63. Tung A, Hou J, Han J (2001) Spatial clustering in the presence of obstacles. In: Proceedings of the 17th ICDE, Heidelberg, pp 359–367
64. Voorhees EM (1986) Implementing agglomerative hierarchical clustering algorithms for use in document retrieval. Inf Proc Manag 22(6):465–476
65. Wang W, Yang J, Muntz R (1997) STING: a statistical information grid approach to spatial data mining. In: Proceedings of the 23rd Conference on VLDB, Athens, pp 186–195
66. Xu X, Ester M, Kriegel HP, Sander J (1998) A distribution-based clustering algorithm for mining large spatial datasets. In: Proceedings of the 14th ICDE, Orlando, pp 324–331
67. Yao A (1982) On constructing minimum spanning trees in k-dimensional space and related problems. SIAM J Comput 11(4):721–736
68. Yu SX, Shi J (2003) Multiclass spectral clustering. In: International Conference on Computer Vision, Nice
69. Zhang B (2001) Generalized k-harmonic means – dynamic weighting of data in unsupervised learning. In: Proceedings of the 1st SIAM International Conference on Data Mining, Chicago
70. Zhang T, Ramakrishnan R, Livny M (1997) BIRCH: A new data clustering algorithm and its applications. J Data Mining Knowl Discov 1(2):141–182

Kolmogorov–Arnold–Moser (KAM) Theory

LUIGI CHIERCHIA
Dipartimento di Matematica, Università "Roma Tre", Roma, Italy

Article Outline

Glossary

Action-angle variables A particular set of variables $(y, x) = ((y_1, \ldots, y_d), (x_1, \ldots, x_d))$, x_i ("angles") defined modulus 2π, particularly suited to describe the general behavior of an integrable system.

Fast convergent (Newton) method Super-exponential algorithms, mimicking Newton's method of tangents, used to solve differential problems involving small divisors.

Hamiltonian dynamics The dynamics generated by a Hamiltonian system on a symplectic manifold, i. e., on an even-dimensional manifold endowed with a symplectic structure.

Hamiltonian system A time reversible, conservative (without dissipation or expansion) dynamical system, which generalizes classical mechanical systems (solutions of Newton's equation $m_i \ddot{x}_i = f_i(x)$, with $1 \leq i \leq d$ and $f = (f_1, \ldots, f_d)$ a conservative force field); they are described by the flow of differential equations (i. e., the time t map associating to an initial condition, the solution of the initial value problem at time t) on a symplectic manifold and, locally, look like the flow associated with the system of differential equation $\dot{p} = -H_q(p, q)$, $\dot{q} = H_p(p, q)$ where $p = (p_1, \ldots, p_d)$, $q = (q_1, \ldots, q_d)$.

Integrable Hamiltonian systems A very special class of Hamiltonian systems, whose orbits are described by linear flows on the standard d-torus: $(y, x) \to (y, x + \omega t)$ where (y, x) are action-angle variables and t is time; the ω_i's are called the "frequencies" of the orbit.

Invariant tori Manifolds diffeomorphic to tori invariant for the flow of a differential equation (especially of Hamiltonian differential equations); establishing the existence of tori invariant for Hamiltonian flows is the main object of KAM theory.

KAM Acronym from the names of Kolmogorov (Andrey Nikolaevich Kolmogorov, 1903–1987), Arnold (Vladimir Igorevich Arnold, 1937) and Moser (Jürgen K. Moser, 1928–1999), whose results, in the 1950's and 1960's, in Hamiltonian dynamics, gave rise to the theory presented in this article.

Nearly-integrable Hamiltonian systems Hamiltonian systems which are small perturbations of an integrable system and which, in general, exhibit a much richer dynamics than the integrable limit. Nevertheless, KAM theory asserts that, under suitable assumptions, the majority (in the measurable sense) of the initial data of a nearly-integrable system behaves as in the integrable limit.

Quasi-periodic motions Trajectories (solutions of a system of differential equations), which are conjugate to linear flow on tori.

Small divisors/denominators Arbitrary small combinations of the form $\omega \cdot k := \sum_{j=1}^{d} \omega_j k_j$ with $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d$ a real vector and $k \in \mathbb{Z}^d$ an integer vector different from zero; these combinations arise in the denominators of certain expansions appearing in the perturbation theory of Hamiltonian systems, making (when $d > 1$) convergent arguments very delicate. Physically, small divisors are related to "resonances", which are a typical feature of conservative systems.

Stability The property of orbits of having certain properties similar to a reference limit; more specifically, in the context of KAM theory, stability is normally referred to as the property of action variables of staying close to their initial values.

Symplectic structure A mathematical structure (a differentiable, non-degenerate, closed 2-form) apt to describe, in an abstract setting, the main geometrical features of conservative differential equations arising in mechanics.

Definition of the Subject

KAM theory is a mathematical, quantitative theory which has as its primary object the persistence, under small (Hamiltonian) perturbations, of typical trajectories of integrable Hamiltonian systems. In integrable systems with bounded motions, the typical trajectory is quasi-periodic, i. e., may be described through the linear flow $x \in \mathbb{T}^d \to x + \omega t \in \mathbb{T}^d$ where $\mathbb{T}^d$ denotes the standard d-dimensional torus (see Sect. "Introduction" below), t is time, and $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d$ is the set of frequencies of the trajectory (if $d = 1$, $2\pi/\omega$ is the *period* of the motion).

The main motivation for KAM theory is related to stability questions arising in celestial mechanics which were addressed by astronomers and mathematicians such as Kepler, Newton, Lagrange, Liouville, Delaunay, Weierstrass, and, from a more modern point of view, Poincaré, Birkhoff, Siegel, ...

The major breakthrough in this context, was due to Kolmogorov in 1954, followed by the fundamental work of Arnold and Moser in the early 1960s, who were able to overcome the formidable technical problem related to the appearance, in perturbative formulae, of arbitrarily small divisors[1]. Small divisors make the use of classical analytical tools (such as the standard Implicit Function Theorem, fixed point theorems, etc.) impossible and could be controlled only through a "fast convergent method" of Newton-type[2], which allowed, in view of the super-exponential rate of convergence, counterbalancing the divergences introduced by small divisors.

Actually, the main bulk of KAM theory is a set of *techniques* based, as mentioned, on fast convergent methods, and solving various questions in Hamiltonian (or generalizations of Hamiltonian) dynamics. By now, there are excellent reviews of KAM theory – especially Sect. 6.3 of [6] and [60] – which should complement the reading of this article, whose main objective is not to review but rather to explain the main fundamental ideas of KAM theory. To do this, we re-examine, in modern language, the main ideas introduced, respectively, by the founders of KAM theory, namely Kolmogorov (in Sect. "Kolmogorov Theorem"), Arnold (in Sect. "Arnold's Scheme") and Moser (Sect. "The Differentiable Case: Moser's Theorem").

In Sect. "Future Directions" we briefly and informally describe a few developments and applications of KAM theory: this section is by no means exhaustive and is meant to give a non technical, short introduction to some of the most important (in our opinion) extensions of the original contributions; for more detailed and complete reviews we recommend the above mentioned articles Sect. 6.3 of [6] and [60].

Appendix A contains a quantitative version of the classical Implicit Function Theorem.

A set of technical notes (such as notes 17, 18, 19, 21, 24, 26, 29, 30, 31, 34, 39), which the reader not particularly interested in technical mathematical arguments may skip, are collected in Appendix B and complete the mathematical expositions. Appendix B also includes several other complementary notes, which contain either standard material or further references or side comments.

Introduction

In this article we will be concerned with Hamiltonian flows on a symplectic manifold $(\mathcal{M}, \mathrm{d}y \wedge \mathrm{d}x)$; for general information, see, e. g., [5] or Sect. 1.3 of[6]. Notation, main definitions and a few important properties are listed in the following items.

(a) As symplectic manifold ("phase space") we shall consider $\mathcal{M} := B \times \mathbb{T}^d$ with $d \geq 2$ (the case $d = 1$ is trivial for the questions addressed in this article) where: B is an open, connected, bounded set in $\mathbb{R}^d$; $\mathbb{T}^d := \mathbb{R}^d/(2\pi\mathbb{Z}^d)$ is the standard flat d-dimensional torus with periods[3] 2π

(b) $\mathrm{d}y \wedge \mathrm{d}x := \sum_{i=1}^d \mathrm{d}y_i \wedge \mathrm{d}x_i$, $(y \in B,\ x \in \mathbb{T}^d)$ is the standard symplectic form[4]

(c) Given a real-analytic (or smooth) function $H\colon \mathcal{M} \to \mathbb{R}$, the *Hamiltonian flow governed by* H is the one-parameter family of diffeomorphisms $\phi_H^t\colon \mathcal{M} \to \mathcal{M}$, which to $z \in \mathcal{M}$ associates the solution at time t of the differential equation[5]

$$\dot{z} = J_{2d} \nabla H(z)\,, \quad z(0) = z\,, \tag{1}$$

where $\dot{z} = \mathrm{d}z/\mathrm{d}t$, J_{2d} is the standard symplectic $(2d \times 2d)$-matrix

$$J_{2d} = \begin{pmatrix} 0 & -\mathbb{1}_d \\ \mathbb{1}_d & 0 \end{pmatrix},$$

$\mathbb{1}_d$ denotes the unit $(d \times d)$-matrix, 0 denotes a $(d \times d)$ block of zeros, and ∇ denotes gradient; in the symplectic coordinates $(y, x) \in B \times \mathbb{T}^d$, equations (1) reads

$$\begin{cases} \dot{y} = -H_x(y, x) \\ \dot{x} = H_y(y, x) \end{cases}, \qquad \begin{cases} y(0) = y \\ x(0) = x \end{cases} \tag{2}$$

Clearly, the flow ϕ_H^t is defined until $y(t)$ eventually reaches the border of B.

Equations (1) and (2) are called the *Hamilton's equations* with *Hamiltonian* H; usually, the symplectic (or "conjugate") variables (y, x) are called *action-angle* variables[6]; the number d (= half of the dimension of the phase space) is also referred to as "the number of degrees of freedom[7]".

The Hamiltonian H is constant over trajectories $\phi_H^t(z)$, as it follows immediately by differentiating $t \to H(\phi_H^t(z))$. The constant value $E = H(\phi_H^t(z))$ is called the energy of the trajectory $\phi_H^t(z)$.

Hamilton equations are left invariant by *symplectic* (or "canonical") change of variables, i. e., by diffeomorphisms on $\mathcal{M}$ which preserve the 2-form $\mathrm{d}y \wedge \mathrm{d}x$; i. e., if $\phi\colon (y, x) \in \mathcal{M} \to (\eta, \xi) = \phi(y, x) \in \mathcal{M}$ is a diffeomorphism such that $\mathrm{d}\eta \wedge \mathrm{d}\xi = \mathrm{d}y \wedge \mathrm{d}x$, then

$$\phi \circ \phi_H^t \circ \phi^{-1} = \phi_{H \circ \phi^{-1}}^t \,. \tag{3}$$

An equivalent condition for a map ϕ to be symplectic is that its Jacobian ϕ' is a *symplectic matrix*, i. e.,

$$\phi'^T J_{2d} \phi' = J_{2d} \tag{4}$$

where J_{2d} is the standard symplectic matrix introduced above and the superscript T denotes matrix transposition.

By a (generalization of a) theorem of Liouville, the Hamiltonian flow is symplectic, i. e., the map $(y, x) \to (\eta, \xi) = \phi_H^t(y, x)$ is symplectic for any H and any t; see Corollary 1.8, [6].

A classical way of producing symplectic transformations is by means of *generating functions*. For example, if $g(\eta, x)$ is a smooth function of $2d$ variables with

$$\det \frac{\partial^2 g}{\partial \eta \partial x} \neq 0 \,,$$

then, by the Implicit Function Theorem (IFT; see [36] or Sect. "A The Classical Implicit Function Theorem" below), the map $\phi\colon (y, x) \to (\eta, \xi)$ defined implicitly by the relations

$$y = \frac{\partial g}{\partial x} \,, \qquad \xi = \frac{\partial g}{\partial \eta} \,,$$

yields a local symplectic diffeomorphism; in such a case, g is called the generating function of the transformation ϕ. For example, the function $\eta \cdot x$ is the generating function of the identity map.

For general information about symplectic changes of coordinates, generating functions and, in general, about symplectic structures we refer the reader to [5] or [6].

(d) A solution $z(t) = (y(t), x(t))$ of (2) is a *maximal quasi-periodic solution* with frequency vector $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$ if ω is a rationally-independent vector, i. e.,

$$\exists n \in \mathbb{Z}^d \ \text{ s.t. } \ \omega \cdot n := \sum_{i=1}^{d} \omega_i n_i = 0 \quad \Longrightarrow \quad n = 0 \,, \tag{5}$$

and if there exist smooth (periodic) functions $v, u\colon \mathbb{T}^d \to \mathbb{R}^d$ such that[8]

$$\begin{cases} y(t) = v(\omega t) \\ x(t) = \omega t + u(\omega t) \,. \end{cases} \tag{6}$$

(e) Let ω, u and v be as in the preceding item and let U and ϕ denote, respectively, the maps

$$\begin{cases} U\colon \theta \in \mathbb{T}^d \to U(\theta) := \theta + u(\theta) \in \mathbb{T}^d \\ \phi\colon \theta \in \mathbb{T}^d \to \phi(\theta) := (v(\theta), U(\theta)) \in \mathcal{M} \end{cases}$$

If U is a smooth diffeomorphism of $\mathbb{T}^d$ (so that, in particular[9] $\det U_\theta \neq 0$) then ϕ is an embedding of $\mathbb{T}^d$ into $\mathcal{M}$ and the set

$$\mathcal{T}_\omega = \mathcal{T}_\omega^d := \phi(\mathbb{T}^d) \tag{7}$$

is an embedded d-torus invariant for ϕ_H^t and on which the motion is conjugated to the linear (Kronecker) flow $\theta \to \theta + \omega t$, i. e.,

$$\phi^{-1} \circ \phi_H^t \circ \phi(\theta) = \theta + \omega t \,, \quad \forall \theta \in \mathbb{T}^d \,. \tag{8}$$

Furthermore, the invariant torus $\mathcal{T}_\omega$ is a graph over $\mathbb{T}^d$ and is *Lagrangian*, i. e., ($\mathcal{T}_\omega$ has dimension d and) the restriction of the symplectic form $\mathrm{d}y \wedge \mathrm{d}x$ on $\mathcal{T}_\omega$ vanishes[10].

(f) In KAM theory a major role is played by the numerical properties of the frequencies ω. A typical assumption is that ω is a (homogeneously) *Diophantine vector*: $\omega \in \mathbb{R}^d$ is called Diophantine or (κ, τ)-Diophantine if, for some constants $0 < \kappa \leq \min_i |\omega_i|$ and

$\tau \geq d-1$, it verifies the following inequalities:

$$|\omega \cdot n| \geq \frac{\kappa}{|n|^{\tau}}, \quad \forall n \in \mathbb{Z}^d \setminus \{0\}, \tag{9}$$

(normally, for integer vectors n, $|n|$ denotes $|n_1|+\cdots+|n_d|$, but other norms may well be used). We shall refer to κ and τ as the Diophantine constants of ω. The set of Diophantine numbers in $\mathbb{R}^d$ with constants κ and τ will be denoted by $\mathcal{D}^d_{\kappa,\tau}$; the union over all $\kappa > 0$ of $\mathcal{D}^d_{\kappa,\tau}$ will be denoted by $\mathcal{D}^d_{\tau}$ and the union over all $\tau \geq d-1$ of $\mathcal{D}^d_{\tau}$ will be denoted by $\mathcal{D}^d$. Basic facts about these sets are[11]: if $\tau < d-1$ then $\mathcal{D}^d_{\tau} = \emptyset$; if $\tau > d-1$ then the Lebesgue measure of $\mathbb{R}^d \setminus \mathcal{D}^d_{\tau}$ is zero; if $\tau = d-1$, the Lebesgue measure of $\mathcal{D}^d_{\tau}$ is zero but its intersection with any open set has the cardinality of $\mathbb{R}$.

(g) The tori $\mathcal{T}_\omega$ defined in (e) with $\omega \in \mathcal{D}^d$ will be called *maximal KAM tori* for H.

(h) A Hamiltonian function $(\eta,\xi) \in \mathcal{M} \to H(\eta,\xi)$ having a maximal KAM torus (or, more generally, a maximal invariant torus as in (e) with ω rationally independent) $\mathcal{T}_\omega$, can be put into the form[12]

$$\begin{aligned} &K(y,x) := E + \omega \cdot y + Q(y,x), \\ &\text{with } \partial_y^\alpha Q(0,x) = 0, \ \forall \alpha \in \mathbb{N}^d, \ |\alpha| \leq 1; \end{aligned} \tag{10}$$

compare, e. g., Sect. 1 of [59]; in the variables (y,x), the torus $\mathcal{T}_\omega$ is simply given by $\{y=0\} \times \mathbb{T}^d$ and E is its (constant) energy. A Hamiltonian in the form (10) is said to be in *Kolmogorov normal form*.

If

$$\det\langle \partial_y^2 Q(0,\cdot)\rangle \neq 0, \tag{11}$$

(where the brackets denote an average over $\mathbb{T}^d$ and ∂_y^2 the Hessian with respect to the y-variables) we shall say that the Kolmogorov normal form K in (10) is *non-degenerate*; similarly, we shall say that the KAM torus $\mathcal{T}_\omega$ for H is non-degenerate if H can be put in a non-degenerate Kolmogorov normal form.

Remark 1

(i) A classical theorem by H. Weyl says that the flow

$$\theta \in \mathbb{T}^d \to \theta + \omega t \in \mathbb{T}^d, \quad t \in \mathbb{R}$$

is dense (ergodic) in $\mathbb{T}^d$ if and only if $\omega \in \mathbb{R}^d$ is rationally independent (compare [6], Theorem 5.4 or Sect. 1.4 of [33]). Thus, trajectories on KAM tori fill them densely (i. e., pass in any neighborhood of any point).

(ii) In view of the preceding remark, it is easy to see that if ω is rationally independent, $(y(t),x(t))$ in (6) is a solution of (2) if and only if the functions v and u satisfy the following quasi-linear system of PDEs on $\mathbb{T}^d$:

$$\begin{cases} D_\omega v = -H_x(v(\theta), \theta + u(\theta)) \\ \omega + D_\omega u = H_y(v(\theta), \theta + u(\theta)) \end{cases} \tag{12}$$

where D_ω denotes the directional derivative $\omega \cdot \partial_\theta = \sum_{i=1}^d \omega_i \frac{\partial}{\partial \theta_i}$.

(iii) Probably, the main motivation for studying quasi-periodic solutions of Hamiltonian systems on $\mathbb{R}^d \times \mathbb{T}^d$ comes from perturbation theory of *nearly-integrable* Hamiltonian systems: a completely integrable system may be described by a Hamiltonian system on $\mathcal{M} := B(y_0,r) \times \mathbb{T}^d \subset \mathbb{R}^d \times \mathbb{T}^d$ with Hamiltonian $H = K(y)$ (compare Theorem 5.8, [6]); here $B(y_0,r)$ denotes the open ball $\{y \in \mathbb{R}^d : |y-y_0| < r\}$ centered at $y_0 \in \mathbb{R}^d$. In such a case the Hamiltonian flow is simply

$$\begin{aligned} &\phi_K^t(y,x) = \big(y, x + \omega(y)t\big), \\ &\qquad \omega(y) := K_y(y) := \frac{\partial K}{\partial y}(y). \end{aligned} \tag{13}$$

Thus, if the "frequency map" $y \in B \to \omega(y)$ is a diffeomorphism (which is guaranteed if $\det K_{yy}(y_0) \neq 0$, for some $y_0 \in B$ and B is small enough), in view of (f), for almost all initial data, the trajectories (13) belong to maximal KAM tori $\{y\} \times \mathbb{T}^d$ with $\omega(y) \in \mathcal{D}^d$.

The main content of (classical) KAM theory, in our language, is that, *if the frequency map $\omega = K_y$ of a (real-analytic) integrable Hamiltonian $K(y)$ is a diffeomorphism, KAM tori persist under small (smooth enough) perturbations of K*; compare Remark 7–(iv) below.

The study of the dynamics generated by the flow of a one-parameter family of Hamiltonians of the form

$$K(y) + \varepsilon P(y,x;\varepsilon), \quad 0 < \varepsilon \ll 1, \tag{14}$$

was called by H. Poincaré *le problème général de la dynamique*, to which he dedicated a large part of his monumental *Méthodes Nouvelles de la Mécanique Céleste* [49].

(iv) A big chapter in KAM theory, strongly motivated by applications to PDEs with Hamiltonian structure (such as nonlinear wave equation, Schrödinger equation, KdV, etc.), is concerned with quasi-periodic solutions with $1 \le n < d$ frequencies, i. e., solutions of (2) of the form

$$\begin{cases} y(t) = v(\omega t) \\ x(t) = U(\omega t) , \end{cases} \tag{15}$$

where $v \colon \mathbb{T}^n \to \mathbb{R}^d$, $U \colon \mathbb{T}^n \to \mathbb{T}^d$ are smooth functions, $\omega \in \mathbb{R}^n$ is a rationally independent n-vector. Also in this case, if the map U is a diffeomorphism onto its image, the set

$$\mathcal{T}^n_\omega := \{(y,x) \in \mathcal{M} \colon y = v(\theta) , \ x = U(\theta) , \ \theta \in \mathbb{T}^n\} \tag{16}$$

defines an invariant n-torus on which the flow ϕ^t_H acts by the linear translation $\theta \to \theta + \omega t$. Such tori are normally referred to as *lower dimensional tori*.

Even though this article will be mainly focused on "classical KAM theory" and on maximal KAM tori, we will briefly discuss lower dimensional tori in Sect. "Future Directions".

Kolmogorov Theorem

In the 1954 International Congress of Mathematicians, in Amsterdam, A.N. Kolmogorov announced the following fundamental (for the terminology, see (f), (g) and (h) above).

Theorem 1 (Kolmogorov [35]) *Consider a one-parameter family of real-analytic Hamiltonian functions on $\mathcal{M} := B(0,r) \times \mathbb{T}^d$ given by*

$$H := K + \varepsilon P \quad (\varepsilon \in \mathbb{R}) , \tag{17}$$

where: (i) K is a non-degenerate Kolmogorov normal form (10)–(11); (ii) $\omega \in \mathcal{D}^d$ is Diophantine. Then, there exists $\varepsilon_0 > 0$ and for any $|\varepsilon| \le \varepsilon_0$ a real-analytic symplectic transformation $\phi_* \colon \mathcal{M}_* := B(0,r_*) \times \mathbb{T}^d \to \mathcal{M}$, for some $0 < r_* < r$, putting H in non-degenerate Kolmogorov normal form, $H \circ \phi_* = K_*$, with $K_* := E_* + \omega \cdot y' + Q_*(y', x')$. Furthermore[13]*, $\|\phi_* - \mathrm{id}\|_{C^1(\mathcal{M}_*)}$, $|E_* - E|$, and $\|Q_* - Q\|_{C^1(\mathcal{M}_*)}$ are small with ε.*

Remark 2

(i) From Theorem 1 it follows that the torus

$$\mathcal{T}_{\omega,\varepsilon} := \phi_*(0, \mathbb{T}^d)$$

is a maximal non-degenerate KAM torus for H 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 $\mathcal{T}_{\omega,0} := \{0\} \times \mathbb{T}^d$, while the energy of $\mathcal{T}_{\omega,\varepsilon}$, namely E_*, is in general different from the energy E of $\mathcal{T}_{\omega,0}$. The idea of keeping the frequency fixed is a key idea introduced by Kolmogorov and its importance will be made clear in the analysis of the proof.

(ii) In fact, *the dependence upon ε is analytic* and therefore the torus $\mathcal{T}_{\omega,\varepsilon}$ is an analytic deformation of the unperturbed torus $\mathcal{T}_{\omega,0}$ (which is invariant for K); see Remark 7–(iii) below.

(iii) Actually, Kolmogorov not only stated the above result but gave also a precise outline of its proof, which is based on a *fast convergent "Newton" scheme*, as we shall see below; compare also [17].

The map ϕ_* is obtained as

$$\phi_* = \lim_{j \to \infty} \phi_1 \circ \cdots \circ \phi_j ,$$

where the ϕ_j's are (ε-dependent) symplectic transformations of $\mathcal{M}$ closer and closer to the identity. It is 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.

We proceed to analyze the scheme of Kolmogorov's proof, which will be divided into three main steps.

Step 1: Kolmogorov Transformation

The map ϕ_1 is close to the identity and is generated by

$$g(y',x) := y' \cdot x + \varepsilon \left(b \cdot x + s(x) + y' \cdot a(x) \right)$$

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

$$\begin{aligned} \beta_0 &= \beta_0(x) := b + s_x , \\ A &= A(x) := a_x \quad \text{and} \\ \beta &= \beta(y',x) := \beta_0 + Ay' , \end{aligned} \tag{18}$$

($s_x = \partial_x s = (s_{x_1}, \ldots, s_{x_d})$ and a_x denotes the matrix $(a_x)_{ij} := \frac{\partial a_i}{\partial x_j}$) ϕ_1 is implicitly defined by

$$\begin{cases} y = y' + \varepsilon \beta(y',x) := y' + \varepsilon(\beta_0(x) + A(x)y') \\ x' = x + \varepsilon a(x) . \end{cases} \tag{19}$$

Thus, 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

$$x = \varphi(x') := x' + \varepsilon \alpha(x'; \varepsilon) , \tag{20}$$

for a suitable real-analytic function α, and ϕ_1 is explicitly given by

$$\phi_1 \colon (y', x') \to \begin{cases} y = y' + \varepsilon\beta\left(y', \varphi(x')\right) \\ x = \varphi(x') \,. \end{cases} \tag{21}$$

Remark 3

(i) Kolmogorov transformation ϕ_1 is actually the composition of two "elementary" symplectic transformations: $\phi_1 = \phi_1^{(1)} \circ \phi_1^{(2)}$ where $\phi_1^{(2)} \colon (y', x') \to (\eta, \xi)$ is the symplectic lift of the $\mathbb{T}^d$-diffeomorphism given by $x' = \xi + \varepsilon a(\xi)$ (i. e., $\phi_1^{(2)}$ is the symplectic map generated by $y' \cdot \xi + \varepsilon y' \cdot a(\xi)$), while $\phi_1^{(1)} \colon (\eta, \xi) \to (y, x)$ is the angle-dependent action translation generated by $\eta \cdot x + \varepsilon(b \cdot x + s(x))$; $\phi_1^{(2)}$ acts in the "angle direction" and will be needed to straighten out the flow up to order $O(\varepsilon^2)$, while $\phi_1^{(1)}$ acts in the "action direction" and will be needed to keep the frequency of the torus fixed.

(ii) The inverse of ϕ_1 has the form

$$(y, x) \to \begin{cases} y' = M(x) y + c(x) \\ x' = \phi(x) \end{cases} \tag{22}$$

with M a $(d \times d)$-invertible matrix and ϕ a diffeomorphism of $\mathbb{T}^d$ (in the present case $M = (\mathbb{1}_d + \varepsilon A(x))^{-1} = \mathbb{1}_d + O(\varepsilon)$ and $\phi = \mathrm{id} + \varepsilon a$) and it is easy to see that the symplectic diffeomorphisms of the form (22) form a subgroup of the symplectic diffeomorphisms, which we shall call *the group of Kolmogorov transformations*.

Determination of Kolmogorov transformation Following Kolmogorov, we now try to determine b, s and a so that the "new Hamiltonian" (better: "the Hamiltonian in the new symplectic variables") takes the form

$$H_1 := H \circ \phi_1 = K_1 + \varepsilon^2 P_1 \,, \tag{23}$$

with K_1 in the Kolmogorov normal form

$$K_1 = E_1 + \omega \cdot y' + Q_1(y', x') \,, \qquad Q_1 = O(|y'|^2) \,. \tag{24}$$

To proceed we insert $y = y' + \varepsilon\beta(y', x)$ into H and, after some elementary algebra and using Taylor formula, we find[14]

$$H(y' + \varepsilon\beta, x) = E + \omega \cdot y' + Q(y', x) + \varepsilon Q'(y', x) + \varepsilon F'(y', x) + \varepsilon^2 P'(y', x) \tag{25}$$

where, defining

$$\begin{cases} Q^{(1)} := Q_y(y', x) \cdot (a_x y') \\ Q^{(2)} := [Q_y(y', x) - Q_{yy}(0, x) y'] \cdot \beta_0 \\ \qquad = \int_0^1 (1 - t) Q_{yyy}(t y', x) y' \cdot y' \cdot \beta_0 \mathrm{d}t \\ Q^{(3)} := P(y', x) - P(0, x) - P_y(0, x) y' \\ \qquad = \int_0^1 (1 - t) P_{yy}(t y', x) y' \cdot y' \mathrm{d}t \\ P^{(1)} := \dfrac{1}{\varepsilon^2}[Q(y' + \varepsilon\beta, x) - Q(y', x) \\ \qquad - \varepsilon Q_y(y', x) \cdot \beta] \\ \qquad = \int_0^1 (1 - t) Q_{yy}(y' + t\varepsilon\beta, x) \beta \cdot \beta \mathrm{d}t \\ P^{(2)} := \dfrac{1}{\varepsilon}[P(y' + \varepsilon\beta, x) - P(y', x)] \\ \qquad = \int_0^1 P_y(y' + t\varepsilon\beta, x) \cdot \beta \mathrm{d}t \,, \end{cases} \tag{26}$$

(recall that $Q_y(0, x) = 0$) and denoting the ω-directional derivative

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

one sees that $Q' = Q'(y', x)$, $F' = F'(y', x)$ and $P' = P'(y', x)$ are given by, respectively

$$\begin{cases} Q'(y', x) := Q^{(1)} + Q^{(2)} + Q^{(3)} = O(|y'|^2) \\ F'(y', x) := \omega \cdot b + D_\omega s + P(0, x) \\ \qquad + \left\{ D_\omega a + Q_{yy}(0, x) \beta_0 + P_y(0, x) \right\} \cdot y' \\ P' := P^{(1)} + P^{(2)} \,, \end{cases} \tag{27}$$

where $D_\omega a$ is the vector function with kth entry $\sum_{j=1}^d \omega_j \frac{\partial a_k}{\partial x_j}$; $D_\omega a \cdot y' = \omega \cdot (a_x y') = \sum_{j,k=1}^d \omega_j \frac{\partial a_k}{\partial x_j} y'_k$; recall, also, that $Q = O(|y|^2)$ so that $Q_y = O(y)$ and $Q' = O(|y'|^2)$.

Notice that, as an intermediate step, we are considering H as a function of mixed variables y' and x (and this causes no problem, as it will be clear along the proof).

Thus, recalling that x is related to x' by the (y'-independent) diffeomorphism $x = x' + \varepsilon\alpha(x'; \varepsilon)$ in (21), we see that in order to achieve relations (23)–(24), we have to determine b, s and a so that

$$F'(y', x) = \text{const} \,. \tag{28}$$

Remark 4

(i) F' is a first degree polynomial in y' so that (28) is equivalent to

$$\begin{cases} \omega \cdot b + D_\omega s + P(0,x) = \text{const}\,, \\ D_\omega a + Q_{yy}(0,x)\beta_0 + P_y(0,x) = 0\,. \end{cases} \tag{29}$$

Indeed, the second equation is necessary to keep the torus frequency fixed and equal to ω (which, as we shall see in more detail later, is a key ingredient introduced by Kolmogorov).

(ii) In solving (28) or (29), we shall encounter differential equations of the form

$$D_\omega u = f\,, \tag{30}$$

for some given function f real-analytic on $\mathbb{T}^d$. Taking the average over $\mathbb{T}^d$ shows that $\langle f \rangle = 0$, and we see that (30) can be solved only if f has vanishing mean value

$$\langle f \rangle = f_0 = 0\,;$$

in such a case, expanding in Fourier series[15], one sees that (30) is equivalent to

$$\sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} i\omega\cdot n u_n e^{in\cdot x} = \sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} f_n e^{in\cdot x}\,, \tag{31}$$

so that the solutions of (30) are given by

$$u = u_0 + \sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} \frac{f_n}{i\omega\cdot n} e^{in\cdot x}\,, \tag{32}$$

for an arbitrary u_0. Recall that for a continuous function$f$ over $\mathbb{T}^d$ to be analytic it is necessary and sufficient that its Fourier coefficients f_n decay exponentially fast in n, i. e., that there exist positive constants M and ξ such that

$$|f_n| \le M e^{-\xi|n|}\,, \quad \forall n\,. \tag{33}$$

Now, since $\omega \in \mathcal{D}^d_{\kappa,\tau}$ one has that (for $n \neq 0$)

$$\frac{1}{|\omega\cdot n|} \le \frac{|n|^\tau}{\kappa} \tag{34}$$

and one sees that if f is analytic so is u in (32) (although the decay constants of u will be different from those of f; see below)

Summarizing, *if f is real-analytic on $\mathbb{T}^d$ and has vanishing mean value f_0, then there exists a unique real-analytic solution of (30) with vanishing mean value, which is given by*

$$D_\omega^{-1} f := \sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} \frac{f_n}{i\omega\cdot n} e^{in\cdot x}\,; \tag{35}$$

all other solutions of (30) are obtained by adding an arbitrary constant to $D_\omega^{-1} f$ as in (32) with u_0 arbitrary.

Taking the average of the first relation in (29), we may the determine the value of the constant denoted const, namely,

$$\text{const} = \omega\cdot b + P_0(0) := \omega\cdot b + \langle P(0,\cdot)\rangle\,. \tag{36}$$

Thus, by (ii) of Remark 4, we see see that

$$s = -D_\omega^{-1}\left(P(0,x) - P_0(0)\right) = -\sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} \frac{P_n(0)}{i\omega\cdot n} e^{in\cdot x}\,, \tag{37}$$

where $P_n(0)$ denote the Fourier coefficients of $x \to P(0,x)$; indeed s is determined only up to a constant by the relation in (29) but *we select the zero-average solution*. Thus, s has been completely determined.

To solve the second (vector) equation in (29) we first have to require that the left hand side (l.h.s.) has vanishing mean value, i. e., recalling that $\beta_0 = b + s_x$ (see (18)), we must have

$$\langle Q_{yy}(0,\cdot)\rangle b + \langle Q_{yy}(0,\cdot)s_x\rangle + \langle P_y(0,\cdot)\rangle = 0\,. \tag{38}$$

In view of (11) this relation is equivalent to

$$b = -\langle Q_{yy}(0,\cdot)\rangle^{-1}\left(\langle Q_{yy}(0,\cdot)s_x\rangle + \langle P_y(0,\cdot)\rangle\right)\,, \tag{39}$$

which uniquely determines b. Thus β_0 is completely determined and the l.h.s. of the second equation in (29) has average zero; thus its unique zero-average solution (again zero-average of a is required as a normalization condition) is given by

$$a = -D_\omega^{-1}\left(Q_{yy}(0,x)\beta_0 + P_y(0,x)\right)\,. \tag{40}$$

Finally, if $\varphi(x') = x' + \varepsilon\alpha(x';\varepsilon)$ is the inverse diffeomorphism of $x \to x + \varepsilon a(x)$ (compare (20)), then, by Taylor's formula,

$$Q(y',\varphi(x')) = Q(y',x') + \varepsilon\int_0^1 Q_x(y',x'+\varepsilon\alpha t)\cdot\alpha\,dt\,.$$

In conclusion, we have

Proposition 1 *If ϕ_1 is defined in (19)–(18) with s, b and a given in (37), (39) and (40) respectively, then (23) holds with*

$$\begin{cases} E_1 := E + \varepsilon\widetilde{E} \\ \qquad \widetilde{E} := \omega \cdot b + P_0(0) \\ Q_1(y', x') := Q(y', x') + \varepsilon\widetilde{Q}(y', x') \\ \qquad \widetilde{Q} := \int_0^1 Q_x(y', x' + t\varepsilon\alpha) \cdot \alpha dt + Q'(y', \varphi(x')) \\ P_1(y', x') := P'(y', \varphi(x')) \end{cases} \tag{41}$$

with Q' and P' defined in (26), (27) and φ in (20).

Remark 5 The main technical problem is now transparent: because of the appearance of the *small divisors* $\omega \cdot n$ (which may become arbitrarily small), the solution $D_\omega^{-1} f$ is *less regular* than f so that the approximation scheme cannot work on a fixed function space. To overcome this fundamental problem – which even Poincaré was unable to solve notwithstanding his enormous efforts (see, e. g., [49]) – three ingredients are necessary:

(i) To set up a Newton scheme: this step has just been performed and it has been summarized in the above Proposition 1; such schemes have the following features: they are "quadratic" and, furthermore, after one step one has reproduced the initial situation (i. e., the form of H_1 in (23) has the same properties of H_0). It is important to notice that the new perturbation $\varepsilon^2 P_1$ is proportional to the *square* ε; thus, if one could iterate *j* times, at the *j*th step, would find

$$H_j = H_{j-1} \circ \phi_j = K_j + \varepsilon^{2^j} P_j \,. \tag{42}$$

The appearance of the exponential of the exponential of ε justifies the term "super-convergence" used, sometimes, in connection with Newton schemes.

(ii) One needs to introduce a *scale of Banach function spaces* $\{\mathcal{B}_\xi : \xi > 0\}$ with the property that $\mathcal{B}_{\xi'} \subset \mathcal{B}_\xi$ when $\xi < \xi'$: the generating functions ϕ_j will belong to $\mathcal{B}_{\xi_j}$ for a suitable decreasing sequence ξ_j;

(iii) One needs to control the small divisors at each step and this is granted by Kolmogorov's idea of keeping the frequency *fixed* in the normal form so that one can systematically use the Diophantine estimate (9).

Kolmogorov in his paper very neatly explained steps (i) and (iii) but did not provide the details for step (ii); in this regard he added: "*Only the use of condition* (9) *for proving the convergence of the recursions, ϕ_j, to the analytic limit for the recursion ϕ_* is somewhat more subtle*". In the next paragraph we shall introduce classical Banach spaces and discuss the needed straightforward estimates.

Step 2: Estimates

For $\xi \leq 1$, we denote by $\mathcal{B}_\xi$ the space of function $f : B(0, \xi) \times \mathbb{T}^d \to \mathbb{R}$ analytic on

$$W_\xi := D(0, \xi) \times \mathbb{T}^d_\xi \,, \tag{43}$$

where

$$\begin{aligned} D(0, \xi) &:= \{y \in \mathbb{C}^d : |y| < \xi\} \quad \text{and} \\ \mathbb{T}^d_\xi &:= \{x \in \mathbb{C}^d : |\mathrm{Im} x_j| < \xi\}/(2\pi\mathbb{Z}^d) \end{aligned} \tag{44}$$

with finite sup-norm

$$\|f\|_\xi := \sup_{D(0,\xi)\times\mathbb{T}^d_\xi} |f| \,, \tag{45}$$

(in other words, $\mathbb{T}^d_\xi$ denotes the complex points x with real parts $\mathrm{Re} x_j$ defined modulus 2π and imaginary part $\mathrm{Im} x_j$ with absolute value less than ξ).

The following properties are elementary:

(P1) $\mathcal{B}_\xi$ equipped with the $\|\cdot\|_\xi$ norm is a Banach space;

(P2) $\mathcal{B}_{\xi'} \subset \mathcal{B}_\xi$ when $\xi < \xi'$ and $\|f\|_\xi \leq \|f\|_{\xi'}$ for any $f \in \mathcal{B}_{\xi'}$;

(P3) if $f \in \mathcal{B}_\xi$, and $f_n(y)$ denotes the n-Fourier coefficient of the periodic function $x \to f(y, x)$, then

$$|f_n(y)| \leq \|f\|_\xi e^{-|n|\xi} \,, \quad \forall n \in \mathbb{Z}^d, \ \forall y \in D(0, \xi) \,. \tag{46}$$

Another elementary property, which together with **(P3)** may found in any book of complex variables (e. g., [1]), is the following "Cauchy estimate" (which is based on Cauchy's integral formula):

(P4) let $f \in \mathcal{B}_\xi$ and let $p \in \mathbb{N}$ then there exists a constant $B_p = B_p(d) \geq 1$ such that, for any multi-index $(\alpha, \beta) \in \mathbb{N}^d \times \mathbb{N}^d$ with $|\alpha| + |\beta| \leq p$ (as above for integer vectors α, $|\alpha| = \sum_j |\alpha_j|$) and for any $0 \leq \xi' < \xi$ one has

$$\|\partial_y^\alpha \partial_x^\beta f\|_{\xi'} \leq B_p \|f\|_\xi (\xi - \xi')^{-(|\alpha|+|\beta|)} \,. \tag{47}$$

Finally, we shall need estimates on $D_\omega^{-1} f$, i. e., on solutions of (30):

(P5) Assume that $x \to f(x) \in \mathcal{B}_\xi$ has a zero average (all above definitions may be easily adapted to functions depending only on x); assume that $\omega \in \mathcal{D}^d_{\kappa,\tau}$ (recall Sect. "Introduction", point (f)), and let $p \in \mathbb{N}$. Then, there exist constants $\bar{B}_p = \bar{B}_p(d, \tau) \geq 1$ and $k_p = k_p(d, \tau) \geq 1$ such that, for any multi-index $\beta \in \mathbb{N}^d$ with $|\beta| \leq p$ and for any $0 \leq \xi' < \xi$ one has

$$\|\partial_x^\beta D_\omega^{-1} f\|_{\xi'} \leq \bar{B}_p \frac{\|f\|_\xi}{\kappa} (\xi - \xi')^{-k_p} \,. \tag{48}$$

Remark 6

(i) A proof of (48) is easily obtained observing that by (35) and (46), calling $\delta := \xi - \xi'$, one has

$$\begin{aligned}\|\partial_x^\beta D_\omega^{-1} f\|_{\xi'} &\le \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} \frac{|n|^{|\beta|}|f_n|}{|\omega\cdot n|}\mathrm{e}^{\xi'|n|}\\ &\le \|f\|_\xi \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} \frac{|n|^{|\beta|+\tau}}{\kappa}\mathrm{e}^{-\delta|n|}\\ &= \frac{\|f\|_\xi}{\kappa}\delta^{-(|\beta|+\tau+d)}\\ &\quad\cdot \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} [\delta|n|]^{|\beta|+\tau}\mathrm{e}^{-\delta|n|}\delta^d\\ &\le \mathrm{const}\frac{\|f\|_\xi}{\kappa}(\xi-\xi')^{-(|\beta|+\tau+d)}\,,\end{aligned}$$

where the last estimate comes from approximating the sum with the Riemann integral

$$\int_{\mathbb{R}^d} |y|^{|\beta|+\tau}\mathrm{e}^{-|y|}\mathrm{d}y\,.$$

More surprising (and much more subtle) is that (48) holds with $k_p = |\beta| + \tau$; such an estimate has been obtained by Rüssmann [54,55]. For other explicit estimates see, e. g., [11] or [12].

(ii) If $|\beta| > 0$ it is not necessary to assume that $\langle f\rangle = 0$.

(iii) Other norms may be used (and, sometimes, are more useful); for example, rather popular are Fourier norms

$$\|f\|'_\xi := \sum_{n\in\mathbb{Z}^d} |f_n|\mathrm{e}^{\xi|n|}\,; \tag{49}$$

see, e. g., [13] and references therein.

By the hypotheses of Theorem 1 it follows that there exist $0 < \xi \le 1$, $\kappa > 0$ and $\tau \ge d-1$ such that $H \in \mathcal{B}_\xi$ and $\omega \in \mathcal{D}^d_{\kappa,\tau}$. Denote

$$T := \langle Q_{yy}(0,\cdot)\rangle^{-1}\,,\quad M := \|P\|_\xi\,. \tag{50}$$

and let $C > 1$ be a constant such that[16]

$$|E|, |\omega|, \|Q\|_\xi, \|T\| < C \tag{51}$$

(i. e., each term on the l.h.s. is bounded by the r.h.s.); finally, fix

$$0<\delta<\xi \quad \text{and define } \bar\xi := \xi - \frac{2}{3}\delta\,,\ \xi' := \xi - \delta\,. \tag{52}$$

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 an intermediate domain where we shall bound various functions of y' and x.

*By **(P4)** and **(P5)**, it follows that there exist constants $\bar c = \bar c(d,\tau,\kappa) > 1$, $\bar\mu \in \mathbb{Z}_+$ and $\bar\nu = \bar\nu(d,\tau) > 1$ such that*[17]

$$\begin{cases} \|s_x\|_{\bar\xi}, |b|, |\widetilde{E}|, \|a\|_{\bar\xi}, \|a_x\|_{\bar\xi}, \|\beta_0\|_{\bar\xi}, \|\beta\|_{\bar\xi},\\ \qquad \|Q'\|_{\bar\xi}, \|\partial^2_{y'}Q'(0,\cdot)\|_0 \le \bar c C^{\bar\mu}\delta^{-\bar\nu}M =: \bar L\,,\\ \|P'\|_{\bar\xi} \le \bar c C^{\bar\mu}\delta^{-\bar\nu}M^2 = \bar L M\,. \end{cases} \tag{53}$$

The estimate in the first line of (53) allows us to construct, for ε small enough, the symplectic transformation ϕ_1, whose main properties are collected in the following

Lemma 1 *If $|\varepsilon| \le \varepsilon_0$ and ε_0 satisfies*

$$\varepsilon_0 \bar L \le \frac{\delta}{3}\,, \tag{54}$$

then the map $\psi_\varepsilon(x) := x + \varepsilon a(x)$ has an analytic inverse $\varphi(x') = x' + \varepsilon\alpha(x';\varepsilon)$ such that, for all $|\varepsilon| < \varepsilon_0$,

$$\|\alpha\|_{\xi'} \le \bar L \quad \textit{and} \quad \varphi = \mathrm{id} + \varepsilon\alpha\colon \mathbb{T}^d_{\xi'} \to \mathbb{T}^d_{\bar\xi}\,. \tag{55}$$

Furthermore, for any $(y', x) \in W_{\bar\xi}$, $|y' + \varepsilon\beta(y', x)| < \xi$, so that

$$\begin{aligned}&\phi_1 = \big(y' + \varepsilon\beta(y',\varphi(x')), \varphi(x')\big) : W_{\xi'} \to W_\xi\,, \quad \textit{and}\\ &\|\phi_1 - \mathrm{id}\|_{\xi'} \le |\varepsilon|\bar L\,;\end{aligned} \tag{56}$$

finally, the matrix $\mathbb{1}_d + \varepsilon a_x$ is, for any $x \in \mathbb{T}^d_{\bar\xi}$, invertible with inverse $\mathbb{1}_d + \varepsilon S(x;\varepsilon)$ satisfying

$$\|S\|_{\bar\xi} \le \frac{\|a_x\|_{\bar\xi}}{1 - |\varepsilon|\|a_x\|_{\bar\xi}} < \frac{3}{2}\bar L\,, \tag{57}$$

so that ϕ_1 defines a symplectic diffeomorphism.

The simple proof[18] of this statement is based upon standard tools in mathematical analysis such as the contraction mapping theorem or the inversion of close-to-identity matrices by Neumann series (see, e. g., [36]).

From the Lemma and the definition of P_1 in (41), it follows immediately that

$$\|P_1\|_{\xi'} \le \bar L\,. \tag{58}$$

Next, by the same technique used to derive (53), one can easily check that

$$\|\widetilde{Q}\|_{\xi'}\,,\quad 2C^2\|\partial^2_{y'}\widetilde{Q}(0,\cdot)\|_0 \le cC^\mu\delta^{-\nu}M = L\,, \tag{59}$$

for suitable constants $c \geq \bar{c}, \bar{\mu} \geq \mu, \bar{\nu} \geq \nu$ (the factor $2C^2$ has been introduced for later convenience; notice also that $L \geq \bar{L}$). But, then, if

$$\varepsilon_0 L := \varepsilon_0 c C^{\mu} \delta^{-\nu} M \leq \frac{\delta}{3}\,, \tag{60}$$

there follows that[19] $\|\widetilde{T}\| \leq L$; this bound, together with (53), (59), (56), and (58), shows that

$$\begin{cases} |\tilde{E}|, \|\widetilde{Q}\|_{\xi'}, \|\widetilde{T}\|, \|\phi_1 - \mathrm{id}\|_{\xi'} \leq L \\ \|P_1\|_{\xi'} \leq LM\,; \end{cases} \tag{61}$$

provided (60) holds (notice that (60) implies (54)).

One step of the iteration has been concluded and the needed estimates obtained. The idea is to iterate the construction infinitely many times, as we proceed to describe.

Step 3: Iteration and Convergence

In order to iterate Kolmogorov's construction, analyzed in Step 2, so as to construct a sequence of symplectic transformations

$$\phi_j\colon W_{\xi_{j+1}} \to W_{\xi_j}\,, \tag{62}$$

closer and closer to the identity, and such that (42) hold, the first thing to do is to choose the sequence ξ_j: this sequence must be convergent, so that $\delta_j = \xi_j - \xi_{j+1}$ has to go to zero rather quickly. Inverse powers of δ_j (which, at the jth step will play the role of δ in the previous paragraph) appear in the smallness conditions (see, e. g., (54)): this "divergence" will, however, be beaten by the super-fast decay of ε^{2^j}.

Fix $0 < \xi_* < \xi$ (ξ_* will be the domain of analyticity of ϕ_* and K_* in Theorem 1 and, for $j \geq 0$, let

$$\begin{cases} \xi_0 := \xi \\ \delta_0 := \dfrac{\xi - \xi_*}{2} \end{cases} \qquad \begin{cases} \delta_j := \dfrac{\delta_0}{2^j} \\ \xi_{j+1} := \xi_j - \delta_j = \xi_* + \dfrac{\delta_0}{2^j} \end{cases} \tag{63}$$

and observe that $\xi_j \downarrow \xi_*$. With this choice[20], Kolmogorov's algorithm can be iterated infinitely many times, provided ε_0 is small enough. To be more precise, let c, μ and ν be as in (59), and define

$$C := 2 \max\left\{|E|, |\omega|, \|Q\|_{\xi}, \|T\|, 1\right\}\,. \tag{64}$$

Smallness Assumption: *Assume that $|\varepsilon| \leq \varepsilon_0$ and that ε_0 satisfies*

$$\varepsilon_0 D B \|P\|_{\xi} \leq 1 \quad \text{where } D := 3c\delta_0^{-(\nu+1)} C^{\mu}\,,\quad B := 2^{\nu+1}\,; \tag{65}$$

notice that the constant C in (64) satisfies (51) and that (65) implies (54). Then the following claim holds.

Claim C: *Under condition (65) one can iteratively construct a sequence of Kolmogorov symplectic maps ϕ_j as in (62) so that (42) holds in such a way that $\varepsilon^{2^j} P_j$, $\Phi_j := \phi_1 \circ \phi_2 \circ \cdots \circ \phi_j$, E_j, K_j, Q_j converge uniformly on W_{ξ_*} to, respectively, 0, ϕ_*, E_*, K_*, Q_*, which are real-analytic on W_{ξ_*} and $H \circ \phi_* = K_* = E_* + \omega \cdot y + Q_*$ with $Q_* = O(|y|^2)$. Furthermore, the following estimates hold for any $|\varepsilon| \leq \varepsilon_0$ and for any $i \geq 0$:*

$$|\varepsilon|^{2^i} M_i := |\varepsilon|^{2^i} \|P_i\|_{\xi_i} \leq \frac{(|\varepsilon| D B M)^{2^i}}{D B^{i+1}}\,, \tag{66}$$

$$\|\phi_* - \mathrm{id}\|_{\xi_*}, |E - E_*|, \|Q - Q_*\|_{\xi_*}, \|T - T_*\| \leq |\varepsilon| D B M\,, \tag{67}$$

where $T_* := \langle \partial_y^2 Q_*(0, \cdot)\rangle^{-1}$, showing that K_* is non-degenerate.

Remark 7

(i) From Claim **C** Kolmogorov Theorem 1 follows at once. In fact we have proven the following quantitative statement: *Let $\omega \in \mathcal{D}^d_{\kappa,\tau}$ with $\tau \geq d-1$ and $0 < \kappa < 1$; let Q and P be real-analytic on $W_\xi = D^d(0,\xi) \times \mathbb{T}^d_\xi$ for some $0 < \xi \leq 1$ and let $0 < \theta < 1$; let T and C be as in, respectively, (50) and (64). There exist $c_* = c_*(d, \tau, \kappa, \theta) > 1$ and positive integers $\sigma = \sigma(d,\tau)$, μ such that if*

$$|\varepsilon| \leq \varepsilon_* := \frac{\xi^{\sigma}}{c_* \|P\|_{\xi} C^{\mu}} \tag{68}$$

then one can construct a near-to-identity Kolmogorov transformation (Remark 3–(ii)) $\phi_*\colon W_{\theta\xi} \to W_\xi$ such that the thesis of Theorem 1 holds together with the estimates

$$\|\phi_* - \mathrm{id}\|_{\theta\xi}, |E - E_*|, \|Q - Q_*\|_{\theta\xi}, \|T - T_*\| \leq \frac{|\varepsilon|}{\varepsilon_*} = |\varepsilon| c_* \|P\|_{\xi} C^{\mu} \xi^{-\sigma}\,. \tag{69}$$

(The correspondence with the above constants being: $\xi_* = \theta\xi, \delta_0 = \xi(1-\theta)/2, \sigma = \nu + 1, D = 3c(2/(1-\theta))^{\nu+1} C^{\mu}, c_* = 3c(4/(1-\theta))^{\nu+1}$).

(ii) From Cauchy estimates and (67), it follows that $\|\phi_* - \mathrm{id}\|_{C^p}$ and $\|Q - Q_*\|_{C^p}$ are small for any p (small in $|\varepsilon|$ but not uniformly in[21] p).

(iii) All estimates are uniform in ε, therefore, from Weierstrass theorem (compare note 18) it follows that ϕ_* and K_* are *analytic in ε in the complex ball of radius* ε_0.

Power series expansions in ε were very popular in the nineteenth and twentieth centuries[22], however convergence of the formal ε-power series of quasi-periodic solutions was proved for the first time only in the 1960s thanks to KAM theory [45]. Some of this matter is briefly discussed in Sect. "Future Directions" below.

(iv) **The Nearly-Integrable Case** In [35] it is pointed out that Kolmogorov's Theorem easily yields the existence of many KAM tori for nearly-integrable systems (14) for $|\varepsilon|$ small enough, provided K is non-degenerate in the sense that

$$\det K_{yy}(y_0) \neq 0 \,. \tag{70}$$

In fact, without loss of generality we may assume that $\omega := H_0'$ is a diffeomorphism on $B(y_0, 2r)$ and $\det K_{yy}(y) \neq 0$ for all $y \in B(y_0, 2r)$. Furthermore, letting $B = B(y_0, r)$, fixing $\tau > d-1$ and denoting by ℓ_d the Lebesgue measure on $\mathbb{R}^d$, from the remark in note 11 and from the fact that ω is a diffeomorphism, there follows that there exists a constant $c_\#$ depending only on d, τ and r such that

$$\ell_d(\omega(B)\backslash \mathcal{D}^d_{\kappa,\tau}), \ell_d(\{y \in B\colon \omega(y) \notin \mathcal{D}^d_{\kappa,\tau}\}) < c_\# \kappa \,. \tag{71}$$

Now, let $B_{\kappa,\tau} := \{y \in B\colon \omega(y) \in \mathcal{D}^d_{\kappa,\tau}\}$ (which, by (71) has Lebesgue measure $\ell_d(B_{\kappa,\tau}) \geq \ell_d(B) - c_\#\kappa$), then for any $\bar{y} \in B_{\kappa,\tau}$ we can make the trivial symplectic change of variables $y \to \bar{y} + y$, $x \to x$ so that K can be written as in (10) with

$$E := K(\bar{y}) \,, \quad \omega := K_y(\bar{y}) \,,$$
$$Q(y,x) = Q(y) := K(y) - K(\bar{y}) - K_y(\bar{y}) \cdot y \,,$$

(where, for ease of notation, we did not change names to the new symplectic variables) and $P(\bar{y} + y, x)$ replacing (with a slight abuse of notation) $P(y,x)$. By Taylor's formula, $Q = O(|y|^2)$ and, furthermore (since $Q(y,x) = Q(y)$, $\langle \partial_y^2 Q(0,x)\rangle = Q_{yy}(0) = K_{yy}(\bar{y})$, which is invertible according to out hypotheses. Thus K is Kolmogorov non-degenerate and Theorem 1 can be applied yielding, for $|\varepsilon| < \varepsilon_0$, a KAM torus $\mathcal{T}_{\omega,\varepsilon}$, with $\omega = K_y(\bar{y})$, for each $\bar{y} \in B_{\kappa,\tau}$. Notice that the measure of initial phase points, which, perturbed, give rise to KAM tori, has a small complementary bounded by $c_\#\kappa$ (see (71)).

(v) In the nearly-integrable setting described in the preceding point, the union of KAM tori is usually called the **Kolmogorov set**. It is not difficult to check that the dependence upon $\bar{y}$ of the Kolmogorov transformation ϕ_* is Lipschitz[23], implying that the measure of the complementary of Kolmogorov set itself is also bounded by $\hat{c}_\#\kappa$ with a constant $\hat{c}_\#$ depending only on d, τ and r.
Indeed, the estimate on the measure of the Kolmogorov set can be made more quantitative (i. e., one can see how such an estimate depends upon ε as $\varepsilon \to 0$). In fact, revisiting the estimates discussed in **Step 2** above one sees easily that the constant c defined in (53) has the form[24]

$$c = \hat{c}\kappa^{-4} \,. \tag{72}$$

where $\hat{c} = \hat{c}(d,\tau)$ depends only on d and τ (here the Diophantine constant κ is assumed, without loss of generality, to be smaller than one). Thus the smallness condition (65) reads $\varepsilon_0\kappa^{-4}\bar{D} \leq 1$ with some constant $\bar{D}$ independent of κ: such condition is satisfied by choosing $\kappa = (\bar{D}\varepsilon_0)^{1/4}$ and since $\hat{c}_\#\kappa$ was an upper bound on the complementary of Kolmogorov set, we see that *the set of phase points which do not lie on KAM tori may be bounded by a constant times* $\sqrt[4]{\varepsilon_0}$. Actually, it turns that this bound is not optimal, as we shall see in the next section: see Remark 10.

(vi) The proof of claim **C** follows easily by induction on the number j of the iterative steps[25].

Arnold's Scheme

The first detailed proof of Kolmogorov Theorem, in the context of nearly-integrable Hamiltonian systems (compare Remark 1–(iii)), was given by V.I. Arnold in 1963.

Theorem 2 (Arnold [2]) *Consider a one-parameter family of nearly-integrable Hamiltonians*

$$H(y,x;\varepsilon) := K(y) + \varepsilon P(y,x) \quad (\varepsilon \in \mathbb{R}) \tag{73}$$

with K and P real-analytic on $\mathcal{M} := B(y_0,r) \times \mathbb{T}^d$ (endowed with the standard symplectic form $dy \wedge dx$) satisfying

$$K_y(y_0) = \omega \in \mathcal{D}^d_{\kappa,\tau} \,, \quad \det K_{yy}(y_0) \neq 0 \,. \tag{74}$$

Then, if ε is small enough, there exists a real-analytic embedding

$$\phi\colon \theta \in \mathbb{T}^d \to \mathcal{M} \tag{75}$$

close to the trivial embedding (y_0, id), such that the d-torus

$$\mathcal{T}_{\omega,\varepsilon} := \phi\left(\mathbb{T}^d\right) \tag{76}$$

is invariant for H and

$$\phi_H^t \circ \phi(\theta) = \phi(\theta + \omega t)\,, \tag{77}$$

showing that such a torus is a non-degenerate KAM torus for H.

Remark 8

(i) The above Theorem is a corollary of Kolmogorov Theorem 1 as discussed in Remark 7–(iv).
(ii) Arnold's proof of the above Theorem *is not based* upon Kolmogorov's scheme and is rather different in spirit – although still based on a Newton method – and introduces several interesting technical ideas.
(iii) Indeed, the iteration scheme of Arnold is more classical and, from the algebraic point of view, easier than Kolmogorov's, but the estimates involved are somewhat more delicate and introduce a logarithmic correction, so that, in fact, the smallness parameter will be

$$\epsilon := |\varepsilon|(\log|\varepsilon|^{-1})^{\rho} \tag{78}$$

(for some constant $\rho = \rho(d,\tau) \geq 1$) rather than $|\varepsilon|$ as in Kolmogorov's scheme; see, also, Remark 9–(iii) and (iv) below.

Arnold's Scheme

Without loss of generality, one may assume that K and P have analytic and bounded extension to $W_{r,\xi}(y_0) := D(y_0,r) \times \mathbb{T}^d_\xi$ for some $\xi > 0$, where, as above, $D(y_0,r)$ denotes the complex ball of center y_0 and radius r. We remark that, in what follows, the analyticity domains of actions and angles play a different role

The Hamiltonian H in (73) admits, for $\varepsilon = 0$ the (KAM) invariant torus $\mathcal{T}_{\omega,0} = \{y_0\} \times \mathbb{T}^d$ on which the K-flow is given by $x \to x + \omega t$. Arnold's basic idea is to find a symplectic transformation

$$\phi_1\colon W_1 := D(y_1,r_1)\times\mathbb{T}^d_{\xi_1} \to W_0 := D(y_0,r)\times\mathbb{T}^d_\xi\,, \tag{79}$$

so that $W_1 \subset W_0$ and

$$\begin{cases} H_1 := H\circ\phi_1 = K_1 + \varepsilon^2 P_1\,, \quad K_1 = K_1(y)\,, \\ \partial_y K_1(y_1) = \omega\,, \quad \det \partial_y^2 K_1(y_1) \neq 0 \end{cases} \tag{80}$$

(with abuse of notation we denote here the new symplectic variables with the same name of the original variables; as above, dependence on ε will, often, not be explicitly indicated). In this way the initial set up is reconstructed and, for ε small enough, one can iterate the scheme so as to build a sequence of symplectic transformations

$$\phi_j\colon W_j := D(y_j,r_j)\times\mathbb{T}^d_{\xi_j} \to W_{j-1} \tag{81}$$

so that

$$\begin{cases} H_j := H_{j-1}\circ\phi_j = K_j + \varepsilon^{2^j} P_j\,, \quad K_j = K_j(y)\,, \\ \partial_y K_j(y_j) = \omega\,, \quad \det \partial_y^2 K_j(y_j) \neq 0\,. \end{cases} \tag{82}$$

Arnold's transformations, as in Kolmogorov's case, are closer and closer to the identity, and the limit

$$\begin{aligned} &\phi(\theta) := \lim_{j\to\infty} \Phi_j(y_j,\theta)\,, \\ &\quad \Phi_j := \phi_1\circ\cdots\circ\phi_j\colon W_j \to W_0\,, \end{aligned} \tag{83}$$

defines a real-analytic embedding of $\mathbb{T}^d$ into the phase space $B(y_0,r)\times\mathbb{T}^d$, which is close to the trivial embedding (y_0, id); furthermore, the torus

$$\mathcal{T}_{\omega,\varepsilon} := \phi(\mathbb{T}^d) = \lim_{j\to\infty} \Phi_j(y_j,\mathbb{T}^d) \tag{84}$$

is invariant for H and (77) holds as announced in Theorem 2. Relation (77) follows from the following argument. The radius r_j will turn out to tend to 0 but in a much slower way than $\varepsilon^{2^j} P_j$. This fact, together with the rapid convergence of the symplectic transformation Φ_j in (83) implies

$$\begin{aligned} \phi_H^t\circ\phi(\theta) &= \lim_{j\to\infty} \phi_H^t\left(\Phi_j(y_j,\theta)\right) \\ &= \lim_{j\to\infty} \Phi_j\circ\phi_{H_j}^t(y_j,\theta) \\ &= \lim_{j\to\infty} \Phi_j(y_j,\theta+\omega t) \\ &= \phi(\theta+\omega t) \end{aligned} \tag{85}$$

where: the first equality is just smooth dependence upon initial data of the flow ϕ_H^t together with (83); the second equality is (3); the third equality is due to the fact that (see (82)) $\phi_{H_j}^t(y_j,\theta) = \phi_{K_j}^t(y_j,\theta) + O(\varepsilon^{2^j}\|P_j\|) = (y_j, \theta+\omega t) + O(\varepsilon^{2^j}\|P_j\|)$ and $O(\varepsilon^{2^j}\|P_j\|)$ goes very rapidly to zero; the fourth equality is again (83).

Arnold's Transformation

Let us look for a near-to-the-identity transformation ϕ_1 so that the first line of (80) holds; this transformation will be determined by a generating function of the form

$$y'\cdot x + \varepsilon g(y',x)\,, \quad \begin{cases} y = y' + \varepsilon g_x(y',x) \\ x' = x + \varepsilon g_{y'}(y',x)\,. \end{cases} \tag{86}$$

Inserting $y = y' + \varepsilon g_x(y', x)$ into H, one finds

$$H(y' + \varepsilon g_x, x) = K(y') + \varepsilon\left[K_y(y') \cdot g_x + P(y', x)\right] + \varepsilon^2\left(P^{(1)} + P^{(2)}\right) \quad (87)$$

with (compare (26))

$$\begin{aligned} P^{(1)} &:= \frac{1}{\varepsilon^2}\left[K(y' + \varepsilon g_x) - K(y') - \varepsilon K_y(y') \cdot g_x\right] \\ &= \int_0^1 (1-t) K_{yy}(y' + t\varepsilon g_x, x)\, g_x \cdot g_x \, \mathrm{d}t \\ P^{(2)} &:= \frac{1}{\varepsilon}\left[P(y' + \varepsilon g_x, x) - P(y', x)\right] \\ &= \int_0^1 P_y(y' + t\varepsilon g_x, x) \cdot g_x \, \mathrm{d}t\,. \end{aligned} \quad (88)$$

Remark 9

(i) The (naive) idea is to try determine g so that

$$K_y(y') \cdot g_x + P(y', x) = \text{function of } y' \text{ only}, \quad (89)$$

however, such a relation is impossible to achieve. First of all, by taking the x-average of both sides of (89) one sees that the "function of y' only" has to be the mean of $P(y', \cdot)$, i. e., the zero-Fourier coefficient $P_0(y')$, so that the *formal* solution of (89), is (by Fourier expansion)

$$\begin{cases} g = \sum_{n \neq 0} \dfrac{-P_n(y')}{iK_y(y') \cdot n} \mathrm{e}^{in \cdot x}\,, \\ K_y(y') \cdot g_x + P(y', x) = P_0(y')\,. \end{cases} \quad (90)$$

But, (at difference with Kolmogorov's scheme) the frequency $K_y(y')$ is *a function of the action* y' and since, by the Inverse Function Theorem (Appendix "A The Classical Implicit Function Theorem"), $y \to K_y(y)$ is a local diffeomorphism, it follows that, in any neighborhood of y_0, there are points y such that $K_y(y) \cdot n = 0$ for some[26] $n \in \mathbb{Z}^d$. Thus, in any neighborhood of y_0, some divisors in (90) will actually vanish and, therefore, *an analytic solution g cannot exist*[27].

(ii) On the other hand, since $K_y(y_0)$ is rationally independent, it is clearly possible (simply by continuity) to control a finite number of divisors in a suitable neighborhood of y_0, more precisely, for any $N \in \mathbb{N}$ one can find $\bar{r} > 0$ such that

$$K_y(y) \cdot n \neq 0\,, \quad \forall y \in D(y_0, \bar{r})\,, \quad \forall 0 < |n| \leq N\,; \quad (91)$$

the important quantitative aspects will be shortly discussed below.

(iii) Relation (89) is also one of the main "identity" in *Averaging Theory* and is related to the so-called *Hamilton–Jacobi equation*. Arnold's proof makes such a theory rigorous and shows how a Newton method can be built upon it in order to establish the existence of invariant tori. In a sense, Arnold's approach is more classical than Kolmogorov's.

(iv) When (for a given y and n) it occurs that $K_y(y) \cdot n = 0$ one speaks of an (exact) *resonance*. As mentioned at the end of point (i), in the general case, *resonances are dense*. This represents the main problem in Hamiltonian perturbation theory and is a typical feature of *conservative systems*. For generalities on Averaging Theory, Hamilton–Jacobi equation, resonances etc. see, e. g., [5] or Sects. 6.1 and 6.2 of [6].

The key (simple!) idea of Arnold is to split the perturbation into two terms

$$P = \hat{P} + \check{P} \quad \text{where} \quad \begin{cases} \hat{P} := \sum_{|n| \leq N} P_n(y) \mathrm{e}^{in \cdot x} \\ \check{P} := \sum_{|n| > N} P_n(y) \mathrm{e}^{in \cdot x} \end{cases} \quad (92)$$

choosing N so that

$$\check{P} = O(\varepsilon) \quad (93)$$

(this is possible because of the fast decay of the Fourier coefficients of P; compare (33)). Then, for $\varepsilon \neq 0$, (87) can be rewritten as follows

$$H(y' + \varepsilon g_x, x) = K(y') + \varepsilon\left[K_y(y') \cdot g_x + \hat{P}(y', x)\right] + \varepsilon^2\left(P^{(1)} + P^{(2)} + P^{(3)}\right) \quad (94)$$

with $P^{(1)}$ and $P^{(2)}$ as in (88) and

$$P^{(3)}(y', x) := \frac{1}{\varepsilon}\check{P}(y', x)\,. \quad (95)$$

Thus, letting[28]

$$g = \sum_{0 < |n| \leq N} \frac{-P_n(y')}{iK_y(y') \cdot n} \mathrm{e}^{in \cdot x}\,, \quad (96)$$

one gets

$$H(y' + \varepsilon g_x, x) = K_1(y') + \varepsilon^2 P'(y', x) \quad (97)$$

where

$$\begin{aligned} K_1(y') &:= K(y') + \varepsilon P_0(y')\,, \\ P'(y', x) &:= P^{(1)} + P^{(2)} + P^{(3)}\,. \end{aligned} \quad (98)$$

Now, by the IFT (Appendix "A The Classical Implicit Function Theorem"), for ε small enough, the map $x \to x + g_{y'}(y', x)$ can be inverted with a real-analytic map of the form

$$\varphi(y', x'; \varepsilon) := x' + \varepsilon\alpha(y', x'; \varepsilon) \tag{99}$$

so that Arnold's symplectic transformation is given by

$$\phi_1 : (y', x') \to \begin{cases} y = y' + \varepsilon g_x\left(y', \varphi(y', x'; \varepsilon)\right) \\ x = \varphi(y', x'; \varepsilon) \\ \quad = x' + \varepsilon\alpha(y', x', ; \varepsilon) \end{cases} \tag{100}$$

(compare (21)). To finish the construction, observe that, from the IFT (see Appendix "A The Classical Implicit Function Theorem" and the quantitative discussion below) it follows that there exists a (unique) point $y_1 \in B(y_0, \bar{r})$ so that the second line of (80) holds, provided ε is small enough.

In conclusion, the analogue of Proposition 1 holds, describing Arnold's scheme:

Proposition 2 *If ϕ_1 is defined in (100) with g given in (96) (with N so that (93) holds) and φ given in (99), then (80) holds with K_1 as in (98) and $P_1(y', x') := P'(y', \varphi(y', x'))$ with P' defined in (98), (95) and (88).*

Estimates and Convergence

If f is a real-analytic function with analytic extension to $W_{r,\xi}$, we denote, for any $r' \le r$ and $\xi' \le \xi$,

$$\|f\|_{r',\xi'} := \sup_{W_{r',\xi'}(y_0)} |f(y, x)| \, ; \tag{101}$$

furthermore, we define

$$T := K_{yy}(y_0)^{-1} \, , \quad M := \|P\|_{r,\xi} \, , \tag{102}$$

and assume (without loss of generality)

$$\kappa < 1 \, , \quad r < 1 \, , \quad \xi \le 1 \, , \qquad \max\{1, \|K_y\|_r, \|K_{yy}\|_r, \|T\|\} < C \, , \tag{103}$$

for a suitable constant C (which, as above, will not change during the iteration).

We begin by discussing how N and $\bar{r}$ depend upon ε. From the exponential decay of the Fourier coefficients (33), it follows that, choosing

$$N := 5\delta^{-1}\lambda \, , \quad \text{where } \lambda := \log|\varepsilon|^{-1} \, , \tag{104}$$

then

$$\|\check{P}\|_{r,\xi-\frac{\delta}{2}} \le |\varepsilon| M \tag{105}$$

provided

$$|\varepsilon| \le \text{const}\, \delta \tag{106}$$

for a suitable[29] const = const(d).

The second key inequality concerns the control of the small divisors $K_y(y') \cdot n$ appearing in the definition of g (see (96)), in a neighborhood $D(y_0, \bar{r})$ of y_0: this will determine the size of $\bar{r}$.

Recalling that $K_y(y_0) = \omega \in \mathcal{D}^d_{\tau,\kappa}$, by Taylor's formula and (9), one finds, for any $0 < |n| \le N$ and any $y' \in D(y_0, \bar{r})$,

$$\begin{aligned} |K_y(y') \cdot n| &= \left|\omega \cdot n + \left(K_y(y') - K_y(y_0)\right) \cdot n\right| \\ &\ge |\omega \cdot n| \left(1 - \frac{\|K_{yy}\|_r}{|\omega \cdot n|}|n|\bar{r}\right) \\ &\ge \frac{\kappa}{|n|^\tau}\left(1 - \frac{C}{\kappa}|n|^{\tau+1}\bar{r}\right) \\ &\ge \frac{\kappa}{|n|^\tau}\left(1 - \frac{C}{\kappa}N^{\tau+1}\bar{r}\right) \\ &\ge \frac{1}{2}\frac{\kappa}{|n|^\tau} \, , \end{aligned} \tag{107}$$

provided $\bar{r} \le r$ satisfies also

$$\bar{r} \le \frac{\kappa}{2CN^{\tau+1}} \overset{(104)}{=} \frac{\kappa}{2 \cdot 5^{\tau+1} C (\delta^{-1}\lambda)^{\tau+1}} \, . \tag{108}$$

Equation (107) allows us to easily control Arnold's generating function g. For example:

$$\begin{aligned} \|g_x\|_{\bar{r},\xi-\frac{\delta}{2}} &= \sup_{D(y_0,\bar{r})\times\mathbb{T}^d_{\xi-\frac{\delta}{2}}} \left| \sum_{0<|n|\le N} \frac{nP_n(y')}{K_y(y') \cdot n} e^{in\cdot x}\right| \\ &\le \sum_{0<|n|\le N} \frac{\sup_{D(y_0,r)}|P_n(y')|}{|K_y(y') \cdot n|}|n| e^{(\xi-\frac{\delta}{2})|n|} \\ &\le \sum_{n\in\mathbb{Z}^d} M \frac{2|n|^{\tau+1}}{\kappa} e^{-\frac{\delta}{2}|n|} \\ &\le \text{const}\frac{M}{\kappa}\delta^{-(\tau+1+d)} \, , \end{aligned} \tag{109}$$

where "const" denotes a constant depending on d and τ only; compare also Remark 6–(i).

Let us now discuss, from a quantitative point of view, how to choose the new "center" of the action variables y_1, which is determined by the requirements in (80). Assuming that

$$\bar{r} \le \frac{r}{2} \tag{110}$$

(allowing the use of Cauchy estimates for y-derivatives of K or P in $D(y_0, \bar{r})$), it is not difficult to see that the quantitative IFT of Appendix "A The Classical Implicit Function Theorem" implies that there exists a unique $y_1 \in D(y_0, \bar{r})$ such that (80) holds and, furthermore

$$|y_1 - y_0| \le 4CMr^{-1}|\varepsilon| \,, \tag{111}$$

and

$$\begin{aligned}\partial_y^2 K_1(y_1) :=& \; K_{yy}(y_1) + \varepsilon\partial_y^2 P_0(y_1)\\ =:& \; T^{-1}(\mathbb{1}_d + A)\end{aligned} \tag{112}$$

with a matrix A satisfying

$$\|A\| \le 10C^3 M|\varepsilon| \le \frac{1}{2} \tag{113}$$

provided[30]

$$\begin{cases} 8C^2 \dfrac{\bar{r}}{r} \le 1 \,, \\ 8CM\bar{r}^{-2}|\varepsilon| \le 1 \end{cases} \tag{114}$$

Equation (113) shows that $\partial_y^2 K(y_1)$ is invertible (Neumann series) and that[31]

$$\partial_y^2 K_1(y_1)^{-1} = T + \varepsilon\widetilde{T} \,, \quad \|\widetilde{T}\| \le 20C^3 M \,. \tag{115}$$

Finally, notice that the second conditions in (114) and (111) imply that $|y_1 - y_0| < \bar{r}/2$ so that

$$D(y_1, \bar{r}/2) \subset D(y_0, \bar{r}) \,. \tag{116}$$

Now, all the estimating tools are set up and, writing

$$\begin{aligned} K_1 &:= K + \varepsilon\widetilde{K} = K + \varepsilon P_0(y') \,, \\ y_1 &:= y_0 + \varepsilon\tilde{y} \,, \end{aligned} \tag{117}$$

one can easily prove (along the lines that led to (53)) the following estimates, where as in Sect. "Kolmogorov Theorem", $\bar{\xi} := \xi - \frac{2}{3}\delta$ and $\bar{r}$ is as above:

$$\begin{cases} \dfrac{\|g_x\|_{\bar{r},\bar{\xi}}}{r}, \|g_y\|_{\bar{r},\bar{\xi}}, \|\widetilde{K}_y\|_{\bar{r}}, \|\widetilde{K}_{yy}\|, |\tilde{y}|, \|\widetilde{T}\| \\ \qquad \le c\kappa^{-2}C^\mu \delta^{-\nu}\lambda^\rho M =: L \,, \\ \|P'\|_{\bar{r},\bar{\xi}} \le c\kappa^{-2}C^\mu\delta^{-\nu}\lambda^\rho M^2 = LM \,, \end{cases} \tag{118}$$

where $c = c(d, \tau) > 1$, $\mu \in \mathbb{Z}_+$, ν and ρ are positive integers depending on d and τ. Now, by[32] Lemma 1 and (118), one has that map $x \to x + \varepsilon g_y(y', x)$ has, for any $y' \in D_{\bar{r}}(y_0)$, an analytic inverse $\varphi = x' + \varepsilon\alpha(x'; y', \varepsilon) =: \varphi(y', x')$ on $\mathbb{T}^d_{\bar{\xi}-\frac{\delta}{3}}$ provided (54), with $\bar{L}$ replaced by L in (118), holds, in which case (55) holds (for any $|\varepsilon| \le \varepsilon_0$ and any $y' \in D_r(y_0)$). Furthermore, under the above hypothesis, it follows that[33]

$$\begin{cases} \phi_1 := \big(y' + \varepsilon g_x(y', \varphi(y', x')), \varphi(y', x')\big) : \\ \qquad W_{\bar{r}/2,\xi-\delta}(y_1) \to W_{r,\xi}(y_0) \\ \|\phi_1 - \mathrm{id}\|_{\bar{r}/2,\xi-\delta} \le |\varepsilon| L \,. \end{cases} \tag{119}$$

Finally, letting $P_1(y', x') := P'(y', \varphi(y', x'))$ one sees that P_1 is real-analytic on $W_{\bar{r}/2,\bar{\xi}-\delta}(y_1)$ and bounded on that domain by

$$\|P_1\|_{\bar{r}/2,\xi-\delta} \le LM \,. \tag{120}$$

In order to iterate the above construction, we fix $0 < \xi_* < \xi$ and set

$$\begin{aligned} C &:= 2\max\{1, \|K_y\|_r, \|K_{yy}\|_r, \|T\|\} \,, \\ \gamma &:= 3C \,, \\ \delta_0 &:= \frac{(\gamma - 1)(\xi - \xi_*)}{\gamma} \,; \end{aligned} \tag{121}$$

ξ_j and δ_j as in (63) *but with* δ_0 *as in* (121); we also define, for any $j \ge 0$,

$$\begin{aligned} \lambda_j &:= 2^j\lambda = \log \varepsilon_0^{-2^j} \,, \\ r_j &:= \frac{\kappa}{4 \cdot 5^{\tau+1} C (\delta_j^{-1}\lambda_j)^{\tau+1}} \,; \end{aligned} \tag{122}$$

(this part is adapted from **Step 3** in Sect. "Kolmogorov Theorem"; see, in particular, (103)). With such choices it is not difficult to check that the iterative construction may be carried out infinitely many times yielding, as a byproduct, Theorem 2 with ϕ real-analytic on $\mathbb{T}^d_{\xi_*}$, provided $|\varepsilon| \le \varepsilon_0$ with ε_0 satisfying[34]

$$\begin{cases} \varepsilon_0 \le \mathrm{e}^{-\beta} & \text{with} \quad \beta := \dfrac{\delta_0}{5}\left(\dfrac{\kappa}{Cr}\right)^{\frac{1}{\tau+1}} \\ \varepsilon_0 DB\|P\|_\xi \le 1 & \text{with} \quad D := 3c\kappa^{-2}\delta_0^{-(\nu+1)}C^\mu \,, \\ & \qquad\quad B := \gamma^{\nu+1}(\log \varepsilon_0^{-1})^\rho \,. \end{cases} \tag{123}$$

Remark 10 Notice that the power of κ^{-1} (the inverse of the Diophantine constant) in the second smallness condition in (123) is two, which implies (compare Remark 7–(v)) that the measure of the complement of the Kolmogorov set may be bounded by a constant times $\sqrt{\varepsilon_0}$. This bound is optimal as the trivial example $(y_1^2 + y_2^2)/2 +$

$\varepsilon \cos(x_1)$ shows: the Hamiltonian is integrable and the phase portrait shows that the separatrices of the pendulum $y_1^2/2 + \varepsilon \cos x_1$ bound a region of area $\sqrt{|\varepsilon|}$ with no KAM tori (as the librational curves within such region are not graphs over the angles).

The Differentiable Case: Moser's Theorem

J.K. Moser, in 1962, proved a perturbation (KAM) Theorem, in the framework of area-preserving twist mappings of an annulus[35] $[0,1] \times \mathbb{S}^1$, for integrable analytic systems perturbed by a C^k perturbation, [42] and [43]. Moser's original setup corresponded to the Hamiltonian case with $d = 2$ and the required smoothness was C^k with $k = 333$. Later, this number was brought down to 5 by H. Rüssmann, [53].

Moser's original approach, similarly to the approach that led J. Nash to prove its theorem on the smooth embedding problem of compact Riemannian manifolds [48], is based on a smoothing technique (via convolutions), which re-introduces at each step of the Newton iteration a certain number of derivatives which one loses in the inversion of the small divisor operator.

The technique, which we shall describe here, is again due to Moser [46] but is rather different from the original one and is based on a quantitative analytic KAM Theorem (in the style of statement in Remark 7–(i) above) in conjunction with a characterization of differentiable functions in terms of functions, which are real-analytic on smaller and smaller complex strips; see [44] and, for an abstract functional approach, [65], [66]. By the way, this approach, suitably refined, leads to optimal differentiability assumptions (i. e., the Hamiltonian may be assumed to be C^ℓ with $\ell > 2d$); see, [50] and the beautiful exposition [59], which inspires the presentation reported here.

Let us consider a Hamiltonian $H = K + \varepsilon P$ (as in (17)) with K a real-analytic Kolmogorov normal form as in (10) with $\omega \in \mathcal{D}^d_{\kappa,\tau}$ and Q real-analytic; P is assumed to be a $C^\ell(\mathbb{R}^d, \mathbb{T}^d)$ function with $\ell = \ell(d,\tau)$ to be specified later[36].

Remark 11 The analytic KAM theorem, we shall refer to is the quantitative Kolmogorov Theorem as stated in Remark 7–(i) above, with (69) strengthened by including in the left hand side of (69) also[37] $\|\partial(\phi_* - \mathrm{id})\|_{\theta\xi}$ and $\|\partial(Q - Q_*)\|_{\theta\xi}$ (where "∂" denotes, here, "Jacobian" with regard to (y,x) for $(\phi_* - \mathrm{id})$ and "gradient" for $(Q - Q_*)$).

The analytic characterization of differentiable functions, suitable for our purposes, is explained in the following two lemmata[38]

Lemma 2 (Jackson, Moser, Zehnder) *Let $f \in C^l(\mathbb{R}^d)$ with $l > 0$. Then, for any $\xi > 0$ there exists a real-analytic function f: $X^d_\xi := \{x \in \mathbb{C}^d : |\mathrm{Im}\, x_j| < \xi\} \to \mathbb{C}$ such that*

$$\begin{cases} \sup\limits_{X^d_\xi} |f_\xi| \le c\|f\|_{C^0}\,, \\ \sup\limits_{X^d_{\xi'}} |f_\xi - f_{\xi'}| \le c\|f\|_{C^l}\xi^l\,, \quad \forall\, 0 < \xi' < \xi\,, \end{cases} \tag{124}$$

where $c = c(d,l)$ is a suitable constant; if f is periodic in some variable x_j, so is f_ξ.

Lemma 3 (Bernstein, Moser) *Let $l \in \mathbb{R}_+ \backslash \mathbb{Z}$ and $\xi_j := 1/2^j$. Let $f_0 = 0$ and let, for any $j \ge 1$, f_j be real analytic functions on $X^d_j := \{x \in \mathbb{C}^d : |\mathrm{Im}\, x_j| < 2^{-j}\}$ such that*

$$\sup_{X^d_j} |f_j - f_{j-1}| \le A 2^{-jl} \tag{125}$$

for some constant A. Then, f_j tends uniformly on $\mathbb{R}^d$ to a function $f \in C^l(\mathbb{R}^d)$ such that, for a suitable constant $C = C(d,l) > 0$,

$$\|f\|_{C^l(\mathbb{R}^d)} \le CA\,. \tag{126}$$

Finally, if the f_i's are periodic in some variable x_j then so is f.

Now, denote by $X_\xi = X^d_\xi \times \mathbb{T}^d \subset \mathbb{C}^{2d}$ and define (compare Lemma 2)

$$P^j := P_{\xi_j}\,, \quad \xi_j := \frac{1}{2^j}\,. \tag{127}$$

Claim M: *If $|\varepsilon|$ is small enough and if $\ell > \sigma + 1$, then there exists a sequence of Kolmogorov symplectic transformations $\{\Phi_j\}_{j\ge0}$, $|\varepsilon|$-close to the identity, and a sequence of Kolmogorov normal forms K_j such that*

$$H_j \circ \Phi_j = K_{j+1} \quad \text{on} \quad W_{\xi_{j+1}} \tag{128}$$

where

$$\begin{aligned} &H_j := K + \varepsilon P^j \\ &\Phi_0 = \phi_0 \quad \text{and} \quad \Phi_j := \Phi_{j-1} \circ \phi_j\,, (j \ge 1) \\ &\phi_j : W_{\xi_{j+1}} \to W_{\alpha\xi_j}\,, \quad \Phi_{j-1} : W_{\alpha\xi_j} \to X_{\xi_j}\,, \\ &\qquad j \ge 1 \ \text{and} \ \alpha := \frac{1}{\sqrt{2}}\,, \\ &\qquad \sup_{x \in \mathbb{T}^d_{\xi_{j+1}}} |\Phi_j(0,x) - \Phi_{j-1}(0,x)| \le \mathrm{const}|\varepsilon| 2^{-(\ell-\sigma)j}\,. \end{aligned} \tag{129}$$

The proof of Claim **M** follows easily by induction[39] from Kolmogorov's Theorem (compare Remark 11) and Lemma 2.

From Claim **M** and Lemma 3 (applied to $f(x) = \Phi_j(0,x) - \Phi_0(0,x)$ and $l = \ell - \sigma$, which may be assumed not integer) it then follows that $\Phi_j(0,x)$ converges in the C^1 norm to a C^1 function $\phi\colon \mathbb{T}^d \to \mathbb{R}^d \times \mathbb{T}^d$, which is ε-close to the identity, and, because of (128),

$$\begin{aligned}\phi(x+\omega t) &= \lim \Phi_j(0, x+\omega t)\\ &= \lim \phi_H^t \circ \Phi_j(0,x) = \phi_H^t \circ \phi(x)\end{aligned} \qquad (130)$$

showing that $\phi(\mathbb{T}^d)$ is a C^1 KAM torus for H (note that the map ϕ is close to the trivial embedding $x \to (y,x)$).

Future Directions

In this section we review in a schematic and informal way some of the most important developments, applications and possible future directions of KAM theory. For exhaustive surveys we refer to [9], Sect. 6.3 of [6] or [60].

1. *Structure of the Kolmogorov set and Whitney smoothness*

The Kolmogorov set (i. e., the union of KAM tori), in nearly-integrable systems, tends to fill up (in measure) the whole phase space as the strength of the perturbation goes to zero (compare Remark 7–(v) and Remark 10). A natural question is: what is the global geometry of KAM tori?

It turns out that KAM tori smoothly interpolate in the following sense. *For ε small enough, there exists a C^∞ symplectic diffeomorphism ϕ_* of the phase space $\mathcal{M} = B \times \mathbb{T}^d$ of the nearly-integrable, non-degenerate Hamiltonians $H = K(y) + \varepsilon P(y,x)$ and a Cantor set $C_* \subset B$ such that, for each $y' \in C_*$, the set $\phi_*^{-1}(\{y'\} \times \mathbb{T}^d)$ is a KAM torus for H*}; in other words, the Kolmogorov set is a smooth, symplectic deformation of the fiber bundle $C_* \times \mathbb{T}^d$. Still another way of describing this result is that *there exists a smooth function $K_*\colon B \to \mathbb{R}$ such that $(K + \varepsilon P) \circ \phi_*$ and K_* agree, together with their derivatives, on $C_* \times \mathbb{T}^d$*: we may, thus, say that, in general, nearly-integrable Hamiltonian systems are integrable on Cantor sets of relative big measure.

Functions defined on closed sets which admit C^k extensions are called *Whitney smooth*; compare [64], where H. Whitney gives a sufficient condition, based on Taylor uniform approximations, for a function to be Whitney C^k.

The proof of the above result – given, independently, in [50] and [19] in, respectively, the differentiable and the analytic case – follows easily from the following lemma[40]:

Lemma 4 *Let $C \subset \mathbb{R}^d$ a closed set and let $\{f_j\}$, $f_0 = 0$, be a sequence of functions analytic on $W_j := \cup_{y \in C} D(y, r_j)$. Assume that $\sum_{j\ge1} \sup_{W_j} |f_j - f_{j-1}| r_j^{-k} < \infty$. Then, f_j converges uniformly to a function f, which is C^k in the sense of Whitney on C.*

Actually, the dependence upon the angles x' of ϕ_* is analytic and it is only the dependence upon $y' \in C_*$ which is Whitney smooth ("anisotropic differentiability", compare Sect. 2 in [50]).

For more information and a systematic use of Whitney differentiability, see [9].

2. *Power series expansions*

KAM tori $\mathcal{T}_{\omega,\varepsilon} = \phi_\varepsilon(\mathbb{T}^d)$ of nearly-integrable Hamiltonians correspond to quasi-periodic trajectories $z(t;\theta,\varepsilon) = \phi_\varepsilon(\theta + \omega t) = \phi_H^t(z(0;\theta,\varepsilon))$; compare items (d) and (e) of Sect. "Introduction" and Remark 2–(i) above. While the *actual* existence of such quasi-periodic motions was proven, for the first time, only thanks to KAM theory, the *formal* existence, in terms of formal ε-power series[41] was well known in the nineteenth century to mathematicians and astronomers (such as Newcombe, Lindstedt and, especially, Poincaré; compare [49], vol. II). Indeed, formal power solutions of nearly-integrable Hamiltonian equations are not difficult to construct (see, e. g., Sect. 7.1 of [12]) but *direct proofs* of the convergence of the series, i. e., proofs not based on Moser's "indirect" argument recalled in Remark 7–(iii) but, rather, based upon direct estimates on the kth ε-expansion coefficient, are quite difficult and were carried out only in the late eighties by H. Eliasson [27]. The difficulty is due to the fact that, in order to prove the convergence of the Taylor–Fourier expansion of such series, one has to recognize compensations among huge terms with different signs[42]. After Eliasson's breakthrough based upon a semi-direct method (compare the "Postscript 1996" at p. 33 of [27]), fully direct proofs were published in 1994 in [30] and [18].

3. *Non-degeneracy assumptions*

Kolmogorov's non-degeneracy assumption (70) can be generalized in various ways. First of all, Arnold pointed out in [2] that the condition

$$\det\begin{pmatrix} K_{yy} & K_y \\ K_y & 0 \end{pmatrix} \neq 0\,, \qquad (131)$$

(this is a $(d+1)\times(d+1)$ symmetric matrix where last column and last row are given by the $(d+1)$-vector $(K_y, 0)$) which is independent from condition (70), is also sufficient to construct KAM tori. Indeed, (131) may be used to construct *iso-energetic* KAM tori, i. e., tori on a *fixed energy level*[43] E.

More recently, Rüssmann [57] (see, also, [58]), using results of Diophantine approximations on manifolds due to Pyartly [52], formulated the following condition (the "Rüssmann non-degeneracy condition"), which is essentially necessary and sufficient for the existence of a positive measure set of KAM tori in nearly-integrable Hamiltonian systems: *the image* $\omega(B) \subset \mathbb{R}^d$ *of the unperturbed frequency map* $y \to \omega(y) := K_y(y)$ *does not lie in any hyperplane passing through the origin*. We simply add that one of the prices that one has to pay to obtain these beautiful general results is that one cannot fix the frequency ahead of time.

For a thorough discussion of this topic, see Sect. 2 of [60].

4. *Some physical applications*

We now mention a short (and non-exhaustive) list of important physical application of KAM theory. For more information, see Sect. 6.3.9 of [6] and references therein.

4.1. *Perturbation of classical integrable systems*

As mentioned above (Remark 1–(iii)), one of the main original motivations of KAM theory is the perturbation theory for nearly-integrable Hamiltonian systems. Among the most famous classical integrable systems we recall: one-degree-of freedom systems; Keplerian two-body problem, geodesic motion on ellipsoids; rotations of a heavy rigid body with a fixed point (for special values of the parameters: Euler's, Lagrange's, Kovalevskaya's and Goryachev–Chaplygin's cases); Calogero–Moser's system of particles; see, Sect. 5 of [6] and [47].

A first step, in order to apply KAM theory to such classical systems, is to explicitly construct action-angle variables and to determine their analyticity properties, which is in itself a technically non-trivial problem. A second problem which arises, especially in Celestial Mechanics, is that the integrable (transformed) Hamiltonian governing the system may be highly degenerate (*proper degeneracies* – see Sect. 6.3.3, B of [6]), as is the important case of the planetary n-body problem. Indeed, the first complete proof of the existence of a positive measure set of invariant tori[44] for the planetary $(n+1)$ problem (one body with mass 1 and n bodies with masses smaller than ε) has been published only in 2004 [29]. For recent reviews on this topic, see [16].

4.2. *Topological trapping in low dimensions*

The general 2-degree-of-freedom nearly-integrable Hamiltonian exhibits a kind of particularly strong stability: the phase space is 4-dimensional and the energy levels are 3-dimensional; thus KAM tori (which are two-dimensional and which are guaranteed, under condition (131), by the iso-energetic KAM theorem) *separate* the energy levels and orbits lying between two KAM tori will remain forever trapped in the invariant region. In particular the evolution of the action variables stays forever close to the initial position ("total stability").

This observation is originally due to Arnold [2]; for recent applications to the stability of three-body problems in celestial mechanics see [13] and item **4.4** below.

In higher dimension this topological trapping is no longer available, and in principle nearby any point in phase space it may pass an orbit whose action variables undergo a displacement of order one ("Arnold's diffusion"). A rigorous complete proof of this conjecture is still missing[45].

4.3. *Spectral Theory of Schrödinger operators*

KAM methods have been applied also very successfully to the spectral analysis of the one-dimensional Schrödinger (or "Sturm–Liouville") operator on the real line $\mathbb{R}$

$$L := -\frac{d^2}{dt^2} + v(t)\,, \quad t \in \mathbb{R}\,. \tag{132}$$

If the "potential" v is bounded then there exists a unique self-adjoint operator on the real Hilbert space $\mathcal{L}^2(\mathbb{R})$ (the space of Lebesgue square-integrable functions on $\mathbb{R}$) which extends L above on C_0^2 (the space of twice differentiable functions with compact support). The problem is then to study the spectrum $\sigma(L)$ of L; for generalities, see [23].

If v is periodic, then $\sigma(L)$ is a continuous band spectrum, as it follows immediately from Floquet theory [23]. Much more complicated is the situation for quasi-periodic potentials $v(t) := V(\omega t) = V(\omega_1 t, \ldots, \omega_n t)$, where V is a (say) real-analytic function on $\mathbb{T}^n$, since small-divisor problems appear and the spectrum can be nowhere dense. For a beautiful classical exposition, see [47], where, in particular, interesting connections with mechanics are discussed[46]; for deep developments of generalization of Floquet theory to quasi-periodic Schrödinger operators ("reducibility"), see [26] and [7].

4.4. *Physical stability estimates and break-down thresholds*

KAM Theory is perturbative and works if the parameter ε measuring the strength of the perturbation is small enough. It is therefore a fundamental question: *how small ε has to be in order for KAM results to hold*. The first concrete applications were extremely discouraging: in 1966, the French astronomer M. Hénon [32] pointed out that Moser's theorem applied to the restricted three-body problem (i. e., the motion of an asteroid under the gravitational influence of two unperturbed primary bodies revolving on a given Keplerian ellipse) yields existence of invariant tori if the mass ratio of the primaries is less than[47] 10^{-52}. Since then, much progress has been made and very recently, in [13], it has been shown via a computer-assisted proof[48], that, for a restricted-three body model of a subsystem of the Solar system (namely, Sun, Jupiter and Asteroid Victoria), KAM tori exist for the "actual" physical values (in that model the Jupiter/Sun mass ratio is about 10^{-3}) and, in this mathematical model – thanks to the trapping mechanism described in item **4.2** above – they trap the actual motion of the subsystem.

From a more theoretical point of view, we notice that, (compare Remark 2–(ii)) KAM tori (with a fixed Diophantine frequency) are analytic in ε; on the other hand, it is known, at least in lower dimensional settings (such as twist maps), that above a certain critical value KAM tori (curves) cannot exist ([39]). Therefore, there must exist a critical value $\varepsilon_c(\omega)$ ("breakdown threshold") such that, for $0 \le \varepsilon < \varepsilon_c\omega$, the KAM torus (curve) $\mathcal{T}_{\omega,\varepsilon}$ exists, while for $\varepsilon > \varepsilon_c(\omega)$ does not. The mathematical mechanism for the breakdown of KAM tori is far from being understood; for a brief review and references on this topic, see, e. g., Sect. 1.4 in [13].

5. *Lower dimensional tori*

In this item we consider (very) briefly, the existence of quasi-periodic solutions with a number of frequencies smaller than the number of degrees of freedom[49]. Such solutions span *lower dimensional* (non Lagrangian) *tori*. Certainly, this is one of the most important topics in modern KAM theory, not only in view of applications to classical problems, but especially in view of extensions to infinite dimensional systems, namely PDEs (Partial Differential Equations) with a Hamiltonian structure; see, item **6** below. For a recent, exhaustive review on lower dimensional tori (in finite dimensions), we refer the reader to [60].

In 1965 V.K. Melnikov [41] stated a precise result concerning the persistence of *stable* (or "elliptic") lower dimensional tori; the hypotheses of such results are, now, commonly referred to as "Melnikov conditions". However, a proof of Melnikov's statement was given only later by Moser [45] for the case $n = d - 1$ and, in the general case, by H. Eliasson in [25] and, independently, by S.B. Kuksin [37]. The *unstable* ("partially hyperbolic") case (i. e., the case for which the lower dimensional tori are linearly unstable and lie in the intersection of stable and unstable Lagrangian manifolds) is simpler and a complete perturbation theory was already given in [45], [31] and [66] (roughly speaking, the normal frequencies to the torus do not resonate with the inner (or "proper") frequencies associated with quasi-periodic motion). Since then, Melnikov conditions have been significantly weakened and much technical progress has been made; see [60], Sects. 5, 6 and 7, and references therein.

To illustrate a typical situation, let us consider a Hamiltonian system with $d = n + m$ degrees of freedom, governed by a Hamiltonian function of the form

$$H(y, x, v, u; \xi) = K(y, v, u; \xi) + \varepsilon P(y, x, v, u; \xi), \quad (133)$$

where $(y, x) \in \mathbb{T}^n \times \mathbb{R}^n$, $(v, u) \in \mathbb{R}^{2m}$ are pairs of standard symplectic coordinates and ξ is a real parameter running over a compact set $\Pi \subset \mathbb{R}^n$ of positive Lebesgue measure[50]; K is a Hamiltonian admitting the n-torus

$$\mathcal{T}_0^n(\xi) := \{y = 0\} \times \mathbb{T}^n \times \{v = u = 0\}, \quad \xi \in \Pi,$$

as invariant linearly stable invariant torus and is assumed to be in the normal form:

$$K = E(\xi) + \omega(\xi) \cdot y + \frac{1}{2} \sum_{j=1}^{m} \Omega_j(\xi)(u_j^2 + v_j^2). \quad (134)$$

The ϕ_K^t flow decouples in the linear flow $x \in \mathbb{T}^n \to x + \omega(\xi)t$ times the motion of m (decoupled) harmonic oscillators with characteristic frequencies $\Omega_j(\xi)$ (sometimes referred to as *normal frequencies*). Melnikov's conditions (in the form proposed in [51]) reads as follows: assume that ω is a Lipschitz homeomorphism; let $\Pi_{k,l}$ denote the "resonant parameter set" $\{\xi \in \Pi : \omega(\xi) \cdot k + \Omega \cdot (\xi) = 0\}$ and assume

$$\begin{cases} \Omega_i(\xi) > 0, \quad \Omega_i(\xi) \neq \Omega_j(\xi), \quad \forall \xi \in \Pi, \forall i \neq j \\ \text{meas } \Pi_{k,l} = 0, \quad \forall k \in \mathbb{Z}^n \backslash \{0\}, \quad \forall l \in \mathbb{Z}^m : |l| \le 2. \end{cases} \quad (135)$$

Under these assumptions and if $|\varepsilon|$ is small enough, there exists a (Cantor) subset of parameters $\Pi_* \subset \Pi$ of posi-

tive Lebesgue measure such that, to each $\xi \in \Pi_*$, there corresponds a n-dimensional, linearly stable H-invariant torus $\mathcal{T}^n_\varepsilon(\xi)$ on which the H flow is analytically conjugated to $x \to x + \omega_*(\xi)t$ where ω_* is a Lipschitz homeomorphism of Π_* assuming Diophantine values and close to ω.

This formulation has been borrowed from [51], to which we refer for the proof; for the differentiable analog, see [22].

Remark 12 The small-divisor problems arising in the perturbation theory of the above lower dimensional tori are of the form

$$\omega \cdot k - l \cdot \Omega\,, \quad |l| \le 2\,, \quad |k| + |l| \ne 0\,, \tag{136}$$

where one has to regard the normal frequency Ω as *functions of the inner frequencies* ω and, at first sight, one has – in J. Moser words – a *lack-of-parameter* problem. To overcome this intrinsic difficulty, one has to give up full control of the inner frequencies and construct, iteratively, n-dimensional sets (corresponding to smaller and smaller sets of ξ-parameters) on which the small divisors are controlled; for more motivations and informal explanations on lower dimensional small divisor problems, see, Sects. 5, 6 and 7 of [60].

6. *Infinite dimensional systems*

As mentioned above, the most important recent developments of KAM theory, besides the full applications to classical n-body problems mentioned above, is the successful extension to infinite dimensional settings, so as to deal with certain classes of partial differential equations carrying a Hamiltonian structure. As a typical example, we mention the non-linear wave equation of the form

$$u_{tt} - u_{xx} + V(x)u = f(u)\,, \qquad f(u) = O(u^2)\,, \quad 0 < x < 1\,, \quad t \in \mathbb{R}\,. \tag{137}$$

These extensions allowed, in the pioneering paper [63], establishing the existence of small-amplitude quasi-periodic solutions for (137), subject to Dirichlet or Neumann boundary conditions (on a finite interval for odd and analytic nonlinearities f); the technically more difficult periodic boundary condition case was considered later; compare [38] and references therein.

A technical discussion of these topics goes far beyond the scope of the present article and, for different equations, techniques and details, we refer the reader to the review article [38].

A The Classical Implicit Function Theorem

Here we discuss the classical Implicit Function Theorem for complex functions from a quantitative point of view. The following Theorem is a simple consequence of the Contraction Lemma, which asserts that a contraction Φ on a closed, non-empty metric space[51] X has a unique fixed point, which is obtained as $\lim_{j\to\infty} \Phi^j(u_0)$ for any[52] $u_0 \in X$. As above, $D^n(y_0, r)$ denotes the ball in $\mathbb{C}^n$ of center y_0 and radius r.

Theorem 3 (Implicit Function Theorem) *Let*

$$F\colon (y,x) \in D^n(y_0,r) \times D^m(x_0,s) \subset \mathbb{C}^{n+m} \to F(y,x) \in \mathbb{C}^n$$

be continuous with continuous Jacobian matrix F_y; assume that $F_y(y_0,x_0)$ is invertible and denote by T its inverse; assume also that

$$\sup_{D(y_0,r)\times D(x_0,s)} \|\mathbb{1}_n - TF_y(y,x)\| \le \frac{1}{2}\,, \qquad \sup_{D(x_0,s)} |F(y_0,x)| \le \frac{r}{2\|T\|}\,. \tag{138}$$

Then, all solutions $(y,x) \in D(y_0,r)\times D(x_0,s)$ of $F(y,x) = 0$ are given by the graph of a unique continuous function $g\colon D(x_0,s) \to D(y_0,r)$ satisfying, in particular,

$$\sup_{D(x_0,s)} |g| \le 2\|T\| \sup_{D(x_0,s)} |F(y_0,\cdot)|\,. \tag{139}$$

Proof Let $X = C(D^m(x_0,s), D^n(y_0,r))$ be the closed ball of continuous function from $D^m(x_0,s)$ to $D^n(y_0,r)$ with respect to the sup-norm $\|\cdot\|$ (X is a non-empty metric space with distance $d(u,v) := \|u-v\|$) and denote $\Phi(y;x) := y - TF(y,x)$. Then, $u \to \Phi(u) := \Phi(u,\cdot)$ maps $C(D^m(x_0,s))$ into $C(\mathbb{C}^m)$ and, since $\partial_y\Phi = \mathbb{1}_n - TF_y(y,x)$, from the first relation in (138), it follows that $u \to \Phi(u)$ is a contraction. Furthermore, for any $u \in C(D^m(x_0,s), D^n(y_0,r))$,

$$\begin{aligned}|\Phi(u) - y_0| &\le |\Phi(u) - \Phi(y_0)| + |\Phi(y_0) - y_0| \\ &\le \frac{1}{2}\|u - y_0\| + \|T\|\|F(y_0,x)\| \\ &\le \frac{1}{2}r + \|T\|\frac{r}{2\|T\|} = r\,,\end{aligned}$$

showing that $\Phi\colon X \to X$. Thus, by the Contraction Lemma, there exists a unique $g \in X$ such that $\Phi(g) = g$, which is equivalent to $F(g,x) = 0$ $\forall x$. If $F(y_1,x_1) = 0$ for some $(y_1,x_1) \in D(y_0,r)\times D(x_0,s)$, it follows that $|y_1 - g(x_1)| = |\Phi(y_1;x_1) - \Phi(g(x_1),x_1)| \le \alpha|y_1 - g(x_1)|$, which implies that $y_1 = g(x_1)$ and that all solutions of $F = 0$ in $D(y_0, r) \times D(x_0, s)$ coincide with the graph og g. Finally, (139) follows by observing that $\|g - y_0\| = \|\Phi(g) - y_0\| \le \|\Phi(g) - \Phi(y_0)\| + \|\Phi(y_0) - y_0\| \le \frac{1}{2}\|g - y_0\| + \|T\|\|F(y_0, \cdot)\|$, finishing the proof. □

Additions:

(i) If F is periodic in x or/and real on reals, then (by uniqueness) so is g;
(ii) If F is analytic, then so is g (Weierstrass Theorem, since g is attained as uniform limit of analytic functions);
(iii) The factors 1/2 appearing in the right-hand sides of (138) may be replaced by, respectively, α and β for any positive α and β such that $\alpha + \beta = 1$.

Taking $n = m$ and $F(y, x) = f(y) - x$ for a given $C^1(D(y_0, r), \mathbb{C}^n)$ function, one obtains the

Theorem 4 (Inverse Function Theorem) *Let $f\colon y \in D^n(y_0, r) \to \mathbb{C}^n$ be a C^1 function with invertible Jacobian $f_y(y_0)$ and assume that*

$$\sup_{D(y_0,r)} \|\mathbb{1}_n - T f_y\| \le \frac{1}{2}\,, \quad T := f_y(y_0)^{-1}\,, \tag{140}$$

then there exists a unique C^1 function $g\colon D(x_0, s) \to D(y_0, r)$ with $x_0 := f(y_0)$ and $s := r/(2\|T\|)$ such that $f \circ g(x) = \mathrm{id} = g \circ f$.

Additions analogous to the above also hold in this case.

B Complementary Notes

1 Actually, the first instance of a small divisor problem solved analytically is the linearization of the germs of analytic functions and is due to C.L. Siegel [61].

2 The well-known Newton's tangent scheme is an algorithm, which allows us to find roots (zeros) of a smooth function f in a region where the derivative f' is bounded away from zero. More precisely, if x_n is an "approximate solution" of $f(x) = 0$, i. e., $f(x_n) := \varepsilon_n$ is small, then the next approximation provided by Newton's tangent scheme is $x_{n+1} := x_n - f(x_n)/f'(x_n)$ [which is the intersection with x-axis of the tangent to the graph of f passing through $(x_n, f(x_n))$] and, in view of the definition of ε_n and Taylor's formula, one has that $\varepsilon_{n+1} := f(x_{n+1}) = \frac{1}{2} f''(\xi_n)\varepsilon_n{}^2/(f'(x_n))^2$ (for a suitable ξ_n) so that $\varepsilon_{n+1} = O(\varepsilon_n^2) = O(\varepsilon_1^{2^n})$ and, in the iteration, x_n will converge (at a super-exponential rate) to a root $\bar{x}$ of f. This type of extremely fast convergence will be typical in the analyzes considered in the present article.

3 The elements of $\mathbb{T}^d$ are equivalence classes $x = \bar{x} + 2\pi\mathbb{Z}^d$ with $\bar{x} \in \mathbb{R}^d$. If $x = \bar{x} + 2\pi\mathbb{Z}^d$ and $y = \bar{y} + 2\pi\mathbb{Z}^d$ are elements of $\mathbb{T}^d$, then their distance $d(x, y)$ is given by $\min_{n \in \mathbb{Z}^d} |\bar{x} - \bar{y} + 2\pi n|$ where $|\cdot|$ denotes the standard euclidean norm in $\mathbb{R}^n$; a smooth (analytic) function on $\mathbb{T}^d$ may be viewed as ("identified with") a smooth (analytic) function on $\mathbb{R}^d$ with period 2π in each variable. The torus $\mathbb{T}^d$ endowed with the above metric is a real-analytic, compact manifold. For more information, see [62].

4 A symplectic form on an (even dimensional) manifold is a closed, non-degenerate differential 2-form. The symplectic form $\alpha = \mathrm{d}y \wedge \mathrm{d}x$ is actually *exact* symplectic, meaning that $\alpha = \mathrm{d}(\sum_{i=1} y_i \mathrm{d}x_i)$. For general information see [5].

5 For general facts about the theory of ODE (such as Picard theorem, smooth dependence upon initial data, existence times, …) see, e. g., [23].

6 This terminology is due to that fact the the x_j are "adimensional" angles, while analyzing the physical dimensions of the quantities appearing in Hamilton's equations one sees that $\dim(y) \times \dim(x) = \dim H \times \dim(t)$ so that y has the dimension of an energy (the Hamiltonian) times the dimension of time, i. e., by definition, the dimension of an action.

7 This terminology is due to the fact that a classical mechanical system of d particles of masses $m_i > 0$ and subject to a potential $V(q)$ with $q \in A \subset \mathbb{R}^d$ is governed by a Hamiltonian of the form $\sum_{j=1}^d p_j^2/2m_j + V(q)$ and d may be interpreted as the (minimal) number of coordinates necessary to physically describe the system.

8 To be precise, (6) should be written as $y(t) = v(\pi_{\mathbb{T}^d}(\omega t))$, $x(t) = \pi_{\mathbb{T}^d}(\omega t + u(\pi_{\mathbb{T}^d}(\omega t)))$ where $\pi_{\mathbb{T}^d}$ denotes the standard projection of $\mathbb{R}^d$ onto $\mathbb{T}^d$, however we normally omit such a projection.

9 As standard, U_θ denotes the $(d \times d)$ Jacobian matrix with entries $(\partial U_i)/(\partial\theta_j) = \delta_{ij} + (\partial u_i)/(\partial\theta_j)$.

10 For generalities, see [5]; in particular, a Lagrangian manifold $L \subset \mathcal{M}$ which is a graph over $\mathbb{T}^d$ admits a "generating function", i. e., there exists a smooth function $g\colon \mathbb{T}^d \to \mathbb{R}$ such that $L = \{(y, x)\colon y = g_x(x), x \in \mathbb{T}^d\}$.

11 Compare [54] and references therein. We remark that, if $B(\omega_0, r)$ denote the ball in $\mathbb{R}^d$ of radius r centered at ω_0 and fix $\tau > d - 1$, then one can prove that the Lebesgue measure of $B(y_0, r)\backslash\mathcal{D}^d_{\kappa,\tau}$ can be bounded by $c_d \kappa r^{d-1}$ for a suitable constant c_d depending only on d; for a simple proof, see, e.g, [21].

12 The sentence "can be put into the form" means "there exists a symplectic diffeomorphism $\phi\colon (y, x) \in \mathcal{M} \to$

$(\eta, \xi) \in \mathcal{M}$ such that $H \circ \phi$ has the form (10)"; for multi-indices α, $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and $\partial_y^\alpha = \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d}$; the vanishing of the derivatives of a function $f(y)$ up to order k in the origin will also be indicated through the expression $f = O(|y|^{k+1})$.

13 **Notation:** If A is an open set and $p \in \mathbb{N}$, then the C^p-norm of a function $f: x \in A \to f(x)$ is defined as $\|f\|_{C^p(A)}$: $\sup_{|\alpha| \le p} \sup_A |\partial_x^\alpha f|$.

14 **Notation:** If f is a scalar function f_y is a d-vector; f_{yy} the Hessian matrix $(f_{y_i y_j})$; f_{yyy} the symmetric 3-tensor of third derivatives acting as follows: $f_{yyy} a \cdot b \cdot c := \sum_{i,j,k=1}^d (\partial^3 f)/(\partial y_i \partial y_j \partial y_k) a_i b_j c_k$.

15 **Notation:** If f is (a regular enough) function over $\mathbb{T}^d$, its Fourier coefficients are defined as $f_n := \int_{\mathbb{T}^d} f(x) e^{-in\cdot x} dx/(2\pi)^d$; where, as usual, $i = \sqrt{-1}$ denotes imaginary unit; for general information about Fourier series see, e. g., [34].

16 The choice of norms on finite dimensional spaces ($\mathbb{R}^d$, $\mathbb{C}^d$, space of matrices, tensors, etc.) is not particularly relevant for the analysis in this article (since changing norms will change d-depending constants); however for matrices, tensors (and, in general, linear operators), it is convenient to work with the "operator norm", i. e., the norm defined as $\|L\| = \sup_{u \ne 0} \|Lu\|/\|u\|$, so that $\|Lu\| \le \|L\|\|u\|$, an estimate, which will be constantly be used; for a general discussion on norms, see, e. g., [36].

17 As an example, let us work out the first two estimates, i. e., the estimates on $\|s_x\|_{\bar\xi}$ and $|b|$: actually these estimates will be given on a larger intermediate domain, namely, $W_{\xi-\delta/3}$, allowing to give the remaining bounds on the smaller domain $W_{\bar\xi}$ (recall that W_s denotes the complex domain $D(0, s) \times \mathbb{T}_s^d$). Let $f(x) := P(0, x) - \langle P(0, \cdot)\rangle$. By definition of $\|\cdot\|_\xi$ and M, it follows that $\|f\|_\xi \le \|P(0, x)\|_\xi + \|\langle P(0, \cdot)\rangle\|_\xi \le 2M$. By **(P5)** with $p = 1$ and $\xi' = \xi - \delta/3$, one gets

$$\|s_x\|_{\xi - \frac{\delta}{3}} \le \bar B_1 \frac{2M}{\kappa} 3^{k_1} \delta^{-k_1} ,$$

which is of the form (53), provided $\bar c \ge (\bar B_1 2 \cdot 3^{k_1})/\kappa$ and $\bar\nu \ge k_1$. 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 estimate: by **(P4)** with $p = 2$ and, respectively, $p = 1$, and $\xi' = 0$, we get

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

where $m = m(d) \ge 1$ is a constant which depend 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 (53) provided $\bar c \ge m(B_2 \bar B_1 2 \cdot 3^{k_1} \kappa^{-1} + B_1)$, $\mu \ge 2$ and $\bar\nu \ge k_1 + 2$. The other bounds in (53) follow easily along the same lines.

18 We sketch here the **proof of Lemma 1**. The defining relation $\psi_\varepsilon \circ \varphi = \text{id}$ implies that $\alpha(x') = -a(x' + \varepsilon\alpha(x'))$, where $\alpha(x')$ is short for $\alpha(x'; \varepsilon)$ and that equation is a fixed point equation for the non-linear operator $f: u \to f(u) := -a(\text{id} + \varepsilon u)$. To find a fixed point for this equation one can use a standard contraction Lemma (see [36]). Let Y denote the closed ball (with respect to the sup-norm) of continuous functions $u: \mathbb{T}_{\xi'}^d \to \mathbb{C}^d$ such that $\|u\|_{\xi'} \le \bar L$. By (54), $|\text{Im}(x' + \varepsilon u(x'))| < \xi' + \varepsilon_0 \bar L < \xi' + \delta/3 = \bar\xi$, for any $u \in Y$, and any $x' \in \mathbb{T}_{\xi'}^d$; thus, $\|f(u)\|_{\xi', \varepsilon_*} \le \|a\|_{\bar\xi} \le \bar L$ by (53), so that $f: Y \to Y$; notice that, in particular, this means that f sends periodic functions into periodic functions. Moreover, (54) implies also that f is a contraction: if $u, v \in Y$, then, by the mean value theorem, $|f(u) - f(v)| \le \bar L |\varepsilon| |u - v|$ (with a suitable choice of norms), so that, by taking the sup-norm, one has $\|f(u) - f(v)\|_{\xi'} < \varepsilon_0 \bar L \|u - v\|_{\xi'} < \frac{1}{3}\|u - v\|_{\xi'}$ 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 uniform limit of analytic function (see [1]), the limit α itself is analytic. In conclusion, $\varphi \in \mathcal{B}_{\xi'}$ and (55) holds.

Next, for $(y', x) \in W_{\bar\xi}$, by (53), one has $|y' + \varepsilon\beta(y', x)| < \bar\xi + \varepsilon_0 \bar L < \bar\xi + \delta/3 = \xi$ so that (56) holds. Furthermore, since $\|\varepsilon a_x\|_{\bar\xi} < \varepsilon_0 \bar L < 1/3$ 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 (57) holds. The proof is finished.

19 From (59), it follows immediately that $\langle \partial_{y'}^2 Q_1(0, \cdot)\rangle = \langle \partial_y^2 Q(0, \cdot)\rangle + \varepsilon \langle \partial_{y'}^2 \widetilde Q(0, \cdot)\rangle = T^{-1}(\mathbb{1}_d + \varepsilon T \langle \partial_{y'}^2 \widetilde Q(0, \cdot)\rangle) =: T^{-1}(\mathbb{1}_d + \varepsilon R)$ and, in view of (51) and (59), we see that $\|R\| < L/(2C)$. Therefore, by (60), $\varepsilon_0 \|R\| < 1/6 < 1/2$, implying that $(1 + \varepsilon R)$ is invertible and $(\mathbb{1}_d + \varepsilon R)^{-1} = \mathbb{1}_d + \sum_{k=1}^\infty (-1)^k \varepsilon^k R^k =: 1 + \varepsilon D$ with $\|D\| \le \|R\|/(1 - |\varepsilon|\|R\|) < L/C$. In conclusion, $T_1 = (1 + \varepsilon R)^{-1} T = T + \varepsilon D T =: T + \varepsilon \widetilde T$, $\|\widetilde T\| \le \|D\| C \le (L/C) C = L$.

20 Actually, there is quite some freedom in choosing the sequence $\{\xi_j\}$ provided the convergence is not too fast; for general discussion, see, [56], or, also, [10] and [14].

21 In fact, denoting by B_* the real d-ball centered at 0 and of radius $\theta\xi_*$ for $\theta \in (0, 1)$, from Cauchy estimate (47) with $\xi = \xi_*$ and $\xi' = \theta\xi_*$, one has $\|\phi_* -$

$\mathrm{id}\|_{C^p(B_*\times\mathbb{T}^d)} = \sup_{B_*\times\mathbb{T}^d}\sup_{|\alpha|+|\beta|\le p}|\partial_y^\alpha\partial_x^\beta(\phi_* - \mathrm{id})| \le \sup_{|\alpha|+|\beta|\le p}\|\partial_y^\alpha\partial_x^\beta(\phi_* - \mathrm{id})\|_{\theta\xi_*} \le B_p\|\phi_* - \mathrm{id}\|_{\xi_*}1/(\theta\xi_*)^p \le \mathrm{const}_p|\varepsilon|$ with $\mathrm{const}_p := B_pDBM1/(\theta\xi_*)^p$. An identical estimate holds for $\|Q_* - Q\|_{C^p(B_*\times\mathbb{T}^d)}$.

22 Also very recently ε-power series expansions have been shown to be a very powerful tool; compare [13].

23 A function $f\colon A\subset\mathbb{R}^n\to\mathbb{R}^n$ is Lipschitz on A if there exists a constant ("Lipschitz constant") $L>0$ such that $|f(x)-f(y)|\le L|x-y|$ for all $x,y\in A$. For a general discussion on how Lebesgue measure changes under Lipschitz mappings, see, e. g., [28]. In fact, the dependence of ϕ_* on $\bar y$ is much more regular, compare Remark 11.

24 In fact, notice that inverse powers of κ appear through (48) (inversion of the operator D_ω), therefore one sees that the terms in the first line of (53) may be replaced by $\bar c\kappa^{-2}$ (in defining a one has to apply the operator D_ω^{-1} twice) but then in $P^{(1)}$ (see (26)) there appears $\|\beta\|^2$, so that the constant c in the second line of (53) has the form (72); since $\kappa<1$, one can replace in (53) c with $\hat c\kappa^{-4}$ as claimed.

25 **Proof of Claim C** Let $H_0 := H$, $E_0 := E$, $Q_0 := Q$, $K_0 := K$, $P_0 := P$, $\xi_0 := \xi$ and let us assume (*inductive hypothesis*) that we can iterate the Kolmogorov transformation j times obtaining j symplectic transformations $\phi_{i+1}\colon W_{\xi_{i+1}}\to W_{\xi_i}$, for $0\le i\le j-1$, and j Hamiltonians $H_{i+1} = H_i\circ\phi_{i+1} = K_i + \varepsilon^{2^i}P_i$ real-analytic on W_{ξ_i} such that

$$\begin{aligned}&|\omega|,|E_i|,\|Q_i\|_{\xi_i},\|T_i\|<C\,,\\ &|\varepsilon|^{2^i}L_i := |\varepsilon|^{2^i}cC^\mu\delta_0^{-\nu}2^{\nu i}M_i\le\frac{\delta_i}{3}\,,\\ &\quad\forall\,0\le i\le j-1\,.\end{aligned}\tag{$*$}$$

By $(*)$, Kolmogorov iteration (**Step 2**) can be applied to H_i and therefore all the bounds described in paragraph **Step 2** hold (having replaced $H,E,\dots,\xi,\delta,H',E',\dots,\xi'$ with, respectively, $H_i,E_i,\dots,\xi_i,\delta_i,H_{i+1},E_{i+1},\dots,\xi_{i+1}$); in particular (see (61)) one has, for $0\le i\le j-1$ (and for any $|\varepsilon|\le\varepsilon_0$),

$$\begin{cases}|E_{i+1}|\le|E_i|+|\varepsilon|^{2^i}L_i\,,\\ \quad\|Q_{i+1}\|_{\xi_{i+1}}\le\|Q_i\|_{\xi_i}+|\varepsilon|^{2^i}L_i\,,\\ \quad\|\phi_{i+1}-\mathrm{id}\|_{\xi_{i+1}}\le|\varepsilon|^{2^i}L_i\\ M_{i+1}\le M_iL_i\end{cases}\tag{C.1}$$

Observe that the definition of D, B and L_I, $|\varepsilon|^{2^j}L_j(3C\delta_j^{-1}) =: DB^j|\varepsilon|^{2^j}M_j$, so that $L_i < DB^iM_i$, thus by the second line in $(C.1)$, for any $0\le i\le j-1$, $|\varepsilon|^{2^{i+1}}M_{i+1} < DB_i(M_i|\varepsilon|^{2^i})^2$, which iterated, yields (66) for $0\le i\le j$. Next, we show that, thanks to (65), $(*)$ holds also for $i=j$ (and this means that Kolmogorov's step can be iterated an infinite number of times). In fact, by $(*)$ and the definition of C in (64): $|E_j|\le|E|+\sum_{i=0}^{j-1}\varepsilon_0^{2^i}L_i\le|E|+\frac13\sum_{i\ge0}\delta_i < |E|+\frac16\sum_{i\ge1}2^{-i} < |E|+1 < C$. The bounds for $\|Q_i\|$ and $\|T_i\|$ are proven in an identical manner. Now, by $(66)_{i=j}$ and (65), $|\varepsilon|^{2^j}L_j(3\delta_j^{-1}) = DB^j|\varepsilon|^{2^j}M_j \le DB^j(DB\varepsilon_0M)^{2^j}/(DB^{j+1}) \le 1/B < 1$, which implies the second inequality in $(*)$ with $i=j$; the proof of the induction is finished and one can construct an *infinite sequence* of Kolmogorov transformations satisfying $(*)$, $(C.1)$ and (66) *for all* $i\ge0$. To check (67), we observe that $|\varepsilon|^{2^i}L_i = \delta_0/(3\cdot2^i)DB^i|\varepsilon|^{2^i}M_i \le (1/2^{i+1})(|\varepsilon|DBM)^{2^i} \le (|\varepsilon|DBM/2)^{i+1}$ and therefore $\sum_{i\ge0}|\varepsilon|^{2^i}L_i\le\sum_{i\ge1}(|\varepsilon|DBM/2)^i\le|\varepsilon|DBM$. Thus, $\|Q-Q_*\|_{\xi_*}\le\sum_{i\ge0}\|\tilde Q_i\|_{\xi_i}\le|\varepsilon|^{2^i}L_i\le|\varepsilon|DBM$; and analogously for $|E-E_*|$ and $\|T-T_*\|$. To estimate $\|\phi_*-\mathrm{id}\|_{\xi_*}$, observe that $\|\Phi_i-\mathrm{id}\|_{\xi_i}\le\|\Phi_{i-1}\circ\phi_i-\phi_i\|_{\xi_i}+\|\phi_i-\mathrm{id}\|_{\xi_i}\le\|\Phi_{i-1}-\mathrm{id}\|_{\xi_{i-1}}+|\varepsilon|^{2^i}L_i$, which iterated yields $\|\Phi_i-\mathrm{id}\|_{\xi_i}\le\sum_{k=0}^i|\varepsilon|^{2^k}L_k\le|\varepsilon|DBM$: taking the limit over i completes the proof of (67) and the proof of Claim **C**.

26 In fact, observe: (i) given any integer vector $0\ne n\in\mathbb{Z}^d$ with $d\ge2$, one can find $0\ne m\in\mathbb{Z}^d$ such $n\cdot m=0$; (ii) the set $\{tn\colon t>0 \text{ and } n\in\mathbb{Z}^d\}$ is dense in $\mathbb{R}^d$; (iii) if U is a neighborhood of y_0, then $K_y(U)$ is a neighborhood of $\omega=K_y(y_0)$. Thus, by (ii) and (iii), in $K_y(U)$ there are infinitely many points of the form tn with $t>0$ and $n\in\mathbb{Z}^d$ to which correspond points $y(t,n)\in U$ such that $K_y(y(t,n)) = tn$ and for any of such points one can find, by (i), $m\in\mathbb{Z}$ such that $m\cdot n=0$, whence $K_y(y(t,n))\cdot m = tn\cdot m = 0$.

27 This fact was well known to Poincaré, who based on the above argument his non-existence proof of integral of motions in the general situation; compare Sect. 7.1.1, [6].

28 Compare (90) but observe, that, since $\hat P$ is a trigonometric polynomial, in view of Remark 9–(ii), g in (96) defines a real-analytic function on $D(y_0,\bar r)\times\mathbb{T}^d_{\xi'}$ with a suitable $\bar r=\bar r(\varepsilon)$ and $\xi'<\xi$. Clearly it is important to see explicitly how the various quantities depend upon ε; this is shortly discussed after Proposition 2.

29 In fact: $\|\check P\|_{r,\xi-\delta/2}\le M\sum_{|n|>N}\mathrm{e}^{-|n|\delta/2}\le M\mathrm{e}^{-(\delta/4)N}\sum_{|n|>N}\mathrm{e}^{-|n|\delta/4}\le M\mathrm{e}^{-(\delta/4)N}\sum_{|n|>0}\mathrm{e}^{-|n|\delta/4}\le \mathrm{const}\,M\mathrm{e}^{-(\delta/4)N}\delta^{-d}\le|\varepsilon|M$ if (106) holds and N is taken as in (104).

30 Apply the IFT of Appendix "A The Classical Implicit Function Theorem" to $F(y,\eta) := K_y(y) + \eta\partial_y P_0(y) - K_y(y_0)$ defined on $D^d(y_0,\bar{r}) \times D^1(0,|\varepsilon|)$: using the mean value theorem, Cauchy estimates and (114), $\|\mathbb{1}_d - TF_y\| \le \|\mathbb{1}_d - TK_{yy}\| + |\varepsilon|\|\partial_y^2 P_0\| \le \|T\|\|K_{yyy}\|\bar{r} + \|T\||\varepsilon|\|\partial_y^2 P_0\| \le C^2 2\bar{r}/r + C|\varepsilon|4/r^2 M \le \frac{1}{4} + \frac{1}{8} < \frac{1}{2}$; also: $2\|T\| \|F(y_0,\eta\| = 2\|T\||\eta|\partial_y P_0(y_0)\| < 2C|\varepsilon|M2/r \le 2CM\bar{r}^{-1}|\varepsilon| < \frac{1}{4}\bar{r}$ (where last inequality is due to (114)), showing that conditions (138) are fulfilled. Equation (111) comes from (139) and (113) follows easily by repeating the above estimates.

31 Recall note 18 and notice that $(\mathbb{1}_d + A)^{-1} = \mathbb{1}_d + D$ with $\|D\| \le \|A\|/(1-\|A\|) \le 2\|A\| \le 20C^3M|\varepsilon|$, where last two inequalities are due to (113).

32 Lemma 1 can be immediately extended to the y'-dependent case (which appear as a dummy parameter) as far as the estimates are uniform in y' (which is the case).

33 By (118) and (54), $|\varepsilon|\|g_x\|_{\bar{r},\bar{\xi}} \le |\varepsilon|rL \le r/2$ so that, by (116), if $y' \in D_{\bar{r}/2}(y_1)$, then $y' + \varepsilon g_x(y', \varphi(y',x')) \in D_r(y_0)$.

34 The first requirement in (123) is equivalent to require that $r_0 \le r$, which implies that if $\bar{r}$ is *defined* as the r.h.s. of (108), then $\bar{r} \le r/2$ as required in (110). Next, the first requirement in (114) at the $(j+1)$th step of the iteration translates into $16C^2 r_{j+1}/r_j \le 1$, which is satisfied, since, by definition, $r_{j+1}/r_j = (1/(2\gamma))^{\tau+1} \le (1/(2\gamma))^2 = 1/(36C^2) < 1/(16C^2)$. The second condition in (114), which at the $(j+1)$th step, reads $2CM_j r_{j+1}^{-2}|\varepsilon|^{2^j}$ is implied by $|\varepsilon|^{2^j}L_j \le \delta_j/(3C)$ (corresponding to (54)), which, in turn, is easily controlled along the lines explained in note 25.

35 An area-preserving twist mappings of an annulus $A = [0,1] \times \mathbb{S}^1$, $(\mathbb{S}^1 = \mathbb{T}^1)$, is a symplectic diffeomorphism $f = (f_1, f_2)\colon (y,x) \in A \to f(y,x) \in A$, leaving invariant the boundary circles of A and satisfying the twist condition $\partial_y f_2 > 0$ (i. e., f twists clockwise radial segments). The theory of area preserving maps, which was started by Poincaré (who introduced such maps as section of the dynamics of Hamiltonian systems with two degrees of freedom), is, in a sense, the simplest nontrivial Hamiltonian context. After Poincaré the theory of area-preserving maps became, in itself, a very rich and interesting field of Dynamical Systems leading to very deep and important results due to Herman, Yoccoz, Aubry, Mather, etc; for generalities and references, see, e. g., [33].

36 It is not necessary to assume that K is real-analytic, but it simplify a little bit the exposition. In our case, we shall see that ℓ is related to the number σ in (66). We recall the definition of Hölder norms: If $\ell = \ell_0 + \mu$ with $\ell_0 \in \mathbb{Z}_+$ and $\mu \in (0,1)$, then $\|f\|_{C^\ell} := \|f\|_{C^\ell} + \sup_{|\alpha|=\ell_0}\sup_{0<|x-y|<1}|\partial^\alpha f(x) - \partial^\alpha f(y)|/|x-y|^\mu$; $C^\ell(\mathbb{R}^d)$ denotes the Banach space of functions with finite C^ℓ norm.

37 To obtain these new estimates, one can, first replace θ by $\sqrt{\theta}$ and then use the remark in the note 21 with $p = 1$; clearly the constant σ has to be increased by one unit with respect to the constant σ appearing in (69).

38 For general references and discussions about Lemma 2 and 3, see, [44] and [65]; an elementary detailed proof can be found, also, in [15].

39 **Proof of Claim M** The first step of the induction consists in constructing $\Phi_0 = \phi_0$: this follows from Kolmogorov's Theorem (i. e., Remark 7–(i) and Remark 11) with $\xi = \xi_1 = 1/2$ (assume, for simplicity, that Q is analytic on W_1 and note that $|\varepsilon|\|P^0\|_{\xi_1} \le |\varepsilon|\|P\|_{C^0}$ by the first inequality in (124)). Now, assume that (128) and (129) holds together with $C_i < 4C$ and $\|\partial(\Phi_i - \mathrm{id})\|_{\alpha\xi_{i+1}} < (\sqrt{2}-1)$ for $0 \le i \le j$ ($C_0 = C$ and C_i are as in (64) for, respectively, $K_0 := K$ and K_i). To determine ϕ_{j+1}, observe that, by (128), one has $H_{j+1}\circ\Phi_{j+1} = (K_{j+1}+\varepsilon P_{j+1})\circ\phi_{j+1}$ where $P_j := (P^{j+1} - P^j) \circ \Phi_j$, which is real-analytic on $W_{\alpha\xi_{j+1}}$; thus we may apply Kolmogorov's Theorem to $K_{j+1} + \varepsilon P_{j+1}$ with $\xi = \alpha\xi_{j+1}$ and $\theta = \alpha$; in fact, by the second inequality in (124), $\|P_{j+1}\|_{\alpha\xi_{j+1}} \le \|P^{j-1} - P^j\|_{X_{j+1}} \le c\|P\|_{C^\ell}\xi_{j+1}^\ell$ and the smallness condition (66) becomes $|\varepsilon|D\xi_{j+1}^{\ell-\sigma}$ (with $D := c_* c\|P\|_{C^\ell}(4C)^b 2^{\sigma/2}$), which is clearly satisfied for $|\varepsilon| < D^{-1}$. Thus, ϕ_{j+1} has been determined and (notice that $\alpha^2\xi_{j+1} = \xi_{j+1}/2 = \xi_{j+2}$) $\|\phi_{j+1} - \mathrm{id}\|_{\xi_{j+2}}$, $\partial(\|\phi_{j+1} - \mathrm{id})\|_{\xi_{j+2}} \le |\varepsilon|D\xi_{j+1}$. Let us now check the domain constraint $\Phi_j\colon W_{\alpha\xi_{j+1}} \to X_{\xi_{j+1}}$. By the inductive assumptions and the real-analyticity of Φ_j, one has that, for $z \in W_{\alpha\xi_{j+1}}$, $|\mathrm{Im}\Phi_j(z)| = |\mathrm{Im}(\Phi_j(z) - \Phi_j(\mathrm{Re}z))| \le |\Phi_j(z) - \Phi_j(\mathrm{Re}z)| \le \|\partial\Phi_j\|_{\alpha\xi_{j+1}}|\mathrm{Im}z| \le (1 + \|\partial(\Phi_i - \mathrm{id})\|_{\alpha\xi_{i+1}})\alpha\xi_{j+1} < \sqrt{2}\alpha\xi_{j+1} = \xi_{j+1}$ so that $\Phi_j\colon W_{\alpha\xi_{j+1}} \to X_{\xi_{j+1}}$. The remaining inductive assumptions in (129) with j replaced by $j+1$ are easily checked by arguments similar to those used in the induction proof of Claim **C** above.

40 See, e. g., the Proposition at page 58 of [14] with $g_j = f_j - f_{j-1}$. In fact, the lemma applies to the Hamiltonians H_j and to the symplectic map ϕ_j in (82) in Arnold's scheme with W_j in (81) and taking $C = C_* := \{y' = \lim_{j\to\infty} y_j(\omega)\colon \omega \in B \cap K_y^{-1}(\mathcal{D}^d_{\kappa,\tau})\}$ and $y_j(\omega) := y_j$ is as in (82).

41 A formal ε-power series quasi-periodic trajectory, with rationally-independent frequency ω, for a nearly-integrable Hamiltonian $H(y,x;\varepsilon) := K(y) + \varepsilon P(y,x)$

is, by definition, a sequence of functions $\{z_k\} := (\{v_k\}, \{u_k\})$, real-analytic on $\mathbb{T}^d$ and such that $D_\omega z_k = J_{2d}\pi_k(\nabla H(\sum_{j=0}^{k-1}\varepsilon^j z_j))$ where $\pi_k(\cdot) := \frac{1}{k!}\partial_\varepsilon^k(\cdot)|_{\varepsilon=0}$; compare Remark 1–(ii) above.

42 In fact, Poincaré was not at all convinced of the convergence of such series: see chapter XIII, n° 149, entitled "Divergence des séries de M. Lindstedt", of his book [49].

43 Equation (70) guarantees that the map from y in the $(d-1)$-dimensional manifold $\{K = E\}$ to the $(d-1)$-dimensional real projective space $\{\omega_1 : \omega_2 : \cdots : \omega_d\} \subset \mathbb{RP}^{d-1}$ (where $\omega_i = K_{y_i}$) is a diffeomorphism. For a detailed proof of the "iso-energetic KAM Theorem", see, e. g., [24].

44 Actually, it is not known if such tori are KAM tori in the sense of the definitions given above!

45 The first example of a nearly-integrable system (with two parameters) exhibiting Arnold's diffusion(in a certain region of phase space) was given by Arnold in [4]; a theory for "a priori unstable systems" (i. e., the case in which the integrable system carries also a partially hyperbolic structure) has been worked out in [20] and in recent years a lot of literature has been devoted to study the "a priori unstable" case and to to try to attack the general problem (see, e. g., Sect. 6.3.4 of [6] for a discussion and further references). We mention that J. Mather has recently announced a complete proof of the conjecture in a general case [40].

46 Here, we mention briefly a different and very elementary connection with classical mechanics. To study the spectrum $\sigma(L)$ (L as above with a quasi-periodic potential $V(\omega_1 t, \dots, \omega_n t)$) one looks at the equation $\ddot{q} = (V(\omega t) - \lambda)q$, which is the q-flow of the Hamiltonian ϕ_H^t $H = H(p, q, I, \varphi; \lambda) := p^2/2 + [\lambda - V(\varphi)]q^2/2$ where $(p, q) \in \mathbb{R}^2$ and $(I, \varphi) \in \mathbb{R}^n \times \mathbb{T}^n$ w.r.t. the standard form $\mathrm{d}p \wedge \mathrm{d}q + \mathrm{d}I \wedge \mathrm{d}\varphi$ and λ is regarded as a parameter. Notice that $\dot{\varphi} = \omega$ so that $\varphi = \varphi_0 + \omega t$ and that the (p, q) decouples from the I-flow, which is, then, trivially determined one the (p, q) flow is known. Now, the action-angle variables(J, θ) for the harmonic oscillator $p^2/2 + \lambda q^2/2$ are given by $J = r^2/\sqrt{\lambda}$ and (r, θ) are polar coordinates in the $(p, \sqrt{\lambda}q)$-plane; in such variables, H takes the form $H = \omega \cdot I + \sqrt{\lambda} J - V(\varphi)/\sqrt{\lambda}\sin^2\theta$. Now, if, for example V is small, this Hamiltonian is seen to be a perturbation of $(n+1)$ harmonic oscillator with frequencies $(\omega, \sqrt{\lambda})$ and it is remarkable that one can provide a KAM scheme, which preserves the linear-in-action structure of this Hamiltonian and selects the (Cantor) set of values of the frequency $\alpha = \sqrt{\lambda}$ for which the KAM scheme can be carried out so as to conjugate H to a Hamiltonian of the form $\omega \cdot I + \alpha J$, proving the existence of (generalized) quasi-periodic eigen-functions. For more details along these lines, see [14].

47 The value 10^{-52} is about the proton-Sun mass ratio: the mass of the Sun is about $1.991 \cdot 10^{30}$ kg, while the mass of a proton is about $1.672 \cdot 10^{-21}$ kg, so that (mass of a proton)/(mass of the Sun) $\simeq 8.4 \cdot 10^{-52}$.

48 "Computer-assisted proofs" are mathematical proofs, which use the computers to give rigorous upper and lower bounds on chains of long calculations by means of so-called "interval arithmetic"; see, e. g., Appendix C of [13] and references therein.

49 Simple examples of such orbits are equilibria and periodic orbits: in such cases there are no small-divisor problems and existence was already established by Poincaré by means of the standard Implicit Function Theorem; see [49], Volume I, chapter III.

50 Typically, ξ may indicate an initial datum y_0 and y the distance from such point or (equivalently, if the system is non-degenerate in the classical Kolmogorov sense) $\xi \to \omega(\xi)$ might be simply the identity, which amounts to consider the unperturbed frequencies as parameter; the approach followed here is that in [51], where, most interestingly, m is allowed to be ∞.

51 I. e., a map $\Phi \colon X \to X$ for which $\exists 0 < \alpha < 1$ such that $d(\Phi(u), \Phi(v)) \le \alpha d(u, v)$, $\forall u, v \in X$, $d(\cdot, \cdot)$ denoting the metric on X; for generalities on metric spaces, see, e. g., [36].

52 $\Phi^j = \Phi \circ \cdots \circ \Phi$ j-times. In fact, let $u_j := \Phi^j(u_0)$ and notice that, for each $j \ge 1$ $d(u_{j+1}, u_j) \le \alpha d(u_j, u_{j-1}) \le \alpha^j d(u_1, u_0) =: \alpha^j\beta$, so that, for each $j, h \ge 1$, $d(u_{j+h}, u_j) \le \sum_{i=0}^{h-1} d(u_{j+i+1}, u_{j+i}) \le \sum_{i=0}^{h-1}\alpha^{j+i}\beta \le \alpha^j\beta/(1-\alpha)$, showing that $\{u_j\}$ is a Cauchy sequence. Uniqueness is obvious.

Bibliography

Primary Literature

1. Ahlfors LV (1978) Complex analysis. An introduction to the theory of analytic functions of one complex variable, 3rd edn. International Series in Pure and Applied Mathematics. McGraw–Hill, New York, xi+331 pp
2. Arnold VI (1963) Proof of a Theorem by Kolmogorov AN on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian. Russ Math Surv 18:13–40
3. Arnold VI (1963) Small divisor problems in classical and Celestial Mechanics. Russ Math Surv 18:85–191
4. Arnold VI (1964) Instability of dynamical systems with many degrees of freedom (Russian). Dokl Akad Nauk SSSR 156:9–12
5. Arnold VI (1974) Mathematical methods of classical mechanics series: graduate texts in mathematics, vol 60. Originally published by Nauka, Moscow, 1974. 2nd edn. 1989. Corr. 4th printing, 1997, pp XVI,509

6. Arnold VI, Kozlov VV, Neishtadt AI (ed) (2006) Mathematical Aspects of Classical and Celestial Mechanics, Dynamical Systems III, 3rd edn. Series: Encyclopedia of Mathematical Sciences, vol 3. Springer, Berlin
7. Avila A, Krikorian R (2006) Reducibility or nonuniform hyperbolicity for quasiperiodic Schrödinger cocycles. Ann Math 164:911–940
8. Broer HW (2004) KAM theory: the legacy of AN Kolmogorov's 1954 paper. Comment on: "The general theory of dynamic systems and classical mechanics". (French in: Proceedings of the International Congress of Mathematicians, Amsterdam, 1954, vol 1, pp 315–333, Erven P, Noordhoff NV, Groningen, 1957). Bull Amer Math Soc (N.S.) 41(4):507–521 (electronic)
9. Broer HW, Huitema GB, Sevryuk MB (1996) Quasi-periodic motions in families of dynamical systems. Order amidst chaos. Lecture Notes in Mathematics, vol 1645. Springer, Berlin
10. Celletti A, Chierchia L (1987) Rigorous estimates for a computer-assisted KAM theory. J Math Phys 28:2078–2086
11. Celletti A, Chierchia L (1988) Construction of analytic KAM surfaces and effective stability bounds. Commun Math Phys 118:119–161
12. Celletti A, Chierchia L (1995) A Constructive Theory of Lagrangian Tori and Computer-assisted Applications. In: Jones CKRT, Kirchgraber U, Walther HO (eds) Dynamics Reported, vol 4 (New Series). Springer, Berlin, pp 60–129
13. Celletti A, Chierchia L (2007) KAM stability and celestial mechanics. Mem Amer Math Soc, 187(878), viii+134 pp
14. Chierchia L (1986) Quasi-periodic Schrödinger operators in one dimension, Absolutely continuous spectra, Bloch waves and integrable Hamiltonian systems. Quaderni del Consiglio Nazionale delle Ricerche, Firenze, Italia, 127 pp http://www.mat.uniroma3.it/users/chierchia/REPRINTS/CNR86.pdf
15. Chierchia L (2003) KAM Lectures. In: Dynamical Systems. Part I: Hamiltonian Systems and Celestial Mechanics. Pubblicazioni della Classe di Scienze, Scuola Normale Superiore, Pisa, pp 1–56. Centro di Ricerca Matematica "Ennio De Giorgi"
16. Chierchia L (2006) KAM theory and Celestial Mechanics. In: Françoise J-P, Naber GL, Tsou ST (eds) Encyclopedia of Mathematical Physics, vol 3. Elsevier, Oxford, pp 189–199. ISBN 978-0-1251-2666-3
17. Chierchia L (2008) A.N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. A comment on: On conservation of conditionally periodic motions for a small change in Hamilton's function [Dokl Akad Nauk SSSR (N.S.) 98:527–530 (1954)]. Regul Chaotic Dyn 13(2):130–139
18. Chierchia L, Falcolini C (1994) A direct proof of a theorem by Kolmogorov in Hamiltonian systems. Ann Sc Norm Su Pisa IV(XXI):541–593
19. Chierchia L, Gallavotti G (1982) Smooth prime integrals for quasi-integrable Hamiltonian systems. Il Nuovo Cimento 67(B)277–295
20. Chierchia L, Gallavotti G (1994) Drift and diffusion in phase space. Ann Inst Henri Poincaré, Phys Théor 60:1–144; see also Erratum, Ann Inst Henri Poincaré, Phys Théor 68(1):135 (1998)
21. Chierchia L, Perfetti P (1995) Second order Hamiltonian equations on $\mathbb{T}^\infty$ and almost-periodic solutions. J Differ Equ 116(1):172–201
22. Chierchia L, Qian D (2004) Moser's theorem for lower dimensional tori. J Differ Equ 206:55–93
23. Coddington EA, Levinson N (1955) Theory of Ordinary Differential Equations. McGraw–Hill, New York. 14+429 pp
24. Delshams A, Gutierrez P (1996) Effective stability and KAM theory. J Differ Equ 128(2):415–490
25. Eliasson L (1988) Perturbations of stable invariant tori for Hamiltonian systems. Ann Scuola Norm Sup Pisa Cl Sci 15:115–147
26. Eliasson LH (1992) Floquet solutions for the 1-dimensional quasi-periodic Schrödinger equation. Comm Math Phys 146:447–482
27. Eliasson LH (1996) Absolutely convergent series expansion for quasi-periodic motions. Math Phys Electron J 2(Paper 4):33 (electronic); see also: Report 2–88, Dept. of Math., University of Stockholm (1988)
28. Evans LC, Gariepy RF (1992) Measure theory and fine properties of functions. Studies in Advanced Mathematics. CRC Press, Boca Raton, pp viii+268
29. Féjoz J (2004) Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman). (French) Ergod Theor Dyn Syst 24(5):1521–1582
30. Gallavotti G (1994) Twistless KAM tori, quasi flat homoclinic intersections and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review. Rev Math Phys 6:343–411
31. Graff S (1974) On the continuation of stable invariant tori for Hamiltonian systems, J Differ Equ 15:1–69
32. Hénon M (1966) Explorationes numérique du problème restreint IV: Masses egales, orbites non periodique. Bull Astronom 3(1, fasc. 2):49–66
33. Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. In: Encyclopedia of Mathematics and its Applications, vol 54, Cambridge University Press, Cambridge, xviii+802 pp
34. Katznelson Y (2004) An introduction to harmonic analysis, 3rd edn. Cambridge Mathematical Library. Cambridge University Press, Cambridge, pp xviii+314 pp
35. Kolmogorov AN (1954) On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian. Dokl Akad Nauk SSR 98:527–530
36. Kolmogorov AN, Fomin SV (1999) Elements of the Theory of Functions and Functional Analysis. Courier Dover Publications, 288 pp
37. Kuksin SB (1988) Perturbation theory of conditionally periodic solutions of infinite-dimensional Hamiltonian systems and its applications to the Korteweg–de Vries equation. (Russian) Mat Sb (N.S.) 136,178(3):396–412,431; translation in Math USSR-Sb 64(2):397–413
38. Kuksin SB (2004) Fifteen years of KAM for PDE. Geometry, topology, and mathematical physics. Am Math Soc Transl Ser 212(2):237–258
39. Mather J (1984) Nonexistence of invariant circles. Erg Theor Dyn Syst 4:301–309
40. Mather JN (2003) Arnold diffusion. I. Announcement of results. (Russian) Sovrem Mat Fundam Napravl 2:116–130 (electronic); translation in J Math Sci 124(5):5275–5289
41. Melnikov VK (1965) On certain cases of conservation of almost periodic motions with a small change of the Hamiltonian function. (Russian) Dokl Akad Nauk SSSR 165:1245–1248
42. Moser JK (1961) A new technique for the construction of solutions of nonlinear differential equations. Proc Nat Acad Sci USA 47:1824–1831

43. Moser JK (1962) On invariant curves of area-preserving mappings of an annulus. Nach Akad Wiss Göttingen, Math Phys Kl II 1:1–20
44. Moser J (1966) A rapidly convergent iteration method and non-linear partial differential equations. Ann Scuola Norm Sup Pisa 20:499–535
45. Moser J (1967) Convergent series expansions for quasi-periodic motions. Math Ann 169:136–176
46. Moser J (1970) On the construction of almost periodic solutions for ordinary differential equations. In: Proc. Internat. Conf. on Functional Analysis and Related Topics, Tokyo, 1969. University of Tokyo Press, Tokyo, pp 60–67
47. Moser J (1983) Integrable Hamiltonian systems and spectral theory, Lezioni Fermiane. [Fermi Lectures]. Scuola Normale Superiore, Pisa, iv+85 pp
48. Nash J (1956) The imbedding problem for Riemannian manifolds. Ann Math 63(2):20–63
49. Poincaré H (1892) Les méthodes nouvelles de la Mécanique céleste. Tome 1: Gauthier–Villars, Paris, 1892, 385 pp; tome 2: Gauthier–Villars, Paris 1893, viii+479 pp; tome 3: Gauthier–Villars, Paris, 1899, 414 pp
50. Pöschel J (1982) Integrability of Hamiltonian sytems on Cantor sets. Comm Pure Appl Math 35:653–695
51. Pöschel J (1996) A KAM-theorem for some nonlinear PDEs Ann Scuola Norm Sup Pisa Cl Sci 23:119–148
52. Pyartli AS (1969) Diophantine approximations of submanifolds of a Euclidean space. Funktsional Anal Prilozhen 3(4):59–62 (Russian). English translation in: Funct Anal Appl 3:303–306
53. Rüssmann H (1970) Kleine Nenner. I. Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes. (German) Nachr Akad Wiss Göttingen Math-Phys Kl II 1970:67–105
54. Rüssmann H (1975) On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In: Moser J (ed) Dynamical Systems, Theory and Applications. Lecture Notes in Physics vol 38. Springer, Berlin, pp 598–624
55. Rüssmann H (1976) Note on sums containing small divisors. Comm Pure Appl Math 29(6):755–758
56. Rüssmann H (1980) On the one-dimensional Schrödinger equation with a quasiperiodic potential. In: Nonlinear dynamics, Internat. Conf., New York, 1979, Ann New York Acad Sci, vol 357. New York Acad Sci, New York, pp 90–107
57. Rüssmann H (1989) Nondegeneracy in the perturbation theory of integrable dynamical systems, Number theory and dynamical systems (York, 1987). London Math Soc. Lecture Note Ser, vol 134. Cambridge University Press, Cambridge, pp 5–18
58. Rüssmann H (2001) Invariant tori in non-degenerate nearly integrable Hamiltonian systems. Regul Chaot Dyn 6(2):119–204
59. Salamon D (2004) The Kolmogorov–Arnold–Moser theorem. Math Phys Electron J 10(Paper 3):37 (electronic)
60. Sevryuk MB (2003) The Classical KAM Theory and the dawn of the twenty-first century. Mosc Math J 3(3):1113–1144
61. Siegel CL (1942) Iteration of analytic functions. Ann Math 43:607–612
62. Spivak M (1979) A Comprehensive Introduction to Differential Geometry, vol 1, 2nd edn. Publish or Perish Inc, Wilmington
63. Wayne CE (1990) Periodic and quasi-periodic solutions for nonlinear wave equations via KAM theory. Commun Math Phys 127:479–528
64. Whitney H (1934) Analytic extensions of differentiable functions defined in closed sets. Trans Amer Math Soc 36(1):63–89
65. Zehnder E (1975) Generalized implicit function theorems with applications to some small divisor problems, vol I. Comm Pure Appl Math 28:91–14
66. Zehnder E (1976) Generalized implicit function theorems with applications to some small divisor problems, vol II. Comm Pure Appl Math 29(1):49–111

Books and Reviews

Arnold VI, Novikov SP (eds) (1985) Dynamical Systems IV, Symplectic Geometry and its Applications. In: Encyclopaedia of Mathematical Sciences, vol 4, Volume package Dynamical Systems. Original Russian edition published by VINITI, Moscow. 2nd expanded and rev. edn., 2001

Berretti A, Celletti A, Chierchia L, Falcolini C (1992) Natural Boundaries for Area-Preserving Twist Maps. J Stat Phys 66:1613–1630

Bost JB (1984/1985) Tores invariants des systèmes dynamiques hamiltoniens. Séminaire Bourbaki, exp. 639

Bourgain J (1997) On Melnikov's persistency problem. Math Res Lett 4(4):445–458

Broer HW, Sevryuk MB KAM theory: Quasi-periodicity in dynamical systems. In: Handbook of Dynamical Systems. Elsevier, to appear

Celletti A, Chierchia L (2006) KAM tori for *N*-body problems: a brief history. Celest Mech Dynam Astronom 95(1–4):117–139

Chierchia L, Falcolini C (1996) Compensations in small divisor problems. Comm Math Phys 175:135–160

Gallavotti G (1983) The Elements of Mechanics. Springer, New York

Giorgilli A, Locatelli U (1997) On classical series expansion for quasi-periodic motions. MPEJ 3(5):1–25

Herman M (1986) Sur les courbes invariantes pour les difféomorphismes de l'anneau. Astérisque 2:248

Herman M (1987) Recent results and some open questions on Siegel's linearization theorems of germs of complex analytic diffeomorphisms of $\mathbf{C}^n$ near a fixed point. In: VIIIth International Congress on Mathematical Physics, Marseille, 1986. World Sci. Publishing, Singapore, pp 138–184

Herman M (1989) Inégalités a priori pour des tores Lagrangiens invariants par des difféomorphismes symplectiques. Publ Math IHÈS 70:47–101

Hofer H, Zehnder E (1995) Symplectic invariants and Hamiltonian dynamics. The Floer memorial volume. Progr. Math., 133. Birkhäuser, Basel, pp 525–544

Kappeler T, Pöschel J (2003) KdV & KAM. Springer, Berlin, XIII,279 pp

Koch H (1999) A renormalization group for Hamiltonians, with applications to KAM tori. Ergod Theor Dyn Syst 19:475–521

Kuksin SB (1993) Nearly integrable infinite dimensional Hamiltonian systems. Lecture Notes in Mathematics, vol 1556. Springer, Berlin

Lanford OE III (1984) Computer assisted proofs in analysis. Phys A 124:465–470

Laskar J (1996) Large scale chaos and marginal stability in the solar system. Celest Mech Dyn Astronom 64:115–162

de la Llave R (2001) A tutorial on KAM theory. Smooth ergodic theory and its applications. In: Proc. Sympos. Pure Math., vol 69, Seattle, WA, 1999. Amer Math Soc, Providence, pp 175–292

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

de la Llave R, Rana D (1990) Accurate strategies for small divisor problems. Bull Amer Math Soc (N.S.) 22:(1)85–90
Lazutkin VF (1993) KAM theory and semiclassical approximations to eigenfunctions. Springer, Berlin
MacKay RS (1985) Transition to chaos for area-preserving maps. Lect Notes Phys 247:390–454
Meyer KR, Schmidt D (eds) (1991) Computer Aided Proofs in Analysis. Springer, Berlin
Moser J (1973) Stable and random motions in dynamical systems. With special emphasis on celestial mechanics. Hermann Weyl Lectures, the Institute for Advanced Study, Princeton, N. J Annals of Mathematics Studies, No. 77. Princeton University Press, Princeton; University of Tokyo Press, Tokyo
Pöschel J (2001) A lecture on the classical KAM theorem. Smooth ergodic theory and its applications. In: Proc Sympos Pure Math, vol 69 Seattle,WA, 1999. Amer Math Soc, Providence, pp 707–732
Salamon D, Zehnder E (1989) KAM theory in configuration space. Comment Math Helvetici 64:84–132
Sevryuk MB (1995) KAM-stable Hamiltonians. J Dyn Control Syst 1(3):351–366
Sevryuk MB (1999) The lack-of-parameters problem in the KAM theory revisited, Hamiltonian systems with three or more degrees of freedom (S'Agaró, 1995). NATO Adv Sci Inst Ser C Math Phys Sci, vol 533. Kluwer, Dordrecht, pp 568–572
Siegel CL, Moser JK (1971) Lectures on Celestial Mechanics. Springer, Berlin
Weinstein A (1979) Lectures on symplectic manifolds, 2nd edn., CBMS Regional Conference Series in Mathematics, vol 29. American Mathematical Society, Providence
Xu J, You J, Qiu Q (1997) Invariant tori for nearly integrable Hamiltonian systems with degeneracy. Math. Z 226(3):375–387
Yoccoz J-C (1992) Travaux de Herman sur les tores invariants. Séminaire Bourbaki, 1991/92, Exp. 754. Astérisque 206:311–344
Yoccoz J-C (1995) Théorème de Siegel, pôlynomes quadratiques et nombres de Brjuno. Astérisque 231:3–88

Korteweg–de Vries Equation (KdV), Different Analytical Methods for Solving the

YU-JIE REN[1,2,3]
[1] Department of Mathematics, Shantou University, Shantou, People's Republic of China
[2] Department of Mathematics and Physics, Dalian Polytechnic University, Dalian, People's Republic of China
[3] Beijing Institute of Applied Physics and Computational Mathematics, Beijing, People's Republic of China

Article Outline

Glossary

Analytical method for solving an equation The method for obtaining the exact solutions of an equation.

Solitons The special solitary waves which retain their original shapes and speeds after collision and exhibit only a small overall phase shift.

ε ± 1

t the time

(x, y) Cartesian coordinates of a point

∂_x^{-1} an indefinite integrate operator $\int \mathrm{d}x$.

Definition of the Subject

This paper presents the analytical methods for obtaining the exact solutions of the Korteweg–de Vries Equation (KdV equation).

The KdV equation and its exact solutions can describe and explain many physical problems. In addition, it is a typical, relatively simple and classical equation among the many nonlinear equations in physics. Much of the literature of nonlinear equation theory customarily uses solving soliton solutions of the KdV equation as an example to introduce the nonlinear theory, method and character of soliton solutions. During the last five decades, the construction of exact solution for a wide class of nonlinear equations has been an exciting and extremely active area of research. This includes the most famous nonlinear example of the KdV equation.

In the family of the KdV equations, a well known KdV equation is expressed in its simplest form as [1,2,3,4]:

$$u_t(x,t) + \alpha u(x,t) u_x(x,t) + \beta u_{xxx}(x,t) = 0\,. \qquad (1)$$