Available online at www.sciencedirect.com
ScienceDirect
J. Differential Equations 244 (2008) 2251–2276
Journal of Differential Equations
www.elsevier.com/locate/jde

Kolmogorov Theorem revisited

Jordi Villanueva

Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Received 19 September 2006; revised 11 January 2008

Available online 14 March 2008

Abstract

Kolmogorov Theorem on the persistence of invariant tori of real analytic Hamiltonian systems is revisited. In this paper we are mainly concerned with the lower bound on the constant of the Diophantine condition required by the theorem. From the existing proofs in the literature, this lower bound turns to be of $\mathcal{O}(\varepsilon^{1/4})$, where ε is the size of the perturbation. In this paper, by means of careful estimates on Kolmogorov's method, we show that this lower bound can be weakened to be of $\mathcal{O}(\varepsilon^{1/2})$. This condition coincides with the optimal one of KAM Theorem. Moreover, we also obtain optimal estimates for the distance between the actions of the perturbed and unperturbed tori. We believe that some ideas contained in this paper may be used for improving several estimates in the general KAM context.

MSC: 37J40

Keywords: Hamiltonian systems; Invariant tori; Kolmogorov Theorem

1. Introduction

Kolmogorov Theorem is one of the most celebrated results of the mechanics. This result was a great step forward in order to obtain a clear picture for the dynamics of a nearly-integrable Hamiltonian system. Kolmogorov Theorem ensures, under generic hypotheses of non-resonance and non-degeneracy, the persistence of a Lagrangian invariant torus of a Hamiltonian system with n degrees of freedom, carrying quasi-periodic motion, under the effect of a small (enough) perturbation. Thus, there is an invariant Lagrangian torus of the perturbed Hamiltonian system close to the unperturbed one with the same vector of basic frequencies.

E-mail address: jordi.villanueva@upc.edu.
URL: http://www.ma1.upc.edu/~jordi.

doi:10.1016/j.jde.2008.02.010

The formulation of this result admits many generalizations: maps or flows, analytic or smooth systems, generic or degenerate systems, etc. (see [6] and references therein). However, in this paper we focus on the original exposition of Kolmogorov in [10] and consider real analytic Hamiltonian systems, written in action-angle variables and non-degenerate in Kolmogorov's sense. See [3,4,8] for proofs of this result following Kolmogorov's original outline. We also quote [5] for a proof performed without using "transformation theory," and [7] for an approach valid for Hamiltonian systems not written in action-angle variables.

Among the conditions needed to prove this result, we are mainly concerned with the Diophantine (non-resonance) condition on the vector of basic frequencies $\omega \in \mathbb{R}^n$,

$$\bigl|\langle k, \omega\rangle\bigr| \geqslant \gamma |k|^{-\tau}, \quad \forall k \in \mathbb{Z}^n \setminus \{0\}. \tag{1}$$

It is well known that if we consider a fixed $\tau > n-1$, then, for almost every $\omega \in \mathbb{R}^n$, there is $\gamma > 0$ for which (1) is fulfilled (see [6]). But γ can be very small for some of those ω. From the existing proofs of Kolmogorov Theorem in the literature, we have that the non-resonance condition on the unperturbed torus to persist is a lower bound for γ of $\mathcal{O}(\varepsilon^{1/4})$, where ε is the size of the perturbation. Let us explain why this estimate is not satisfactory at all.

In [10] Kolmogorov also stated a "global version" of this result, first proved by Arnol'd in [2] and later tackled by many other authors, presently known as KAM Theorem (Kolmogorov–Arnol'd–Moser). We refer to [6] for a nice survey on KAM Theory. The starting point of KAM Theorem is a nearly-integrable Hamiltonian,

$$H_\varepsilon(\theta, I) = h(I) + \varepsilon f(\theta, I), \tag{2}$$

that for our purposes we assume given by a real analytic function. Here, h is an integrable Hamiltonian written in action-angle variables, $(\theta, I) \in \mathbb{T}^n \times V$, with $\mathbb{T}^n = (\mathbb{R}/2\pi\mathbb{Z})^n$ and $V \subset \mathbb{R}^n$ open and bounded, with the symplectic structure $d\theta \wedge dI$ (see Liouville–Arnol'd Theorem, for instance in [1]). The Hamilton equations for h are

$$\dot\theta = \Omega(I) := \nabla_I h(I), \qquad \dot I = 0. \tag{3}$$

For any $I_0 \in V$ the Lagrangian torus $\mathbb{T}^n \times \{I_0\}$ is invariant by the flow of (3), with linear quasi-periodic dynamics for the variable θ, having $\Omega(I_0)$ as vector of basic frequencies. We assume that the frequency map of this integrable system, $I \in V \mapsto \Omega(I)$, is non-degenerate in Kolmogorov's sense, i.e., $\det(D_I\Omega(I)) \neq 0$ $\forall I \in V$. Under these hypotheses, KAM Theorem ensures that, for any ε small enough, there is a (Cantor) set $\mathcal{V}_\varepsilon \subset \mathbb{T}^n \times V$ foliated by Lagrangian tori invariant by the flow of H_ε. More precisely, for any $I_0 \in V$ such that $\omega = \Omega(I_0)$ verifies (1), with γ bounded away from zero by $\mathcal{O}(\varepsilon^{1/2})$, the n-dimensional invariant torus of (3) labeled by this frequency is not destroyed by the perturbation, but only slightly deformed. The distance (in action variables) between the perturbed and unperturbed torus is of $\mathcal{O}(\varepsilon/\gamma)$. We refer to [11,12] for these quantitative estimates.

The lower bound on γ for which KAM tori persist has an important dynamical meaning, because in the proof of KAM Theorem it is translated into an upper bound for the Lebesgue measure of the set defined by the "gaps" between invariant tori ("resonance or stochastic zones"). Thus, the Lebesgue measure of the complementary set of $\mathcal{V}_\varepsilon$, $(\mathbb{T}^n \times V) \setminus \mathcal{V}_\varepsilon$, is bounded by $\mathcal{O}(\varepsilon^{1/2})$. We point out that there are dynamical reasons that imply that this estimate of $\mathcal{O}(\varepsilon^{1/2})$ for the size of these gaps cannot be improved in general (due to the breakdown of resonant tori and the hyperbolic dynamics thus generated, see for instance [15]).

As a summary of this exposition, the lower bound on γ required by Kolmogorov Theorem is greater than the one of KAM Theorem. Consequently, there are invariant tori of $h(I)$ that persist in KAM Theorem, but to which we cannot apply Kolmogorov Theorem. In principle it is not unnatural, because the two proofs follow different strategies. The usual methodology for proving both results is based on the construction of an infinite sequence of canonical transformations, that are recursively applied to the perturbed system (2). But Kolmogorov's method looks for only one invariant torus and Arnol'd's method looks for all the tori in $\mathcal{V}_\varepsilon$ simultaneously. On the one hand, in Kolmogorov's method this sequence is constructed so that the limit Hamiltonian has the most simple form for which the torus $\mathbb{T}^n \times \{0\}$ is invariant, having quasi-periodic motion with the selected vector of basic frequencies (Kolmogorov normal form, see (9)). On the other hand, Arnol'd's method computes, at any step of the iterative procedure, a canonical transformation removing as most non-integrable terms as possible from the Hamiltonian obtained after the previous step (thus obtaining at the limit an "integrable system on a Cantor set").

Although KAM Theorem looks simultaneously for more objects than Kolmogorov's one, and thus we can legitimately think that poor estimates can be expected for it, there are several reasons leading to a better condition for γ in KAM Theorem than in Kolmogorov Theorem. The most remarkable one is that any step of Arnol'd's method only involves one small divisors equation, but in Kolmogorov's method we have to solve two (see (15)). Moreover, as the sequence of transformed Hamiltonian of Arnol'd's proof converges to an integrable one, there are several "couplings," between expressions depending on θ, appearing in Kolmogorov's method but not in Arnol'd's one. As a summary, Arnol'd's scheme seems more "powerful" in order to remove the perturbation from the torus if γ is small.

But Kolmogorov's method presents several advantages making it suitable to be applied when looking for only one invariant torus: it is more simple to perform, does not require the initial system to be nearly-integrable (only a small perturbation of a Kolmogorov normal form), leads to a limit Hamiltonian analytic with respect to the actions and provides bounds with "better constants" (for explicit formulas of these constants and a computer assisted application of Kolmogorov Theorem, see [5]). Hence, for values of γ "not too small," the condition on ε of Kolmogorov Theorem is better (for numerical applications) than KAM's one. Our purpose is to show that the same happens when γ is small.

In this paper we prove that Kolmogorov's method is convergent under a condition for γ analogous to the one of KAM Theorem, that is, a lower bound of $\mathcal{O}(\varepsilon^{1/2})$. Moreover, we also obtain an estimate of $\mathcal{O}(\varepsilon/\gamma)$ for the distance (in action variables) between the perturbed and unperturbed torus. The price we pay is that when proving this theorem we have to be very careful with the estimates on the iterative scheme, which become more involved and tedious, but, in some sense, "more natural."

There are two main points we want to stress from our proof. The most important one is that when quantifying how far is a given Hamiltonian from being in Kolmogorov normal form, there are two different terms of the system to be taken into account. At the beginning of the proof, the size of both terms is of the same order, but during the iterative procedure they follow "different scales" with respect to γ. The usual way to quantify the distance from the Hamiltonian to Kolmogorov normal form is to consider the maximum between the size of both terms. In this paper, we take advantage from their natural scale to define a suitable normalized expression to measure this distance (see (32)). This approach "resembles" the one used in [9] to "optimize" (here with respect to τ) the exponent of an exponentially small estimate for the remainder of a (partial) normal form around a lower dimensional torus.

The second point refers to the particular (simple) form of the canonical transformations used in Kolmogorov's method (see Section 4). By using "closed formulas" for these transformations several estimates can be improved.

The contents of this paper are organized as follows. In Section 2 we introduce some basic notations and definitions we use throughout the paper. The precise formulation of Kolmogorov Theorem is given in Section 3 (Theorem 2). The proof of Theorem 2 goes from Section 4, where the canonical transformations we use are described, through Section 5, where a general step of the iterative method is controlled (Lemma 5), till Section 6, where Theorem 2 is properly proved. Some technical aspects of the paper are postponed until Appendix A.

2. Basic notations and definitions

We consider action-angle variables $(\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n$ and the compact notation $z = (\theta, I)$, where $\mathbb{T}^n = (\mathbb{R}/2\pi\mathbb{Z})^n$, with $n \geqslant 2$ fixed from now on. We denote by Id_θ, Id_I and Id the identity functions of θ, I and z, respectively. Analogously, ∇_θ, ∇_I and ∇ are the gradient operators with respect to θ, I and z, respectively, and $D_\theta(\cdot)$, $D_I(\cdot)$ and $D(\cdot)$ denote the differential matrices of $(\cdot)$ with respect to θ, I and z, respectively. The notation D_I^2 is used for the Hessian matrix with respect to I. If $\omega \in \mathbb{R}^n$ we define

$$L_\omega(\cdot) = \sum_{j=1}^{n} \omega_j \partial_{\theta_j}(\cdot). \tag{4}$$

Given a Hamilton function $H(\theta, I)$, we consider Hamiltonian systems of the form

$$\dot{\theta} = \nabla_I H(\theta, I), \qquad \dot{I} = -\nabla_\theta H(\theta, I),$$

or equivalently, $\dot{z} = J_n \nabla H(z)$, where

$$J_n = \begin{pmatrix} 0 & \mathrm{Id}_n \\ -\mathrm{Id}_n & 0 \end{pmatrix}$$

and Id_n is the n-dimensional identity matrix.

In the following definitions we consider real analytic functions depending on θ or (θ, I), 2π-periodic in θ and taking values in $\mathbb{C}$, $\mathbb{C}^l$ or in a space of complex matrices. In the forthcoming, the name "function" (with no extra information) is reserved for complex-valued functions.

Given $f(\theta)$, then $\bar{f} = \langle f \rangle_\theta = (2\pi)^{-n} \int_{\mathbb{T}^n} f(\theta)\, d\theta$ means the average of f. If we expand f in Fourier series,

$$f(\theta) = \sum_{k \in \mathbb{Z}^n} f_k \mathrm{e}^{\mathrm{i}\langle k, \theta \rangle}, \tag{5}$$

then $\bar{f} = f_0$. We also set $\tilde{f}(\theta) = f(\theta) - \bar{f}$.

Given a function $F(\theta, I)$, we consider the following Taylor expansion with respect to I,

$$F(\theta, I) = F(\theta, 0) + \big\langle \nabla_I F(\theta, 0), I \big\rangle + \frac{1}{2}\big\langle I, D_I^2 F(\theta, 0) I \big\rangle + [F]_3(\theta, I),$$

where $\langle u, v\rangle = u^\top \cdot v$ is the inner product of $\mathbb{C}^n$ and $\top$ means the transposition of a vector or a matrix. The notation $[F]_3$ refers to the terms of $\mathcal{O}_3(I)$.

We denote by $|\cdot|$ the supremum norm of a vector, $|x| = \sup_{j=1,\dots,n}\{|x_j|\}$ if $x \in \mathbb{C}^n$, and we extend this notation to the associated matrix norm. We use this norm to define the domains

$$\Delta(\rho) = \{\theta \in \mathbb{C}^n:\ |\mathrm{Im}(\theta)| \leqslant \rho\}, \qquad \mathbb{B}(r) = \{I \in \mathbb{C}^n:\ |I| \leqslant r\}, \qquad \mathbb{D}(\rho, r) = \Delta(\rho) \times \mathbb{B}(r).$$

Let $f(\theta)$ and $F(\theta, I)$ be bounded functions defined in $\Delta(\rho)$ and $\mathbb{D}(\rho, r)$, respectively. We introduce the norms:

$$\|f\|_\rho = \sup_{\Delta(\rho)} |f(\theta)|, \qquad \|F\|_{\rho,r} = \sup_{\mathbb{D}(\rho,r)} |F(\theta, I)|.$$

If f takes values in $\mathbb{C}^l$, $f = (f_1, \dots, f_l)$, we set $\|f\|_\rho = |(\|f_1\|_\rho, \dots, \|f_l\|_\rho)|$. If f is a matrix-valued function, we extend the notation $\|f\|_\rho$ by computing the $|\cdot|$-norm of the (constant) matrix defined by the $\|\cdot\|_\rho$-norms of the entries of f. Analogously, we define $\|F\|_{\rho,r}$ if F is vector or matrix-valued. We observe that if the (matrix) product $f_1 \cdot f_2$ is defined then $\|f_1 \cdot f_2\|_\rho \leqslant \|f_1\|_\rho \|f_2\|_\rho$, and the same holds for $\|\cdot\|_{\rho,r}$. If f and F are analytic in the interior of their domains, we have (Cauchy estimates)

$$\|\partial_{\theta_j} f\|_{\rho-\delta} \leqslant \frac{\|f\|_\rho}{\delta}, \qquad \|\partial_{\theta_j} F\|_{\rho-\delta,r} \leqslant \frac{\|F\|_{\rho,r}}{\delta}, \qquad \|\partial_{I_j} F\|_{\rho,r-\delta} \leqslant \frac{\|F\|_{\rho,r}}{\delta}, \tag{6}$$

for any $j = 1, \dots, n$. Moreover, we also have the following bounds for the average of f:

$$|\bar{f}| \leqslant \|f\|_\rho, \qquad \|\tilde{f}\|_\rho \leqslant 2\|f\|_\rho. \tag{7}$$

Finally, if we restrict again to complex-valued functions, we have, for any $\theta, \theta' \in \Delta(\rho)$ and $I, I' \in \mathbb{B}(r)$ (Mean Value Theorem),

$$\begin{aligned} |f(\theta) - f(\theta')| &\leqslant \|D_\theta f\|_\rho |\theta - \theta'|, \\ |F(\theta, I) - F(\theta', I')| &\leqslant \|D_\theta F\|_{\rho,r} |\theta - \theta'| + \|D_I F\|_{\rho,r} |I - I'|. \end{aligned} \tag{8}$$

3. Formulation of the result

Let $\mathcal{H}$ be a Hamiltonian written in action-angle variables $(\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n$. We suppose that the Lagrangian torus $\mathbb{T}^n \times \{0\}$ is invariant under the flow of $\mathcal{H}$, with θ carrying linear quasi-periodic motion, with vector of basic frequencies $\omega \in \mathbb{R}^n$. This implies that $\mathcal{H}$ takes the form

$$\mathcal{H}(\theta, I) = \lambda + \langle \omega, I\rangle + \frac{1}{2}\langle I, \mathcal{A}(\theta) I\rangle + [\mathcal{H}]_3(\theta, I), \tag{9}$$

with 2π-periodic dependence in θ, $\lambda \in \mathbb{R}$ and $\mathcal{A}^\top = \mathcal{A}$. We say that the system (9) is in Kolmogorov normal form. We require this normal form to be non-resonant and non-degenerate (in Kolmogorov's sense), i.e., we ask ω to verify the Diophantine condition (1) and $\det(\bar{\mathcal{A}}) \neq 0$.

Remark 1. By using the Diophantine assumption on ω, the dependence on θ of $\mathcal{A}(\theta)$ can be removed by means of a canonical change (see [9]). Thus, the non-degeneracy condition $\det(\bar{\mathcal{A}}) \neq 0$ on $\mathcal{H}$ is analogous to Kolmogorov's one on the integrable system (3), $\det(D_I^2 h(I)) \neq 0$, for $I = 0$.

Now we state Kolmogorov Theorem.

Theorem 2 *(Kolmogorov Theorem). Let H be a Hamiltonian written in action-angle variables, $(\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n$, $n \geqslant 2$, 2π-periodic in θ, real analytic in the interior of $\mathbb{D}(\rho_0, r_0)$ and bounded on the closure, for certain $\rho_0 > 0$ and $0 < r_0 \leqslant 1$. We suppose that H is a (small) perturbation of a non-resonant and non-degenerate Kolmogorov normal form* (9),

$$H(\theta, I) = \mathcal{H}(\theta, I) + \mathcal{F}(\theta, I). \tag{10}$$

More concretely, we assume that ω verifies the Diophantine condition (1), *for certain $0 < \gamma \leqslant 1$ and $\tau \geqslant n-1$, and the following quantitative estimates on H:*

$$\|\mathcal{A}\|_{\rho_0} \leqslant \alpha, \qquad \left|(\bar{\mathcal{A}})^{-1}\right| \leqslant \bar{\alpha}, \qquad \left\|[\mathcal{H}]_3\right\|_{\rho_0, R} \leqslant \tilde{\alpha} R^3, \qquad \|\mathcal{F}\|_{\rho_0, r_0} \leqslant \varepsilon,$$

for any $0 \leqslant R \leqslant r_0$, with $\alpha, \bar{\alpha}, \tilde{\alpha}$ greater than one and $0 < \varepsilon \leqslant 1$.

Given a fixed $0 < \delta_0 \leqslant \min\{\rho_0/16, r_0/32\}$, we suppose ε small enough so that

$$3 \times 2^{9+2\tau} n^3 \alpha^4 \bar{\alpha}^2 \tilde{\alpha} \sigma^3 \frac{\varepsilon}{\gamma^2 \delta_0^{2\tau+6}} \leqslant 1, \tag{11}$$

where $\sigma = \sigma(\tau, n, \omega) \geqslant 1$ is provided by Lemma 8. *Then, there exists a canonical transformation $\Psi^*(\theta, I) = (\Xi^*(\theta), \mathcal{J}^*(\theta, I))$, with $\mathcal{J}^*$ of the form*

$$\mathcal{J}^*(\theta, I) = \big(\mathrm{Id}_n + \mathcal{B}^*(\theta)\big) I + h^*(\theta),$$

where $\Xi^*(\theta) - \theta$, $\mathcal{B}^*(\theta)$ and $h^*(\theta)$ are real analytic in the interior of $\Delta(\rho^*)$, with $\rho^* = \rho_0 - 11\delta_0$, bounded on the closure and 2π-periodic in θ, such that the Hamiltonian $H^* = H \circ \Psi^*$ is in Kolmogorov normal form,

$$H^*(\theta, I) = \bar{a}^* + \langle \omega, I\rangle + \frac{1}{2}\big\langle I, A^*(\theta) I\big\rangle + [H^*]_3(\theta, I).$$

Hence, the torus $\mathbb{T}^n \times \{0\}$ is invariant under the flow of H^*, with linear quasi-periodic motion for the variable θ, having ω as vector of basic frequencies. Then, $\theta \in \mathbb{T}^n \mapsto \Psi^*(\theta, 0) = (\Xi^*(\theta), h^*(\theta))$ gives the parameterization of an invariant Lagrangian torus of H, with ω-quasi-periodic dynamics on θ. Moreover, we have the following estimates

$$\|\Psi^* - \mathrm{Id}\|_{\rho^*, r^*} \leqslant \delta_0, \qquad \|h^*\|_{\rho^*} \leqslant 3 \times 2^8 \alpha \bar{\alpha} \sigma \frac{\varepsilon}{\gamma \delta_0^{\tau+1}},$$

$$\|\Xi^* - \mathrm{Id}_\theta\|_{\rho^*} \leqslant 9 \times 2^7 \alpha^2 \bar{\alpha} \sigma^2 \frac{\varepsilon}{\gamma^2 \delta_0^{2\tau+1}}, \qquad \|\mathcal{B}^*\|_{\rho^*} \leqslant 9 \times 2^9 n \alpha^2 \bar{\alpha} \sigma^2 \frac{\varepsilon}{\gamma^2 \delta_0^{2(\tau+1)}},$$

$$\|A^*\|_{\rho^*} \leqslant 4\alpha, \qquad \left\|(\bar{A}^*)^{-1}\right\|_{\rho^*} \leqslant 4\bar{\alpha}, \qquad \left\|[H^*]_3\right\|_{\rho^*, R} \leqslant 4\tilde{\alpha} R^3,$$

for any $0 \leqslant R \leqslant r^* = r_0/2 - 11\delta_0$.

Remark 3. As we claimed, condition (11) is fulfilled if γ is bounded away from zero by $\mathcal{O}(\sqrt{\varepsilon})$. Moreover, the displacement in action variables of the perturbed torus with respect to the unperturbed one is given by $h^*(\theta)$, for which we have an estimate of $\mathcal{O}(\varepsilon/\gamma)$. We point out that the displacement in the angular variables, given by $\Xi^* - \mathrm{Id}_\theta$, can be bigger if γ is small.

4. Canonical transformations

Let us start by describing (formally) a general step of Kolmogorov's (iterative) method. For this purpose we consider a Hamiltonian of the form

$$H(\theta, I) = a(\theta) + \langle \omega + b(\theta), I\rangle + \frac{1}{2}\langle I, A(\theta)I\rangle + F(\theta, I), \tag{12}$$

with $A^\top = A$ and $F = [H]_3$. We also suppose that ω verifies (1) and that $\det(\bar{A}) \neq 0$. For (12) to be in Kolmogorov normal form (9) we require $\tilde{a} = 0$ and $b = 0$. If we assume $\tilde{a}$ and b small (but nonzero), Kolmogorov's method looks for a canonical transformation, Φ, such that "squares" the size of these terms. To explain it more precisely, let us assume for the moment that H is the Hamiltonian (10) on the statement of Theorem 2 ("the initial system"). In this case, we have that $\tilde{a}$ and b are of $\mathcal{O}(\varepsilon)$ and that $\det(\bar{A}) \neq 0$, if ε is small. We set $H^{(1)} = H \circ \Phi$ and expand $H^{(1)}$ as in (12) (see (24)). We want $\tilde{a}^{(1)}$ and $b^{(1)}$ to become of $\mathcal{O}(\varepsilon^2)$ after this first transformation. If we iterate this process, the size of these "error terms" after s steps turns to be ("roughly speaking") of $\mathcal{O}(\varepsilon^{2^s})$. This "super-convergence," introduced by Siegel [14], is used to overcome the small divisors of the problem (see (39)).

We use canonical transformations defined as the flow time $t = 1$ of a suitable Hamiltonian system, $\dot{z} = J_n \nabla G(z)$, where $G(\theta, I)$ is called the generating function of the transformation (Lie method). We denote by $\Phi_t^G(z)$ the flow of this system. If we set $t = 1$, we simply write $\Phi^G(z)$. We point out that if ∇G is 2π-periodic in θ, then $\Phi_t^G - \mathrm{Id}$ is also 2π-periodic in θ. Moreover, under the assumption of smallness of ∇G, then Φ_t^G is close to the identity. This smallness also implies the convergence of the following expansion,

$$H \circ \Phi_t^G = H + \frac{t}{1!}\{H, G\} + \frac{t^2}{2!}\big\{\{H, G\}, G\big\} + \frac{t^3}{3!}\big\{\big\{\{H, G\}, G\big\}, G\big\} + \cdots, \tag{13}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket, $\{f, g\} = (\nabla f)^\top J_n \nabla g$. Unfortunately, if we use (13) to bound the transformed Hamiltonian $H \circ \Phi^G$, without taking into account the particular simple form of the canonical transformations we use, it seems not possible to obtain the estimates of Theorem 2. What we have done is to compute explicitly the expressions of the different terms of $H^{(1)} = H \circ \Phi^G$ (see (24)) as function of the expansion (12) of H and of the particular form of G. For this purpose, let us start by discussing Φ_t^G more precisely.

By using the Lie method to prove Kolmogorov Theorem, it is natural to use generating functions of the form (see [4])

$$G(\theta, I) = \langle \xi, \theta\rangle + c(\theta) + \langle d(\theta), I\rangle, \tag{14}$$

where $\xi \in \mathbb{R}^n$ and c, d are 2π-periodic in θ, with the normalizations $\bar{c} = 0$ and $\bar{d} = 0$. To construct G we proceed as follows. We denote by $\hat{H}$ the Hamiltonian obtained by setting $\tilde{a} = 0$ and $b = 0$ in (12). It is clear that $\hat{H}$ is in Kolmogorov normal form and close to H. Then, we define G

by asking $H + \{\hat{H}, G\}$ to be also in Kolmogorov normal form. We point out that this condition for G can be motivated, in the aim of the classical Newton method, by considering the linear approximation of $H \circ \Phi^G$ with respect to G (see (13)). We obtain the following small divisors equations for ξ, c and d (see (4)),

$$L_\omega c(\theta) = \tilde{a}(\theta), \qquad L_\omega d(\theta) + A(\theta)\big(\xi + \nabla_\theta c(\theta)\big) = b(\theta). \tag{15}$$

These equations can be solved (formally) by expanding them in Fourier series (5) and by using the assumption $\det(\bar{A}) \neq 0$. The convergence of c and d is guaranteed by the Diophantine condition (1) on ω (see Lemma 8). For more details, see the proof of Lemma 5.

The Hamiltonian system $\dot{z} = J_n \nabla G(z)$ takes the form:

$$\dot{\theta} = \nabla_I G(\theta, I) = d(\theta), \qquad \dot{I} = -\nabla_\theta G(\theta, I) = -\xi - \nabla_\theta c(\theta) - \big(D_\theta d(\theta)\big)^\top I. \tag{16}$$

Remark 4. Two remarkable observations about (16) are that the equation for θ is uncoupled and that the equation for I is linear with respect to I. Both are keystones of our proof.

We write the solutions of (16) as $\Phi_t^G(\theta^0, I^0) = (\Theta^t(\theta^0), \mathcal{I}^t(\theta^0, I^0))$, where (θ^0, I^0) mean the initial conditions at $t = 0$, with $\Theta^t(\theta^0) - \theta^0$ and $\mathcal{I}^t(\theta^0, I^0)$ 2π-periodic in θ^0. By assuming smallness of ∇G, the components of Φ_t^G can be expanded as

$$\Theta^t\big(\theta^0\big) = \theta^0 + f^t\big(\theta^0\big), \tag{17}$$

$$\begin{aligned} \mathcal{I}^t\big(\theta^0, I^0\big) &= \big(\mathrm{Id}_n + B^t\big(\theta^0\big)\big)I^0 + g^t\big(\theta^0\big) \\ &= \big(\mathrm{Id}_n - \big(D_\theta d\big(\theta^0\big)\big)^\top t + \hat{B}^t\big(\theta^0\big)\big)I^0 - \big(\xi + \nabla_\theta c\big(\theta^0\big)\big)t + \hat{g}^t\big(\theta^0\big). \end{aligned} \tag{18}$$

We skip the superscript t of (17) and (18) if $t = 1$. By using the integral expressions of (16),

$$\theta(t) = \theta^0 + \int_0^t d\big(\theta(s)\big)\,ds, \qquad I(t) = I^0 - \int_0^t \big(\xi + \nabla_\theta c\big(\theta(s)\big) + \big(D_\theta d\big(\theta(s)\big)\big)^\top I(s)\big)\,ds,$$

we derive the following equations for f^t, B^t, $\hat{B}^t$, g^t and $\hat{g}^t$:

$$f^t\big(\theta^0\big) = \int_0^t d\big(\Theta^s\big(\theta^0\big)\big)\,ds, \tag{19}$$

$$B^t\big(\theta^0\big) = -\int_0^t \big(D_\theta d\big(\Theta^s\big(\theta^0\big)\big)\big)^\top \big(\mathrm{Id}_n + B^s\big(\theta^0\big)\big)\,ds, \tag{20}$$

$$\hat{B}^t\big(\theta^0\big) = -\int_0^t \big(D_\theta d\big(\Theta^s\big(\theta^0\big)\big) - D_\theta d\big(\theta^0\big)\big)^\top ds - \int_0^t \big(D_\theta d\big(\Theta^s\big(\theta^0\big)\big)\big)^\top B^s\big(\theta^0\big)\,ds, \tag{21}$$

$$g^t(\theta^0) = -\int_0^t \big(\xi + \nabla_\theta c\big(\Theta^s(\theta^0)\big)\big)\,ds - \int_0^t \big(D_\theta d\big(\Theta^s(\theta^0)\big)\big)^\top g^s(\theta^0)\,ds, \tag{22}$$

$$\hat{g}^t(\theta^0) = -\int_0^t \big(\nabla_\theta c\big(\Theta^s(\theta^0)\big) - \nabla_\theta c(\theta^0)\big)\,ds - \int_0^t \big(D_\theta d\big(\Theta^s(\theta^0)\big)\big)^\top g^s(\theta^0)\,ds. \tag{23}$$

Now, we show the action of Φ^G on the Hamiltonian (12). As Φ^G is canonical this means that, if we perform the change of coordinates $(\theta, I) = \Phi^G(\theta', I')$ on the Hamiltonian system defined by $H(\theta, I)$, we obtain again a Hamiltonian system, with Hamilton function $H^{(1)}(\theta', I') = H \circ \Phi^G(\theta', I')$. Let us compute (formally) the expression of $H^{(1)}$. For simplicity we skip the primes of the new variables. We obtain

$$\begin{aligned} H^{(1)}(\theta, I) &= a\big(\theta + f(\theta)\big) + \big\langle \omega + b\big(\theta + f(\theta)\big), \big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta)\big\rangle \\ &\quad + \frac{1}{2}\big\langle \big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta), A\big(\theta + f(\theta)\big)\big(\big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta)\big)\big\rangle \\ &\quad + F\big(\theta + f(\theta), \big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta)\big) \\ &= a^{(1)}(\theta) + \big\langle \omega + b^{(1)}(\theta), I\big\rangle + \frac{1}{2}\big\langle I, A^{(1)}(\theta)I\big\rangle + F^{(1)}(\theta, I), \end{aligned} \tag{24}$$

with

$$\begin{aligned} a^{(1)} &= a\big(\theta + f(\theta)\big) + \big\langle \omega + b\big(\theta + f(\theta)\big), g(\theta)\big\rangle + \frac{1}{2}\big\langle g(\theta), A\big(\theta + f(\theta)\big)g(\theta)\big\rangle \\ &\quad + F\big(\theta + f(\theta), g(\theta)\big), \\ b^{(1)} &= B(\theta)^\top \omega + \big(\mathrm{Id}_n + B(\theta)\big)^\top \big(b\big(\theta + f(\theta)\big) + A\big(\theta + f(\theta)\big)g(\theta) + \nabla_I F\big(\theta + f(\theta), g(\theta)\big)\big), \\ A^{(1)} &= \big(\mathrm{Id}_n + B(\theta)\big)^\top \big(A\big(\theta + f(\theta)\big) + D_I^2 F\big(\theta + f(\theta), g(\theta)\big)\big)\big(\mathrm{Id}_n + B(\theta)\big), \\ F^{(1)} &= \big[F\big(\theta + f(\theta), \big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta)\big)\big]_3. \end{aligned}$$

Now, we recall that ξ, c and d verify (15). Then, we have (see the definition of L_ω)

$$\begin{aligned} \langle \omega, g\rangle &= -\big\langle \omega, \xi + \nabla_\theta c(\theta)\big\rangle + \big\langle \omega, \hat{g}(\theta)\big\rangle = -\langle \omega, \xi\rangle - \tilde{a}(\theta) + \big\langle \omega, \hat{g}(\theta)\big\rangle, \\ \langle \omega, \hat{g}\rangle &= -\left\langle \omega, \int_0^1 \big(\nabla_\theta c\big(\Theta^s(\theta)\big) - \nabla_\theta c(\theta)\big)\,ds \right\rangle - \left\langle \omega, \int_0^1 \big(D_\theta d\big(\Theta^s(\theta)\big)\big)^\top g^s(\theta)\,ds \right\rangle \\ &= -\int_0^1 \big(\tilde{a}\big(\Theta^s(\theta)\big) - \tilde{a}(\theta)\big)\,ds - \int_0^1 \big(b\big(\Theta^s(\theta)\big) - A\big(\Theta^s(\theta)\big)\big(\xi + \nabla_\theta c\big(\Theta^s(\theta)\big)\big)\big)^\top g^s(\theta)\,ds, \end{aligned} \tag{25}$$

$$B^\top \omega = -D_\theta d(\theta)\omega + \hat{B}(\theta)^\top \omega = -b(\theta) + A(\theta)\big(\xi + \nabla_\theta c(\theta)\big) + \hat{B}(\theta)^\top \omega,$$

$$\hat{B}^\top\omega = -\int_0^1 \big(D_\theta d\big(\Theta^s(\theta)\big) - D_\theta d(\theta)\big)\omega\, ds - \int_0^1 B^s(\theta)^\top D_\theta d\big(\Theta^s(\theta)\big)\omega\, ds$$

$$= -\int_0^1 \big(b\big(\Theta^s(\theta)\big) - b(\theta)\big)\, ds + \int_0^1 \big(A\big(\Theta^s(\theta)\big)\big(\xi + \nabla_\theta c\big(\Theta^s(\theta)\big)\big) - A(\theta)\big(\xi + \nabla_\theta c(\theta)\big)\big)\, ds$$

$$- \int_0^1 B^s(\theta)^\top \big(b\big(\Theta^s(\theta)\big) - A\big(\Theta^s(\theta)\big)\big(\xi + \nabla_\theta c\big(\Theta^s(\theta)\big)\big)\big)\, ds. \qquad (26)$$

Using these expressions we can rewrite $H^{(1)}$ into the following form:

$$a^{(1)} = \bar{a} - \langle\omega, \xi\rangle + \big(\tilde{a}\big(\theta + f(\theta)\big) - \tilde{a}(\theta)\big) + \big\langle\omega, \hat{g}(\theta)\big\rangle + \big\langle b\big(\theta + f(\theta)\big), g(\theta)\big\rangle$$
$$+ \frac{1}{2}\big\langle g(\theta), A\big(\theta + f(\theta)\big)g(\theta)\big\rangle + F\big(\theta + f(\theta), g(\theta)\big), \qquad (27)$$

$$b^{(1)} = \hat{B}(\theta)^\top\omega + \big(b\big(\theta + f(\theta)\big) - b(\theta)\big) + B(\theta)^\top b\big(\theta + f(\theta)\big) + \big(A\big(\theta + f(\theta)\big) - A(\theta)\big)g(\theta)$$
$$+ A(\theta)\hat{g}(\theta) + B(\theta)^\top A\big(\theta + f(\theta)\big)g(\theta) + \big(\mathrm{Id}_n + B(\theta)\big)^\top \nabla_I F\big(\theta + f(\theta), g(\theta)\big), \qquad (28)$$

$$A^{(1)} = A(\theta) + \big(A\big(\theta + f(\theta)\big) - A(\theta)\big) + B(\theta)^\top A\big(\theta + f(\theta)\big)\big(\mathrm{Id}_n + B(\theta)\big)$$
$$+ A\big(\theta + f(\theta)\big)B(\theta) + \big(\mathrm{Id}_n + B(\theta)\big)^\top D_I^2 F\big(\theta + f(\theta), g(\theta)\big)\big(\mathrm{Id}_n + B(\theta)\big), \qquad (29)$$

$$F^{(1)} = F(\theta, I) + \big[F\big(\theta + f(\theta), \big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta)\big) - F(\theta, I)\big]_3. \qquad (30)$$

5. The Iterative Lemma

In this section we state and prove a result controlling a general step of Kolmogorov's iterative process (Lemma 5 can be seen as a quantitative version of Section 4).

Lemma 5 *(Iterative Lemma). Let H be a Hamiltonian written in action-angle variables, $(\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n$, $n \geqslant 2$, 2π-periodic in θ, real analytic in the interior of $\mathbb{D}(\rho, r)$ and bounded on the closure, for certain $\rho > 0$ and $0 < r \leqslant 1$. We expand H as*

$$H(\theta, I) = a(\theta) + \big\langle\omega + b(\theta), I\big\rangle + \frac{1}{2}\big\langle I, A(\theta)I\big\rangle + F(\theta, I),$$

where $F = [H]_3$, and we assume $\det(\bar{A}) \neq 0$ and the following bounds:

$$\|\tilde{a}\|_\rho \leqslant \tilde{\mu}, \qquad \|b\|_\rho \leqslant \bar{\mu}, \qquad \|A\|_\rho \leqslant m, \qquad \big|(\bar{A})^{-1}\big| \leqslant \bar{m}, \qquad \|F\|_{\rho,R} \leqslant \tilde{m}R^3, \quad (31)$$

for any $0 \leqslant R \leqslant r$, with $m, \bar{m}, \tilde{m}$ greater than one. Moreover, we suppose that $\omega \in \mathbb{R}^n$ verifies the Diophantine condition (1), *for certain $0 < \gamma \leqslant 1$ and $\tau \geqslant n - 1$.*

Given a fixed $0 < \delta \leqslant \min\{\rho, r\}/6$, we define

$$\mu = \bar{\mu} + \frac{\tilde{\mu}}{\gamma\delta^{\tau+1}}, \qquad (32)$$

and suppose

$$13n^2m^3\bar{m}^2\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+2}}\leqslant 1, \tag{33}$$

where $\sigma=\sigma(n,\tau,\omega)\geqslant 1$ *is provided by Lemma* 8. *Then, there exists a canonical transformation* $\Phi(\theta,I)=(\Theta(\theta),\mathcal{I}(\theta,I))$, *with* $\mathcal{I}$ *of the form*

$$\mathcal{I}(\theta,I)=\big(\mathrm{Id}_n+B(\theta)\big)I+g(\theta),$$

such that $\Theta(\theta)-\theta$, $B(\theta)$ *and* $g(\theta)$ *are real analytic in the interior of* $\Delta(\rho^{(1)})$, *where* $\rho^{(1)}=\rho-5\delta$, *bounded on the closure and* 2π*-periodic in* θ, *that transforms* H *into* $H^{(1)}=H\circ\Phi$, *that we expand as*

$$H^{(1)}(\theta,I)=a^{(1)}(\theta)+\big\langle\omega+b^{(1)}(\theta),I\big\rangle+\frac{1}{2}\big\langle I,A^{(1)}(\theta)I\big\rangle+F^{(1)}(\theta,I),$$

being $F^{(1)}=[H^{(1)}]_3$, *with bounds in the domain* $\mathbb{D}(\rho^{(1)},r^{(1)})$, *where* $r^{(1)}=r-5\delta$, *given by*

$$\tilde{\mu}^{(1)}=50nm^3\bar{m}^2\sigma^2\mu^2,\qquad \bar{\mu}^{(1)}=80n^2m^4\bar{m}^2\tilde{m}\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}, \tag{34}$$

$$m^{(1)}=m+39n^2m^3\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}},\qquad \bar{m}^{(1)}=\bar{m}+78n^2m^3\bar{m}^3\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}, \tag{35}$$

$$\tilde{m}^{(1)}=\tilde{m}+2n^5m^2\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+5}}, \tag{36}$$

where the expressions displayed above play the same rôle for $H^{(1)}$ *than the ones without the superscript* (1) *for* H *in* (31). *Moreover,* Φ *satisfies the bounds*

$$\|\Phi-\mathrm{Id}\|_{\rho^{(1)},r^{(1)}}\leqslant\delta,\qquad \|g\|_{\rho^{(1)}}\leqslant 4m\bar{m}\sigma\mu, \tag{37}$$

$$\|\Theta-\mathrm{Id}_\theta\|_{\rho^{(1)}}\leqslant 3m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau}},\qquad \|B\|_{\rho^{(1)}}\leqslant 6nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}. \tag{38}$$

Proof. We look for $\Phi=\Phi^G$ with $G(\theta,I)$ given by (14), where ξ, c and d are the solutions of (15) with the normalizations $\bar{c}=0$ and $\bar{d}=0$. The existence and uniqueness of these solutions is easy to establish. First, if we expand $a(\theta)$ and $c(\theta)$ in Fourier series (5) and consider the action of the differential operator L_ω (4) on c, it is immediate to derive the (formal) solution

$$c(\theta)=\sum_{k\in\mathbb{Z}^n\setminus\{0\}}\frac{a_k}{\mathrm{i}\langle k,\omega\rangle}\mathrm{e}^{\mathrm{i}\langle k,\theta\rangle}, \tag{39}$$

where $\{a_k\}_{k\in\mathbb{Z}^n}$ are the Fourier coefficients of $a(\theta)$. To ensure the convergence of c, we use the estimates on small divisors of Lemma 8, thus obtaining

$$\|c\|_{\rho-\delta}\leqslant\sigma\frac{\|\tilde{a}\|_\rho}{\gamma\delta^\tau}\leqslant\sigma\frac{\tilde{\mu}}{\gamma\delta^\tau},$$

and by applying Cauchy estimates (6),

$$\|\nabla_\theta c\|_{\rho-2\delta} \leqslant \frac{\|c\|_{\rho-\delta}}{\delta} \leqslant \sigma \frac{\tilde{\mu}}{\gamma\delta^{\tau+1}}.$$

Cauchy estimates on the derivatives are extensively used along the paper without any explicit mention. If we compute the average with respect to θ of the second equation of (15), we obtain

$$\xi = (\bar{A})^{-1}(\bar{b} - \overline{A\nabla_\theta c}), \tag{40}$$

and hence (see (7)),

$$|\xi| \leqslant \left|(\bar{A})^{-1}\right|\left(\|b\|_\rho + \|A\|_\rho \|\nabla_\theta c\|_{\rho-2\delta}\right) \leqslant \bar{m}\left(\bar{\mu} + m\sigma\frac{\tilde{\mu}}{\gamma\delta^{\tau+1}}\right) \leqslant m\bar{m}\sigma\mu.$$

Consequently,

$$|\xi| + \|\nabla_\theta c\|_{\rho-2\delta} \leqslant m\bar{m}\sigma\mu + \sigma\frac{\tilde{\mu}}{\gamma\delta^{\tau+1}} \leqslant 2m\bar{m}\sigma\mu.$$

If we take ξ defined by (40) we have $\langle b - A(\xi + \nabla_\theta c)\rangle_\theta = 0$. Then, by expanding $b - A(\xi + \nabla_\theta c)$ in Fourier series, we derive the formal expression of d in analogy to c in (39). Next to that, by using Lemma 8 again, we obtain the following estimate for d:

$$\begin{aligned} \|d\|_{\rho-3\delta} &\leqslant \frac{\sigma}{\gamma\delta^\tau}\left(\|b\|_\rho + \|A\|_\rho\left(|\xi| + \|\nabla_\theta c\|_{\rho-2\delta}\right)\right) \\ &\leqslant \frac{\sigma}{\gamma\delta^\tau}\left(\bar{\mu} + 2m^2\bar{m}\sigma\mu\right) \leqslant 3m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^\tau}. \end{aligned} \tag{41}$$

From (41) we have

$$\max\left\{\|D_\theta d\|_{\rho-4\delta}, \left\|(D_\theta d)^\top\right\|_{\rho-4\delta}\right\} \leqslant n\frac{\|d\|_{\rho-3\delta}}{\delta} \leqslant 3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}.$$

Now we control the action of the canonical transformation Φ^G. For this purpose, we consider the system of ordinary differential equations (16). First we solve the equation $\dot{\theta} = d(\theta)$. We observe that from (41) and condition (33) we have $\|d\|_{\rho-3\delta} \leqslant \delta$. Then, by using elemental arguments on the solutions of an ordinary differential equation (more precisely, the solutions are defined until they reach the boundary of the domain), we have that $\Theta^t(\theta)$ is defined for any $\theta \in \Delta(\rho - 5\delta)$ and $0 \leqslant t \leqslant 1$, and verifies (see (17) and (19))

$$\sup_{t\in[0,1]}\left\{\left\|\Theta^t - \mathrm{Id}_\theta\right\|_{\rho-5\delta}\right\} = \sup_{t\in[0,1]}\left\{\left\|f^t\right\|_{\rho-5\delta}\right\} \leqslant \sup_{t\in[0,1]}\left\{t\cdot\|d\|_{\rho-3\delta}\right\} \leqslant 3m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^\tau}. \tag{42}$$

We observe that, in particular, the expression (42) is smaller than δ. After we have solved the equation for θ, we replace $\theta \equiv \Theta^t(\theta)$ in the equation for I, which becomes a linear differential equation on I. Thus, for any $\theta \in \Delta(\rho - 5\delta)$ and $I \in \mathbb{C}^n$, the solution of this equation,

$$\mathcal{I}^t(\theta, I) = \left(\mathrm{Id}_n + B^t(\theta)\right)I + g^t(\theta),$$

is also defined for any $0 \leqslant t \leqslant 1$. By using the integral equation of g^t (see (22)) we have

$$\begin{aligned}\|g^t\|_{\rho-5\delta} &\leqslant t\big(|\xi| + \|\nabla_\theta c\|_{\rho-2\delta}\big) + \int_0^t \big\|(D_\theta d)^\top\big\|_{\rho-4\delta}\|g^s\|_{\rho-5\delta}\,ds \\ &\leqslant 2m\bar{m}\sigma\mu + \int_0^t 3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\|g^s\|_{\rho-5\delta}\,ds,\end{aligned}$$

for any $0 \leqslant t \leqslant 1$. If we apply Gronwall's inequality we obtain (see (33))

$$\sup_{t\in[0,1]}\big\{\|g^t\|_{\rho-5\delta}\big\} \leqslant 2m\bar{m}\sigma\mu\exp\left\{3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\right\} \leqslant 4m\bar{m}\sigma\mu. \tag{43}$$

We repeat the same process for B^t (see (20)). Then, for any $0 \leqslant t \leqslant 1$, we have

$$\begin{aligned}\|B^t\|_{\rho-5\delta} &\leqslant \int_0^t \big\|(D_\theta d)^\top\big\|_{\rho-4\delta}\big(1 + \|B^s\|_{\rho-5\delta}\big)\,ds \\ &\leqslant 3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}} + \int_0^t 3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\|B^s\|_{\rho-5\delta}\,ds.\end{aligned}$$

Using Gronwall again we obtain

$$\sup_{t\in[0,1]}\big\{\|B^t\|_{\rho-5\delta}\big\} \leqslant 3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\exp\left\{3nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\right\} \leqslant 6nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}.$$

Then, we have

$$\begin{aligned}&\sup_{t\in[0,1]}\big\{\|\mathcal{I}^t - \mathrm{Id}_I\|_{\rho-5\delta,r-\delta}\big\} \\ &\quad\leqslant \sup_{t\in[0,1]}\big\{\|B^t\|_{\rho-5\delta}\big\}(r-\delta) + \sup_{t\in[0,1]}\big\{\|g^t\|_{\rho-5\delta}\big\} \\ &\quad\leqslant 6nm^2\bar{m}\sigma^2(r-\delta)\frac{\mu}{\gamma\delta^{\tau+1}} + 4m\bar{m}\sigma\mu \leqslant 8nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}},\end{aligned} \tag{44}$$

which from (33) is smaller than δ. Bounds (42) and (44) on the components of Φ_t^G give the inequality $\|\Phi - \mathrm{Id}\|_{\rho^{(1)},r^{(1)}} \leqslant \delta$ on the statement, and justify that the compositions involved in the computation of $H^{(1)}$ (according to formulas of Section 4) are well defined in $\mathbb{D}(\rho-5\delta, r-\delta)$.

To control $H^{(1)}$ we also have to consider $\hat{g}$ and $\hat{B}$. For $\hat{g}$ we have (see (23) for $t=1$ and (25))

$$\begin{aligned} \|\hat{g}\|_{\rho-5\delta} &\leqslant \sup_{t\in[0,1]}\bigl\{\bigl\|\nabla_\theta c\bigl(\Theta^t(\theta)\bigr)-\nabla_\theta c(\theta)\bigr\|_{\rho-5\delta}\bigr\}+\bigl\|(D_\theta d)^\top\bigr\|_{\rho-4\delta}\sup_{t\in[0,1]}\bigl\{\bigl\|g^t\bigr\|_{\rho-5\delta}\bigr\} \\ &\leqslant 3nm^2\bar{m}\sigma^3\frac{\tilde{\mu}\mu}{\gamma^2\delta^{2\tau+2}}+12nm^3\bar{m}^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}\leqslant 15nm^3\bar{m}^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}, \end{aligned} \tag{45}$$

$$\begin{aligned} \bigl\|\langle\omega,\hat{g}\rangle\bigr\|_{\rho-5\delta} &\leqslant \sup_{t\in[0,1]}\bigl\{\bigl\|\tilde{a}\bigl(\Theta^t(\theta)\bigr)-\tilde{a}(\theta)\bigr\|_{\rho-5\delta}\bigr\} \\ &\quad + n\bigl(\|b\|_\rho+\|A\|_\rho\bigl(|\xi|+\|\nabla_\theta c\|_{\rho-2\delta}\bigr)\bigr)\sup_{t\in[0,1]}\bigl\{\bigl\|g^t\bigr\|_{\rho-5\delta}\bigr\} \\ &\leqslant 3nm^2\bar{m}\sigma^2\frac{\mu\tilde{\mu}}{\gamma\delta^{\tau+1}}+4nm\bar{m}\sigma\mu\bigl(\bar{\mu}+2m^2\bar{m}\sigma\mu\bigr)\leqslant 12nm^3\bar{m}^2\sigma^2\mu^2, \end{aligned} \tag{46}$$

where we use the Mean Value Theorem (8) and Cauchy estimates to bound

$$\sup_{t\in[0,1]}\bigl\{\bigl\|\tilde{a}\bigl(\Theta^t(\theta)\bigr)-\tilde{a}(\theta)\bigr\|_{\rho-5\delta}\bigr\}\leqslant\|D_\theta\tilde{a}\|_{\rho-\delta}\sup_{t\in[0,1]}\bigl\{\bigl\|f^t\bigr\|_{\rho-5\delta}\bigr\}\leqslant n\frac{\|\tilde{a}\|_\rho}{\delta}\sup_{t\in[0,1]}\bigl\{\bigl\|f^t\bigr\|_{\rho-5\delta}\bigr\},$$

and, bounding component by component,

$$\sup_{t\in[0,1]}\bigl\{\bigl\|\nabla_\theta c\bigl(\Theta^t(\theta)\bigr)-\nabla_\theta c(\theta)\bigr\|_{\rho-5\delta}\bigr\}\leqslant n\frac{\|c\|_{\rho-\delta}}{\delta^2}\sup_{t\in[0,1]}\bigl\{\bigl\|f^t\bigr\|_{\rho-5\delta}\bigr\}.$$

Moreover, we also use that $|(\cdot)^\top|\leqslant n|(\cdot)|$, being $(\cdot)$ an n-dimensional vector or a matrix with n rows. For further uses we observe that this implies the bound $|\langle(\cdot)_1,(\cdot)_2\rangle|\leqslant n|(\cdot)_1|\cdot|(\cdot)_2|$.

To bound $\hat{B}^\top\omega$ we consider (26) for $t=1$. Then, we add and subtract $A(\theta)(\xi+\nabla_\theta c(\Theta^s(\theta)))$ inside the second integral, thus obtaining

$$\begin{aligned} \bigl\|\hat{B}^\top\omega\bigr\|_{\rho-5\delta} &\leqslant \sup_{t\in[0,1]}\bigl\{\bigl\|b\bigl(\Theta^t(\theta)\bigr)-b(\theta)\bigr\|_{\rho-5\delta}\bigr\} \\ &\quad+\sup_{t\in[0,1]}\bigl\{\bigl\|A\bigl(\Theta^t(\theta)\bigr)-A(\theta)\bigr\|_{\rho-5\delta}\bigr\}\bigl(|\xi|+\|\nabla_\theta c\|_{\rho-2\delta}\bigr) \\ &\quad+\|A\|_\rho\sup_{t\in[0,1]}\bigl\{\bigl\|\nabla_\theta c\bigl(\Theta^t(\theta)\bigr)-\nabla_\theta c(\theta)\bigr\|_{\rho-5\delta}\bigr\} \\ &\quad+\sup_{t\in[0,1]}\bigl\{\bigl\|\bigl(B^t\bigr)^\top\bigr\|_{\rho-5\delta}\bigr\}\bigl(\|b\|_\rho+\|A\|_\rho\bigl(|\xi|+\|\nabla_\theta c\|_{\rho-2\delta}\bigr)\bigr) \\ &\leqslant 3m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^\tau}\biggl(n\frac{\bar{\mu}}{\delta}+2n^2m^2\bar{m}\sigma\frac{\mu}{\delta}+nm\sigma\frac{\tilde{\mu}}{\gamma\delta^{\tau+2}}\biggr) \\ &\quad+6n^2m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\bigl(\bar{\mu}+2m^2\bar{m}\sigma\mu\bigr) \\ &\leqslant 26n^2m^4\bar{m}^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}. \end{aligned} \tag{47}$$

Here we use (bounding again component by component)

$$\sup_{t\in[0,1]}\big\{\big\|b\big(\Theta^t(\theta)\big)-b(\theta)\big\|_{\rho-5\delta}\big\}\leqslant n\frac{\|b\|_\rho}{\delta}\sup_{t\in[0,1]}\big\{\big\|f^t\big\|_{\rho-5\delta}\big\},$$

$$\sup_{t\in[0,1]}\big\{\big\|A\big(\Theta^t(\theta)\big)-A(\theta)\big\|_{\rho-5\delta}\big\}\leqslant n^2\frac{\|A\|_\rho}{\delta}\sup_{t\in[0,1]}\big\{\big\|f^t\big\|_{\rho-5\delta}\big\}.$$

At this time we have all the ingredients needed to bound the (transformed) Hamiltonian $H^{(1)}=H\circ\Phi^G$ (see (24)) in the domain $\mathbb{D}(\rho-5\delta,r-5\delta)$. For $a^{(1)}$ we have (see (27))

$$\begin{aligned}\big\|a^{(1)}-\bar a+\langle\omega,\xi\rangle\big\|_{\rho-5\delta}&\leqslant\big\|\tilde a\big(\Theta(\theta)\big)-\tilde a(\theta)\big\|_{\rho-5\delta}+\big\|\langle\omega,\hat g(\theta)\rangle\big\|_{\rho-5\delta}+n\|b\|_\rho\|g\|_{\rho-5\delta}\\&\quad+\frac{n}{2}\|A\|_\rho\|g\|^2_{\rho-5\delta}+\big\|F\big(\theta+f(\theta),g(\theta)\big)\big\|_{\rho-5\delta}\\&\leqslant 3nm^2\bar m\sigma^2\frac{\mu\tilde\mu}{\gamma\delta^{\tau+1}}+12nm^3\bar m^2\sigma^2\mu^2+4nm\bar m\sigma\mu\bar\mu\\&\quad+8nm^3\bar m^2\sigma^2\mu^2+64m^3\bar m^3\tilde m\sigma^3\mu^3\\&\leqslant 25nm^3\bar m^2\sigma^2\mu^2,\end{aligned}$$

where we use (see hypothesis (31) on F)

$$\big\|F\big(\theta+f(\theta),g(\theta)\big)\big\|_{\rho-5\delta}\leqslant\tilde m\|g\|^3_{\rho-5\delta}\tag{48}$$

and that $64\bar m\tilde m\sigma\mu\leqslant n$ (see (33) and recall that $\delta\leqslant 1/6$). To obtain (48) we require $\|g\|_{\rho-5\delta}\leqslant r$, which is guaranteed by (33) and (43). From here, we derive the estimate (see (7))

$$\big\|\tilde a^{(1)}\big\|_{\rho-5\delta}\leqslant 2\big\|a^{(1)}-\bar a+\langle\omega,\xi\rangle\big\|_{\rho-5\delta}\leqslant 50nm^3\bar m^2\sigma^2\mu^2.$$

Remark 6. If instead of the bound (46) for $\|\langle\omega,\hat g\rangle\|_{\rho-5\delta}$, that follows from formula (25), we use the direct (but worst) estimate $n|\omega|\|\hat g\|_{\rho-5\delta}$, with the bound on $\hat g$ given by (45), then the estimate of $\mathcal{O}(\mu^2)$ for $\|\tilde a^{(1)}\|_{\rho-5\delta}$ turns to be of $\mathcal{O}(\mu^2/(\gamma\delta^{\tau+1}))$.

For $b^{(1)}$ we have (see (28))

$$\begin{aligned}\big\|b^{(1)}\big\|_{\rho-5\delta}&\leqslant\big\|\hat B^\top\omega\big\|_{\rho-5\delta}+\big\|b\big(\Theta(\theta)\big)-b(\theta)\big\|_{\rho-5\delta}+\big\|B^\top\big\|_{\rho-5\delta}\|b\|_\rho\\&\quad+\big\|A\big(\Theta(\theta)\big)-A(\theta)\big\|_{\rho-5\delta}\|g\|_{\rho-5\delta}+\|A\|_\rho\|\hat g\|_{\rho-5\delta}+\big\|B^\top\big\|_{\rho-5\delta}\|A\|_\rho\|g\|_{\rho-5\delta}\\&\quad+\big(1+\big\|B^\top\big\|_{\rho-5\delta}\big)\big\|\nabla_I F\big(\theta+f(\theta),g(\theta)\big)\big\|_{\rho-5\delta}\\&\leqslant 26n^2m^4\bar m^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}+3nm^2\bar m\sigma^2\frac{\mu\bar\mu}{\gamma\delta^{\tau+1}}+6n^2m^2\bar m\sigma^2\frac{\mu\bar\mu}{\gamma\delta^{\tau+1}}\\&\quad+12n^2m^4\bar m^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}+15nm^4\bar m^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}+24n^2m^4\bar m^2\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}}\end{aligned}$$

$$
\begin{aligned}
&+ 36m^2\bar{m}^2\tilde{m}\sigma^2 r\frac{\mu^2}{\delta}\left(1+6n^2m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\right)\\
&\leqslant 80n^2m^4\bar{m}^2\tilde{m}\sigma^3\frac{\mu^2}{\gamma\delta^{\tau+1}},
\end{aligned}
\tag{49}
$$

where we use Lemma 9 to bound

$$
\left\|\nabla_I F\big(\theta+f(\theta),g(\theta)\big)\right\|_{\rho-5\delta} \leqslant \frac{9}{4}\tilde{m}\frac{r}{\delta}\|g\|^2_{\rho-5\delta}. \tag{50}
$$

To write (50) we recall that $0<\delta\leqslant r/3$ and observe that from (33) we have $\|g\|_{\rho-5\delta}\leqslant r-\delta$. Moreover, in (49) we also use $6n^2m^2\bar{m}\sigma^2\mu/(\gamma\delta^{\tau+1})\leqslant 1$ and $1/\delta\leqslant 1/(6\delta^{\tau+1})$.

Remark 7. Another technical comment refers to the bound on $\|\hat{B}^\top\omega\|_{\rho-5\delta}$ we use to obtain (49). From formula (26) we have obtained in (47) an estimate for $\|\hat{B}^\top\omega\|_{\rho-5\delta}$ of $\mathcal{O}(\mu^2/(\gamma\delta^{\tau+1}))$. However, we observe that the direct bound $n|\omega|\|\hat{B}\|_{\rho-5\delta}$, by taking $\hat{B}$ from (21), produces an estimate of $\mathcal{O}(\mu^2/(\gamma^2\delta^{2\tau+2}))$, leading to a bound for $\|b^{(1)}\|_{\rho-5\delta}$ of the same order (worst than (49)).

Now it is the turn of $A^{(1)}$ (see (29)),

$$
\begin{aligned}
&\left\|A^{(1)}-A\right\|_{\rho-5\delta}\\
&\quad\leqslant \left\|A\big(\Theta(\theta)\big)-A(\theta)\right\|_{\rho-5\delta}+\|A\|_\rho\big(\|B\|_{\rho-5\delta}+\left\|B^\top\right\|_{\rho-5\delta}+\|B\|_{\rho-5\delta}\left\|B^\top\right\|_{\rho-5\delta}\big)\\
&\qquad+\big(1+\|B\|_{\rho-5\delta}\big)\big(1+\left\|B^\top\right\|_{\rho-5\delta}\big)\left\|D_I^2F\big(\theta+f(\theta),g(\theta)\big)\right\|_{\rho-5\delta}\\
&\quad\leqslant 3n^2m^3\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}+6nm^3\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\left(1+n+6n^2m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\right)\\
&\qquad+12nm\bar{m}\tilde{m}r^2\sigma\frac{\mu}{\delta^2}\left(1+6nm^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\right)\left(1+6n^2m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\right)\\
&\quad\leqslant 39n^2m^3\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}},
\end{aligned}
$$

where we use Lemma 9 again, combined with the inequality $\|g\|_{\rho-5\delta}\leqslant r-2\delta$, to obtain

$$
\left\|D_I^2F\big(\theta+f(\theta),g(\theta)\big)\right\|_{\rho-5\delta}\leqslant 3n\tilde{m}\frac{r^2}{\delta^2}\|g\|_{\rho-5\delta}.
$$

Moreover, we also point out that $1/\delta^2\leqslant 1/\delta^{\tau+1}$.

At this point we can control $(\bar{A}^{(1)})^{-1}$. We start by writing

$$
\bar{A}^{(1)}=\bar{A}+\big(\overline{A^{(1)}-A}\big)=\bar{A}\big(\mathrm{Id}_n+(\bar{A})^{-1}\big(\overline{A^{(1)}-A}\big)\big).
$$

By using again hypothesis (33), we have

$$
\left|(\bar{A})^{-1}\big(\overline{A^{(1)}-A}\big)\right|\leqslant\left|(\bar{A})^{-1}\right|\left\|A^{(1)}-A\right\|_{\rho-5\delta}\leqslant 39n^2m^3\bar{m}^2\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}\leqslant\frac{1}{2}.
$$

Hence, we obtain (using Neumann's series to compute the inverse)

$$\begin{aligned}\left|\left(\bar{A}^{(1)}\right)^{-1}\right| &\leqslant \frac{|(\bar{A})^{-1}|}{1-|(\bar{A})^{-1}|\|A^{(1)}-A\|_{\rho-5\delta}} = \left|(\bar{A})^{-1}\right| + \frac{|(\bar{A})^{-1}|^2\|A^{(1)}-A\|_{\rho-5\delta}}{1-|(\bar{A})^{-1}|\|A^{(1)}-A\|_{\rho-5\delta}} \\ &\leqslant \left|(\bar{A})^{-1}\right| + 2\left|(\bar{A})^{-1}\right|^2\left\|A^{(1)}-A\right\|_{\rho-5\delta} \leqslant \bar{m} + 78n^2m^3\bar{m}^3\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+1}}. \qquad (51)\end{aligned}$$

To finish the proof of Lemma 5 it only remains to consider $F^{(1)}$ (see (30)). We define

$$\mathcal{F}(\theta, I) = F\big(\theta + f(\theta), \big(\mathrm{Id}_n + B(\theta)\big)I + g(\theta)\big) - F(\theta, I),$$

and we have $F^{(1)} = F + [\mathcal{F}]_3$. From part (b) of Lemma 9 we can express $[\mathcal{F}]_3$ as

$$[\mathcal{F}]_3 = \int_0^1 \frac{1}{2}(1-s)^2 D_I^3\mathcal{F}(\theta, sI)[I, I, I]\, ds. \qquad (52)$$

To control $[\mathcal{F}]_3$ first we use the Mean Value Theorem (8) to bound $\mathcal{F}$. We have

$$\begin{aligned}\|\mathcal{F}\|_{\rho-5\delta,r-2\delta} &\leqslant \|D_\theta F\|_{\rho-\delta,r}\|f\|_{\rho-5\delta} + \|D_I F\|_{\rho,r-\delta}\left\|B(\theta)I + g(\theta)\right\|_{\rho-5\delta,r-2\delta} \\ &\leqslant n\frac{\|F\|_{\rho,r}}{\delta}\big(\|f\|_{\rho-5\delta} + \|B\|_{\rho-5\delta}(r-2\delta) + \|g\|_{\rho-5\delta}\big) \\ &\leqslant n\tilde{m}\frac{r^3}{\delta}\left(3m^2\bar{m}\sigma^2\frac{\mu}{\gamma\delta^\tau} + 6nm^2\bar{m}\sigma^2(r-2\delta)\frac{\mu}{\gamma\delta^{\tau+1}} + 4m\bar{m}\sigma\mu\right) \\ &\leqslant 12n^2m^2\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+2}}.\end{aligned}$$

To obtain this estimate for $\mathcal{F}$ we recall that $\|f\|_{\rho-5\delta} \leqslant \delta$ and $\|\mathcal{I} - \mathrm{Id}_I\|_{\rho-5\delta,r-2\delta} \leqslant \delta$. Thus, we derive the following bound for the third order derivatives of $\mathcal{F}$,

$$\left\|\partial^3_{I_j I_k I_l}\mathcal{F}\right\|_{\rho-5\delta,r-5\delta} \leqslant 12n^2m^2\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+5}}, \qquad (53)$$

and hence, for any $0 \leqslant R \leqslant r - 5\delta$, we obtain

$$\begin{aligned}\left\|F^{(1)} - F\right\|_{\rho-5\delta,r-5\delta} &\leqslant \int_0^1 6n^5m^2\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+5}}R^3(1-s)^2\,ds \\ &\leqslant 2n^5m^2\bar{m}\tilde{m}\sigma^2\frac{\mu}{\gamma\delta^{\tau+5}}R^3, \qquad (54)\end{aligned}$$

where we observe that bounding the multi-linear operator $D_I^3\mathcal{F}$ contributes with a factor n^3. □

6. Proof of Theorem 2

This section is devoted to prove Kolmogorov Theorem itself. This result follows from an inductive application of Lemma 5.

For this purpose, the first step is to arrange the initial Hamiltonian $H = \mathcal{H} + \mathcal{F}$ of (10) so that Lemma 5 can be applied to it. Thus, we set $H^{(0)} = H$ and expand $H^{(0)}$ as in (12),

$$H^{(0)}(\theta, I) = a^{(0)}(\theta) + \langle \omega + b^{(0)}(\theta), I \rangle + \frac{1}{2}\langle I, A^{(0)}(\theta) I \rangle + F^{(0)}(\theta, I),$$

with

$$a^{(0)} = H(\theta, 0), \qquad b^{(0)} = \nabla_I \mathcal{F}(\theta, 0), \qquad A^{(0)} = D_I^2 H(\theta, 0), \qquad F^{(0)} = [H]_3.$$

Moreover, we observe that $\tilde{a}^{(0)} = \mathcal{F}(\theta, 0) - \langle \mathcal{F}(\theta, 0) \rangle_\theta$. If we define $\rho^{(0)} = \rho_0$ and $r^{(0)} = r_0/2$, we obtain the following bounds (we use (6) and (7)),

$$\begin{aligned}
\|\tilde{a}^{(0)}\|_{\rho^{(0)}} &\leqslant 2\|\mathcal{F}\|_{\rho_0, r_0} \leqslant 2\varepsilon, \\
\|b^{(0)}\|_{\rho^{(0)}} &\leqslant \frac{\|\mathcal{F}\|_{\rho_0, r_0}}{r_0} \leqslant \frac{\varepsilon}{r_0}, \\
\|A^{(0)} - \mathcal{A}\|_{\rho^{(0)}} &\leqslant \frac{n}{(r_0/2)^2} \|\mathcal{F}\|_{\rho_0, r_0} \leqslant 4n \frac{\varepsilon}{r_0^2}, \\
\|A^{(0)}\|_{\rho^{(0)}} &\leqslant \|\mathcal{A}\|_{\rho_0} + \|A^{(0)} - \mathcal{A}\|_{\rho^{(0)}} \leqslant \alpha + 4n \frac{\varepsilon}{r_0^2} \leqslant 2\alpha,
\end{aligned} \tag{55}$$

where we use that $4n\varepsilon/r_0^2 \leqslant 1 \leqslant \alpha$ (see (11)). We also observe that (55) implies the following bound on the average of $A^{(0)}$ (compare (51)),

$$\left|\left(\bar{A}^{(0)}\right)^{-1}\right| \leqslant \frac{|(\bar{\mathcal{A}})^{-1}|}{1 - |(\bar{\mathcal{A}})^{-1}| \|A^{(0)} - \mathcal{A}\|_{\rho^{(0)}}} \leqslant 2\bar{\alpha},$$

where we use now that $4n\bar{\alpha}\varepsilon/r_0^2 \leqslant 1/2$. Moreover, if we use part (b) of Lemma 9 we obtain, for any $0 \leqslant R \leqslant r^{(0)}$ (compare (52), (53) and (54)),

$$\|[\mathcal{F}]_3\|_{\rho^{(0)}, R} \leqslant \int_0^1 \frac{1}{2} n^3 \frac{\|\mathcal{F}\|_{\rho_0, r_0}}{(r_0/6)^3} R^3 (1-s)^2 \, ds = 36 n^3 \frac{\varepsilon}{r_0^3} R^3.$$

Hence, using that $36n^3\varepsilon/r_0^3 \leqslant 1 \leqslant \tilde{\alpha}$, we have

$$\|F^{(0)}\|_{\rho^{(0)}, R} \leqslant \|[\mathcal{H}]_3\|_{\rho^{(0)}, R} + \|[\mathcal{F}]_3\|_{\rho^{(0)}, R} + \leqslant \left(\tilde{\alpha} + 36 n^3 \frac{\varepsilon}{r_0^3}\right) R^3 \leqslant 2\tilde{\alpha} R^3.$$

From these bounds on $H^{(0)}$ we introduce the following quantities,

$$\tilde{\mu}^{(0)} = 2\varepsilon, \qquad \bar{\mu}^{(0)} = \frac{\varepsilon}{r_0}, \qquad m^{(0)} = 2\alpha,$$
$$\bar{m}^{(0)} = 2\bar{\alpha}, \qquad \tilde{m}^{(0)} = 2\tilde{\alpha}, \qquad \delta^{(0)} = \delta_0. \tag{56}$$

Moreover, we also define (see (32))

$$\mu^{(0)} = \bar{\mu}^{(0)} + \frac{\tilde{\mu}^{(0)}}{\gamma(\delta^{(0)})^{\tau+1}} = \frac{\varepsilon}{r_0} + \frac{2\varepsilon}{\gamma\delta_0^{\tau+1}} \leqslant 3\frac{\varepsilon}{\gamma\delta_0^{\tau+1}}.$$

Now we proceed by induction. We suppose that we have applied s times Lemma 5, for certain $s \geqslant 0$, and we verify that we can apply it again. This means that we have computed, recursively, $H^{(j)} = H^{(j-1)} \circ \Phi^{(j-1)}$, for $j = 1, \ldots, s$, starting with $H^{(0)}$. Here, $\Phi^{(j-1)}$ is the canonical transformation associated to the j-application of the lemma. When applying Lemma 5 to $H^{(0)}$, we take the quantities ρ, r, δ, $\bar{\mu}$, $\tilde{\mu}$, μ, m, $\bar{m}$ and $\tilde{m}$ on the statement of the lemma as the values with the superscript (0) introduced in (56). When applying Lemma 5 to $H^{(j)}$, for $j \geqslant 1$, these quantities are replaced by different ones, now labeled with the superscript (j), where

$$\delta^{(j)} = \frac{\delta_0}{2^j}, \qquad \rho^{(j)} = \rho^{(j-1)} - 5\delta^{(j-1)}, \qquad r^{(j)} = r^{(j-1)} - 5\delta^{(j-1)}, \quad j \geqslant 1,$$

and $\bar{\mu}^{(j)}$, $\tilde{\mu}^{(j)}$, $\mu^{(j)}$, $m^{(j)}$, $\bar{m}^{(j)}$, $\tilde{m}^{(j)}$ are defined recursively from the ones of the step $j-1$, by using the bounds on the transformed Hamiltonian given by Lemma 5 (see (34)–(36)). Moreover, to simplify the discussion of these recursive bounds, we assume

$$m^{(j)} \leqslant 2m^{(0)}, \qquad \bar{m}^{(j)} \leqslant 2\bar{m}^{(0)}, \qquad \tilde{m}^{(j)} \leqslant 2\tilde{m}^{(0)}, \quad j = 1, \ldots, s. \tag{57}$$

To start the induction process, we remark that the condition $\delta^{(j)} \leqslant \min\{\rho^{(j)}, r^{(j)}\}/6$ of Lemma 5 is automatically fulfilled from the hypothesis $\delta_0 \leqslant \min\{\rho_0/16, r_0/32\}$. In order to check this, we observe that $\lim_{j\to+\infty} \rho^{(j)} = \rho_0 - 10\delta_0 = \rho^* + \delta_0$ and $\lim_{j\to+\infty} r^{(j)} = r_0/2 - 10\delta_0 = r^* + \delta_0$. Furthermore, condition (33) of Lemma 5 written for $H^{(j)}$ reads as

$$13n^2\big(m^{(j)}\big)^3\big(\bar{m}^{(j)}\big)^2\tilde{m}^{(j)}\sigma^2\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\tau+2}} \leqslant 1. \tag{58}$$

By assuming (56) and (57), we have that (58) holds provided that

$$13 \times 2^{12}n^2\alpha^3\bar{\alpha}^2\tilde{\alpha}\sigma^2 2^{j(\tau+2)}\frac{\mu^{(j)}}{\gamma\delta_0^{\tau+2}} \leqslant 1. \tag{59}$$

If (57) and (59) are fulfilled, for certain $j \geqslant 0$, then Lemma 5 can be applied to $H^{(j)}$ and we obtain the following bound for $\mu^{(j+1)}$ (see (32) and (34)),

$$\mu^{(j+1)} = \bar{\mu}^{(j+1)} + \frac{\tilde{\mu}^{(j+1)}}{\gamma(\delta^{(j+1)})^{\tau+1}}$$
$$= 80n^2\big(m^{(j)}\big)^4\big(\bar{m}^{(j)}\big)^2\tilde{m}^{(j)}\sigma^3\frac{(\mu^{(j)})^2}{\gamma(\delta^{(j)})^{\tau+1}} + 50n\big(m^{(j)}\big)^3\big(\bar{m}^{(j)}\big)^2\sigma^2\frac{(\mu^{(j)})^2}{\gamma(\delta^{(j+1)})^{\tau+1}}$$

$$\leqslant \big(5 \times 2^{18} n^2 \alpha^4 \bar{\alpha}^2 \tilde{\alpha} \sigma^3 + 25 \times 2^{\tau+12} n \alpha^3 \bar{\alpha}^2 \sigma^2\big) 2^{j(\tau+1)} \frac{(\mu^{(j)})^2}{\gamma \delta_0^{\tau+1}}$$

$$\leqslant \tilde{\chi} 2^{j(\tau+1)} \frac{(\mu^{(j)})^2}{\gamma \delta_0^{\tau+1}},$$

where we define

$$\tilde{\chi} = 345 \times 2^{\tau+12} n \alpha^4 \bar{\alpha}^2 \tilde{\alpha} \sigma^3, \tag{60}$$

and observe that $n \leqslant \tau + 1 \leqslant 2^\tau$. From the inductive hypotheses we are assuming, this recurrence can be iterated backwards, thus obtaining for any $j = 0, \dots, s$,

$$\mu^{(j)} \leqslant \left(2^{\tau+1} \frac{\tilde{\chi}}{\gamma \delta_0^{\tau+1}}\right)^{2^j - 1} 2^{-j(\tau+1)} \big(\mu^{(0)}\big)^{2^j} = \frac{\gamma \delta_0^{\tau+1}}{\tilde{\chi}} 2^{-(j+1)(\tau+1)} \left(3 \times 2^{\tau+1} \tilde{\chi} \frac{\varepsilon}{\gamma^2 \delta_0^{2(\tau+1)}}\right)^{2^j}.$$

This expression motivates to introduce

$$\chi = 3 \times 2^{\tau+1} \tilde{\chi} \frac{\varepsilon}{\gamma^2 \delta_0^{2(\tau+1)}}, \tag{61}$$

which from condition (11) is bounded by $\chi \leqslant 1/2$ (we recall that $\delta_0 \leqslant 1/32$). Hence, if we prove that we can iterate for any $j \geqslant 0$, then $\lim_{j \to +\infty} \mu^{(j)} = 0$.

At this point, condition (59) can be re-written in the following form,

$$13 \times 2^{10-\tau} n^2 \alpha^3 \bar{\alpha}^2 \tilde{\alpha} \sigma^2 \frac{2^{j+1}}{\tilde{\chi} \delta_0} \chi^{2^j} \leqslant 1. \tag{62}$$

By using the definition of $\tilde{\chi}$ (see (60)) and the inequalities $j + 1 \leqslant 2^j$ and $n \leqslant 2^\tau$, then (62) holds if, for instance, $(2\chi/\delta_0)^{2^j} \leqslant 1$. Thus, we observe that (11) implies that $2\chi/\delta_0 \leqslant 1$.

To finish the induction process we verify the assumptions (57) for $j = s + 1$. By using the recursive definition of $m^{(j)}$ (see (35)), we obtain

$$m^{(s+1)} = m^{(0)} + \sum_{j=0}^{s} 39 n^2 \big(m^{(j)}\big)^3 \bar{m}^{(j)} \tilde{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma (\delta^{(j)})^{\tau+1}}$$

$$\leqslant m^{(0)} + m^{(0)} \sum_{j=0}^{s} 39 \times 2^{8-\tau} n^2 \alpha^2 \bar{\alpha} \tilde{\alpha} \sigma^2 \frac{\chi^{2^j}}{\tilde{\chi}}$$

$$\leqslant m^{(0)} + m^{(0)} \sum_{j=0}^{s} \chi^{j+1} \leqslant 2m^{(0)},$$

where we use again that $\chi \leqslant 1/2$. The bound $\bar{m}^{(s+1)} \leqslant 2m^{(0)}$ follows from similar computations (see (35) again). Finally, the bound $\tilde{m}^{(s+1)} \leqslant 2\tilde{m}^{(0)}$ is just slightly different (see (36)),

$$
\begin{aligned}
\tilde{m}^{(s+1)} &= \tilde{m}^{(0)} + \sum_{j=0}^{s} 2n^5 \big(m^{(j)}\big)^2 \bar{m}^{(j)} \tilde{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma (\delta^{(j)})^{\tau+5}} \\
&\leqslant \tilde{m}^{(0)} + \tilde{m}^{(0)} \sum_{j=0}^{s} 2^{3-\tau} n^5 \alpha^2 \bar{\alpha} \sigma^2 2^{4(j+1)} \frac{\chi^{2^j}}{\tilde{\chi} \delta_0^4} \\
&\leqslant \left(1 + 2^{8-\tau} n^5 \alpha^2 \bar{\alpha} \sigma^2 \frac{\chi}{\tilde{\chi} \delta_0^4}\right) \tilde{m}^{(0)} \\
&\leqslant \left(1 + 3 \times 2^{9+2\tau} n^3 \alpha^2 \bar{\alpha} \sigma^2 \frac{\varepsilon}{\gamma^2 \delta_0^{2\tau+6}}\right) \tilde{m}^{(0)} \leqslant 2\tilde{m}^{(0)}.
\end{aligned}
$$

Here, we have bounded $2^{4(j+1)} \chi^{2^j}$ by $(2^4 \chi)^{j+1}$ and the ratio of this geometric progression by $1/2$ (see (11)). Then we compute $\chi/\tilde{\chi}$ from (61) and use again condition (11) to control the expression inside the brackets (recall that $n \leqslant 2^\tau$).

To finish the proof of Theorem 2, it only remains to study the convergence of the composition of the infinite sequence of canonical transformations $\{\Phi^{(s)}\}_{s \geqslant 0}$. We define

$$
\Psi^{(s)} = \Phi^{(0)} \circ \cdots \circ \Phi^{(s)}, \quad s \geqslant 0,
$$

and we want to prove that there exists $\Psi^* = \lim_{s \to +\infty} \Psi^{(s)}$. We recall that

$$
\Phi^{(s)}(\theta, I) = \big(\Theta^{(s)}(\theta), \mathcal{I}^{(s)}(\theta, I)\big), \qquad \mathcal{I}^{(s)}(\theta, I) = \big(\mathrm{Id}_n + B^{(s)}(\theta)\big) I + g^{(s)}(\theta).
$$

This implies that we can write $\Psi^{(s)}$ and Ψ^* as

$$
\begin{aligned}
\Psi^{(s)}(\theta, I) &= \big(\Xi^{(s)}(\theta), \mathcal{J}^{(s)}(\theta, I)\big), & \mathcal{J}^{(s)}(\theta, I) &= \big(\mathrm{Id}_n + \mathcal{B}^{(s)}(\theta)\big) I + h^{(s)}(\theta), \\
\Psi^*(\theta, I) &= \big(\Xi^*(\theta), \mathcal{J}^*(\theta, I)\big), & \mathcal{J}^*(\theta, I) &= \big(\mathrm{Id}_n + \mathcal{B}^*(\theta)\big) I + h^*(\theta).
\end{aligned}
$$

Thus, what we have to prove is the convergence of $\Xi^{(s)}$, $\mathcal{B}^{(s)}$ and $h^{(s)}$ to Ξ^*, $\mathcal{B}^*$ and h^*, respectively. First, we obtain the formula

$$
\Xi^{(s)}(\theta) = \Theta^{(0)} \circ \Theta^{(1)} \circ \cdots \circ \Theta^{(s)}(\theta),
$$

and the following recurrences for $\mathcal{B}^{(s)}$ and $h^{(s)}$:

$$
\mathcal{B}^{(s)}(\theta) = \mathcal{B}^{(s-1)}\big(\Theta^{(s)}(\theta)\big) + \big(\mathrm{Id}_n + \mathcal{B}^{(s-1)}\big(\Theta^{(s)}(\theta)\big)\big) B^{(s)}(\theta), \tag{63}
$$

$$
h^{(s)}(\theta) = h^{(s-1)}\big(\Theta^{(s)}(\theta)\big) + \big(\mathrm{Id}_n + \mathcal{B}^{(s-1)}\big(\Theta^{(s)}(\theta)\big)\big) g^{(s)}(\theta), \tag{64}
$$

starting up with $\mathcal{B}^{(0)} = B^{(0)}$ and $h^{(0)} = g^{(0)}$.

To prove the convergence of $\Xi^{(s)}$ in $\Delta(\rho^*)$ we write it as

$$
\Xi^{(s)}(\theta) - \theta = \sum_{j=1}^{s} \big(\Xi^{(j)}(\theta) - \Xi^{(j-1)}(\theta)\big) + \big(\Theta^{(0)}(\theta) - \theta\big), \tag{65}
$$

and discuss the absolute convergence of the sum $\sum_{j=1}^{+\infty} \| \Xi^{(j)} - \Xi^{(j-1)} \|_{\rho^*}$. For this purpose, we consider the following expression for $\Xi^{(j-1)}$, if $j \geqslant 2$,

$$\Xi^{(j-1)}(\theta) - \theta = \sum_{l=0}^{j-2} \big(\Theta^{(l)} \circ \cdots \circ \Theta^{(j-1)}(\theta) - \Theta^{(l+1)} \circ \cdots \circ \Theta^{(j-1)}(\theta) \big) + \big(\Theta^{(j-1)}(\theta) - \theta \big).$$

From the iterative application of Lemma 5, we have that $\Theta^{(l+1)} \circ \cdots \circ \Theta^{(j-1)}(\theta) \in \Delta(\rho^{(l+1)})$ if $\theta \in \Delta(\rho^{(j)})$, for any $l = 0, \ldots, j-2$. Thus, we obtain (see (38))

$$\begin{aligned} \big\| \Xi^{(j-1)} - \mathrm{Id}_\theta \big\|_{\rho^{(j)}} &\leqslant \sum_{l=0}^{j-1} \big\| \Theta^{(l)} - \mathrm{Id}_\theta \big\|_{\rho^{(l+1)}} \leqslant \sum_{l=0}^{j-1} 3 \big(m^{(l)} \big)^2 \bar{m}^{(l)} \sigma^2 \frac{\mu^{(l)}}{\gamma (\delta^{(l)})^\tau} \\ &\leqslant \sum_{l=0}^{j-1} 3 \times 2^{6-\tau} \alpha^2 \bar{\alpha} \sigma^2 \delta_0 2^{-(l+1)} \frac{\chi^{2^l}}{\tilde{\chi}} \leqslant 9 \times 2^7 \alpha^2 \bar{\alpha} \sigma^2 \frac{\varepsilon}{\gamma^2 \delta_0^{2\tau+1}}, \end{aligned} \tag{66}$$

where we control $2^{-(l+1)} \chi^{2^l}$ by the geometric progression $(\chi/2)^{l+1}$, of ratio smaller than 1/2. We observe that if we assume *a priori* convergence of $\Xi^{(j)}$ to Ξ^* in $\Delta(\rho^*)$, then (66) also holds for $\| \Xi^* - \mathrm{Id}_\theta \|_{\rho^*}$. In particular, as (66) is smaller than δ_0/n (see (11)), then $\| \Xi^* - \mathrm{Id}_\theta \|_{\rho^*} \leqslant \delta_0$. Moreover, if we take $\theta, \theta' \in \Delta(\rho^{(j)} - \delta_0)$, for $j \geqslant 1$, and use (6) and (8), we obtain

$$\begin{aligned} \big| \Xi^{(j-1)}(\theta') - \Xi^{(j-1)}(\theta) \big| &\leqslant |\theta' - \theta| + \big| \big(\Xi^{(j-1)}(\theta') - \theta' \big) - \big(\Xi^{(j-1)}(\theta) - \theta \big) \big| \\ &\leqslant \left(1 + n \frac{\| \Xi^{(j-1)} - \mathrm{Id}_\theta \|_{\rho^{(j)}}}{\delta_0} \right) |\theta' - \theta| \leqslant 2 |\theta' - \theta|. \end{aligned} \tag{67}$$

Now, we pick up $\theta \in \Delta(\rho^{(j+1)} - \delta_0)$ and set $\theta' = \Theta^{(j)}(\theta)$ in (67). From the bound (37) on $\Phi^{(j)}$ we have $\| \Theta^{(j)} - \mathrm{Id}_\theta \|_{\rho^{(j+1)}} \leqslant \delta^{(j)}$ and hence $\theta, \theta' \in \Delta(\rho^{(j)} - \delta_0)$. Then, by using now (38) to bound $\| \Theta^{(j)} - \mathrm{Id}_\theta \|_{\rho^{(j+1)}}$ and by recalling that $\rho^{(j)} - \delta_0 \geqslant \rho^*$, we obtain

$$\big\| \Xi^{(j)} - \Xi^{(j-1)} \big\|_{\rho^*} \leqslant 6 \big(m^{(j)} \big)^2 \bar{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma (\delta^{(j)})^\tau} \leqslant 3 \times 2^{7-\tau} \alpha^2 \bar{\alpha} \sigma^2 \delta_0 2^{-(j+1)} \frac{\chi^{2^j}}{\tilde{\chi}} \leqslant \chi^{j+1},$$

giving the absolute convergence of (65), as wanted.

To study the convergence of $\mathcal{B}^{(s)}$ we repeat the same strategy. First, let us suppose *a priori* that $\| \mathcal{B}^{(j)} \|_{\rho^{(j+1)}} \leqslant 1$, for any $j = 0, \ldots, s-1$. By using the recurrent expression (63), we obtain the following inductive bound (see (38)),

$$\begin{aligned} \big\| \mathcal{B}^{(s)} \big\|_{\rho^{(s+1)}} &\leqslant \sum_{j=0}^{s} 2 \big\| B^{(j)} \big\|_{\rho^{(j+1)}} \leqslant \sum_{j=0}^{s} 12 n \big(m^{(j)} \big)^2 \bar{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma (\delta^{(j)})^{\tau+1}} \\ &\leqslant \sum_{j=0}^{s} 3 \times 2^{7-\tau} n \alpha^2 \bar{\alpha} \sigma^2 \frac{\chi^{2^j}}{\tilde{\chi}} \leqslant 9 \times 2^9 n \alpha^2 \bar{\alpha} \sigma^2 \frac{\varepsilon}{\gamma^2 \delta_0^{2(\tau+1)}}, \end{aligned} \tag{68}$$

where we bound again χ^{2^j} by χ^{j+1}. In particular (68) proves, by induction, the *a priori* assumption $\|\mathcal{B}^{(j)}\|_{\rho^{(j+1)}} \leqslant 1$. Moreover, if we also assume *a priori* convergence of $\mathcal{B}^{(s)}$ to $\mathcal{B}^*$ in $\Delta(\rho^*)$, then (68) also bounds $\|\mathcal{B}^*\|_{\rho^*}$.

At this point we rewrite (63) as

$$\mathcal{B}^{(j)}(\theta) - \mathcal{B}^{(j-1)}(\theta) = \left(\mathcal{B}^{(j-1)}\left(\Theta^{(j)}(\theta)\right) - \mathcal{B}^{(j-1)}(\theta)\right) + \left(\mathrm{Id}_n + \mathcal{B}^{(j-1)}\left(\Theta^{(j)}(\theta)\right)\right) B^{(j)}(\theta).$$

By recalling again that $\Theta^{(j)}(\Delta(\rho^{(j+1)} - \delta_0)) \subset \Delta(\rho^{(j)} - \delta_0)$, we obtain

$$\begin{aligned}\left\|\mathcal{B}^{(j)} - \mathcal{B}^{(j-1)}\right\|_{\rho^{(j+1)}-\delta_0} &\leqslant n \frac{\|\mathcal{B}^{(j-1)}\|_{\rho^{(j)}}}{\delta_0} \left\|\Theta^{(j)} - \mathrm{Id}_\theta\right\|_{\rho^{(j+1)}} + 2\left\|B^{(j)}\right\|_{\rho^{(j+1)}} \\ &\leqslant 3n\left(m^{(j)}\right)^2 \bar{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma \delta_0 (\delta^{(j)})^\tau} + 12n\left(m^{(j)}\right)^2 \bar{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma (\delta^{(j)})^{\tau+1}} \\ &\leqslant 15 \times 2^{5-\tau} n \alpha^2 \bar{\alpha} \sigma^2 \frac{\chi^{2^j}}{\tilde{\chi}} \leqslant \chi^{j+1}. \end{aligned} \tag{69}$$

This bound implies the convergence of $\sum_{j=1}^{+\infty} \|\mathcal{B}^{(j)} - \mathcal{B}^{(j-1)}\|_{\rho^*}$ and hence, the convergence of $\mathcal{B}^{(s)}$ to $\mathcal{B}^*$ in $\Delta(\rho^*)$ (compare (65)).

Now it is the turn of $h^{(s)}$. From the recurrence (64) and the bound (37) on g we have

$$\begin{aligned}\left\|h^{(s)}\right\|_{\rho^{(s+1)}} &\leqslant \sum_{j=0}^{s} 2\left\|g^{(j)}\right\|_{\rho^{(j+1)}} \leqslant \sum_{j=0}^{s} 8 m^{(j)} \bar{m}^{(j)} \sigma \mu^{(j)} \\ &\leqslant \sum_{j=0}^{s} 2^7 \alpha \bar{\alpha} \sigma \gamma \delta_0^{\tau+1} 2^{-(j+1)(\tau+1)} \frac{\chi^{2^j}}{\tilde{\chi}} \leqslant 3 \times 2^8 \alpha \bar{\alpha} \sigma \frac{\varepsilon}{\gamma \delta_0^{\tau+1}}, \end{aligned} \tag{70}$$

where we control the expression $2^{-(j+1)(\tau+1)}\chi^{2^j}$ by the geometric progression $(\chi/2^{\tau+1})^{j+1}$, of ratio smaller than $1/2$. We observe that, in particular, $\|h^{(s)}\|_{\rho^{(s+1)}} \leqslant 1$. If $h^{(s)}$ converges to h^* in $\Delta(\rho^*)$, then (70) also gives a bound for $\|h^*\|_{\rho^*}$.

By using (64) again we have

$$h^{(j)}(\theta) - h^{(j-1)}(\theta) = \left(h^{(j-1)}\left(\Theta^{(j)}(\theta)\right) - h^{(j-1)}(\theta)\right) + \left(\mathrm{Id}_n + \mathcal{B}^{(j-1)}\left(\Theta^{(j)}(\theta)\right)\right) g^{(j)}(\theta).$$

By applying to this formula the same arguments used in (69), we obtain

$$\begin{aligned}\left\|h^{(j)} - h^{(j-1)}\right\|_{\rho^{(j+1)}-\delta_0} &\leqslant n \frac{\|h^{(j-1)}\|_{\rho^{(j)}}}{\delta_0} \left\|\Theta^{(j)} - \mathrm{Id}_\theta\right\|_{\rho^{(j+1)}} + 2\left\|g^{(j)}\right\|_{\rho^{(j+1)}} \\ &\leqslant 3n\left(m^{(j)}\right)^2 \bar{m}^{(j)} \sigma^2 \frac{\mu^{(j)}}{\gamma \delta_0 (\delta^{(j)})^\tau} + 8 m^{(j)} \bar{m}^{(j)} \sigma^2 \mu^{(j)} \\ &\leqslant 2^6 \alpha^2 \bar{\alpha} \sigma^2 \frac{\chi^{2^j}}{\tilde{\chi}} \leqslant \chi^{j+1}. \end{aligned}$$

Thus, the convergence of $h^{(s)}$ follows from the convergence of the sum $\sum_{j=1}^{+\infty} \|h^{(j)} - h^{(j-1)}\|_{\rho^*}$. To finish the proof it only remains to prove that $\|\mathcal{J}^* - \mathrm{Id}_I\|_{\rho^*, r^*} \leqslant \delta_0$, which follows from

$$\begin{aligned}\|\mathcal{J}^* - \mathrm{Id}_I\|_{\rho^*,r^*} &\leqslant \|\mathcal{B}^*\|_{\rho^*} r^* + \|h^*\|_{\rho^*} \\ &\leqslant 9 \times 2^9 n\alpha^2\bar{\alpha}\sigma^2 \frac{\varepsilon}{\gamma^2\delta_0^{2(\tau+1)}}\left(\frac{r_0}{2} - 11\delta_0\right) + 3 \times 2^8 \alpha\bar{\alpha}\sigma \frac{\varepsilon}{\gamma\delta_0^{\tau+1}} \\ &\leqslant 3 \times 2^{10} n\alpha^2\bar{\alpha}\sigma^2 \frac{\varepsilon}{\gamma^2\delta_0^{2(\tau+1)}}.\end{aligned}$$

Acknowledgment

This work has been partially supported by the Spanish MCyT/FEDER grant MTM2006-00478.

Appendix A. Some technical results

In this section we present some basic results used throughout the paper.

Lemma 8 *(Rüssmann estimates). (See [13].) Let $g(\theta)$, with $\theta \in \mathbb{T}^n$, $n \geqslant 2$, be a 2π-periodic function, analytic in the interior of $\Delta(\rho)$ and bounded on the closure, and $\omega \in \mathbb{R}^n$ a Diophantine vector verifying* (1), *for certain $\gamma > 0$ and $\tau \geqslant n-1$. If $\bar{g} = 0$, there is a unique 2π-periodic function $f(\theta)$, with the normalization $\bar{f} = 0$, solving the equation $L_\omega f = g$ (see* (4)*), that is analytic in the interior of $\Delta(\rho)$ and satisfies the bound*

$$\|f\|_{\rho-\delta} \leqslant \sigma(n,\tau,\omega)\frac{\|g\|_\rho}{\gamma\delta^\tau},$$

for any $0 < \delta \leqslant \rho$, where $\sigma \geqslant 1$ can be taken as

$$\sigma(n,\tau,\omega) = \frac{3\pi}{2^{\tau+1}} 6^{n/2}\sqrt{\tau\Gamma(2\tau)}\left(\frac{|\omega_1| + \cdots + |\omega_n|}{|\omega|}\right)^\tau,$$

with $\Gamma(\cdot)$ the Gamma function.

Lemma 9. *Let $F(I)$ be an analytic function in the interior of $\mathbb{B}(r)$ and bounded on the closure.*

(a) *If $|F(I)| \leqslant A|I|^3$, $\forall I \in \mathbb{B}(r)$, and $0 < \delta \leqslant r/3$, then if $j, k \in \{1,\ldots,n\}$,*

$$\left|\partial_{I_j} F(I)\right| \leqslant \frac{9}{4} A\frac{r}{\delta}|I|^2, \qquad \left|\partial^2_{I_j I_k} F(I)\right| \leqslant 3A\frac{r^2}{\delta^2}|I|,$$

for any $I \in \mathbb{B}(r-\delta)$ and any $I \in \mathbb{B}(r-2\delta)$, respectively.

(b) *For any $I \in \mathbb{B}(r)$ we have*

$$\left[F(I)\right]_3 = \int_0^1 \frac{1}{2}(1-s)^2 D_I^3 F(sI)[I,I,I]\,ds,$$

where $D_I^3 F(x)[I,I,I] = \sum_{j,k,l \in \{1,\ldots,n\}} \partial^3_{I_j I_k I_l} F(x) I_j I_k I_l$.

Proof. We start by proving (a). If $|I| \leqslant 2r/3$ we have, by using Cauchy estimates (6),

$$\left|\partial_{I_j} F(I)\right| \leqslant A \frac{(|I| + |I|/2)^3}{|I|/2} = \frac{27}{4} A |I|^2.$$

If $2r/3 \leqslant |I| \leqslant r - \delta$ we have

$$\left|\partial_{I_j} F(I)\right| \leqslant A \frac{r^3}{r - |I|} = A \frac{r^3}{(r - |I|)|I|^2} |I|^2 \leqslant A \frac{r^3}{\delta (r-\delta)^2} |I|^2 \leqslant \frac{9}{4} A \frac{r}{\delta} |I|^2,$$

where we use that $\min_{x \in [2r/3, r-\delta]} \{(r-x)x^2\} = \delta (r-\delta)^2$ and $0 < \delta \leqslant r/3$. Hence, the desired estimate on $\partial_{I_j} F$ is straightforward from both inequalities. The bound on the second order derivatives is obtained from the two next inequalities. If $|I| \leqslant r/3$ we have

$$\left|\partial^2_{I_j I_k} F(I)\right| \leqslant A \frac{(|I| + 2|I|)^3}{|I|^2} = 27 A |I|,$$

and if $r/3 \leqslant |I| \leqslant r - 2\delta$ then

$$\left|\partial^2_{I_j I_k} F(I)\right| \leqslant A \frac{r^3}{((r - |I|)/2)^2} = 4A \frac{r^3}{(r-|I|)^2 |I|} |I| \leqslant A \frac{r^3}{\delta^2 (r - 2\delta)} |I| \leqslant 3A \frac{r^2}{\delta^2} |I|.$$

Part (b) of the statement is obtained from the following integral expression:

$$\begin{aligned} \left[F(I)\right]_3 &= F(I) - F(0) - \left\langle \nabla_I F(0), I \right\rangle - \frac{1}{2} \left\langle I, D_I^2 F(0) I \right\rangle \\ &= \int_0^1 \frac{d}{ds} \left(F(sI) + (1-s) \left\langle \nabla_I F(sI), I \right\rangle + \frac{1}{2} (1-s)^2 \left\langle I, D_I^2 F(sI) I \right\rangle \right) ds. \qquad \square \end{aligned}$$

References

[1] R. Abraham, J. Marsden, Foundations of Mechanics, Adv. Book Program, Benjamin/Cummings, Reading, MA, 1978.

[2] V.I. Arnol'd, Proof of a theorem of A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations, Russian Math. Surveys 18 (5) (1963) 9–36.

[3] R.B. Barrar, Convergence of the von Zeipel procedure, Celestial Mech. 2 (4) (1970) 494–504.

[4] G. Benettin, L. Galgani, A. Giorgilli, J.-M. Strelcyn, A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the Lie method, Nuovo Cimento B (11) 79 (2) (1984) 201–223.

[5] A. Celletti, L. Chierchia, On the stability of realistic three-body problems, Comm. Math. Phys. 186 (2) (1997) 413–449.

[6] R. de la Llave, A tutorial on KAM theory, in: Smooth Ergodic Theory and Its Applications, Seattle, WA, 1999, in: Proc. Sympos. Pure Math., vol. 69, Amer. Math. Soc., 2001, pp. 175–292.

[7] R. de la Llave, A. González, À. Jorba, J. Villanueva, KAM theory without action-angle variables, Nonlinearity 18 (2) (2005) 855–895.

[8] J. Hubbard, Y. Ilyashenko, A proof of Kolmogorov's theorem, Discrete Contin. Dyn. Syst. 10 (1–2) (2004) 367–385, partial differential equations and applications.

[9] À. Jorba, J. Villanueva, On the normal behaviour of partially elliptic lower-dimensional tori of Hamiltonian systems, Nonlinearity 10 (4) (1997) 783–822.

[10] A.N. Kolmogorov, On conservation of conditionally periodic motions for a small change in Hamilton's function, Dokl. Akad. Nauk SSSR (N.S.) 98 (1954) 527–530.
[11] A.I. Neĭshtadt, Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions, J. Appl. Math. Mech. 45 (6) (1981) 1016–1025.
[12] J. Pöschel, Integrability of Hamiltonian systems on Cantor sets, Comm. Pure Appl. Math. 35 (5) (1982) 653–696.
[13] H. Rüssmann, On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus, in: Dynamical Systems, Theory and Applications, Rencontres, Battelle Res. Inst., Seattle, WA, 1974, in: Lecture Notes in Phys., vol. 38, Springer, Berlin, 1975, pp. 598–624.
[14] C.L. Siegel, Iteration of analytic functions, Ann. of Math. (2) 43 (1942) 607–612.
[15] D.V. Treshchëv, A mechanism for the destruction of resonance tori in Hamiltonian systems, Mat. Sb. 180 (10) (1989) 1325–1346, 1439.