ISSN 1560-3547, Regular and Chaotic Dynamics, 2008, Vol. 13, No. 2, pp. 130–139. © Pleiades Publishing, Ltd., 2008.

REVIEW ARTICLES

Kolmogorov's 1954 Paper on Nearly-Integrable Hamiltonian Systems

L. Chierchia[1*]

[1]*Dipartimento di Matematica, Università "Roma Tre"*
Largo S. L. Murialdo 1, I-00146 Roma, Italy

Received October 17, 2007; accepted November 2, 2007

Abstract—Following closely Kolmogorov's original paper [1], we give a complete proof of his celebrated Theorem on perturbations of integrable Hamiltonian systems by including few "straightforward" estimates.

MSC2000 numbers: 37J40, 70H08, 37J05, 37J25

DOI: 10.1134/S1560354708020056

Key words: Kolmogorov's theorem, KAM theory, small divisors, Hamiltonian systems, perturbation theory, symplectic transformations, nearly-integrable systems

1. INTRODUCTION

In 1954 A. N. Kolmogorov ([1, 2]), in occasion of the International Congress of Mathematicians in Amsterdam, presented a fundamental theorem concerning the existence, under suitable assumptions, of quasi–periodic trajectories in the dynamics generated by nearly–integrable Hamiltonian systems. Kolmogorov's theorem is the *fulcrum* of the well–known KAM (Kolmogorov–Arnol'd–Moser) theory; for reviews, see, e.g., [3, §6.3].

KAM proofs are typically based on two clearly separated steps:

(i) a "Newton like" scheme[1)];

(ii) quantitative estimates and proof of convergence of the scheme.

In [1], Kolmogorov, besides stating precisely the result, gives complete details for step **(i)** and, as for step **(ii)**, simply says[2)]: "*Only the use of condition (3) for proving the convergence of the recursions, $K_\theta^{(k)}$, to the analytical limit for the recursion K_θ is somewhat more subtle*".

In this paper, we review and comment Kolmogorov's original paper [1]. In particular, we recall Kolmogorov's statement and give a complete and elementary proof of it, by recalling step **(i)** as given in [1] and giving complete details for step **(ii)**.

We point out that step **(ii)** – which consists in introducing a "scale of Banach spaces[3)]", giving recursive estimates and deducing from such estimates the convergence of the scheme – is based on very classical tools (such as Cauchy estimates for analytic functions or the classical Implicit Function Theorem) obviously well known to Kolmogorov.

*E-mail: luigi@mat.uniroma3.it

1) Loosely speaking, a "Newton like" scheme (which takes its name from the elementary Newton's tangent algorithm for finding zeros of real functions) consists in finding solutions of a (functional) equation $F(u) = 0$ by successive approximate solutions u_n so that the error at the n^{th} step, $\epsilon_n := \|F(u_n)\|$, is (roughly) quadratic in the preceding error: $\epsilon_n \sim \epsilon_{n-1}^2$.

2) The quotation is from p. 55 of [1]: Eq. (3) corresponds to the Diophantine assumption (compare (2) below); θ is the perturbative parameter that, in this paper, will be denoted ε; finally K refers to symplectic transformations (while here the letter K will denote a Hamiltonian in "Kolmogorov's normal form").

3) Compare **(c)** of Remark 1 below.

It is our belief that the deepest part of the proof is, by far, step **(i)** so that, in a sense, Kolmogorov's paper might be regarded as "essentially complete".

However, Kolmogorov's approach has been, somehow, overlooked. In particular, neither V.I. Arnol'd nor J.K. Moser followed up Kolmogorov's outline: Arnol'd, in 1963 [4], published a lengthy, detailed proof of Kolmogorov's theorem using a rather different approach, which might be called a "renormalization approach", leading, in the limit, to a "pointwise integrable" Hamiltonian (or more precisely, to a Hamiltonian integrable on a Cantor set); Moser's first complete proof ([5], 1962), which is also rather different, deals with the finitely differentiable case for twist maps.

We mention that Moser wrote a review [6] of Kolmogorov's article [2] saying also: *The proof of this theorem was published in Dokl. Akad. Nauk SSSR 98 (1954), 527–530 [MR0068687 (16,924c)], but the convergence discussion does not seem convincing to the reviewer.*

Indeed, Moser's opinion is certainly correct, since Kolmogorov did not discuss the convergence of the algorithm (apart from the sentence quoted above) and our belief should not appear in contrast to Moser's statement: we simply argue that step **(ii)** – although it involves several (elementary) estimates – is rather straightforward and based on well known tools[4)] while the *real* breakthrough in the solution of the problem is based on step **(i)**.

We hope that this paper will help to appreciate, once more, the beauty of Kolmogorov's proof.

2. KOLMOGOROV'S THEOREM

Kolmogorov considers a one–parameter family of Hamiltonian systems governed by a real–analytic Hamiltonian $(y,x,\varepsilon)\in\mathcal{M}^{2d}\times(-\varepsilon_0,\varepsilon_0):=B\times\mathbb{T}^d\times(-\varepsilon_0,\varepsilon_0)\to H(y,x;\varepsilon)$, where B is a d-ball around the origin in $\mathbb{R}^d$, $\mathbb{T}^d:=\mathbb{R}^d/(2\pi\mathbb{Z}^d)$, $\varepsilon_0>0$ and ε is a real "small" parameter; the phase space $\mathcal{M}^{2d}$ is endowed with the standard symplectic form $dy\wedge dx=\sum_{j=1}^d dy_j\wedge dx_j$ so that for each ε the Hamiltonian flow $\phi_H^t:\mathcal{M}^{2d}\to\mathcal{M}^{2d}$ generated by H is the solution at time t of the Cauchy problem

$$\begin{cases}\dot y=-H_x(y,x;\varepsilon)\\ \dot x=H_y(y,x;\varepsilon)\end{cases},\qquad \begin{cases}y(0)=y\\ x(0)=x\end{cases}$$

where dot is derivative with respect to time $t\in\mathbb{R}$ and H_x and H_y denote the gradients with respect to x and y (for generalities see [3]).

Theorem (Komogorov [1], p. 52). *Let H be as above and assume that $K(y,x):=H(y,x;0)$ has the form*[5)]

$$K:=E+\omega\cdot y+Q(y,x)\quad\textit{and}\quad Q=O(|y|^2);\tag{1}$$

$E\in\mathbb{R}$ and $\omega\in\mathbb{R}^d$ is a (homogeneously) Diophantine vector, i.e., there exist positive constants κ and τ such that[6)]

$$|\omega\cdot n|\geqslant\frac{\kappa}{|n|^\tau}\,,\qquad\forall\, n\in\mathbb{Z}^d\backslash\{0\}.\tag{2}$$

Furthermore, K in (1) *is assumed to be non–degenerate in the sense that*

$$\det\langle\partial_y^2Q(0,\cdot)\rangle\neq0,$$

4)Compare **(d)** and **(c)** in Remark 1 below.

5)As usual, $\omega\cdot y=\sum_{j=1}^d\omega_jy_j$; $Q=O(|y|^2)$ means that $\partial_y^\alpha Q(0,x)=0$ for all $\alpha\in\mathbb{N}^d$ with $|\alpha|\leqslant1$, where $\partial_y^\alpha=\frac{\partial^{|\alpha|}}{\partial y_1^{\alpha_1}\cdots\partial y_d^{\alpha_d}}$ and $|\alpha|=\alpha_1+\cdots+\alpha_d$.

6)Normally, for integer vectors n, $|n|$ denotes $|n_1|+\cdots+|n_d|$, but other norms may as well be used.

where $\langle\cdot\rangle$ *denotes average over* $\mathbb{T}^d$. *Then, there exist* $0 < \varepsilon_* \leqslant \varepsilon_0$, *a ball* $B_* \subset B$ *centered at the origin of* $\mathbb{R}^d$ *and a real–analytic symplectic transformation* $\phi_* : B_* \times \mathbb{T}^d \to \mathcal{M}^{2d}$, *depending analytically also on* $\varepsilon \in (-\varepsilon_*, \varepsilon_*)$, *such that* $\phi_*|_{\varepsilon=0}$ *is the identity map and, for any* $|\varepsilon| < \varepsilon_*$,

$$H \circ \phi_*(y', x') = K_*(y', x'; \varepsilon) = E_*(\varepsilon) + \omega \cdot y' + Q_*(y', x'; \varepsilon), \quad \textit{with} \quad Q_* = O(|y'|^2).$$

Remark 1. (a) A Hamiltonian K of the form (1) is said to be in Kolmogorov's normal form; the Lagrangian torus $\mathcal{T}_\omega := \{0\} \times \mathbb{T}^d$ is invariant for K, as $\phi_K^t(0, x) = (0, x + \omega t)$; *viceversa* a Hamiltonian admitting a d-dimensional invariant torus on which the flow is conjugated to the translation $x \to x + \omega t$ (with ω_j rationally independent) can be put in Kolmogorov's normal form. From Kolmogorov's Theorem it follows that the torus $\mathcal{T}_{\omega,\varepsilon} := \phi_*(0, \mathbb{T}^d)$ is a Lagrangian torus invariant for ϕ_H^t and the H-flow on $\mathcal{T}_{\omega,\varepsilon}$ is analytically conjugated (by ϕ_*) to the translation $x' \to x' + \omega t$ with the *same frequency* vector of $\mathcal{T}_\omega = \mathcal{T}_{\omega,0}$ (while the energy of $\mathcal{T}_{\omega,\varepsilon}$, namely E_*, is in general different from the energy E of $\mathcal{T}_\omega$). The idea of keeping fixed the frequency is one of the fundamental ideas introduced by Kolmogorov.

(b) The map ϕ_* is obtained as $\phi_* = \lim_{k\to\infty} \phi_1 \circ \cdots \circ \phi_k$, where the ϕ_k's are (ε-dependent) symplectic transformations of $\mathcal{M}^{2d}$ closer and closer to the identity. It will be enough to describe the construction of ϕ_1; ϕ_2 is then obtained by replacing $H_0 := H$ with $H_1 = H \circ \phi_1$ and so on. The construction of ϕ_1 is carried out in step **(i)** following closely Kolmogorov.

(c) The analytical tools on which step **(ii)** is based are classical (and well established at the beginning of the XXth century). Let us recall them here. For $\xi > 0$, let

$$D_\xi^d := \{y \in \mathbb{C}^d : |y| < \xi\} \quad \text{and} \quad \mathbb{T}_\xi^d := \{x \in \mathbb{C}^d : \ |\operatorname{Im} x_j| < \xi,\ 1 \leqslant j \leqslant d\}/(2\pi\mathbb{Z}^d);$$

let, then, $\mathcal{B}_\xi = \mathcal{B}_{\xi,\varepsilon_0}$ be the Banach spaces of real–analytic functions f on

$$W_{\xi,\varepsilon_0} := D_\xi^d \times \mathbb{T}_\xi^d \times \{\varepsilon \in \mathbb{C} : |\varepsilon| < \varepsilon_0\}$$

with finite sup–norm[7)]

$$\|f\|_\xi = \|f\|_{\xi,\varepsilon_0} := \sup_{W_{\xi,\varepsilon_0}} |f|. \tag{3}$$

The standard sup-norm of real-analytic functions on $\mathbb{T}^d$ depending analytically on $\varepsilon \in D^1_{\varepsilon_0} = \{\varepsilon \in \mathbb{C}\colon |\varepsilon| < \varepsilon_0\}$ will be denoted by $\|\cdot\|_0 = \|\cdot\|_{0,\varepsilon_0} = \sup_{\mathbb{T}^d \times D^1_{\varepsilon_0}} |\cdot|$. Then

1. $\mathcal{B}_{\xi'} \subset \mathcal{B}_\xi$ whenever $\xi < \xi'$ and $\|f\|_\xi \leqslant \|f\|_{\xi'}$ for any $f \in \mathcal{B}_{\xi'}$;

2. if $f \in \mathcal{B}_\xi$ and $f_n(y; \varepsilon)$ denotes the n-th Fourier coefficient of the periodic function $x \to f(y, x; \varepsilon)$, then, by Cauchy integral formula of complex variables,

$$|f_n(y; \varepsilon)| \leqslant \|f\|_\xi e^{-|n|\xi}, \quad \forall\, n \in \mathbb{Z}^d,\ \forall\, y \in D_\xi^d, \quad \forall \varepsilon \in \mathbb{C}\colon |\varepsilon| < \varepsilon_0;$$

3. let $f \in \mathcal{B}_\xi$ and let $p \in \mathbb{N}$ then there exists a constant $B_p = B_p(d) \geqslant 1$ such that, for any multi–index $(\alpha, \beta) \in \mathbb{N}^d \times \mathbb{N}^d$ with $|\alpha| + |\beta| \leqslant p$ and for any $0 \leqslant \xi' < \xi$, one has[8)]

$$\|\partial_y^\alpha \partial_x^\beta f\|_{\xi'} \leqslant B_p \|f\|_\xi\ (\xi - \xi')^{-(|\alpha|+|\beta|)}; \tag{4}$$

[7)] Sometimes, the explicit dependence on ε_0 will not be denoted in the norms or in the $\mathcal{B}$–spaces, as ε_0 (or ε_* below) will not be changed during the iteration.

[8)] These are the so called Cauchy estimates, which follow immediately by Cauchy integral formula.

4. assume that[9] $x \to f(x) \in \mathcal{B}_\xi$ has zero average; assume that ω satisfies (2) and let $p \in \mathbb{N}$; denote by

$$D_\omega := \omega \cdot \partial_x := \sum_{j=1}^{d} \omega_j \frac{\partial}{\partial x_j} \tag{5}$$

the derivative on $\mathbb{T}^d$ in the direction ω and let

$$u = \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} \frac{f_n}{i\omega \cdot n}\, e^{in\cdot x} =: D_\omega^{-1} f$$

denote the unique solution with zero average of the equation

$$D_\omega u = f\ , \qquad \langle u \rangle = 0.$$

Then, $D_\omega^{-1} f \in \mathcal{B}_{\xi'}$ for any $0 \leqslant \xi' < \xi$ and there exist constants $\bar{B}_p = \bar{B}_p(d,\tau) \geqslant 1$ and $k_p = k_p(d,\tau) \geqslant 1$ such that, for any multi–index $\beta \in \mathbb{N}^d$ with $|\beta| \leqslant p$, one has[10]

$$\|\partial_x^\beta D_\omega^{-1} f\|_{\xi'} \leqslant \bar{B}_p \frac{\|f\|_\xi}{\kappa}\ (\xi - \xi')^{-k_p}. \tag{6}$$

(d) Cauchy estimates and the "small divisor estimate" (6) are based on the idea of giving bounds of analytic functions, on which there act operators diagonal in Fourier space, on smaller complex domains. This idea goes back to Cauchy and it seems unlikely that it was not obvious to Kolmogorov.

Proof of the Theorem. By hypothesis (eventually reducing ε_0), it follows that there exist $0 < \xi \leqslant 1$, such that $H \in \mathcal{B}_{\xi,\varepsilon_0}$. Write $H(y,x;\varepsilon)$ as $H = K + \varepsilon P$, then $K, P \in \mathcal{B}_{\xi,\varepsilon_0}$.

Step (i) Kolmogorov's idea is to construct a near–to–the–identity symplectic transformation ϕ_1, such that

$$H_1 := H \circ \phi_1 = K_1 + \varepsilon^2 P_1, \quad K_1 = E_1 + \omega \cdot y' + Q_1(y', x'), \quad Q_1 = O(|y'|^2); \tag{7}$$

if this is achieved, the Hamiltonian K_1 has the same basic properties of K (the linear part in y is the same and, being ϕ_1 close to the identity, K_1 is non–degenerate), and the procedure can be iterated.

To carry out this strategy, Kolmogorov indicates the form of the generating function of ϕ_1, namely[11]

$$g(y', x) := y' \cdot x + \varepsilon\big(b \cdot x + s(x) + y' \cdot a(x)\big), \tag{8}$$

where s and a are (respectively, scalar and vector–valued, ε–dependent) real–analytic functions on $\mathbb{T}^d$ with zero average and $b \in \mathbb{R}^d$.

---

[9] Clearly, all definitions may be easily adapted to functions depending only on x.

[10] This estimate follows by expanding in Fourier series. In fact, if $\delta := \xi - \xi'$, by (2), $\|\partial_x^\beta D_\omega^{-1} f\|_{\xi'} \leqslant \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} \frac{|n|^{|\beta|}\,|f_n|}{|\omega\cdot n|} e^{\xi'|n|} \leqslant \|f\|_\xi \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} \frac{|n|^{|\beta|+\tau}}{\kappa}\ e^{-\delta|n|} = \frac{\|f\|_\xi}{\kappa}\,\delta^{-(|\beta|+\tau+d)} \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} [\delta|n|]^{|\beta|+\tau}\ e^{-\delta|n|}\ \delta^d \leqslant \bar{B}_p\ \frac{\|f\|_\xi}{\kappa}\,(\xi-\xi')^{-(|\beta|+\tau+d)}$, where last estimate comes from approximating the sum with the Riemann integral $\int_{\mathbb{R}^d} |y|^{|\beta|+\tau}\, e^{-|y|}\, dy$.

[11] Compare [3] for generalities on symplectic transformations and their generating functions. For simplicity, we do not report in the notation the dependence of various functions on ε, but, in fact, $P = P(y,x;\varepsilon)$, $s = s(x;\varepsilon)$, $a = a(x;\varepsilon)$, etc.

Remark 2. **(a)** If we denote[12)]

$$\beta_0 = \beta_0(x) := b + s_x \ , \qquad A = A(x) := a_x \quad \text{and} \quad \beta = \beta(y', x) := \beta_0 + A\, y',$$

then ϕ_1 is implicitly defined by

$$\begin{cases} y = y' + \varepsilon\beta(y', x) := y' + \varepsilon(\beta_0(x) + A(x)y') \\ \\ x' = x + \varepsilon a(x). \end{cases}$$

(b) For ε small, $x \in \mathbb{T}^d \to x + \varepsilon a(x) \in \mathbb{T}^d$ defines a diffeomorphism of $\mathbb{T}^d$ with inverse

$$x = \varphi(x') := x' + \varepsilon\alpha(x'; \varepsilon),$$

for a suitable real–analytic function α. Thus ϕ_1 is explicitly given by

$$\phi_1 : (y', x') \to \begin{cases} y = y' + \varepsilon\beta\big(y', \varphi(x')\big) \\ \\ x = \varphi(x'). \end{cases} \tag{9}$$

To determine b, s and a, observe that by Taylor's formula

$$H(y' + \varepsilon\beta, x) = E + \omega \cdot y' + Q(y', x) + \varepsilon\Big[\omega \cdot \beta + Q_y(y', x) \cdot \beta + P(y', x)\Big] + \varepsilon^2 P'(y', x) \tag{10}$$

where $P' := P'(y', x; \varepsilon) := P^{(1)} + P^{(2)}$ with

$$\begin{cases} P^{(1)} := \frac{1}{\varepsilon^2}[Q(y' + \varepsilon\beta, x) - Q(y', x) - \varepsilon Q_y(y', x) \cdot \beta] = \displaystyle\int_0^1 (1-t) Q_{yy}(y' + t\varepsilon\beta, x)\, \beta \cdot \beta\, dt \\ P^{(2)} := \frac{1}{\varepsilon}[P(y' + \varepsilon\beta, x) - P(y', x)] = \displaystyle\int_0^1 P_y(y' + t\varepsilon\beta, x) \cdot \beta\, dt. \end{cases} \tag{11}$$

Note that

$$Q_y(y', x) \cdot (a_x y') =: Q^{(1)}(y', x) = O(|y'|^2), \tag{12}$$

and that (again by Taylor's formula)

$$\begin{cases} Q_y(y', x) \cdot \beta_0 = Q_{yy}(0, x) y' \cdot \beta_0 + Q^{(2)}(y', x)\ , \quad Q^{(2)} := \displaystyle\int_0^1 (1-t) Q_{yyy}(ty', x) y' \cdot y' \cdot \beta_0\, dt \\ \\ P(y', x) = P(0, x) + P_y(0, x) \cdot y' + Q^{(3)}(y', x)\ , \quad Q^{(3)} := \displaystyle\int_0^1 (1-t) P_{yy}(ty', x)\, y' \cdot y'\, dt\ . \end{cases} \tag{13}$$

Thus, since[13)] $\omega \cdot \beta = \omega \cdot b + D_\omega s + D_\omega a \cdot y'$, we find

$$H(y' + \varepsilon\beta, x) = E + \omega \cdot y' + Q(y', x) + \varepsilon Q'(y', x) + \varepsilon F'(y', x) + \varepsilon^2 P'(y', x) \tag{14}$$

with P' as in (10)–(11) and

$$\begin{cases} Q'(y', x) := Q^{(1)} + Q^{(2)} + Q^{(3)} = O(|y'|^2) \\ \\ F'(y', x) := \omega \cdot b + D_\omega s + P(0, x) + \Big\{D_\omega a + Q_{yy}(0, x) b + Q_{yy}(0, x) s_x + P_y(0, x)\Big\} \cdot y' \end{cases} \tag{15}$$

[12)] As usual, we denote $s_x = \partial_x s = (s_{x_1}, ..., s_{x_d})$ and a_x denotes the matrix $(a_x)_{ij} := \frac{\partial a_j}{\partial x_i}$; as above, we often do not report in the notation the dependence upon ε (but β_0, A and β do depend also on ε).

[13)] Recall that $\omega \cdot s_x = D_\omega s$ and $\omega \cdot (a_x y') = (D_\omega a) \cdot y'$.

It is now easy to see that *there exist a unique constant b and unique functions s and a* (with zero average) *such that F' is constant*. In fact, if[14)]

$$\begin{cases} s := -D_\omega^{-1}\Big(P(0,x) - P_0(0)\Big) = -\sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} \dfrac{P_n(0)}{i\omega\cdot n} e^{in\cdot x} \\ b := -\langle Q_{yy}(0,\cdot)\rangle^{-1}\Big(\langle Q_{yy}(0,\cdot)s_x\rangle + \langle P_y(0,\cdot)\rangle\Big) \\ a := -D_\omega^{-1}\Big(Q_{yy}(0,x)(b+s_x) + P_y(0,x)\Big) \end{cases}$$

then $F' = \omega\cdot b + P_0(0)$. Thus, with this determination of g in (8), recalling **(b)** of Remark 2, we find that (7) holds with

$$\begin{cases} E_1 := E + \varepsilon\widetilde{E}\ , & \widetilde{E} := \omega\cdot b + P_0(0) \\ Q_1(y',x') := Q(y',x') + \varepsilon\widetilde{Q}(y',x')\ , & \widetilde{Q} := \displaystyle\int_0^1 Q_x(y', x' + t\varepsilon\alpha)\cdot\alpha\, dt\ + Q'(y',\varphi(x')) \\ P_1(y',x') := P'(y',\varphi(x'))\ . & \end{cases}$$

Clearly, for ε small enough $\langle\partial^2_{y'}Q_1(0,\cdot)\rangle$ is invertible and, if $T := \langle Q_{yy}(0,\cdot\rangle^{-1}$, we may write

$$T_1 := \langle\partial^2_{y'}Q_1(0,\cdot)\rangle^{-1} =: T + \varepsilon\widetilde{T}\ . \tag{16}$$

Step (ii) This step (missing in Kolmogorov's paper [1]) will be divided in two lemmata: the first Lemma consists in equipping with estimates the construction of Kolmogorov's transformation ϕ_1 discussed in step **(i)** and the second one describes the iteration and its convergence.

Let $M := \|P\|_{\xi,\varepsilon_0}$; let $C > 1$ be a constant such that[15)]

$$|E|\ ,\ |\omega|\ ,\ \|Q\|_\xi, \|T\| < C\ ; \tag{17}$$

fix[16)]

$$0 < \delta < \xi \quad \text{and define} \quad \bar{\xi} := \xi - \frac{2}{3}\delta\ , \quad \xi' := \xi - \delta\ .$$

Lemma 4. *There exist constants $\bar{c} = \bar{c}(d,\tau,\kappa) > 1$, $\bar{\mu}\in\mathbb{Z}_+$ and $\bar{\nu} = \nu(d,\tau) > 1$ such that*[17)]

$$\begin{cases} \|s_x\|_{\bar{\xi}}, |b|, |\widetilde{E}|, \|a\|_{\bar{\xi}}, \|a_x\|_{\bar{\xi}}, \|\beta_0\|_{\bar{\xi}}, \|\beta\|_{\bar{\xi}}, \|Q'\|_{\bar{\xi}}, \|\partial^2_{y'}Q'(0,\cdot)\|_0 \leqslant \bar{c}C^{\bar{\mu}}\delta^{-\bar{\nu}}\ M =: \bar{L}\ , \\ \\ \|P'\|_{\bar{\xi}} \leqslant \bar{c}C^{\bar{\mu}}\delta^{-\bar{\nu}}\ M^2 = \bar{L}M\ . \end{cases} \tag{18}$$

Furthermore, if ε_* satisfies

$$\varepsilon_*\,\bar{L} \leqslant \frac{\delta}{3}\ , \tag{19}$$

14) Recall point **(c)** in Remark 1 above for the notation.

15) The notation in Eq. (17) means that each term on the l.h.s. is bounded by the r.h.s. The choice of norms on finite dimensional spaces ($\mathbb{R}^d$, $\mathbb{C}^d$, space of matrices, tensors, etc.) is not particularly relevant, however for matrices, tensors (and, in general, linear operators) we shall work with the "operator norm", i.e., the norm defined as $\|L\| = \sup_{u\neq 0}\|Lu\|/\|u\|$, so that $\|Lu\| \leqslant \|L\|\|u\|$, an estimate, which will be constantly used.

16) The parameter ξ' will be the size of the domain of analyticity of the new symplectic variables (y',x'), domain on which we shall bound the Hamiltonian $H_1 = H\circ\phi_1$, while $\bar{\xi}$ is the size of an intermediate domain where we shall bound various functions of y' and x.

17) Here $\|\cdot\|_{\bar{\xi}} = \|\cdot\|_{\bar{\xi},\varepsilon_0}$.

the following is true. For $|\varepsilon| < \varepsilon_*$, *the map* $\psi_\varepsilon(x) := x + \varepsilon a(x)$ *has an analytic inverse* $\varphi(x') = x' + \varepsilon\alpha(x';\varepsilon)$ *such that, for all* $|\varepsilon| < \varepsilon_*$,

$$\|\alpha\|_{\xi',\varepsilon_*} \leqslant \bar{L} \quad \textit{and}\ \forall\ |\varepsilon| < \varepsilon_*, \quad \varphi = \mathrm{id} + \varepsilon\alpha : \mathbb{T}^d_{\xi'} \to \mathbb{T}^d_{\bar{\xi}}; \tag{20}$$

for any $(y', x, \varepsilon) \in W_{\bar{\xi},\varepsilon_*}$, $|y' + \varepsilon\beta(y', x)| < \xi$; *the map* ϕ_1 *is a symplectic diffeomorphism and*

$$\phi_1 = \big(y' + \varepsilon\beta(y', \varphi(x')), \varphi(x')\big) : W_{\xi',\varepsilon_*} \to D^d_\xi \times \mathbb{T}^d_\xi, \quad \textit{and} \quad \|\tilde{\phi}\|_{\xi',\varepsilon_*} \leqslant \bar{L}, \tag{21}$$

where $\tilde{\phi}$ *is defined by the relation* $\phi_1 =: \mathrm{id} + \varepsilon\tilde{\phi}$.

Finally, there exist $c \geqslant \bar{c}$, $\mu \geqslant \bar{\mu}$ *and* $\nu \geqslant \bar{\nu}$ *such that if*[18]

$$\varepsilon_* L := \varepsilon_* c C^\mu \delta^{-\nu}\ M \leqslant \frac{\delta}{3}, \tag{22}$$

then

$$\begin{cases} |\widetilde{E}|\ ,\ \|\widetilde{Q}\|_{\xi',\varepsilon_*}\ ,\ \|\widetilde{T}\|\ ,\ \|\tilde{\phi}\|_{\xi',\varepsilon_*} \leqslant L, \\ \\ \|P_1\|_{\xi',\varepsilon_*} \leqslant LM. \end{cases} \tag{23}$$

Proof. The estimates in (18) follow easily from **(c)** of Remark 1: as an example, let us work out the first two estimates in (18), i.e., the estimates on $\|s_x\|_{\bar{\xi}}$ and $|b|$: actually these estimates will be given on a larger intermediate domain, namely, $W_{\xi-\frac{\delta}{3},\varepsilon_0}$, allowing to give the remaining bounds on the smaller domain $W_{\bar{\xi},\varepsilon_0}$. Let $f(x) := P(0,x) - \langle P(0,\cdot)\rangle$. By definition of $\|\cdot\|_\xi$ and M, it follows that $\|f\|_\xi \leqslant \|P(0,x)\|_\xi + \|\langle P(0,\cdot)\rangle\|_\xi \leqslant 2M$. By (6) with $p = 1$ and $\xi' = \xi - \frac{\delta}{3}$, one gets

$$\|s_x\|_{\xi-\frac{\delta}{3}} \leqslant \bar{B}_1 \frac{2M}{\kappa}\, 3^{k_1}\delta^{-k_1},$$

which is of the form (18), provided $\bar{c} \geqslant \bar{B}_1 \cdot 2 \cdot 3^{k_1}/\kappa$ and $\bar{\nu} \geqslant k_1$. To estimate b, we need to bound first $|Q_{yy}(0,x)|$ and $|P_y(0,x)|$ for real x. To do this we can use Cauchy estimates: by (4) with $p = 2$ and, respectively, $p = 1$, and $\xi' = 0$, we get

$$\|Q_{yy}(0,\cdot)\|_0 \leqslant m\ B_2 C\xi^{-2} < m\ B_2 C\delta^{-2} \quad \text{and} \quad \|P_y(0,x)\|_0 < m B_1 M\delta^{-1},$$

where $m = m(d) \geqslant 1$ is a constant which depends on the choice of the norms (recall also that $\delta < \xi$). Putting these bounds together, one gets that $|b|$ can be bounded by the r.h.s. of (18) provided $\bar{c} \geqslant m(B_2\bar{B}_1 \cdot 2 \cdot 3^{k_1}\kappa^{-1} + B_1)$, $\bar{\mu} \geqslant 2$ and $\bar{\nu} \geqslant k_1 + 2$. No new ideas are required to work out the other estimates in (18).

Next, we show how (19) implies the existence of the inverse of ψ_ε satisfying (20). The defining relation $\psi_\varepsilon \circ \varphi = \mathrm{id}$ implies that $\alpha(x') = -a(x' + \varepsilon\alpha(x'))$, where $\alpha(x')$ is short for $\alpha(x';\varepsilon)$, and such relation is a fixed point equation for the non–linear operator $f : u \to f(u) := -a(\mathrm{id} + \varepsilon u)$. To find a fixed point for this equation one can use a standard contraction Lemma (see [7]). Let Y denote the closed ball (with respect to the sup–norm) of continuous functions $u : \mathbb{T}^d_{\xi'} \times \{|\varepsilon| < \varepsilon_*\} \to \mathbb{C}^d$ such that $\|u\|_{\xi',\varepsilon_*} \leqslant \bar{L}$. By (19), $|\operatorname{Im}(x' + \varepsilon u(x'))| \leqslant \xi' + \varepsilon_*\bar{L} \leqslant \xi' + \frac{\delta}{3} = \bar{\xi}$, for any $u \in Y$ and any $x' \in \mathbb{T}^d_{\xi'}$; thus, $\|f(u)\|_{\xi',\varepsilon_*} \leqslant \|a\|_{\bar{\xi}} \leqslant \bar{L}$ by (18), so that $f : Y \to Y$; notice that, in particular, this means that f sends x–periodic functions into x–periodic functions. Moreover, (19) implies also that f is a contraction: if $u, v \in Y$, then, by the mean value theorem and (18), $|f(u) - f(v)| \leqslant \|a_x\|_{\bar{\xi}}\ |\varepsilon|\ |u - v| \leqslant \bar{L}|\varepsilon|\ |u - v|$, so that, by taking the sup–norm, one has $\|f(u) - f(v)\|_{\xi'} \leqslant \varepsilon_*\bar{L}\|u - v\|_{\xi'} \leqslant \frac{1}{3}\|u - v\|_{\xi'}$ showing that f is a contraction. Thus, there exists a unique $\alpha \in Y$ such that $f(\alpha) = \alpha$. Furthermore, recalling that the fixed point is achieved as the uniform limit $\lim_{n\to\infty} f^n(0)$ $(0 \in Y)$ and since $f(0) = -a$ is analytic, so is $f^n(0)$ for any n and, hence, by Weierstrass Theorem on the

[18] Notice that, since $L \geqslant \bar{L}$, (22) implies (19).

uniform limit of analytic functions, the limit α itself is analytic. In conclusion, $\varphi \in \mathcal{B}_{\xi',\varepsilon_*}$ and (20) holds. Next, for $(y', x, \varepsilon) \in W_{\bar{\xi},\varepsilon_*}$, by (18), one has $|y' + \varepsilon\beta(y', x)| < \bar{\xi} + \varepsilon_* \bar{L} \leqslant \bar{\xi} + \frac{\delta}{3} = \xi - \frac{\delta}{3} < \xi$ so that (21) holds and ϕ_1 defines a symplectic diffeomorphism[19] satisfying (21) and the fourth inequality in the first line of (23).

It remains to show the other estimates in (23). Since $L \geqslant \bar{L}$, the bound on $|\widetilde{E}|$ comes straightforwardly from (18). By (20) and (18), one has $\|P_1\|_{\xi',\varepsilon_*} \leqslant \|P'\|_{\bar{\xi},\varepsilon_*} \leqslant \bar{L}M \leqslant LM$. Now, by Cauchy estimates, (18) and (21), it follows immediately that[20]

$$\|\widetilde{Q}\|_{\xi',\varepsilon_*} \ , \ 2C^2 \, \|\partial^2_{y'} \widetilde{Q}(0,\cdot)\|_{0,\varepsilon_*} \leqslant cC^{\mu}\delta^{-\nu} \ M = L \ , \tag{24}$$

for suitable constants $c \geqslant \bar{c}$, $\bar{\mu} \geqslant \mu$, $\bar{\nu} \geqslant \nu$. Thus,

$$\begin{aligned} \langle \partial^2_{y'} Q_1(0,\cdot)\rangle &= \langle \partial^2_{y} Q(0,\cdot)\rangle + \varepsilon \langle \partial^2_{y'} \widetilde{Q}(0,\cdot)\rangle = T^{-1}(\mathbb{1}_d + \varepsilon T \langle \partial^2_{y'} \widetilde{Q}(0,\cdot)\rangle) \\ &=: T^{-1}(\mathbb{1}_d + \varepsilon R) \ , \end{aligned} \tag{25}$$

and, in view of (17) and (24), we see that $\|R\| < L/(2C)$. Therefore, by (22), $\varepsilon_* \|R\| < 1/6 < 1/2$, implying that $(\mathbb{1} + \varepsilon R)$ is invertible and

$$(\mathbb{1}_d + \varepsilon R)^{-1} = \mathbb{1}_d + \sum_{k=1}^{\infty} (-1)^k \varepsilon^k R^k =: \mathbb{1} + \varepsilon D$$

with $\|D\| \leqslant \|R\|/(1 - |\varepsilon| \, \|R\|) < L/C$. In conclusion, by (25), and the estimate on $\|D\|$,

$$T_1 = (\mathbb{1} + \varepsilon R)^{-1} T = T + \varepsilon DT =: T + \varepsilon \widetilde{T} \ , \qquad \|\widetilde{T}\| \leqslant \|D\| C \leqslant \frac{L}{C} C = L \ ,$$

proving last estimate in (23) and, hence, Lemma 4. □

Next lemma shows that, for $|\varepsilon|$ small enough, Kolmogorov's construction can be iterated and convergence proven.

Lemma 5. *Let*

$$C := 2 \max \Big\{ |E|, |\omega|, \|Q\|_{\xi}, \|T\|, 1 \Big\} \ . \tag{26}$$

Fix $0 < \xi_* < \xi$ and, for $j \geqslant 0$, let

$$\begin{cases} \xi_0 := \xi \\ \delta_0 := \dfrac{\xi - \xi_*}{2} \end{cases} \qquad \begin{cases} \delta_j := \dfrac{\delta_0}{2^j} \\ \xi_{j+1} := \xi_j - \delta_j = \xi_* + \frac{\delta_0}{2^j} \end{cases}$$

(so that $\xi_j \downarrow \xi_*$). Let, also, $H_0 := H$, $E_0 := E$, $Q_0 := Q$, $K_0 := K$, $P_0 := P$ and let c, μ and ν be as in (22) with $\delta = \delta_0$, and assume that ε_* satisfies

$$\varepsilon_* \, c_* \, d_* \, \|P\|_{\xi,\varepsilon_0} \leqslant 1 \quad \text{where} \quad c_* := 3c \, \delta_0^{-(\nu+1)} \, C^{\mu} \ , \ \ d_* := 2^{\nu+1} \ \geqslant 1. \tag{27}$$

Then, one can construct a sequence of symplectic transformations

$$\phi_j : W_{\xi_j,\varepsilon_*} \to D^d_{\xi_{j-1}} \times \mathbb{T}^d_{\xi_{j-1}} \quad (j \geqslant 1), \tag{28}$$

so that

$$H_j := H_{j-1} \circ \phi_j =: K_j + \varepsilon^{2^j} P_j \tag{29}$$

[19] Notice, in particular, that the matrix $\mathbb{1}_d + \varepsilon a_x$ is, for any $x \in \mathbb{T}^d_{\bar{\xi}}$, invertible with inverse $\mathbb{1}_d + \varepsilon S(x;\varepsilon)$; in fact, since $\|\varepsilon a_x\|_{\bar{\xi}} < \varepsilon_* \bar{L} < 1/3$ the matrix $\mathbb{1}_d + \varepsilon a_x$ is invertible with inverse given by the "Neumann series" $(\mathbb{1}_d + \varepsilon a_x)^{-1} = \mathbb{1}_d + \sum_{k=1}^{\infty} (-1)^k (\varepsilon a_x)^k =: \mathbb{1}_d + \varepsilon S(x;\varepsilon)$, so that $\|S\|_{\bar{\xi},\varepsilon_*} \leqslant \|a_x\|_{\bar{\xi},\varepsilon_*}/(1 - |\varepsilon| \|a_x\|_{\bar{\xi},\varepsilon_*}) < \frac{3}{2}\bar{L}$.

[20] It is only here that a constant $L > \bar{L}$ is needed; the (irrelevant) factor $2C^2$ has been introduced for later convenience.

converges uniformly to a Kolmogorov's normal form. More precisely, $\varepsilon^{2^j} P_j$, $\Phi_j := \phi_1 \circ \phi_2 \circ \cdots \circ \phi_j$, E_j, K_j, Q_j *converge uniformly on* W_{ξ_*,ε_*} *to, respectively,* 0, ϕ_*, E_*, K_*, Q_*, *which are real–analytic on* W_{ξ_*,ε_*} *and* $H \circ \phi_* = K_* = E_* + \omega \cdot y + Q_*$ *with* $Q_* = O(|y|^2)$. *Finally, the following estimates hold for any* $|\varepsilon| < \varepsilon_*$ *and for any* $i \geqslant 0$*:*

$$|\varepsilon|^{2^i} M_i := |\varepsilon|^{2^i} \|P_i\|_{\xi_i,\varepsilon_*} \leqslant \frac{(\,|\varepsilon| c_*\, d_* M)^{2^i}}{c_*\, {d_*}^{i+1}}\ , \tag{30}$$

$$\|\phi_* - \mathrm{id}\,\|_{\xi_*}, |E - E_*|, \|Q - Q_*\|_{\xi_*}, \|T - T_*\| \leqslant |\varepsilon|\, c_*\, d_* M\ , \tag{31}$$

where $M := \|P\|_{\xi,\varepsilon_0} = M_0$ *and* $T_* := \langle \partial_y^2 Q_*(0,\cdot)\rangle^{-1}$.

Proof. Notice that the constant C defined in (26) satisfies (17) and that (27) implies (22) (and, hence, (19)).

Let us assume (*inductive hypothesis*) that we can iterate $j \geqslant 1$ times Kolmogorov transformation obtaining j symplectic transformations $\phi_{i+1} : W_{\xi_{i+1},\varepsilon_*} \to D^d_{\xi_i} \times \mathbb{T}^d_{\xi_i}$, for $0 \leqslant i \leqslant j-1$, and j Hamiltonians $H_{i+1} = H_i \circ \phi_{i+1} = K_{i+1} + \varepsilon^{2^i} P_{i+1}$ real–analytic on $W_{\xi_{i+1},\varepsilon_*}$ such that, for any $0 \leqslant i \leqslant j-1$,

$$\begin{cases} |\omega|, |E_i|, \|Q_i\|_{\xi_i}, \|T_i\| < C \\ \\ |\varepsilon|^{2^i} L_i := |\varepsilon|^{2^i} c C^\mu \delta_0^{-\nu}\, 2^{\nu i} M_i \leqslant \frac{\delta_i}{3}\ . \end{cases} \tag{32}$$

Observe that for $j = 1$, it is $i = 0$ and (32) is implied by the definition of C in (26) and by condition (27).

Because of (32), (22) holds[21] for H_i and Lemma 4 can be applied to H_i, and one has, for $0 \leqslant i \leqslant j-1$ and for any $|\varepsilon| < \varepsilon_*$ (compare (23)):

$$\begin{aligned} &|E_{i+1}| \leqslant |E_i| + |\varepsilon|^{2^i} L_i\ , \quad \|Q_{i+1}\|_{\xi_{i+1}} \leqslant \|Q_i\|_{\xi_i} + |\varepsilon|^{2^i} L_i\ , \quad \|T_{i+1}\| \leqslant \|T_i\| + |\varepsilon|^{2^i} L_i\ , \\ &\|\phi_{i+1} - \mathrm{id}\,\|_{\xi_{i+1}} \leqslant |\varepsilon|^{2^i} L_i\ , \quad M_{i+1} \leqslant M_i L_i\ . \end{aligned} \tag{33}$$

Observe that, by definition of c_*, d_* in (27) and of L_i in (32), one has $|\varepsilon|^{2^i} L_i(3\delta_i^{-1}) = c_*\, {d_*}^i |\varepsilon|^{2^i} M_i$, so that $L_i < c_*\, {d_*}^i M_i$, thus by last relation in (33), for any $0 \leqslant i \leqslant j-1$, $|\varepsilon|^{2^{i+1}} M_{i+1} < c_*\, {d_*}^i (M_i |\varepsilon|^{2^i})^2$, which, after having been iterated, yields (30) for $0 \leqslant i \leqslant j$.

Next, we show that, thanks to (27), (32) holds also for $i = j$. In fact, by (32), (33) and the definition of C in (26):

$$|E_j| \leqslant |E| + \sum_{i=0}^{j-1} \varepsilon_*^{2^i} L_i \leqslant |E| + \frac{1}{3} \sum_{i\geqslant 0} \delta_i < |E| + \frac{1}{3} \sum_{i \geqslant 0} 2^{-i} = |E| + \frac{2}{3} < C\ .$$

The bounds for $\|Q_j\|$ and $\|T_j\|$ are proven in an identical manner. Define L_j as in (32) for $i = j$. Now, by $(30)_{i=j}$ and (27),

$$|\varepsilon|^{2^j} L_j(3\delta_j^{-1}) = c_*\, {d_*}^j |\varepsilon|^{2^j} M_j \leqslant c_*\, {d_*}^j (c_*\, d_* \varepsilon_* M)^{2^j} / (c_*\, {d_*}^{j+1}) \leqslant 1/\, d_* < 1\ ,$$

which implies the second inequality in (32) with $i = j$; the proof of the induction is finished and one can construct an *infinite sequence* of Kolmogorov transformations satisfying (32), (33) and (30) *for all* $i \geqslant 0$.

To check (31), we observe that for any $i \geqslant 0$

$$|\varepsilon|^{2^i} L_i = \frac{\delta_0}{3 \cdot 2^i}\ c_*\, {d_*}^i |\varepsilon|^{2^i} M_i \leqslant \frac{1}{2^{i+1}} (|\varepsilon| c_*\, d_* M)^{2^i} \leqslant \Big(\frac{|\varepsilon| c_*\, d_* M}{2}\Big)^{i+1},$$

[21] *Mutatis mutandis.*

and therefore

$$\sum_{i\geqslant 0} |\varepsilon|^{2^i} L_i \leqslant \sum_{i\geqslant 1} \Big(\frac{|\varepsilon| c_* \, d_* M}{2}\Big)^i \leqslant |\varepsilon| c_* \, d_* M \ .$$

Thus,

$$\|Q - Q_*\|_{\xi_*} \leqslant \sum_{i\geqslant 0} \|\widetilde{Q}_i\|_{\xi_i} \leqslant \sum_{i\geqslant 0} |\varepsilon|^{2^i} L_i \leqslant |\varepsilon| c_* \, d_* M;$$

and analogously for $|E - E_*|$ and $\|T - T_*\|$.

To estimate $\|\phi_* - \mathrm{id}\,\|_{\xi_*}$, observe that for any $i \geqslant 2$

$$\|\Phi_i - \mathrm{id}\,\|_{\xi_i} \leqslant \|\Phi_{i-1} \circ \phi_i - \phi_i\|_{\xi_i} + \|\phi_i - \mathrm{id}\,\|_{\xi_i} \leqslant \|\Phi_{i-1} - \mathrm{id}\,\|_{\xi_{i-1}} + |\varepsilon|^{2^{i-1}} L_{i-1},$$

which, after having been iterated, yields $\|\Phi_i - \mathrm{id}\,\|_{\xi_i} \leqslant \sum_{k=0}^{i} |\varepsilon|^{2^k} L_k \leqslant |\varepsilon| c_* \, d_* M$; taking the limit over i completes the proof of (31), of Lemma 5 and, hence, of Kolmogorov's Theorem. □

3. ACKNOWLEDGMENTS

The work is partially supported by the Italian MIUR project"Metodi variazionali ed equazioni differenziali nonlineari".

REFERENCES

1. Kolmogorov, A.N., On the Conservation of Conditionally Periodic Motions under Small Perturbation of the Hamiltonian, *Dokl. akad. nauk SSSR*, 1954, vol. 98, pp. 527–530. Engl. transl.: *Stochastic Behavior in Classical and Quantum Hamiltonian Systems*, Volta Memorial conference, Como, 1977, Lecture Notes in Physics, vol. 93, Springer, 1979, pp. 51–56.
2. Kolmogorov, A.N., Théorie générale des systèmes dynamiques et mécanique classique (French, 1957), *Proceedings of the International Congress of Mathematicians, Amsterdam*, 1954, Vol. 1, pp. 315–333, Erven P. Noordhoff N.V., Groningen; North-Holland Publishing Co., Amsterdam.
3. Arnol'd, V.I., Kozlov, V.V., and Neishtadt, A.I., *Mathematical Aspects of Classical and Celestial Mechanics*, 3rd ed., Dynamical Systems III, Encyclopaedia of Mathematical Sciences, vol. 3, Berlin: Springer-Verlag, 2006.
4. Arnol'd, V.I., Proof of a Theorem by A. N. Kolmogorov on the Invariance of Quasi-Periodic Motions under Small Perturbations of the Hamiltonian, *Uspehi Mat. Nauk*, 1963, vol. 18, no. 5(113), pp. 13–40 [*Russ. Math. Surv.* (Engl. transl.), 1963, vol. 18, no. 5, pp. 9–36.]
5. Moser, J.K., On Invariant Curves of Area-Preserving Mappings of an Annulus, *Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl. II*, 1962, pp. 1–20.
6. Moser, J.K., AMS Mathematical Review MR0097598 (20 #4066) of the article [2]
7. Kolmogorov, A.N. and Fomin, S.V., *Elements of the Theory of Functions and Functional Analysis*, Courier Dover Publications, 1999 [Translated from the original Russian edition (Izdat. Moskov. Univ., Moscow, 1954)].