KAM Theory
for
finite and infinite dimensional systems

Luigi Chierchia, Michela Procesi

Dipartimento di Matematica e Fisica , Università “Roma Tre”

Largo S. L. Murialdo 1, I-00146 Roma (Italy)

luigi@mat.uniroma3.it, procesi@mat.unrioma3.it

http://www.mat.uniroma3.it/users/chierchia

http://www.mat.uniroma3.it/users/procesi

February 28, 2018

Contents

Glossary

Action–angles variables

A particular set of variables $(y, x) = ((y_1, ..., y_d), (x_1, ..., x_d))$, x_i ("angles") defined modulus 2π, particularly suited to describe the general behavior of a finite dimensional 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 differential equation on a symplectic space/manifold (in the finite dimensional case, 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 $f = (f_1, ..., f_d)$ 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 space/manifold.

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, 1937–2010) 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.

Integrable Hamiltonian systems

Very special class of Hamiltonian systems, whose flow can be "explicitly computed" for all initial data and typically is described through a linear flow on a (in)finite dimensional torus.

Nearly–integrable Hamiltonian systems

Hamiltonian systems which are small perturbations of an integrable systems and which, in general, exhibits a much richer dynamics than the integrable limit. Nevertheless, finite dimensional KAM theory asserts that, under suitable assumptions, the majority (in measure 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 $x \in \mathbb{T}^d \mapsto x + \omega t$ with $\omega = (\omega_1, ..., \omega_d) \in \mathbb{R}^d$ called frequency vector.

Small divisors/denominators

Arbitrary small combinations of the form $\omega \cdot k := \sum_{j=1}^{d} \omega_i k_i$ with $\omega = (\omega_1, ..., \omega_d) \in \mathbb{R}^d$ 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 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.

Complex symplectic variables

The identification of the real symplectic space $(\mathbb{R}^{2d}, dp \wedge dq)$ with the complex space $\mathbb{C}^d$: one sets $z_j = (q_j + \mathrm{i} p_j)/\sqrt{2}$ for $j = 1, \ldots d$, in this way the symplectic two-form is the imaginary part of the hermitian product.

1 Definition of the subject

KAM theory is a mathematical, quantitative theory which has as primary object the persistence, under small (Hamiltonian) perturbations, of quasi–periodic trajectories of integrable Hamiltonian systems. Quasi–periodic motions 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 $\omega = (\omega_1, ..., \omega_d) \in \mathbb{R}^d$ is the set of frequencies of the trajectory (if $d = 1$, $2\pi/\omega$ is the *period* of the motion).

In finite dimensional integrable systems with bounded motions, the typical trajectory is indeed quasi–periodic and KAM theory is apt to describe the behavior of "most" initial data. In general, this is not the case in infinite dimensional systems and PDEs. Still, the search for periodic and quasi–periodic solutions is obviously an interesting and challenging task.

2 Introduction

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 1960's, who were able to overcome the formidable technical problem related to the appearance, in perturbative formulae, of arbitrarily small divisors[1]. Small divisors make impossible the use of classical analytical tools (such as the standard Implicit Function Theorem, fixed point theorems, etc.) 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, to counterbalance the divergences introduced by small divisors.

KAM theory was extended to the context of Hamiltonian PDEs starting from the early 1990's by Kuksin, Wayne, Pöschel, with the purpose of proving the existence and linear stability of *small*–quasi–periodic solutions for semi–linear PDEs with Dirichlet boundary conditions. Although there is no general theory available, the KAM approach has been successively developed in order to cover also examples of PDEs on tori and compact Lie groups, quasi-linear and fully non-linear PDEs on the circle and PDEs on the line with a coercive potential.

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. There are excellent reviews of KAM theory – especially Sect. 6.3 of [6] and [108] – 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. "Arnol'd's scheme") and Moser (Sect. "The differentiable case: Moser's Theorem").

While the tools and techniques in finite dimensions are by now quite well understood, the situation in infinite dimensions is, as can be expected, significantly more complicated and there are many fundamental open issues, such as, e.g., the "general" behavior of a "nearly-integrable" system. Therefore, in discussing the finite dimensional case, we shall try to give a quite complete and quantitative description of results and, especially, the techniques used in order to obtain them. On the other hand, in infinite dimension we mainly focus on specific examples, trying to convey the main ideas and the similarities and differences with the finite dimensional case.

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 exposition. Appendix B includes also several other complementary notes, which contain either standard material or further references or side comments.

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.

2.1 Finite dimensional context

In the finite dimensional setting, we will be concerned with Hamiltonian flows on the symplectic manifold $(\mathcal{M}, dy \wedge dx)$; for general information, see, e.g., [5] or Sect. 1.3 of [6]. Notation, main definitions and properties are listed in the following items.

(a) $\mathcal{M} := B \times \mathbb{T}^d$ with $d \geq 2$ (the case $d = 1$ is trivial for the questions addressed in this article); 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) $dy \wedge dx := \sum_{i=1}^{d} dy_i \wedge dx_i$, $(y \in B,\ x \in \mathbb{T}^d)$ is the standard symplectic form[4];

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

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

where $\dot{z} = \frac{dz}{dt}$, 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 and 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 $y(t)$ reaches eventually the border of B.

Eq.n's (1) or (2) are called the *Hamilton's equations* with *Hamiltonian* H; usually, the symplectic (or "conjugate") variables (y, x) are called *action–angles* 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 of $\mathcal{M}$ which preserve the 2–form $dy \wedge dx$; i.e., if $\phi : (\eta, \xi) \in \mathcal{M} \to (y, x) \in \mathcal{M}$ is a diffeomorphism such that $d\eta \wedge d\xi = dy \wedge dx$, then

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

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 transpose.

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 IFT (Implicit Function Theorem; see [73] or Appendix A below), the map $\phi : (y, x) \to (\eta, \xi)$ 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; 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 $\omega = (\omega_1, ..., \omega_d) \in \mathbb{R}^d$ if ω is a rationally–independent vector, i.e.,

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

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

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

is a solution of (2).

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

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

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

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

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

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

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

(f) In KAM theory a major rôle is played by the numerical properties of the frequencies ω. A typical assumption is that ω is a *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:

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

(normally, for integer vectors n, $|n|$ denotes $|n_1| + \cdots + |n_d|$, but other norms may as 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 $\mathcal{D}_{\kappa,\tau}$ or $\mathcal{D}^d_{\kappa,\tau}$, while the union over all $\kappa > 0$ of $\mathcal{D}_{\kappa,\tau}$ will be denoted by $\mathcal{D}_\tau = \mathcal{D}^d_\tau$. 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 $\mathbb{R}^d \backslash \mathcal{D}_\tau$ is zero; if $\tau = d-1$, the Lebesgue measure of $\mathcal{D}_\tau$ is zero but its intersection with any open set has the cardinality of $\mathbb{R}$. The union over all $\tau \geq d-1$ of $\mathcal{D}^d_\tau$ will be denoted by $\mathcal{D}^d$.

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

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

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

compare, e.g., Sect. 1 of [107]. 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 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 $\mathcal{T}_\omega$ is non–degenerate if it admits a non–degenerate Kolmogorov normal.

(i) Quasi–periodic solutions with $1 \leq n < d$ frequencies, i.e., solutions of (2) of the form

$$\begin{cases} y(t) = v(\omega t) \\ x(t) = U(\omega t) \end{cases} \tag{12}$$

where $v : \mathbb{T}^n \to \mathbb{R}^d$, $U : \mathbb{T}^n \to \mathbb{T}^d$ 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

$$\mathcal{T}^n_\omega := \Big\{ (y, x) \in \mathcal{M} : \ y = v(\theta) \ , \ x = U(\theta) \ , \ \ \theta \in \mathbb{T}^n \Big\} \tag{13}$$

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*.

Remark 1 (i) A classical theorem by H. Weyl says that the flow

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

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 [70], Sect. 1.4). 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 PDE's on $\mathbb{T}^d$:

$$\begin{cases} D_\omega v = -H_x(v(\theta), \theta + u(\theta)) \\ \\ \omega + D_\omega u = H_y(v(\theta), \theta + u(\theta)) \end{cases} \tag{14}$$

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

(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 for *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 $y_0 \in \mathbb{R}^d$; we shall also denote by $D(y_0, r)$ the complex ball in $\mathbb{C}^d$ of radius r centered in $y_0 \in \mathbb{C}^d$. In such a case the Hamiltonian flow is simply

$$\phi^t_K(y, x) = \Big(y, x + \omega(y)t\Big) \ , \qquad \omega(y) := K_y(y_0) := \frac{\partial K}{\partial y}(y) \ . \tag{15}$$

Thus, if the frequency map $y \in B \to \omega(y)$ is a diffeomorphism (which is guaranteed if $\det K_{yy}(x_0) \neq 0$, for some $x_0 \in B$ and B is small enough), in view of (f), for almost all initial data, the trajectories (15) 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 ω is a diffeomorphism, KAM tori persist under small 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

$$K(y) + \varepsilon P(y, x; \varepsilon) \ , \qquad 0 < \varepsilon \ll 1 \ , \tag{16}$$

was called by H. Poincarè *le problème général de la dynamique*, to which he dedicated a large part of his monumental *Méthodes Nouvelles de la Mécanique Céleste* [91].

2.2 Infinite dimensional context

In the infinite dimensional setting, we will be concerned with Hamiltonian flows on a scale of Banach or Hilbert spaces, for a more detailed presentation we refer the reader to [77, 78], [69], [64]; for properties of analytic functions on Hilbert spaces, see [97].

(A) A symplectic structure on scale of real Hilbert spaces $\mathcal{H}_s, (\cdot,\cdot)_s$ is defined by an antisymmetric morphism $J : H_s \to H_{s+d}$ of order $d \geq 0$ so that the symplectic two form defined by $\omega(u, v) := (u, J^{-1}v)_0$, is a skew-symmetric bilinear form.

(B) Given a complex Hilbert space H with a Hermitian product $\langle\cdot,\cdot\rangle$, its realification is a real symplectic Hilbert space with scalar product and symplectic form given by

$$(u, v) = 2\mathrm{Re}\langle u, v\rangle \, , \quad \omega(u, v) = 2\mathrm{Im}\langle u, v\rangle \, .$$

(C) Given a real–analytic (or smooth) function $H : O_s \subseteq \mathcal{H}_s \to \mathbb{R}$, O_s open, the *Hamiltonian flow governed by* H is, if the equation is at least locally well posed, the one–parameter family of diffeomorphisms $\phi^t_H : \mathcal{H}_s \to \mathcal{H}_s$, which to $z \in \mathcal{H}_s$ associates the solution at time t of the differential equation

$$\dot{z} = J\nabla H(z) \, , \qquad z(0) = z \in \mathcal{H}_s, \tag{17}$$

where J is the symplectic morphism and ∇H is identified through the bilinear product $(,)_0$ namely

$$dH[\cdot] = (\nabla H, \cdot)_0.$$

Note that in general in the infinite dimensional case the fact that H is smooth does not guarantee that the equation is even locally well-posed. As in the finite dimensional counterpart, the Hamiltonian H is constant over trajectories $\phi^t_H(z)$, and the constant value $E = H(\phi^t_H(z))$ is called the energy of the trajectory $\phi^t_H(z)$.

(D) A Hamiltonian H whose Hamiltonian vector field is an analytic map $\mathcal{H}_s \to \mathcal{H}_s$ is called a *regular* Hamiltonian; in this case the Hamiltonian flow is at least locally well posed.

(E) Hamilton equations are left invariant by *symplectic* (or "canonical") change of variables, i.e., by diffeomorphisms of $\mathcal{H}_s$ which preserve the 2–form ω. A classical way to generate such changes of coordinates is as the time-one flow of an auxiliary Hamiltonian function, say S. We recall that given a smooth Hamiltonian $S : O_s \to \mathbb{R}$ if the corresponding Hamilton equation

$$w_\tau = J\nabla S(w) \, , \qquad w|_{\tau=0} = z$$

is well-posed for $\tau \leq 1$ then the flow $\phi^1_H(z)$ defines a symplectic change of variables.

(F) A solution $z(t)$ of (17) is *quasi–periodic* with frequency vector $\omega = (\omega_1, ..., \omega_d) \in \mathbb{R}^d$ if ω is a rationally–independent vector, see (5), and there exists an embedding $U : \mathbb{T}^d \to \mathcal{H}_s$ such that $z(t) = U(\omega t)$ is a solution of (17).

(G) A partial differential equation, supplemented by some boundary conditions, is called a Hamiltonian partial differential equation, or an HPDE, if under a suitable choice of a symplectic Hilbert scale, domain and Hamiltonian, it can be written in the form (17).

(H) A non-linear PDE is called fully non-linear if the highest order derivatives appear with degree higher than one, it is called quasi-linear if the highest order derivatives appear with degree one both in the linear and in the nonlinear terms of the equation. It is called semi–linear if the linear term contains derivatives of higher order with respect to the nonlinear terms.

(I) Integrable PDE's is a fascinating, deep and interesting field by its own, and it has been widely studied starting from the 1960's with a variety of methods (formal algebraic methods, algebraic geometry, inverse spectral methods,...). For the connection of infinite integrable systems and KAM methods; see, e.g., [78] and [69].

3 Finite dimensional KAM Theory

3.1 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 [72]) *Consider a one–parameter family of real–analytic Hamiltonian functions on $\mathcal{M} := B(0,r) \times \mathbb{T}^d$ given by*

$$H := K + \varepsilon P \ , \qquad (\varepsilon \in \mathbb{R}) \ , \tag{18}$$

where: (i) K is a non–degenerate Kolmogorov normal form; (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 $\phi_* : \mathcal{M}_* := B(0, r_*) \times \mathbb{T}^d \to \mathcal{M}$, 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 fixed the frequency 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.

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 it is generated by

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

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

$$\beta_0 = \beta_0(x) := b + s_x \ , \qquad A = A(x) := a_x \quad \text{and} \quad \beta = \beta(y', x) := \beta_0 + A\, y' \ , \tag{19}$$

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

$$\begin{cases} y = y' + \varepsilon\beta(y', x) := y' + \varepsilon(\beta_0(x) + A(x)y') \\[2ex] x' = x + \varepsilon a(x) \ . \end{cases} \tag{20}$$

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

$$x = \varphi(x') := x' + \varepsilon\alpha(x';\varepsilon) \ , \tag{21}$$

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

$$\phi_1 : (y', x') \to \begin{cases} y = y' + \varepsilon\beta(y', \varphi(x')) \\ \\ x = \varphi(x') \ . \end{cases} \tag{22}$$

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 $\phi_1^{(2)} : (y', x') \to (\eta, \xi)$ 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 $\phi_1^{(1)} : (\eta, \xi) \to (y, x)$ 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

$$(y, x) \to \begin{cases} y' = M(x)y + c(x) \\ \\ x' = \varphi(x) \end{cases} \tag{23}$$

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 $\varphi = \text{ id} + \varepsilon a$) and it is easy to see that the symplectic diffeomorphisms of the form (23) form a subgroup of the symplectic diffeomorphisms, which we shall call *the group of Kolmogorov transformation*.

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

$$H_1 := H \circ \phi_1 = K_1 + \varepsilon^2 P_1 \ , \tag{24}$$

with K_1 in the Kolmogorov normal form

$$K_1 = E_1 + \omega \cdot y' + Q_1(y', x') \ , \qquad (Q_1 = O(|y'|^2) \ . \tag{25}$$

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

$$H(y' + \varepsilon\beta, x) = E + \omega \cdot y' + Q'(y', x) + \varepsilon F(y', x) + \varepsilon^2 P'(y', x) \tag{26}$$

where, letting

$$
\begin{aligned}
Q^{(1)} &:= [Q_y(y',x) - Q_{yy}(0,x)\, y'] \cdot \beta_0 = \frac{1}{2}\int_0^1 Q_{yyy}(ty',x) y' \cdot y' \cdot \beta_0 \, dt \\
Q^{(2)} &:= P(y',x) - P(0,x) - P_y(0,x)\, y' = \frac{1}{2}\int_0^1 P_{yy}(ty',x)\, y' \cdot y' \, dt \\
P^{(1)} &:= \frac{1}{\varepsilon^2}[Q(y'+\varepsilon\beta,x) - Q(y',x) - \varepsilon Q_y(y',x)\cdot\beta] = \frac{1}{2}\int_0^1 Q_{yy}(y'+t\varepsilon\beta,x)\,\beta\cdot\beta\, dt \\
P^{(2)} &:= \frac{1}{\varepsilon}[P(y'+\varepsilon\beta,x) - P(y',x)] = \int_0^1 P_y(y'+t\varepsilon\beta,x)\cdot\beta\, dt \ , \qquad (27)
\end{aligned}
$$

(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

$$
\begin{aligned}
&Q' := Q(y',x) + \varepsilon \widetilde{Q}(y',x) \ , \qquad \widetilde{Q}(y',x) := Q_y(y',x)\cdot(a_x\, y') + Q^{(1)} + Q^{(2)} \\
&F := b\cdot\omega + D_\omega s + D_\omega a \cdot y' + Q_{yy}(0,x) y' \cdot \beta_0 + P(0,x) + P_y(0,x)\cdot y' \\
&P' := P^{(1)} + P^{(2)} \ , \qquad (28)
\end{aligned}
$$

where $D_\omega a$ is the vector function with $k^{\rm th}$–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 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 (22), we see that in order to achieve relations (24)–(25), we have to determine b, s and a so that

$$
F(y',x) = \text{const} \ . \qquad (29)
$$

Remark 4 (i) F is a first degree polynomial in y' so that (29) is equivalent to

$$
\begin{cases}
b\cdot\omega + D_\omega s + P(0,x) = \text{const} \ , \\
D_\omega a + Q_{yy}(0,x)\beta_0 + P_y(0,x) = 0 \ .
\end{cases} \qquad (30)
$$

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 (29) or (30), we shall encounter differential equations of the form

$$D_\omega u = f \ , \tag{31}$$

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 Eq.n (31) 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 (31) is equivalent to

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

so that the solutions of (31) are given by

$$u = u_0 + \sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} \frac{f_n}{i\omega\cdot n} e^{in\cdot x} \ , \tag{33}$$

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

$$|f_n| \le M e^{-\sigma|n|} \ , \qquad \forall\, n \ . \tag{34}$$

Now, since $\omega \in \mathcal{D}_{\kappa,\tau}$ one has that

$$\frac{1}{|\omega\cdot n|} \le \frac{|n|^\tau}{\gamma} \tag{35}$$

and one sees that if f is analytic so is u in (33) (although the decay constants of u will be different to 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* (31) *with vanishing mean value, which is given by*

$$D_\omega^{-1} f := \sum_{\substack{n\in\mathbb{Z}^d \\ n\neq 0}} \frac{f_n}{i\omega\cdot n} e^{in\cdot x} \ ; \tag{36}$$

all other solutions of (31) are obtained by adding an arbitrary constant to $D_\omega^{-1} f$ as in (33) with u_0 arbitrary.

Taking the average of the first relation in (30), we may the determine the value of the constant denoted const , namely,

$$\text{const} = b\cdot\omega + P_0(0) := b\cdot\omega + \langle P(0,\cdot)\rangle \ . \tag{37}$$

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

$$s = D_\omega^{-1}\Big(P(0,x) - P_0(0)\Big) = \sum_{\substack{n\in\mathbb{Z}^d\\ n\neq 0}} \frac{P_n(0)}{i\omega\cdot n} e^{in\cdot x} \ , \tag{38}$$

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 (30) but *we select the natural zero–average solution*. Thus, s has been completely determined.

To solve the second (vector) equation in (30) we first have to require that the l.h.s. (left hand side) has vanishing mean value, i.e., recalling that $\beta_0 = b + s_x$ (see (19)), we must have

$$\langle Q_{yy}(0,\cdot)\rangle b + \langle Q_{yy}(0,\cdot)s_x\rangle + \langle P_y(0,\cdot)\rangle = 0 \ . \tag{39}$$

In view of (11) this relation is equivalent to

$$b = -\langle Q_{yy}(0,\cdot)\rangle^{-1}\Big(\langle Q_{yy}(0,\cdot)s_x\rangle + \langle P_y(0,\cdot)\rangle\Big) \ , \tag{40}$$

which determines uniquely b. Thus β_0 is completely determined, the l.h.s. of the second equation in (30) has zero average and the unique zero–average solution (again zero–average of a is required as a normalization condition) is given by

$$a = -D_\omega^{-1}\Big(Q_{yy}(0,x)\beta_0 + P_y(0,x)\Big) \ . \tag{41}$$

Finally, if $\varphi(x') = x' + \varepsilon\alpha(x';\varepsilon)$ is the inverse diffeomorphism of $x \to x + \varepsilon a(x)$ (compare (21)), 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 proved

Proposition 1 *If ϕ_1 is defined in* (20)–(19) *with s, b and a given in* (38), (40) *and* (41) *respectively, then* (24) *holds with*

$$\begin{cases} E_1 := E + \varepsilon\widetilde{E} \ , \\ \qquad \widetilde{E} := b \cdot \omega + P_0(0) \\ Q := \int_0^1 Q_x(y', x' + t\varepsilon\alpha) \cdot \alpha dt + Q'(y', \varphi(x')) \ , \\ P_1(y', x') := P'(y', \varphi(x')) \end{cases} \tag{42}$$

with Q' and P' defined in (27), (28) *and φ in* (21).

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., [91]) – 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 fundamental advantages: they are "quadratic" and, furthermore, after one step one has reproduced the initial situation (i.e., the form of H_1 in (24) 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, at the j^{th} step, would find

$$H_j = H_{j-1} \circ \phi_j = K_j + \varepsilon^{2^j} P_j \ . \tag{43}$$

The appearance of the exponential of the exponential of ε justifies the term "super–converge" 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 *fixed* the frequency in the normal form so that one can systematically use the Diophantine estimate (9).

Kolmogorov in his paper explained very neatly steps (i) and (iii) but did not provide the details for step (ii); at 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 $\xi \le 1$, we denote by $\mathcal{B}_\xi$ the space of function $f : B(0,\xi) \times \mathbb{T}^d \to \mathbb{R}$ analytic on

$$W_\xi := D(0,\xi) \times \mathbb{T}^d_\xi \ , \tag{44}$$

where

$$D(0,\xi) := \{y \in \mathbb{C}^d : |y| < \xi\} \quad \text{and} \quad \mathbb{T}^d_\xi := \{x \in \mathbb{C}^d : \ |\operatorname{Im} x_j| < \xi\}/(2\pi\mathbb{Z}^d) \tag{45}$$

with finite sup–norm

$$\|f\|_\xi := \sup_{D(0,\xi)\times\mathbb{T}^d_\xi} |f| \ , \tag{46}$$

(in other words, $\mathbb{T}^d_\xi$ denotes the complex points x with real parts $\operatorname{Re} x_j$ defined modulus 2π and imaginary part $\operatorname{Im} x_j$ 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 $\|f\|_\xi \le \|f\|_{\xi'}$ 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

$$|f_n(y)| \le \|f\|_\xi e^{-|n|\xi} \ , \quad \forall \ n \in \mathbb{Z}^d \ , \ \forall \ y \in D(0,\xi) \ . \tag{47}$$

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

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

$$\|\partial_y^\alpha \partial_x^\beta f\|_{\xi'} \le B_p \|f\|_\xi \ (\xi - \xi')^{-(|\alpha|+|\beta|)} \ . \tag{48}$$

Finally, we shall need estimate on $D_\omega^{-1} f$, i.e., on solutions of (31):

(P5) Assume that $x \to f(x) \in \mathcal{B}_\xi$ has zero average; assume that $\omega \in \mathcal{D}_{\kappa,\tau}$ (recall point (f) of Sect. "Introduction"), 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

$$\|\partial_x^\beta D_\omega^{-1} f\|_{\xi'} \leq \bar{B}_p \, \frac{\|f\|_\xi}{\kappa} \, (\xi - \xi')^{-k_p} \ . \tag{49}$$

Remark 6 (i) A proof of (49) is easily obtained observing that by (36) and (47), calling $\delta := \xi - \xi'$, one has

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

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

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

More surprising (and much more subtle) is that (49) holds with $k_p = |\beta| + \tau$; such estimate has been obtained by Rüssmann [102], [103]. For other explicit estimates see, e.g., [31] or [32].

(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

$$\|f\|'_\xi := \sum_{n\in\mathbb{Z}^d} |f_n| e^{\xi|n|} \ ; \tag{50}$$

see, e.g., [33] 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 $\omega \in \mathcal{D}_{\kappa,\tau}$. Denote

$$T := \langle Q_{yy}(0,\cdot)\rangle^{-1} \ , \qquad\qquad M := \|P\|_\xi \ . \tag{51}$$

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

$$|E| \ , \ |\omega| \ , \ \|Q\|_\xi, \|T\| < C \tag{52}$$

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

$$0 < \delta < \xi \quad \text{and define} \quad \bar\xi := \xi - \frac{2}{3}\delta \ , \quad \xi' := \xi - \delta \ . \tag{53}$$

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]

$$\begin{cases} \|s_x\|_{\bar\xi}, |b|, |\widetilde E|, \|a\|_{\bar\xi}, \|a_x\|_{\bar\xi}, \|\beta_0\|_{\bar\xi}, \|\beta\|_{\bar\xi}, \\ \qquad \|Q'\|_{\bar\xi}, \|\partial^2_{y'} Q'(0,\cdot)\|_0 \le \bar c C^{\bar\mu}\delta^{-\bar\nu} \ M =: \bar L \ , \\ \|P'\|_{\bar\xi} \le \bar c C^{\bar\mu}\delta^{-\bar\nu} \ M^2 = \bar L M \ . \end{cases} \tag{54}$$

The estimate in the first line of (54) allow 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*

$$\varepsilon_0 \, \bar L \le \frac{\delta}{3} \ , \tag{55}$$

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

$$\|\alpha\|_{\xi'} \le \bar L \qquad \text{and} \qquad \varphi = \text{ id } + \varepsilon\alpha : \mathbb{T}^d_{\xi'} \to \mathbb{T}^d_{\bar\xi} \ . \tag{56}$$

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

$$\phi_1 = \Big(y' + \varepsilon\beta(y', \varphi(x')), \varphi(x')\Big) : W_{\xi'} \to W_\xi \ , \quad \text{and} \quad \|\phi_1 - \text{ id }\|_{\xi'} \le |\varepsilon|\bar L \ ; \tag{57}$$

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*

$$\|S\|_{\bar\xi} \le \frac{\|a_x\|_{\bar\xi}}{1-|\varepsilon|\|a_x\|_{\bar\xi}} < \frac{3}{2}\bar L \ , \tag{58}$$

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., [73]).

From the Lemma and the definition of P_1 in (42), it follows immediately that

$$\|P_1\|_{\xi'} \le \bar L M \ . \tag{59}$$

Next, by the same technique used to derive (54), one can easily check that

$$\|\widetilde Q\|_{\xi'} \ , \ \ 2C^2\|\partial^2_{y'}\widetilde Q(0,\cdot)\|_0 \le cC^\mu\delta^{-\nu}M = L \ , \tag{60}$$

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

$$\varepsilon_0 L := \varepsilon_0 cC^\mu\delta^{-\nu}M \le \frac{\delta}{3} \ , \tag{61}$$

there follows that[19] $\|\widetilde T\| \le L$; this bound, together with (54), (60), (57), and (59), shows that

$$\begin{cases} |\tilde E|, \|\widetilde Q\|_{\xi'}, \|\widetilde T\|, \|\phi_1 - \mathrm{id}\|_{\xi'} \le L \\ \\ \|P_1\|_{\xi'} \le LM \ ; \end{cases} \tag{62}$$

provided (61) holds (notice that (61) implies (55)).

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

$$\phi_j : W_{\xi_{j+1}} \to W_{\xi_j} \ , \tag{63}$$

closer and closer to the identity, and such that (43) hold, the first thing to do is to choose the sequence ξ_j: such sequence has to be convergent, so that $\delta_j = \xi_j - \xi_{j+1}$ has to go to zero rather fast. Inverse power of δ_j (which, at the $j^{\rm th}$ step will play the rôle of δ in the previous paragraph) appear in the smallness conditions (see, e.g., (55)): 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

$$\begin{cases} \xi_0 := \xi \\ \\ \delta_0 := \dfrac{\xi-\xi_*}{2} \end{cases} \qquad \begin{cases} \delta_j := \dfrac{\delta_0}{2^j} \\ \\ \xi_{j+1} := \xi_0 - \delta_j = \xi_* + \dfrac{\delta_0}{2^j} \end{cases} \tag{64}$$

and observe that $\xi_j \downarrow \xi_*$. With this choice[20], Kolmogorov 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 (54), and define

$$C := 2\max\Big\{|E|, |\omega|, \|Q\|_\xi, \|T\|, 1\Big\}\ . \tag{65}$$

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

$$\varepsilon_0\, D\, B\, \|P\|_\xi \leq 1 \quad \text{where} \quad D := 3c\,\delta_0^{-(\nu+1)}\, C^\mu\ ,\ B := 2^{\nu+1}\ ; \tag{66}$$

notice that the constant C in (65) satisfies (52) and that (66) implies (55). Then the following claim holds.

Claim C: *Under condition* (66) *one can iteratively construct a sequence of Kolmogorov symplectic maps ϕ_j as in* (63) *so that* (43) *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 $i\geq 0$:*

$$|\varepsilon|^{2^i} M_i := |\varepsilon|^{2^i}\|P_i\|_{\xi_i} \leq \frac{(\,|\varepsilon| DBM)^{2^i}}{DB^{i+1}}\ , \tag{67}$$

$$\|\phi_* - \mathrm{id}\,\|_{\xi_*}, |E-E_*|, \|Q-Q_*\|_{\xi_*}, \|T-T_*\| \leq |\varepsilon|\, DBM\ , \tag{68}$$

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,* (51) *and* (65)*. There exist $c_* = c_*(d,\tau,\kappa,\theta)>1$ and positive integers $\sigma=\sigma(d,\tau)$, b such that if*

$$|\varepsilon| \leq \varepsilon_* := \frac{\xi^\sigma}{c_*\|P\|_\xi\ C^b} \tag{69}$$

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

$$\|\phi_* - \mathrm{id}\,\|_{\theta\xi}, |E - E_*|, \|Q - Q_*\|_{\theta\xi}, \|T - T_*\| \le \frac{|\varepsilon|}{\varepsilon_*} = |\varepsilon|\ c_*\|P\|_\xi C^b \xi^{-\sigma}\ . \tag{70}$$

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

(ii) From Cauchy estimates and (68), 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. Analyticity in ε and ε–power series expansions were very popular in the XIX and XX century[22], however was only J. Moser, within the framework of KAM theory, who proved rigorously (but "indirectly") for the first time, the convergence of such expansions in 1967: see [87]. Some of this matter is briefly discussed in Sect. 3.5 below.

(iv) **The nearly–integrable case.** In [72] it is pointed out that Kolmogorov Theorem yields easily the existence of many KAM tori for nearly–integrable systems (16) for $|\varepsilon|$ small enough, provided K is non–degenerate in the sense that

$$\det K_{yy}(y_0) \neq 0\ . \tag{71}$$

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

$$\ell_d(\omega(B)\backslash\mathcal{D}_{\kappa,\tau}), \ell_d(\{y \in B : \omega(y) \notin \mathcal{D}_{\kappa,\tau}\}) < c_\#\, \kappa\ . \tag{72}$$

Now, let $B_{\kappa,\tau} := \{y \in B : \omega(y) \in \mathcal{D}_{\kappa,\tau}\}$ (which, by (72) has Lebesgue measure $\ell_d(B_{\kappa,\tau}) \ge \ell_d(B) - c_\#\,\kappa$), 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

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

(where, for ease of notation, we did not change name 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 our 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 (72)).

(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 Kolmogorov set can be made more quantitative (i.e., one can see how such 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 (54) has the form[24]

$$c = \hat{c}\kappa^{-4} \ . \tag{73}$$

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 (66) reads $\varepsilon_0\kappa^{-4}\bar{D} \le 1$ 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].

3.2 Arnol'd'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. Arno'ld in 1963.

Theorem 2 (Arnol'd [2]) *Consider a one–parameter family of nearly–integrable Hamiltonians*

$$H(y,x;\varepsilon) := K(y) + \varepsilon P(y,x) \ , \qquad (\varepsilon \in \mathbb{R}) \tag{74}$$

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

$$K_y(y_0) = \omega \in \mathcal{D}_{\kappa,\tau} \ , \qquad \det K_{yy}(y_0) \neq 0 \ . \tag{75}$$

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

$$\phi : \theta \in \mathbb{T}^d \to \mathcal{M} \tag{76}$$

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

$$\mathcal{T}_{\omega,\varepsilon} := \phi(\mathbb{T}^d) \tag{77}$$

is invariant for H *and*

$$\phi_H^t \circ \phi(\theta) = \phi(\theta + \omega t) \ , \tag{78}$$

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) Arnol'd'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 Arnol'd's is more classical and, from the algebraic point of view, easier to construct than Kolmogorov's one but the estimates involved are somewhat more delicate and introduce a logarithmic correction, so that, in fact, the smallness parameter will be

$$\epsilon := |\varepsilon|(\log |\varepsilon|^{-1})^\rho \tag{79}$$

(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.

Arnol'd'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 rôle

The Hamiltonian H in (74) 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 $x \to x + \omega t$. Arno'ld's basic idea is to find a symplectic transformation

$$\phi_1 : W_1 := D(y_1, r_1) \times \mathbb{T}^d_{\xi_1} \to W_0 := D(y_0, r) \times \mathbb{T}^d_\xi \ , \tag{80}$$

so that $W_1 \subseteq W_0$ and

$$\begin{cases} H_1 := H \circ \phi_1 = K_1 + \varepsilon^2 P_1 \ , \qquad K_1 = K_1(y) \ , \\ \partial_y K_1(y_1) = \omega \ , \quad \det \partial_y^2 K_1(y_1) \neq 0 \end{cases} \tag{81}$$

(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

$$\phi_j : W_j := D(y_j, r_j) \times \mathbb{T}^d_{\xi_j} \to W_{j-1} \tag{82}$$

so that

$$\begin{cases} H_j := H_{j-1} \circ \phi_j = K_j + \varepsilon^{2^j} P_j \ , \qquad K_j = K_j(y) \ , \\ \\ \partial_y K_j(y_j) = \omega \ , \quad \det \partial_y^2 K_j(y_j) \neq 0 \ . \end{cases} \tag{83}$$

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

$$\phi(\theta) := \lim_{j\to\infty} \Phi_j(y_i, \theta) \ , \qquad \Phi_j := \phi_1 \circ \cdots \circ \phi_j : W_j \to W_0 \ \ , \tag{84}$$

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 $(y_0, \mathrm{id}\,)$; furthermore, the torus

$$\mathcal{T}_{\omega,\varepsilon} := \phi(\mathbb{T}^d) = \lim_{j\to\infty} \Phi_j(y_j, \mathbb{T}^d) \tag{85}$$

is invariant for H and (78) holds as announced in Theorem 2. Relation (78) 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 (84) implies

$$\begin{aligned} \phi_H^t \circ \phi(\theta) &= \lim_{j\to\infty} \phi_H^t \Big(\Phi_j(y_j, \theta) \Big) \\ &= \lim_{j\to\infty} \Phi_j \circ \phi_{H_j}^t (y_j, \theta) \\ &= \lim_{j\to\infty} \Phi_j(y_j, \theta + \omega t) \\ &= \phi(\theta + \omega t) \end{aligned} \tag{86}$$

(the first equality is just smooth dependence upon initial data of the flow ϕ_H^t together with (84); the second equality is (3); the third equality is due to the fact that $\phi_{H_j}^t(y_j, \theta) = (y_j, \theta + \omega t) + \epsilon_n$ where ϵ_n goes very rapidly to zero and the fourth equality is again (84)).

Arnol'd's transformation. Let us look for a near–to–the–identity transformation ϕ_1 so that the first line of (81) holds; such transformation will be determined by a generating function of the form

$$y' \cdot x + \varepsilon g(y', x) \ , \qquad \begin{cases} y = y' + \varepsilon g_x(y', x) \\ x' = x + \varepsilon g_{y'}(y', x) \ . \end{cases} \tag{87}$$

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

$$H(y' + \varepsilon g_x, x) = K(y') + \varepsilon\Big[K_y(y')\cdot g_x + P(y', x)\Big] + \varepsilon^2\Big(P^{(1)} + P^{(2)}\Big) \tag{88}$$

with (compare (27))

$$\begin{aligned} P^{(1)} &:= \frac{1}{\varepsilon^2}[K(y' + \varepsilon g_x) - K(y') - \varepsilon K_y(y')\cdot g_x] = \frac{1}{2}\int_0^1 K_{yy}(y' + t\varepsilon g_x)\, g_x\cdot g_x\, dt \\ P^{(2)} &:= \frac{1}{\varepsilon}[P(y' + \varepsilon g_x, x) - P(y', x)] = \int_0^1 P_y(y' + t\varepsilon g_x, x)\cdot g_x\, dt\ . \end{aligned} \tag{89}$$

Remark 9 (i) The (naive) idea is to try determine g so that

$$K_y(y')\cdot g_x + P(y', x) = \text{ function of } y' \text{ only }, \tag{90}$$

however, such relation is impossible to achieve. First of all, by taking the x–average of both sides of (90) 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 (90), is (by Fourier expansion)

$$\begin{cases} g = \displaystyle\sum_{n\neq 0} \frac{-P_n(y')}{iK_y(y')\cdot n}\ e^{in\cdot x}\ , \\ \\ K_y(y')\cdot g_x + P(y', x) = P_0(y')\ . \end{cases} \tag{91}$$

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), $y \to K_y(y)$ 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 (91) 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

$$K_y(y)\cdot n \neq 0\ , \qquad \forall\ y \in D(y_0, \bar{r})\ ,\ \forall\ 0 < |n| \leq N\ ; \tag{92}$$

the important quantitative aspects will be shortly discussed below.

(iii) Relation (90) is also one of the main "identity" in *Averaging Theory* and is related to the so–called *Hamilton–Jacobi equation*. Arnol'd's proof makes rigorous such theory and shows how a Newton method can be built upon it in order to establish the existence of invariant tori. In a sense, Arno'ld's approach is much more classical than Kolmogorov's one.

(iv) When (for a given y and n) it occurs that $K_y(y)\cdot n = K_y(y)\cdot n = 0$ 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 Sect. 6.1 and Sect. 6.2 of [6].

The key (simple!) idea of Arno'ld is to split the perturbation in two terms

$$P = \hat{P} + \check{P} \qquad \text{where} \qquad \begin{cases} \hat{P} := \sum_{|n|\le N} P_n(y)\, e^{in\cdot x} \\ \\ \check{P} := \sum_{|n|> N} P_n(y)\, e^{in\cdot x} \end{cases} \tag{93}$$

choosing N so that

$$\check{P} = O(\varepsilon) \tag{94}$$

(this is possible because of the fast decay of the Fourier coefficients of P; compare (34)). Then, for $\varepsilon \neq 0$, (88) can be rewritten as follows

$$H(y' + \varepsilon g_x, x) = K(y') + \varepsilon\Big[K_y(y')\cdot g_x + \hat{P}(y', x)\Big] + \varepsilon^2\Big(P^{(1)} + P^{(2)} + P^{(3)}\Big) \tag{95}$$

with $P^{(1)}$ and $P^{(2)}$ as in (89) and

$$P^{(3)}(y', x) := \frac{1}{\varepsilon}\check{P}(y', x)\ . \tag{96}$$

Thus, letting[28]

$$g = \sum_{0<|n|\le N} \frac{-P_n(y')}{iK_y(y')\cdot n}\ e^{in\cdot x}\ , \tag{97}$$

one gets

$$H(y' + \varepsilon g_x, x) = K_1(y') + \varepsilon^2 P'(y', x) \tag{98}$$

where

$$K_1(y') := K(y') + \varepsilon P_0(y')\ ,\quad P'(y', x) := P^{(1)} + P^{(2)} + P^{(3)}\ . \tag{99}$$

Now, by the IFT (Appendix A), for ε small enough, the map $x \to x + \varepsilon g_{y'}(y', x)$ can be inverted with a real–analytic map of the form

$$\varphi(y', x'; \varepsilon) := x' + \varepsilon\alpha(y', x'; \varepsilon) \tag{100}$$

so that Arnol'd's symplectic transformation is given by

$$\phi_1 : (y', x') \to \begin{cases} y = y' + \varepsilon g_x(y', \varphi(y', x'; \varepsilon)) \\ \\ x = \varphi(y', x'; \varepsilon) = x' + \varepsilon\alpha(y', x, ; \varepsilon) \end{cases} \tag{101}$$

(compare (22)). To finish the construction, observe that, from the IFT (see Appendix A 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 (81) holds, provided ε is small enough.

In conclusion, the analogous of Proposition 1 holds, describing Arnol'd's scheme:

Proposition 2 *If* ϕ_1 *is defined in* (101) *with* g *given in* (97) *(with* N *so that* (94) *holds) and* φ *given in* (100)*, then* (81) *holds with* K_1 *as in* (99) *and* $P_1(y', x') := P'(y', \varphi(y', x'))$ *with* P' *defined in* (99)*,* (96) *and* (89)*.*

Estimates and convergence. If f is a real–analytic function with analytic extension to $W_{r,\xi}$, we denote, for any $r' \leq r$ and $\xi' \leq \xi$,

$$\|f\|_{r',\xi'} := \sup_{W_{r',\xi'}(y_0)} |f(y, x)| \ ; \tag{102}$$

furthermore, we define

$$T := K_{yy}(y_0)^{-1} \ , \qquad M := \|P\|_{r,\xi} \ , \tag{103}$$

and assume (without loss of generality)

$$\kappa < 1 \ , \quad r < 1 \ , \quad \xi \leq 1 \ , \quad \max\{1, \|K_y\|_r, \|K_{yy}\|_r, \|T\|\} < C \ , \tag{104}$$

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 (34), it follows that, choosing

$$N := 5\delta^{-1}\lambda \ , \quad \text{where} \quad \lambda := \log|\varepsilon|^{-1} \ , \tag{105}$$

then

$$\|\check{P}\|_{r,\xi-\frac{\delta}{2}} \leq |\varepsilon| M \tag{106}$$

provided

$$|\varepsilon| \leq \ \text{const}\, \delta^{4d} \tag{107}$$

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 (97)), in a neighborhood $D(y_0, \bar{r})$ of y_0: this will determine the size of $\bar{r}$.

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

$$
\begin{aligned}
|K_y(y') \cdot n| &= \Big|\omega \cdot n + \Big(K_y(y') - K_y(y_0)\Big) \cdot n\Big| \\
&\ge |\omega \cdot n| \left(1 - \frac{\|K_{yy}\|_r}{|\omega \cdot n|} |n| \bar{r}\right) \\
&\ge \frac{\kappa}{|n|^\tau} \left(1 - \frac{C}{\kappa} |n|^{\tau+1} \bar{r}\right) \\
&\ge \frac{\kappa}{|n|^\tau} \left(1 - \frac{C}{\kappa} N^{\tau+1} \bar{r}\right) \\
&\ge \frac{1}{2} \frac{\kappa}{|n|^\tau}, \qquad (108)
\end{aligned}
$$

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

$$\bar{r} \le \frac{\kappa}{2C\ N^{\tau+1}} \overset{(105)}{=} \frac{\kappa}{2 \cdot 5^{\tau+1}\ C\ \ (\delta^{-1}\lambda)^{\tau+1}}\ . \qquad (109)$$

Eq. (108) allows easily to control Arnol'd's generating function g. For example:

$$
\begin{aligned}
\|g_x\|_{\bar{r},\xi-\frac{\delta}{2}} &= \sup_{D(y_0,\bar{r}) \times \mathbb{T}^d_{\xi-\frac{\delta}{2}}} \Big| \sum_{0<|n|\le N} \frac{n P_n(y')}{K_y(y') \cdot n}\ e^{in\cdot x} \Big| \\
&\le \sum_{0<|n|\le N} \frac{\sup_{D(y_0,r)} |P_n(y')|}{|K_y(y') \cdot n|} |n|\ e^{(\xi-\frac{\delta}{2})\ |n|} \\
&\le \sum_{n\in\mathbb{Z}^d} M\ \frac{2\,|n|^\tau}{\kappa}\ e^{-\frac{\delta}{2}|n|} \\
&\le \text{const}\ \frac{M}{\kappa}\ \delta^{-(\tau+1+d)}\ , \qquad (110)
\end{aligned}
$$

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 (81). Assuming that

$$\bar{r} \le \frac{r}{2} \qquad (111)$$

(allowing to use 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 implies that there exists a unique $y_1\in D(y_0,\bar r)$ such that (81) holds. In fact, assuming

$$\begin{cases} 8C^2\ \dfrac{\bar r}{r}\le 1\ , \\ \\ \dfrac{8CM}{r\bar r}\,|\varepsilon|\le 1 \end{cases} \tag{112}$$

one can show that[30]

$$|y_1-y_0|\le 4CMr^{-1}|\varepsilon|\le\frac{\bar r}{2}\ , \tag{113}$$

and

$$\partial_y^2K_1(y_1):=K_{yy}(y_1)+\varepsilon\partial_y^2P_0(y_1)=:T^{-1}(\mathbb{1}_d+A) \tag{114}$$

with a matrix $A:=T(K_{yy}(y_1)-K_{yy}(y_0)+\varepsilon\partial_y^2P_0(y_1))$ satisfying

$$\|A\|\le\frac{12\ C^3M}{r^2}|\varepsilon|\overset{(112)}{\le}\frac{12\ C^3M}{r}\frac{1}{8C^2\bar r}|\varepsilon|\overset{(112)}{\le}\frac{3}{16}\ . \tag{115}$$

Eq.'s (114) and (115) imply that $\partial_y^2K(y_1)$ is invertible (Neumann series) and that[31]

$$\partial_y^2K_1(y_1)^{-1}=T+\varepsilon\widetilde T\ ,\qquad \|\widetilde T\|\le 15\frac{C^4M}{r^2}\ . \tag{116}$$

Finally, notice that by (113),

$$D(y_1,\bar r/2)\subseteq D(y_0,\bar r)\ . \tag{117}$$

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

$$K_1:=K+\varepsilon\widetilde K=K+\varepsilon P_0(y')\ ,\quad y_1:=y_0+\varepsilon\tilde y\ , \tag{118}$$

one can easily prove (along the lines that led to (54)) the following estimates, where as in Sect. "Kolmogorov Theorem", $\bar\xi:=\xi-\frac{2}{3}\delta$ and $\bar r$ is as above:

$$\begin{cases} \dfrac{\|g_x\|_{\bar r,\bar\xi}}{r},\|g_{y'}\|_{\bar r,\bar\xi},\dfrac{|\tilde y|}{r},\|\tilde K_y\|_{\bar r},\|\tilde K_{yy}\|,\|\widetilde T\|\le c\kappa^{-2}\,C^\mu\delta^{-\nu}\lambda^\rho\ M=:L\ , \\ \\ \|P'\|_{\bar\xi}\le c\kappa^{-2}C^\mu\delta^{-\nu}\lambda^\rho M^2=LM\ , \end{cases} \tag{119}$$

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 (119), 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 (55) holds (with L as in (119)), in which case (56) holds (for any $|\varepsilon| \le \varepsilon_0$ and any $y' \in D_r(y_0)$). Furthermore, under the above hypothesis, it follows that[33]

$$\begin{cases} \phi_1 := \Big(y' + \varepsilon g_x(y', \varphi(y', x')), \varphi(y', x')\Big) : W_{\bar r/2, \xi-\delta}(y_1) \to W_{r,\xi}(y_0) \\ \\ \|\phi_1 - \mathrm{id}\,\|_{\bar r/2, \xi-\delta} \le |\varepsilon| L \ . \end{cases} \tag{120}$$

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 such domain by

$$\|P_1\|_{\bar r/2, \xi-\delta} \le LM \ . \tag{121}$$

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

$$C := 2\max\{1, \|K_y\|_r, \|K_{yy}\|_r, \|T\|\} \ , \quad \gamma := 3C \ , \quad \delta_0 := \frac{(\gamma-1)(\xi-\xi_*)}{\gamma} \ ; \tag{122}$$

ξ_j and δ_j as in (64) *but with* δ_0 *as in* (122); we also define, for any $j \ge 0$,

$$\lambda_j := 2^j\lambda = \log \varepsilon_0^{-2^j} \ , \qquad r_j := \frac{\kappa}{4\cdot 5^{\tau+1} C (\delta_j^{-1}\lambda_j)^{\tau+1}} \ ; \tag{123}$$

(this part is adapted from **Step 3** in Sect. "Kolmogorov Theorem"; see, in particular, (104)). 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]

$$\begin{cases} \varepsilon_0 \le e^{-\beta} & \text{with} \quad \beta := \dfrac{\delta_0}{5}\left(\dfrac{\kappa}{Cr}\right)^{\frac{1}{\tau+1}} \\ \\ \varepsilon_0\, D\, B\, \|P\|_{r,\xi} \le 1 & \text{with} \quad D := 3c\kappa^{-2}\,\delta_0^{-(\nu+1)}\, C^{\mu+1} \ , \ B := \gamma^{\nu+1}(\log \varepsilon_0^{-1})^{\rho} \ . \end{cases} \tag{124}$$

Remark 10 Notice that the power of κ^{-1} (the inverse of the Diophantine constant) in the second smallness condition in (124) is two, which implies (compare Remark 7–(v)) that the measure of the complementary of Kolmogorov set may be bounded by a constant times $\sqrt{\epsilon}$, where $\epsilon := \varepsilon(\log\varepsilon)^{-\rho}$. This bound is almost optimal (i.e., optimal, up to logarithmic corrections) as the trivial example $(y_1^2 + y_2^2)/2 + \varepsilon\cos(x_1)$ shows: such 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).
Taking out the logarithm is not a completely trivial matter and even though in the literature is normally claimed that the sharp estimate holds, a complete proof of this fact is hard to find. For a recent detailed proof, see [20].

3.3 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, [84] and [85]. Moser's original set up corresponds 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, [101].

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 [90], 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 looses in the inversion of the small divisor operator.

The technique, which we shall describe here, is again due to Moser ([88]) but is rather different from the original one and it 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 [86] and, for an abstract functional approach, [113], [114]. 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, [92] and the beautiful exposition [107], which inspires the presentation reported here.

Let us consider a Hamiltonian $H = K + \varepsilon P$ (as in (18)) with K a real–analytic Kolmogorov normal form as in (10) with $\omega \in \mathcal{D}_{\kappa,\tau}$ and Q real–analytic; P is assumed to be a $C^\ell(\mathbb{R}^d \times \mathbb{T}^d)$ 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 in Remark 7–(i) above, with (70) strengthened by including in the left hand side of (70) also[37] $\|\partial(\phi_* - \mathrm{id}\,)\|_{\theta\xi}$ and $\|\partial(Q - Q_*)\|_{\theta\xi}$ (where "∂" denotes, here, "Jacobian" with respect 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 $0 < \xi \le 1$ there exists a real–analytic function $f_\xi : X_\xi^d := \{x \in \mathbb{C}^d : \ |\operatorname{Im} x_j| < \xi\} \to \mathbb{C}$ such that*

$$\begin{cases} \sup_{X_\xi^d} |f_\xi| \le c\|f\|_{C^0} \ , \quad \|f_\xi - f\|_{C^s} \le c\|f\|_{C^l}\xi^{l-s} \ , (s \in \mathbb{N}, s \le l) \\ \\ \sup_{X_{\xi'}^d} |f_\xi - f_{\xi'}| \le c\|f\|_{C^l}\xi^l \ , \quad \forall\ 0 < \xi' < \xi \ , \end{cases} \tag{125}$$

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

Lemma 3 (Bernstein, Moser) *Let $l \in \mathbb{R}_+ \backslash \mathbb{Z}$; let $f_0 = 0$ and let, for any $j \geq 1$, f_j be real analytic functions on $X_j^d := \{x \in \mathbb{C}^d : \ |\operatorname{Im} x_k| < 2^{-j}\}$ such that*

$$\sup_{X_j^d} |f_j - f_{j-1}| \leq A \ 2^{-jl} \tag{126}$$

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 $C = C(d, l) > 0$,

$$\|f\|_{C^l(\mathbb{R}^d)} \leq CA \ . \tag{127}$$

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

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

$$P^j := P_{\xi_j} \ , \qquad \xi_j := \frac{1}{2^j} \ . \tag{128}$$

Claim M: *If $|\varepsilon|$ is small enough and if $\ell > \sigma + 1$, then there exists a sequence of Kolmogorov symplectic transformations $\{\Phi_j\}_{j \geq 0}$, $|\varepsilon|$–close to the identity, and a sequence of Kolmogorov normal forms K_j such that*

$$H_j \circ \Phi_j = K_{j+1} \qquad \text{on} \quad W_{\xi_{j+1}} \tag{129}$$

where

$$\begin{aligned}
&H_j := K + \varepsilon P^j \\
&\Phi_0 = \phi_0 \quad \text{and} \quad \Phi_j := \Phi_{j-1} \circ \phi_j \ , (j \geq 1) \\
&\phi_j : W_{\xi_{j+1}} \to W_{\alpha \xi_j} \ , \qquad \Phi_{j-1} : W_{\alpha \xi_j} \to X_{\xi_j} \ \ , \ \ j \geq 1 \text{ and } \alpha := \frac{1}{\sqrt{2}}, \\
&\sup_{x \in \mathbb{T}^d_{\xi_{j+1}}} |\Phi_j(0, x) - \Phi_{j-1}(0, x)| \leq \ \text{const} \ |\varepsilon| 2^{-(\ell - \sigma) j} \ .
\end{aligned} \tag{130}$$

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

From Claim **M** and Lemma 3 (applied to $f_j(x) = \Phi_j(0, x) - \Phi_0(0, x)$ 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 $\phi : \mathbb{T}^d \to \mathbb{R}^d \times \mathbb{T}^d$, which is ε–close to the identity, and, because of (129),

$$\phi(x + \omega t) = \lim \Phi_j(0, x + \omega t) = \lim \phi^t_{H_j} \circ \Phi_j(0, x) = \phi^t_H \circ \phi(x) \tag{131}$$

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 $x \to (0, x)$).

3.4 Lower dimensional KAM tori

We consider 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 PDE's (Partial Differential Equations) with a Hamiltonian structure. For a review on lower dimensional tori (in finite dimensions), we refer the reader to [108].

In 1965 V.K. Melnikov [83] stated a precise result concerning the persistence of *stable* (or "elliptic") lower dimensional tori; the hypotheses of such result are, now, commonly referred to as "Melnikov conditions". However, a proof of Melnikov's statement was given only later by Moser [87] for the case $n = d - 1$ and, in the general case, by H. Eliasson in [52] and, independently, by S.B. Kuksin [75]. 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 [87], [63] and [114] (roughly speaking, the normal frequencies to the torus do not resonate with the inner (or "proper") frequencies associated to the quasi–periodic motion). Since then, Melnikov conditions have been significantly weakened and a lot of technical progress has been done; see [108], Sect's 5, 6 and 7, and references therein.

As an example we consider a system with $n + m$ degrees of freedom with Hamiltonian

$$H = K(x, y, z; \xi) + \varepsilon P(x, y, z; \xi) \tag{132}$$

where $(x, y) \in \mathbb{T}^n \times \mathbb{R}^n$, $z = \frac{1}{\sqrt{2}}(p + \mathrm{i}q) \in \mathbb{C}^m$ are pairs of standard symplectic coordinates,while ξ is a real parameter running over a compact set $\Pi \subset \mathbb{R}^n$ of positive Lebesgue measure**50**. K, P are Lipschitz in ξ and analytic with respect to the dynamical variables $x, y, z, \bar{z}$ (note that when we complexify (p, q) the variables $z, \bar{z}$ become independent) in the complexified domain

$$(x, y, z, \bar{z}) \in D(s, r) := \mathbb{T}^n_s \times B_{r^2}(\mathbb{C}^n) \times B_r(\mathbb{C}^m) \times B_r(\mathbb{C}^m)$$

namely they can be written in totally convergent Taylor Fourier series as

$$K = \sum_{d \in \mathbb{N}} K^{(d)} = \sum_{d \in \mathbb{N}} \sum_{\substack{\ell \in \mathbb{Z}^n, l \in \mathbb{N}^n, \alpha, \beta \in \mathbb{N}^m \\ 2|l| + |\alpha| + |\beta| = d}} K_{\ell,\alpha,\beta} e^{\mathrm{i}\ell \cdot x} y^l z^\alpha \bar{z}^\beta \,,$$

where $y^l z^\alpha \bar{z}^\beta = \prod_{i=1}^n y_i^{l_i} \prod_{j=1}^m z_j^{\alpha_j} \bar{z}_j^{\beta_j}$.

Here the apex (d) denotes the homogeneous components of *degree* d, provided that we

assign degree two to the variables y, degree one to the variables z and degree zero to the variables x. Note that this choice of degrees is the one that makes the symplectic form homogeneous of degree two, since the variables x, which are not close to zero, must have degree zero.
We shall assume that, for all $\xi \in \Pi$, K admits the n-torus

$$\mathcal{T}_0(\xi) := \{y = 0\} \times \mathbb{T}^n \times \{z = 0\}$$

as a linearly stable invariant torus and is written in normal form

$$K = K^{(0)}(\xi) + \omega^{(0)}(\xi) \cdot y + \sum_{j=1}^{m} \Omega_j^{(0)}(\xi)|z_j|^2 + K^{(\geq 3)} \tag{133}$$

here $K^{(\geq 3)}$ is an analytic Hamiltonian with minimal degree at least three while $K^{(0)}(\xi)$ is a constant. The ϕ_K^t flow decouples in the linear flow $x \in \mathbb{T}^n \to x + \omega^{(0)}(\xi)t$ times the motion of m (decoupled) harmonic oscillators with characteristic frequencies $\Omega_j^{(0)}(\xi)$ (sometimes referred to as *normal frequencies*).
We have the following result:

Theorem 3 (Pöschel [93]) *Fix $\gamma > 0, \tau > n$ then for all $|\varepsilon|$ sufficiently small there exists Lipschitz functions $\omega(\xi), \Omega(\xi) : \Pi \to \mathbb{R}^{n+m}$ ε-close to $\omega^{(0)}(\xi), \Omega^{(0)}(\xi)$ such that setting*

$$\Pi^* := \left\{ \xi \in \Pi : |\omega \cdot \ell + k \cdot \Omega| \geq \frac{\gamma}{|\ell|^\tau}, \forall (\ell, k) \in \mathbb{Z}^{n+m} \setminus \{0\} : |k| \leq 2 \right\} \tag{134}$$

then for all $\xi \in \Pi^*$ there exists a change of variables Φ, ε-close to the identity, such that

$$H \circ \Phi = \omega(\xi) \cdot y + \sum_{j=1}^{m} \Omega_j(\xi)|z_j|^2 + H^{(\geq 3)} \tag{135}$$

namely it is in normal form with frequencies $\omega(\xi), \Omega(\xi)$.

Now in order to make this result interesting we have to give conditions which ensure that the set Π^* has positive Lebesgue measure. This follows for instance by requiring that $\xi \to \omega^{(0)}(\xi)$ is a Lipeomorphism and that the Melnikov conditions hold. Explicitly, for any $(\ell, k) \in \mathbb{Z}^{n+m} \setminus \{0\}$ with $|k| \leq 2$, we define:

$$\mathcal{R}_{\ell,k}^{(0)} := \{\xi \in \Pi : \omega^{(0)} \cdot \ell + k \cdot \Omega^{(0)} = 0\}, \quad \text{and assume that} \quad |\mathcal{R}_{\ell,k}^{(0)}| = 0, \mathcal{R}_{0,k}^{(0)} = \emptyset \tag{136}$$

This formulation has been borrowed from [93], to which we refer for a complete proof; the description of the set Π^* in terms of the *final frequencies* is the one given in [12]; for the differentiable analog, see [46].

In order to give a sketch of the proof let us introduce some notation: we define the degree projections $\Pi^{\leq j}$, Π^j, $\Pi^{>j}$ as

$$\Pi^j H = H^{(j)}\,, \quad \Pi^{\leq j} H = \sum_{0\leq d\leq j} H^{(d)}\,,$$

in the same way $H^{\geq 3}$ is a Hamiltonian with minimal degree at lest three, while $H^{\leq 2}$ is a polynomial Hamiltonian of maximal degree two etc...

We endow the space of Hamiltonians with a structure of scale of Banach spaces with respect to the norm defined as follows.

We represent a vector field on $\mathbb{R}^{2n}\times\mathbb{C}^m$ as

$$X = \sum_{i=1}^n X^{(x_i)}(x,y,z)\frac{\partial}{\partial x_i} + X^{(y_i)}(x,y,z)\frac{\partial}{\partial y_i} + \sum_{j=1}^m X^{(z_j)}(x,y,z)\frac{\partial}{\partial z_j}\,,$$

where each component is an analytic function,

$$X^{(\mathtt{v})}(x,y,z) = \sum_{\ell\in\mathbb{Z}^n, l\in\mathbb{N}^n, \alpha,\beta\in\mathbb{N}^m} X^{(\mathtt{v})}_{\ell,\alpha,\beta} e^{\mathrm{i}\ell\cdot x} y^l z^\alpha \bar{z}^\beta\,, \quad \mathtt{v} = x_1,\dots x_n, y_1,\dots,y_n, z_1,\dots,z_m\,.$$

Finally we define the majorant vector field $\underline{X}$ by setting $\underline{X}^{(\mathtt{v})}_{\ell,\alpha,\beta} = |X^{(\mathtt{v})}_{\ell,\alpha,\beta}|$.

$$|H|_{s,r} := \sup_{D(s,r)}\left(|H| + |\underline{X}^{(x)}_H|^{\mathrm{lip}} + \frac{1}{r^2}|\underline{X}^{(y)}_H|^{\mathrm{lip}} + \frac{1}{r}|\underline{X}^{(z)}_H|^{\mathrm{lip}}\right) \tag{137}$$

where given a Lipschitz map $f:\Pi\to E$ with E a Banach space we denote by $|f|^{\mathrm{lip}}_E$ the *inhomogeneous Lipschitz norm*

$$|f|^{\mathrm{lip}}_E := \sup_{\xi\in\Pi}|f(\xi)|_E + \sup_{\xi\neq\xi'\in\Pi}\frac{|f(\xi)-f(\xi')|_E}{|\xi-\xi'|}$$

This norm is less natural than the one defined in (46), in particular due to the presence of the majorant it is not coordinate independent. However it is closed with respect to Poisson brackets, projection onto the components of homogeneous degree and has exponentially small smoothing estimates for the ultraviolet terms, i.e. has properties similar to **(P1)-(P5)**, moreover it turns out that with this definition the smallness assumptions on ε in the KAM theorem 3 are independent of m. Now our goal is to find:

(A) a sequence $\omega^{(n)}, \Omega^{(n)}$ defined and Lipschitz for $\xi \in \Pi$ and tending to ω, Ω super–exponentially;

(B) a sequence ε_n rapidly converging to zero and a (rapidly converging) sequence of changes of variables Ψ_n, well defined and Lipschitz for ξ in a nested sequence of domains which contains Π^* defined by

$$\Pi_n^* := \left\{ \xi \in \Pi_{n-1}^* : |\omega^{(n)} \cdot \ell + k \cdot \Omega^{(n)}| \geq \frac{\gamma}{|\ell|^\tau}, \forall (\ell, k) \in \mathbb{Z}^{n+m} \setminus \{0\} : |k| \leq 2, \quad |\ell| \leq N_n \right\}, \tag{138}$$

with $N_n \sim \ln(\varepsilon_n^{-1})$, such that

$$H_0 = K_0 + \varepsilon_0 P_0 , \quad H_{n+1} = \Psi_n H_n = K_{n+1} + \varepsilon_{n+1} P_{n+1} ,$$

$K_n = K_n^{(0)} + K_n^{(2)} + K_n^{(\geq 3)}$ with $K_n^{(0)}$ depending only on ξ,

$$K_n^{(2)} = \omega^{(n)} \cdot y + \sum_j \Omega_j^{(n)} |z_j|^2 ,$$

Finally P_n is a polynomial of maximal degree two and $\varepsilon_n \sim \varepsilon_{n-1}^{3/2}$ tends to zero super–exponentially.
Let us show how to perform one step of this procedure. We claim that Ψ_n is the time-one flow of a generating function $S_n \sim O(N_n^{3\tau} \varepsilon_n)$, where the closeness in in the norm (137) with an appropriate choice of parameters s_n, r_n . Recalling the Lie exponentiation formula, we have

$$H_n \circ \Psi_n = H_n + \{S_n, H_n\} + O(\varepsilon_n^2) = K_n + \varepsilon_n P_n + \{S_n, K_n\} + O(N_n^{6\tau} \varepsilon_n^2) .$$

Our goal is achieved provided that we fix S_n so that

$$\Pi^{(\leq 2)}(K_n + \varepsilon_n P_n + \{S_n, K_n\}) = \Pi^{(\leq 2)} K_{n+1} + O(\varepsilon_n^{3/2}) ,$$

recall that we only want the terms of degree at most two to be in normal form, this is why we apply the projection $\Pi^{(\leq 2)}$. We assume that S_n is a polynomial of maximal degree two and solve the equations above in increasing homogeneous degrees, recalling that $\{F^{(d_1)}, G^{(d_2)}\}$ has degree $d_1 + d_2 - 2$. We get a triangular system:

$$\begin{aligned} \{S_n^{(0)}, K_n^{(2)}\} &= -\varepsilon_n P_n^{(0)} + K_{n+1}^{(0)} - K_n^{(0)} + O(\varepsilon_n^{3/2}) \\ \{S_n^{(1)}, K_n^{(2)}\} &= -\varepsilon_n P_n^{(1)} - \{S_n^{(0)}, K_n^{(3)}\} + O(\varepsilon_n^{3/2}) \\ \{S_n^{(2)}, K_n^{(2)}\} &= -\varepsilon_n P_n^{(2)} - \{S_n^{(0)}, K_n^{(4)}\} - \{S_n^{(1)}, K_n^{(3)}\} + K_{n+1}^{(2)} - K_n^{(2)} + O(\varepsilon_n^{3/2}) . \end{aligned} \tag{139}$$

which we solve for $\xi \in \Pi_n$, just like we did for equation (31), by noticing that

$$\{K_n^{(2)}, e^{\mathrm{i}\ell\cdot x} y^l z^\alpha \bar{z}^\beta\} = \mathrm{i}\big(\omega^{(n)} \cdot \ell + \Omega^{(n)} \cdot (\alpha - \beta)\big) e^{\mathrm{i}\ell\cdot x} y^l z^\alpha \bar{z}^\beta$$

and that all the ultraviolet terms with frequency $\geq N_n$ can be ignored since they are $O(\varepsilon_n^{3/2})$.

One can also consider more general cases, for instance where the conditions (136) hold only for (ℓ, k) with $|k| \leq 1$, namely the *second* Melnikov conditions do not hold. Then one can still prove the existence of a torus, for ξ in some positive measure Cantor-like set, this was done by J. Bourgain in [22]. We state his theorem (written with our notations):

Theorem 4 (Bourgain [22]) *Let $H(x, y, z)$ be of he form*

$$H = \omega^{(0)}(\xi) \cdot y + \frac{1}{2}|y|^2 + \sum_{j=1}^{m} \Omega_j^{(0)}(\xi)|z_j|^2 + \varepsilon P$$

where we assume that $\omega^{(0)}$ is diophantine and that condition (136) holds for (ℓ, k) with $|k| \leq 1$. Then, for any fixed small $\varepsilon > 0$ and for λ taken in a set of positive measure, there exists a perturbed torus with frequency vector $\omega = \lambda\omega^{(0)}$, parametrized as

$$x = \omega t + X(\omega t), \qquad y = Y(\omega t), \qquad z = Z(\omega t)$$

with (X, Y, Z) quasi-periodic and of size, say, $O(\varepsilon^{1/2})$ in a suitable real analytic function space norm.

We remark that in general in this case one does not have information on the stability in the z directions.

Bourgain's approach to this problem was to look directly for the quasi-periodic solution. This amounts to looking for a map

$$\mathbf{i} : \mathbb{T}^n \to \mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}^m, \quad \varphi \to \mathbf{i}(\varphi) = (\varphi + X(\varphi), Y(\varphi), Z(\varphi)),$$

and for a frequency $\omega \in \mathbb{R}^n$, which solve the functional equation

$$\mathcal{F}(\mathbf{i}) := \omega \cdot \partial_\varphi \mathbf{i}(\varphi) - X_H(\mathbf{i}(\varphi)) = 0.$$

Now in order to solve this functional problem we apply a Nash-Moser quadratic algorithm, starting from the approximate solution $\mathbf{i} = \mathbf{i}_0(\varphi) = (\varphi, 0, 0)$ and $\omega = \omega^{(0)}$ and constructing a super–exponentially convergent sequence of approximate solutions $\mathbf{i}_n(\varphi), \omega^{(n)}$. The key point is to invert (with some quantitative control on the bounds) the linearized operator at an approximate solution. This is in general a much more difficult task with respect to solving the homological equations (139), since it involves a linear operator which depends quasi-periodically on time. In the case of maximal tori this problem can be overcome, e.g., [31], by exploiting the symplectic structure. In the more difficult elliptic lower–dimensional case, Bourgain solves the problem by a "multiscale theorem", which he first developed in the context of KAM for PDE's, see [26] or [16]. Actually as shown in [14] this approach is completely parallel to a KAM scheme. Indeed the existence of a quasi periodic solution $\mathbf{i}(\varphi)$ implies the existence of a symplectic change of variables which puts the Hamiltonian in the normal form:

$$H \circ \Phi = \omega(\xi) \cdot y + Q(x, z; \xi) + H^{(\geq 3)} \tag{140}$$

where Q is a quadratic form in z which depends on the angles x.

3.5 Other Chapters in classical KAM Theory

In this section we review in a schematic and informal way some developments and applications of KAM theory; for other more exhaustive surveys we refer to [29], [6] (Sect. 6.3) or [108].

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 $\mathcal{C}_* \subset B$ such that, for each $y' \in \mathcal{C}_*$, 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 $\mathcal{C}_* \times \mathbb{T}^d$. Still another way of describing this result is that *there exists a smooth function $K_* : B \to \mathbb{R}$ such that $(K + \varepsilon P) \circ \phi_*$ and K_* agree, together with their derivatives, on $\mathcal{C}_* \times \mathbb{T}^d$* : 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 admits C^k extensions are called *Whitney smooth*; compare [112], 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 [92] and [39] in, respectively, the differentiable and the analytic case – follows easily from the following lemma[40]:

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

Actually, the dependence upon the angles x' of ϕ_* is analytic and it is only the dependence upon $y' \in \mathcal{C}_*$ which is Whitney smooth ("anisotropic differentiability", compare Sect. 2 in [92]).

For more information and a systematic use of Whitney differentiability, see [29].

2 *Power series expansions*

KAM tori $\mathcal{T}_{\omega,\varepsilon} = \phi_\varepsilon(\mathbb{T}^d)$ of nearly–integrable Hamiltonians correspond to quasi–periodic trajectories $z(t;\theta,\varepsilon) = \phi_\varepsilon(\theta + \omega t) = \phi_H^t(z(0;\theta,0))$; 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 XIX century to mathematicians and astronomers (such as Newcombe, Lindstedt and, especially, Poincaré; compare [91], Vol. II). Indeed, formal power solutions of nearly–integrable Hamiltonian equations are not difficult to construct (see, e.g., Sect. 7.1 of [32]) 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 k^{th} ε–expansion coefficient, are quite difficult and were carried out only in the late eighties by H. Eliasson [54]. 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 [54]), fully direct proofs were published in 1994 in [60] and [38].

3 *Non–degeneracy assumptions*

Kolmogorov's non–degeneracy assumption (71) can be generalized in various ways. First of all, Arnold pointed out in [2] that the condition

$$\det \begin{pmatrix} K_{yy} & K_y \\ K_y & 0 \end{pmatrix} \neq 0 \ , \tag{141}$$

(this is a $(d+1)\times(d+1)$ matrix where last column and last row are given by the $(d+1)$–vector $(K_y, 0)$) which is independent from condition (71), is also sufficient to construct KAM tori. Indeed, (141) may be used to construct *iso–energetic* KAM tori, i.e., tori on a *fixed energy level*[43] E.

More recently, Rüssmann [104] (see, also, [106]), using results of Diophantine approximations on manifolds due to Pyartly [98], formulated the following condition ("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 ahead the frequency

For a thorough discussion of this topic, see Sect. 2 of [108].

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 motivation 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 [89].

A first highly non–trivial step, in order to apply KAM theory to such classical systems, is to construct explicitly action–angle variables and to determine their analyticity properties, which is in itself a technical 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 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 [59] (see, also, [36]); a completion of Arnold's project [3] (where Arnold proved the first nontrivial case of the circular planar three body problem and gave a sketch of how to generalize to the general case) has been completed in [42]; see also, [43], [44] and [45].

4.2 *Topological trapping in low dimensions*

The general 2–degree–of–freedom nearly–integrable Hamiltonian exhibits a kind of stability particularly strong: 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 (141), by the iso–energetic KAM theorem) *separate* the energy levels and orbits lying between two KAM tori will remain forever trapped in such 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 applications to the stability of three–body problems in celestial mechanics see [33] and item **4.4** below.

In higher dimension this topological trapping is no more 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 $\mathbb{R}$

$$L := -\frac{d^2}{dt^2} + v(t) \ , \qquad t \in \mathbb{R} \ . \tag{142}$$

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 [48].

If v is periodic, then $\sigma(L)$ is a continuous band spectrum, as it follows immediately from Floquet theory [48]. Much more complicate is the situation for quasi–periodic potentials $v(t) := V(\omega t) = V(\omega_1 t, ..., \omega_n t)$, 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 [89], where, in particular, interesting connections with mechanics are discussed[46]; for deep developments of generalization of Floquet theory ("reducibility") to quasi–periodic Schrödinger operators, see [53] 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 [67] 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 as given Keplerian ellipse) yields existence of invariant tori if the mass ratio of the primaries is less than[47] 10^{-50}. Since then, a lot of progress has been done and, in [33], 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 such 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 – 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 ([81]). Therefore, there must exist a critical value $\varepsilon_c(\omega)$ ("breakdown threshold") such that, for $0 \leq \varepsilon < \varepsilon_c(\omega)$, 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 [33].

4 Infinite dimensional KAM theory

One of the most important 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 classes of partial differential equations carrying a Hamiltonian or a reversible structure. The concept of integrability for a Hamiltonian PDE has been studied widely since the 1960's. Most of the literature on KAM theory for PDEs however is on the existence of small quasi-periodic solutions for PDEs with an elliptic fixed point at zero and such that the equation linearized at zero has a numerable basis of eigenvectors, either on a *compact manifold* or on $\mathbb{R}^d$ with a confining potential. Regarding the construction of large quasi-periodic solutions for PDE's close to a nonlinear integrable model, the results are much fewer; see, however, [17] .

It must be remarked that quasi-periodic solutions are the infinite dimensional analogue of lower dimensional tori, hence they are expected to cover a set of measure zero in phase space. Results on *maximal tori* for PDEs are very few and mostly on ad-hoc models; see, e.g., [41], [96], [27].

The first results on quasi-periodic solutions were obtained by using an adaptation of Theorem 3 (see, for instance, [75], citeWa90, [94]), and were for semi–linear PDEs with Dirichlet boundary conditions in $[0, \pi]$. As an example we state the result for the NLS equation

$$\mathrm{i}u_t - u_{xx} + |u|^2 u + f(|u|^2)u \tag{143}$$

Theorem 5 (Kuksin-Pöschel [79]) *Suppose the nonlinearity $f(y)$ is analytic and has a zero of degree at least two in $y = 0$. Then for all $n \in \mathbb{N}$ and all $\mathcal{S} = \{j_1, \dots, j_n\} \subset \mathbb{N}$ there exists a Cantor manifold $\mathcal{E}_{\mathcal{S}}$ of real analytic, linearly stable, diophantine n-tori for equation (143). More precisely there exists a Cantor set $\mathcal{C}$, with asymptotically full density at zero, such that for all $\xi \in \mathcal{C}$ there exists a linearly stable solution of (143) of the form*

$$u(t, x; \xi) = \sum\nolimits_{j \in S} 2\sqrt{\xi_j}\, \sin(\omega_j t + jx) + o(\sqrt{|\xi|}), \quad \omega_j := j^2 + O(|\xi|)\,, \tag{144}$$

where $o(\sqrt{|\xi|})$ is small in some appropriate analytic norm and the map $\xi \to u(\xi)$ is Lipschitz continuous.

A first remark is that in this equation there are *no parameters*, hence it is not directly written in the setting of Theorem 3. Just as one would do in the finite dimensional case, this problem is overcome by performing a step of Birkhoff normal form, in order to start from an unperturbed system which has a twist, and then using the initial actions as parameters.

In order to concentrate on the problems connected with small divisors we shall outline the proof only in the simplified case

$$\begin{cases} \mathrm{i}u_t - u_{xx} + V * u + |u|^2 u \\ u(t, 0) = u(t, \pi) \end{cases} \tag{145}$$

where $V * u$ is convolution with an even function $V = V(x)$, and we consider its Fourier coefficients $\{V_j\}_{j \geq 0}$ as parameters.

Just as (143), this is a Hamiltonian system with respect to the symplectic form

$$\omega(u, v) = 2\mathrm{Im} \int_0^\pi u\bar{v}.$$

In order to highlight the equivalence with the problem of lower dimensional tori, we pass this equation in sin-Fourier series

$$u(t, x) := \sqrt{\frac{2}{\pi}} \sum_{j \in \mathbb{N}} u_j(t) \sin(jx)$$

and we write the Hamiltonian as

$$H = \sum_{j\in\mathbb{N}} (j^2 + V_j)|u_j|^2 + \int_0^\pi |\sum_{j\in\mathbb{N}} u_j \sin(jx)|^4$$

We choose:

- any finite set $\mathcal{S} := \{j_1, \dots, j_n\} \subset \mathbb{N}$;
- any *initial actions* $\mathcal{I} := \{I_1, \dots, I_n\} \in \mathbb{R}_+$;

we fix all the V_j with $j \notin \mathcal{S}$ and keep the rest as free parameters. For example we might fix $V_j = 0\ \forall j \notin \mathcal{S}$ and denote $V_{j_i} = \xi_i$ for $i = 1, \dots, n$.
Now we look for small quasi-periodic solutions of the form

$$\sqrt{\varepsilon} \sum_{i=1}^n \sqrt{I_i} e^{\mathrm{i}\omega_i t} \sin(j_i x) + o(\varepsilon)\,, \quad \omega_i = j_i^2 + \xi_i + O(\varepsilon) \tag{146}$$

For this purpose we pass to action-angle variables all the $\{u_j\}_{j\in\mathcal{S}}$, by writing

$$u_{j_i} = \sqrt{\varepsilon}\sqrt{I_i + y_i} e^{\mathrm{i}x_i}\,,\ i = 1, \dots, n \qquad u_j = \sqrt{\varepsilon} z_j\,, \quad \forall j \notin \mathcal{S}\,.$$

After rescaling the time the Hamiltonian becomes

$$K^{(0)}(I, \xi) + \sum_{i=1}^n (j_i^2 + \xi_i) y_i + \sum_{j\notin\mathcal{S}} j^2 |z_j|^2 + \varepsilon P(y, x, z)\,,$$

namely it has the form (132) with $m = \infty$. Now in order to apply Theorem 3 we have to specify in which space the sequence $\{z_j\}_{j\notin\mathcal{S}}$ lives. Typically one uses a weighted Hilbert space such as

$$\ell_{a,p} := \{z = (z_j)_{j\in\mathbb{N}\setminus\mathcal{S}} : \quad |z|^2_{a,p} := \sum_{j\in\mathbb{N}\setminus\mathcal{S}} \langle j\rangle^{2p} e^{2a|j|} |z_j|^2\}$$

and redefines the domain $D(s, r)$ accordingly by substituting $\mathbb{C}^m$ with $\ell_{a,p}$; see [97] for an analysis of the properties of analytic functions on a Banach space. One also defines the regular Hamiltonians as those analytic Hamiltonians for which the norm (137) (again substituting $\mathbb{C}^m$ with $\ell_{a,p}$) is finite. Also in the infinite dimensional case this class of Hamiltonians is a scale of Banach spaces closed with respect to Poisson brackets, homogeneous projections and which satisfies smoothing estimates for the ultraviolet cut off in the variables x. Since the proof of Theorem 3 depends only on such properties and is uniform in m, we have the same result also in this case.

We have proved that for any choice of $\mathcal{S}, \mathcal{I}$ there exists a Cantor-like set Π^*(explicitly defined in (134), and depending on $\mathcal{S}, \mathcal{I}$) such that for all $\xi \in \Pi^*$ there exists quasi-periodic solutions for (145) of the form (146).

One easily verifies that the conditions (136) hold. However in this infinite dimensional setting this is not enough in order to ensure that the measure of Π^* is positive. By exploiting the fact that $\Omega_j = j^2 + O(\varepsilon)$, one can however verify directly that $|\Pi^*| \sim \gamma$, provided that $\tau > n + 1$.

The same kind of result can be formulated in the more natural case where the potential is multiplicative, and one can prove that for any $\mathcal{S}$ and for most choices of potential there exists analytic solutions such as (146).

This strategy for proving the existence of finite dimensional invariant tori is quite general and can be applied to many dispersive PDEs on an interval with Dirichlet boundary conditions. A similar strategy can be used also for the Klein-Gordon equation, even though the linear dispersion law makes the measure estimates more complex, see [95].

This approach, based on applying Theorem 3 in an infinite dimensional setting, has two main drawbacks:

(A) It relies on the fact that the unperturbed normal frequencies are distinct or at least have finite and uniformly bounded multiplicity as in [47]. In the context of PDEs this gives strong restrictions on the domains. For instance it cannot be applied to equations such as the Nonlinear Schrödinger or the Nonlinear Wave on compact domains without boundary, except in the simplest case of the circle, since the eigenvalues are multiple with unbounded multiplicity.

(B) By construction the change of coordinates which puts the Hamiltonian in normal form must be the time one flow of a *regular* Hamiltonian i.e. with finite norm (137). In the infinite dimensional case this creates unnecessary restrictions, since there exist bounded symplectic changes of variables which are not of this form.

The first results in the direction of removing the assumptions on the multiplicity of eigenvalues, were obtained by using a different strategy, proposed by Craig, Wayne for periodic solutions and then developed by Bourgain. This approach is the infinite dimensional analogue of Theorem 5. Actually the first results were in the infinite dimensional setting, and the applications to finite dimensional systems came afterwards. As in the KAM approach, in order to work in an infinite dimensional setting, one needs some knowledge on the asymptotics of the normal sites; we refer to [24], [26], [15] or [16] for details. We remark that these type of results do not imply any stability of the quasi-periodic solutions, nor the existence of a constant coefficients normal form such as the one in (135).

The first results on stable KAM tori on $\mathbb{T}^d$ were given by Geng and You in [61], for Nonlinear Wave and Beam equations with a convolution potential. The main ideas were: 1.

to exploit the translation invariance of such equations, and to use the consequent constants of motion in order to simplify the small divisor problem; 2. to exploit the fact that the nonlinearities in the Wave and Beam equations are 1-*smoothing* in order to prove the measure estimates (in our notation, this amounts to proving that Π^* has positive Lebesgue measure). The more difficult case of the Nonlinear Schrödinger equation was studied by Eliasson and Kuksin see [57, 56], where the authors deal with an equation with external parameters, like (145), but with $x \in \mathbb{T}^d$. In these papers the authors do not require translation invariance, instead they deal with clusters of multiple eigenvalues. Moreover they introduce the notion of *Töplitz-Lipschitz* hamiltonians in order to handle the measure estimates. We mention also the papers [99],[100], which prove existence and stability of quasi-periodic solutions for the NLS equation without outer parameters. The statement of the result is essentially identical to the one of Theorem 3, but there are two main differences:

1. Due to the complicated resonant structure of the NLS on $\mathbb{T}^d$, there are some *pathological* choices of tangential sites $\mathcal{S}$ on which one is not able to prove existence of quasi-periodic solutions and which are the basis of the construction of *weakly turbulent* solutions as in [49]. More precisely the existence of quasi-periodic solutions is proved for *generic* choices of the tangential sites, i.e., all $\mathcal{S}$ which are not on the zero set of an explicit but very complicated polynomial.

2. There exist positive measure sets of actions in which solutions exist but there are a finite number of linearly unstable directions.

Concerning results on more complicated manifolds, we mention [16], [65] and finally [66] which deals with a non-linear quantum harmonic oscillator.
A breakthrough step in overcoming the restrictions explained in point (B) above was first proposed for the much simpler case of periodic solutions in [68], in order to study Euler's equations of water waves. This strategy was developed and extended to the quasi-periodic case by Baldi Berti and Montalto, who started by considering an equation of the form

$$u_t + u_{xxx} - 6uu_x - \partial_x[(\partial_u f)(x, u, u_x) - \partial_x((\partial_{u_x} f)(x, u, u_x)) = 0\,, \qquad (147)$$

under periodic boundary conditions $x \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$, and assuming that $f(x, u, v) \in C^q$ has a zero of order at least five in $u, v = 0$.

Theorem 6 (Baldi Berti and Montalto, [10]) *For any $\nu \geq 1$ and for all generic choices of tangential sites $S = \{j_1, \ldots, j_\nu\} \subset \mathbb{N}$, the KdV equation (147) possesses small amplitude quasi-periodic solutions of the form*

$$u(t, x) = \sum\nolimits_{j \in S} 2\sqrt{\xi_j}\, \cos(\omega_j t + jx) + o(\sqrt{|\xi|}), \quad \omega_j := j^3 - 6\xi_j j^{-1}\,, \qquad (148)$$

for a "Cantor-like" set of small amplitudes $\xi \in \mathbb{R}^\nu_+$ *with density* 1 *at* $\xi = 0$. *The term* $o(\sqrt{|\xi|})$ *is small in some* H^s*-Sobolev norm,* $s < q$. *These quasi-periodic solutions are linearly stable.*

The proof is done by applying a Nash-Moser scheme, as explained in the proof of Theorem 4. The key problem is in inverting a linear unbounded operator $\mathcal{L}$ of the form

$$\mathcal{L} = \partial_t + (1 + \varepsilon a(x, \omega t))\partial_x^3 + \varepsilon b(x, \omega t)\partial_x + \varepsilon c(x, \omega t)$$

with $\omega \in \mathbb{R}^n$ a diophantine vector. The simplest way to invert $\mathcal{L}$ is to diagonalize it by a bounded change of variables. One could try to construct such change of variables by a KAM scheme: recall that a linear Hamiltonian vector field corresponds to a quadratic Hamiltonian, hence one can try to apply Theorem 3 (putting the Hamiltonian in normal form corresponds to diagonalizing $\mathcal{L}$). This approach however fails, indeed even in the simplest cases it may not be possible to diagonalize $\mathcal{L}$ by using the flow of a regular Hamiltonian. As an example, assume for simplicity that $\mathcal{L} = \partial_t + (\frac{1}{1+\varepsilon a(x)}\partial_x)^3$ where $a(x)$ has zero mean. Then, clearly, the diagonalizing change of variables is

$$u(x) \to v(x) = u(x + \varepsilon\beta(x)), \quad \beta_x = a(x), ,$$

which is not the flow of a regular Hamiltonian but is bounded from H^s to itself for all s. In this simple case the change of variables is constructed by hand directly, in more complicated examples the main feature that one exploits is that $\mathcal{L}$ is a *pseudo-differential* operator. Then the strategy proposed in [10] is:

(i) apply changes of variables which are the flow of pseudo-differential vector fields, in order to conjugate $\mathcal{L}$ to an operator, say $\hat{\mathcal{L}}$, sum of a diagonal operator plus a correction which is a bounded operator of size ε.

(ii) Use a KAM scheme like the one in Theorem 3 in order to diagonalize $\hat{\mathcal{L}}$.

This approach is quite general, it can be adapted to cover also autonomous equations and has allowed to prove existence and stability for quasi-periodic solutions for many fully non linear PDEs on the circle. We mention, among others, the paper [8] where the authors show the existence of quasi-periodic solutions for water waves with gravity.

5 Future Directions

Many natural questions, especially in infinite dimensions, remain widely open in KAM theory. In this final section we briefly mention a (very) few of them.

(i) In finite dimensional Hamiltonian system a basic question is still to fully understand the "Kolmogorov set", i.e., the set of *all* Diophantine invariant maximal tori, including the maximal invariant tori, which are not deformation of integrable tori and which, in general, arise near resonances. In [6] it conjectured that the complement of the Kolmogorov set is, in general, bounded by ε if ε is the size of the perturbation (recall that, as mentioned above, the complementary of the KAM primary tori – namely, the invariant tori which are deformation of integrable tori – may be bounded by a constant times $\sqrt{\varepsilon}$). In [19] it is announced a partial proof of this conjecture in the special case of "mechanical systems", i.e., Hamiltonian systems of the form $|p|^2/2 + \varepsilon f(q)$ with $(p, q) \in \mathbb{R}^d \times \mathbb{T}^d$.

(ii) A very interesting and widely open topic in the study of Hamiltonian PDEs, is the study of maximal tori, or possibly even lower dimensional tori of infinite dimension, see, e.g., [41], [96], [27]. Such results concern problems with external parameters, of the form say (145). The application to more natural *parameterless* equations, such as (143), is still beyond our reach. In particular it would be interesting to understand the regularity of such almost-periodic solutions and whether they can cover positive measure (in any reasonable sense) sets, as in the finite dimensional case.

(iii) Another important open problem is that of proving existence of quasi-periodic solutions for general compact Riemannian manifolds. Up to now the, few, results are confined to the case of Zoll manifolds [65], or Lie groups [16], where there is a very good knowledge of the harmonic analysis.

(iv) The strategy proposed in [10] has allowed to prove existence and stability for many fully non linear PDEs on the circle and has been developed, in the similar setting of reducibility, in order to tackle various classes of PDEs in one space variable. Whether this strategy can be generalized in order to cover higher dimensional cases, both on the torus $\mathbb{T}^d$ or on the line, is a very challenging open problem, in this direction we mention [11], [50].

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 on a closed, non–empty metric space[51] has a unique fixed point, which is obtained as $\lim_{j\to\infty} \Phi^j(u_0)$ for any[52] $u_0 \in X$.

Implicit Function Theorem. *Let*

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

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

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

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 $g : D(x_0,s) \to D(y_0,r)$ satisfying, in particular,

$$\sup_{D(x_0,s)} |g - y_0| \leq 2\|T\| \sup_{D(x_0,s)} |F(y_0,x)| \ . \tag{A.2}$$

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 (A.1), it follows that is a contraction. Furthermore, for any $u \in C(D^m(x_0,s), D^n(y_0,r))$,

$$|\Phi(u) - y_0| \leq |\Phi(u) - \Phi(y_0)| + |\Phi(y_0) - y_0| \leq \frac{1}{2}\|u - y_0\| + \|T\|\|F(y_0,x)\| \leq \frac{1}{2}r + \|T\| \ \frac{r}{2\|T\|} = r \ ,$$

showing that $\Phi : X \to X$. 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 $|y_1 - g(x_1)| = |\Phi(y_1;x_1) - \Phi(g(x_1),x_1)| \leq \alpha|y_1 - g(x_1)|$, 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 of g. Finally, (A.2) follows by observing that $\|g - y_0\| = \|\Phi(g) - y_0\| \leq \|\Phi(g) - \Phi(y_0)\| + \|\Phi(y_0) - y_0\| \leq \frac{1}{2}\|g - y_0\| + \|T\|\|F(y_0,\cdot)\|$, 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 r.h.s.'s of (A.1) 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

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

$$\sup_{D(y_0,r)} \|\mathbb{1}_n - Tf_y\| \leq \frac{1}{2} \ , \qquad T := f_y(y_0)^{-1} \ , \tag{A.3}$$

then there exists a unique C^1 function $g : D(x_0,s) \to D(y_0,r)$ 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 ones holds also in this case.

B Complementary notes

1 Actually, the first instance of small divisor problem solved analytically is the linearization of the germs of analytic functions and it due to C.L. Siegel [109]. [Page 5]

[2] The well–known Newton's tangent scheme is an algorithm, which allows 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 $x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)}$ [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 analysis considered in the present article. [Page 5]

[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 [110]. [Page 6]

[4] A symplectic form on a (even dimensional) manifold is a closed, non–degenerate differential 2–form. The symplectic form $\alpha = dy \wedge dx$ is actually *exact* symplectic, meaning that $\alpha = d(\sum_{i=1} y_i dx_i)$. For general information see [5]. [Page 6]

[5] For general facts about the theory of ODE (such as Picard theorem, smooth dependence upon initial data, existence times,...) see, e.g., [48]. [Page 6]

[6] This terminology is due to the fact that 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. [Page 7]

[7] This terminology is due to the fact that a classical mechanical systems 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 $\sum_{j=1}^d \frac{p_j^2}{2m_i} + V(q)$ and d may be interpreted as the (minimal) number of coordinates necessary to physically describe the system. [Page 7]

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

[9] As standard, U_θ denotes the $(d \times d)$ Jacobian matrix with entries $\dfrac{\partial U_i}{\partial \theta_j} = \delta_{ij} + \dfrac{\partial u_i}{\partial \theta_j}$. [Page 8]

[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 $g : \mathbb{T}^d \to \mathbb{R}$ such that $L = \{(y, x) : \ y = g_x(x) \ , \ x \in \mathbb{T}^d\}$. [Page 8]

[11] Compare [102] 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 $B(y_0, r)\backslash\mathcal{D}_{\kappa,\tau}$ can be bounded by $c_d \kappa r^{d-1}$ for a suitable constant c_d depending only on d; for the simple proof, see, e.g, [41]. [Page 9]

12 The sentence "can be put into the form" means "there exists a symplectic diffeomorphism $\phi : (y,x) \in \mathcal{M} \to (\eta,\xi) \in \mathcal{M}$ 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 $f = O(|y|^{k+1})$. [Page 9]

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 $\|f\|_{C^p(A)} : \sup_{|\alpha| \le p} \sup_A |\partial_x^\alpha f|$. [Page 12]

14 Standard 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: $f_{yyy}\ a \cdot b \cdot c := \sum_{i,j,k=1}^{d} \frac{\partial^3 f}{\partial y_i\, \partial y_j\, \partial y_k}\ a_i b_j c_k$. [Page 14]

15 Standard notation: If f is (a regular enough) function over $\mathbb{T}^d$, its Fourier coefficients are defined as $f_n := \int_{\mathbb{T}^d} f(x)\ e^{-in\cdot x} \frac{dx}{(2\pi)^d}$; where, as usual, $i = \sqrt{-1}$ denotes imaginary unit; for general information about Fourier series see, e.g., [71]. [Page 16]

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 $\|L\| = \sup_{u \neq 0} \|Lu\|/\|u\|$, so that $\|Lu\| \le \|l\|\|u\|$, an estimate, which will be constantly be used; for a general discussion on norms, see, e.g., [73]. [Page 21]

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, $W_{\xi - \frac{\delta}{3}}$, allowing to give the remaining bounds on the smaller domain $W_{\bar\xi}$ (recall that W_s denotes the complex domain $D(0,s) \times \mathbb{T}^d_s$). 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 $\xi' = \xi - \frac{\delta}{3}$, one gets

$$\|s_x\|_{\xi - \frac{\delta}{3}} \le \bar B_0 \frac{2M}{\kappa}\, 3^{k_0} \delta^{-k_0} \ ,$$

which is of the form (54), provided $c \ge (\bar B_0 2 \cdot 3^{k_0})/\kappa$ and $\nu \ge k_0$. 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

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

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 (54) provided $c \ge m(B_2 \bar B_0 2 \cdot 3^{k_0} + B_1)$, $\mu \ge 1$, $\mu \ge 2$ and $\nu \ge k_0 + 2$. The other bounds in (54) follow easily along the same lines. The factor $3C$ in front of $\|\partial_{y'}^2 \widetilde Q\|_0$ has been inserted to simplify later estimates. [Page 21]

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 such 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 [73]). Let Y denote the closed ball (with respect

to the sup–norm) of continuous functions $u : \mathbb{T}^d_{\xi'} \to \mathbb{C}^d$ such that $\|u\|_{\xi'} \le L$. By (55), $|\,\mathrm{Im}\,(x' + \varepsilon u(x'))| < \xi' + \varepsilon_0 L < \xi' + \frac{\delta}{3} = \bar{\xi}$, for any $u \in Y$, and any $x' \in \mathbb{T}^d_{\xi'}$, so that $f : Y \to Y$; notice that, in particular, this means that f sends periodic functions into periodic functions. Moreover, (55) implies also that f is a contraction: if $u, v \in Y$, then, by the mean value theorem, $|f(u) - f(v)| \le L|\varepsilon|\ |u - v|$ (with a suitable choice of norms), so that, by taking the sup–norm, one has $\|f(u) - f(v)\|_{\xi'} < \varepsilon_0 L \|u - v\|_{\xi'}$ 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 (56) holds .
Next, for $(y', x) \in W_{\bar{\xi}}$, by (54), one has $|y' + \varepsilon\beta(y', x)| < \bar{\xi} + \varepsilon_0 L < \bar{\xi} + \frac{\delta}{3} = \xi$ so that (57) holds. Furthermore, since $\|\varepsilon a_x\|_{\bar{\xi}} < \varepsilon_0 L < 1/3$ the matrix $\mathbb{1}_d + \varepsilon a_x$ is invertible with inverse given by the "Neumann series" $(\mathbb{1}_d + \varepsilon a_x)^{-1} = \mathbb{1}_d + \displaystyle\sum_{k=1}^{\infty} (-1)^k (\varepsilon a_x)^k =: \mathbb{1}_d + \varepsilon S(x; \varepsilon)$, so that (58) holds. The proof is finished. [Page 22]

19 From (60) it follows immediately that:
$\langle \partial^2_{y'} Q_1(0, \cdot)\rangle = \langle \partial^2_y Q(0, \cdot)\rangle + \varepsilon \langle \partial^2_{y'} \widetilde{Q}(0, \cdot)\rangle = T^{-1}(\mathbb{1}_d + \varepsilon T \langle \partial^2_{y'} \widetilde{Q}(0, \cdot)\rangle) =: T^{-1}(\mathbb{1}_d + \varepsilon R)$
and, in view of (52) and (60), we see that $\|R\| < L/(2C)$. Therefore, by (61), $\varepsilon_0\|R\| < 1/6 < 1/2$, implying that $(\mathbb{1} + \varepsilon R)$ is invertible and $(\mathbb{1}_d + \varepsilon R)^{-1} = \mathbb{1}_d + \sum_{k=1}^{\infty} (-1)^k \varepsilon^k R^k =: 1 + \varepsilon D$ with $\|D\| \le \|R\|/(\mathbb{1} - |\varepsilon|\|R\|) < L/C$. In conclusion, $T_1 = (\mathbb{1} + \varepsilon R)^{-1} T = T + \varepsilon D T =: T + \varepsilon \widetilde{T}$, $\|\widetilde{T}\| \le \|D\| C \le (L/C) C = L$. [Page 22]

20 Actually, there is quite some freedom in choosing the sequence $\{\xi_j\}$ provided the convergence is not too fast; for general discussion, see, [105], or, also, [30] and [34]. [Page 23]

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 (48) with $\xi = \xi_*$ and $\xi' = \theta\xi_*$, one has $\|\phi_* - \mathrm{id}\,\|_{C^p(B_* \times \mathbb{T}^d)} = \sup_{B_* \times \mathbb{T}^d} \sup_{|\alpha|+|\beta| \le p} |\partial^\alpha_y \partial^\beta_x (\phi_* - \mathrm{id}\,)| \le \sup_{|\alpha|+|\beta| \le p} \|\partial^\alpha_y \partial^\beta_x (\phi_* - \mathrm{id}\,)\|_{\theta\xi_*} \le B_p \|\phi_* - \mathrm{id}\,\|_{\xi_*} 1/(\theta\xi_*)^p \le \mathrm{const}_p\,|\varepsilon|$ with $\mathrm{const}_p := B_p D B M 1/(\theta\xi_*)^p$. An identical estimate hold for $\|Q_* - Q\|_{C^p(B_* \times \mathbb{T}^d)}$. [Page 24]

22 Also in third millennium, however, ε–power expansions turned out to be an important and efficient tool; see [33]. [Page 24]

23 A function $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$ 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., [58]. In fact, the dependence of ϕ_* on $\bar{y}$ is much more regular, compare Remark 11. [Page 25]

24 In fact, notice that inverse powers of κ appear through (49) (inversion of the operator D_ω), therefore one sees that the terms in the first line of (54) may be bounded by $\tilde{c}\kappa^{-2}$ (in defining a one has to apply the operator D_ω^{-1} twice) but then in $P^{(1)}$ (see (27)) there appears $\|\beta\|^2$, so that the constant c in the second line of (54) has the form (73); since $\kappa < 1$, one can replace in (54) c with $\hat{c}\kappa^{-4}$ as claimed. [Page 25]

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 j times Kolmogorov transformation obtaining

j symplectic transformations $\phi_{i+1} : W_{\xi_{i+1}} \to W_{\xi_i}$, 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

$$|\omega|, |E_i|, \|Q_i\|_{\xi_i}, \|T_i\| < C \ , \quad |\varepsilon|^{2^i} L_i := |\varepsilon|^{2^i} c C^\mu \delta_0^{-\nu} 2^{\nu i} M_i \le \frac{\delta_i}{3} \ , \quad \forall\, 0 \le i \le j-1 \ . \qquad (*)$$

By $(*)$, Kolmogorov iteration (**Step 2**) can be applied to H_i and therefore all the bounds described in paragraph **Step 2** holds (having replaced $H, E, ..., \xi, \delta, H', E', ..., \xi'$ with, respectively, $H_i, E_i, ..., \xi_i, \delta_i, H_{i+1}, E_{i+1}, ..., \xi_{i+1}$); in particular (see (62)) one has, for $0 \le i \le j-1$ (and for any $|\varepsilon| \le \varepsilon_0$),

$$|E_{i+1}| \le |E_i| + |\varepsilon|^{2^i} L_i \ , \|Q_{i+1}\|_{\xi_{i+1}} \le \|Q_i\|_{\xi_i} + |\varepsilon|^{2^i} L_i \ , \ \|\phi_{i+1} - \mathrm{id}\,\|_{\xi_{i+1}} \le |\varepsilon|^{2^i} L_i \ , M_{i+1} \le M_i L_i \qquad \text{(C.1)}$$

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 $L_i < DB^i M_i$, 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 (67) for $0 \le i \le j$. Next, we show that, thanks to (66), $(*)$ 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 (65): $|E_j| \le |E| + \sum_{i=0}^{j-1} \varepsilon_0^{2^i} L_i \le \frac{1}{3C} \sum_{i\ge 0} \delta_i < |E| + \frac{1}{6}\sum_{i\ge 1} 2^{-i} < |E| + 1 < C$. The bounds for $\|Q_i\|$ and $\|T_i\|$ are proven in an identical manner. Now, by $(67)_{i=j}$ and (66), $|\varepsilon|^{2^j} L_j (3C\delta_j^{-1}) = DB^j |\varepsilon|^{2^j} M_j \le DB^j (DB\varepsilon_0 M)^{2^j} / (DB^{j+1}) \le 1/B < 1$, 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 (67) *for all* $i \ge 0$. To check (68), we observe that $|\varepsilon|^{2^i} L_i = \frac{\delta_0}{3C 2^i}\, DB^i |\varepsilon|^{2^i} M_i \le \frac{1}{2^{i+1}} (|\varepsilon| DBM)^{2^i} \le \left(\frac{|\varepsilon| DBM}{2}\right)^{i+1}$ and therefore $\sum_{i\ge 0} |\varepsilon|^{2^i} L_i \le \sum_{i\ge 1} \left(\frac{|\varepsilon| DBM}{2}\right)^i \le |\varepsilon| DBM$. Thus, $\|Q - Q_*\|_{\xi_*} \le \sum_{i\ge 0} \|\tilde{Q}_i\|_{\xi_i} \le |\varepsilon|^{2^i} L_i \le |\varepsilon| DBM$; 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 (68) and the proof of Claim **C**. [Page 25]

26 In fact, observe: (i) given any integer vector $0 \ne n \in \mathbb{Z}^d$ with $d \ge 2$, one can find $0 \ne m \in \mathbb{Z}^d$ such $n \cdot m = 0$; (ii) the set $\{tn : \ t > 0 \text{ and } n \in \mathbb{Z}^d\}$ 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 $K_y(y(t, n)) \cdot m = tn \cdot m = 0$. [Page 28]

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]. [Page 28].

28 Compare (91) but observe, that, since $\hat{P}$ is a trigonometric polynomial, in view of Remark 9–(ii), g in (97) 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 is important to see explicitly how the various quantities depend upon ε; this is shortly discussed after Proposition 2. [Page 29]

29 In fact: $\|\check{P}\|_{r, \xi - \frac{\delta}{2}} \le M \sum_{|n|>N} e^{-|n|\frac{\delta}{2}} \le M e^{-\frac{\delta}{4}N} \sum_{|n|>N} e^{-|n|\frac{\delta}{4}} \le M e^{-\frac{\delta}{4}N} \sum_{|n|>0} e^{-|n|\frac{\delta}{4}} \le \mathrm{const}\, M e^{-\frac{\delta}{4}N} \delta^{-d} \le |\varepsilon| M$ if (107) holds and N is taken as in (105). [Page 30]

30 Apply the IFT of Appendix A (with r replaced by $\bar r$, x_0 by 0 and s by $|\varepsilon|$) 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 (112), $\|\mathbb{1}_d - TF_y\| \le \|\mathbb{1}_d - TK_{yy}\| + |\varepsilon|\|\partial_y^2 P_0\| \le \|T\|\|K_{yyy}\|\bar r + \|T\||\varepsilon|\|\partial_y^2 P_0\| \le C^2 2\frac{\bar r}{r} + C|\varepsilon|\frac{4}{r^2}M \le 2C^2\frac{\bar r}{r} + \frac{|\varepsilon|M}{2r\bar r} \le \frac14 + \frac1{16} < \frac12$; also: $2\|T\|\|F(y_0,\eta)\| = 2\|T\||\eta|\partial_y P_0(y_0)\| < 2C|\varepsilon|M\frac{2}{r} \le \frac{\bar r}{2}$ (where last inequality is due to the second condition in (112)), showing that conditions (A.1) are fulfilled. Eq. (113) comes from (A.2). Finally, by Cauchy estimates and (113), $\|A\| \le C(C\frac{2}{r}\cdot\frac{4CM}{r}|\varepsilon| + |\varepsilon|\frac{4}{r^2}M) \le \frac{12C^3M}{r^2}|\varepsilon|$ and (115) follows. [Page 32]

31 Recall note 18 and notice that $(\mathbb{1}_d + A)^{-1} = \mathbb{1}_d + D$ with $\|D\| \le \frac{\|A\|}{1-\|A\|} \le \frac{16}{13}\|A\| < \frac{192}{13}C^3M|\varepsilon|/r^2$, where last two inequalities are due to (115). [Page 32]

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). [Page 32]

33 By (119) and (55), $|\varepsilon|\|g_x\|_{\bar r,\bar\xi} \le |\varepsilon|rL \le r/2$ so that, by (117), if $y' \in D_{\bar r/2}(y_1)$, then $y' + \varepsilon g_x(y',\varphi(y',x')) \in D_r(y_0)$. [Page 33]

34 The first requirement in (124) is equivalent to require that $r_0 \le r$, which implies that if $\bar r$ is *defined* as the r.h.s. of (109), then $\bar r \le r/2$ as required in (111). Next, the first requirement in (112) at the $(j+1)^{\text{th}}$ 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 (112), which at the $(j+1)^{\text{th}}$ 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 (55)), which, in turn, is easily controlled along the lines explained in the note 25. [Page 33]

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 $f = (f_1,f_2) : (y,x) \in A \to f(y,x) \in A$, 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 Hamiltonian context. After Poincaré the theory of area–preserving maps became, in itself, a very reach 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., [70]. [Page 34]

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 (69). We recall the definition of Hölder norms: If $\ell = \ell_0 + \mu$ with $\ell_0 \in \mathbb{Z}_+$ and $\mu \in (0,1)$, then $\|f\|_{C^\ell} := \|f\|_{C^{\ell_0}} + \sup_{|\alpha|=\ell_0}\sup_{0<|x-y|<1}\frac{|\partial^\alpha f(x)-\partial^\alpha f(y)|}{|x-y|^\mu}$; $C^\ell(\mathbb{R}^d)$ denotes the Banach space of functions with finite C^ℓ norm. [Page 34]

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 (70). [Page 34]

38 For general references and discussions about Lemma 2 and 3, see, [86] and [113]; an elementary detailed proof can be found, also, in [35]. [Page 34].

39 **Proof of claim M** The first step of the induction consists in constructing $\Phi_0 = \phi_0$: this follows from Kolmogorov 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 $|\varepsilon|\|P^1\|_{\xi_1} \le C|\varepsilon|\|P\|_{C^0}$ by the first inequality in (125)). Now, assume that (129) and (130) 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 (65) for, respectively, $K_0 := K$

and K_i). To determine ϕ_{j+1}, observe that, by (129), 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 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 (125), $\|P_{j+1}\|_{\alpha\xi_{j+1}} \le \|P^{j+2} - P^{j+1}\|_{\xi_{j+2}} \le c\|P\|_{C^\ell}\xi_{j+1}^\ell$ and the smallness condition (69) 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 $|\varepsilon| < D^{-1}\xi^a$, for some $a > 0$. 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 $\Phi_j : W_{\alpha\xi_{j+1}} \to X_{\xi_{j+1}}$. By the inductive assumptions and the real–analyticity of Φ_j, one has that, for $z \in W_{\alpha\xi_{j+1}}$, $|\,\mathrm{Im}\,\Phi_j(z)| = |\,\mathrm{Im}\,(\Phi_j(z) - \Phi_j(\,\mathrm{Re}\,z))| \le |\Phi_j(z) - \Phi_j(\,\mathrm{Re}\,z)| \le \|\partial\Phi_j\|_{\alpha\xi_{j+1}}|\,\mathrm{Im}\,z| \le (1 + \|\partial(\Phi_{j+1} - \mathrm{id}\,)\|_{\alpha\xi_{j+2}})\alpha\xi_{j+1} < \sqrt{2}\alpha\xi_{j+1} = \xi_{j+1}$ so that $\Phi_j : W_{\alpha\xi_{j+1}} \to X_{\xi_{j+1}}$. The remaining inductive assumptions in (130) with j replaced by $j+1$ are easily checked by arguments similar to those used in the induction proof of Claim **C** above. [Page 35]

40 See, e.g., the Proposition at page 58 of [34] 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 (83) in Arnold's scheme with W_j in (82) and taking $\mathcal{C} = \mathcal{C}_* := \{y' = \lim_{j\to\infty} y_j(\omega) : \omega \in B \cap K_y^{-1}(\mathcal{D}_{\kappa,\tau})\}$ and $y_j(\omega) := y_j$ is as in (83). [Page 41]

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. [Page 42]

42 In fact, Poincaré was not at all convinced of the convergence of such series: see chapter XIII, n$^{\mathrm{o}}$ 149, entitled "Divergence des séries de M. Lindstedt", of his book [91]. [Page 42]

43 (71) 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., [51]. [Page 43]

44 Actually, it is not known if such tori are KAM tori in the sense of the definitions given above! [Page 43]

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 [40] 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 [82]. [Page 44]

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 $V(\omega_1 t, ..., \omega_n t)$) one looks at the equation $\ddot{q} = (V(\omega t) - \lambda)q$, which is the q–flow of the Hamiltonian ϕ_H^t $H = H(p, q, I, \varphi; \lambda) := \frac{p^2}{2} + [\lambda - V(\varphi)]\frac{q^2}{2}$ where $(p, q) \in \mathbb{R}^2$ and $(I, \varphi) \in \mathbb{R}^n \times \mathbb{T}^n$ (with respect to the standard form $dp \wedge dq + dI \wedge d\varphi$) 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 $\frac{p^2}{2} + \lambda\frac{q^2}{2}$ 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 $H = \omega \cdot I + \sqrt{\lambda} J - \frac{V(\varphi)}{\sqrt{\lambda}} \sin^2 \theta$. 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 [34]. [Page 44]

47 The value 10^{-50} 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) $\simeq 8.4 \cdot 10^{-52}$. [Page 45]

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 the so–called "interval arithmetic"; see, e.g., Appendix C of [33] and references therein. [Page 45]

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 [91], Volume I, chapter III. [Page 36].

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 [94], where, most interestingly, m is allowed to be ∞. [Page 36]

51 I.e., a map $\Phi : X \to X$ for which $\exists\, 0 < \alpha < 1$ such that $d(\Phi(u), \Phi(v)) \leq \alpha d(u, v)$, $\forall u, v \in X$, $d(\cdot, \cdot)$ denoting the metric on X; for generalities on metric spaces, see, e.g., [73]. [Page 51]

52 $\Phi^j = \Phi \circ \cdots \circ \Phi$ j–times. In fact, let $u_j := \Phi^j(u_0)$ and notice that, for each $j \geq 1$ $d(u_{j+1}, u_j) \leq \alpha d(u_j, u_{j-1}) \leq \alpha^j d(u_1, u_0) =: \alpha^j \beta$, so that, for each $j, h \geq 1$, $d(u_{j+h}, u_j) \leq \sum_{i=0}^{h-1} d(u_{j+i-1}, u_{j+i}) \leq \sum_{i=0}^{h-1} \alpha^{j+i} \beta \leq \alpha^{j+1} \beta / (1 - \alpha)$, showing that $\{u_j\}$ is a Cauchy sequence. Uniqueness is obvious. [Page 51]

References

[1] L.V. Ahlfors, *Complex analysis. An introduction to the theory of analytic functions of one complex variable.* Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York, 1978. xi+331 pp.

[2] V. I. Arnol'd, *Proof of a Theorem by A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian*, Russian Math. Survey **18** (1963), 13-40.

[3] V. I. Arnol'd, *Small divisor problems in classical and Celestial Mechanics*, Russian Math. Survey **18** (1963), 85-191.

[4] V. I. Arnold, *Instability of dynamical systems with many degrees of freedom.* (Russian) Dokl. Akad. Nauk SSSR **156** (1964) 9–12

[5] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics* Series: Graduate Texts in Mathematics **60**, Originally published by Nauka, Moscow, 1974 2nd ed. 1989. Corr. 4th printing, 1997, XVI, 509

[6] V. I. Arnol'd, V.V. Kozlov, and A.I. Neishtadt (editor), *Mathematical Aspects of Classical and Celestial Mechanics*, Dynamical Systems III, Series: Encyclopaedia of Mathematical Sciences **3** (2006), Springer-Verlag 3rd ed.

[7] A. Avila and R. Krikorian, *Reducibility or nonuniform hyperbolicity for quasiperiodic Schrödinger cocycles* , Annals of Mathematics, **164** (2006), 911–940

[8] P. Baldi, M. Berti, E. Haus and R. Montalto, *Time quasi-periodic gravity water waves in finite depth*, preprint, arXiv:1708.01517 (2017).

[9] P. Baldi, M. Berti and R. Montalto, *KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation*, Math Annalen **359** (2014), 471-536.

[10] P. Baldi, M. Berti and R. Montalto, *KAM for autonomous quasi-linear perturbations of KdV*, Ann. I. H. Poincaré (C) Anal. Non Linéaire (2015).

[11] D. Bambusi, B. Grebert, A. Maspero, D. Robert *Reducibility of the Quantum Harmonic Oscillator in d-dimensions with Polynomial Time Dependent Perturbation.* Analysis and PDEs, to appear.

[12] M. Berti, L. Biasco, *Branching of Cantor Manifolds of Elliptic Tori and Applications to PDEs*, Communications in Mathematical Physics, **305**, 3, (2011), pp 741-796.

[13] M. Berti, L. Biasco, M. Procesi, *KAM theory for the reversible derivative wave equation*, Arch. Rational Mech. Anal., 212 , 905-955 (2014).

[14] M. Berti, P. Bolle, *A Nash-Moser approach to KAM theory*, Fields Institute Communications, volume 75 "Hamiltonian PDEs and Applications", 255-284 (2014).

[15] M. Berti, P. Bolle, *Quasi-periodic solutions with Sobolev regularity of NLS on* $\mathbb{T}^d$ *with a multiplicative potential*, Eur. Jour. Math. 15, 229-286 (2013).

[16] M. Berti, L. Corsi, M. Procesi, *An abstract Nash-Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous spaces* , Commun. Math. Phys. 334, no. 3, 1413-1454 (2015).

[17] M. Berti, T. Kappeler, R. Montalto, *Large KAM tori for perturbations of the dNLS equation* , preprint arXiv:1603.09252, to appear on Astérisque

[18] M. Berti, R. Montalto, *Quasi-periodic standing wave solutions of gravity-capillary water waves*, preprint arXiv:1602.02411, to appear in Memoirs of the AMS, MEMO 891 (2016).

[19] L. Biasco; L. Chierchia *On the measure of Lagrangian invariant tori in nearly-integrable mechanical systems.* Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. 26 (2015), no. 4, 423–432

[20] L. Biasco; L. Chierchia; *Explicit estimates on the measure of primary KAM tori.* Ann. Mat. Pura Appl. (4) 197 (2018), no. 1, 261–281.

[21] J. Bourgain, *Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE*, Internat. Math. Res. Notices, 11 (1994).

[22] J. Bourgain, *On Melnikov's persistency problem*. Math. Res. Lett. 4 (1997), no. **4**, 445–458

[23] J. Bourgain, *Gibbs measures and quasi-periodic solutions for nonlinear Hamiltonian partial different equations*, 23-43, Gelfand Math. Sem., Birkhäuser Boston, Boston, MA (1996).

[24] J. Bourgain, *Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations*, Annals of Math. 148, 363-439 (1998).

[25] J. Bourgain, *Periodic solutions of nonlinear wave equations*, Harmonic analysis and partial differential equations, Chicago Lectures in Math., Univ. Chicago Press, 69-97 (1999).

[26] J. Bourgain, *Green's function estimates for lattice Schrödinger operators and applications*, Annals of Mathematics Studies 158, Princeton University Press, Princeton (2005).

[27] J. Bourgain, *On invariant tori of full dimension for the 1d periodic NLS*, J. Funct. Anal. 229 (2005).

[28] H.W. Broer, *KAM theory: the legacy of A. N. 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, 315–333, Erven P. Noordhoff N.V., Groningen, 1957]. Bull. Amer. Math. Soc. (N.S.) 41 (2004), no. **4**, 507–521 (electronic).

[29] H.W. Broer, G.B. Huitema, M.B. Sevryuk, *Quasi-periodic motions in families of dynamical systems. Order amidst chaos*, Lecture Notes in Mathematics, 1645. Springer-Verlag, Berlin, 1996.

[30] A. Celletti, and L. Chierchia, *Rigorous estimates for a computer-assisted KAM theory*, J. Math. Phys. **28** (1987), 2078-2086.

[31] A. Celletti, and L. Chierchia, *Construction of analytic KAM surfaces and effective stability bounds*, Commun. in Math. Physics **118** (1988), 119-161.

[32] A. Celletti, and L. Chierchia, *A Constructive Theory of Lagrangian Tori and Computer-assisted Applications*, Dynamics Reported (C.K.R.T. Jones, U. Kirchgraber, H.O. Walther Managing Editors), Springer-Verlag, **4** (New Series) (1995), 60-129.

[33] A. Celletti, and L. Chierchia, *KAM stability and celestial mechanics*, Mem. Amer. Math. Soc. **187** (2007), no. 878, viii+134 pp.

[34] L. Chierchia, *Quasi-periodic Schrdinger operators in one dimension, Absolutely continuous spectra, Bloch waves and integrable Hamiltonian systems.* "Quaderni del Consiglio Nazionale delle Ricerche", (Firenze, Italia), 127 pp. (1986)
download: http://www.mat.uniroma3.it/users/chierchia/REPRINTS/CNR86.pdf

[35] L. Chierchia, *KAM Lectures*, in "Dynamical Systems. Part I: Hamiltonian Systems and Celestial Mechanics", Pubblicazioni della Classe di Scienze, Scuola Normale Superiore, Pisa. Pagg. 1-56. Centro di Ricerca Matematica "Ennio De Giorgi": Proceedings (2003)

[36] L. Chierchia, *KAM theory and Celestial Mechanics*, Encyclopedia of Mathematical Physics, eds. J.-P. Franoise, G.L. Naber and S.T. Tsou Oxford: Elsevier, 2006 (ISBN 978-0-1251-2666-3), vol. 3, pp. 189-199

[37] L. Chierchia, *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** *(1954), 527–530; MR0068687]*, Regul. Chaotic Dyn., **13** (2008), no. 2, 130–139.

[38] L. Chierchia, and C. Falcolini, *A direct proof of a theorem by Kolmogorov in Hamiltonian systems*, Ann. Sc. Norm. Su Pisa, Serie IV, **X**XI (1994), 541-593.

[39] L. Chierchia, and G. Gallavotti, *Smooth prime integrals for quasi- integrable Hamiltonian systems*, Il Nuovo Cimento, **67 B** (1982), 277–295.

[40] L. Chierchia, and G. Gallavotti, *Drift and diffusion in phase space.* Ann. Inst. Henri Poincaré, Phys. Théor., **60**, 1–144 (1994); see also *Erratum*, Ann. Inst. Henri Poincaré, Phys. Théor., **68**, no. 1, 135 (1998)

[41] L. Chierchia, and P. Perfetti, *Second order Hamiltonian equations on $\mathbb{T}^\infty$ and almost-periodic solutions*, Journal of Diff. Equations, **116**, No. 1, 172–201 (1995)

[42] L. Chierchia, and G. Pinzari *The planetary N-body problem: symplectic foliation, reductions and invariant tori.* Invent. Math. 186 (2011), no. 1, 177

[43] L. Chierchia, and G. Pinzari *Properly–degenerate KAM theory (following V. I. Arnold).* Discrete Contin. Dyn. Syst. Ser. S 3 (2010), no. 4, 545–578

[44] L. Chierchia, and G. Pinzari *Planetary Birkhoff normal forms.* J. Mod. Dyn. 5 (2011), no. 4, 623–664

[45] L. Chierchia, and G. Pinzari *Metric stability of the planetary N–body problem. Proceedings of the International Congress of Mathematicians–Seoul 2014.* Vol. III, 547–570, Kyung Moon Sa, Seoul, 2014

[46] L. Chierchia, and D. Qian, *Moser's theorem for lower dimensional tori* Journal of Diff. Equations, **206** (2004) 55–93

[47] L. Chierchia, and J. You, *KAM Tori for 1D Nonlinear Wave Equations with Periodic Boundary Conditions*, Commun. Math. Phys. 211, 497-525 (2000).

[48] E.A. Coddington, and N. Levinson, *Theory of Ordinary Differential Equations*, New York, McGraw-Hill, 1955. 14+429 pp.

[49] J Colliander, M Keel, G Staffilani, H Takaoka, and T. Tao: *Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation*, Inventiones mathematicae **181** (1), 39–113, 2010.

[50] L. Corsi, R. Montalto, *Quasi-periodic solutions for the forced Kirchhoff equation on $\mathbb{T}^d$*, arXiv:1802.04139 (2018)

[51] A. Delshams, P. Gutierrez, *Effective stability and KAM theory.* J. Differential Equations, **128** (1996), no. 2, 415–490

[52] L. H. Eliasson: *Perturbations of stable invariant tori for Hamiltonian systems*, Ann. Scuola Norm. Sup. Pisa, Cl. Sci., **15** (1988), 115–147.

[53] L. H. Eliasson, *Floquet solutions for the 1-dimensional quasi-periodic Schrödinger equation*, Comm. Math. Phys. 146 (1992), 447–482

[54] L. H. Eliasson, *Absolutely convergent series expansion for quasi-periodic motions*, Math. Phys. Electron. J. **2** (1996), Paper 4, 33 pp. (electronic); see also: Report 2–88, Dept. of Math., University of Stockholm (1988)

[55] L. H. Eliasson, Grebert B., and S. B. Kuksin, *KAM for the nonlinear Beam equation*, Geometric and Functional Analysis, Volume 26, Issue 6, pp 1588-1715 (2016).

[56] L. H. EliassonH., and S. B. Kuksin, *On reducibility of Schrödinger equations with quasiperiodic in time potentials*, Comm. Math. Phys. 286, 125-135 (2009).

[57] L. H. Eliasson, and S. B. Kuksin, *KAM for non-linear Schrödinger equation*, Annals of Math., 172, 371-435 (2010).

[58] L. C. Evans, and R. F. Gariepy, *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992. viii+268 pp.

[59] J. Féjoz, *Démonstration du 'théorème d'Arnol'd' sur la stabilité du système planétaire (d'après Herman).* (French), Ergodic Theory Dynam. Systems **2**4 (2004), no. 5, 1521-1582.

[60] G. Gallavotti, *Twistless KAM tori, quasi flat homoclinic intersections and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review*, Reviews on Mathematical Physics **6** (1994), 343-411.

[61] J. Geng, and J. You, *A KAM Theorem for Hamiltonian Partial Differential Equations in Higher Dimensional Spaces*, Communications in Mathematical Physics, Volume 262, Issue 2, pp 343-372 (2006).

[62] J. Geng, X. Xu, and J. You, *An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation*, Adv. Math. 226, 5361-5402 (2011).

[63] S. Graff: *On the continuation of stable invariant tori for Hamiltonian systems*, J. Differential Equations, **15** (1974), 1–69.

[64] B. Grébert, and T. Kappeler, *The defocusing NLS equation and its normal form*, European Mathematical Society (EMS), Zrich, 2014.

[65] B. Grébert and E. Paturel, *KAM for the Klein Gordon equation on* $\mathbb{S}^d$. Boll. Unione Mat. Ital. 9 (2016), no. 2, 237–288.

[66] B. Grébert and E. Paturel, *KAM for the quantum harmonic oscillator.* Comm. Math. Phys. 307 (2011), no. 2, 383–427.

[67] M. Hénon, *Explorationes numérique du problème restreint IV: Masses egales, orbites non periodique*, Bullettin Astronomique **3**, 1, fasc. 2 (1966), 49-66.

[68] G. Iooss, P. I. Plotnikov, and J. F. Toland, *Standing waves on an infinitely deep perfect fluid under gravity*, Arch. Rational Mech. Anal. 177 no. 3, 367-478 (2005).

[69] T. Kappeler, and Pöschel J., *KAM and KdV*, Springer, 2003.

[70] A. Katok, and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. **54**, Cambridge Univ. Press, 1995, xviii+802

[71] Y. Katznelson, *An introduction to harmonic analysis.* Third edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2004. xviii+314 pp.

[72] A. N. Kolmogorov, *On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian*, Dokl. Akad. Nauk. SSR **9**8 (1954), 527-530

[73] A. N. Kolmogorov, and S.V. Fomin *Elements of the Theory of Functions and Functional Analysis*, Courier Dover Publications, (1999) 288 pp.

[74] S. B. Kuksin, *Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum*, Funktsional. Anal. i Prilozhen. 21, no. 3, 22–37, 95, 1987.

[75] S.B. Kuksin, *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)** (1988), no. 3, 396–412, 431; translation in Math. USSR-Sb. **64** (1989), no. 2, 397–413

[76] S. B. Kuksin, *A KAM theorem for equations of the Korteweg-de Vries type*, Rev. Math. Phys., 10, 3, (1998), 1-64.

[77] S. B. Kuksin, *Analysis of Hamiltonian PDEs*, Oxford Lecture Series in Mathematics and its Applications, 19. Oxford University Press (2000).

[78] S.B. Kuksin, *Fifteen years of KAM for PDE*. Geometry, topology, and mathematical physics, 237–258, Amer. Math. Soc. Transl. Ser. 2, **212**, Amer. Math. Soc., Providence, RI, 2004

[79] S. B. Kuksin, and J. Pöschel, *Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation*, Annals of Math. 2 143, (1996), 149-179.

[80] J. Liu, and X. Yuan, *A KAM theorem for Hamiltonian Partial Differential Equations with Unbounded Perturbations*, Comm. Math. Phys., 307 (3), 629-673 (2011).

[81] J. N. Mather, *Nonexistence of invariant circles*, Erg. Theory and Dynam. Systems **4** (1984), 301-309

[82] J.N. Mather, *Arnol'd diffusion. I. Announcement of results.* (Russian) Sovrem. Mat. Fundam. Napravl. 2 (2003), 116–130 (electronic); translation in J. Math. Sci. (N. Y.) 124 (2004), no. 5, 5275–5289

[83] V. K. Melnikov: *On certain cases of conservation of almost periodic motions with a small change of the Hamiltonian function.* (Russian) Dokl. Akad. Nauk SSSR **165** 1965 1245–1248

[84] J. K. Moser, *A new technique for the construction of solutions of nonlinear differential equations.* Proc. Nat. Acad. Sci. U.S.A. **47** 1961 1824–1831

[85] J. K. Moser, *On invariant curves of area-preserving mappings of an annulus*, Nach. Akad. Wiss. Göttingen, Math. Phys. Kl. II **1** (1962), 1-20

[86] J. K. Moser, *A rapidly convergent iteration method and non-linear partial differential equations*, Ann. Scuola Norm. Sup. Pisa **2**0 (1966), 499-535

[87] J. K. Moser, *Convergent series expansions for quasi-periodic motions*, Math. Ann. **1**69 (1967), 136–176

[88] J. K. Moser, *On the construction of almost periodic solutions for ordinary differential equations.* 1970 Proc. Internat. Conf. on Functional Analysis and Related Topics (Tokyo, 1969) pp. 60–67 Univ. of Tokyo Press, Tokyo

[89] J. K. Moser,*Integrable Hamiltonian systems and spectral theory*, Lezioni Fermiane. [Fermi Lectures] Scuola Normale Superiore, Pisa, 1983. iv+85 pp.

[90] J. Nash, *The imbedding problem for Riemannian manifolds.* Ann. of Math. (2) **63** (1956), 20–63

[91] H. Poincarè, *Les méthodes nouvelles de la Mécanique céleste*, tome 1, Paris, Gauthier-Villars, 1892, 385 pp.; tome 2, Paris, Gauthier-Villars, 1893, VIII + 479 pp.; tome 3, Paris, Gauthier-Villars, 1899, 414 pp.

[92] J. Pöschel, *Integrability of Hamiltonian sytems on Cantor sets*, Comm. Pure Appl. Math., **35** (1982), 653–695.

[93] J. Pöschel, *On Elliptic Lower Dimensional Tori in Hamiltonian Systems*, Math. Z., **202** (1989), 559–608.

[94] J. Pöschel, *A KAM-theorem for some nonlinear PDEs*, Ann. Scuola Norm. Sup. Pisa, Cl. Sci., **23** (1996), 119–148.

[95] J. Pöschel, *Quasi-periodic solutions for a nonlinear wave equation*, Comment. Math. Helv., 71, no. 2, (1996) 269-296.

[96] J. Pöschel, *On the construction of almost periodic solutions for a nonlinear Schrödinger equation*, Ergodic Theory Dynam. Systems 22, no. 5, 1537-1549 (2002).

[97] J. Pöschel, E. Trubowitz, *Inverse Spectral Theory.* Academic Press, Boston, (1987).

[98] A. S. Pyartli, *Diophantine approximations of submanifolds of a Euclidean space*, Funktsional. Anal. i Prilozhen. 3 (1969), no. 4, 59–62 (Russian). English translation in: Funct. Anal. Appl. 3 (1969), 303–306.

[99] C. Procesi, and M. Procesi, *A KAM algorithm for the non-linear Schrödinger equation*, Advances in Math. 272 , 399-470 (2015).

[100] C. Procesi, and M. Procesi, *Reducible quasi-periodic solutions for the Non Linear Schrödinger equation*, BUMI, (2016), Vol. 9, N. 2.

[101] H. Rüssmann, *Kleine Nenner. I. Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes.* (German) Nachr. Akad. Wiss. G̈ttingen Math.-Phys. Kl. II **1970** 1970 67–105

[102] H. Rüssmann, *On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus* In: Jürgen Moser (editor): Dynamical Systems, Theory and Applications. Lecture Notes in Physics **38**, 598-624. Springer, Berlin, Heidelberg, New York 1975

[103] H. Rüssmann, *Note on sums containing small divisors*, Comm. Pure Appl. Math. **2**9, 6 (1976), 755-758.

[104] H. Rüssmann, *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 Univ. Press, Cambridge, 1989, pp. 5–18.

[105] H. Rüssmann, *On the one-dimensional Schrödinger equation with a quasiperiodic potentia*l. Nonlinear dynamics (Internat. Conf., New York, 1979), pp. 90–107, Ann. New York Acad. Sci., **357**, New York Acad. Sci., New York, 1980.

[106] H. Rüssmann, *Invariant tori in non-degenerate nearly integrable Hamiltonian systems*, Regul. Chaotic Dyn. **6**, 2 (2001), 119-204.

[107] D. Salamon, *The Kolmogorov-Arnold-Moser theorem* Math. Phys. Electron. J. **10** (2004), Paper 3, 37 pp. (electronic)

[108] M. B. Sevryuk, *The Classical KAM Theory and the dawn of the twenty–first century*, Moscow Mathematical Journal Volume 3, Number 3, July–September 2003, Pages 1113–1144

[109] C. L. Siegel, *Iteration of analytic functions*, Ann. Math. **4**3 (1942), 607-612

[110] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Volume 1, 3rd Edition, Publish or Perish, inc; 3rd Ed edition, 1999

[111] C. E. Wayne, *Periodic and quasi–periodic solutions for nonlinear wave equations via KAM theory.* Commun. Math. Phys. **127**, 479528 (1990)

[112] H. Whitney, *Analytic extensions of differentiable functions defined in closed sets.* Trans. Amer. Math. Soc. **36** (1934), no. 1, 63–89

[113] E. Zehnder, *Generalized implicit function theorems with applications to some small divisor problems. I*, Comm. Pure Appl. Math. **2**8 (1975), 91-14.

[114] E. Zehnder, *Generalized implicit function theorems with applications to some small divisor problems. II*, Comm. Pure Appl. Math. **2**9 (1976), no. 1, 49-111.

[115] Zakharov et. al *What Is Integrability?* Springer series in Nonlinear Dynamics (1991)

Books and Reviews

V. I. Arnol'd, S.P. Novikov, (Eds.), Dynamical Systems IV, *Symplectic Geometry and its Applications*, Series: Encyclopaedia of Mathematical Sciences , Vol. **4**, Volume package Dynamical Systems, Original Russian edition published by VINITI, Moscow, 1985, 2nd expanded and rev. ed., 2001

A. Berretti, A. Celletti, L. Chierchia, and C. Falcolini, *Natural Boundaries for Area-Preserving Twist Maps*, Journal of Statistical Physics **66** (1992), 1613–1630

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

H. W. Broer, and M. B. Sevryuk, *KAM Theory: quasi-periodicity in dynamical systems*, in "Handbook of Dynamical Systems" (Elsevier), to appear

A. Celletti, and L. Chierchia, *KAM tori for N-body problems: a brief history*, Celestial Mech. Dynam. Astronom. **95** (2006), no. 1-4, 117–139

L. Chierchia, and C. Falcolini, *Compensations in small divisor problems*, Comm. Math. Phys. **1**75 (1996), 135-160

H. S. Dumas, *The KAM story. A friendly introduction to the content, history, and significance of classical Kolmogorov—Arnold–Moser theory.* World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014. xvi+361 pp. ISBN: 978–981–4556–58–3

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

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

B. Grébert, and T. Kappeler *The defocusing NLS equation and its normal form*, European Mathematical Society (EMS), Zrich, 2014.

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

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

M. Herman, *Inégalités a priori pour des tores lagrangiens invariants par des difféomorphismes symplectiques*, Publications Mathématiques de l'IHÉS, **70** (1989), p. 47–101

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

T. Kappeler, J. Pöschel, *KdV & KAM*, Springer, 2003. XIII, 279 pp.

H. Koch, *A renormalization group for Hamiltonians, with applications to KAM tori*, Erg. Theor. Dyn. Syst. **19** (1999), 475-521

S. B. Kuksin, *Nearly integrable infinite dimensional Hamiltonian systems*. Lecture Notes in Mathematics, **1556**, Berlin: Springer, 1993

O. E. Lanford III, *Computer assisted proofs in analysis*, Physics A **124** (1984), 465–470

J. Laskar, *Large scale chaos and marginal stability in the solar system*, Celestial Mechanics and Dynamical Astronomy **64** (1996), 115–162

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

R. de la Llave, A. Gonzàlez, À. Jorba, and J. Villanueva, *KAM theory without action-angle variables*, Nonlinearity 18 (2005), no. **2**, 855–895

R. de la Llave, and D. Rana, *Accurate strategies for small divisor problems*, Bull. Amer. Math. Soc. (N. S.) **22**, 1 (1990), 85–90

V. F. Lazutkin, *KAM theory and semiclassical approximations to eigenfunctions*, Springer-Verlag, Berlin, 1993

R. S. MacKay, *Transition to chaos for area-preserving maps*, Lectures Notes in Physics **247** (1985), 390–454

K. R. Meyer, and D. Schmidt (editors), *Computer Aided Proofs in Analysis*, Springer-Verlag, Berlin, 1991

J. K. Moser, *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, N. J.; University of Tokyo Press, Tokyo, 1973

J. Pöschel, *A lecture on the classical KAM theorem, Smooth ergodic theory and its applications* (Seattle,WA, 1999), Proc. Sympos. Pure Math., vol. 69, Amer. Math. Soc., Providence, RI, 2001, pp. 707–732

D. Salamon, and E. Zehnder, *KAM theory in configuration space*, Comment. Math. Helvetici **6**4 (1989), 84-132.

M. B. Sevryuk, *The lack-of-parameters problem in the KAM theory revisited*, Hamiltonian systems with three or more degrees of freedom (SAgaró, 1995), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 533, Kluwer Acad. Publ., Dordrecht, 1999, pp. 568–572

M. B. Sevryuk, *KAM-stable Hamiltonians*, J. Dynam. Control Systems 1 (1995), no. **3**, 351–366

C. L. Siegel, and J. K. Moser, *Lectures on Celestial Mechanics*, Springer-Verlag, Berlin, 1971

A. Weinstein, *Lectures on symplectic manifolds*, 2nd ed., CBMS Regional Conference Series in Mathematics, vol. 29, American Mathematical Society, Providence, R.I., 1979

J. Xu, J. You, and Q. Qiu, *Invariant tori for nearly integrable Hamiltonian systems with degeneracy*, Math. Z. 226 (1997), no. **3**, 375–387

J.-C. Yoccoz, *Travaux de Herman sur les tores invariants*, Séminaire Bourbaki, 1991/92, Exp. 754, Astérisque 206 (1992), 311–344

J.-C. Yoccoz, *Théorème de Siegel, pôlynomes quadratiques et nombres de Brjuno*, Astérisque **2**31 (1995), 3-88