The authors would like to thank G. Benettin, A. I. Neishtadt, J. Pöschel, C. Simó and D. V. Treschev for their very useful comments, suggestions and general interest.

2 The common part

2.1 Normal forms via the Lie series method

We describe in this section the iterative process leading our Hamiltonian $H(\phi, I) = h(I) + f(\phi, I)$ to normal form. This setup provides a common environment for the proofs of Nekhoroshev and KAM theorems. According to the approach described in the introduction, we restrict our Hamiltonian H to a subset $G \subset \mathcal{G}$, where it is assumed that the frequency set $\omega(G)$ is nonresonant modulo $\mathcal{M}$ up to order K for given $\mathcal{M}$ and K. For notational convenience, we consider the starting Hamiltonian H written, on the set G, in the form

$$H(\phi, I) = h(I) + Z(\phi, I) + R(\phi, I), \tag{3}$$

with $Z \in \mathbf{R}(\mathcal{M}, K)$. For instance, we may choose $Z = 0$ and $R = f$. However, if the starting Hamiltonian is already near to the normal form, we may write it in the form (3), with a small R with respect to Z, and seek for a better approximation to the normal form.

The transformation Ψ leading to normal form is constructed as a product of canonical transformations $\Phi^{(1)}, \Phi^{(2)}, \ldots$ We put $\Psi^{(q)} = \Phi^{(1)} \circ \cdots \circ \Phi^{(q)}$. At the step q, the transformed Hamiltonian is written in the form

$$H^{(q)} = H \circ \Psi^{(q)} = h + Z^{(q)} + R^{(q)},$$

with $Z^{(q)} \in \mathbf{R}(\mathcal{M}, K)$. Obviously we start with $Z^{(0)} = Z$ and $R^{(0)} = R$.

Now, to describe a concrete iterative step, we write H, Z, R, $\tilde{Z}$, $\tilde{R}$, Φ, instead of $H^{(q-1)}$, $Z^{(q-1)}$, $R^{(q-1)}$, $Z^{(q)}$, $R^{(q)}$, $\Phi^{(q)}$, respectively. Following the Lie series method, as in [7], we construct Φ as the flow at time 1 associated to a generating Hamiltonian W to be determined.

More precisely, if Φ_t denotes the flow at time t of an autonomous Hamiltonian W, it is known from the Hamiltonian theory that, for any function f, the derivative of $f \circ \Phi_t$ with respect to t can be expressed in terms of the Poisson bracket of f and W:

$$\frac{d}{dt}(f \circ \Phi_t) = \{f, W\} \circ \Phi_t.$$

So, assuming analyticity in t and taking the Taylor expansion, one has the Lie series

$$f \circ \Phi_t = \sum_{m=0}^{\infty} \frac{t^m}{m!} L_W^m f,$$

where we denote $L_W^0 f = f$ and $L_W^m f = \left\{ L_W^{m-1} f, W \right\}$ for $m \geq 1$. For the remainders of the Lie series, we use the notation

$$r_m(f, W, t) := f \circ \Phi_t - \sum_{l=0}^{m-1} \frac{t^l}{l!} L_W^l f = \sum_{l=m}^{\infty} \frac{t^l}{l!} L_W^l f \tag{4}$$

for $m \geq 0$.

With this notation, we have for the transformed Hamiltonian the following expression:

$$H \circ \Phi = h + Z + R + \{h, W\} + r_2(h, W, 1) + r_1(Z + R, W, 1). \tag{5}$$

We want $\tilde{R}$ to be smaller than R in order to get $H \circ \Phi$ closer to normal form than H. Consequently W should be chosen in such a way that $R + \{h, W\}$ be in normal form. As a matter of fact, this can only be guaranteed up to order K because the nonresonance condition on $\omega(G)$ does not avoid the small divisors corresponding to higher orders. Thus, we seek for $\Delta Z \in \mathbf{R}(\mathcal{M}, K)$ and W solving the linear functional equation

$$\{W, h\} + \Delta Z = R_{\leq K}\,, \tag{6}$$

where we write $R_{\leq K}(\phi, I) = \sum_{|k|_1 \leq K} R_k(I)\, e^{i\, k \cdot \phi}$.

The resolution of equation (6) is standard. In terms of Fourier coefficients, we have the solution

$$\begin{array}{lll} W_k(I) = \dfrac{R_k(I)}{i\, k \cdot \omega(I)}\,, & \Delta Z_k(I) = 0, & \text{for } k \in \mathbf{Z}^n \setminus \mathcal{M},\ |k|_1 \leq K; \\ W_k(I) = 0, & \Delta Z_k(I) = R_k(I), & \text{for } k \in \mathcal{M},\ |k|_1 \leq K; \\ W_k(I) = 0, & \Delta Z_k(I) = 0, & \text{for } |k|_1 > K. \end{array} \tag{7}$$

This is the only solution of equation (6) if we require W to have no resonant terms with respect to $\mathcal{M}$ and to be of degree K. We denote by $\mathbf{NR}(\mathcal{M},K)$ the set of the functions satisfying these requirements. If h and R are real functions, we see from (7) that ΔZ and W are also real.

The new Hamiltonian can be put as

$$H\circ\Phi = h+\tilde{Z}+\tilde{R},$$

with

$$\begin{aligned}\tilde{Z} &= Z+\Delta Z \in \mathbf{R}(\mathcal{M},K), &(8)\\ \tilde{R} &= R_{>K}+r_2(h,W,1)+r_1(Z+R,W,1), &(9)\end{aligned}$$

where we write $R_{>K}=R-R_{\le K}$. If h, Z and R are real, then the transformation Φ preserves real domains, and the new Hamiltonian is also real. Recall that the algorithm explicited in (7–9) is just one step of the iterative process. It describes how to get $H^{(q)}=H^{(q-1)}\circ\Phi^{(q)}$.

Roughly speaking, this procedure can be considered as linear if we ignore the term $R^{(q-1)}_{>K}$. Indeed, if $Z^{(q-1)}=\mathcal{O}(\varepsilon)$, $R^{(q-1)}=\mathcal{O}(\varepsilon^q)$, we see from equations (8–9) that $R^{(q)}=\mathcal{O}(\varepsilon^{q+1})$, since the generating Hamiltonian for $\Phi^{(q)}$ is taken of the same order as $R^{(q-1)}$. We use this procedure for the proof of Nekhoroshev theorem (section 3.5). The term $R^{(q-1)}_{>K}$ is exponentially small in K. So its influence can be overcome by choosing K large enough.

We remark that the canonical transformation $\Phi^{(q)}$ could also be constructed by means of a quadratic procedure: if $R^{(q-1)}=\mathcal{O}\left(\varepsilon^{2^{q-1}}\right)$, then $R^{(q)}=\mathcal{O}\left(\varepsilon^{2^q}\right)$. We can attain this aim by taking another term of the Lie series in (5), which gives rise to an alternative algorithm for the reduction to normal form. However, for an arbitrary module $\mathcal{M}$, the linear equation substituting (6) is not easily resoluble, and an approximate solution does not seem to improve the estimates of Nekhoroshev theorem.

In the case of KAM theorem (section 4), where $\mathcal{M}=0$, the linear procedure described in (7–9) is almost quadratic, provided we take $Z^{(q-1)}=0$ at each step (the procedure can never be strictly quadratic because of the presence of small divisors). This forces us to a little change in the algorithm: we include $\Delta Z^{(q-1)}$ in the integrable part in order to have $Z^{(q)}=0$ for the following step. In this way, the integrable part changes at every step: we begin the step q with $H^{(q-1)}(\phi,I)=h^{(q-1)}(I)+R^{(q-1)}(\phi,I)$, and the new Hamiltonian can be written as

$$H^{(q)}(\phi,I)=h^{(q)}(I)+R^{(q)}(\phi,I), \qquad (10)$$

where, by (7),

$$h^{(q)}=h^{(q-1)}+\Delta Z^{(q-1)}=h^{(q-1)}+R_0^{(q-1)} \qquad (11)$$

(note that the function $R_0^{(q-1)}(I)$ is the angular average of $R^{(q-1)}(\phi,I)$). The new remainder $R^{(q)}$ is obtained like in (9), which then gives a fastly convergent procedure if the term $R^{(q-1)}_{>K}$ is ignored. To solve the difficulty caused by this term, we take increasing orders K_q, tending to infinity for $q\to\infty$. We can then see the convergence to zero of the remainders $R^{(q)}$, and hence the existence of invariant tori. However, resonances of successive higher orders have to be removed along the procedure and hence the final domain is reduced to a Cantorian set.

Another remark is that in section 4.4 we prove KAM theorem without showing explicitly that the remainders $R^{(q)}$ converge in a fast way, but linear. Nevertheless, we use the almost quadratic convergence of the remainders to show the existence of nearly-invariant tori, with exponential estimates (see section 4.5).

2.2 A norm for Hamiltonian vectorfields

In order to obtain rigorous estimates for the successive remainders, we need to define norms for the functions taking part in the iterative process introduced in section 2.1.

An important remark is that a Hamiltonian function H does not take part directly in the Hamiltonian equations, but rather its derivative

$$DH = \left(\frac{\partial H}{\partial \phi} \, , \, \frac{\partial H}{\partial I} \right) = \left(\frac{\partial H}{\partial \phi_1} \, , \, \ldots \, , \, \frac{\partial H}{\partial \phi_n} \, , \, \frac{\partial H}{\partial I_1} \, , \, \ldots \, , \, \frac{\partial H}{\partial I_n} \right).$$

Then, to obtain the stability estimates leading to the proof of Nekhoroshev and KAM theorems, we do not need to obtain estimates for the successive remainders provided by (9), but estimates for the derivatives of these remainders suffice.

Looking carefully at equations (8–9), one realizes that it is possible to bound the derivatives $D\tilde{Z}$ and $D\tilde{R}$ from the derivatives DZ and DR, since the Lie remainders r_1, r_2 have been defined in (4) in terms of Poisson brackets. So it would be a nice tool to work with a suitable vectorfield norm for the derivatives, which would avoid unnecessary uses of the Cauchy inequalities in estimating derivatives. This idea was suggested to us by A. I. Neishtadt, although it goes back to [10], where estimates for the Lie series method for not necessarily Hamiltonian vectorfields are fully developed.

However, we cannot avoid all uses of the Cauchy inequalities, since the remainders r_1, r_2 in (9) have to be differentiated in order to estimate $D\tilde{R}$. Moreover, a differentiation has to be done before starting the first iterative step. Thus, we work with analytic functions on complex neighborhoods of the domain $\mathbf{T}^n \times G$. Given $\rho = (\rho_1, \rho_2) \geq 0$ (i.e. $\rho_j \geq 0$, $j = 1, 2$), we first introduce the sets:

$$\begin{aligned} \mathcal{W}_{\rho_1}(\mathbf{T}^n) &:= \{\phi \ : \ \operatorname{Re}\phi \in \mathbf{T}^n, \ |\operatorname{Im}\phi|_\infty \leq \rho_1\}, \\ \mathcal{V}_{\rho_2}(G) &:= \{I \in \mathbf{C}^n \ : \ |I - I'| \leq \rho_2 \text{ for some } I' \in G\}, \end{aligned}$$

where $|\cdot|_\infty$ and $|\cdot| = |\cdot|_2$ denote, respectively, the maximum norm and the Euclidean norm for vectors. We then define:

$$\mathcal{D}_\rho(G) := \mathcal{W}_{\rho_1}(\mathbf{T}^n) \times \mathcal{V}_{\rho_2}(G).$$

Several kinds of norms are used along this paper. First, we consider functions of the n action variables. Given a (real or complex) function $f(I)$, defined on a complex neighborhood $\mathcal{V}_\eta(G)$, $\eta \geq 0$, we introduce the supremum norm:

$$|f|_{G,\eta} := \sup_{I \in \mathcal{V}_\eta(G)} |f(I)| \, , \qquad |f|_G := |f|_{G,0} \ .$$

In this way, the subscript η is removed from the notation if $\eta = 0$. This remark applies throughout this section.

In an analogous way, we consider the supremum norm for vector-valued functions, i.e. vectorfields. Given $F : \mathcal{V}_\eta(G) \longrightarrow \mathbf{C}^n$ and $1 \le p \le \infty$, we define

$$|F|_{G,\eta,p} := \sup_{I \in \mathcal{V}_\eta(G)} |F(I)|_p \ , \qquad |F|_{G,\eta} := |F|_{G,\eta,2} \ .$$

In this definition, $|\cdot|_p$ means the p-norm for vectors in $\mathbf{C}^n$, i.e.: $|v|_p = \left(\sum_{j=1}^n |v_j|^p\right)^{1/p}$ for $1 \le p < \infty$, and $|v|_\infty = \max_{1\le j\le n} |v_j|$. Note that we remove the subscript p to mean the Euclidean norm ($p = 2$).

We also define the supremum norm for matrix-valued fuctions or even tensor-valued functions (e.g. successive total derivatives of a function). The definition is analogous, taking for matrices and tensors the norm induced by the Euclidean norm for vectors (we only consider the case $p = 2$).

Next we consider functions of the action–angle variables. For a given complex function $f(\phi, I)$ (2π-periodic in ϕ) defined on the neighborhood $\mathcal{D}_\rho(G)$, $\rho = (\rho_1, \rho_2) \ge 0$, we may consider its supremum norm:

$$|f|_{G,\rho} := \sup_{(\phi,I)\in\mathcal{D}_\rho(G)} |f(\phi, I)| \, . \tag{12}$$

But if f is *analytic* on (a neighborhood of) the set $\mathcal{D}_\rho(G)$, we may define an exponentially weighted norm in terms of the Fourier series of f. Writing $f(\phi, I) = \sum_{k\in\mathbf{Z}^n} f_k(I)\, e^{i\,k\cdot\phi}$, we introduce

$$\|f\|_{G,\rho} := \sum_{k\in\mathbf{Z}^n} |f_k|_{G,\rho_2} \cdot e^{|k|_1 \rho_1}. \tag{13}$$

Note that $|f|_{G,\rho} \le \|f\|_{G,\rho}$. This exponentially weighted norm, analogous to the one used in [27], allows to carry out easily a separate control of harmonics in estimating the solution of the linear functional equation (6), in proposition 4 of section 2.3. This would be more difficult by using the supremum norm.

Exactly in the same way as before we may extend the definitions of the norms (12–13) to the case of vector-valued functions. Given $F : \mathcal{D}_\rho(G) \longrightarrow \mathbf{C}^n$ and $1 \le p \le \infty$, and writing $F(\phi, I) = \sum_{k\in\mathbf{Z}^n} F_k(I)\, e^{i\,k\cdot\phi}$, where $F_k : \mathcal{V}_{\rho_2}(G) \longrightarrow \mathbf{C}^n$, we define

$$\|F\|_{G,\rho,p} := \sum_{k\in\mathbf{Z}^n} |F_k|_{G,\rho_2,p} \cdot e^{|k|_1\rho_1}, \qquad \|F\|_{G,\rho} := \|F\|_{G,\rho,2} \ .$$

Let us recall the Cauchy inequalities for the ϕ-derivatives and the I-derivatives (see also [27]). Given f analytic on $\mathcal{D}_\rho(G)$, for $0 < \delta < \rho$ (i.e. $0 < \delta_j < \rho_j$, $j = 1, 2$) one has

$$\left\|\frac{\partial f}{\partial \phi}\right\|_{G,(\rho_1-\delta_1,\rho_2),1} \le \frac{1}{e\delta_1}\,\|f\|_{G,\rho} \ , \qquad \left\|\frac{\partial f}{\partial I}\right\|_{G,(\rho_1,\rho_2-\delta_2),\infty} \le \frac{1}{\delta_2}\,\|f\|_{G,\rho} \ .$$

To have a more compact writing and to avoid to carry out separate estimates for the ϕ-derivatives and the I-derivatives along the iterative process, we introduce for $Df = \left(\frac{\partial f}{\partial\phi}\, , \, \frac{\partial f}{\partial I}\right)$ the vectorfield norm

$$\|Df\|_{G,\rho,c} := \max\left(\left\|\frac{\partial f}{\partial \phi}\right\|_{G,\rho,1} , \ c\left\|\frac{\partial f}{\partial I}\right\|_{G,\rho,\infty}\right), \tag{14}$$

where $c > 0$ is a parameter to be fixed in subsequent sections. This parameter (having the physical dimension of the action variables) is introduced in order to compensate the difference between the Cauchy inequalities for ϕ-derivatives and I-derivatives.

Lemma 1 *Let f, g be analytic functions on $\mathcal{D}_\rho(G)$. For $0 < \delta = (\delta_1, \delta_2) < \rho$ and $c > 0$ given, let us denote*

$$\hat{\delta}_c := \min(c\,\delta_1, \delta_2).$$

Then,

a) $\|Df\|_{G,\rho-\delta,c} \leq \dfrac{c}{\hat{\delta}_c} \|f\|_{G,\rho}$.

b) $\|\{f, g\}\|_{G,\rho} \leq \dfrac{2}{c} \|Df\|_{G,\rho,c} \cdot \|Dg\|_{G,\rho,c}$.

c) $\|D\,(f_{>K})\|_{G,(\rho_1-\delta_1,\rho_2),c} \leq e^{-K\delta_1} \cdot \|Df\|_{G,\rho,c}$.

The proof of the properties contained in this lemma is very simple. In subsequent sections, we shall see that an appropiate choice for the parameter c makes possible to obtain better estimates, even in case of very different δ_1, δ_2.

More notation is required. At every step of the iterative process described in section 2.1, the canonical transformation Φ leading our Hamiltonian to normal form is constructed as a flow associated to the generating Hamiltonian W defined in (7). To know how near to the identity map this canonical transformation is, we need to define a norm for maps like $\Phi - \mathrm{id}$. This map is defined in $\mathcal{D}_\rho(G)$, and we may consider it taking values in $\mathbf{C}^{2n}$. First, we take the parameter $c > 0$ of definition (14) and, for a $2n$-vector $x = (\phi, I)$, we introduce its "c-norm" as

$$|x|_c := \max\left(c\,|\phi|_\infty, |I|\right).$$

Then, for a map $\Upsilon : \mathcal{D}_\rho(G) \longrightarrow \mathbf{C}^{2n}$, we define the norms:

$$|\Upsilon|_{G,\rho,c} := \sup_{x\in\mathcal{D}_\rho(G)} |\Upsilon(x)|_c\,, \qquad |D\Upsilon|_{G,\rho,c} := \sup_{x\in\mathcal{D}_\rho(G)} |D\Upsilon(x)|_c\,,$$

where, for the second definition, the matrix c-norm is the one induced by the c-norm for vectors. With these notations, it is easy to prove the following property: if Υ is analytic on $\mathcal{D}_\rho(G)$, then

$$|D\Upsilon|_{G,\rho-\delta,c} \leq \frac{|\Upsilon|_{G,\rho,c}}{\hat{\delta}_c}. \tag{15}$$

In the following lemma, the effect of the flow associated to a generating Hamiltonian is estimated in terms of the norms introduced above. Moreover, a bound for the remainder of a Lie series is found. The proof is given in section 5.

Lemma 2 *Let W be an analytic function on $\mathcal{D}_\rho(G)$, $\rho > 0$, and Φ_t its associated Hamiltonian flow at time t, $t \geq 0$. Let $\delta = (\delta_1, \delta_2) > 0$ and $c > 0$ given. Assume that $\|DW\|_{G,\rho,c} \leq \hat{\delta}_c$. Then, Φ_t maps $\mathcal{D}_{\rho-t\delta}(G)$ into $\mathcal{D}_\rho(G)$ and one has:*

a) $|\Phi_t - \mathrm{id}|_{G,\rho-t\delta,c} \leq t\,\|DW\|_{G,\rho,c}$.

b) $\Phi_t(\mathcal{D}_{\rho'}(G)) \supset \mathcal{D}_{\rho'-t\delta}(G)$ *for* $\rho' \le \rho - t\delta$.

c) *Assuming that* $\|DW\|_{G,\rho,c} < \hat{\delta}_c/2e$, *for any given function* f, *analytic on* $\mathcal{D}_\rho(G)$, *and for any integer* $m \ge 0$, *the following bound holds:*

$$\begin{aligned} \|r_m(f,W,t)\|_{G,\rho-t\delta} &\le \left[\sum_{l=0}^{\infty} \frac{1}{\binom{l+m}{m}} \cdot \left(\frac{2e\,\|DW\|_{G,\rho,c}}{\hat{\delta}_c}\right)^l\right] \cdot \frac{t^m}{m!}\,\|L_W^m f\|_{G,\rho} \\ &= \gamma_m\left(\frac{2e\,\|DW\|_{G,\rho,c}}{\hat{\delta}_c}\right) \cdot t^m\,\|L_W^m f\|_{G,\rho}\ , \end{aligned}$$

where, for $0 \le x < 1$, *we define*

$$\gamma_m(x) := \sum_{l=0}^{\infty} \frac{l!}{(l+m)!}\, x^l.$$

2.3 The Iterative Lemma

Now we are going to obtain estimates for one step of the procedure of section 2.1, with the help of the norm introduced in section 2.2. We consider the Hamiltonian (3), real analytic on $\mathcal{D}_\rho(\mathcal{G})$, with $Z \in \mathbf{R}(\mathcal{M},K)$, and we restrict it to a subset $G \subset \mathcal{G}$ such that $\omega(G)$ is nonresonant modulo $\mathcal{M}$ up to order K (see the introduction).

We first introduce, following [27], a quantitative version of this nonresonance condition. Given a module $\mathcal{M}$, an integer K and $\alpha > 0$, a subset F of the n-dimensional frequency space is said to be *α, K-nonresonant modulo $\mathcal{M}$* if

$$|k\cdot v| \ge \alpha \qquad \forall k \in \mathbf{Z}^n \setminus \mathcal{M},\ |k|_1 \le K,\ \ \forall v \in F. \tag{16}$$

We begin by seeing that this nonresonance condition on the set $\omega(G)$ can be extended to a complex neighborhood of small enough radius ρ_2.

Lemma 3 *Let* $h(I)$ *be a real analytic function on* $\mathcal{V}_{\rho_2}(G)$, *let* $\omega = \operatorname{grad} h$, *and assume that* $\omega(G)$ *is* α, K*-nonresonant modulo* $\mathcal{M}$. *Assume that* $\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \le M$. *If*

$$\rho_2 \le \frac{\alpha}{2MK}\ , \tag{17}$$

then $\omega\left(\mathcal{V}_{\rho_2}(G)\right)$ *is* $\frac{\alpha}{2}, K$*-nonresonant modulo* $\mathcal{M}$.

The proof is a simple application of the mean value theorem. We point out that, as shown in section 3.5, condition (17) on ρ_2 imposes an important restriction on the domain. An exception is the very special case of a system of harmonic oscillators, where $M = 0$ (see section 3.2).

The next result provides estimates for the functions ΔZ and W solving the linear functional equation (6).

Proposition 4 *Let $h(I)$, $Z(\phi, I)$, $R(\phi, I)$ be real analytic functions on $\mathcal{D}_\rho(G)$, let $\omega = \operatorname{grad} h$, and assume that $\omega(G)$ is α, K-nonresonant modulo $\mathcal{M}$, and that $Z \in \mathbf{R}(\mathcal{M}, K)$. Assume that $\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \le M$, and $\rho_2 \le \dfrac{\alpha}{2MK}$. Let $c > 0$ given. Then the functions $\Delta Z \in \mathbf{R}(\mathcal{M}, K)$ and $W \in \mathbf{NR}(\mathcal{M}, K)$ given in (7), which solve the linear equation (6), are both real analytic on $\mathcal{D}_\rho(G)$, and the following bounds hold:*

$$\|D(\Delta Z)\|_{G,\rho,c} \le \|DR\|_{G,\rho,c}\,, \quad \|D(R - \Delta Z)\|_{G,\rho,c} \le \|DR\|_{G,\rho,c}\,,$$
$$\|DW\|_{G,\rho,c} \le \frac{2A}{\alpha}\|DR\|_{G,\rho,c}\,,$$

where we define

$$A := 1 + \frac{2Mc}{\alpha}\,. \tag{18}$$

Proof We obtain the estimates from the explicit solution given in (7), in terms of Fourier coefficients. The two first ones are clear, since ΔZ and $R - \Delta Z$ are obtained from R just removing the appropiate Fourier harmonics. To estimate DW, we bound $\frac{\partial W}{\partial \phi}$ and $\frac{\partial W}{\partial I}$. Using lemma 3, it is easy to see that

$$\left\|\frac{\partial W}{\partial \phi}\right\|_{G,\rho,1} \le \frac{2}{\alpha}\left\|\frac{\partial R}{\partial \phi}\right\|_{G,\rho,1}.$$

Next we write, for $k \in \mathbf{Z}^n \setminus \mathcal{M}$, $|k|_1 \le K$,

$$\frac{\partial W_k}{\partial I} = \frac{\frac{\partial R_k}{\partial I}}{i\,k\cdot\omega(I)} - \frac{R_k(I)\,\frac{\partial}{\partial I}(i\,k\cdot\omega(I))}{(i\,k\cdot\omega(I))^2} = \frac{\frac{\partial R_k}{\partial I}}{i\,k\cdot\omega(I)} + \frac{\left[\frac{\partial R}{\partial \phi}\right]_k \cdot \frac{\partial \omega}{\partial I}}{(k\cdot\omega(I))^2}\,,$$

where we have used that $\left[\frac{\partial R}{\partial \phi}\right]_k = i\,R_k(I)\,k$ (differentiating the Fourier expansion of R). From lemma 3, we obtain

$$\left|\frac{\partial W_k}{\partial I}\right|_{G,\rho_2,\infty} \le \frac{2}{\alpha}\left|\frac{\partial R_k}{\partial I}\right|_{G,\rho_2,\infty} + \frac{4M}{\alpha^2}\left|\left[\frac{\partial R}{\partial \phi}\right]_k\right|_{G,\rho_2}.$$

Thus,

$$\left\|\frac{\partial W}{\partial I}\right\|_{G,\rho,\infty} \le \frac{2}{\alpha}\left\|\frac{\partial R}{\partial I}\right\|_{G,\rho,\infty} + \frac{4M}{\alpha^2}\left\|\frac{\partial R}{\partial \phi}\right\|_{G,\rho}$$

and finally

$$\|DW\|_{G,\rho,c} \le \left(\frac{2}{\alpha} + \frac{4Mc}{\alpha^2}\right)\|DR\|_{G,\rho,c} = \frac{2A}{\alpha}\|DR\|_{G,\rho,c}\,. \quad \square$$

Remarks

1. These estimates do not involve a reduction of the domain $\mathcal{D}_\rho(G)$. This becomes more difficult if we use a norm that does not take into account the explicit expansion in Fourier series (for example, the supremum norm). One exception is the case $\dim \mathcal{M} = n - 1$, i.e. near periodic orbits, where integral expressions for the solution of equation (6) are available (see [15] and also [14]).

2. The value of A could be big (of the order of $1/\alpha$). Therefore, it would be an obstruction to the obtainment of the optimal exponent, unless we chose c small. But we shall see in the subsequent sections that our choice of c allows to bound A by a constant not depending on α.

Theorem A (Iterative Lemma) *Let $H(\phi, I) = h(I) + Z(\phi, I) + R(\phi, I)$ real analytic on $\mathcal{D}_\rho(G)$, let $\omega = \operatorname{grad} h$, and assume that $\omega(G)$ is α, K-nonresonant modulo $\mathcal{M}$, and that $Z \in \mathbf{R}(\mathcal{M}, K)$. Assume that $\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \leq M$. Let $\delta < \rho$ and $c > 0$ given, and let A defined as in (18). Assume:*

$$\rho_2 \leq \frac{\alpha}{2MK}\ , \qquad \|DR\|_{G,\rho,c} \leq \frac{\alpha\hat{\delta}_c}{74A}\ . \tag{19}$$

Then, there exists a real analytic canonical transformation $\Phi : \mathcal{D}_{\rho-\frac{\delta}{2}}(G) \longrightarrow \mathcal{D}_\rho(G)$ such that $H \circ \Phi = h + \tilde{Z} + \tilde{R}$, with $\tilde{Z} \in \mathbf{R}(\mathcal{M}, K)$, and one has:

a) $\left\|D\tilde{Z}\right\|_{G,\rho,c} \leq \|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c}\ .$

b) $\left\|D\tilde{R}\right\|_{G,\rho-\delta,c} \leq e^{-K\delta_1} \cdot \|DR\|_{G,\rho,c} + \dfrac{14A}{\alpha\hat{\delta}_c}\left(\|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c}\right) \cdot \|DR\|_{G,\rho,c}\ .$

c) $|\Phi - \mathrm{id}|_{G,\rho-\frac{\delta}{2},c} \leq \dfrac{2A}{\alpha}\|DR\|_{G,\rho,c}\ .$

d) $\Phi(\mathcal{D}_{\rho'}(G)) \supset \mathcal{D}_{\rho'-\frac{\delta}{2}}(G)$ *for* $\rho' \leq \rho - \frac{\delta}{2}$.

Proof We take ΔZ, W and Φ as constructed in section 2.1. Then the bounds of proposition 4 for $D(\Delta Z)$, $D(R - \Delta Z)$ and DW hold. In particular,

$$\|DW\|_{G,\rho,c} \leq \frac{2A}{\alpha}\|DR\|_{G,\rho,c} \leq \frac{\hat{\delta}_c}{37} < \frac{\hat{\delta}_c}{4e}\ ,$$

and therefore lemma 2 applies, with $t = 1$ and with $\delta/2$ instead of δ. We obtain $\Phi : \mathcal{D}_{\rho-\frac{\delta}{2}}(G) \longrightarrow \mathcal{D}_\rho(G)$ and expressions (8–9) hold for the transformed Hamiltonian.

From (8) and proposition 4, we easily get estimate (a). On the other hand, from (9) and parts (a) and (c) of lemma 1,

$$\left\|D\tilde{R}\right\|_{G,\rho-\delta,c} \leq e^{-K\delta_1} \cdot \|DR\|_{G,\rho,c} + \frac{2c}{\hat{\delta}_c}\left(\|r_2(h, W, 1)\|_{G,\rho-\frac{\delta}{2}} + \|r_1(Z + R, W, 1)\|_{G,\rho-\frac{\delta}{2}}\right).$$

From part (c) of lemma 2,

$$\|r_2(h, W, 1)\|_{G,\rho-\frac{\delta}{2}} \leq \gamma_2\left(\frac{4e\|DW\|_{G,\rho,c}}{\hat{\delta}_c}\right) \cdot \|\{\{h, W\}, W\}\|_{G,\rho}\ ,$$

$$\|r_1(Z + R, W, 1)\|_{G,\rho-\frac{\delta}{2}} \leq \gamma_1\left(\frac{4e\|DW\|_{G,\rho,c}}{\hat{\delta}_c}\right) \cdot \|\{Z + R, W\}\|_{G,\rho}\ .$$

We estimate the Poisson brackets using part (b) of lemma 1:

$$\|\{Z+R,W\}\|_{G,\rho} \le \frac{2}{c}\left(\|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c}\right)\cdot\|DW\|_{G,\rho,c} \ ,$$
$$\|\{\{h,W\},W\}\|_{G,\rho} = \|\{\Delta Z - R_{\le K},W\}\|_{G,\rho} \le \frac{2}{c}\|DR\|_{G,\rho,c}\cdot\|DW\|_{G,\rho,c}$$

where, in the second estimate, we have used proposition 4 to ensure that

$$\|D(\Delta Z - R_{\le K})\|_{G,\rho,c} \le \|D(\Delta Z - R)\|_{G,\rho,c} \le \|DR\|_{G,\rho,c} \ .$$

For $0 < x < 1$, one has

$$\gamma_1(x) = -\frac{\ln(1-x)}{x} \ , \qquad \gamma_2(x) = \frac{x + (1-x)\ln(1-x)}{x^2} \ .$$

Using that these functions are increasing and evaluating them at $x = 4e/37$, we obtain

$$\begin{aligned}
&\|r_2(h,W,1)\|_{G,\rho-\frac{\delta}{2}} + \|r_1(Z+R,W,1)\|_{G,\rho-\frac{\delta}{2}} \\
\le\ & \frac{2}{c}\left(\gamma_2\left(\frac{4e}{37}\right) + \gamma_1\left(\frac{4e}{37}\right)\right)\cdot\left(\|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c}\right)\cdot\|DW\|_{G,\rho,c} \\
\le\ & \frac{7A}{c\alpha}\left(\|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c}\right)\cdot\|DR\|_{G,\rho,c} \ . \qquad (20)
\end{aligned}$$

By putting all of these estimates together, we get (b). Finally, we deduce from lemma 2 (with $\delta/2$ instead of δ) the statements (c) and (d), concerning the distance from Φ to the identity. □

Remarks

1. The Iterative Lemma provides a description for one step of the transformation to normal form constructed in section 2.1. The improvement of this result, with respect to related papers (for instance, [27]), is the main contribution of the vectorfield norm (14). It avoids a subsequent application of the Cauchy inequalities, which would cause an extra division by $\hat{\delta}_c$ in estimate (b).

2. In the statement of the Iterative Lemma the value of the parameter c is still free. From now onwards, we shall take

$$c = \frac{\delta_2}{\delta_1} \ ,$$

and hence $\hat{\delta}_c = \delta_2$. This choice of c seems to be the best because it leads to the smallest possible value for the quotient

$$\frac{\|DR\|_{G,\rho,c}}{\hat{\delta}_c} \ ,$$

appearing implicitly in condition (19) and estimate (b).

3 Nekhoroshev estimates and related results

3.1 Estimates for the normal form

Now, starting with $H(\phi, I) = h(I) + Z(\phi, I) + R(\phi, I)$ on $\mathcal{D}_\rho(G)$, we apply Q times the Iterative Lemma and obtain an estimate for the remainder. By choosing $Q = Q(K)$ adequately, we get an exponentially small remainder in the next theorem.

Theorem B (Normal Form Theorem) *Let $H(\phi, I) = h(I) + Z(\phi, I) + R(\phi, I)$ real analytic on $\mathcal{D}_\rho(G)$, let $\omega = \operatorname{grad} h$, and assume that $\omega(G)$ is α, K-nonresonant modulo $\mathcal{M}$, and that $Z \in \mathbf{R}(\mathcal{M}, K)$. Assume that $\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \leq M$. Let $\delta < \rho$ given, $c = \delta_2/\delta_1$, and let A the constant defined in (18). Assume:*

$$\rho_2 \leq \frac{\alpha}{2MK}\,, \qquad \|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c} \leq \frac{\alpha\delta_2}{61AK\delta_1}\,. \tag{21}$$

Then, there exists a real analytic canonical transformation $\Psi : \mathcal{D}_{\rho-\delta}(G) \longrightarrow \mathcal{D}_\rho(G)$ such that $H \circ \Psi = h + Z^* + R^*$, with $Z^* \in \mathbf{R}(\mathcal{M}, K)$, and one has:

a) $\|DZ^*\|_{G,\rho-\delta,c} + \|DR^*\|_{G,\rho-\delta,c} \leq \|DZ\|_{G,\rho,c} + 2\,\|DR\|_{G,\rho,c}\,.$

b) $\|DR^*\|_{G,\rho-\delta,c} \;\leq\; 3\,e^{-\frac{K\delta_1}{2}} \cdot \|DR\|_{G,\rho,c}\,.$

c) $|\Psi - \mathrm{id}|_{G,\rho-\delta,c} \leq \dfrac{4A}{\alpha}\,\|DR\|_{G,\rho,c}\,.$

d) $\Psi\left(\mathcal{D}_{\rho'}(G)\right) \supset \mathcal{D}_{\rho'-\frac{\delta}{2}}(G)$ *for* $\rho' \leq \rho - \delta$.

Proof Let $Q \geq 1$ be an integer to be chosen below, and let us introduce the sequence

$$\rho^{(q)} = \rho - \frac{q\delta}{Q}\,, \quad 0 \leq q \leq Q.$$

We take $\delta^{(q)} = \delta/Q$ for every $1 \leq q \leq Q$. Next we shall construct a sequence of real analytic canonical transformations $\Phi^{(q)} : \mathcal{D}_{\rho^{(q)}}(G) \longrightarrow \mathcal{D}_{\rho^{(q-1)}}(G)$, $1 \leq q \leq Q$. Denoting $\Psi^{(q)} = \Phi^{(1)} \circ \cdots \circ \Phi^{(q)}$, the successive transformed Hamiltonians will be written in the form $H^{(q)} = H \circ \Psi^{(q)} = h + Z^{(q)} + R^{(q)}$, with $Z^{(q)} \in \mathbf{R}(\mathcal{M}, K)$. Moreover, we are going to show that, if

$$K\delta_1 \geq 2Q, \tag{22}$$

the following statements are true for $0 \leq q \leq Q$:

1_q) $\left\|DZ^{(q)}\right\|_{G,\rho^{(q)},c} \leq \|DZ\|_{G,\rho,c} + \displaystyle\sum_{s=0}^{q-1} \left\|DR^{(s)}\right\|_{G,\rho^{(s)},c}\,.$

2_q) $\left\|DR^{(q)}\right\|_{G,\rho^{(q)},c} \leq \dfrac{1}{e^q}\,\|DR\|_{G,\rho,c}\,.$

We proceed by induction. The results are obviously true for $q = 0$. For $1 \le q \le Q$, note that, by (2_{q-1}) and condition (21),

$$\left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \le \frac{1}{e^{q-1}} \left\| DR \right\|_{G,\rho,c} \le \frac{\alpha\delta_2}{61AK\delta_1} \le \frac{\alpha\delta_2}{122AQ}$$

and hence the Iterative Lemma applies, with δ/Q instead of δ, and we obtain the canonical transformation $\Phi^{(q)}$. We immediately get (1_q). The bound (2_q) comes from the following estimate:

$$\begin{aligned}
&\left\| DR^{(q)} \right\|_{G,\rho^{(q)},c} \\
&\quad \le e^{-\frac{K\delta_1}{Q}} \cdot \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \\
&\qquad + \frac{14AQ}{\alpha\delta_2} \left(\left\| DZ^{(q-1)} \right\|_{G,\rho^{(q-1)},c} + \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \right) \cdot \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \\
&\quad \le \left(\frac{1}{e^2} + \frac{28AQ}{\alpha\delta_2} \left(\left\| DZ \right\|_{G,\rho,c} + \left\| DR \right\|_{G,\rho,c} \right) \right) \cdot \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \\
&\quad \le \left(\frac{1}{e^2} + \frac{28}{122} \right) \cdot \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \le \frac{1}{e} \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} .
\end{aligned}$$

Now, let us assume that $K\delta_1 \ge 2$ (if $K\delta_1 < 2$, all results are obvious if we take Ψ as the identity map). Then, we may choose $Q = Q(K)$ as the maximum integer satisfying (22), i.e. $Q = \left[\frac{K\delta_1}{2} \right]$. Denoting $\Psi = \Psi^{(Q)}$, $Z^* = Z^{(Q)}$, $R^* = R^{(Q)}$, we have $H \circ \Psi = h + Z^* + R^*$. Then, part (a) comes from (1_Q). For part (b), we use (2_Q):

$$\left\| DR^* \right\|_{G,\rho-\delta,c} \le \frac{1}{e^Q} \left\| DR \right\|_{G,\rho,c} \le \frac{1}{e^{\frac{K\delta_1}{2}-1}} \left\| DR \right\|_{G,\rho,c} \le 3\, e^{-\frac{K\delta_1}{2}} \cdot \left\| DR \right\|_{G,\rho,c} .$$

The proof of (c) is very simple from the analogous bound in the Iterative Lemma and the inequalities (2_q):

$$\left| \Psi^{(Q)} - \mathrm{id} \right|_{G,\rho-\delta,c} \le \sum_{q=1}^{Q} \left| \Phi^{(q)} - \mathrm{id} \right|_{G,\rho^{(q)},c} \le \sum_{q=1}^{Q} \frac{2A}{\alpha} \left\| DR^{(q-1)} \right\|_{G,\rho^{(q-1)},c} \le \frac{4A}{\alpha} \left\| DR \right\|_{G,\rho,c} .$$

Finally, to get (d) it suffices to prove that, for $0 \le q \le Q$,

$$\Psi^{(q)} \left(\mathcal{D}_{\rho'}(G) \right) \supset \mathcal{D}_{\rho' - \frac{q\delta}{2Q}}(G) \qquad \text{if} \quad \rho' \le \rho - \tfrac{q\delta}{Q} .$$

Indeed, this inclusion is obvious for $q = 0$. By induction, we assume it for $q - 1$:

$$\Psi^{(q-1)} \left(\mathcal{D}_{\rho''}(G) \right) \supset \mathcal{D}_{\rho'' - \frac{(q-1)\delta}{2Q}}(G) \qquad \text{if} \quad \rho'' \le \rho - \tfrac{(q-1)\delta}{Q} .$$

Then, taking $\rho'' = \rho' - \frac{\delta}{2Q}$ in this inclusion and applying part (d) of the Iterative Lemma, we get:

$$\Psi^{(q)} \left(\mathcal{D}_{\rho'}(G) \right) = \Psi^{(q-1)} \left(\Phi^{(q)} \left(\mathcal{D}_{\rho'}(G) \right) \right) \supset \Psi^{(q-1)} \left(\mathcal{D}_{\rho' - \frac{\delta}{2Q}}(G) \right) \supset \mathcal{D}_{\rho' - \frac{q\delta}{2Q}}(G). \qquad \square$$

Remark This result is essentially equivalent to the analogous one in [27], and seems to be "optimal" in the sense that the exponent for K in the second condition of (21) is 1. The difference is that our proof is much simpler because the Iterative Lemma is also optimal, whereas the proof appearing in [27] relies in a very careful choice of the size of the successive reductions of the domain (see also [23]).

3.2 Nonresonant stability estimates: application to the harmonic oscillators case

From theorem B, one can obtain estimates for the variation of the action variables on the set G where this theorem is applied. This is very simple in a nonresonant region ($\mathcal{M} = 0$), where no extra geometric condition on the unperturbed Hamiltonian h is required.

Lemma 5 *Let $H(\phi, I) = h(I) + Z(I) + R(\phi, I)$ real analytic on $\mathcal{D}_\rho(G)$, let $\omega = \operatorname{grad} h$, and assume that $\omega(G)$ is α, K-nonresonant modulo 0. Assume that $\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \leq M$. Let $c = \rho_2/\rho_1$, and assume:*

$$\rho_2 \leq \frac{\alpha}{2MK}\,, \qquad \|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c} \leq \frac{\alpha\rho_2}{122K\rho_1}\,. \tag{23}$$

Then, for every trajectory $(\phi(t), I(t))$ of H, with $(\phi(0), I(0)) \in \mathbf{T}^n \times G$, one has

$$|I(t) - I(0)| \leq \frac{24}{\alpha}\,\|DR\|_{G,\rho,c} \qquad \textit{for} \;\; |t| \leq \frac{2}{\alpha}\, e^{\frac{K\rho_1}{6}}. \tag{24}$$

The proof is deferred to section 5. Now, as a simple application, we consider the case $h(I) = \omega \cdot I$, i.e. H is a perturbation of a system of n harmonic oscillators. The frequency vector $\omega \in \mathbf{R}^n$ is assumed to be τ, γ-Diophantine (see (2)), for $\tau \geq n-1$ and $\gamma > 0$ given. This case, where no geometric part is required, is also considered in [12], [11], [10], [27] and [7]. We obtain, like in the last four of the quoted papers, the "optimal" stability exponent $a = 1/(\tau+1)$. We remark that, since $M = 0$ in this case, condition (17) does not impose any restriction on ρ_2.

Theorem C *Let $H(\phi, I) = \omega \cdot I + f(\phi, I)$ real analytic on $\mathcal{D}_\rho(\mathcal{G})$, and assume that the vector ω is τ, γ-Diophantine for some $\tau \geq n-1$ and $\gamma > 0$. Assume:*

$$\varepsilon := \|f\|_{\mathcal{G},\rho} \;\leq\; \varepsilon_0 := \frac{\gamma\rho_2}{244}\,.$$

Then, for every trajectory $(\phi(t), I(t))$ of H, with $(\phi(0), I(0)) \in \mathbf{T}^n \times \mathcal{G}$, one has

$$|I(t) - I(0)| \leq \frac{\rho_2}{5\rho_1}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/(\tau+1)} \qquad \textit{for} \;\; |t| \leq \frac{2}{\gamma}\exp\left\{\frac{\rho_1}{24}\left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/(\tau+1)}\right\}.$$

Proof Let $c = \rho_2/\rho_1$. We notice that we may take $M = 0$ in lemma 5. For a fixed K to be chosen, the set $\{\omega\}$ is clearly $\frac{\gamma}{K^\tau}, K$-nonresonant modulo 0. We are going to apply lemma 5 with $Z = 0$, $R = f$, and $\rho/2$ instead of ρ. Since

$$\|Df\|_{\mathcal{G},\frac{\rho}{2},c} \leq \frac{2\varepsilon}{\rho_1}\,,$$

the second condition of (23) is satisfied for

$$\varepsilon \le \frac{\gamma \rho_2}{244\, K^{\tau+1}} \, .$$

We then choose $K = \left[\left(\frac{\varepsilon_0}{\varepsilon} \right)^{1/(\tau+1)} \right]$ and obtain:

$$|I(t) - I(0)| \le \frac{24 K^\tau}{\gamma} \cdot \frac{2\varepsilon}{\rho_1} \le \frac{\rho_2}{5\rho_1} \left(\frac{\varepsilon}{\varepsilon_0} \right)^{1/(\tau+1)} .$$

Concerning the stability time, it is easily obtained from the one of lemma 5 if we take into account that, since $\varepsilon \le \varepsilon_0$, we have $K \ge \frac{1}{2} \left(\frac{\varepsilon_0}{\varepsilon} \right)^{1/(\tau+1)}$. □

Remarks

1. One may notice that our results in the case of a perturbation of a system of harmonic oscillators are slightly worse than the ones obtained in [10]. Indeed, the stability exponent $a = 1/(\tau+1)$ is the same but we have obtained $b = 1/(\tau+1)$ instead of $b = 1$. This difference comes from a different performance of the iterations leading to normal form. Indeed, in [10] the linear functional equation (6) is solved without cutting the Fourier expansions at order K (but making a reduction of the domain). This approach makes the estimates of proposition 4 better, but it is limited to the nonresonant case ($\mathcal{M} = 0$). Although our approach leads to worse estimates, it avoids dealing with infinitely many small divisors, and also allows to treat the resonant case.

2. Even in the harmonic oscillators case, our approach looks more significative from a practical point of view. Indeed, if we consider ε fixed (i.e. a concrete Hamiltonian), then the result of theorem C still holds if Diophantine condition (2) is required just for $|k|_1 \le \left(\frac{\varepsilon_0}{\varepsilon} \right)^{1/(\tau+1)}$. For instance, if the frequency ω is known only up to a finite precision then it has no sense to check the Diophantine condition farther than a certain finite order, but our estimates could also be applied.

3.3 Resonant stability estimates

Now we restrict ourselves to a neighborhood of the resonance associated to a given module $\mathcal{M} \subset \mathbf{Z}^n$, and afterwards the whole domain $\mathcal{G}$ will be divided in resonant and nonresonant regions corresponding to the different modules. A set of frequencies $F \subset \mathbf{R}^n$ is said to be *η-close to $\mathcal{M}$-resonances* if

$$|v - \Pi_{\mathcal{M}} v| \le \eta \qquad \forall I \in F,$$

where $\Pi_{\mathcal{M}}$ denotes the orthogonal projection onto the space of exact $\mathcal{M}$-resonant frequencies

$$\mathcal{M}^\perp = \{ v \in \mathbf{R}^n \ : \ k \cdot v = 0 \ \ \forall k \in \mathcal{M} \} \, .$$

To obtain stability estimates for the trajectories with initial condition in a set G such that $\omega(G)$ is close to a resonance, we need to impose some geometric condition on the unperturbed Hamiltonian h. In the original Nekhoroshev's proof [24], a general steepness condition was imposed. But the main geometric ideas of the proof are contained in the simpler quasiconvex case, considered for instance in [2] (convex case), [15] and [27]. Following [15], we say the function $h(I)$ to be *m-quasiconvex* on a set U if

$$\left|\frac{\partial^2 h}{\partial I^2}(I)\,(v,v)\right| \geq m\,|v|^2 \qquad \forall v \in \langle\omega(I)\rangle^{\perp}, \quad \forall I \in U$$

(this definition is slightly different from the one given in [27]). One remarks that the level hypersurfaces of h are convex if h is m-quasiconvex. Moreover, the quasiconvexity implies that, for every module $\mathcal{M}$, the resonant manifold $S_{\mathcal{M}}$ and the vector subspace generated by $\mathcal{M}$ are always transversal.

Under this condition, the next lemma (called Resonant Stability Lemma in [27]) provides stability estimates on a region $G \subset \mathcal{G}$ such that $\omega(G)$ is assumed to be close to the resonance associated to a given module $\mathcal{M}$ and satisfying a nonresonant condition modulo $\mathcal{M}$. Our proof is standard. It follows [15] and [27] in the main ideas, which go back (for convex systems) to [3]. The basic point is that, for a Hamiltonian in normal form with respect to $\mathcal{M}$ with an exponentially small remainder, the speed of variation of the action variables along the $\mathcal{M}^{\perp}$-direction is exponentially small. On the other hand, the quasiconvexity condition forces the energy hypersurface of h passing through a point of the resonance $S_{\mathcal{M}}$ to have a contact of order two with the $\mathcal{M}$-direction. Then, by energy conservation, one may bound the variation of the actions along the $\mathcal{M}$-direction, giving rise to the stability estimate. It has to be noticed that this approach differs from the one of [2], where a different feature of (quasi)convex Hamiltonians is used: the transversality between the resonant manifold $S_{\mathcal{M}}$ and the $\mathcal{M}$-direction.

We define a real neighborhood of the domain G as

$$\mathcal{U}_{\rho_2}(G) := \{I \in \mathbf{R}^n \ : \ |I - I'| \leq \rho_2 \text{ for some } I' \in G\} = \mathcal{V}_{\rho_2}(G) \cap \mathbf{R}^n. \tag{25}$$

Lemma 6 *Let* $H(\phi, I) = h(I) + Z(\phi, I) + R(\phi, I)$ *real analytic on* $\mathcal{D}_\rho(G)$*, and let* $\omega = \operatorname{grad} h$*. For a given module* $\mathcal{M} \neq \mathbf{Z}^n$*, assume that* $\omega(G)$ *is* η*-close to* $\mathcal{M}$*-resonances, and* α, K*-nonresonant modulo* $\mathcal{M}$*, with* $K \geq 1$*, and assume also that* $Z \in \mathbf{R}(\mathcal{M}, K)$*. Assume that*

$$\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \leq M, \qquad |\omega|_G \leq L,$$

and that h *is* m*-quasiconvex on* $\mathcal{U}_{\rho_2}(G)$*. Let* $c = \rho_2/\rho_1$*, and assume:*

$$\rho_2 \leq \frac{m\alpha}{48M^2K}\,, \qquad \eta \leq \frac{m\rho_2}{60}\,, \qquad \|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c} \leq \frac{m\sigma\rho_2^2}{350}\,, \tag{26}$$

where we write $\sigma := \min\left(1, \dfrac{1}{\sqrt{n}\,\rho_1}\right)$*. Then, for every trajectory* $(\phi(t), I(t))$ *of* H*, with* $(\phi(0), I(0)) \in \mathbf{T}^n \times G$*, one has*

$$|I(t) - I(0)| \leq \rho_2 \qquad \text{for} \ \ |t| \leq \frac{m\rho_2^2}{74L\,\|DR\|_{G,\rho,c}}\, e^{\frac{mK\rho_1}{6M}}. \tag{27}$$

We give the proof of this result in section 5.

Remark If $R = 0$, this lemma implies that, if the quasiconvexity condition is fulfilled, all trajectories starting in $\mathbf{T}^n \times G$ have perpetual stability. Nevertheless, one could deduce this fact in a more direct way, because in this case the Hamiltonian $H = h + Z$ is already in normal form with respect to $\mathcal{M}$.

3.4 Geometry of resonances

Next we return to the Hamiltonian (1), which we assume real analytic on $\mathcal{D}_\rho(\mathcal{G})$. The stability estimates obtained in lemmas 5 and 6 only apply to the trajectories starting in a subset $G \subset \mathcal{G}$ where the frequencies are assumed to be close to the resonance characterized by a fixed module $\mathcal{M} \subset \mathbf{Z}^n$ and satisfying a nonresonance condition modulo $\mathcal{M}$. In order to obtain stability estimates for all trajectories starting in $\mathbf{T}^n \times \mathcal{G}$, the whole action domain $\mathcal{G}$ is covered by a family of sets $G_\mathcal{M}$, called resonant and nonresonant blocks (for $\mathcal{M} \neq 0$ and $\mathcal{M} = 0$, respectively). For each module $\mathcal{M}$, the frequencies on the block $G_\mathcal{M}$ have to be close to $\mathcal{M}$-resonances, and to satisfy the nonresonance condition (16) up to a fixed order K.

A construction of such a covering is carried out in [24] and [2]. The quantitative aspect was improved in [27]. The Geometric Lemma stated below has been taken from [27] with no changes.

Actually one may work in frequency space. We obtain for this space a covering $\{B_\mathcal{M}\}$, which can be pulled back by the frequency map ω to a covering $\{G_\mathcal{M}\}$ for $\mathcal{G}$. Before stating the Geometric Lemma we recall some concepts and terminology introduced in [27].

For each module $\mathcal{M} \subset \mathbf{Z}^n$, we consider the space of $\mathcal{M}$-resonant frequencies $\mathcal{M}^\perp$. Note that there are a lot of modules giving rise to the same resonant space. Obviously we only have to consider the *maximal* one. A module $\mathcal{M} \subset \mathbf{Z}^n$ is said to be maximal if it is not properly contained in any other module of the same dimension. See appendix 3 of the book [16] for an explicit characterization of the maximal modules in $\mathbf{Z}^n$.

Given a maximal d-dimensional module $\mathcal{M} \subset \mathbf{Z}^n$, the set $B_\mathcal{M}$ is constructed by taking a neighborhood of the space $\mathcal{M}^\perp$ and removing from it a neighborhood of the resonant spaces associated to the $(d+1)$-dimensional modules. The set constructed in this way would not contain any open set. However, one remarks that, to satisfy the nonresonance condition (16) up to order K, it suffices to consider *K-modules.* A module $\mathcal{M} \subset \mathbf{Z}^n$ is said to be a K-module if it is generated by vectors of order less or equal than K.

To make these ideas quantitative, one requires the notion of *volume* of a module. For a d-dimensional module $\mathcal{M} \subset \mathbf{Z}^n$, $1 \le d \le n$, let C be the $(n \times d)$-matrix obtained by choosing a basis of $\mathcal{M}$ and putting its vectors as columns. The volume of $\mathcal{M}$ is then defined as

$$|\mathcal{M}| := \sqrt{\det(C^\top C)},$$

i.e. the d-dimensional volume of the parallelepiped spanned by the vectors of the basis. The choice of the basis does not have influence in this definition.

Let $\lambda_d > 0$, for $1 \le d \le n$, be fixed parameters. For each maximal d-dimensional

K-module $\mathcal{M}$, one introduces

$$\eta_{\mathcal{M}} := \frac{\lambda_d}{|\mathcal{M}|} ,$$

and the resonant zone associated to $\mathcal{M}$ is defined as a neighborhood of radius $\eta_{\mathcal{M}}$ around $\mathcal{M}^{\perp}$. Recalling that $\Pi_{\mathcal{M}}$ denotes the orthogonal projection onto $\mathcal{M}^{\perp}$, one defines

$$A_{\mathcal{M}} := \{ v \in \mathbf{R}^n \ : \ |v - \Pi_{\mathcal{M}} v| < \eta_{\mathcal{M}} \} .$$

Then, the *resonant block* associated to $\mathcal{M}$ is defined as

$$B_{\mathcal{M}} := A_{\mathcal{M}} \setminus A^*_{d+1},$$

where $A^*_l, \ 1 \le l \le n,$ stands for the union of all resonant zones corresponding to maximal l-dimensional K-modules, and $A^*_{n+1} = \emptyset$. Note that every resonant block $B_{\mathcal{M}}$ is $\eta_{\mathcal{M}}$-close to $\mathcal{M}$-resonances. For the trivial module one defines the *nonresonant block:*

$$B_0 := \mathbf{R}^n \setminus A^*_1.$$

It is easy to see that the whole frequency space is covered by the blocks $B_{\mathcal{M}}$.

Lemma 7 (Geometric Lemma) *Let us fix* $K \ge 1, \ E > 0$ *and* $F \ge E + \sqrt{2}$. *Assume:*

$$\frac{\lambda_{d+1}}{\lambda_d} \ge FK$$

for $1 \le d < n$. *Then, the blocks* $B_{\mathcal{M}}$ *defined above are* $\alpha_{\mathcal{M}}, K$*-nonresonant modulo* $\mathcal{M}$*, with*

$$\alpha_{\mathcal{M}} := EK\eta_{\mathcal{M}} \quad \textit{for} \ \ \mathcal{M} \ne 0,$$
$$\alpha_0 := \lambda_1.$$

For the proof, see [27]. The desired covering for $\mathcal{G}$ is then obtained from this lemma by putting $G_{\mathcal{M}} = \omega^{-1}(B_{\mathcal{M}})$ for each maximal K-module $\mathcal{M}$, except for the ones giving rise to an empty set.

3.5 Global effective stability

Our main result on effective stability concerns estimates holding for all motions in phase space. Like in [27], these estimates are obtained by considering the covering supplied by lemma 7 with a fixed order K (which is chosen as a suitable function of the size ε of perturbation) and then applying the stability estimates to each block of the covering.

Theorem D (Nekhoroshev Theorem) *Let* $H(\phi, I) = h(I) + f(\phi, I)$ *real analytic on* $\mathcal{D}_\rho(\mathcal{G})$*, let* $\omega = \operatorname{grad} h$*, and assume that*

$$\left| \frac{\partial^2 h}{\partial I^2} \right|_{\mathcal{G}, \rho_2} \le M, \qquad |\omega|_{\mathcal{G}} \le L.$$

Assume also that h is m-quasiconvex on $\mathcal{U}_{\rho_2}(\mathcal{G})$. Let $\lambda > 0$ given, and assume:

$$\lambda \leq \frac{23M^2\rho_2}{m}, \qquad \varepsilon := \|f\|_{\mathcal{G},\rho} \;\leq\; \varepsilon_0 := \frac{m^{4n-1}\hat{\rho}\lambda^2}{2^{24n-2}M^{4n}}, \tag{28}$$

where we write $\hat{\rho} := \min\left(\rho_1, \frac{2}{\sqrt{n}}\right)$. Then, for every trajectory $(\phi(t), I(t))$ of H, with $(\phi(0), I(0)) \in \mathbf{T}^n \times \mathcal{G}$ and satisfying $|\omega(I(0))| \geq \lambda$, one has

$$|I(t) - I(0)| \leq \rho_2 \cdot \left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2n} \qquad \textit{for} \;\; |t| \leq \frac{4}{L} \exp\left\{\frac{m\rho_1}{24M}\left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/2n}\right\}. \tag{29}$$

Proof Fix $K \geq 1$ to be chosen below. Let

$$F = \frac{2882M^2}{m^2}, \qquad E = F - \sqrt{2}.$$

For $1 \leq d \leq n$, we put:

$$\lambda_d = \frac{\lambda}{(FK)^{n-d}}.$$

Then, lemma 7 provides a covering $\{G_{\mathcal{M}}\}$ of $\mathcal{G}$, with $G_{\mathcal{M}} = \omega^{-1}(B_{\mathcal{M}})$, and its parameters are

$$\eta_{\mathcal{M}} = \frac{\lambda}{|\mathcal{M}|\;(FK)^{n-d}}, \qquad \alpha_{\mathcal{M}} = \frac{E\lambda}{|\mathcal{M}|\;F^{n-d}K^{n-d-1}}$$

for every maximal d-dimensional K-module $\mathcal{M}$, $1 \leq d \leq n$, and

$$\alpha_0 = \frac{\lambda}{(FK)^{n-1}}$$

for the trivial module. We also put $\eta_0 = 0$.

We are going to apply lemma 6 with $Z = 0$ and $R = f$ on all blocks, except on the one corresponding to $\mathcal{M} = \mathbf{Z}^n$. In this way the estimates hold for all initial conditions satisfying $|\omega(I(0))| \geq \eta_{\mathbf{Z}^n} = \lambda$. Unlike the case of theorem C (harmonic oscillators), the smallness condition (17) on ρ_2 makes us restrict the domain. For every $\mathcal{M}$ we take $\rho^{(\mathcal{M})} = \left(\rho_1^{(\mathcal{M})}, \rho_2^{(\mathcal{M})}\right)$, with

$$\rho_1^{(\mathcal{M})} = \frac{\rho_1}{2}, \quad \rho_2^{(\mathcal{M})} = \frac{m\alpha_{\mathcal{M}}}{48M^2K}, \qquad c_{\mathcal{M}} = \frac{\rho_2^{(\mathcal{M})}}{\rho_1^{(\mathcal{M})}}. \tag{30}$$

For every d-dimensional K-module $\mathcal{M}$, $0 \leq d \leq n-1$, one has

$$\frac{60\lambda}{mF^{n-d}K^n} \leq \rho_2^{(\mathcal{M})} \leq \frac{61\lambda}{m(FK)^{n-d}} \leq \frac{\rho_2}{2},$$

where we used that $1 \leq |\mathcal{M}| \leq K^d$. We have

$$\|Df\|_{G_{\mathcal{M}},\rho^{(\mathcal{M})},c_{\mathcal{M}}} \leq \frac{\varepsilon}{\rho_1^{(\mathcal{M})}} \leq \frac{2\varepsilon}{\rho_1}.$$

To apply lemma 6 on $G_{\mathcal{M}}$, we must verify the three inequalities of (26). The two first ones are easily verified and the last one is fulfilled for all $\mathcal{M}$ if

$$\frac{2\varepsilon}{\rho_1} \le \frac{2\varepsilon_0}{\rho_1 K^{2n}} \le \frac{m\sigma}{350}\left(\frac{60\lambda}{m(FK)^n}\right)^2 \le \frac{m\sigma}{350}\left(\rho_2^{(\mathcal{M})}\right)^2,$$

with $\sigma = \min\left(1, \dfrac{2}{\sqrt{n}\,\rho_1}\right)$. Thus, we choose $K = \left[\left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/2n}\right]$. For every $\mathcal{M}$, the stability radius and the stability time for the trajectories starting in $\mathbf{T}^n \times G_{\mathcal{M}}$ are obtained from lemma 6:

$$\rho_2^{(\mathcal{M})} \le \frac{61\lambda}{mFK} \le \frac{122\lambda}{mF}\left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2n} \le \rho_2 \cdot \left(\frac{\varepsilon}{\varepsilon_0}\right)^{1/2n},$$

$$\frac{m\left(\rho_2^{(\mathcal{M})}\right)^2}{74L \cdot \frac{2\varepsilon}{\rho_1}} e^{\frac{mK\rho_1}{12M}} \ge \frac{4}{L}\exp\left\{\frac{m\rho_1}{24M}\left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/2n}\right\}. \quad \square$$

Remarks

1. These estimates would have been a little better if, on the nonresonant block G_0, we had used lemma 5 instead of lemma 6. But the stability exponents obtained would have been the same, so we used lemma 6 on all blocks for sake of simplicity.

2. We also point out that, actually, condition (28) on ε is not essential. It can be removed with some additional effort, but we omit the details. Note, however, that for a large ε Nekhoroshev estimate (29) is meaningless. The same remark holds for theorem C.

In the proof of theorem D, we have obtained the stability exponents

$$a = b = \frac{1}{2n}$$

by carrying out the estimates on every block $G_{\mathcal{M}}$ and always taking the worst possible case. The key point is to find greater and lower bounds for $\rho_2^{(\mathcal{M})}$, valid for all modules $\mathcal{M}$. However, the stability exponents can be improved by means of a particular analysis, if one is only interested in a given region.

In particular, one remarks the case of the nonresonant block G_0. This case corresponds to the smallest $\rho_2^{(0)}$, which gives rise to the smallest stability radius. It is not hard to see that, as stated in theorem 2 of [27] and theorem 3 of [7], the stability exponents obtained for this case are

$$a = \frac{1}{2n}, \qquad b = \frac{1}{2}.$$

However, if lemma 5 were used to obtain the stability estimate, one may check that the exponents would be

$$a = \frac{1}{2n}, \qquad b = \frac{n+1}{2n}.$$

It is also interesting to consider, for a fixed module $\mathcal{M}_0$, a neighborhood of the resonant manifold $S_{\mathcal{M}_0}$. This set can be covered by the blocks $G_{\mathcal{M}}$ associated to the modules $\mathcal{M}$ containing $\mathcal{M}_0$. If we restrict ourselves to these modules, the lower bound for $\rho_2^{(\mathcal{M})}$ is greater than the one obtained in theorem D. This allows to choose K greater, and leads to the exponents:

$$a = b = \frac{1}{2\nu_0} ,$$

where ν_0 is the codimension of $\mathcal{M}_0$. A precise statement of this result is given in theorem 3 of [27].

4 KAM theorem and nearly-invariant tori

4.1 Nondegeneracy conditions

Now our aim is to prove that, for a nearly-integrable Hamiltonian $H(\phi, I) = h(I) + f(\phi, I)$, analytic on $\mathcal{D}_\rho(\mathcal{G})$, most orbits lie in n-dimensional invariant tori if the perturbation f is small enough. To reach this result, a suitable nondegeneracy condition has to be fulfilled by unperturbed system. In the usual statements of KAM theorem, two sorts of nondegeneracy conditions are imposed on $\omega = \operatorname{grad} h$. These are the *(standard) nondegeneracy* and the *isoenergetic nondegeneracy*. The frequency map ω is said to be nondegenerate if

$$\det\left(\frac{\partial\omega}{\partial I}(I)\right) \neq 0 \qquad \forall I \in \mathcal{G}, \tag{31}$$

and isoenergetically nondegenerate if

$$\det\begin{pmatrix} \frac{\partial\omega}{\partial I}(I) & \omega(I) \\ \omega(I)^\top & 0 \end{pmatrix} \neq 0 \qquad \forall I \in \mathcal{G}. \tag{32}$$

An equivalent formulation for the isoenergetic nondegeneracy is to require that ω is non-vanishing on $\mathcal{G}$ and

$$\frac{\partial\omega}{\partial I}(I)\, v + \lambda\omega(I) \neq 0 \qquad \forall v \in \langle\omega(I)\rangle^\perp \setminus \{0\},\ \forall\lambda \in \mathbf{R},\ \ \forall I \in \mathcal{G}, \tag{33}$$

In action space, condition (33) can be interpreted as transversality, at every point, between any energy level $M_E = h^{-1}(E)$ and the hypersurfaces $\omega(I) \cdot v = 0$ (which include the resonant ones). The interpretation in frequency space is that the image $\omega\left(M_E\right)$ of any energy level and the subspace $\langle\omega(I)\rangle$ are always transversal.

It is easy to construct examples showing that conditions (31) and (32) are independent:

$$h(I_1, I_2) = \ln\frac{I_2}{I_1} , \qquad h(I_1, I_2) = \frac{1}{2}I_1^2 + I_2 . \tag{34}$$

The first one is only nondegenerate on its whole domain, and the second one is only isoenergetically nondegenerate.

We give in the next sections a direct and quantitative proof of KAM theorem under the assumption of isoenergetic nondegeneracy, although the same approach holds for the standard nondegeneracy. Our setup differs from the one which can be found in [9], [4], where the isoenergetic KAM theorem is proved from the standard one.

We first remind that the standard version of KAM theorem (under the standard condition (31)) states that, given $\tau > n-1$ and $\gamma > 0$ previously fixed, and assuming for the size of the perturbation f a smallness condition of the type

$$\varepsilon = \mathcal{O}\left(\gamma^2\right), \tag{35}$$

then every invariant torus of the unperturbed Hamiltonian h having τ, γ-Diophantine frequency (i.e. satisfying (2)) is preserved in the perturbed system with the same frequency vector. Moreover, the measure of the complement of the set filled with the invariant tori is $\mathcal{O}(\gamma)$. In proving this statement it is crucial to use that, under condition (31), the map ω is a local diffeomorphism and therefore the unperturbed invariant tori can be locally parametrized by their frequency vector.

The preservation of the invariant tori with the same frequency vector can be false under the isoenergetic condition (32). Indeed, it suffices to consider $h(I)$ as in the second example of (34), where the frequency map $\omega(I_1, I_2) = (I_1, 1)$ maps the whole plane into a straight line, and $f(\phi, I) = \varepsilon\, h(I)$ as a perturbation.

Nevertheless, it is known that in the isoenergetic case the unperturbed invariant tori can be locally parametrized on each energy level M_E by their frequency ratios. More precisely, if we assume, with no loss of generality, that the component ω_n does not vanish on $\mathcal{G}$, then the isoenergetic condition is equivalent to requiring that the map

$$\Omega(I) := \left(\frac{\overline{\omega}(I)}{\omega_n(I)}\,,\, h(I)\right) = \left(\frac{\omega_1(I)}{\omega_n(I)}\,, \ldots\,, \frac{\omega_{n-1}(I)}{\omega_n(I)}\,,\, h(I)\right) \tag{36}$$

is a local diffeomorphism on $\mathcal{G}$. We use, in this section and in the subsequent ones, the notation $\overline{v} = (v_1, \ldots, v_{n-1})$ for $v = (v_1, \ldots, v_{n-1}, v_n)$. Note that, including the last component $h(I)$ in the definition of Ω, we avoid to consider each energy level separately. Using the nondegeneracy of the map Ω, we are able to state that, if a smallness condition on ε like (35) is fulfilled, then for every τ, γ-Diophantine torus of the unperturbed Hamiltonian there exists an invariant torus of the perturbation, with the same frequency ratios (though the frequency itself can vary) and the same energy. Moreover, like in the standard case, we get that the measure of the complement of the invariant set can be estimated as $\mathcal{O}(\gamma)$. One deduces that in the isoenergetic case most of the invariant tori on any energy level are preserved under the perturbation. Indeed, since $\omega_n(I) \neq 0$ for $I \in \mathcal{G}$, the frequency vector $\omega(I)$ associated to a given torus is Diophantine if the vector $\left(\frac{\overline{\omega}(I)}{\omega_n(I)}, 1\right)$ is also Diophantine. This occurs for most of the unperturbed tori on a fixed energy level, since these tori can always be parametrized by their frequency ratios.

It is worth noting that the isoenergetic version of KAM theorem is more significative from the point of view of stability, because in this case the existence of a large family of invariant tori is ensured on every fixed energy level. For two degrees of freedom ($n = 2$), it follows the stability of the system (in the sense that all motions are bounded), since a given energy level is always separated by the invariant tori. Under the standard nondegeneracy condition, one cannot deduce from KAM theorem that on a given energy level any invariant tori is preserved. On the other hand, for more than two degrees of freedom, stability cannot be guaranteed under any nondegeneracy condition but in the isoenergetic case the preserved tori seem to be stronger barriers to Arnold diffusion.

Another remark is that the isoenergetic KAM theorem can eventually be applied to periodic nonautonomous Hamiltonians, by taking the time variable as an additional angular variable.

Let us make an outline of the method we use for the proof of KAM theorem and the main technical problems found. As established in section 2.1, we use an iterative procedure leading to successive Hamiltonians of the form $H^{(q)} = h^{(q)} + R^{(q)}$, with the integrable part $h^{(q)}$ changing from every step to the next one. The domain is restricted by removing at every step resonant strips up to successive orders K_q increasing to infinity. A first problem is that we need to guarantee the nondegeneracy condition for the frequency map $\omega^{(q)} = \operatorname{grad} h^{(q)}$ at every iteration. Hence the analytic part and the geometric part cannot be separated and must be carried out simultaneously along the proof.

Another technical problem is that, to move the bounds of the measure of the resonant zones from frequency space to phase space, we need to estimate from below the Jacobian determinant of a suitable diffeomorphism. In the standard case, the frequency map ω itself is a local diffeomorphism, and one can see that the successive maps $\omega^{(q)}$ are still diffeomorphisms on their domains. One can assume the starting map ω to be one-to-one (restricting the domain if necessary). To guarantee the perturbed maps $\omega^{(q)}$ to be also one-to-one, their domains still have to be slightly restricted after having removed resonances.

The isoenergetic case is, in this aspect, more cumbersome. As showed above, the frequency map ω is not necessarily a local diffeomorphism, but we may use the map Ω introduced in (36), which we shall assume to be one-to-one on $\mathcal{G}$. Then, like one would do in the standard case, successive perturbations $\Omega^{(q)}$ of this map will be guaranteed to be also one-to-one provided their domain is slightly restricted.

We notice that the resonant strips to be removed along the successive iterations are better expressed in the image space $\omega(\mathcal{G})$ or $\Omega(\mathcal{G})$. Indeed, the strips can then be taken as linear and are thus easier to handle. To be more precise, in the isoenergetic case we remove from $\Omega(\mathcal{G})$ resonant strips of the form

$$\Delta(k,\alpha) := \left\{ J \in \mathbf{R}^n \ : \ \left| \overline{k} \cdot \overline{J} + k_n \right| < \alpha \right\}, \tag{37}$$

with $\alpha > 0$. This is very appropiate for the isoenergetic case since Ω maps every resonant zone $|k \cdot \omega(I)| < \delta$ into such a linear strip. The only exception is the case $k = (0, \ldots, 0, 1)$, since $\Delta(k,\alpha)$ is then empty if $\alpha < 1$, but this integer vector corresponds to the resonance $\omega_n(I) = 0$, which has previously been excluded from the domain. We also remark that it is easy to bound the measure of a linear strip like (37), and that this bound can be pulled back to action space by estimating from below the Jacobian determinant of the diffeomorphism.

It is possible to construct a common environment for both the standard and the isoenergetic nondegeneracy conditions. Indeed, let us consider the condition:

$$\frac{\partial \omega}{\partial I}(I)\, v \neq 0 \qquad \forall v \in \langle \omega(I) \rangle^{\perp} \setminus \{0\}, \quad \forall I \in \mathcal{G}, \tag{38}$$

which means that the restriction of ω to each energy level M_E is a local diffeomorphism. It is not hard to check that this condition is equivalent to that, at every point $I \in \mathcal{G}$, the standard or the isoenergetic condition holds.

It is known that the measure of the set of vectors which do not satisfy Diophantine condition (2), with given $\tau > n-1$ and γ, is $\mathcal{O}(\gamma)$. Note also that the nondegeneracy condition (38) implies that every resonance $k \cdot \omega(I) = 0$, with $k \neq 0$, is a regular hypersurface. Then, the measure of the complement of the invariant set is also $\mathcal{O}(\gamma)$, since this estimate is obtained by pulling back to action space the measure of every resonant strip removed along the iterative process. This is the idea of the measure estimates for both nondegeneracy conditions (31) and (32), and the only reason for carrying out the proofs separately is the technical problem concerning the appropiate diffeomorphisms described above.

A higher-order nondegeneracy condition has been announced by Rüssmann [28]: the Taylor expansion of ω at a point I is required to contain n linearly independent coefficients. As a matter of fact, it is easy to see (using that $\frac{\partial\omega}{\partial I}(I)$ is a symmetric matrix) that condition (38) is equivalent to imposing that n of the vectors $\omega(I), \frac{\partial\omega}{\partial I_1}(I), \ldots, \frac{\partial\omega}{\partial I_n}(I)$ are linearly independent, namely Rüssmann's condition at order 1. As said in [28], KAM theorem remains true even under the most general Rüssmann's condition. However, the estimate one would get for the measure of the complement of the invariant set would then be larger: $\mathcal{O}(\gamma^b)$, with $b < 1$. Very recent results along these lines can be found in [29], [5] and [30].

4.2 Isoenergetic nondegeneracy: quantitative results

We give in this section several quantitative results related to the isoenergetic nondegeneracy. All proofs are technical and we postpone them to section 5.

We need to work with a quantitative version of the isoenergetic nondegeneracy in its form (33). If $h(I)$ is defined on a set $G \subset \mathbf{R}^n$, we say the frequency map $\omega = \operatorname{grad} h$ to be *μ-isoenergetically nondegenerate* if ω does not vanish on G and

$$\left| \frac{\partial\omega}{\partial I}(I)\, v + \lambda\omega(I) \right| \geq \mu\, |v| \qquad \forall v \in \langle \omega(I) \rangle^{\perp},\ \forall \lambda \in \mathbf{R},\quad \forall I \in G.$$

Moreover, we modify slightly the map defined in (36). For a fixed constant $a > 0$, we define

$$\Omega_{\omega,h,a}(I) := \left(\frac{\overline{\omega}(I)}{\omega_n(I)}\,,\ a\, h(I) \right), \qquad I \in G. \tag{39}$$

The constant a is introduced for quantitative reasons. It has the appropiate dimensions to make the components of the map (39) dimensionally coherent. But its main motivation is that the estimates given in the next lemma are better with a good choice of a.

Lemma 8 *Let h be a real function of class $\mathcal{C}^3$ on $G \subset \mathbf{R}^n$, and $\omega = \operatorname{grad} h$. Assume the bounds:*

$$\left| \frac{\partial^2 h}{\partial I^2} \right|_G \leq M, \qquad \left| \frac{\partial^3 h}{\partial I^3} \right|_G \leq M', \qquad |\omega|_G \leq L \qquad \text{and} \qquad |\omega_n(I)| \geq l \ \ \forall I \in G.$$

Assume also that ω is μ-isoenergetically nondegenerate on G. Let $a \geq 2M/l^2$ be a fixed constant, and denote $\Omega = \Omega_{\omega,h,a}$. One has:

a) $\left|\dfrac{\partial\Omega}{\partial I}\right|_G \le 2La.$

b) $\left|\dfrac{\partial\Omega}{\partial I}(I)\,v\right| \ge \dfrac{\mu}{2L}\,|v| \quad \forall v\in\mathbf{R}^n,\ \forall I\in G.$

c) $\left|\det\left(\dfrac{\partial\Omega}{\partial I}(I)\right)\right| \ge \dfrac{\mu^{n-1}a}{L^{n-2}} \quad \forall I\in G.$

d) $\left|\dfrac{\partial^2\Omega}{\partial I^2}\right|_G \le \left(\dfrac{M'}{2M}+\dfrac{3M}{l}\right)La.$

Next we state how a slight variation of the frequency map affects the constant μ of the isoenergetic nondegeneracy condition. This result may be expressed in terms of vectors and matrices.

Lemma 9 *Let $\omega,\tilde\omega\in\mathbf{R}^n$, and let $A,\tilde A$ be $(n\times n)$-matrices. Let $\varepsilon=|\tilde\omega-\omega|$, $\varepsilon'=\left|\tilde A-A\right|$, and define $l=\min\left(|\omega|,|\tilde\omega|\right)$, $M=\max\left(|A|,\left|\tilde A\right|\right)$. For some $\mu>0$, assume that*

$$|Av+\lambda\omega|\ge\mu\,|v| \qquad \forall v\in\langle\omega\rangle^{\perp},\quad \forall\lambda\in\mathbf{R}.$$

Then,

$$\left|\tilde Av+\lambda\tilde\omega\right|\ge\left(\mu-\frac{4M\varepsilon}{l}-\varepsilon'\right)|v| \qquad \forall v\in\langle\tilde\omega\rangle^{\perp},\quad \forall\lambda\in\mathbf{R}.$$

Finally, we see that a small perturbation of a one-to-one map is still one-to-one provided the domain is slightly restricted. Previously, for $b\ge 0$ we define the set

$$G-b:=\left\{I\in G\ :\ \mathcal{U}_b(I)\subset G\right\},$$

where $\mathcal{U}_b(I)$ means the closed ball of radius b centered at I, according to (25).

Lemma 10 *Let $G\subset\mathbf{R}^n$ a compact, and let $\Omega,\tilde\Omega:G\longrightarrow\mathbf{R}^n$ maps of class $\mathcal{C}^2$, with $\left|\tilde\Omega-\Omega\right|_G\le\varepsilon$. Assume that Ω is one-to-one on G, and let $F=\Omega(G)$. Assume the bounds:*

$$\left|\frac{\partial\Omega}{\partial I}\right|_G\le M,\qquad \left|\frac{\partial\tilde\Omega}{\partial I}\right|_G\le\tilde M,\qquad \left|\frac{\partial^2\tilde\Omega}{\partial I^2}\right|_G\le\tilde M',$$
$$\left|\frac{\partial\Omega}{\partial I}(I)\,v\right|\ge m\,|v|,\qquad \left|\frac{\partial\tilde\Omega}{\partial I}(I)\,v\right|\ge\tilde m\,|v| \quad \forall v\in\mathbf{R}^n,\quad \forall I\in G,$$

with $0<\tilde m<m$, $\tilde M>M$. Assume also that

$$\varepsilon\le\frac{\tilde m^2}{4\tilde M'}\,. \tag{40}$$

Then, given a subset $\tilde F\subset F-\dfrac{4M\varepsilon}{\tilde m}$ and writing $\tilde G=\left(\tilde\Omega\right)^{-1}\left(\tilde F\right)$, the map $\tilde\Omega$ is one-to-one from $\tilde G$ to $\tilde F$, and one has the inclusions

$$\tilde G\subset G-\frac{2\varepsilon}{\tilde m}\,,\qquad \Omega\left(\tilde G\right)\supset\tilde F-\varepsilon.$$

Moreover, the following estimate holds:

$$\left|\left(\tilde{\Omega}\right)^{-1} - \Omega^{-1}\right|_{\tilde{F}} \leq \frac{\varepsilon}{m} .$$

4.3 Analytic and geometric estimates for one step

We provide in proposition 11 below quantitative estimates for one concrete step of the iterative process described in section 2.1 for the isoenergetic KAM theorem. A parallel result for the standard theorem is called Inductive Lemma in [1].

Let us outline what we call the "analytic part". If the starting Hamiltonian in a step of the iterative process is written in the form $H(\phi, I) = h(I) + R(\phi, I)$, we get, from the Iterative Lemma (section 2.3), estimates for the new Hamiltonian $\tilde{H}(\phi, I) = \tilde{h}(I) + \tilde{R}(\phi, I)$. Assuming the starting frequencies $\omega(I)$ to be Diophantine up to a given order K, we ensure the Iterative Lemma to apply with $\mathcal{M} = 0$. To be more precise, the Diophantine condition up to order K is slightly modified (see condition (41) below) in view of the resonant strips introduced in (37). In this way, we obtain the estimates of parts (a–e) of proposition 11, which could be stated without the hypothesis of isoenergetic nondegeneracy.

On the other hand, some "geometric part" is required. Indeed, assuming that ω is isoenergetically nondegenerate on the domain G, we need to guarantee that the new frequency map $\tilde{\omega} = \operatorname{grad} \tilde{h}$ is also isoenergetically nondegenerate, with a new parameter, in order to allow further iterations. This is established in part (f) of proposition 11. Moreover, if the map $\Omega_{\omega,h,a}$ introduced in (39) is one-to-one, we see in part (g) that the new map $\Omega_{\tilde{\omega},\tilde{h},a}$ is also one-to-one on a new domain $\tilde{G} \subset G$.

Proposition 11 (Inductive Lemma) *Let $G \subset \mathbf{R}^n$ a compact, and $H(\phi, I) = h(I) + R(\phi, I)$ real analytic on $\mathcal{D}_\rho(G)$. Let $\omega = \operatorname{grad} h$, and assume the bounds:*

$$\left|\frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2} \leq M, \qquad |\omega|_G \leq L \qquad \text{and} \qquad |\omega_n(I)| \geq l \ \ \forall I \in G.$$

Assume also that ω is μ-isoenergetically nondegenerate on G. Let $\tilde{M} > M$, $\tilde{L} > L$, $\tilde{l} < l$ and $\tilde{\mu} < \mu$ given. For a fixed constant $a \geq 2\tilde{M}/\tilde{l}^2$, assume that the map $\Omega = \Omega_{\omega,h,a}$ is one-to-one on G, and let $F = \Omega(G)$. For $\tau > 0$, $0 < \beta \leq 1$ and K given, K an integer, assume the nonresonance condition:

$$F \cap \Delta\left(k, \frac{\beta}{|k|_1^\tau}\right) = \emptyset \qquad \forall k = \left(\overline{k}, k_n\right) \in \mathbf{Z}^n,\ |k|_1 \leq K,\ \overline{k} \neq 0. \tag{41}$$

Let $\delta < \rho$ given, $c = \delta_2/\delta_1$ and

$$A = 1 + \frac{2McK^\tau}{l\beta} .$$

Let $\varepsilon := \|DR\|_{G,\rho,c}$, $\eta := |R_0|_{G,\rho_2}$ and $\xi := \left|\frac{\partial R_0}{\partial I}\right|_{G,\rho_2}$ (where $R_0(I)$ is the angular average of $R(\phi, I)$), and assume:

$$\rho_2 \leq \frac{\tilde{l}\beta}{2\tilde{M}K^{\tau+1}} , \tag{42}$$

$$\varepsilon \leq \frac{l\beta\delta_2}{74AK^\tau}\,, \qquad \xi \leq \min\left(\left(\tilde{M}-M\right)\delta_2\,,\ \tilde{L}-L\,,\ l-\tilde{l}\,,\ \frac{(\mu-\tilde{\mu})\,\rho_2}{3}\right), \tag{43}$$

$$\eta' := \frac{L\xi}{2\tilde{M}} + \eta \leq \frac{\tilde{\mu}^2(\rho_2-\delta_2)}{32\tilde{L}^3a^2}\,. \tag{44}$$

Then, there exists a real analytic canonical transformation $\Phi : \mathcal{D}_{\rho-\frac{\delta}{2}}(G) \longrightarrow \mathcal{D}_\rho(G)$ *and a descomposition* $H\circ\Phi = \tilde{h}(I) + \tilde{R}(\phi, I)$ *such that, writing* $\tilde{\omega} = \operatorname{grad}\tilde{h}$ *and* $\tilde{\Omega} = \Omega_{\tilde{\omega},\tilde{h},a}$, *one has:*

a) $|\tilde{\omega}-\omega|_{G,\rho_2} = \xi, \quad \left|\tilde{h}-h\right|_{G,\rho_2} = \eta.$

b) $\tilde{\varepsilon} := \left\|D\tilde{R}\right\|_{G,\rho-\delta,c} \leq e^{-K\delta_1}\cdot\varepsilon + \dfrac{14AK^\tau}{l\beta\delta_2}\cdot\varepsilon^2.$

c) $\tilde{\eta} := \left|\tilde{R}_0\right|_{G,\rho_2-\frac{\delta_2}{2}} \leq \dfrac{7AK^\tau}{c\,l\beta}\cdot\varepsilon^2.$

d) $|\Phi - \mathrm{id}|_{G,\rho-\frac{\delta}{2},c} \leq \dfrac{2AK^\tau}{l\beta}\cdot\varepsilon\,.$

e) $\left|\dfrac{\partial^2\tilde{h}}{\partial I^2}\right|_{G,\rho_2-\delta_2} \leq \tilde{M}, \quad |\tilde{\omega}|_G \leq \tilde{L}$ *and* $|\tilde{\omega}_n(I)| \geq \tilde{l}\ \ \forall I \in G.$

f) $\tilde{\omega}$ *is* $\tilde{\mu}$*-isoenergetically nondegenerate on* G.

g) *Given a subset* $\tilde{F} \subset F - \dfrac{16L\tilde{L}a^2\eta'}{\tilde{\mu}}$ *and writing* $\tilde{G} = \left(\tilde{\Omega}\right)^{-1}\left(\tilde{F}\right)$, *the map* $\tilde{\Omega}$ *is one-to-one from* $\tilde{G}$ *to* $\tilde{F}$, *and one has the inclusions*

$$\tilde{G} \subset G - \frac{4\tilde{L}a\eta'}{\tilde{\mu}}\,, \qquad \Omega\left(\tilde{G}\right) \supset \tilde{F} - a\eta'.$$

Moreover, the following estimates hold:

$$\left|\tilde{\Omega}-\Omega\right|_G \leq a\eta', \qquad \left|\left(\tilde{\Omega}\right)^{-1} - \Omega^{-1}\right|_{\tilde{F}} \leq \frac{2La\eta'}{\mu}\,.$$

Proof By condition (41) we have, for every $I \in G$ and $0 < |k|_1 \leq K$, $\overline{k} \neq 0$,

$$|k\cdot\omega(I)| \geq \frac{\beta}{|k|_1^\tau}\,|\omega_n(I)| \geq \frac{l\beta}{K^\tau}\,.$$

This estimate holds even if $\overline{k} = 0$ since $|\omega_n(I)| \geq l$ and $\beta \leq 1$. Then, the set $\omega(G)$ is $\frac{l\beta}{K^\tau}$, K-nonresonant modulo 0. This fact and conditions (42) and (43) on ρ_2 and ε allow to apply the Iterative Lemma with $Z = 0$, $\mathcal{M} = 0$ and $\alpha = \frac{l\beta}{K^\tau}$. We obtain the canonical transformation Φ and, according to (10–11), the new Hamiltonian may be written as $H\circ\Phi = \tilde{h} + \tilde{R}$, where $\tilde{h} = h + \tilde{Z}$, and we have $\tilde{Z} = R_0$. One sees the estimates of (a) using that

$$|\tilde{\omega}-\omega|_{G,\rho_2} = \left|\frac{\partial R_0}{\partial I}\right|_{G,\rho_2} = \xi. \tag{45}$$

The estimates of (b) and (d) are provided directly by the Iterative Lemma. The bound (c) comes from the inequality (20) in the Iterative Lemma. The estimates of part (e) come from (45), condition (43) on ξ and the Cauchy inequality

$$\left|\frac{\partial^2 \tilde{h}}{\partial I^2} - \frac{\partial^2 h}{\partial I^2}\right|_{G,\rho_2-\delta_2} \leq \frac{\xi}{\delta_2} .$$

We prove (f) using lemma 9, which tells us that $\tilde{\omega}$ is μ'-isoenergetically nondegenerate on G, where

$$\mu' = \mu - \frac{4\tilde{M}}{\tilde{l}} \left|\tilde{\omega} - \omega\right|_G - \left|\frac{\partial^2 \tilde{h}}{\partial I^2} - \frac{\partial^2 h}{\partial I^2}\right|_G \geq \mu - \frac{4\tilde{M}}{\tilde{l}} \cdot \xi - \frac{\xi}{\rho_2} \geq \mu - \frac{3\xi}{\rho_2} \geq \tilde{\mu}.$$

We have used (45), condition (43) on ξ, the inequality

$$\left|\frac{\partial^2 \tilde{h}}{\partial I^2} - \frac{\partial^2 h}{\partial I^2}\right|_G \leq \frac{\xi}{\rho_2}$$

and the inequality $\rho_2 \leq \tilde{l}/2\tilde{M}$, deduced from (42).

Before going on, we remark that the bounds on the derivatives of Ω given in lemma 8 are also valid for $\tilde{\Omega}$, on G, provided one replaces M, L, l, μ, by $\tilde{M}$, $\tilde{L}$, $\tilde{l}$, $\tilde{\mu}$, respectively. We do not need to change the constant a because we are assuming that $a \geq 2\tilde{M}/\tilde{l}^2$. We shall now apply lemma 10 but we have to take, instead of the parameters of that lemma M, m, $\tilde{M}$, $\tilde{m}$, the values obtained applying lemma 8 to Ω and to $\tilde{\Omega}$. We first find the value which will replace ε in lemma 10, i.e. an estimate for $\left|\tilde{\Omega} - \Omega\right|_G$. We have

$$\begin{aligned}
\left|\overline{\tilde{\Omega}}(I) - \overline{\Omega}(I)\right| &\leq \frac{|\tilde{\omega}_n(I) - \omega_n(I)| \cdot |\overline{\omega}(I)| + \left|\overline{\tilde{\omega}}(I) - \overline{\omega}(I)\right| \cdot |\omega_n(I)|}{|\omega_n(I)| \cdot |\tilde{\omega}_n(I)|} \\
&\leq \frac{|\tilde{\omega}(I) - \omega(I)| \cdot |\omega(I)|}{|\omega_n(I)| \cdot |\tilde{\omega}_n(I)|} \leq \frac{L\xi}{l\tilde{l}} , \\
\left|\tilde{\Omega}_n(I) - \Omega_n(I)\right| &\leq a \left|\tilde{h}(I) - h(I)\right| \leq a\eta.
\end{aligned}$$

Therefore,

$$\left|\tilde{\Omega} - \Omega\right|_G \leq \sqrt{\left(\frac{L\xi}{l\tilde{l}}\right)^2 + (a\eta)^2} \leq a\eta',$$

where the condition on a has been used. Then, lemma 10 applies if the next inequality (which replaces (40) of that lemma) is fulfilled:

$$a\eta' \leq \frac{\left(\frac{\tilde{\mu}}{2\tilde{L}}\right)^2}{4\left(\frac{\tilde{M}'}{2\tilde{M}} + \frac{3\tilde{M}}{\tilde{l}}\right)\tilde{L}a} , \tag{46}$$

where we have taken into account the inequality

$$\left|\frac{\partial^3 \tilde{h}}{\partial I^3}\right|_G \leq \tilde{M}' := \frac{\tilde{M}}{\rho_2 - \delta_2} .$$

It is easy to check that the inequality (46) is guaranteed by condition (44) on η'. Then, lemma 10 gives part (g). □

4.4 Invariant tori

In order to estimate the measure of the resonant strips which we remove along the successive steps, a reasonable condition has to be imposed on the domain. Given $F \subset \mathbf{R}^n$ and $D > 0$ we say F to be a *D-set* if, for any $0 \le b_1 < b_2$,

$$\operatorname{mes}\left[(F - b_1) \setminus (F - b_2)\right] \le D(b_2 - b_1).$$

We remark that the constant D is a rough upper bound of the "area" of the boundary of F, which is forced to be finite.

The next technical lemma provides the necessary estimates for the measure when resonant strips of the type introduced in (37) are removed from F. We point out that it suffices to remove resonant strips corresponding to integer vectors $k = \left(\overline{k}, k_n\right) \in \mathbf{Z}^n$ with $\overline{k} \neq 0$, since it is assumed that $\omega_n(I) \neq 0$ throughout the domain. The proof of this result is deferred to section 5.

Lemma 12 *Let $F \subset \mathbf{R}^n$ a D-set. For $d \ge 0$, $\tau > 0$, $\beta \ge 0$ and $K \ge 0$ given, K an integer, let us denote*

$$F(d, \beta, K) := (F - d) \setminus \bigcup_{\substack{|k|_1 \le K \\ \overline{k} \neq 0}} \Delta\left(k, \frac{\beta}{|k|_1^\tau}\right).$$

One has:

a) *Given $d' \ge d$, $\beta' \ge \beta$ and $K' \ge K$,*

$$\begin{aligned}
&\operatorname{mes}\left[F(d, \beta, K) \setminus F(d', \beta', K')\right] \\
&\qquad \le D(d' - d) + 2(\operatorname{diam} F)^{n-1} \left(\sum_{\substack{|k|_1 \le K \\ \overline{k} \neq 0}} \frac{\beta' - \beta}{|k|_1^\tau \cdot \left|\overline{k}\right|} + \sum_{\substack{K < |k|_1 \le K' \\ \overline{k} \neq 0}} \frac{\beta'}{|k|_1^\tau \cdot \left|\overline{k}\right|} \right).
\end{aligned}$$

b) *For every $b \ge 0$,*

$$\operatorname{mes}\left[F(d, \beta, K) \setminus (F(d, \beta, K) - b)\right] \le \left(D + 2^{n+1} (\operatorname{diam} F)^{n-1} K^n\right) b\,.$$

Next we give the proof of KAM theorem under the assumption of isoenergetic nondegeneracy. We remark that the basic scheme of the proof would be the same for the standard nondegeneracy.

Our approach consists, like the original Arnold's proof [1], of iterating the estimates of the Inductive Lemma, which give rise to a rapidly convergent (i.e. more than linear) procedure. However, conversely to [1], for the proof of KAM theorem we only need to show explicitly that the remainders decrease in a linear way. This approach already appears in [26]. However, in the next section (on nearly-invariant tori) we show the rapid convergence in order to obtain exponential stability estimates.

A comment has to be made on the parameter ν appearing in theorem E below. Our statement and its proof have been slightly complicated because of the presence of this parameter. Actually, the freedom on the choice of ν is not strictly necessary but we make use of it in the next section, where it is shown that a small ν gives rise to an almost quadratic procedure (with exponent $2^{1-\nu}$) and hence to better stability estimates. It is not possible to choose $\nu = 0$, which would actually provide a quadratic procedure.

Theorem E (Isoenergetic KAM Theorem) *Let $\mathcal{G} \subset \mathbf{R}^n$, $n \geq 2$, a compact, and let $H(\phi, I) = h(I) + f(\phi, I)$ real analytic on $\mathcal{D}_\rho(\mathcal{G})$. Let $\omega = \operatorname{grad} h$, and assume the bounds:*

$$\left|\frac{\partial^2 h}{\partial I^2}\right|_{\mathcal{G},\rho_2} \leq M, \qquad |\omega|_{\mathcal{G}} \leq L \qquad \text{and} \qquad |\omega_n(I)| \geq l \ \ \forall I \in \mathcal{G}.$$

Assume also that ω is μ-isoenergetically nondegenerate on $\mathcal{G}$. For $a = 16M/l^2$, assume that the map $\Omega = \Omega_{\omega,h,a}$ is one-to-one on $\mathcal{G}$, and that its range $F = \Omega(\mathcal{G})$ is a D-set. Let $\tau > n-1$, $\gamma > 0$ and $0 < \nu < 1$ given, and assume:

$$\varepsilon := \|f\|_{\mathcal{G},\rho} \ \leq \ \frac{\nu^2 l^6 \mu^2 \hat{\rho}^{2\tau+2}}{2^{4\tau+32} L^6 M^3} \cdot \gamma^2, \qquad \gamma \leq \min\left(\frac{8LM\rho_2}{\nu l \hat{\rho}^{\tau+1}}\,,\, l\right), \tag{47}$$

where we write $\hat{\rho} := \min\left(\dfrac{\nu\rho_1}{12(\tau+2)}\,,\, 1\right)$. *Define the set*

$$\widehat{G} = \widehat{G}_\gamma := \left\{ I \in \mathcal{G} - \frac{2\gamma}{\mu} \ : \ \omega(I) \ \textit{is}\ \tau,\gamma\textit{-Diophantine} \right\}.$$

Then, there exists a real continuous map $\mathcal{T} : \mathcal{W}_{\frac{\rho_1}{4}}(\mathbf{T}^n) \times \widehat{G} \longrightarrow \mathcal{D}_\rho(\mathcal{G})$, analytic with respect to the angular variables, such that:

a) *For every $I \in \widehat{G}$, the set $\mathcal{T}(\mathbf{T}^n \times \{I\})$ is an invariant torus of H, its frequency vector is colinear to $\omega(I)$ and its energy is $h(I)$.*

b) *Writing*

$$\mathcal{T}(\phi, I) = (\phi + \mathcal{T}_\phi(\phi, I), I + \mathcal{T}_I(\phi, I)),$$

one has the estimates

$$|\mathcal{T}_\phi|_{\widehat{G},(\frac{\rho_1}{4},0),\infty} \leq \frac{2^{2\tau+15} L^2 M}{\nu^2 l^2 \hat{\rho}^{2\tau+1}} \cdot \frac{\varepsilon}{\gamma^2}\,, \qquad |\mathcal{T}_I|_{\widehat{G},(\frac{\rho_1}{4},0)} \leq \frac{2^{\tau+16} L^3 M}{\nu l^3 \mu \hat{\rho}^{\tau+1}} \cdot \frac{\varepsilon}{\gamma}\,.$$

c) $\operatorname{mes}\left[(\mathbf{T}^n \times \mathcal{G}) \setminus \mathcal{T}\left(\mathbf{T}^n \times \widehat{G}\right)\right] \leq C\gamma$, *where C is a (very complicated) constant depending on n, τ, $\operatorname{diam} F$, D, $\hat{\rho}$, M, L, l, μ.*

Proof

A. Choice of the parameters. Since we make iterative use of proposition 11, we first introduce appropiate sequences of parameters to replace the constants of that proposition. For $q \geq 0$, we define

$$M_q = \left(2 - \frac{1}{2^q}\right) M, \quad L_q = \left(2 - \frac{1}{2^q}\right) L, \quad l_q = \left(1 + \frac{1}{2^q}\right)\frac{l}{2}\,, \quad \mu_q = \left(1 + \frac{1}{2^q}\right)\frac{\mu}{2}\,.$$

Note that M_q, L_q increase from M, l to $2M$, $2L$, respectively, for $q \to \infty$, and that l_q, μ_q decrease from l, μ to $l/2$, $\mu/2$. We also put

$$K_0 = 0, \qquad K_q = K \cdot 2^{q-1}, \ q \geq 1,$$

where $K \geq 1$ is an integer to be fixed below. Moreover, we define

$$\beta := \frac{\gamma}{L} \leq 1$$

and, for $q \geq 0$, we put $\rho^{(q)} = \left(\rho_1^{(q)}, \rho_2^{(q)}\right)$, with

$$\rho_1^{(q)} = \left(1 + \frac{1}{2^{\nu q}}\right) \frac{\rho_1}{4}, \qquad \rho_2^{(q)} = \frac{\nu l \beta}{32 M K_{q+1}^{\tau+1}} .$$

We notice that $\rho_1^{(q)}$ decreases from $\rho_1/2$ to $\rho_1/4$, and that $\rho_2^{(q)}$ decreases to 0. We also write

$$\delta_1^{(q)} = \rho_1^{(q-1)} - \rho_1^{(q)}, \qquad \delta_2^{(q)} = \rho_2^{(q-1)} - \rho_2^{(q)}, \qquad c_q = \frac{\delta_2^{(q)}}{\delta_1^{(q)}} .$$

Taking into account that the inequalities $\frac{\nu}{2} \leq 1 - \frac{1}{2^\nu} \leq \nu$ and $1 - \frac{1}{2^{\tau+1}} \geq \frac{1}{2}$, hold for $0 < \nu < 1$ and $\tau > 0$ respectively, it is easy to see that, for every $q \geq 1$,

$$\frac{\nu \rho_1}{8 \cdot 2^{\nu(q-1)}} \leq \delta_1^{(q)} \leq \frac{\nu \rho_1}{4 \cdot 2^{\nu(q-1)}}, \qquad \delta_2^{(q)} \geq \frac{\nu l \beta}{64 M K_q^{\tau+1}}, \tag{48}$$

$$\frac{l\beta \cdot 2^{\nu(q-1)}}{16 M K_q^{\tau+1} \rho_1} \leq c_q \leq \frac{l\beta \cdot 2^{\nu(q-1)}}{4 M K_q^{\tau+1} \rho_1} . \tag{49}$$

Finally, we define

$$\beta_q = \left(1 - \frac{1}{2^{\nu q}}\right) \beta, \qquad \beta'_q = \frac{\beta_q + \beta_{q+1}}{2},$$

which are both increasing with limit β. It is also easy to check that $\beta'_q \geq \nu\beta/4$ for every $q \geq 0$.

We choose K as the minimum integer such that $K\hat{\rho} \geq 1$. Then, one has $K \leq 2/\hat{\rho}$. Using this inequality, our choice $\beta = \gamma/L$ and the inequality $\hat{\rho} \leq \nu\rho_1$, we deduce from conditions (47) the inequalites

$$\varepsilon \leq \min\left(\frac{\nu^3 l^2 \rho_1 \beta^2}{2^{2\tau+20} M K^{2\tau+1}}, \frac{\nu^2 l^6 \mu^2 \beta^2}{2^{\tau+30} L^4 M^3 K^{2\tau+2}}\right), \qquad \beta \leq \frac{8 M K^{\tau+1} \rho_2}{\nu l} . \tag{50}$$

B. Induction. Starting with $G_0 = \mathcal{G}$, we shall now construct a decreasing sequence of compacts $G_q \subset \mathcal{G}$ and a sequence of real analytic canonical transformations $\Phi^{(q)} : \mathcal{D}_{\rho^{(q)}}(G_q) \longrightarrow \mathcal{D}_{\rho^{(q-1)}}(G_{q-1})$, $q \geq 1$. Denoting $\Psi^{(q)} = \Phi^{(1)} \circ \cdots \circ \Phi^{(q)}$, the transformed Hamiltonians will be written in the form $H^{(q)} = H \circ \Psi^{(q)} = h^{(q)}(I) + R^{(q)}(\phi, I)$. Moreover, we write $\omega^{(q)} = \operatorname{grad} h^{(q)}$ and $\Omega^{(q)} = \Omega_{\omega^{(q)}, h^{(q)}, a}$. We are going to show that the following statements hold for every $q \geq 0$:

$1_q)$ $\varepsilon_q := \left\| DR^{(q)} \right\|_{G_q, \rho^{(q)}, c_{q+1}} \leq \dfrac{8\varepsilon}{\nu \rho_1 \cdot 2^{(2\tau+2)q}}$.

$2_q)$ $\eta_q := \left|R_0^{(q)}\right|_{G_q,\rho_2^{(q)}} \le \dfrac{\varepsilon}{2^{(2\tau+3)q}}\,, \quad \xi_q := \left|\dfrac{\partial R_0^{(q)}}{\partial I}\right|_{G_q,\rho_2^{(q)}} \le \dfrac{2^6 M K^{\tau+1}\,\varepsilon}{\nu l\beta\cdot 2^{(\tau+2)q}}\,.$

$3_q)$ $\left|\dfrac{\partial^2 h^{(q)}}{\partial I^2}\right|_{G_q,\rho_2^{(q)}} \le M_q, \quad \left|\omega^{(q)}\right|_{G_q} \le L_q \quad$ and $\quad \left|\omega^{(q)}(I)\right| \ge l_q \ \ \forall I \in G_q.$

$4_q)$ $\omega^{(q)}$ is μ_q-isoenergetically nondegenerate on G_q.

$5_q)$ $\Omega^{(q)}$ is one-to-one on G_q, and $\Omega^{(q)}(G_q) = F_q$, where we define

$$F_q := (F-\beta_q) \setminus \bigcup_{\substack{|k|_1 \le K_q \\ k\ne 0}} \Delta\left(k, \frac{\beta_q}{|k|_1^\tau}\right). \tag{51}$$

We proceed by induction. For $q=0$, we choose $G_0 = \mathcal{G}$, $h^{(0)} = h$, $R^{(0)} = f$. Note that

$$\rho_1^{(0)} = \frac{\rho_1}{2}\,, \qquad \rho_2^{(0)} = \frac{\nu l\beta}{32MK^{\tau+1}} \le \frac{\rho_2}{2}$$

by (50). Then,

$$\varepsilon_0 = \|Df\|_{\mathcal{G},\rho^{(0)},c_1} \le \frac{\varepsilon}{\delta_1^{(1)}} \le \frac{8\varepsilon}{\nu\rho_1}\,, \tag{52}$$

namely (1_0). The first estimate of (2_0) is obvious, and the second one comes from the Cauchy inequality

$$\xi_0 \le \frac{2}{\rho_2}\left|R_0^{(0)}\right|_{\mathcal{G},\rho_2} \le \frac{2\varepsilon}{\rho_2} \le \frac{\varepsilon}{\rho_2^{(0)}}\,.$$

The remaining statements $(3_0$–$5_0)$ are clear.

For $q \ge 1$, we assume the statements true for $q-1$ and we prove them for q. We attain this aim by applying proposition 11 to $H^{(q-1)} = h^{(q-1)} + R^{(q-1)}$, with K_q instead of K. However, in order to fulfil condition (41), we must replace the domains G_{q-1} and F_{q-1} by suitable subsets where the resonances from order $K_{q-1}+1$ to order K_q have been removed. More precisely, we define

$$F'_{q-1} := (F-\beta_{q-1}) \setminus \bigcup_{\substack{|k|_1 \le K_q \\ k\ne 0}} \Delta\left(k, \frac{\beta'_{q-1}}{|k|_1^\tau}\right), \qquad G'_{q-1} := \left(\Omega^{(q-1)}\right)^{-1}\left(F'_{q-1}\right). \tag{53}$$

The nonresonance condition (41) is then fulfilled by F'_{q-1}, β'_{q-1} and K_q instead of F, β and K, respectively (taking β'_{q-1} instead of β_{q-1}, we avoid to apply proposition 11 with $\beta_0 = 0$ for $q=1$). The remaining parameters taking part in proposition 11 are M_{q-1}, L_{q-1}, l_{q-1}, μ_{q-1}, $\rho^{(q-1)}$, $\delta^{(q)}$, c_q, M_q, L_q, l_q, μ_q, replacing M, L, l, μ, ρ, δ, c, $\tilde M$, $\tilde L$, $\tilde l$, $\tilde\mu$, respectively, and

$$a = \frac{16M}{l^2} \ge \frac{2M_q}{l_q^2}\,.$$

It is clear that condition (42) on $\rho_2^{(q-1)}$ is satisfied with our choice of the parameters. Concerning condition (43) on ε_{q-1}, we first notice that, by (49) and our choice of K,

$$A_q := 1 + \frac{2M_{q-1}c_qK_q^\tau}{l_{q-1}\beta'_{q-1}} \le 1 + \frac{32M}{\nu l\beta}\cdot\frac{l\beta}{4MK\rho_1} = 1 + \frac{8}{\nu K\rho_1} \le 2.$$

Then, to see that condition (43) is fulfilled we have to check the inequality

$$\varepsilon_{q-1} \le \frac{l_{q-1}\beta'_{q-1}\delta_2^{(q)}}{148K_q^\tau} .$$

Indeed, recalling the definitions of the parameters and applying (1_{q-1}) and (48), it suffices to check that

$$\frac{8\varepsilon}{\nu\rho_1} \le \frac{\nu l\beta}{2^{11}K^\tau} \cdot \frac{\nu l\beta}{64MK^{\tau+1}} ,$$

which can be deduced from (50). Let us verify the second condition of (43), namely

$$\xi_{q-1} \le \min\left((M_q - M_{q-1})\delta_2^{(q)} , \, L_q - L_{q-1} , \, l_{q-1} - l_q , \, \frac{(\mu_{q-1} - \mu_q)\rho_2^{(q-1)}}{3} \right).$$

By (2_{q-1}) and (48), this condition holds, since we may deduce from (50) the inequality

$$\frac{2^6 MK^{\tau+1}\varepsilon}{\nu l\beta} \le \min\left(\frac{M}{2} \cdot \frac{\nu l\beta}{64MK^{\tau+1}} , \frac{L}{2} , \frac{l}{4} , \frac{\mu}{12} \cdot \frac{\nu l\beta}{32MK^{\tau+1}} \right).$$

Finally, we have to check condition (44):

$$\eta'_{q-1} := \frac{L_{q-1}\xi_{q-1}}{2M_q} + \eta_{q-1} \;\le\; \frac{\mu_q^2\rho_2^{(q)}}{32L_q^3a^2} . \tag{54}$$

By (2_{q-1}), we have the estimate

$$\eta'_{q-1} \le \frac{L}{2M} \cdot \frac{2^6 MK^{\tau+1}\varepsilon}{\nu l\beta \cdot 2^{(\tau+2)(q-1)}} + \frac{\varepsilon}{2^{(2\tau+3)(q-1)}} \le \frac{2^6 LK^{\tau+1}\varepsilon}{\nu l\beta \cdot 2^{(\tau+2)(q-1)}} . \tag{55}$$

Taking into account the value of a, the inequality (54) holds, since

$$\frac{2^6 LK^{\tau+1}\varepsilon}{\nu l\beta} \le \frac{l^4\mu^2}{2^{18}L^3M^2} \cdot \frac{\nu l\beta}{2^{\tau+6}MK^{\tau+1}} ,$$

as deduced from (50).

Applying proposition 11 with the parameters considered above, we obtain a canonical transformation $\Phi^{(q)}$ and a new Hamiltonian $H^{(q)} = h^{(q)} + R^{(q)}$. The new domain $G_q \subset G'_{q-1}$ is chosen below. First, we prove $(1_q$–$4_q)$. By the bound (b) of proposition 11 and the inequality $c_{q+1} \le c_q$,

$$\varepsilon_q \le \left\| DR^{(q)} \right\|_{G_q,\rho^{(q)},c_q} \le e^{-K_q\delta_1^{(q)}} \cdot \varepsilon_{q-1} + \frac{14A_qK_q^\tau}{l_{q-1}\beta'_{q-1}\delta_2^{(q)}} \cdot \varepsilon_{q-1}^2 . \tag{56}$$

Let us bound the terms of this sum. By (48) and the inequality $K\hat\rho \ge 1$,

$$K_q\delta_1^{(q)} \ge \frac{\nu K\rho_1}{8} \cdot 2^{(1-\nu)(q-1)} \ge (2\tau+3)\ln 2, \tag{57}$$

and therefore

$$e^{-K_q\delta_1^{(q)}} \le \frac{1}{2^{2\tau+3}} .$$

Moreover, applying (48), (1_{q-1}) and (50),

$$\frac{14A_qK_q^\tau}{l_{q-1}\beta'_{q-1}\delta_2^{(q)}}\cdot\varepsilon_{q-1}\le\frac{2^8K_q^\tau}{\nu l\beta}\cdot\frac{64MK_q^{\tau+1}}{\nu l\beta}\cdot\frac{8\varepsilon}{\nu\rho_1\cdot 2^{(2\tau+2)(q-1)}}\le\frac{1}{2^{2\tau+3}\cdot 2^{q-1}}\,. \tag{58}$$

Then, we deduce from (56) that

$$\varepsilon_q\le\frac{\varepsilon_{q-1}}{2^{2\tau+2}}\,,$$

which gives (1_q). For (2_q), we use part (c) of proposition 11. Writing $\sigma_2^{(q)}=\rho_2^{(q-1)}-\delta_2^{(q)}/2$, we have:

$$\eta_q\le\left|R_0^{(q)}\right|_{G_q,\sigma_2^{(q)}}\le\frac{7A_qK_q^\tau}{c_ql_{q-1}\beta'_{q-1}}\cdot\varepsilon_{q-1}^2\le\frac{\delta_1^{(q)}\varepsilon_{q-1}}{2}\cdot\frac{1}{2^{2\tau+3}\cdot 2^{q-1}}\le\frac{\varepsilon}{2^{(2\tau+3)q}}\,, \tag{59}$$

where we have used the inequalities (58), (48) and (1_{q-1}). We obtain the other estimate of (2_q) using the Cauchy inequality and (48):

$$\xi_q\le\frac{2}{\delta_2^{(q)}}\left|R_0^{(q)}\right|_{G_q,\sigma_2^{(q)}}\le\frac{2}{\delta_2^{(q)}}\cdot\frac{\varepsilon}{2^{(2\tau+3)q}}\,.$$

The statements (3_q–4_q) are clear from proposition 11. For (5_q), we apply part (g) of proposition 11 to the subset F_q. We have to check that

$$F_q\subset F'_{q-1}-\frac{16L_{q-1}L_qa^2\eta'_{q-1}}{\mu_q}\,. \tag{60}$$

Defining

$$d_q:=\frac{\beta_q-\beta_{q-1}}{2K_q^{\tau+1}}$$

and looking at (53) and (51), we have

$$F'_{q-1}-d_q\supset(F-(\beta_{q-1}+d_q))\setminus\bigcup_{\substack{|k|_1\le K_q\\ \overline{k}\ne 0}}\Delta\left(k\,,\,\frac{\beta'_{q-1}}{|k|_1^\tau}+\left|\overline{k}\right|d_q\right)\supset F_q\,, \tag{61}$$

where we have used the inequalities

$$\beta_{q-1}+d_q\le\beta_q\,,\qquad\frac{\beta'_{q-1}}{|k|_1^\tau}+\left|\overline{k}\right|d_q\le\frac{\beta_q}{|k|_1^\tau}\,.$$

Moreover, applying estimate (55) for η'_{q-1} and (50),

$$\frac{16L_{q-1}L_qa^2\eta'_{q-1}}{\mu_q}\le\frac{2^{15}L^2M^2}{l^4\mu}\cdot\frac{2^6LK^{\tau+1}\,\varepsilon}{\nu l\beta\cdot 2^{(\tau+2)(q-1)}}\le\frac{\nu\beta}{4\cdot 2^{\nu(q-1)}K_q^{\tau+1}}\le d_q\,.$$

This bound and the inclusion (61) imply (60). Hence, part (g) of proposition 11 says that $\Omega^{(q)}$ is one-to-one on $G_q=\left(\Omega^{(q)}\right)^{-1}(F_q)$. This gives ($5_q$), and the induction is completed.

C. Convergence of the diffeomorphisms. We now see the convergence of the successive maps $\Omega^{(q)} : G_q \longrightarrow F_q$. Applying part (g) of proposition 11 again, we obtain for $q \geq 1$ the estimates

$$\left|\Omega^{(q)} - \Omega^{(q-1)}\right|_{G_q} \leq a\eta'_{q-1}\,, \qquad \left|\left(\Omega^{(q)}\right)^{-1} - \left(\Omega^{(q-1)}\right)^{-1}\right|_{F_q} \leq \frac{2L_{q-1}a\eta'_{q-1}}{\mu_{q-1}}\,. \tag{62}$$

Therefore, by the bound (55) on η'_{q-1}, the sequences $\Omega^{(q)}$ and $\left(\Omega^{(q)}\right)^{-1}$ converge respectively to maps which we name Ω^* and Υ, defined on the sets

$$G^* := \bigcap_{q\geq 0} G_q, \qquad F^* := \bigcap_{q\geq 0} F_q = (F-\beta) \setminus \bigcup_{\substack{k\in\mathbf{Z}^n\\ k\neq 0}} \Delta\left(k, \frac{\beta}{|k|_1^\tau}\right). \tag{63}$$

Note that the compacity of F^* has been used to establish the second equality. From (62) we deduce:

$$\left|\Omega^* - \Omega^{(q)}\right|_{G^*} \leq \sum_{s\geq q} a\eta'_s\,, \qquad \left|\Upsilon - \left(\Omega^{(q)}\right)^{-1}\right|_{F^*} \leq \sum_{s\geq q} \frac{2L_s a\eta'_s}{\mu_s}\,. \tag{64}$$

Note also that, for every q,

$$G_q \subset G_{q-1} - \frac{4L_q a\eta'_{q-1}}{\mu_q}\,, \qquad F_q \subset F_{q-1} - \frac{16L_{q-1}L_q a^2\eta'_{q-1}}{\mu_q}\,.$$

Iterating these inclusions we deduce the following two ones:

$$G^* \subset G_q - \sum_{s\geq q} \frac{4L_{s+1}a\eta'_s}{\mu_{s+1}}\,, \qquad F^* \subset F_q - \sum_{s\geq q} \frac{16L_s L_{s+1}a^2\eta'_s}{\mu_{s+1}}\,, \tag{65}$$

to be used below. We are now going to see that Ω^* is one-to-one on G^* and that $\Omega^*(G^*) = F^*$. Given $I \in G^*$, we have $\Omega^{(q)}(I) \in F_q$ for every q. Hence $\Omega^*(I) \in F^*$, and we deduce that $\Omega^*(G^*) \subset F^*$. Analogously, we have $\Upsilon(F^*) \subset G^*$. Moreover, we prove that $\Upsilon(\Omega^*(I)) = I$ for every $I \in G^*$. Indeed, for every q,

$$\begin{aligned}
&|\Upsilon(\Omega^*(I)) - I| \\
&\quad\leq \left|\Upsilon(\Omega^*(I)) - \left(\Omega^{(q)}\right)^{-1}(\Omega^*(I))\right| + \left|\left(\Omega^{(q)}\right)^{-1}(\Omega^*(I)) - \left(\Omega^{(q)}\right)^{-1}\left(\Omega^{(q)}(I)\right)\right| \\
&\quad\leq \left|\Upsilon - \left(\Omega^{(q)}\right)^{-1}\right|_{F^*} + \frac{2L_q}{\mu_q}\left|\Omega^* - \Omega^{(q)}\right|_{G^*}.
\end{aligned} \tag{66}$$

For the estimate of the second term, we have used the following two facts: on one hand, part (b) of lemma 8 gives the bound

$$\left|\frac{\partial\Omega^{(q)}}{\partial I}(I')\,v\right| \geq \frac{\mu_q}{2L_q}|v| \qquad \forall v \in \mathbf{R}^n, \quad \forall I' \in G_q\,;$$

on the other hand, by (64–65) the segment joining $\Omega^{(q)}(I)$ and $\Omega^*(I)$ is contained in F_q. Then, since the bound obtained in (66) tends to zero for $q \to \infty$, we deduce that $\Upsilon(\Omega^*(I)) = I$, and therefore Ω^* is one-to-one. A symmetric argument allows to prove that $\Omega^*(\Upsilon(J)) = J$ for every $J \in F^*$. This implies that $\Omega^*(G^*) \supset F^*$. Thus, the map Ω^* is one-to-one and $\Omega^*(G^*) = F^*$.

We also see, from proposition 11, that

$$\left|\omega^{(q)} - \omega^{(q-1)}\right|_{G_q,\rho_2^{(q-1)}} \le \xi_{q-1}, \qquad \left|h^{(q)} - h^{(q-1)}\right|_{G_q,\rho_2^{(q-1)}} \le \eta_{q-1}.$$

This implies that the sequences $\omega^{(q)}$ and $h^{(q)}$ converge to some continuous maps ω^* and h^*, respectively. Note also that, for $I \in G^*$,

$$\Omega^*(I) = \left(\frac{\overline{\omega^*}(I)}{\omega_n^*(I)}\,,\; a\,h^*(I)\right). \tag{67}$$

From (2_q) we deduce, for every $q \ge 0$, the following bound to be used later:

$$\left|\omega^* - \omega^{(q)}\right|_{G^*} \le \sum_{s\ge q} \xi_s \le \frac{2^7 M K^{\tau+1}\,\varepsilon}{\nu l \beta \cdot 2^{(\tau+2)q}}\,. \tag{68}$$

D. Convergence of the canonical transformations. Next we estimate how near to the identity map the transformations $\Psi^{(q)}$ are. Applying part (d) of proposition 11 and using (1_{q-1}), (48) and (50), one deduces that, for every $q \ge 1$,

$$\begin{aligned}\left|\Phi^{(q)} - \mathrm{id}\right|_{G_q,\sigma^{(q)},c_q} &\le \frac{2A_q K_q^\tau}{l_{q-1}\beta'_{q-1}} \cdot \varepsilon_{q-1} \le \frac{2^5 K_q^\tau}{\nu l \beta} \cdot \frac{8\varepsilon}{\nu\rho_1 \cdot 2^{(2\tau+2)(q-1)}} \\ &= \frac{2^8 K^\tau\,\varepsilon}{\nu^2 l \rho_1 \beta \cdot 2^{(\tau+2)(q-1)}} \le \frac{\delta_2^{(q)}}{8 \cdot 2^{q-1}}\,,\end{aligned} \tag{69}$$

where we write $\sigma^{(q)} = \rho^{(q-1)} - \delta_2^{(q)}/2$. Then, applying property (15) of section 2.2,

$$\left|D\Phi^{(q)} - \mathrm{Id}\right|_{G_q,\rho^{(q)},c_q} \le \frac{2}{\delta_2^{(q)}} \left|\Phi^{(q)} - \mathrm{id}\right|_{G_q,\sigma^{(q)},c_q} \le \frac{1}{4 \cdot 2^{q-1}}\,.$$

Let x, y such that the segment joining them is contained $\mathcal{D}_{\rho^{(q)}}(G_q)$. The mean value theorem gives the bound:

$$\left|\Phi^{(q)}(x) - \Phi^{(q)}(y)\right|_{c_q} \le \left|D\Phi^{(q)}\right|_{G_q,\rho^{(q)},c_q} \cdot |x-y|_{c_q}\,.$$

By (69),

$$\left|\Phi^{(q)}(x) - x\right|_{c_q} \le \delta_2^{(q)}, \qquad \left|\Phi^{(q)}(y) - y\right|_{c_q} \le \delta_2^{(q)}.$$

Then, since $\rho^{(q)} + \delta^{(q)} = \rho^{(q-1)}$, it turns out that the segment joining $\Phi^{(q)}(x)$ and $\Phi^{(q)}(y)$ is contained in $\mathcal{D}_{\rho^{(q-1)}}(G_{q-1})$. Therefore,

$$\begin{aligned}&\left|\Phi^{(q-1)}\left(\Phi^{(q)}(x)\right) - \Phi^{(q-1)}\left(\Phi^{(q)}(y)\right)\right|_{c_{q-1}} \\ &\quad\le \left|D\Phi^{(q-1)}\right|_{G_{q-1},\rho^{(q-1)},c_{q-1}} \cdot \left|\Phi^{(q)}(x) - \Phi^{(q)}(y)\right|_{c_{q-1}} \\ &\quad\le 2^{\tau+1-\nu}\left|D\Phi^{(q-1)}\right|_{G_{q-1},\rho^{(q-1)},c_{q-1}} \cdot \left|\Phi^{(q)}(x) - \Phi^{(q)}(y)\right|_{c_q}\,,\end{aligned}$$

where we have used that $c_{q-1}/c_q = 2^{\tau+1-\nu}$. Iterating this argument and putting the successive bounds obtained together, we arrive at the estimate

$$\begin{aligned}
&\left|\Psi^{(q)}(x) - \Psi^{(q)}(y)\right|_{c_1} \\
&\quad \le 2^{(\tau+1-\nu)(q-1)} \left|D\Phi^{(1)}\right|_{G_1,\rho^{(1)},c_1} \cdot \left|D\Phi^{(2)}\right|_{G_2,\rho^{(2)},c_2} \cdots \left|D\Phi^{(q)}\right|_{G_q,\rho^{(q)},c_q} \cdot |x-y|_{c_q} \\
&\quad \le 2^{(\tau+1-\nu)(q-1)} \left(1+\frac{1}{4}\right)\left(1+\frac{1}{4\cdot 2}\right)\cdots\left(1+\frac{1}{4\cdot 2^{q-1}}\right) \cdot |x-y|_{c_q} \\
&\quad \le 2^{(\tau+1-\nu)(q-1)} \cdot e^{1/2}\,|x-y|_{c_q} \le 2^{(\tau+1-\nu)(q-1)} \cdot 2\,|x-y|_{c_q}\,,
\end{aligned} \tag{70}$$

which holds for $q \ge 1$, and for every x, y such that the segment joining them is contained in $\mathcal{D}_{\rho^{(q)}}(G_q)$. Now, given $q \ge 2$ and $x \in \mathcal{D}_{\rho^{(q)}}(G_q)$, we put $y = \Phi^{(q)}(x)$ and apply (70) with $q-1$ instead of q. We obtain:

$$\begin{aligned}
\left|\Psi^{(q)}(x) - \Psi^{(q-1)}(x)\right|_{c_1} &= \left|\Psi^{(q-1)}\left(\Phi^{(q)}(x)\right) - \Psi^{(q-1)}(x)\right|_{c_1} \\
&\le 2^{(\tau+1-\nu)(q-2)} \cdot 2\left|\Phi^{(q)}(x) - x\right|_{c_{q-1}} \\
&\le 2^{(\tau+1-\nu)(q-1)} \cdot 2\left|\Phi^{(q)}(x) - x\right|_{c_q} \le \frac{2^9 K^\tau\,\varepsilon}{\nu^2 l \rho_1 \beta \cdot 2^{(1+\nu)(q-1)}}\,,
\end{aligned} \tag{71}$$

where (69) has been used. This estimate holds for $q \ge 2$, but one readily sees from (69) that it is also true for $q = 1$ (we put $\Psi^{(0)} = \mathrm{id}$). Clearly, estimate (71) implies that the sequence of transformations $\Psi^{(q)}$ converges to a map

$$\Psi^* : \mathcal{D}_{\left(\frac{\rho_1}{4},0\right)}(G^*) = \mathcal{W}_{\frac{\rho_1}{4}}(\mathbf{T}^n) \times G^* \longrightarrow \mathcal{D}_\rho(G)$$

and we deduce, for every $q \ge 0$, the estimate

$$\left|\Psi^* - \Psi^{(q)}\right|_{G^*,\left(\frac{\rho_1}{4},0\right),c_1} \le \frac{2^{10} K^\tau\,\varepsilon}{\nu^2 l \rho_1 \beta \cdot 2^{(1+\nu)q}}\,. \tag{72}$$

Moreover, by carrying to the limit the equation $H \circ \Psi^{(q)} = h^{(q)} + R^{(q)}$, we see that $H \circ \Psi^* = h^*(I)$ on $\mathcal{D}_{\left(\frac{\rho_1}{4},0\right)}(G^*)$.

E. Stability estimates. Next we see that, for $q \to \infty$, the motions associated to the transformed Hamiltonian $H^{(q)} = h^{(q)} + R^{(q)}$ and the quasiperiodic motions of $h^{(q)}$ become closer and closer. Let us denote $x^{(q)}(t) = \left(\phi^{(q)}(t), I^{(q)}(t)\right)$ the trajectory of $H^{(q)}$ corresponding to a given initial condition $x^{(q)}(0) = x_0^* = (\phi_0^*, I_0^*) \in \mathbf{T}^n \times G_q$, and let $\hat{x}^{(q)}(t) := \left(\hat{\phi}^{(q)}(t), I_0^*\right) = \left(\phi_0^* + \omega^{(q)}(I_0^*)\,t, I_0^*\right)$ the corresponding trajectory of the integrable part $h^{(q)}$. It is clear that $\hat{x}^{(q)}(t)$ is defined for all $t \in \mathbf{R}$. Like in lemma 5, let us denote

$$T_q := \inf\left\{t > 0 \;:\; \left|I^{(q)}(t) - I_0^*\right| > \delta_2^{(q+1)} \text{ or } \left|\phi^{(q)}(t) - \hat{\phi}^{(q)}(t)\right|_\infty > \delta_1^{(q+1)}\right\}. \tag{73}$$

Clearly, $x^{(q)}(t)$ is defined and belongs to $\mathcal{D}_{\rho^{(q)}}(G_q)$ for $0 \le t \le T_q$ (we remark that our use of $\delta^{(q+1)}$ instead of $\rho^{(q)}$ is just due to the fact that we shall take some advantage of the "c_{q+1}-norm"). From the Hamiltonian equations associated to $H^{(q)}$

$$\dot{I}^{(q)}(t) = -\frac{\partial R^{(q)}}{\partial \phi}\left(x^{(q)}(t)\right), \qquad \dot{\phi}^{(q)}(t) = \omega^{(q)}\left(I^{(q)}(t)\right) + \frac{\partial R^{(q)}}{\partial I}\left(x^{(q)}(t)\right),$$

we get the bounds:

$$\left|\dot{I}^{(q)}(t)\right| \le \left\|\frac{\partial R^{(q)}}{\partial \phi}\right\|_{G_q,\rho^{(q)}} \le \varepsilon_q\,, \tag{74}$$

$$\begin{aligned}\left|\dot{\phi}^{(q)}(t) - \omega^{(q)}\left(I_0^*\right)\right|_\infty &\le M_q\left|I^{(q)}(t) - I_0^*\right| + \left\|\frac{\partial R^{(q)}}{\partial I}\right\|_{G_q,\rho^{(q)},\infty}\\ &\le 2M\left|I^{(q)}(t) - I_0^*\right| + \frac{\varepsilon_q}{c_{q+1}}\\ &\le 2M\delta_2^{(q+1)} + \frac{\varepsilon_q}{c_{q+1}} \le 3M\delta_2^{(q+1)}.\end{aligned} \tag{75}$$

In the second bound, we have used the inequality

$$\frac{\varepsilon_q}{c_{q+1}} \le M\delta_2^{(q+1)}, \tag{76}$$

which comes from the bounds (1_q) and (48–50). Thus, since one of the inequalities defining (73) is an equality for $t = T_q$, we have

$$\delta_2^{(q+1)} = \left|I^{(q)}\left(T_q\right) - I_0^*\right| \le T_q\varepsilon_q$$

or

$$\delta_1^{(q+1)} = \left|\phi^{(q)}\left(T_q\right) - \hat{\phi}^{(q)}\left(T_q\right)\right|_\infty \le T_q \cdot 3M\delta_2^{(q+1)}\,. \tag{77}$$

Therefore,

$$T_q \ge \min\left(\frac{\delta_2^{(q+1)}}{\varepsilon_q}\,,\, \frac{\delta_1^{(q+1)}}{3M\delta_2^{(q+1)}}\right) \;\ge\; T_q' := \frac{1}{3Mc_{q+1}}\,, \tag{78}$$

where the inequality (76) has been used again. This implies:

$$\left|x^{(q)}(t) - \hat{x}^{(q)}(t)\right|_{c_{q+1}} \le \delta_2^{(q+1)} \qquad \text{for} \quad |t| \le T_q'\,. \tag{79}$$

Since $H^{(q)} = H \circ \Psi^{(q)}$ and $\Psi^{(q)}$ is canonical, it turns out that $\Psi^{(q)}\left(x^{(q)}(t)\right)$ is a trajectory of H, defined for $|t| \le T_q'$. For large values of q, this trajectory remains near the torus $\Psi^{(q)}\left(\mathbf{T}^n \times \{I_0^*\}\right)$. Note that T_q' tends to infinity for $q \to \infty$.

F. Invariant tori. Assume now that $x_0^* \in \mathbf{T}^n \times G^*$, and write $x^*(t) = \left(\phi_0^* + \omega^*\left(I_0^*\right)t, I_0^*\right)$, $t \in \mathbf{R}$. Note that

$$\left|\hat{x}^{(q)}(t) - x^*(t)\right|_{c_{q+1}} \le c_{q+1}\left|\omega^{(q)}\left(I_0^*\right) - \omega^*\left(I_0^*\right)\right|_\infty \cdot |t| \le c_{q+1}\left|\omega^{(q)} - \omega^*\right|_{G^*,\infty} \cdot T_q'' \le \delta_2^{(q+1)}$$

for

$$|t| \le T_q'' := \frac{\delta_1^{(q+1)}}{\left|\omega^{(q)} - \omega^*\right|_{G^*,\infty}}\,,$$

which also tends to infinity, by (68). Then,

$$\left|x^{(q)}(t) - x^*(t)\right|_{c_{q+1}} \le 2\delta_2^{(q+1)} \qquad \text{for} \quad |t| \le T_q''' := \min\left(T_q', T_q''\right).$$

Next, we see that the trajectory $\Psi^{(q)}\left(x^{(q)}(t)\right)$ is very close to $\Psi^*\left(x^*(t)\right)$ for large values of q. Indeed, for $|t| \leq T_q'''$,

$$\begin{aligned}
&\left|\Psi^{(q)}\left(x^{(q)}(t)\right) - \Psi^*\left(x^*(t)\right)\right|_{c_1} \\
\leq\ & \left|\Psi^{(q)}\left(x^{(q)}(t)\right) - \Psi^{(q)}\left(x^*(t)\right)\right|_{c_1} + \left|\Psi^{(q)}\left(x^*(t)\right) - \Psi^*\left(x^*(t)\right)\right|_{c_1} \\
\leq\ & 2^{(\tau+1-\nu)(q-1)} \cdot 4\delta_2^{(q+1)} + \left|\Psi^{(q)} - \Psi^*\right|_{G^*,\left(\frac{\rho_1}{4},0\right),c_1} \\
=\ & 4c_1\,\delta_1^{(q+1)} + \left|\Psi^{(q)} - \Psi^*\right|_{G^*,\left(\frac{\rho_1}{4},0\right),c_1}, \qquad (80)
\end{aligned}$$

where we have applied (70). The bound obtained in (80) tends to zero, by (72). Then, for every fixed t, we see that $\Psi^{(q)}\left(x^{(q)}(t)\right)$ exists for q large enough, and its limit is $\Psi^*\left(x^*(t)\right)$. This fact and the continuity of the flow of H imply that $\Psi^*\left(x^*(t)\right)$ is also a trajectory of H, which is defined for all $t \in \mathbf{R}$. This holds for every initial condition $x_0^* = (\phi_0^*, I_0^*) \in \mathbf{T}^n \times G^*$. Hence $\Psi^*\left(\mathbf{T}^n \times \{I_0^*\}\right)$ is an invariant torus of H, with frequency vector $\omega^*\left(I_0^*\right)$. Moreover, the energy on the torus is $H\left(\Psi^*\left(\phi_0^*, I_0^*\right)\right) = h^*\left(I_0^*\right)$.

The preserved invariant tori are thus parametrized by the transformed actions $I_0^* \in G^*$. We are now going to parametrize the preserved tori by their original actions. First, let us see that $\Omega\left(\widehat{G}\right) \subset F^*$ (the Diophantine set F^* has been introduced in (63)). Indeed, using part (b) of lemma 8 and the fact that $\beta = \gamma/L$, we see that $\Omega\left(\mathcal{G} - \frac{2\gamma}{\mu}\right) \subset F - \beta$. On the other hand, given $I \in \widehat{G}$, we deduce from the Diophantine condition fulfilled by $\omega(I)$ that

$$\left|\overline{k}\cdot\overline{\Omega}(I) + k_n\right| \geq \frac{\gamma/\left|\omega_n(I)\right|}{|k|_1^\tau} \geq \frac{\beta}{|k|_1^\tau}$$

and therefore $\Omega(I) \notin \Delta\left(k, \frac{\beta}{|k|_1^\tau}\right)$ for every $k \neq 0$ (we point out that this estimate motivated our choice $\beta = \gamma/L$). Hence $\Omega\left(\widehat{G}\right) \subset F^*$ and, since $\Omega^* : G^* \longrightarrow F^*$ is one-to-one, we can take for the set of invariant tori a parameter $I_0 \in \widehat{G}$ (note that some of the invariant tori are thus neglected). We define, for $(\phi_0, I_0) \in \mathcal{W}_{\frac{\rho_1}{4}}\left(\mathbf{T}^n\right) \times \widehat{G}$,

$$\mathcal{T}(\phi_0, I_0) = \Psi^*\left(\phi_0, I_0^*\right),$$

where $I_0^* = \left(\Omega^*\right)^{-1}\left(\Omega(I_0)\right) \in G^*$. One then obtains part (a): the set $\mathcal{T}\left(\mathbf{T}^n \times \{I_0\}\right)$ is an invariant torus of H, with frequency $\omega^*\left(I_0^*\right)$ and energy $h^*\left(I_0^*\right)$. Since $\Omega^*\left(I_0^*\right) = \Omega(I_0)$, we deduce from (67) that $\omega^*\left(I_0^*\right)$ is colinear to $\omega(I_0)$ and that $h^*\left(I_0^*\right) = h(I_0)$.

For (b), let us write, for $(\phi_0, I_0^*) \in \mathcal{W}_{\frac{\rho_1}{4}}\left(\mathbf{T}^n\right) \times G^*$,

$$\Psi^*\left(\phi_0, I_0^*\right) = \left(\phi_0 + \Psi_\phi^*\left(\phi_0, I_0^*\right), I_0^* + \Psi_I^*\left(\phi_0, I_0^*\right)\right).$$

Then, for $(\phi_0, I_0) \in \mathcal{W}_{\frac{\rho_1}{4}}\left(\mathbf{T}^n\right) \times \widehat{G}$,

$$\mathcal{T}_\phi(\phi_0, I_0) = \Psi_\phi^*\left(\phi_0, I_0^*\right), \qquad \mathcal{T}_I(\phi_0, I_0) = \Psi_I^*\left(\phi_0, I_0^*\right) + I_0 - I_0^*.$$

Using (72) and (49), we get the following estimates:

$$\left|\Psi_\phi^*\left(\phi_0, I_0^*\right)\right|_\infty \leq \frac{1}{c_1}\left|\Psi^* - \mathrm{id}\right|_{G^*,\left(\frac{\rho_1}{4},0\right),c_1} \leq \frac{2^{14}MK^{2\tau+1}\,\varepsilon}{\nu^2 l^2 \beta^2},$$

$$\left|\Psi_I^*\left(\phi_0, I_0^*\right)\right| \leq \left|\Psi^* - \mathrm{id}\right|_{G^*,\left(\frac{\rho_1}{4},0\right),c_1} \leq \frac{2^{10}K^\tau\,\varepsilon}{\nu^2 l \rho_1 \beta}.$$

By (64) and (55),

$$\begin{aligned} |I_0^* - I_0| &\leq \left| (\Omega^*)^{-1} - \Omega^{-1} \right|_{F^*} \leq \sum_{s\geq 0} \frac{2L_s a \eta_s'}{\mu_s} \leq \frac{8La}{\mu} \sum_{s\geq 0} \eta_s' \\ &\leq \frac{2^7 LM}{l^2 \mu} \cdot \frac{2^7 L K^{\tau+1} \varepsilon}{\nu l \beta} = \frac{2^{14} L^2 M K^{\tau+1} \varepsilon}{\nu l^3 \mu \beta} . \end{aligned}$$

By putting these bounds together, applying the inequalities $\hat{\rho} \leq \nu \rho_1$ and $K \leq 2/\hat{\rho}$ and writing the estimate in terms of γ instead of β, we get part (b).

G. Estimate of the measure. Finally, we carry out the estimate of part (c). Writing $\widehat{G}^* = (\Omega^*)^{-1} \left(\Omega \left(\widehat{G} \right) \right)$, the invariant tori found fill the set $\mathcal{T} \left(\mathbf{T}^n \times \widehat{G} \right) = \Psi^* \left(\mathbf{T}^n \times \widehat{G}^* \right)$. Since the transformations $\Psi^{(q)}$ are canonical,

$$\operatorname{mes} \left[\Psi^{(q)} \left(\mathbf{T}^n \times \widehat{G}^* \right) \right] = \operatorname{mes} \left(\mathbf{T}^n \times \widehat{G}^* \right) = (2\pi)^n \cdot \operatorname{mes} \left(\widehat{G}^* \right) .$$

Using estimate (72) and the compacity of $\Psi^* \left(\mathbf{T}^n \times \widehat{G}^* \right)$, we get the inequality

$$\operatorname{mes} \left[\Psi^* \left(\mathbf{T}^n \times \widehat{G}^* \right) \right] \geq (2\pi)^n \cdot \operatorname{mes} \left(\widehat{G}^* \right) .$$

Then, to estimate the measure of the complement of the invariant set, it suffices to bound the measure of $\mathcal{G} \setminus \widehat{G}^*$.

First we construct an auxiliary set, included in $\widehat{G}^*$, such that the estimates become easier on it. Let

$$\tilde{\beta} = \frac{64 LM\gamma}{l^2 \mu} , \qquad \tilde{\beta}_q = \left(1 - \frac{1}{2^{\nu q}} \right) \tilde{\beta},$$

for $q \geq 0$, and note that $\tilde{\beta} \geq \beta$. We define the sets

$$\tilde{F}_q = \left(F - \tilde{\beta}_q \right) \setminus \bigcup_{\substack{|k|_1 \leq K_q \\ \overline{k} \neq 0}} \Delta \left(k, \frac{\tilde{\beta}_q}{|k|_1^\tau} \right) , \qquad \tilde{G}_q = \left(\Omega^{(q)} \right)^{-1} \left(\tilde{F}_q \right) ,$$

and

$$\tilde{F}^* = \bigcap_{q\geq 0} \tilde{F}_q = \left(F - \tilde{\beta} \right) \setminus \bigcup_{\substack{k \in \mathbf{Z}^n \\ \overline{k} \neq 0}} \Delta \left(k, \frac{\tilde{\beta}}{|k|_1^\tau} \right) , \qquad \tilde{G}^* = \bigcap_{q\geq 0} \tilde{G}_q .$$

From the fact that $\Omega^{(q)} \left(\tilde{G}_q \right) = \tilde{F}_q$ for every q, we deduce $\Omega^* \left(\tilde{G}^* \right) = \tilde{F}^*$. Let us check that $\Omega \left(\widehat{G} \right) \supset \tilde{F}^*$. Indeed, from part (a) of lemma 8, we get that $\Omega \left(\mathcal{G} - \frac{2\gamma}{\mu} \right) \supset F - \tilde{\beta}$. Moreover, given $J = \Omega(I) \in \tilde{F}^*$, for every $k \in \mathbf{Z}^n$ with $\overline{k} \neq 0$ we have

$$\left| \overline{k} \cdot \overline{J} + k_n \right| \geq \frac{\tilde{\beta}}{|k|_1^\tau}$$

and hence

$$|k \cdot \omega(I)| \geq \frac{\tilde{\beta}\, |\omega_n(I)|}{|k|_1^\tau} \geq \frac{l\tilde{\beta}}{|k|_1^\tau} \geq \frac{\gamma}{|k|_1^\tau} .$$

For $\overline{k}=0$, it is clear that the Diophantine estimate is also fulfilled since $|\omega_n(I)| \geq l \geq \gamma$. Hence $\Omega\left(\widehat{G}\right) \supset \tilde{F}^*$ and therefore $\widehat{G}^* \supset \tilde{G}^*$. Then,

$$\operatorname{mes}\left(\mathcal{G} \setminus \widehat{G}^*\right) \leq \sum_{q=1}^{\infty} \operatorname{mes}\left(\tilde{G}_{q-1} \setminus \tilde{G}_q\right).$$

For $q \geq 1$ we have the estimate

$$\operatorname{mes}\left(\tilde{G}_{q-1} \setminus \tilde{G}_q\right) \leq \frac{2^{2n-7} l^2 L^{n-2}}{\mu^{n-1} M} \cdot \operatorname{mes}\left(\tilde{F}_{q-1} \setminus \left(\tilde{F}_q - a\eta'_{q-1}\right)\right).$$

It has been used that $\Omega^{(q-1)}\left(\tilde{G}_{q-1}\right) = \tilde{F}_{q-1}$ and that $\Omega^{(q-1)}\left(\tilde{G}_q\right) \supset \tilde{F}_q - a\eta'_{q-1}$. This inclusion comes from part (g) of proposition 11. Another point we have used is the bound

$$\left|\det\left(\frac{\partial \Omega^{(q-1)}}{\partial I}(I)\right)\right| \geq \frac{\mu_{q-1}^{n-1} a}{L_{q-1}^{n-2}} \geq \frac{\mu^{n-1} M}{2^{2n-7} l^2 L^{n-2}} \ ,$$

given by part (c) of lemma 8. In accordance to the notation of lemma 12, we have $\tilde{F}_{q-1} = F\left(\tilde{\beta}_{q-1}, \tilde{\beta}_{q-1}, K_{q-1}\right)$, $\tilde{F}_q = F\left(\tilde{\beta}_q, \tilde{\beta}_q, K_q\right)$. Applying that lemma,

$$\begin{aligned}
\operatorname{mes}\left(\tilde{F}_{q-1} \setminus \tilde{F}_q\right) \ &\leq \ D\left(\tilde{\beta}_q - \tilde{\beta}_{q-1}\right) \\
&\quad + 2(\operatorname{diam} F)^{n-1} \left(\sum_{\substack{|k|_1 \leq K_{q-1} \\ \overline{k} \neq 0}} \frac{\tilde{\beta}_q - \tilde{\beta}_{q-1}}{|k|_1^\tau \cdot \left|\overline{k}\right|} + \sum_{\substack{K_{q-1} < |k|_1 \leq K_q \\ \overline{k} \neq 0}} \frac{\tilde{\beta}_q}{|k|_1^\tau \cdot \left|\overline{k}\right|} \right), \\
\operatorname{mes}\left(\tilde{F}_q \setminus \left(\tilde{F}_q - a\eta'_{q-1}\right)\right) \ &\leq \ \left(D + 2^{n+1} (\operatorname{diam} F)^{n-1} K_q^n\right) \cdot a\eta'_{q-1}.
\end{aligned}$$

Putting these estimates together, we get

$$\begin{aligned}
\operatorname{mes}\left(\mathcal{G} \setminus \widehat{G}^*\right) \ &\leq \ \frac{2^{2n-7} l^2 L^{n-2}}{\mu^{n-1} M} \left(D\tilde{\beta} + 2(\operatorname{diam} F)^{n-1} \sum_{\substack{k \in \mathbf{Z}^n \\ \overline{k} \neq 0}} \frac{\tilde{\beta}}{|k|_1^\tau \cdot \left|\overline{k}\right|} \right. \\
&\qquad \left. + D \sum_{q=1}^{\infty} a\eta'_{q-1} + 2^{n+1} (\operatorname{diam} F)^{n-1} \sum_{q=1}^{\infty} K_q^n a\eta'_{q-1} \right). \qquad (81)
\end{aligned}$$

It is crucial to use the condition $\tau > n-1$ in checking that the three series taking part in the right hand side of (81) are convergent. Indeed, for the first series,

$$\sum_{\substack{k \in \mathbf{Z}^n \\ \overline{k} \neq 0}} \frac{1}{|k|_1^\tau \cdot \left|\overline{k}\right|} \leq \sum_{\substack{\overline{k} \in \mathbf{Z}^{n-1} \\ \overline{k} \neq 0}} \sum_{k_n \in \mathbf{Z}} \frac{\sqrt{n}}{\left(\left|\overline{k}\right|_1 + |k_n|\right)^\tau \cdot \left|\overline{k}\right|_1} \leq \sqrt{n}\, 2^{n-1} \sum_{j=1}^{\infty} \sum_{k_n \in \mathbf{Z}} \frac{j^{n-3}}{(j + |k_n|)^\tau} \ ,$$

where we have used that the number of vectors $\overline{k} \in \mathbf{Z}^{n-1}$ with $\left|\overline{k}\right|_1 = j \geq 1$ can be bounded by $2^{n-1} j^{n-2}$. The series indexed by k_n can be bounded by comparing it with an integral:

$$\sum_{k_n \in \mathbf{Z}} \frac{1}{(j + |k_n|)^\tau} \leq \frac{1}{j^\tau} + 2\int_0^\infty \frac{dx}{(j+x)^\tau} = \frac{1}{j^\tau} + \frac{2}{(\tau - 1) j^{\tau - 1}} \leq \frac{\tau + 1}{\tau - 1} \cdot \frac{1}{j^{\tau - 1}}\ .$$

We have used that $\tau > 1$ since $n \geq 2$. Then,

$$\sum_{\substack{k \in \mathbf{Z}^n \\ \overline{k} \neq 0}} \frac{1}{|k|_1^\tau \cdot |\overline{k}|} \leq \frac{\sqrt{n}\, 2^{n-1}(\tau+1)}{\tau - 1} \sum_{j=1}^{\infty} \frac{1}{j^{\tau-n+2}} \,,$$

which converges by the condition $\tau > n-1$. It is easy to check that the second and the third series of (81) converge, using the bound

$$a\eta'_{q-1} \leq \frac{\nu l^3 \mu^2}{2^{\tau+20} L^3 M^2 K^{\tau+1} \cdot 2^{(\tau+2)(q-1)}} \cdot \beta \,,$$

which comes from estimates (55) and (50). Writing all bounds in terms of γ instead of $\tilde{\beta}$ or β, we get from (81) a bound of the type

$$\operatorname{mes}\left(\mathcal{G} \setminus \widehat{G}^*\right) \leq C'\,\gamma,$$

where C' is a constant depending on n, τ, $\operatorname{diam} F$, D, K, M, L, l, μ. We then get estimate (c), with $C = (2\pi)^n C'$. This constant may be explicited if desired. □

Remarks

1. All of the sequences introduced at the beginning of the proof have linear convergence. Of course, alternative choices for those sequences are possible provided the restrictions imposed by proposition 11 are fulfilled.
2. The reason of our choice of K_q and $\delta_1^{(q)}$ will be transparent in the next section, where we see that, for a small ν, the remainders decrease in an almost quadratic way.
3. The estimates of part (b) on the deformation of the perturbed invariant tori from the unperturbed ones are essentially the same of [22].

4.5 Fast convergence and nearly-invariant tori

We have established in the previous section the linear convergence to zero of the sizes of the successive remainders. This kind of convergence is enough (and very suitable) for the proof of the existence of invariant tori. But in the current section we show that the remainders actually decrease much faster, and we take advantage on this fact. Indeed, by stopping the iterative process at an appropriate step, we deduce that the domain obtained is full of nearly-invariant tori, i.e. Nekhoroshev-like estimates hold for the trajectories starting on these tori.

One may look our theorem F on effective stability as an attempt to make KAM and Nekhoroshev theorems closer. Indeed, we provide Nekhoroshev-like estimates which are very near, from a quantitative point of view, to KAM theorem. From the practical point of view, this result is more significative than KAM theorem. Indeed, if the coordinates of a given unperturbed invariant torus are known just approximately, up to a precision r, it is not possible to decide whether the frequency associated to this torus is Diophantine or not, and therefore one cannot deduce that this torus survives in the perturbation. In fact, in checking the Diophantine condition it has no sense to go farther than a finite order $K^{(r)}$ (tending to infinity as $r \to 0$). However, this finite test is enough to ensure that the torus is still included in the domain at an appropiate step of the iterative process and that this torus survives in the perturbation at least in the form of a nearly-invariant torus: a trajectory starting on this torus remains near to it up to a stability time which is exponentially long in $1/r$.

This result is similar to the one of [18], which does not however worry about optimal estimates. Moreover, in that paper the stability estimates are expressed in terms of the stability time, previously fixed, instead of r.

Another related result is obtained in [25] and [19], where it is shown that KAM tori are "sticky": estimates are given for the time to move away from a fixed KAM torus. The estimates are exponential in [25] and "superexponential" (but only for quasiconvex systems) in [19]. This result requires the transformation to normal form to hold in a full neighborhood of the given KAM torus, which is achieved in the quoted papers by carrying out the Kolmogorov's approach to KAM theorem instead of the Arnold's one. But our result seems in practice more useful since the existence of a KAM torus is not used for the estimates.

We notice that theorem E gives a large family of invariant tori if γ is small. But for large values of γ one cannot guarantee the preservation of any invariant tori even if (47) is satisfied, since the set $\widehat{G}_\gamma$ may be empty. Actually, there is a maximum value γ_0 such that $\widehat{G}_\gamma$ is empty for $\gamma > \gamma_0$ (in the case $n = 2$, the set $\widehat{G}_{\gamma_0}$ corresponds to the noble frequencies). Nevertheless, in theorem F the nearly-invariant tori are parametrized by a set $G_\gamma^{(r)}$ (defined below) containing $\widehat{G}_\gamma$ properly. Then, for some interval of values $\gamma > \gamma_0$ we may still ensure the existence of nearly-invariant tori.

Theorem F *Consider notations and hypothesis as in theorem E and assume also that*

$$\varepsilon \ \le\ \frac{\nu^2\sigma l^2\hat\rho^{2\tau+2}}{2^{2\tau+17}L^2M}\cdot\gamma^2, \tag{82}$$

where we define $\sigma := \min\limits_{s\ge 0}\dfrac{e^{\left(2-2^{1-\nu}\right)\cdot 2^{(1-\nu)s}}}{2^{(2\tau+1)s}} > 0.$ *Let*

$$0 \ <\ r\ \le\ r_0 := \frac{\hat\rho^{\tau+1}}{2^{\tau+2}M}\cdot\gamma$$

given, and write

$$\begin{aligned}
\widehat{G}^{(r)} = \widehat{G}^{(r)}_\gamma &:= \left\{ I\in\mathcal{G}-\frac{2\gamma}{\mu}\ :\ |k\cdot\omega(I)| \ge \frac{\gamma}{|k|_1^\tau}\ \ \forall k\in\mathbf{Z}^n,\ 0<|k|_1\le K^{(r)}\right\},\\
G^{(r)} = G^{(r)}_\gamma &:= \mathcal{U}_r\left(\widehat{G}^{(r)}\right),
\end{aligned}$$

where

$$K^{(r)} := \frac{2}{\hat\rho}\left(\frac{r_0}{r}\right)^{1/(\tau+1+\nu)}.$$

Then, there exist an analytic map $A^{(r)} : G^{(r)} \longrightarrow \mathcal{G}$ *and a real analytic canonical transformation* $\Upsilon^{(r)} : \mathcal{D}_{\left(\frac{\rho_1}{4},\frac{\nu l r}{2^{\tau+5}L}\right)}\left(A^{(r)}\left(G^{(r)}\right)\right) \longrightarrow \mathcal{D}_\rho(\mathcal{G})$ *such that, writing* $\mathcal{T}^{(r)} =$

$\Upsilon^{(r)} \circ \left(\mathrm{id} \times A^{(r)}\right)$, *any torus* $\mathcal{T}^{(r)}\left(\mathbf{T}^n \times \{I_0\}\right)$, *with* $I_0 \in G^{(r)}$ *has the property that, for every trajectory* $(\phi(t), I(t)) = \Upsilon^{(r)}\left(\hat{\phi}(t), \hat{I}(t)\right)$ *of* H *with* $(\phi(0), I(0))$ *belonging to this torus, one has:*

$$\left|\hat{I}(t) - A^{(r)}(I_0)\right| \leq \sqrt{\frac{2\varepsilon}{M}} \cdot \exp\left\{-\frac{1}{2}\left(\frac{r_0}{r}\right)^{(1-\nu)/(\tau+1+\nu)}\right\}, \tag{83}$$

$$\left|\hat{\phi}(t) - \left(\hat{\phi}(0) + \lambda^{(r)}(I_0)t\right)\right|_\infty \leq \frac{\nu\rho_1}{4}\left(\frac{r}{r_0}\right)^{\nu/(\tau+1+\nu)}, \tag{84}$$

for

$$|t| \leq \frac{\nu\rho_1}{51\sqrt{M\,\varepsilon}}\left(\frac{r}{r_0}\right)^{\nu/(2\tau+2)} \cdot \exp\left\{\frac{1}{2}\left(\frac{r_0}{r}\right)^{(1-\nu)/(\tau+1+\nu)}\right\}, \tag{85}$$

where the vector $\lambda^{(r)}(I_0)$ *is colinear to* $\omega(I_0)$*. Moreover, if* E *denotes the energy of the trajectory* $(\phi(t), I(t))$,

$$|E - h(I_0)| \leq \frac{2^{2n+1}\varepsilon}{\hat{\rho}} \cdot \exp\left\{-\left(\frac{r_0}{r}\right)^{(1-\nu)/(\tau+1+\nu)}\right\}.$$

Proof We go again into the iterative process of the proof of theorem E. Improving the argument given there for the linear estimate (1_q), we are going to see that the successive sizes of the remainders admit an almost quadratic estimate. We first notice that, choosing K as in the proof of theorem E, and taking into account the definition $\beta = \gamma/L$ and the inequalities $\hat{\rho} \leq 2/K$ and $\hat{\rho} \leq \nu\rho_1/16$, we deduce from (82) the inequality

$$\varepsilon \leq \frac{\nu^3\sigma l^2\rho_1\,\beta^2}{2^{20}MK^{2\tau+1}}\,. \tag{86}$$

We next prove that, for every $q \geq 0$,

$$\varepsilon_q \leq \frac{32\varepsilon}{\nu\rho_1} \cdot e^{-2^{(1-\nu)q}}. \tag{87}$$

Indeed, this is true for $q = 0$ by (52). Given $q \geq 1$ and assuming the estimate true for $q-1$, we are going to establish it for q. From (57) and the inequality $K\hat{\rho} \geq 1$, we deduce

$$K_q\delta_1^{(q)} \geq \frac{\nu K\rho_1}{8} \cdot 2^{(1-\nu)(q-1)} \geq \ln 2 + 2^{(1-\nu)(q-1)},$$

and therefore

$$e^{-K_q\delta_1^{(q)}} \leq \frac{1}{2}\,e^{-2^{(1-\nu)(q-1)}}.$$

Moreover, applying the inequalities (48) and (86), the definition of σ and the hypothesis that (87) holds for ε_{q-1},

$$\begin{aligned} \frac{14A_qK_q^\tau}{l_{q-1}\beta'_{q-1}\delta_2^{(q)}} \cdot \varepsilon_{q-1} &\leq \frac{2^8K_q^\tau}{\nu l\beta} \cdot \frac{64MK_q^{\tau+1}}{\nu l\beta} \cdot \frac{32\varepsilon}{\nu\rho_1} \cdot e^{-2^{(1-\nu)(q-1)}} \\ &\leq \frac{1}{2}\,e^{-\left(2^{1-\nu}-1\right)\cdot 2^{(1-\nu)(q-1)}}. \end{aligned} \tag{88}$$

Then, from (56) we deduce:

$$\varepsilon_q \leq \left(\frac{1}{2}\, e^{-2^{(1-\nu)(q-1)}} + \frac{1}{2}\, e^{-\left(2^{1-\nu}-1\right)\cdot 2^{(1-\nu)(q-1)}}\right) \cdot \varepsilon_{q-1} \leq e^{-\left(2^{1-\nu}-1\right)\cdot 2^{(1-\nu)(q-1)}} \cdot \varepsilon_{q-1}\,,$$

which gives estimate (87) for ε_q.

The stability time given in (79) can also be improved. Recall that we denote $x^{(q)}(t) = \left(\phi^{(q)}(t), I^{(q)}(t)\right)$ the trajectory of $H^{(q)}$ such that $x^{(q)}(0) = (\phi_0^*, I_0^*) \in \mathbf{T}^n \times G_q$. We deduce from the inequality (74) that, for $0 \leq t \leq T_q$,

$$\left|I^{(q)}(t) - I_0^*\right| \leq T_q \varepsilon_q$$

(T_q was introduced in (73)). Then, we may replace (75) by the next alternative bound:

$$\left|\dot{\phi}^{(q)}(t) - \omega^{(q)}\left(I_0^*\right)\right|_\infty \leq 2MT_q\varepsilon_q + \frac{\varepsilon_q}{c_{q+1}} \leq 5MT_q\varepsilon_q,$$

where (78) has been used. Then, the inequality (77) becomes

$$\delta_1^{(q+1)} \leq T_q^2 \cdot 5M\varepsilon_q,$$

and we obtain:

$$T_q \geq \min\left(\frac{\delta_2^{(q+1)}}{\varepsilon_q}, \sqrt{\frac{\delta_1^{(q+1)}}{5M\varepsilon_q}}\right) \;=\; \tilde{T}_q := \sqrt{\frac{\delta_1^{(q+1)}}{5M\varepsilon_q}}\,,$$

where we have used (76) to see that the minimum is given by the second term. Hence, for every initial condition $(\phi_0^*, I_0^*) \in \mathbf{T}^n \times G_q$, the corresponding trajectory of $H^{(q)}$ satisfies:

$$\left|I^{(q)}(t) - I_0^*\right| \;\leq\; \tilde{T}_q\varepsilon_q = \sqrt{\frac{\delta_1^{(q+1)}\varepsilon_q}{5M}}\,, \quad \left|\phi^{(q)}(t) - \left(\phi_0^* + \omega^{(q)}\left(I_0^*\right)t\right)\right|_\infty \leq \delta_1^{(q+1)}, \tag{89}$$

for

$$|t| \leq \tilde{T}_q\,. \tag{90}$$

This stability time is much better than the one of (79), because of the quadratic behaviour of ε_q. This estimate says that $\mathbf{T}^n \times \{I_0^*\}$ is a nearly-invariant torus of $H^{(q)}$ for every $I_0^* \in G_q$, since every trajectory starting at a point on this torus remains near a quasiperiodic motion with frequency vector $\omega^{(q)}\left(I_0^*\right)$ for a long time. Then, the torus $\Psi^{(q)}\left(\mathbf{T}^n \times \{I_0^*\}\right)$ is also nearly-invariant for the flow of H.

We are now going to choose $q = q(r) \geq 1$, as large as possible, such that $F_q \supset \Omega\left(G^{(r)}\right)$. Given $I_0 \in G^{(r)}$, there exists $I_0' \in \widehat{G}^{(r)}$ such that $|I_0' - I_0| \leq r$. We get $I_0 \in \mathcal{G} - \left(\frac{2\gamma}{\mu} - r\right)$ and hence

$$\Omega(I_0) \in F - \left(\beta - \frac{\mu r}{2L}\right) \tag{91}$$

by part (b) of lemma 8. On the other hand, for $0 < |k|_1 \leq K^{(r)}$,

$$|k \cdot \omega(I_0)| \geq \frac{\gamma}{|k|_1^\tau} - |k| \cdot Mr$$

and therefore

$$\left|\overline{k}\cdot\overline{\Omega}(I_0)+k_n\right| \geq \frac{1}{|\omega_n(I_0)|}\left(\frac{\gamma}{|k|_1^{\tau}}-|k|\cdot Mr\right) \geq \frac{\beta}{|k|_1^{\tau}}-\frac{|k|\cdot Mr}{L}\,. \tag{92}$$

We deduce from (91–92) that $J_0 \in F_q$ provided the following inequalities are fulfilled:

$$K^{(r)} \geq K_q\,, \qquad r \leq \frac{L}{M}\cdot\frac{\beta-\beta_q}{K_q^{\tau+1}}\,.$$

Noting that $\beta-\beta_q=\beta/2^{\nu q}$ and reminding that $K \leq 2/\hat{\rho}$, we see that it suffices to choose q such that the inequality

$$2^{(\tau+1+\nu)(q-1)} \leq \frac{r_0}{r} \tag{93}$$

holds. Hence, we choose $q \geq 1$ as the maximum integer such that this happens. We then have also

$$2^{(\tau+1+\nu)q} \geq \frac{r_0}{r}\,. \tag{94}$$

With this choice of q, we have $\Omega\left(G^{(r)}\right) \subset F_q$. We take $A^{(r)} := \left(\Omega^{(q)}\right)^{-1}\circ\Omega$, and note that $A^{(r)}\left(G^{(r)}\right) \subset G_q$. The transformation $\Upsilon^{(r)} := \Psi^{(q)}$ is defined on $\mathcal{D}_{\rho^{(q)}}(G_q)$, and we have the inequalities $\rho_1^{(q)} \geq \rho_1/4$ and

$$\rho_2^{(q)} = \frac{\nu l\beta}{32MK^{\tau+1}\cdot 2^{(\tau+1)q}} \geq \frac{\nu l\hat{\rho}^{\tau+1}\beta}{2^{\tau+6}M\cdot 2^{(\tau+1)q}} = \frac{\nu l r_0}{2^4 L\cdot 2^{(\tau+1)q}} \geq \frac{\nu l\, r}{2^{\tau+5}L}\,,$$

where we have used the inequality (93). This gives, in function of r, the complex domain where $\Upsilon^{(r)}$ is defined. For $(\phi_0, I_0) \in \mathbf{T}^n \times G^{(r)}$, we put

$$\mathcal{T}^{(r)}(\phi_0, I_0) = \Psi^{(q)}\left(\phi_0, I_0^*\right),$$

where $I_0^* = A^{(r)}(I_0) \in G_q$. If $(\phi(t), I(t))$ is a trajectory of H starting on the torus $\mathcal{T}^{(r)}\left(\mathbf{T}^n \times \{I_0\}\right)$, then $\left(\hat{\phi}(t), \hat{I}(t)\right)$ is a trajectory of $H^{(q)}$, with $\hat{I}(0) = I_0^*$. We can thus apply the stability estimate of (89–90). Using the inequalities (87), (48) and (93–94), we get the bounds

$$\begin{aligned}
\varepsilon_q &\leq \frac{32\varepsilon}{\nu\rho_1}\cdot\exp\left\{-\left(\frac{r_0}{r}\right)^{(1-\nu)/(\tau+1+\nu)}\right\},\\
\delta_1^{(q+1)} &\leq \frac{\nu\rho_1}{4\cdot 2^{\nu q}} \leq \frac{\nu\rho_1}{4}\left(\frac{r}{r_0}\right)^{\nu/(\tau+1+\nu)},\\
\delta_1^{(q+1)} &\geq \frac{\nu\rho_1}{16\cdot 2^{\nu(q-1)}} \geq \frac{\nu\rho_1}{16}\left(\frac{r}{r_0}\right)^{\nu/(\tau+1+\nu)}.
\end{aligned}$$

Including these bounds in (89–90), we get (83–85). Note that we put $\lambda^{(r)}(I_0) = \omega^{(q)}\left(I_0^*\right)$, which is colinear to $\omega(I_0)$ since $\Omega^{(q)}\left(I_0^*\right) = \Omega(I_0)$. Finally, the energy of the trajectory $(\phi(t), I(t))$ is

$$E = H(\phi(0), I(0)) = H^{(q)}\left(\hat{\phi}(0), I_0^*\right) = h(I_0) + R^{(q)}\left(\hat{\phi}(0), I_0^*\right)$$

since $h^{(q)}(I_0^*) = h(I_0)$. Then,

$$|E - h(I_0)| \le \left| R^{(q)}\left(\hat{\phi}(0), I_0^*\right)\right| \le \left| R_0^{(q)} \right|_{G_q} + \left| \frac{\partial R^{(q)}}{\partial \phi} \right|_{G_q} \cdot \pi^n \le \eta_q + \pi^n \varepsilon_q \,.$$

To get a quadratic estimate for η_q, we proceed as in (59) but we now use (88), (48) and (87):

$$\eta_q \le \frac{7 A_q K_q^\tau}{c_q l_{q-1} \beta'_{q-1}} \cdot \varepsilon_{q-1}^2 \le \frac{\delta_1^{(q)} \varepsilon_{q-1}}{2} \cdot \frac{1}{2} \, e^{-\left(2^{1-\nu}-1\right) \cdot 2^{(1-\nu)(q-1)}} \le 2\varepsilon \cdot e^{-2^{(1-\nu)q}} .$$

Then,

$$|E - h(I_0)| \le \left(2 + \frac{32 \pi^n}{\nu \rho_1} \right) \varepsilon \cdot e^{-2^{(1-\nu)q}} \le \frac{2^{2n+1} \varepsilon}{\hat{\rho}} \cdot \exp\left\{ - \left(\frac{r_0}{r} \right)^{(1-\nu)/(\tau+1+\nu)} \right\}. \quad \square$$

Remarks

1. To give an idea of how this theorem should be applied in practice, assume that I_0 is the action of a given unperturbed torus, for which we just know an approximation I'_0, with $|I'_0 - I_0| \le r$. If $I'_0 \in \widehat{G}_\gamma^{(r)}$, i.e. the frequency ratios of $\omega(I'_0)$ satisfy the Diophantine condition (2) up to order $K^{(r)}$, then the invariant torus corresponding to the action I_0 survives as a nearly-invariant torus.

2. We have omitted statements like parts (b) and (c) of theorem E because they would be exactly the same, with $\mathcal{T}^{(r)}$ instead of $\mathcal{T}$.

3. If the size ε of the perturbation is fixed, we could take r small and the stability estimate given in (83–85) is then much better than the one provided by Nekhoroshev theorem. This is due to the fact that the estimate has been expressed in the transformed coordinates $\left(\hat{\phi}, \hat{I}\right)$ provided by the canonical transformation $\Upsilon^{(r)}$. These coordinates are better because the nearly-invariant tori are given by equations $\hat{I} = \text{const}$. In Nekhoroshev theorem, the stability estimate includes the coordinate change because it is expressed in the original coordinates.

4. The comparison between the estimates (83–84) for action and angular variables shows that the separation from a given torus remains much smaller than the separation from a linear flow inside this torus.

5. The stability exponent given in (85) is larger (as near to $1/(\tau+1)$ as wanted) if we choose the parameter ν near to zero.

6. Roughly speaking, the set $G_\gamma^{(r)}$ parametrizing the nearly-invariant tori has a very large boundary for r small. However, the "area" of this boundary is finite, which means that the set $G_\gamma^{(r)}$ is not as strange as the Cantorian set $\widehat{G}_\gamma$ provided by KAM theorem. An alternative way to express this fact is used in [18]: the set of nearly-invariant tori contains balls of a suitable radius, and hence it contains inner points.

We can also get "superexponential" stability estimates, like in [19], by means of an alternative approach. However, we need to assume that the unperturbed Hamiltonian h is quasiconvex. Applying the iterative process of KAM theorem, our starting Hamiltonian H is transformed after q steps into $H^{(q)} = h^{(q)} + R^{(q)}$. If h is m-quasiconvex with $m > 0$ and we assume $\varepsilon = \mathcal{O}(m)$, then $h^{(q)}$ is also quasiconvex and Nekhoroshev theorem may be applied to $H^{(q)}$. In this way, we can obtain for every trajectory $\left(\phi^{(q)}(t), I^{(q)}(t)\right)$ of $H^{(q)}$, with initial condition in $\mathbf{T}^n \times G_q$, a stability estimate of the type

$$\left|I^{(q)}(t) - I^{(q)}(0)\right| \le R \qquad \text{for } \; |t| \le T,$$

with

$$R \sim \varepsilon_q^{1/2n}, \qquad T \sim \exp\left\{\left(\frac{1}{\varepsilon_q}\right)^{1/2n}\right\}.$$

Choosing $q = q(r)$ as in (93–94), we get

$$R \sim \left(\varepsilon \cdot \exp\left\{-\left(\frac{1}{r}\right)^c\right\}\right)^{1/2n}, \qquad T \sim \exp\left\{\left(\frac{1}{\varepsilon} \cdot \exp\left\{\left(\frac{1}{r}\right)^c\right\}\right)^{1/2n}\right\},$$

where we have put $c = (1-\nu)/(\tau+1+\nu)$. The stability radius R and the stability time T substitute the ones obtained in (83) and (85), respectively.

5 Appendix: proofs of the technical lemmas

Proof of lemma 2 Given $(\phi_0, I_0) \in \mathcal{D}_{\rho - t\delta}(G)$, let $(\phi(s), I(s)) = \Phi_s(\phi_0, I_0)$. First, we prove that

$$|\phi(s) - \phi_0|_\infty \le t \left\|\frac{\partial W}{\partial I}\right\|_{G,\rho,\infty}, \qquad |I(s) - I_0| \le t \left\|\frac{\partial W}{\partial \phi}\right\|_{G,\rho,1}, \tag{95}$$

for $0 \le s \le t$. Let s_0 be the supremum of the $s \ge 0$ satisfying both inequalities in (95). Clearly $s_0 > 0$, and one of these inequalities is an equality for $s = s_0$. On the other hand, we have $(\phi(s), I(s)) \in \mathcal{D}_\rho(G)$ for $0 \le s \le s_0$. From the mean value theorem,

$$|\phi(s_0) - \phi_0|_\infty \le s_0 \cdot \sup_{0 \le s \le s_0} \left|\frac{\partial W}{\partial I}(\phi(s), I(s))\right|_\infty \le s_0 \left\|\frac{\partial W}{\partial I}\right\|_{G,\rho,\infty}, \tag{96}$$

$$|I(s_0) - I_0| \le s_0 \cdot \sup_{0 \le s \le s_0} \left|\frac{\partial W}{\partial \phi}(\phi(s), I(s))\right| \le s_0 \left\|\frac{\partial W}{\partial \phi}\right\|_{G,\rho} \le s_0 \left\|\frac{\partial W}{\partial \phi}\right\|_{G,\rho,1}. \tag{97}$$

Thus, $s_0 \ge t$ and (95) is true. This implies that Φ_t is defined in $\mathcal{D}_{\rho - t\delta}(G)$ and that the bound (a) holds. We can deduce the inclusion (b) from the fact that Φ_{-t} is the flow at time t of $-W$.

To see (c), note that $f \circ \Phi_t$ is defined in $\mathcal{D}_{\rho - t\delta}(G)$. Since W is analytic, $f \circ \Phi_t$ is also analytic in t, and hence the Lie series expansion (4) for $r_m(f, W, t)$ holds. Given

$l \geq m+1$, let $\eta = \delta/(l-m)$. For $j = m+1, \ldots, l$, we have

$$\begin{aligned} \left\| L_W^j f \right\|_{G,\rho-(j-m)t\eta} &\leq \frac{2}{c} \left\| D\left(L_W^{j-1} f\right) \right\|_{G,\rho-(j-m)t\eta\,,\,c} \cdot \|DW\|_{G,\rho,c} \\ &\leq \frac{2}{t\hat{\eta}_c} \left\| L_W^{j-1} f \right\|_{G,\rho-(j-1-m)t\eta} \cdot \|DW\|_{G,\rho,c} \,. \end{aligned}$$

Thus,

$$\left\| L_W^l f \right\|_{G,\rho-t\delta} \leq \left(\frac{2\,\|DW\|_{G,\rho,c}}{t\hat{\eta}_c} \right)^{l-m} \cdot \|L_W^m f\|_{G,\rho} \leq (l-m)! \left(\frac{2e\,\|DW\|_{G,\rho,c}}{t\hat{\delta}_c} \right)^{l-m} \cdot \|L_W^m f\|_{G,\rho} \,,$$

where we have used that $k^k \leq e^k \cdot k!$ for $k \geq 1$. In this way, the bound that we obtain for $\|r_m(f,W,t)\|_{G,\rho-t\delta}$ is

$$\sum_{l=m}^{\infty} \frac{t^l}{l!} \left\| L_W^l f \right\|_{G,\rho-t\delta} \leq \left[\sum_{l=m}^{\infty} \frac{(l-m)!}{l!} \cdot \left(\frac{2e\,\|DW\|_{G,\rho,c}}{\hat{\delta}_c} \right)^{l-m} \right] \cdot t^m \, \|L_W^m f\|_{G,\rho} \,,$$

this series being convergent for $\|DW\|_{G,\rho,c} < \hat{\delta}_c/2e$. □

Remark The bounds (96–97) are based on the special structure of Hamiltonian equations. Our choice of the ∞-norm for the angular variables was motivated by (96). Concerning the action variables, the best choice would be, according to (97), the 1-norm, but our use of Euclidean geometry in the geometric parts of Nekhoroshev and KAM theorems made us choose the 2-norm.

Proof of lemma 5 Assume first $K\rho_1 \geq 1$. Let $\delta = \rho/3$. From (18), we see that

$$A = 1 + \frac{2M}{\alpha} \cdot \frac{\rho_2}{\rho_1} \leq 1 + \frac{1}{K\rho_1} \leq 2.$$

Then, theorem B provides a canonical transformation $\Psi : \mathcal{D}_{\frac{2\rho}{3}}(G) \longrightarrow \mathcal{D}_\rho(G)$ such that $H \circ \Psi = h + Z^* + R^*$, with $Z^* = Z^*(I)$, and

$$\|DR^*\|_{G,\frac{2\rho}{3},c} \leq 3\,e^{-\frac{K\rho_1}{6}} \cdot \|DR\|_{G,\rho,c} \,.$$

Let us denote

$$\eta = \frac{8}{\alpha} \|DR\|_{G,\rho,c} \,.$$

By estimate (c) of theorem B and the second condition of (23),

$$|\Psi - \mathrm{id}|_{G,\frac{2\rho}{3},c} \leq \eta \leq \frac{\rho_2}{15K\rho_1} \leq \frac{\rho_2}{15} \,.$$

Moreover, by the inclusion (d) of theorem B, we have $\Psi\left(\mathcal{D}_{\frac{2\rho}{3}}(G)\right) \supset \mathcal{D}_{\frac{\rho}{2}}(G) \supset \mathbf{T}^n \times G$ and therefore we can write $(\phi(0), I(0)) = \Psi(\phi_0^*, I_0^*)$ and, since Ψ preserves real domains,

$(\phi_0^*, I_0^*) \in \mathbf{T}^n \times \mathcal{V}_{\frac{\rho_2}{15}}(G)$. Let $(\phi^*(t), I^*(t))$ be the trajectory of $H \circ \Psi$ with $(\phi^*(0), I^*(0)) = (\phi_0^*, I_0^*)$. Let

$$T = \inf \{t > 0 \ : \ |I^*(t) - I^*(0)| > \eta\}$$

(the procedure for negative times is exactly the same). For $0 \le t \le T$, we obtain $(\phi^*(t), I^*(t)) \in \mathcal{D}_{\frac{2\rho}{3}}(G)$. Since Ψ takes the motions of $H \circ \Psi$ into motions of H, we get that $(\phi(t), I(t)) = \Psi(\phi^*(t), I^*(t))$ is also defined for $0 \le t \le T$, and obtain the estimate

$$|I(t) - I(0)| \le |I(t) - I^*(t)| + |I^*(t) - I^*(0)| + |I^*(0) - I(0)| \le 3\eta.$$

Next, we proceed to obtain a lower bound for T. Let $\Delta I^* = I^*(T) - I^*(0)$. Clearly, $|\Delta I^*| = \eta$. On the other hand, by using the form of the Hamiltonian equations and the fact that the normal form Z^* only depends on the action variables, we obtain

$$|\Delta I^*| \le T \cdot \left\| \frac{\partial R^*}{\partial \phi} \right\|_{G, \frac{2\rho}{3}} \le T \cdot \|DR^*\|_{G, \frac{2\rho}{3}, c} \le T \cdot 3\, e^{-\frac{K\rho_1}{6}} \cdot \|DR\|_{G,\rho,c} \ ,$$

and it follows that

$$T \ge \frac{\eta}{3\, \|DR\|_{G,\rho,c}}\, e^{\frac{K\rho_1}{6}} \ge \frac{2}{\alpha}\, e^{\frac{K\rho_1}{6}}.$$

For $K\rho_1 < 1$, one obtains, by working in original coordinates,

$$|I(t) - I(0)| \le \frac{24}{\alpha}\, \|DR\|_{G,\rho,c} \qquad \text{for} \ \ |t| \le \frac{24}{\alpha} \ ,$$

and it is then easy to see that this stability time is longer than the one proclaimed in (24). □

Proof of lemma 6 First we assume $K\rho_1 \ge M/m$. Let $\delta = \rho/3$. Like in the proof of lemma 5, we have $A \le 2$. Note that

$$\|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c} \le \frac{m\rho_2^2}{350\rho_1} \le \frac{\alpha\rho_2}{122K\rho_1} \ . \tag{98}$$

Then, theorem B provides a canonical transformation $\Psi : \mathcal{D}_{\frac{2\rho}{3}}(G) \longrightarrow \mathcal{D}_\rho(G)$ such that $H \circ \Psi = h + Z^* + R^*$, with $Z^* \in \mathbf{R}(\mathcal{M}, K)$, and

$$\|DZ^*\|_{G,\frac{2\rho}{3},c} + \|DR^*\|_{G,\frac{2\rho}{3},c} \le \|DZ\|_{G,\rho,c} + 2\, \|DR\|_{G,\rho,c} \ ,$$

$$\|DR^*\|_{G,\frac{2\rho}{3},c} \le 3\, e^{-\frac{K\rho_1}{6}} \cdot \|DR\|_{G,\rho,c} \ .$$

Moreover, by estimate (c) of theorem B and the inequality (98),

$$|\Psi - \mathrm{id}|_{G,\frac{2\rho}{3},c} \le \frac{8}{\alpha}\, \|DR\|_{G,\rho,c} \le \frac{\rho_2}{15K\rho_1} \le \frac{m\rho_2}{15M} \ .$$

Like in lemma 5, we can write $(\phi(0), I(0)) = \Psi(\phi_0^*, I_0^*)$, with $(\phi_0^*, I_0^*) \in \mathbf{T}^n \times \mathcal{V}_{\frac{m\rho_2}{15M}}(G)$. Let $(\phi^*(t), I^*(t))$ be the trajectory of $H \circ \Psi$ with $(\phi^*(0), I^*(0)) = (\phi_0^*, I_0^*)$. Let

$$T = \inf \left\{ t > 0 \ : \ |I^*(t) - I^*(0)| > \frac{\rho_2}{2} \right\}.$$

For $0 \leq t \leq T$, we have $(\phi^*(t), I^*(t)) \in \mathcal{D}_{\frac{2\rho}{3}}(G)$. We then obtain the estimate

$$|I(t) - I(0)| \leq |I(t) - I^*(t)| + |I^*(t) - I^*(0)| + |I^*(0) - I(0)| \leq \frac{m\rho_2}{15M} + \frac{\rho_2}{2} + \frac{m\rho_2}{15M} \leq \rho_2.$$

We introduce the notations

$$\Delta\phi^* = \phi^*(T) - \phi^*(0), \qquad \Delta I^* = I^*(T) - I^*(0),$$

and, for a function $f(\phi, I)$,

$$\Delta f = f(\phi^*(T), I^*(T)) - f(\phi^*(0), I^*(0)).$$

The definition of T clearly implies that $|\Delta I^*| = \rho_2/2$.

We notice that, since Z^* is in normal form with respect to $\mathcal{M}$, it does not contribute to ΔI^* in any direction lying in $\mathcal{M}^\perp$. More precisely, let P denote the orthogonal projection onto the one-dimensional space $\langle \Pi_{\mathcal{M}} \omega(I(0)) \rangle$. By the specific form of the Hamiltonian equations, we have

$$P\Delta I^* = -\int_0^T P\left(\frac{\partial R^*}{\partial \phi}(\phi^*(t), I^*(t))\right) dt$$

and it follows that the $\Pi_{\mathcal{M}}\omega(I(0))$-component of the vector ΔI^* is small up to an exponentially long time:

$$|P\Delta I^*| \leq T \cdot \left\| \frac{\partial R^*}{\partial \phi} \right\|_{G, \frac{2\rho}{3}} \leq T \cdot \|DR^*\|_{G, \frac{2\rho}{3}, c} \leq T \cdot 3\, e^{-\frac{K\rho_1}{6}} \cdot \|DR\|_{G,\rho,c} \,. \tag{99}$$

To bound the whole vector ΔI^* we use the quasiconvexity condition on h. By Taylor formula, one has

$$\Delta h = \omega(I_0^*) \cdot \Delta I^* + \int_0^1 (1-s) \frac{\partial^2 h}{\partial I^2}(I_0^* + s\Delta I^*)\,(\Delta I^*, \Delta I^*)\, ds. \tag{100}$$

We notice that, since Ψ preserves real domains, $I_0^* + s\Delta I^* \in \mathcal{U}_{\rho_2}(G)$ for $0 \leq s \leq 1$. For a fixed s, we write $\Delta I^* = P_s \Delta I^* + Q_s \Delta I^*$, where P_s and Q_s denote the orthogonal projections onto $\langle \omega(I_0^* + s\Delta I^*) \rangle$ and $\langle \omega(I_0^* + s\Delta I^*) \rangle^\perp$, respectively. Thus, applying the quasiconvexity condition at the point $I_0^* + s\Delta I^*$ (and this is the only time that we use it), we obtain

$$\left| \frac{\partial^2 h}{\partial I^2}(I_0^* + s\Delta I^*)\,(Q_s\Delta I^*, Q_s\Delta I^*) \right| \geq m\, |Q_s \Delta I^*|^2 .$$

We deduce that

$$\begin{aligned}
&\left| \frac{\partial^2 h}{\partial I^2}(I_0^* + s\Delta I^*)\,(\Delta I^*, \Delta I^*) \right| \\
&\qquad \geq m\,|Q_s\Delta I^*|^2 - 2M\,|P_s\Delta I^*| \cdot |Q_s\Delta I^*| - M\,|P_s\Delta I^*|^2 \\
&\qquad = m\,|\Delta I^*|^2 - (m\,|P_s\Delta I^*| + 2M\,|Q_s\Delta I^*| + M\,|P_s\Delta I^*|) \cdot |P_s\Delta I^*| \\
&\qquad \geq m\,|\Delta I^*|^2 - 4M\,|\Delta I^*| \cdot |P_s\Delta I^*| .
\end{aligned}$$

From formula (100) we obtain

$$\frac{m}{2}\left|\Delta I^*\right|^2 \le |\Delta h| + \left|\omega\left(I_0^*\right)\cdot\Delta I^*\right| + 4M\left|\Delta I^*\right|\int_0^1 (1-s)\left|P_s\Delta I^*\right|\,ds,$$

and hence

$$\frac{m\rho_2^2}{8} \le |\Delta h| + \left|\omega\left(I_0^*\right)\cdot\Delta I^*\right| + 2M\rho_2\int_0^1 (1-s)\left|P_s\Delta I^*\right|\,ds. \tag{101}$$

Next we bound the terms appearing in the right hand side of (101). To bound $|P_s\Delta I^*|$ we use that the vector $\omega\left(I_0^*+s\Delta I^*\right)$ is near to $\Pi_{\mathcal{M}}\omega(I(0))$. More precisely, we apply the following property: given $v,v'\in\mathbf{R}^n$, if $P_{(v)}$ and $P_{(v')}$ denote the orthogonal projections onto the one-dimensional spaces $\langle v\rangle$ and $\langle v'\rangle$, respectively, then for every vector $u\in\mathbf{R}^n$ one has:

$$\left|P_{(v)}u - P_{(v')}u\right| \le \frac{4\left|v-v'\right|}{|v|}\cdot|u|\,.$$

First we notice that, from lemma 3 and the fact that $\mathcal{M}\ne\mathbf{Z}^n$ and $K\ge1$, we deduce $|\omega(I)|\ge\alpha/2$ for $I\in\mathcal{U}_{\rho_2}(G)$. Then, applying the property, one has:

$$\begin{aligned}
\left|P_s\Delta I^* - P\Delta I^*\right| &\le \frac{4}{\left|\omega\left(I_0^*+s\Delta I^*\right)\right|}\left|\omega\left(I_0^*+s\Delta I^*\right)-\Pi_{\mathcal{M}}\omega(I(0))\right|\cdot\left|\Delta I^*\right| \\
&\le \frac{4\rho_2}{\alpha}\left|\omega\left(I_0^*+s\Delta I^*\right)-\Pi_{\mathcal{M}}\omega(I(0))\right|.
\end{aligned}$$

Using that $\omega(G)$ is η-close to $\mathcal{M}$-resonances, we get

$$\begin{aligned}
\left|\omega\left(I_0^*+s\Delta I^*\right)-\Pi_{\mathcal{M}}\omega(I(0))\right| &\le M\left|I_0^*+s\Delta I^*-I(0)\right| + \left|\omega(I(0))-\Pi_{\mathcal{M}}\omega(I(0))\right| \\
&\le \left(\frac{m}{15}+\frac{sM}{2}\right)\rho_2+\eta \;\le\; \left(\frac{m}{12}+\frac{sM}{2}\right)\rho_2.
\end{aligned}\tag{102}$$

Thus,

$$\left|P_s\Delta I^*\right| \le \left|P\Delta I^*\right| + \left|P_s\Delta I^*-P\Delta I^*\right| \le \left|P\Delta I^*\right| + \frac{4\rho_2^2}{\alpha}\left(\frac{m}{12}+\frac{sM}{2}\right)$$

and we obtain, using the first condition of (26),

$$\int_0^1(1-s)\left|P_s\Delta I^*\right|\,ds \le \frac{1}{2}\left|P\Delta I^*\right| + \frac{4\rho_2^2}{\alpha}\left(\frac{m}{24}+\frac{M}{12}\right) \le \frac{1}{2}\left|P\Delta I^*\right| + \frac{m\rho_2}{96M}\,.$$

To bound $\left|\omega\left(I_0^*\right)\cdot\Delta I^*\right|$, we put $s=0$ in (102):

$$\left|\omega\left(I_0^*\right)\cdot\Delta I^*\right| \le \left|\Pi_{\mathcal{M}}\omega(I(0))\right|\cdot\left|P\Delta I^*\right| + \left|\omega\left(I_0^*\right)-\Pi_{\mathcal{M}}\omega(I(0))\right|\cdot\left|\Delta I^*\right| \le L\left|P\Delta I^*\right| + \frac{m\rho_2^2}{24}\,.$$

Finally, by energy conservation,

$$\begin{aligned}
\Delta h \;=\; -\Delta\left(Z^*+R^*\right) \;=\; -\int_0^1 \Bigg(&\frac{\partial\left(Z^*+R^*\right)}{\partial\phi}\left(\phi_0^*+s\Delta\phi^*, I_0^*+s\Delta I^*\right)\cdot\Delta\phi^* \\
&+ \frac{\partial\left(Z^*+R^*\right)}{\partial I}\left(\phi_0^*+s\Delta\phi^*, I_0^*+s\Delta I^*\right)\cdot\Delta I^*\Bigg)\,ds
\end{aligned}$$

and we deduce

$$\begin{aligned}
|\Delta h| &\le \left\| \frac{\partial (Z^* + R^*)}{\partial \phi} \right\|_{G,\frac{2\rho}{3},1} \cdot |\Delta \phi^*|_\infty + \left\| \frac{\partial (Z^* + R^*)}{\partial I} \right\|_{G,\frac{2\rho}{3}} \cdot |\Delta I^*| \\
&\le \left(\pi + \frac{\sqrt{n}}{c} \cdot \frac{\rho_2}{2} \right) \cdot \| D\,(Z^* + R^*) \|_{G,\frac{2\rho}{3},c} \\
&\le \left(2\pi + \sqrt{n}\,\rho_1 \right) \cdot \left(\|DZ\|_{G,\rho,c} + \|DR\|_{G,\rho,c} \right) \\
&\le \frac{(2\pi+1) m\rho_2^2}{350} \le \frac{m\rho_2^2}{48} \,.
\end{aligned}$$

We insert all of these estimates in (101) and obtain:

$$\frac{m\rho_2^2}{8} \le \left(\frac{1}{48} + \frac{1}{24} + \frac{1}{48} \right) m\rho_2^2 + (L + M\rho_2)\ |P\Delta I^*| \le \frac{m\rho_2^2}{12} + \frac{49L}{48}\ |P\Delta I^*|\,,$$

where we used that $M\rho_2 \le L/48$ since $\alpha \le L$. By estimate (99), it follows that

$$T \ge \frac{m\rho_2^2}{74L\,\|DR\|_{G,\rho,c}}\, e^{\frac{K\rho_1}{6}} \ge \frac{m\rho_2^2}{74L\,\|DR\|_{G,\rho,c}}\, e^{\frac{mK\rho_1}{6M}} \,.$$

For $K\rho_1 < M/m$, one may work in original coordinates. Then, one obtains:

$$|I(t) - I(0)| \le \rho_2 \qquad \text{for } \ |t| \le \frac{\rho_2}{\|DR\|_{G,\rho,c}} \,.$$

It is not hard to check that this stability time is longer than the one proclaimed in (27). □

Proof of lemma 8 A simple computation gives, for $I \in G$ and $v \in \mathbf{R}^n$,

$$\frac{\partial \Omega}{\partial I}(I)\, v = \left(\frac{\partial \overline{\Omega}}{\partial I}(I)\, v \,,\ \frac{\partial \Omega_n}{\partial I}(I)\, v \right) = \left(\frac{1}{\omega_n(I)} \left(\frac{\partial \overline{\omega}}{\partial I}(I)\, v - \frac{\frac{\partial \omega_n}{\partial I}(I)\, v}{\omega_n(I)}\, \overline{\omega}(I) \right) \,,\ a\,\omega(I) \cdot v \right).$$

We have the bound

$$\begin{aligned}
\left| \frac{\partial \overline{\Omega}}{\partial I}(I)\, v \right| &\le \frac{1}{\omega_n(I)^2} \left(\left| \frac{\partial \overline{\omega}}{\partial I}(I)\, v \right| \cdot |\omega_n(I)| + \left| \frac{\partial \omega_n}{\partial I}(I)\, v \right| \cdot |\overline{\omega}(I)| \right) \\
&\le \frac{1}{\omega_n(I)^2} \left| \frac{\partial \omega}{\partial I}(I)\, v \right| \cdot |\omega(I)| \ \le\ \frac{M\,|\omega(I)|}{l^2}\, |v|
\end{aligned} \tag{103}$$

and we then get estimate (a):

$$\left| \frac{\partial \Omega}{\partial I} \right|_G \le \sqrt{\left(\frac{ML}{l^2} \right)^2 + (aL)^2} \le 2La. \tag{104}$$

To prove parts (b) and (c) we use the isoenergetic condition. For any $v \in \langle \omega(I) \rangle^\perp$, we have

$$\begin{aligned}
\left| \frac{\partial \overline{\Omega}}{\partial I}(I)\, v \right| &= \left| \frac{1}{\omega_n(I)} \left(\frac{\partial \overline{\omega}}{\partial I}(I)\, v - \frac{\frac{\partial \omega_n}{\partial I}(I)\, v}{\omega_n(I)}\, \overline{\omega}(I) \right) \right| \\
&= \left| \frac{1}{\omega_n(I)} \left(\frac{\partial \omega}{\partial I}(I)\, v - \frac{\frac{\partial \omega_n}{\partial I}(I)\, v}{\omega_n(I)}\, \omega(I) \right) \right| \ \ge\ \frac{\mu}{|\omega_n(I)|}\, |v| \,,
\end{aligned} \tag{105}$$

where we have used the fact that the vector

$$\frac{\partial\omega}{\partial I}(I)\,v-\frac{\frac{\partial\omega_n}{\partial I}(I)\,v}{\omega_n(I)}\,\omega(I)$$

has its n-th component vanishing. Moreover, the isoenergetic nondegeneracy has been used to bound the size of this vector from below. Now, consider an orthonormal basis $e_1,\dots,e_n$ of $\mathbf{R}^n$ with the first $n-1$ vectors belonging to $\langle\omega(I)\rangle^\perp$ and the last one belonging to $\langle\omega(I)\rangle$. Let $P=\left(\overline{P},P_n\right)$ be the $(n\times n)$-matrix having these vectors as columns. Let us write:

$$\frac{\partial\Omega}{\partial I}(I)\,P=\begin{pmatrix}\frac{\partial\overline{\Omega}}{\partial I}(I)\,\overline{P} & \frac{\partial\overline{\Omega}}{\partial I}(I)\,P_n\\ \frac{\partial\Omega_n}{\partial I}(I)\,\overline{P} & \frac{\partial\Omega_n}{\partial I}(I)\,P_n\end{pmatrix}=\begin{pmatrix}A & \overline{b}\\ 0 & b_n\end{pmatrix}. \tag{106}$$

It follows directly from (105) that

$$|A\overline{v}|\geq\frac{\mu}{|\omega_n(I)|}\,|\overline{v}|\qquad\forall\overline{v}\in\mathbf{R}^{n-1}.$$

Note also that $\left|\overline{b}\right|\leq M\,|\omega(I)|\,/l^2$ by (103). Moreover, $b_n=a\,\omega(I)\cdot e_n$ and hence $|b_n|=a\,|\omega(I)|$. Then, computing the inverse of the matrix (106) and carrying out a rough bound on its norm,

$$\begin{aligned}\left|\left(\frac{\partial\Omega}{\partial I}(I)\right)^{-1}\right| &\leq \left|\begin{pmatrix}A^{-1} & -\frac{1}{b_n}A^{-1}\overline{b}\\ 0 & \frac{1}{b_n}\end{pmatrix}\right| \leq \left|A^{-1}\right|+\frac{1}{|b_n|}\left|A^{-1}\right|\left|\overline{b}\right|+\frac{1}{|b_n|}\\ &\leq \frac{|\omega_n(I)|}{\mu}+\frac{1}{a\,|\omega(I)|}\cdot\frac{|\omega_n(I)|}{\mu}\cdot\frac{M\,|\omega(I)|}{l^2}+\frac{1}{a\,|\omega(I)|}\\ &\leq \frac{L}{\mu}+\frac{LM}{l^2\mu a}+\frac{1}{al}\;\leq\;\frac{L}{\mu}\left(1+\frac{2M}{l^2a}\right).\end{aligned} \tag{107}$$

This estimate implies (b), by our condition on a. To obtain the lower bound (c) for the determinant, we take into account the expression (106) again:

$$\left|\det\left(\frac{\partial\Omega}{\partial I}(I)\right)\right|\geq\left(\frac{\mu}{|\omega_n(I)|}\right)^{n-1}\cdot a\,|\omega(I)|\geq\frac{\mu^{n-1}a}{L^{n-2}}\,.$$

Finally, we prove (d). For $I\in G$ and $u,v\in\mathbf{R}^n$,

$$\begin{aligned}\frac{\partial^2\Omega}{\partial I^2}(I)\,(u,v)&=\left(\frac{\partial^2\overline{\Omega}}{\partial I^2}(I)\,(u,v),\frac{\partial^2\Omega_n}{\partial I^2}(I)\,(u,v)\right)\\ &=\left(\frac{\frac{\partial^2\overline{\omega}}{\partial I^2}(I)\,(u,v)}{\omega_n(I)}-\frac{\left(\frac{\partial\omega_n}{\partial I}(I)\,u\right)\left(\frac{\partial\overline{\omega}}{\partial I}(I)\,v\right)+\left(\frac{\partial\omega_n}{\partial I}(I)\,v\right)\left(\frac{\partial\overline{\omega}}{\partial I}(I)\,u\right)}{\omega_n(I)^2}\right.\\ &\qquad\left.-\frac{\left(\frac{\partial^2\omega_n}{\partial I^2}(I)\,(u,v)\right)\overline{\omega}(I)}{\omega_n(I)^2}+\frac{2\left(\frac{\partial\omega_n}{\partial I}(I)\,u\right)\left(\frac{\partial\omega_n}{\partial I}(I)\,v\right)\overline{\omega}(I)}{\omega_n(I)^3}\,,\,a\,\frac{\partial^2h}{\partial I^2}(I)\,(u,v)\right).\end{aligned}$$

We can bound this expression like in (103–104):

$$\left|\frac{\partial^2\overline{\Omega}}{\partial I^2}(I)\,(u,v)\right|\;\leq\;\frac{1}{\omega_n(I)^2}\left(\left|\frac{\partial^2\omega}{\partial I^2}(I)\,(u,v)\right|\cdot|\omega(I)|+\left|\frac{\partial\omega}{\partial I}(I)\,u\right|\cdot\left|\frac{\partial\omega}{\partial I}(I)\,v\right|\right)$$

$$
\begin{aligned}
&+\frac{2}{|\omega_n(I)|^3}\left|\frac{\partial\omega_n}{\partial I}(I)\,u\right|\cdot\left|\frac{\partial\omega_n}{\partial I}(I)\,v\right|\cdot|\overline{\omega}(I)| \\
\le\ &\left(\frac{M'L}{l^2}+\frac{3M^2L}{l^3}\right)|u|\,|v|\,,
\end{aligned}
$$

and we deduce estimate (d):

$$
\left|\frac{\partial^2\Omega}{\partial I^2}\right|_G \le \sqrt{\left(\frac{M'L}{l^2}+\frac{3M^2L}{l^3}\right)^2+(aM)^2} \le \left(\frac{M'}{2M}+\frac{3M}{l}\right)La. \quad \square
$$

Remark If the condition $a \ge 2M/l^2$ is removed, then estimate (b) has to be substituted by (107), which is actually worse if a is taken too small.

Proof of lemma 9 Clearly, it suffices to check the result for the vectors $v \in \langle\tilde{\omega}\rangle^{\perp}$ such that $|v| = 1$. One has $|\omega\cdot v| \le \varepsilon$ for these vectors. Writing $v = v_1 + v_2$, with $v_1 \in \langle\omega\rangle^{\perp}$ and $v_2 \in \langle\omega\rangle$, one deduces that

$$
|v_2| \le \frac{\varepsilon}{|\omega|}\,, \qquad |v_1| \ge 1-\frac{\varepsilon}{|\omega|}\,.
$$

By the hypothesis,

$$
|Av+\lambda\omega| \ge |Av_1+\lambda\omega| - |Av_2| \ge \mu\,|v_1| - |A|\cdot|v_2| \ge \mu\left(1-\frac{\varepsilon}{|\omega|}\right) - |A|\frac{\varepsilon}{|\omega|} \ge \mu - \frac{2\,|A|\,\varepsilon}{|\omega|}\,.
$$

Then, if we assume that $|\lambda\tilde{\omega}| \le 2\left|\tilde{A}\right|$, we obtain

$$
\begin{aligned}
\left|\tilde{A}v+\lambda\tilde{\omega}\right| &\ge |Av+\lambda\omega| - \left|\left(\tilde{A}-A\right)v\right| - |\lambda\,(\tilde{\omega}-\omega)| \\
&\ge \mu - \frac{2\,|A|\,\varepsilon}{|\omega|} - \varepsilon' - \frac{2\left|\tilde{A}\right|}{|\tilde{\omega}|}\varepsilon \ \ge\ \mu - \frac{4M\varepsilon}{l} - \varepsilon'.
\end{aligned}
$$

In the case $|\lambda\tilde{\omega}| > 2\left|\tilde{A}\right|$, the proof is much easier:

$$
\left|\tilde{A}v+\lambda\tilde{\omega}\right| \ge |\lambda\tilde{\omega}| - \left|\tilde{A}v\right| \ge \left|\tilde{A}\right| \ge |A| - \varepsilon' \ge \mu - \varepsilon'. \quad \square
$$

Proof of lemma 10 For a fixed $J \in F - \dfrac{4M\varepsilon}{\tilde{m}}$, we are going to prove that there exists a unique point $I^* \in G$ solving $\tilde{\Omega}(I^*) = J$. This aim is attained with the help of a modified Newton algorithm. Let us consider

$$
I^{(0)} = \Omega^{-1}(J) \in G - \frac{4\varepsilon}{\tilde{m}}
$$

as a first approximation, and we have to see that the map

$$\Lambda(I) = I - \left(\frac{\partial\tilde\Omega}{\partial I}\left(I^{(0)}\right)\right)^{-1}\left(\tilde\Omega(I) - J\right)$$

has a unique fixed point in G. We first compute the derivative of this map:

$$\frac{\partial\Lambda}{\partial I}(I) = \mathrm{Id} - \left(\frac{\partial\tilde\Omega}{\partial I}\left(I^{(0)}\right)\right)^{-1}\frac{\partial\tilde\Omega}{\partial I}(I) = \left(\frac{\partial\tilde\Omega}{\partial I}\left(I^{(0)}\right)\right)^{-1}\left(\frac{\partial\tilde\Omega}{\partial I}\left(I^{(0)}\right) - \frac{\partial\tilde\Omega}{\partial I}(I)\right),$$

and therefore

$$\left|\frac{\partial\Lambda}{\partial I}(I)\right| \le \frac{\tilde M'}{\tilde m}\left|I - I^{(0)}\right| \qquad \text{if } \left|I - I^{(0)}\right| \le \frac{4\varepsilon}{\tilde m}, \tag{108}$$

because the segment joining $I^{(0)}$ and I is fully contained in G. Starting at $I^{(0)}$, we consider the sequence defined by $I^{(k)} = \Lambda\left(I^{(k-1)}\right)$, $k \ge 1$. We check by induction that

$$\left|I^{(k)} - I^{(k-1)}\right| \le \frac{\varepsilon}{2^{k-1}\tilde m}$$

and $I^{(k)} \in G$ for every $k \ge 1$. Indeed, this is true for $k = 1$. For $k > 1$, the induction hypothesis implies that the distance from $I^{(0)}$ to $I^{(k-1)}$ or $I^{(k-2)}$ is less than $2\varepsilon/\tilde m$. The same holds for every point in the segment joining $I^{(k-1)}$ and $I^{(k-2)}$. Then, using (108) and (40) we obtain

$$\left|I^{(k)} - I^{(k-1)}\right| = \left|\Lambda\left(I^{(k-1)}\right) - \Lambda\left(I^{(k-2)}\right)\right| \le \frac{\tilde M'}{\tilde m}\cdot\frac{2\varepsilon}{\tilde m}\left|I^{(k-1)} - I^{(k-2)}\right| \le \frac{1}{2}\left|I^{(k-1)} - I^{(k-2)}\right|.$$

Thus, the sequence $I^{(k)}$ converges, for $k \to \infty$, to a fixed point of Λ which we name I^*. This point satisfies

$$\left|I^* - I^{(0)}\right| \le \frac{2\varepsilon}{\tilde m} \tag{109}$$

and therefore

$$I^* \in G - \frac{2\varepsilon}{\tilde m}. \tag{110}$$

Then,

$$\Omega(I^*) \in F - \frac{2\varepsilon m}{\tilde m} \subset F - 2\varepsilon. \tag{111}$$

The point I^* is the unique fixed point of Λ. Indeed, assuming that there is another fixed point $I^{**} \ne I^*$, we have $\tilde\Omega(I^*) = \tilde\Omega(I^{**})$ and hence $|\Omega(I^*) - \Omega(I^{**})| \le 2\varepsilon$. Then, we have $|I^* - I^{**}| \le 2\varepsilon/m$ because the whole segment joining $\Omega(I^*)$ and $\Omega(I^{**})$ is contained in F, by (111). We deduce from (109) that the distance from $I^{(0)}$ to every point in the segment joining I^* and I^{**} is less or equal than $\frac{2\varepsilon}{\tilde m} + \frac{2\varepsilon}{m} < \frac{4\varepsilon}{\tilde m}$. Applying (108), we get a contradiction:

$$|I^* - I^{**}| = |\Lambda(I^*) - \Lambda(I^{**})| < \frac{\tilde M'}{\tilde m}\cdot\frac{4\varepsilon}{\tilde m}|I^* - I^{**}| \le |I^* - I^{**}|.$$

Given a subset $\tilde F \subset F - \dfrac{4M\varepsilon}{\tilde m}$, the map $\tilde\Omega$ is one-to-one on $\tilde G = \left(\tilde\Omega\right)^{-1}\left(\tilde F\right)$. Moreover, one has $\tilde G \subset G - \dfrac{2\varepsilon}{\tilde m}$ by (110). For the proof of the other inclusion, note that $\tilde\Omega\left(G\setminus\tilde G\right)\cap$

$\tilde{F} = \emptyset$. Then, since $\left|\tilde{\Omega} - \Omega\right|_G \le \varepsilon$, we have $\Omega\left(G \setminus \tilde{G}\right) \cap \left(\tilde{F} - \varepsilon\right) = \emptyset$, and we deduce that $\Omega\left(\tilde{G}\right) \supset \tilde{F} - \varepsilon$.

Finally, we check that $\left|\left(\tilde{\Omega}\right)^{-1} - \Omega^{-1}\right|_{\tilde{F}} \le \varepsilon/m$. For a fixed $J \in \tilde{F}$, let $I = \Omega^{-1}(J)$, $\tilde{I} = \left(\tilde{\Omega}\right)^{-1}(J)$. We have

$$\left|\Omega\left(\tilde{I}\right) - \Omega(I)\right| = \left|\Omega\left(\tilde{I}\right) - \tilde{\Omega}\left(\tilde{I}\right)\right| \le \varepsilon,$$

and therefore the segment joining $\Omega(I)$ and $\Omega\left(\tilde{I}\right)$ is contained in F, since $\Omega(I) \in \tilde{F} \subset F - \varepsilon$. Hence, we obtain $\left|\tilde{I} - I\right| \le \varepsilon/m$. □

Proof of lemma 12 The estimate of part (a) comes from the inclusion

$$\begin{aligned} F(d,\beta,K) \setminus F(d',\beta',K') \quad \subset \quad & ((F-d) \setminus (F-d')) \\ & \cup \bigcup_{\substack{|k|_1 \le K \\ \overline{k} \ne 0}} \left((F-d) \cap \left(\Delta\left(k, \frac{\beta'}{|k|_1^\tau}\right) \setminus \Delta\left(k, \frac{\beta}{|k|_1^\tau}\right)\right)\right) \\ & \cup \bigcup_{\substack{K < |k|_1 \le K' \\ \overline{k} \ne 0}} \left((F-d) \cap \Delta\left(k, \frac{\beta'}{|k|_1^\tau}\right)\right) \end{aligned}$$

and the fact that, for $0 \le \alpha \le \alpha'$ and $\overline{k} \ne 0$,

$$\text{mes}\left[(F-d) \cap (\Delta(k,\alpha') \setminus \Delta(k,\alpha))\right] \le (\text{diam}\, F)^{n-1} \cdot \frac{2(\alpha' - \alpha)}{\left|\overline{k}\right|}\,. \tag{112}$$

Concerning part (b), we first remark that, for $b \ge 0$,

$$F(d,\beta,K) - b \supset (F - (d+b)) \setminus \bigcup_{\substack{|k|_1 \le K \\ \overline{k} \ne 0}} \Delta\left(k\,,\ \frac{\beta}{|k|_1^\tau} + \left|\overline{k}\right| b\right).$$

Then, proceeding like in part (a) and applying (112) again,

$$\text{mes}\left[(F(d,\beta,K) \setminus (F(d,\beta,K) - b)\right] \le Db + \sum_{\substack{|k|_1 \le K \\ \overline{k} \ne 0}} (\text{diam}\, F)^{n-1} \cdot 2b\,.$$

From the fact that the number of integer vectors $k \in \mathbf{Z}^n \setminus \{0\}$ with $|k|_1 \le K$ can be estimated by $2^n K^n$, one gets part (b). It is worth noting that this estimate expresses the "growth" of the boundary of the domain when the resonances are removed. □

References

[1] V. I. Arnold. Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian. *Russian Math. Surveys* **18** (1963), 9–36.

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

[3] G. Benettin, G. Gallavotti. Stability of motions near resonances in quasi-integrable Hamiltonian systems. *J. Stat. Phys.* **44** (1986), 293–338.

[4] H. W. Broer, G. B. Huitema. A proof of the isoenergetic KAM–theorem from the 'ordinary' one. *J. Diff. Eq.* **90** (1991), 52–60.

[5] Ch.-Q. Cheng, Y.-S. Sun. Existence of KAM tori in degenerate Hamiltonian systems. Preprint, 1994. To appear in *J. Diff. Eq.*

[6] B. V. Chirikov. A universal instability of many-dimensional oscillator systems. *Physics Reports* **52** (1979), 263–379.

[7] A. Delshams, P. Gutiérrez. Effective stability for nearly integrable Hamiltonian systems, *in* "Proceedings of the International Conference on Differential Equations, Barcelona 1991" (C. Perelló, C. Simó, J. Solà-Morales, ed.), vol. 1, p. 415–420. World Scientific, Singapore, 1993.

[8] A. Delshams, P. Gutiérrez. Nekhoroshev and KAM theorems revisited via a unified approach, *in* "Hamiltonian Mechanics: Integrability and Chaotic Behaviour" (J. Seimenis, ed.), p. 299–306. Plenum, New York, 1994.

[9] R. Douady. "Applications du théorème des tores invariants". Thèse de 3e cycle, Université Paris VII, 1982.

[10] F. Fassò. Lie series method for vector fields and Hamiltonian perturbation theory. *J. Appl. Math. Phys. (ZAMP)* **41** (1990), 843–864.

[11] A. Giorgilli, A. Delshams, E. Fontich, L. Galgani, C. Simó. Effective stability for a Hamiltonian system near an elliptic equilibrium point, with an application to the restricted three body problem. *J. Diff. Eq.* **77** (1989), 167–198.

[12] A. Giorgilli, L. Galgani. Rigorous estimates for the series expansions of Hamiltonian perturbation theory. *Celest. Mech.* **37** (1985), 95–112.

[13] A. N. Kolmogorov. The general theory of dynamical systems and classical mechanics, *in* "Foundations of mechanics" (R. H. Abraham, J. E. Marsden), 2nd edition, p. 741–757 (appendix). Benjamin/Cummings, Reading (Massachusetts), 1978.

[14] R. de la Llave, C. E. Wayne. Whiskered and low dimensional tori in nearly integrable Hamiltonian systems. Preprint, 1988.

[15] P. Lochak. Stabilité en temps exponentiels des systèmes hamiltoniens proches de systèmes integrables: résonances et orbites fermées. Preprint E.N.S., Paris, 1990.

[16] P. Lochak, C. Meunier. "Multiphase averaging for classical systems: with applications to adiabatic theorems". *Applied Mathematical Sciences*, Vol. 72. Springer-Verlag, New York, 1988.

[17] P. Lochak, A. I. Neishtadt. Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian. *Chaos* **2** (1992), 495–499.

[18] A. Morbidelli, A. Giorgilli. Quantitative perturbation theory by successive elimination of harmonics. *Celest. Mech.* **55** (1993), 131–159.

[19] A. Morbidelli, A. Giorgilli. Superexponential stability of KAM tori, 1994.

[20] J. Moser. On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen: Math. Phys. Kl.* **II** (1962), 1–20.

[21] J. Moser. Lectures on Hamiltonian systems. *Memoirs Am. Math. Soc.* **81**, 1–60. American Mathematical Society, Providence (Rhode Island), 1968.

[22] A. I. Neishtadt. Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions. *J. Appl. Math. Mech.* **45** (1982), 766–772.

[23] A. I. Neishtadt. The separation of motions in systems with rapidly rotating phase. *J. Appl. Math. Mech.* **48** (1984), 133–139.

[24] N. N. Nekhoroshev. An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. *Russian Math. Surveys* **32** (1977), 1–65.

[25] A. D. Perry, S. Wiggins. KAM tori are very sticky: rigorous lower bounds on the time to move away from an invariant Lagrangian torus with linear flow. *Physica D* **71** (1994), 102–121.

[26] J. Pöschel. Integrability of Hamiltonian systems on Cantor sets. *Comm. Pure Appl. Math.* **XXXV** (1982), 653–696.

[27] J. Pöschel. Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Math. Zeitschrift* **213** (1993), 187.

[28] H. Rüssmann. Nondegeneracy in the perturbation theory of integrable dynamical systems, *in* “Stochastics, algebra and analysis in classical and quantum dynamics” (S. Albeverio, Ph. Blanchard, D. Testard, ed.), p. 211–223. *Mathematics and its Applications*, Vol. 59. Kluwer Academic Publishers, Dordrecht, 1990.

[29] M. B. Sevryuk. KAM–stable Hamiltonians, 1994. To appear in *J. Dyn. Control Syst.*

[30] J. Xiu, J. You, Q. Qiu. Invariant tori for nearly integrable Hamiltonian systems with degeneracy. Preprint E.T.H. Zürich, 1994.