Introduction to KAM theory with a view to celestial mechanics

Jacques Féjoz

2015

Abstract. The theory of Kolmogorov, Arnold and Moser consists of a set of results regarding the persistence of quasiperiodic solutions, primarily in Hamiltonian systems. We bring forward a "twisted conjugacy" normal form, due to Herman, which contains all the (not so) hard analysis. We focus on the real analytic setting. A variety of KAM results follows, including most classical statements as well as more general ones. This strategy makes it simple to deal with various kinds of degeneracies and (Abelian and non-Abelian) symmetries. As an example of application, we prove the existence of quasiperiodic motions in the spatial 3-body problem.

KAM theory consists of results regarding the existence of quasiperiodic solutions, primarily in Hamiltonian systems. It was initiated by Kolmogorov in 1954, before being further developped by Arnold, Moser, and others.

The phase space of an integrable Hamiltonian system is foliated by Lagrangian invariant tori carrying a resonant or non resonant quasiperiodic dynamics. Kolmogorov's theorem asserts that, for any perturbation of the Hamiltonian, many non-resonant quasiperiodic Lagrangian invariant tori persist [57]. Kolmogorov's proof consists in looking for a strongly non-resonant invariant torus and solving the corresponding functional equation using Newton's algorithm in a (non Banach) functional space of infinite dimension. Poincaré wrote that the convergence of perturation series looked very unlikely. Others, among which Weierstrass or Bogoliubov-Krylov, failed to prove it. Proving the convergence of perturbation series is difficult due to the accumulation, in this context, of "small denominators". On the other hand, looking for invariant tori more geometrically as one would in the general theory of invariant manifolds, without prescribing a precise dynamics on the torus, fails severely because a resonant invariant torus does not persist under a generic perturbation (see [77] or [20, section 2.6 and chapter 11]). It was a stroke of genius of Kolmogorov both to imagine the correct statement and to realize that the Newton algorithm could be implemented in a family of Banach spaces and beat the effects of small denominators.

The invariant torus theorem has many applications in mathematical physics and mechanics. For example, Arnold's theorem tells us that the $1 + n$-body planetary problem, where one body (standing for the Sun) has a (hugely) larger mass than the others (standing for planets), has many quasiperiodic solutions [3, 38]. In the circular restricted three-body problem, the regular limit where one of the masses vanishes while the other two describe circular motions around their center of mass, those invariant tori separate energy levels and thus confine all neighboring solutions, thus resulting in a resounding stability property. In the same line of thought, Herman has shown that Boltzman's ergodic hypothesis fails for a generic Hamiltonian, because codimension-1 invariant tori prevent energy levels to be ergodic [100]. Yet, invariant tori theorems seldom apply directly, due to symmetries and degeneracies, and to the stength of non-integrability. Should KAM theory apply at all (see [16]), refined versions are usually needed.

Bibliographical comments. For background in Hamiltonian systems, excellent references are the books of V. Arnold [4], V. Arnold-V. Koslov-A. Neishtadt [5], V. Guillemin-S. Sternberg [47], A. Knauf [56], K. Meyer-G. Hall [68], C. Siegel-J. Moser [93] or S. Sternberg [96].

There exist many surveys of KAM theory, among which we recommend those of J.-B. Bost [12], L. Chierchia [22], J. Pöschel [79] or M. Sevryuk [89, 90, 92]. S. Dumas's book [29] is an interesting, historical account of the subject. Several results emphasized in the present paper, including the twisted conjugacy theorem 1, were already proved in [38] in the smooth setting.

Here are some examples of refinements or extensions, sometimes spectacular:

- quantitative versions [24]
- persistence of lower dimensional tori [31] even without controling the first order normal dynamics [14]
- persistence under degenerate torsion [33]
- global (non perturbative) versions for diffeomorphisms of the circle [51] or cocycles [7]
- various kinds of linearizations of cocycles [6] or of interval exchange maps [66]
- the dynamics of generic Lagrangian invariant tori [100]
- weak KAM theory [35]
- reversible systems [91]
- non-Hamiltonian perturbations [67]
- Hamiltonian partial differential equations [10, 58]

(see references therein).

Contents

1 Twisted conjugacy normal form

Let $\mathcal{H}$ be the set of germs along $T_0 = \mathbb{T}^n \times \{0\}$ of real analytic functions ("Hamiltonians") in $\mathbb{T}^n \times \mathbb{R}^n = \{(\theta, r)\}$. The Hamiltonian vector field associated with $H \in \mathcal{H}$ is

$$\begin{cases} \dot{\theta} = \partial_r H \\ \dot{r} = -\partial_\theta H. \end{cases}$$

For any given vector $\alpha \in \mathbb{R}^n$, let $\mathcal{K}(\alpha)$ be the affine space of Hamiltonians $K \in \mathcal{H}$ of the form

$$K = c + \alpha \cdot r + O(r^2),$$

for some (non fixed) $c \in \mathbb{R}$; $O(r^2)$ stands for the remainder (depending on θ) of the expansion in power of r. The space $\mathcal{K}(\alpha)$ consists exactly of Hamiltonians for which T_0 is invariant ($\dot{r}|_{r=0} = 0$) and carries a linear flow with velocity α ($\dot{\theta}|_{r=0} = \alpha$).[1]

Let $\mathcal{G}$ be the set of germs along T_0 of exact symplectic real analytic isomorphisms of the form[2]

$$G(\theta, r) = \left(\varphi(\theta), (r + S'(\theta)) \cdot \varphi'(\theta)^{-1}\right),$$

where φ is an isomorphism of $\mathbb{T}^n$ fixing the origin and S is a function on $\mathbb{T}^n$ vanishing at the origin. The goal being to find invariant tori close to T_0 and carrying a linear flow of frequency α, φ allows us to make changes of coordinates at will on the Lagrangian torus T_0, while S allows us to bring back to the zero section any graph over T_0, of 0-average and sufficiently close to the zero section.

In the next theorem, we assume that α is *Diophantine*:

$$|k \cdot \alpha| \geq \frac{\gamma}{|k|^\tau} \quad (\forall k \in \mathbb{Z}^n \setminus \{0\}) \tag{1}$$

for some fixed $\gamma, \tau > 0$; we have set $|k| = |k_1| + \cdots + |k_n|$. We will call $D_{\gamma,\tau}$ the set of such vectors. $D_{\gamma,\tau}$ is non empty if and only if $\tau \geq n-1$ (Dirichlet's theorem) and, if $\tau > n-1$ and $\gamma \to 0$, the complement of $D_{\gamma,\tau}$ within a ball has measure $O(\gamma)$, hence $\cup_\gamma D_{\gamma,\tau}$ has full measure [82].

1 Theorem (Herman)**.** *If $K^o \in \mathcal{K}(\alpha)$ and if $H \in \mathcal{H}$ is close enough to K^o, there is a unique $(K, G, \beta) \in \mathcal{K}(\alpha) \times \mathcal{G} \times \mathbb{R}^n$ such that*

$$H = K \circ G + \beta \cdot r. \tag{2}$$

We will prove theorem 1 in the next two sections.

The statement calls for some remarks.

---

[1] Recall that, in Dynamical Systems, a path $\gamma : \mathbb{R} \to X$ on a manifold is *quasiperiodic* (of some rank $\leq k$) if there exists $\omega \in \mathbb{R}^k$ and a map $\Gamma : \mathbb{T}^k \to X$ such that $\gamma(t) = \Gamma(t\omega)$. Provided γ is smooth enough, γ then admits a Fourier expansion of the form $\gamma(t) = \sum_{j \in \mathbb{Z}^k} \gamma_j e^{i2\pi\,(j\cdot\omega)\,t}$ (in some local coordinates in a tubular neighborhood of $\gamma(\mathbb{R})$).

[2] Conventionally, if f is a map from an open set U of a vector space E into another vector space F, we define $f'(x)$ as an element of $F \otimes E^*$ (as opposed to $E^* \otimes F$), and we write $f'(x) \cdot \xi$ for the contraction with a vector $\xi \in E$. Also, we identify $\mathbb{R}^n$ with its dual, so that $S'(\theta)$ may be imaged as the gradient of S.

- The frequency being a conjugacy invariant of quasi-periodic flows, the counter-term $\beta \cdot r$, which allows us to tune the frequency, is necessary. Yet it breaks the dynamical conjugacy between K and H and does not comply H with having an invariant torus, as K does. We call this normal form a *twisted conjugacy*. The geometrical contents of the theorem is that locally the set of Hamiltonians possessing an α-quasiperiodic torus is a submanifold of finite codimension if α is Diophantine (it has infinite codimension if α is not). The counter-term is the finite dimensional obstruction to conjugacy to a Hamiltonian of $\mathcal{K}(\alpha)$, and can be imaged as a simple control to preserve a torus of the same frequency and cohomology class as that of K^o.

- In general, one cannot expect H to be of the form

$$H = (K + \beta \cdot r) \circ G;$$

  this would show that having a Diophantine invariant torus is an open property, which is wrong, as the following example shows.

  Consider the Hamiltonian $H = \alpha \cdot r$, $\alpha \in \mathbb{R}^2$. All the tori $r = cst$ are invariant. By a first arbitrarily small perturbation, we may assume that α is resonant: $k \cdot \alpha = 0$ for some $k \in \mathbb{Z}^2 \setminus \{0\}$. Then add a resonant monomial:

$$H = \alpha \cdot r - \epsilon \sin(2\pi k \cdot \theta).$$

  The vector field is

$$\begin{cases} \dot\theta = \alpha \\ \dot r = 2\pi\epsilon \cos(2\pi k \cdot \theta)\, k. \end{cases}$$

  So, the solution through $(0, r)$ at time $t = 0$ is

$$t \mapsto (t\alpha, r + 2\pi\epsilon t k).$$

  So, if $\epsilon > 0$, this solution is unbounded and prevents any invariant torus (among graphs over T_0) to exist.

2 Exercise. Deduce Arnold's normal form for vector fields v on the torus $\mathbb{T}^n$ close to a Diophantine rotation [3, 12], from the twisted conjugacy theorem. *Hint:* Apply the twisted conjugacy theorem to the Hamiltonian $H(\theta, r) = r \cdot v$ on $\mathbb{T}^n \times \mathbb{R}^n$. Then check that the zero section is invariant by the corresponding isomorphism G, using an argument of Lagrangian intersection [38].

Bibliographical comments. – Computing the codimension of a group orbit is sometimes imprecisely called the "method of parameters". It is commonplace in singularity theory. In KAM theory, where the codimension often turns out finite, it has been fruitfully used in a number of works, among which: Arnold's normal form of vector fields on the torus (the paradigmatic, founding example) [3, 71, 100], Moser's normal form of vector fields [72] (which encompasses many natural subcases [15, 67, 98] but which has been much overlooked for 30 years), Chenciner's work on bifurcations of elliptic fixed points [17, 18, 19] or Eliasson-Fayad-Krikorian's study of the neighborhood of invariant tori [34]. The method of parameters allows us to first prove a normal form theorem which does not depend on any non-degeneracy assumption, but which contains all the hard analysis; the remaining, finite dimensional problem is then to show that the frequency offset vanishes, using a non-degeneracy hypothesis. This last step was probably not well understood before the late 80s [31, 61, 62, 85, 89]. The method fails for other kinds of dynamics than the quasiperiodic one on the torus because generically there are infinitely many new obstructions (the right hand side of the cohomological equation should have zero-average on periodic orbits) at each step of the Newton algorithm [44].

– The normal form of theorem 1, which was advertised by Herman in the 90s (M. Herman seemingly did not know Moser's normal form), in particular in his lectures on Arnold's theorem at the Dynamical System Seminar in Université Paris VII, can be seen as a particular case of Moser's normal form, when the vector field is Hamiltonian, then giving more precise information [39, 67]. A proof in the smooth category can be found in [38]. The lesser rigidity there allows us not to introduce deformed norms.

2 One step of the Newton algorithm

Let

$$\phi(x) = K \circ G + \beta \cdot r, \quad x = (K, G, \beta).$$

We want to solve the following equation between Hamiltonians:

$$\phi(x) = H, \tag{3}$$

for H close to $\phi(K^o, id, 0) = K^o$. The twisted conjugacy theorem thus reduces to prove that ϕ is invertible, keeping in mind that

– if ϕ is formally defined on the whole space $\mathcal{K}(\alpha) \times \mathcal{G} \times \mathbb{R}^n$, it is only if G is close enough to the identity, with respect to the width of analyticity of K, that $\phi(K, G, \beta)$ is analytic on a neighborhood of T_0,

– equation (3) is really of interest to us only if it holds on a neighborhood of $G^{-1}(T_0)$, a domain depending on the unknown G.

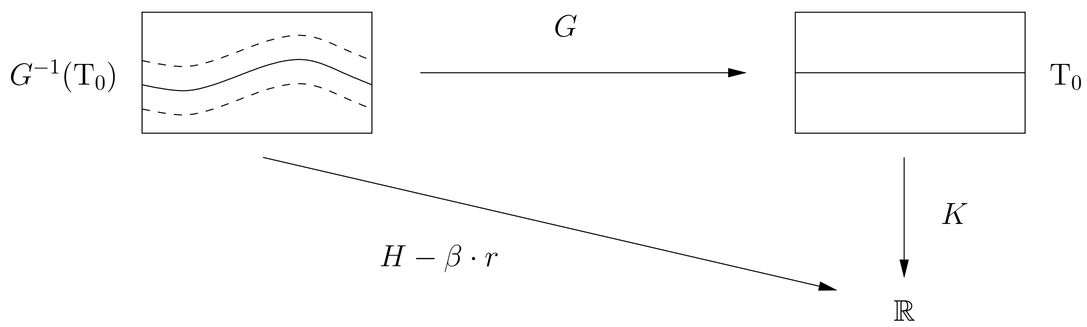


Figure 1:

Note that $\mathcal{H}$ and $\mathbb{R}^n$ are trivially vector spaces, while

- $\mathcal{K}(\alpha)$ is an affine space, directed by the vector space $\mathbb{R} + O(r^2)$
- and $\mathcal{G}$, while being a groupoid with semi-direct product law given by $G_2 \circ G_1 = (\varphi_2 \circ \varphi_1, (r + S_1' + S_2' \cdot \varphi_1') \cdot (\varphi_2 \circ \varphi_1)'^{-1})$, will rather be identified, locally in a neighborhood of the identity, to an open set of the affine space passing through the identity and directed by the linear space $\{(\varphi - \mathrm{id}, S)\}$, where $v = \varphi - \mathrm{id} : \mathbb{T}^n \to \mathbb{R}^n$ and $S : \mathbb{T}^n \to \mathbb{R}$ are analytic and vanish at the origin.

We will invert ϕ using the Newton algorithm, which consists in iterating the operator

$$f : x = (K, G, \beta) \mapsto \hat{x} = x + \phi'(x)^{-1}(H - \phi(x)). \tag{4}$$

Each step of the induction requires to invert the linearized operator $\phi'(x)$, not only at $x_0 = (K^o, \mathrm{id}, 0)$, but at some unknown x in the neighborhood of x_0, i.e. to solve the linearized equation

$$\phi'(K, G, \beta) \cdot (\delta K, \delta G, \delta\beta) = \delta K \circ G + K' \circ G \cdot \delta G + \delta\beta \cdot r = \delta H \tag{5}$$

where δH is the data, (K, G, β) is a parameter, and the unknowns are the "tangent vectors" $\delta K \in \mathbb{R} \oplus O(r^2)$, δG (geometrically, a vector field along G) and $\delta\beta \in \mathbb{R}^n$. Pre-composing with G^{-1} modifies the equation into an equation between germs along the standard torus T_0 (as opposed to the G-dependant torus $G^{-1}(\mathrm{T}_0)$):

$$\delta K + K' \cdot \dot{G} + \delta\beta \cdot r \circ G^{-1} = \dot{H}, \tag{6}$$

where we have set $\dot{G} = \delta G \circ G^{-1}$ (geometrically, a germ along T_0 of tangent vector field) and $\dot{H} = \delta H \circ G^{-1}$. It is a key point in measuring norms that we are interested in the neighborhood of T_0 on one side of the conjugacy, and in the neighborhood of $G^{-1}(\mathrm{T}_0)$ on the other side.

Using the additional notations (in which $*_{\geq k}$ stands for a function in $O(r^k)$, depending on θ):

$$\begin{cases} K = c + \alpha \cdot r + Q(\theta) \cdot r^2 + K_{\geq 3} \\ \delta K = \delta c + \delta K_{\geq 2} \\ \dot{G} = (\dot{\varphi}, -r \cdot \dot{\varphi}' + \dot{S}') \\ \dot{H} = \dot{H}_0 + \dot{H}_1 \cdot r + \dot{H}_{\geq 2} \end{cases}$$

and the fact that

$$\delta\beta \cdot r \circ G^{-1} = \left(r \cdot \varphi' \circ \varphi^{-1} - S' \circ \varphi^{-1}\right) \cdot \delta\beta,$$

and identifying the Taylor coefficients in equation (6) yield the following three equations:

$$\delta c + \dot{S}' \cdot \alpha - S' \circ \varphi^{-1} \cdot \delta\beta = \dot{H}_0 \tag{7}$$

$$\varphi' \circ \varphi^{-1} \cdot \delta\beta - \dot{\varphi}' \cdot \alpha + 2\,\dot{S}' \cdot Q = \dot{H}_1 \tag{8}$$

$$\delta K_{\geq 2} + \partial_\theta K \cdot \dot{\varphi} - r \cdot \dot{\varphi}' \cdot \partial_r K_{\geq 2} = \dot{H}_{\geq 2}. \tag{9}$$

The first equation aims at infinitesimally straightening the would-be invariant torus of $\dot{H}_0$, the second equation at straightening its dynamics, and the third at equating higher order terms. Due to the symplectic constraint, the first two equations are coupled (the Hamiltonian lift of the vector field $\dot{\varphi}$ having a non trivial component in the r-direction), whereas in the context of general vector fields the system is triangular. We will now show the existence of a unique solution to this system of equations, and derive estimates of the solution, within some appropriate functional setting.

Let

$$\mathbb{T}^n_s = \mathbb{T}^n \times i[-s,s]^n$$

be the complex extension of $\mathbb{T}^n$ of "width" s, and

$$|f|_s = \max_{\theta \in \mathbb{T}^n_s} |f(\theta)|$$

for functions f which are real holomorphic on the interior of $\mathbb{T}^n_s$ and continuous on $\mathbb{T}^n_s$; such functions form a space $\mathcal{A}(\mathbb{T}^n_s)$ which is Banach (there are other possible choices here, e.g. one could consider the space of functions which are real holomorphic on $\mathbb{T}^n_s$). We extend this definition to vector-valued functions by taking the maximum of the norms of the components (and, consistently, the ℓ^1-norm for "dual" integer vectors, e.g. $k \in \mathbb{Z}^n$). Similarly, let $\mathbb{R}^n_s$ be a complex neighborhood of the origin in $\mathbb{R}^n$ of "width" s:

$$\mathbb{R}^n_s = \{z \in \mathbb{C}^n,\ |z| \le s\}, \quad |z| = \max(|z_1|, ..., |z_n|).$$

We will call $\mathcal{A}(\mathbb{T}^n_s \times \mathbb{R}^n_s)$ the Banach space of functions which are continuous on $\mathbb{T}^n_s \times \mathbb{R}^n_s$ and real holomorphic on the interior.

Let

- $\mathcal{H}_s = \mathcal{H} \cap \mathcal{A}(\mathbb{T}^n_s \times \mathbb{R}^n_s)$ (endowed with the supremum norm $|\cdot|_s$)
- $\mathcal{K}_s(\alpha) = \mathcal{K}(\alpha) \cap \mathcal{A}(\mathbb{T}^n_s \times \mathbb{R}^n_s) \subset \mathcal{H}_s$.
- $\mathcal{G}^\sigma_s$ be the subset of $\mathcal{G}$ consisting of isomorphisms $G \simeq (\varphi, S)$ such that $\varphi - \mathrm{id} \in \mathcal{A}(\mathbb{T}^n_s, \mathbb{R}^n_1)$ and $S \in \mathcal{A}(\mathbb{T}^n_s, \mathbb{R}_1)$ and G is σ-polynomially-close to the identity, i.e.
$$|G - \mathrm{id}\,|_s \le C_{\mathcal{G}} \sigma^{k_{\mathcal{G}}} \tag{10}$$
for some fixed $C_{\mathcal{G}} > 0$ and $k_{\mathcal{G}} > 0$ to be later determined.

3 Lemma (Linearized equation)**.** *If x is close enough to x^0, equation* (6) *possesses a unique solution $\dot{x} = (\delta K, \dot{G}, \delta\beta)$. Moreover, there exist $C', \tau' > 0$ such that, for all s, σ,*

$$|\dot{x}|_s \le \frac{C'}{\sigma^{\tau'}} |\dot{H}|_{s+\sigma},$$

where C' depends only on n, τ, provided K, G^{-1} and β are bounded on $\mathbb{T}^n_{s+\sigma} \times \mathbb{R}^n_{s+\sigma}$.

Proof. First assume that $\delta\beta \in \mathbb{R}^n$ is given with $|\delta\beta \cdot r \circ G^{-1}| \leq Cst\,|\dot{H}|_{s+\sigma}$, and replace equation (8) by

$$\delta\hat{\beta} + \varphi' \circ \varphi^{-1} \cdot \delta\beta - \dot{\varphi}' \cdot \alpha + 2\,\dot{S}' \cdot Q = \dot{H}_1, \tag{11}$$

where $\delta\hat{\beta} \in \mathbb{R}^n$ is an additional unknown; as elsewhere in this proof, Cst stands for a constant, to which we do not want to give a consistant name, and which depends only on n, τ and $|(K - \alpha \cdot r, G^{-1} - \mathrm{id}, \beta)|_{s+\sigma}$.

- Averaging equation (7) yields $\delta c = \int_{\mathbb{T}^n} \left(\dot{H}_0 + S' \circ \varphi^{-1} \cdot \delta\beta\right) d\theta$, hence

  $$|\delta c| \leq Cst\,|\dot{H}|_{s+\sigma}$$

- According to lemma 45, equation (7) has a unique solution $\delta\tilde{S}$ having zero average, with

  $$|\delta\tilde{S}|_s \leq \frac{Cst}{\gamma\sigma^{\tau_c}}|\dot{H}_0|_{s+\sigma} \leq \frac{Cst}{\gamma\sigma^{\tau_c}}|\dot{H}|_{s+\sigma}.$$

  Then the unique solution vanishing at the origin, $\delta S = \delta\tilde{S} - \delta\tilde{S}(0)$, satisfies the same estimate (up to an unessential factor 2 which we absorb in the constant).

Note that the estimates hold for all s, σ (at the expense of possibly having an infinite right hand side). We now proceed similarly with equation (11):

- The average yields $\delta\hat{\beta} = \int_{\mathbb{T}^n} \left(\dot{H}_1 - 2\dot{S}' \cdot Q - \varphi' \circ \varphi^{-1} \cdot \delta\beta\right)$, hence, using Cauchy's inequality,[3]

  $$|\delta\hat{\beta}| \leq \frac{Cst}{\gamma\sigma^{\tau_c+1}}|\dot{H}|_{s+\sigma}.$$

- The average-free part determines $\delta\varphi$, with

  $$|\dot{\varphi}|_s \leq \frac{Cst}{\gamma\sigma^{\tau_c}}\left(|\dot{H}_1|_{s+\sigma} + 2|\dot{S}' \cdot Q^o|_{s+\sigma}\right) \quad (\forall s, \sigma > 0).$$

  Using Cauchy's inequality and the fact that Q^o is given, we see that

  $$|\dot{\varphi}|_s \leq \frac{Cst}{\gamma^2\sigma^{\tau_c(\tau_c+1)}}|\dot{H}|_{s+\sigma},$$

  where as before the constant depends only on n, τ, and $|Q^o|_{s+\sigma}$.

Equation (9) can then be solved explicitly:

$$\delta K_{\geq 2} = -\partial_\theta K^o \cdot \dot{\varphi} + 2r \cdot \dot{\varphi}' \cdot Q^o \cdot r + \delta H_{\geq 2},$$

[3] We use the ℓ^∞-norm on $\mathbb{R}^n$ (consistantly with the ℓ^1-norm already used on the dual space $\mathbb{Z}^n$).

and whence

$$|\delta K|_s \leq \frac{Cst}{\gamma^2 \sigma^{\tau_c(\tau_c+1)+1}} |\delta H|_{s+\sigma}.$$

We have built a map $\delta\beta \mapsto \delta\hat{\beta}$ in the neighborhood of $\delta\beta = 0$. It is affine and, when φ is close to the identity, invertible. Thus there exists a unique $\delta\beta$ such that $\delta\hat{\beta} = 0$, which satisfies

$$\|\delta\beta\| \leq \frac{Cst}{\gamma \sigma^{\tau_c+1}} |\dot{H}|_{s+\sigma}.$$

The claim follows, with $\tau' = \tau_c(\tau_c+1)+1$ and $C' = Cst/\gamma^2$ for some constant Cst independent of γ.[4] □

The lemma may be rephrased: the linear operator $\phi'(x)$ has a unique local inverse $\phi'(x)^{-1}$, with the given estimates.

Let

$$\delta x = \phi'(x)^{-1}(y - \phi(x)) \quad \text{and} \quad \hat{x} = \phi(x) = x + \delta x.$$

Taylor's formula says that

$$\begin{aligned}\phi(\hat{x}) &= \phi(x) + \phi'(x) \cdot \delta x + Q(x,\hat{x}), \quad Q(x,\hat{x}) = \int_0^1 (1-t)\phi''(x_t) \cdot (\delta x)^2 \, dt \\ &= y + Q(x,\hat{x}),\end{aligned}$$

where we have set $x_t = x + t\,\delta x$ $(0 \leq t \leq 1)$, hence

$$y - \phi(\hat{x}) = -Q(x,\hat{x}).$$

4 Lemma (Remainder)**.** *If* $|\dot{G}|_{s+\sigma} \leq \sigma/2$,

$$|Q(x,\hat{x}) \circ G^{-1}|_s \leq \frac{C''}{\sigma^{\tau''}} |\dot{x}|^2_{s+\sigma}.$$

Proof. Let $\delta^2\phi = \phi''(K,G,\beta) \cdot (\delta K, \delta G, \delta\beta)^2$. We have

$$\delta^2\phi = 2\delta K' \circ G \cdot \delta G + K'' \circ G \cdot (\delta G)^2,$$

hence

$$\delta^2\phi \circ G^{-1} = 2\delta K' \cdot \dot{G} + K'' \cdot \dot{G}^2,$$

so

$$|\delta^2\phi \circ G^{-1}|_s \leq \frac{Cst}{\sigma} |(\delta K, \dot{G})|^2_{s+\sigma}; \tag{12}$$

note here that $\delta^2\phi$ is computed in (K,G,β), and it is then pre-composed by G^{-1}.

Now, if $x_t = (K_t, G_t, \beta_t)$,

$$|Q(x,\hat{x}) \circ G^{-1}|_s \leq \int_0^1 \left| \left(\phi''(x_t) \cdot \delta x^2\right) \circ G^{-1} \right|_s dt.$$

[4] A better, but less transparent, choice of norms lets $C' = Cst/\gamma$ instead.

Since $|(\mathrm{id} + \dot{G})^{-1} - (\mathrm{id} - \dot{G})|_s \le Cst |\dot{G}|_s^2$,

$$|Q(x,\hat{x}) \circ G^{-1}|_s \le \int_0^1 \left| \left(\phi''(x_t) \cdot \delta x^2\right) \circ G_t^{-1} \right|_{s+2|\dot{G}|^2_{s+\sigma}} \, dt,$$

whence the wanted estimate, using (12). □

It remains to show that the iterated images

$$x_0 = (K^o, \mathrm{id}, 0), \quad x_{n+1} = f(x_n)$$

of the Newton map (4) are defined for $n \in \mathbb{N}$ and converge to some $(K, G, \beta) \in \mathcal{K}(\alpha) \times \mathcal{G} \times \mathbb{R}^n$ such that $H = K \circ G + \beta \cdot r$ in the neighborhood of $G^{-1}(\mathrm{T}_0)$, provided H is close enough to K^o. Namely, we will assume that $K^o \in \mathcal{K}_{s+\sigma}(\alpha)$, $H \in \mathcal{H}_{s+\sigma}$ for some fixed s, σ with $0 < s < s + \sigma \le 1$, and

$$|H - K^o|_{s+\sigma} \le \epsilon$$

for some $\epsilon > 0$. This is the goal of the next section.

3 Inverse function theorem

We first give an abstraction of our problem, and will afterwards show how it allows us to complete the proof of the twisted conjugacy theorem.

Let $E = (E_s)_{0<s<1}$ be a decreasing family of Banach spaces with increasing norms $|\cdot|_s$, and $\epsilon B_s^E = \{x \in E_s, |x|_s < \epsilon\}$, $\epsilon > 0$, be its balls centered at 0. Let (F_s) be an analogous family, and $\phi : \sigma B^E_{s+\sigma} \to F_s$, $s < s + \sigma$, $\phi(0) = 0$, be maps of class C^2, commuting with inclusions.

On account of composition operators, we will assume there are additional, deformed norms $|\cdot|_{x,s}$, $x \in \mathrm{Int}(sB_s^E)$, $0 < s < 1$, satisfying

$$|y|_{0,s} = |y|_s \quad \text{and} \quad |y|_{x',s} \le |y|_{x, s+|x'-x|_s},$$

and we will phrase our hypotheses on ϕ in terms of these norms.

Define

$$Q : \sigma B^E_{s+\sigma} \times \sigma B^E_{s+\sigma} \to F_s, \quad (x, \hat{x}) \mapsto \phi(\hat{x}) - \phi(x) - \phi'(x)(\hat{x} - x).$$

Assume that, if $x \in sB^E_{s+\sigma}$, the derivative $\phi'(x) : E_{s+\sigma} \to F_s$ has a right inverse $\phi'(x)^{-1} : F_{s+\sigma} \to E_s$, and

$$\begin{cases} |\phi'(x)^{-1}\eta|_s \le C'\sigma^{-\tau'}|\eta|_{x,s+\sigma} \\ |Q(x,\hat{x})|_{x,s} \le C''\sigma^{-\tau''}|\hat{x} - x|^2_{s+\sigma+|\hat{x}-x|_s} \quad (\forall s, \sigma, x, \hat{x}, \eta) \end{cases} \tag{13}$$

with $C', C'', \tau', \tau'' \ge 1$. Let $C := C'C''$ and $\tau := \tau' + \tau''$.

The important fact in the Newton algorithm below, is that the index loss σ can be chosen arbitrarily small, without s itself being small, provided the deformed norm substitutes for the initial norm of the spaces F_s. The initial norm $|\cdot|_s$ of F_s is here only for the practical purpose of having a fixed target space, to which perturbations belong.

5 Theorem. *ϕ is locally surjective and, more precisely, for any s, η and σ with $\eta < s$,*

$$\epsilon B^F_{s+\sigma} \subset \phi\left(\eta B^E_s\right), \quad \epsilon := 2^{-8\tau} C^{-2} \sigma^{2\tau} \eta.$$

In other words, ϕ has a right-inverse $\psi : \epsilon B^F_{s+\sigma} \to \eta B^E_s$.

Some numbers s, η and σ and $y \in B^F_{s+\eta}$ being given, let

$$f : \sigma B^E_{s+\eta+\sigma} \to E_s, \quad x \mapsto x + \phi'(x)^{-1}(y - \phi(x)).$$

Proof of the theorem. Now, let s, η and σ be fixed, with $\eta < s$ and $y \in \epsilon B^F_{s+\sigma}$ for some ϵ. We will see that if ϵ is small enough, the sequence $x_0 = 0$, $x_n := f^n(0)$ is defined for all $n \geq 0$ and converges towards some preimage $x \in \eta B^E_s$ of y by ϕ.

Let $(\sigma_n)_{n\geq 0}$ be a sequence of positive real numbers such that $3\sum \sigma_n = \sigma$, and $(s_n)_{n\geq 0}$ be the sequence decreasing from $s_0 := s + \sigma$ to s defined by induction by the formula $s_{n+1} = s_n - 3\sigma_n$.

Assuming the existence of $x_0, ..., x_{n+1}$, we see that $\phi(x_k) = y + Q(x_{k-1}, x_k)$, hence

$$x_{k+1} - x_k = \phi'(x_k)^{-1}(y - \phi(x_k)) = -\phi'(x_k)^{-1} Q(x_{k-1}, x_k) \qquad (1 \leq k \leq n).$$

Further assuming that $|x_{k+1} - x_k|_{s_k} \leq \sigma_k$, the estimate of the right inverse and lemma 39 entail that

$$|x_{n+1} - x_n|_{s_{n+1}} \leq c_n |x_n - x_{n-1}|^2_{s_n} \leq \cdots \leq c_n c^2_{n-1} \cdots c_1^{2^{n-1}} |x_1|^{2^{n-1}}_{s_1}, \quad c_k := C\sigma_k^{-\tau}.$$

The estimate

$$|x_1|_{s_1} \leq C'(3\sigma_0)^{-\tau'} |y|_{s_0} \leq C\sigma_0^{-\tau}\epsilon = c_0\epsilon$$

and the fact, to be checked later, that $c_k \geq 1$ for all $k \geq 0$, show :

$$|x_{n+1} - x_n|_{s_{n+1}} \leq \left(\epsilon \prod_{k\geq 0} c_k^{2^{-k}}\right)^{2^n}.$$

Since $\sum_{n\geq 0} \rho^{2^n} \leq 2\rho$ if $2\rho \leq 1$, and using the definition of constants c_k's, we get a sufficient condition to have all x_n's defined and to have $\sum |x_{n+1} - x_n|_s \leq \eta$:

$$\epsilon = \frac{\eta}{2} \prod_{k\geq 0} c_k^{-2^{-k}} = \frac{2\eta}{C^2} \prod_{k\geq 0} \sigma_k^{\tau 2^{-k}}. \tag{14}$$

Maximizing the upper bound of ϵ under the constraint $3\sum_{n\geq 0}\sigma_n = \sigma$ yields $\sigma_k := \frac{\sigma}{6} 2^{-k}$. A posteriori it is straightforward that $|x_{n+1} - x_n|_{s_n} \leq \sigma_n$ (as earlier assumed to apply lemma 39) and $c_n \geq 1$ for all $n \geq 0$. Besides, using that $\sum k2^{-k} = \sum 2^{-k} = 2$ we get

$$\epsilon = \frac{\eta}{2} \prod_{k \geq 0} c_k^{-2^{-k}} = \frac{\eta}{2} \prod_{k \geq 0} \frac{1}{2^{\tau k 2^{-k}}} \left(\frac{1}{C} \left(\frac{\sigma}{6} \right)^{\tau} \right)^{2^{-k}} = \frac{\eta}{C^2} \left(\frac{\sigma}{12} \right)^{2\tau} > \frac{\sigma^{2\tau} \eta}{2^{8\tau} C^2},$$

whence the theorem. □

Remark. The two competing small parameters η and σ being fixed, our choice of the sequence (σ_n) maximizes ϵ for the Newton algorithm. It does not modify the sequence (x_k) but only the information we retain from (x_k).

6 Exercise (End of proof of theorem 1). Complete the proof by checking that

– A similar statement as theorem 5 holds if ϕ is defined only on a ball of polynomial radius with respect to the width of analyticity (see (10)).

– $|K_n|_{s_n}$, $|G_n^{-1}|_{s_n}$ and β_n are bounded along the induction (in order to justify the repeated use of estimates of lemmata 3 and 4, which are not uniform as assumed in (13)). *Hint:* Use the fact that

$$G_{n+1}^{-1} = G_n^{-1} \circ (\mathrm{id} + \dot{G}_n)^{-1},$$

the estimate of $\dot{G}$ in the induction and the estimate of proposition 46 in appendix B.

7 Corollary. *The size of the allowed perturbation is polynomial in the Diophantine constant γ (see* (1)*).*

8 Exercise. What is the domain of ψ in F_S? *Hint:* Optimize the function $\epsilon(\eta, \sigma)$ under the constraint $s + \sigma = S$.

Bibliographical comments. – The seeming detour through Herman's normal form reduces Kolmogorov's theorem to a functionally well posed inversion problem, as opposed to Zehnder's (remarkable) work [101, 102]. One may compare the present stragegy and Zehnder's in the following way. Inverting the operator

$$\phi : (K, G, \beta) \mapsto H = K \circ G + \beta \cdot r$$

(see equation (3)) is equivalent to solving the implicit function

$$F(K, G, \beta; H) = K - (H - \beta \cdot r) \circ G^{-1} = 0.$$

But ϕ happens to be a local diffeomorphism, while $\partial F / \partial (K, G, \beta)$ is invertible in no neighborhood of $(K^o, id, 0)$. This is why Zehnder had to deal with approximate inverses. The drawback of focusing on the equation $\phi(K, G, \beta) = H$ is that we neeed it to be satisfied on a domain which depends on G.

As Zehnder, we have encapsulated the Newton algorithm in an abstract inverse function theorem, à la Nash-Moser. The algorithm indeed converges without very specific hypotheses on the internal structure of the variables (see exercise 6, though). At the expense of some optimality, ignoring this structure allows for simple estimates and control of the bounds, and for solving a whole class of analogous problems with the same toolbox (lower dimensional tori, codimension-one tori, Siegel problem, as well as some problems in singularity theory).

– The fast convergence of the Newton algorithm makes it possible to beat the effect of small denominators and other sources of loss of width of analyticity. Its has proved unreasonably efficient compared to other lines of proof in KAM theory such as direct proofs of convergence of perturbation series [32] or proofs via renormalization [28]. Another approach relies on the method of periodic approximation and on simultaneous Diophantine approximations [13]. Still another alternative to Newtons's algorithm consists, at each step of the induction, in solving a (non-linear) finite dimentional approximation of the functional equation (3) using Ekeland's variational principle [30].

– The arithmetic condition is not optimal. Indeed, solving the exact cohomological equation at each step is inefficient because the small denominators appearing with intermediate-order harmonics deteriorate the estimates, whereas some of these harmonics could have a smaller amplitude than the error terms and thus would better not be taken care of. Even stronger, Rüssmann and Pöschel have noticed that at each step it is worth neglecting part of the low-order harmonics themselves (to some carefully chosen extent). Then the expense, a worse error term, turns out to be cheaper than that the gain –namely, the right hand side of the cohomological equation now has a smaller size over a larger complex extension. This makes it possible, with a slowly converging sequence of approximations, to show the persistence of invariant tori under some arithmetic condition which, in one dimension, is equivalent to Brjuno's condition [80]. Bounemoura-Fischler have found an interesting alternative proof based on periodic approximations [13].

– The analytic (or Gevrey) category is simpler than Hölder or Sobolev categories, in Nash-Moser theory, because the Newton algorithm can be carried out without intercalating smoothing operators (cf. [69, 88, 49, 12]). On the other hand, the analytic category is more complicated because of the absence of cut off functions, which forces us to pay attention to the domain of definition of the Hamiltonian more carefully (cf. [38]).

– The method of Jacobowitz [53] (see [70] also) in order to deduce an inverse function theorem in the smooth category from its analogue in the analytic category does not work directly, here. The idea would be to use Jackson's theorem in approximation theory to caracterize the Hölder spaces by their approximation properties in terms of analytic functions and, then, to find a smooth preimage x by ϕ of a smooth function y as the limit of analytic preimages x_j of analytic approximations y_j of y. However, in our inverse function theorem we require the operator ϕ to be defined only on balls $\sigma B_{s+\sigma}$ with shrinking radii when $s+\sigma$ tends to 0. This domain is too small in general to include the analytic approximations y_j of a smooth y. Such a restriction is inherent in the presence of composition operators. The problem of isometric embeddings is simpler, from this viewpoint.

4 Local uniqueness and regularity of the normal form

In the proof of theorem 5 we have built right inverses $\psi : \epsilon B^F_{s+\eta+\sigma} \to \eta B^E_{s+\eta}$, of ϕ, commuting with inclusions. The proof shows that ψ is continuous at 0; due to the invariance of the hypotheses of the theorem by small translations, ψ is locally continuous.

We further make the following two assumptions:

— The maps $\phi'(x)^{-1} : F_{s+\sigma} \to E_s$ are left (as well as right) inverses (in theorem 1 we have restricted to an adequate class of symplectomorphisms);

— The scale $(|\cdot|_s)$ of norms of (E_s) satisfies some interpolation inequality:

$$|x|^2_{s+\sigma} \leq |x|_s \, |x|_{s+\tilde{\sigma}} \quad \text{for all } s,\ \sigma,\ \tilde{\sigma} = \sigma\left(1 + \frac{1}{s}\right)$$

(according to the sentence after the statement of corollary 48 in appendix C, this estimate is satisfied in the case of interest to us, since $\sigma + \log(1+\sigma/s) \leq \tilde{\sigma}$).

9 Lemma (Lipschitz regularity)**.** *If $\sigma < s$ and $y, \hat{y} \in \epsilon B^F_{s+\sigma}$ with $\epsilon = 2^{-14\tau}C^{-3}\sigma^{3\tau}$,*

$$|\psi(\hat{y}) - \psi(y)|_s \leq C_L |\hat{y} - y|_{\psi(y),s+\sigma}, \quad C_L = 2C'\sigma^{-\tau'}.$$

In particular, ψ is the unique local right inverse of ϕ, i.e. it is also the local left inverse of ϕ.

Proof. Fix $\eta < \zeta < \sigma < s$; the impatient reader can readily look at the end of the proof how to choose the auxiliary parameters η and ζ more precisely.

Let $\epsilon = 2^{-8\tau}C^{-2}\zeta^{2\tau}\eta$, and $y, \hat{y} \in \epsilon B^F_{s+\sigma}$. According to theorem 5, $x := \psi(y)$ and $\hat{x} := \psi(\hat{y})$ are in $\eta B^E_{s+\sigma-\zeta}$, provided the condition, to be checked later, that $\eta < s + \sigma - \zeta$. In particular, we will use a priori that

$$|\hat{x} - x|_{s+\sigma-\zeta} \leq |\hat{x}|_{s+\sigma-\zeta} + |x|_{s+\sigma-\zeta} \leq 2\eta.$$

We have

$$\begin{aligned} \hat{x} - x &= \phi'(x)^{-1}\phi'(x)(\hat{x} - x) \\ &= \phi'(x)^{-1}\left(\hat{y} - y - Q(x, \hat{x})\right) \end{aligned}$$

and, according to the assumed estimate on $\phi'(x)^{-1}$ and to lemma 39,

$$|\hat{x} - x|_s \leq C'\sigma^{-\tau'}|\hat{y} - y|_{x,s+\sigma} + 2^{-1}C\zeta^{-\tau}|\hat{x} - x|^2_{s+2\eta+|\hat{x}-x|_s}.$$

In the norm index of the last term, we will coarsely bound $|\hat{x} - x|_s$ by 2η. Additionally using the interpolation inequality:

$$|\hat{x} - x|^2_{s+4\eta} \leq |\hat{x} - x|_s |\hat{x} - x|_{s+\tilde{\sigma}}, \quad \tilde{\sigma} = 4\eta\left(1 + \frac{1}{s}\right),$$

yields

$$\left(1 - 2^{-1}C\zeta^{-\tau}|\hat{x}-x|_{s+\tilde{\sigma}}\right)|\hat{x}-x|_s \leq C'\sigma^{-\tau'}|\hat{y}-y|_{x,s+\sigma}.$$

Now, we want to choose η small enough so that

— first, $\tilde{\sigma} \leq \sigma - \zeta$, which implies $|\hat{x}-x|_{s+\tilde{\sigma}} \leq 2\eta$. By definition of $\tilde{\sigma}$, it suffices to have $\eta \leq \frac{\sigma-\zeta}{4(1+1/s)}$.

— second, $2^{-1}C\zeta^{-\tau}\,2\eta \leq 1/2$, or $\eta \leq \frac{\zeta^\tau}{2C}$, which implies that $2^{-1}C\zeta^{-\tau}|\hat{x}-x|_{s+\tilde{\sigma}} \leq 1/2$, and hence $|\hat{x}-x|_s \leq 2C'\sigma^{-\tau'}|\hat{y}-y|_{x,s+\sigma}$.

A choice is $\zeta = \frac{\sigma}{2}$ and $\eta = \frac{\sigma^\tau}{16C} < s$, whence the value of ϵ in the statement. □

10 Proposition (Smoothness). *For every $\sigma < s$, there exists ϵ, C_1 such that for every $y, \hat{y} \in \epsilon B^F_{s+\sigma}$,*

$$|\psi(\hat{y}) - \psi(y) - \phi'(\psi(y))^{-1}(\hat{y}-y)|_s \leq C_1|\hat{y}-y|^2_{s+\sigma}.$$

Moreover, the map $\psi' : \epsilon B^F_{s+\sigma} \to L(F_{s+\sigma}, E_s)$ defined locally by $\psi'(y) = \phi'(\psi(y))^{-1}$ is continuous and, if $\phi : \sigma B^E_{s+\sigma} \to F$ is C^k, $2 \leq k \leq \infty$, for all σ, so is $\psi : \epsilon B^F_{s+\sigma} \to E_s$.

Proof. Fix ϵ as in the previous proof and $y, \hat{y} \in \varepsilon B^F_{s+\sigma}$. Let $x = \psi(y)$, $\eta = \hat{y} - y$, $\xi = \psi(y+\eta) - \psi(y)$ (thus $\eta = \phi(x+\xi) - \phi(x)$), and $\Delta := \psi(y+\eta) - \psi(y) - \phi'(x)^{-1}\eta$. Definitions yield

$$\Delta \;=\; \phi'(x)^{-1}\left(\phi'(x)\xi - \eta\right) = -\phi'(x)^{-1}Q(x, x+\xi).$$

Using the estimates on $\phi'(x)^{-1}$ and Q and the latter lemma,

$$|\Delta|_s \leq C_1|\eta|^2_{s+\sigma'}$$

for some σ' tending to 0 when σ itself tends to 0, and for some $C_1 > 0$ depending on σ. Up to the substitution of σ by σ', the estimate is proved.

The inversion of linear operators between Banach spaces being analytic, $y \mapsto \phi'(\psi(y))^{-1}$ has the same degree of smoothness as ϕ'. □

11 Corollary. *If $\pi \in L(E_s, V)$ is a family of linear maps, commuting with inclusions, into a fixed Banach space V, then $\pi \circ \psi$ is C^1 and $(\pi \circ \psi)' = \pi \cdot (\phi' \circ \psi)^{-1}$.*

This corollary is used with $\pi : (K, G, \beta) \mapsto \beta$ in the proof of theorem 1.

It will later be convenient to extend ϕ^{-1} to non-Diophantine vectors α. Whitney-smoothness is a criterion for such an extension to exist [95, 99].

Suppose $\phi(x) = \phi_\alpha(x)$ now depends on some parameter $\alpha \in B^\kappa$ (the unit ball of $\mathbb{R}^\kappa$),

— that the estimates assumed up to now are uniform with respect to α over some closed subset $\mathrm{D} \subset \mathbb{R}^\kappa$,

— and that ϕ is C^1 with respect to α.

We will denote ψ_α the parametrized version of the inverse of ϕ_α.

12 Proposition (Whitney-smoothness). *If s, σ and ϵ are chosen like in proposition 10, the map $\psi : \mathrm{D} \times \epsilon B^F_{s+\sigma} \to E_s$ is C^1-Whitney-smooth and extends to a map $\psi : \mathbb{R}^n \times \epsilon B^F_{s+\sigma}$ of class C^1. If ϕ is C^k, $1 \le k \le \infty$, with respect to α, this extension is C^k.*

Proof. Let $y \in \epsilon B^F_{s+\sigma}$. If $\alpha, \alpha + \beta \in \mathrm{D}$, $x_\alpha = \psi_\alpha(y)$ and $x_{\alpha+\beta} = \psi_{\alpha+\beta}(y)$, we have

$$\phi_{\alpha+\beta}(x_{\alpha+\beta}) - \phi_{\alpha+\beta}(x_\alpha) = \phi_\alpha(x_\alpha) - \phi_{\alpha+\beta}(x_\alpha).$$

Since $\hat{y} \mapsto \psi_{\alpha+\beta}(\hat{y})$ is Lipschitz (lemma 9),

$$|x_{\alpha+\beta} - x_\alpha|_s \le C_L |\phi_\alpha(x_\alpha) - \phi_{\alpha+\beta}(x_\alpha)|_{s+\sigma},$$

and, since $\hat{\alpha} \mapsto \phi_{\hat{\alpha}}(x_\alpha)$ itself is Lipschitz, so is $\alpha \mapsto x_\alpha$.

Moreover, the formal derivative of $\alpha \mapsto x_\alpha$ is

$$\partial_\alpha x_\alpha = -\phi'_\alpha(x_\alpha) \cdot \partial_\alpha \phi(x_\alpha).$$

Expanding $y = \phi_{\alpha+\beta}(x_{\alpha+\beta})$ at $\beta = 0$ and using the same estimates as above, shows that

$$|x_{\alpha+\beta} - x_\alpha - \partial_\alpha x_\alpha \cdot \beta|_s = O(\beta^2)$$

when $\beta \to 0$, locally uniformly with respect to α. Hence $\alpha \mapsto x_\alpha$ is C^1-Whitney-smooth, and, similarly, C^k-Whitney-smooth if $\alpha \mapsto \phi_\alpha$ is.

Thus, by the Whitney extension theorem, the claimed extension exists. Note that Whitney's original theorem needs two straightforward generalizations to be applied here: ψ takes values in a Banach space, instead of $\mathbb{R}$ or a finite dimension vector space (see [45]); and ψ is defined on a Banach space, but the extension directions are of finite dimension. □

13 Exercise (Quasiperiodic time dependant perturbations). Let $\nu \in \mathbb{R}^m$ be fixed. Consider the subspace $\mathcal{H}_\nu$ of $\mathcal{H}$ (in dimension $2(n+m)$) consisting of Hamiltonians in

$$\mathbb{T}^{n+m} \times \mathbb{R}^{n+m} = \mathbb{T}^n_\theta \times \mathbb{T}^m_\psi \times \mathbb{R}^n_r \times \mathbb{R}^m_\Psi$$

of the form

$$\hat{H} = \nu \cdot \Psi + H,$$

where H does not depend on Ψ. Since the corresponding Hamiltonian vector field has the component

$$\dot{\psi} = \nu,$$

$\mathcal{H}_\nu$ may be imaged as the space of Hamiltonians on $\mathbb{T}^n \times \mathbb{R}^n$ with quasiperiodic time dependance. Show that, if $\hat{H} \in \mathcal{H}_\nu$ and $\hat{H} = \phi(K, G, (\beta, \beta'))$ (with $\beta \in \mathbb{R}^n$ and $\beta' \in \mathbb{R}^m$), then

$$\begin{cases} K \in \mathcal{H}_\nu \\ G \text{ leaves } \psi \text{ unchanged} \\ \beta' = 0. \end{cases}$$

Further question: develop the KAM theory below in this particular case.

14 Exercise (Control & persistence of tori). If H is close to an integrable Hamiltonian $K^o = K^o(r)$, show that there is a smooth integrable Hamiltonian $\beta = \beta(r)$ such that for every R such that T_0 is a (γ,τ)-Diophantine invariant torus of K^o, $H - \beta(r)\cdot r$ has an invariant torus carrying a quasiperiodic dynamics with the same frequency.

Hint. Apply the twisted conjugacy theorem to each $H(R+\cdot,\cdot)$, with $R \in \mathbb{R}^n$ close to 0 such that the torus $r = R$ is Diophantine for K^o and, using Proposition 12, extend the so-obtained function $R \mapsto \beta(R)$ as a smooth function.

Bibliographical comments. – It is possible give a proof, patterned on [88, p. 626], that ψ is C^1 without the assumption that $\phi'(x)$ has unique inverse (or right inverse). Yet the proof simplifies and the estimates improve under the combined two additional assumptions. In particular, the existence of a right inverse of $\phi'(x)$ makes the inverse ψ unique and thus allows us to ignore the way it was built (a posteriori regularity result).

– Latzutkin understood, in the case of the standard map, the fundamental importortance of Whitney-smoothness of the invariant circles with respect the rotation number. This is a key point in the method of parameter. The dependance actually is of Gevrey class [78], but we do not need it here.

5 Conditional conjugacy

We now move to a *conditional conjugacy*, the common ground of invariant tori theorems of later sections.

Let

$$\mathcal{K}_s = \cup_{\alpha\in\mathbb{R}^n}\mathcal{K}_s(\alpha) = \left\{c + \alpha\cdot r + O(r^2),\ c \in \mathbb{R}, \alpha \in \mathbb{R}^n\right\}$$

be the set of Hamiltonians on $\mathbb{T}^n_s \times \mathbb{R}^n_s$ for which T_0 is invariant and quasi-periodic, with unprescribed frequency.

15 Theorem (Conditional conjugacy)**.** *For every $K^o \in \mathcal{K}_{s+\sigma}(\alpha^o)$ with $\alpha^o \in D_{\gamma,\tau}$, there is a germ of smooth map*[5]

$$\Theta : \mathcal{H}_{s+\sigma} \to \mathcal{K}_s \times \mathcal{G}_s, \quad H \mapsto (K_H, G_H), \quad K_H = c_H + \alpha_H\cdot r + O(r^2),$$

at $K^o \mapsto (K^o, \mathrm{id})$ such that the following implication holds:

$$(\forall H)\ \alpha_H\ \textit{Diophantine} \Longrightarrow H = K_H \circ G_H$$

and (K_H, G_H) is unique in $\mathcal{K}\times\mathcal{G}$.

Proof. Denote ϕ_α the operator we have been denoting ϕ –because the frequency α was fixed while we now want to vary it. Define the map

$$\begin{array}{rlcl} \hat\Theta : & \mathrm{D}_{\gamma,\tau}\times\mathcal{H}_{s+\sigma} & \to & \mathcal{K}_s\times\mathcal{G}_s\times\mathbb{R}^n \\ & (\alpha, H) & \mapsto & \hat\Theta_\alpha(H) := (\phi_\alpha)^{-1}(H) = (K, G, \beta) \end{array}$$

[5]Thank you to Jean-Christophe Yoccoz for drawing my attention to a mistake in a prior version of this statement.

locally in the neighborhood of (α^o, K^o). Since ϕ is infinitely differentiable, by proposition 12 there exist a C^∞-extension

$$\hat{\Theta} : \mathbb{R}^n \times \mathcal{H}_{s+\sigma} \to \mathcal{K}_s \times \mathcal{G}_s \times \mathbb{R}^n.$$

Write $K^o = \alpha^o \cdot r + \hat{K}$, $\hat{K} = c + O(r^2)$. In particular, since

$$\phi_\alpha(K^o + (\alpha - \alpha^o) \cdot r, \mathrm{id}, \alpha^o - \alpha) \equiv K^o$$

locally for all $\alpha \in \mathbb{R}^n$ close to α^o we have

$$\hat{\Theta}(\alpha, K^o) = (K^o, \mathrm{id}, \beta), \quad \beta(\alpha, K^o) = \alpha^o - \alpha.$$

In particular,

$$\frac{\partial \beta}{\partial \alpha} = -\,\mathrm{id}$$

and, by the implicit function theorem, locally for all H there exists a unique $\hat{\alpha}$ such that $\beta(\hat{\alpha}, H) = 0$. We conclude by letting $\Theta(H) = \hat{\Theta}(\hat{\alpha}, H)$. □

The so-defined vector α_H, which is called a *frequency vector* of H, is unique when belonging to $D_{\gamma,\tau}$. It depends Gevrey-smoothly on H (i.e. their partial derivatives of order r behave like positive powers of $r!$), as discovered by Popov [78], but not analytically (except for a family of integrable Hamiltonians). (For our purpose, the Lipschitz regularity would suffice, in conjunction with the Lipschitz inverse function theorem. For the sake of simplicity, we stick to the C^1 class.)

6 Invariant torus with prescribed frequency

The first invariant torus theorem will be a trivial corollary of the conditional conjugacy theorem. Consider a smooth family $(K_t^o)_{t \in \mathbb{B}^\kappa}$ of Hamiltonians in some $\mathcal{K}_s$. Each K_t^o is of the form $K_t^o = c_t^o + \alpha_t^o \cdot r + O(r^2)$. The frequency map of the family is

$$\alpha^o : \mathbb{B}^\kappa \mapsto \mathbb{R}^n, \quad t \mapsto \alpha_t^o.$$

In this section, we will describe the simplest case, where (the derivative of) α^o has rank n, which, by the submersion theorem, implies that α^o is onto, stably with respect to C^1-perturbations. In celestial mechanics, the parameter t may be masses, semi major axes, eccentricities, inclinations, energy, angular momentum, etc.

Now, let (H_t) be a smooth family of Hamiltonians in $\mathcal{H}_s$ such that, for each t, H_t is close enough to K_t (a condition that we will not repeat in each statement).

16 Theorem. *If the frequency map α^o is a local submersion (that is, of rank n) and if $\alpha^o(0) \in D_{\gamma,\tau}$, there exists $t \in \mathbb{B}^n$ such that H_t has an invariant torus with frequency $\alpha^o(0)$. Moreover, the subset formed by the values of $t \in \mathbb{B}^n$ for which H_t has an invariant torus has positive Lebegue measure.*

Proof. According to the conditional conjugacy theorem, the family (H_t) has some frequency map α which is C^∞-close to α^o. So α itself is a local C^∞-submersion and attains $\alpha^o(0)$. Besides, as soon as $\alpha_t \in D_{\gamma,\tau}$, H_t has an invariant torus, which happens for a subset $\mathcal{B} \subset \mathbb{B}^n$ of positive Lebesgue measure. □

17 Remark. The first part of the conlusion holds under the topological hypothesis that α^o has non-zero degree (which ensures that α^o is locally onto, stably with respect to perturbations), a remark which applies if α^o has a ramification point, for example.

Poincaré [77] introduced the following two transversality conditions (he was considering the particular case, considered next, where t is the action variable, $K^o = K^o(r)$ and $\alpha = \partial_r K^o(r)$).

18 Definition. The Hamiltonian family (K_t^o) is

- *isochronically non-degenerate* if the frequency map has rank n
- *isoenergetically non-degenerate* if the map

$$\mathbb{B}^\kappa \to \mathbb{R} \times \mathbb{P}(\mathbb{R}^n), \quad t \mapsto ([\alpha_t^o], c_t^o)$$

  (where $[\alpha_t^o]$ stands for the homogeneous class of α_t^o) has rank n.

(Neither condition implies the other.)

19 Exercise (Variants of theorem 16). Prove the following two variants.

– Isoenergetic theorem: If the family (K_t) is isoenergetically non-degenerate, and if the frequency vector α^o belongs to $D_{\gamma,\tau}$, there exists $t \in \mathbb{B}^\kappa$ such that H_t has an invariant torus of energy c_0^o and frequency class $[\alpha_0^o]$. Moreover, the subset formed by the values of $t \in \mathbb{B}^\kappa$ for which H_t has an invariant torus of energy c_0^o has positive $(n-1)$-dimensional Lebesgue measure.[6]

– "Iso first integral" theorem: More generally, assume that for all t, f_t is an $\mathbb{R}^\lambda$-valued first integral of K_t^o and H_t (e.g., with $\lambda = 2$, f_t may stand for the energy and the angular momentum of a mechanical system in the plane) and that the frequency vector α^o belongs to $D_{\gamma,\tau}$ The function f_t must be constant on T_0 and we call $f_t(\mathrm{T}_0)$ this constant. If the map

$$\mathbb{B}^\kappa \to \mathbb{R} \times \mathbb{P}(\mathbb{R}^n), \quad t \mapsto (f_t(\mathrm{T}_0), [\alpha_t^o])$$

has maximal rank n, there exists $t \in \mathbb{B}^\kappa$ such that H_t has an invariant torus on which $f_t = f_0(\mathrm{T}_0)$ and with frequency class $[\alpha_0^o]$. More strongly, if the map

$$\mathbb{B}^\kappa \to \mathbb{R} \times \mathbb{R}^n, \quad t \mapsto (f_t(\mathrm{T}_0), \alpha_t^o)$$

has maximal rank n, for every $t^o \in \mathbb{B}^\kappa$ close to 0 there exists $t \in \mathbb{B}^\kappa$ such that H_t has an invariant torus with $f_t = f_0(\mathrm{T}_0)$ and frequency vector α_0^o.

(For a first integral associated with a non-Abelian symmetry, see section 8).

[6] There is no intrinsic $(n-1)$-dimensional Lebesgue measure, but the fact of having positive $(n-1)$-dimensional Lebesgue measure *is* intrinsic.

We now turn to Kolmogorov's theorem, which corresponds to the particular case where the family (H_t) is obtained by mere translation of some initial Hamiltonian $H \in \mathcal{H}$, in the direction of actions: $H_t(\theta, r) = H(\theta, t + r)$, $t \in \mathbb{B}^n$. Call c^o and Q^o the constant and quadratic parts of some $K^o \in \mathcal{K}(\alpha^o)$:

$$K^o = c^o + \alpha^o \cdot r + Q^o(\theta) \cdot r^2 + O(r^3).$$

20 Theorem (Kolmogorov)**.** *If the frequency vector α^o belongs to $D_{\gamma,\tau}$ and if the quadratic form $\int_{\mathbb{T}^n} Q^o(\theta)\, d\theta$ is non-degenerate, there exists a unique $R \in \mathbb{R}^n$ such that $G^{-1}(\mathrm{T}_0) + (0, R)$ is an α^o-quasiperiodic invariant torus of H. Moreover, the invariant tori of H form a set of positive Lebesgue measure in the phase space.*

Proof. Let F be the analytic function taking values among symmetric bilinear forms, which solves the cohomological equation $L_{\alpha^o} F = Q^o - \int_{\mathbb{T}^n} Q^o\, d\theta$ (use lemma (45)), and ψ be the germ along T_0 of the (well defined) time-one map of the flow of the Hamiltonian $F(\theta) \cdot r^2$. The map ψ is symplectic and restricts to the identity on T_0. At the expense of substituting $K^o \circ \psi$ and $H \circ \psi$ for K^o and H respectively, one can thus assume that

$$K^o = c^o + \alpha^o \cdot r + Q^1 \cdot r^2 + O(r^3), \quad Q^1 := \int_{\mathbb{T}^n} Q^o(\theta)\, d\theta.$$

The germs so obtained from the initial K^o and H are close to one another.

Consider the family of trivial perturbations obtained by translating K^o in the direction of actions:

$$K^o_R(\theta, r) := K^o(\theta, R + r), \quad R \in \mathbb{R}^n,\ R \text{ small},$$

and its approximation obtained by truncating the first order jet of K^o_R along T_0 from its terms $O(R^2)$:

$$\hat{K}^o_R(\theta, r) := (c^o + \alpha^o \cdot R) + \big(\alpha^o + 2Q^1 \cdot R\big) \cdot r + O(r^2) = K^o_R + O(R^2).$$

For the Hamiltonian $\hat{K}^o_R$, T_0 is invariant and quasiperiodic of frequency $\alpha^o + 2Q^1 \cdot R$. The first assertion then follows from theorem 16.

What has been done for the torus of frequency α^o can more generally be done for all tori of Diophantine frequency. What remains to be proved is that the collection of perturbed invariant tori has positive measure. Using the map Θ of the conditional conjugacy theorem, now define

$$(K_R, G_R) = \Theta(H_R),$$

with

$$\begin{cases} K_R = c_R + \alpha_R \cdot r + O(r^2) \\ G_R(\theta, r) = (\varphi_R(\theta), (r + S'_R(\theta)) \cdot \varphi'_R(\theta)^{-1}) \end{cases}$$

locally in the neighborhood of $R = 0$, say for $\|R\| < R_0$. Let

$$\mathcal{R} = \{R \in \mathbb{R}^n,\ \|R\| < R_0,\ \alpha_R \in D_{\gamma,\tau}\}.$$

As soon as $R \in \mathcal{R}$, $G_R^{-1}(\mathrm{T}_0)$ is invariant for H_R, hence

$$\mathcal{T}_R = G_R^{-1}(\mathrm{T}_0) + (0, R) = \{(\theta, R - S'_R(\theta)),\ \theta \in \mathbb{T}^n\}$$

is invariant for H. Because of proposition 12, $\mathcal{T}_R$ depends Whitney-smoothly on $R \in \mathcal{R}$. Thus the diffeomorphisms which straighten all the $\mathcal{T}_R$'s individually may be glued together, by Whitney's extension theorem and the last assertion follows. □

21 Remark (Measure of the set of tori). Due to the estimate of the inverse function theorem 5, if $\gamma \ll 1$, the allowed size of $|H - K^o|_s$ (for some $s > 0$) is polynomial in γ (of degree 4). One can actually show that it is $|H - K^o|_s = O(\gamma^2)$ [79]. In other words, for a given H, a torus with frequency vector in $D_{\gamma,\tau}$ is preserved for some $\gamma = O(\sqrt{\epsilon})$, and, as a classical estimate of the measure of the complement of Diophantine vectors shows, the measure of the complement of the invariant tori is of order $O(\gamma) = O(\sqrt{\epsilon})$.

Once one has one invariant torus, it is straightforward to obtain a set of positive measure of invariant tori, as the proof above has shown. (This was not so at the level of generality of theorem 16. Why? If $t_1, t_2 \in \mathcal{B}$, the invariant tori of H_{t_1} and H_{t_2} may meet. In Kolmogorov's theorem, the parameter being the cohomology class of the tori, this cannot happen.) We will see in the next section that a much weaker transversality condition is sufficient for locally finding a positive measure of tori. Yet, in the absence of any transversality hypothesis, the question of the accumulation of a quasiperiodic invariant torus by quasiperiodic invariant tori, and their measure, is the subject of Herman's conjecture [34].

22 Exercise. Instead of applying theorem 16, complete the proof of theorem 20 using the twisted conjugacy theorem.

Hint. The twisted conjugacy normal form of $\hat{K}^o_R$ with respect to the frequency α is

$$\hat{K}^o_R = \left(\hat{K}^o_R - \hat{\beta}^o_R \cdot r\right) \circ \mathrm{id} + \hat{\beta}^o_R \cdot r, \quad \hat{\beta}^o_R := 2Q^1 \cdot R.$$

By assumption the matrix $\frac{\partial \hat{\beta}^o}{\partial R}\Big|_{R=0} = 2Q^1$ is invertible and the map $R \mapsto \hat{\beta}^o_R$ is a local diffeomorphism. Now, there is an analogous map $R \to \beta_R$ for H_R, which is a small C^∞-perturbation of $R \mapsto \hat{\beta}^o_R$, and thus a local diffeomorphism, with a domain having a lower bound locally uniform with respect to H. Hence if H is close enough to K^o there is a unique small R such that $\beta = 0$. For this R the equality $H_R = K \circ G$ holds, hence the torus obtained by translating $G^{-1}(\mathrm{T}_0)$ by R in the direction of actions is invariant and α-quasiperiodic for H.

Bibliographical comments. – Claims that Kolmogorov's proof was incomplete are unfounded in view of the breakthrough: the supposedly missing arguments in Kolmogorov's paper bear upon to Cauchy's inequality and elementary harmonic analysis [57, 23, 40]. Kolmogorov actually gave these details in Moscow's seminar, as Arnold and Sinaï have testified. Arnold later gave an alternative proof. Arnold's statement is equivalent to Kolmogorov's, despite the superficial difference of looking to all neighboring tori at a time. Arnold additionally payed attention to how far H can be from K^o, as the torsion gets close to degenerate [3].

– The remark that parameters are not necessarily action variables adds some flexibility for finding invariant tori, e.g. in the work of Zhao L. [104, 105]. Another example is an analogue of Arnold's theorem where one would be allowed to tune not only the semi major axes but also the masses of the planets.

7 Invariant tori with unprescribed frequencies

There is a KAM theory which assumes only a much weaker non-degeneracy condition than above. Let $\alpha : \mathbb{B}^\kappa \to \mathbb{R}^n$ be a smooth map.

23 Definition. The frequency map α is *skew*[7] if its image is nowhere locally contained in a vector hyperplane.

24 Lemma (Rüssmann [83, 84])**.** *If α is skew and analytic, for all $t \in B^\kappa$ there exist $r \in \mathbb{N}_*$ and $j_1, ..., j_r \in \mathbb{N}^\kappa$ such that*

$$\mathrm{Vect}\left(\partial^{j_1}\alpha_t, ..., \partial^{j_r}\alpha_t\right) = \mathbb{R}^n; \tag{15}$$

the integer $\max_i |j_i|$ *(where $|j_i|$ is the length of j_i) is called the* index of degeneracy *at t. Conversely, if there exists $t \in B^\kappa$, $r \in \mathbb{N}_*$ and $j_1, ..., j_r \in \mathbb{N}^\kappa$ such that* (15) *holds, α is skew.*

The property of being skew is a very weak transversality condition. It is of crucial interest that κ may be smaller than n.

25 Example. The monomial curve

$$\alpha : t \in [0, 1] \mapsto (1, t, ..., t^{n-1}) \in \mathbb{R}^n$$

is skew. Indeed, with the convention that $1/n! = 0$ if $n \in \mathbb{Z}^-$, the matrix

$$\left(\alpha, \alpha', \cdots, \alpha^{(n-1)}\right) = \left(\frac{(j-1)!}{(j-i)!}t^{j-i}\right)_{1\le i,j\le n}.$$

has rank n.

See [87] for a comparison with a dozen conditions which have been used in KAM theory. Here we content ourselves with the following examples, showing in particular that being skew is implied by the traditional conditions of isochronic or isoenergetic non-degeneracy.

26 Example. – If α is isochronically non-degenerate, at every point $t \in B^\kappa$ its local image is an open set of $\mathbb{R}^n$, so α is skew, with index of degeneracy equal to 1.

– Suppose that t is the action variable, $H = H(r)$ and $\alpha = \partial_r H(r)$. If then H is isoenergetically non-degenerate, its frequency map is skew. Indeed, since the

[7] The terminology we have chosen here is not standard. Related (but not always equivalent) conditions have been called *essentially non planar*, *non planar*, *Rüssmann-non-degenerate*, *weakly non-degenerate*, *curved* etc.

determinant of the "bordered torsion"[8] $\begin{pmatrix} \alpha' & \alpha \\ {}^t\alpha & 0 \end{pmatrix}$ is non zero, α' must have rank $n-1$, the bordered torsion is equivalent to

$$\begin{pmatrix} \bar{\tau} & 0 & \bar{\alpha} \\ 0 & 0 & \beta \\ {}^t\alpha & \beta & 0 \end{pmatrix}$$

with $\det \bar{\tau} \neq 0$ and $\beta \in \mathbb{R}$, hence the borderd torsion has determinant $-\beta^2 \det \bar{\tau}$, hence $\beta \neq 0$, hence H is skew with index of degeneracy equal to 1.

27 Example (L. Chierchia). The integrable Hamiltonian defined over $\mathbb{T}^4 \times \mathbb{R}^4$ by

$$H = \frac{1}{4} r_1^4 + \frac{1}{2} r_1^2 r_2 + r_1 r_3 + r_4$$

is isochronically and isoenergetically degenerate, but its frequency, as a function of the action r_1, is skew at $(r_1, 0, 0, 0)$, $r_1 \neq 0$.

We now take up hypotheses of the beginning of section 6, i.e. we consider a smooth family $(K_t^o)_{t \in \mathbb{B}^\kappa}$ of Hamiltonians in $\mathcal{K}$. Each K_t^o is of the form $K_t^o = c_t^o + \alpha_t^o \cdot r + O(r^2)$. The (analytic) frequency map of the family is

$$\alpha^o : \mathbb{B}^\kappa \mapsto \mathbb{R}^n, \quad t \mapsto \alpha_t^o.$$

Let (H_t) be a smooth family of Hamiltonians in $\mathcal{H}_s$ such that, for each t, H_t is close enough to K_t. The conditional conjugacy theorem yields a smooth frequency map $t \mapsto \alpha_t$ of H which is C^∞-close to $t \mapsto \alpha_t^o$.

28 Proposition (Rüssmann [87]). *If α^o is skew, there exists $\mu \in \mathbb{N}_*$ (an affine function of the index of degeneracy of α^o) such that if α is C^μ-close to α^o,*

$$\text{Leb}\,\{t \in \mathbb{B}^\kappa,\ \alpha_t \notin D_{\gamma,\tau}\} \leq C\gamma^{1/\mu}.$$

From the proof of the proposition, it is not hard to see how these estimate deteriorate when there are several time scales (a situation otherwise called *properly degenerate*).

29 Corollary. *Under the hypotheses of proposition 28, if we split α into $\alpha = (\hat{\alpha}, \check{\alpha}) \in \mathbb{R}^{\hat{n}} \times \mathbb{R}^{\check{n}}$, $\hat{n} + \check{n} = n$, then*

$$\text{Leb}\,\{t \in \mathbb{B}^\kappa,\ (\hat{\alpha}_t, \epsilon\check{\alpha}_t) \notin D_{\gamma,\tau}\} \leq C \left(\frac{\gamma}{\epsilon}\right)^{1/\mu}$$

for some affine function μ of the index of degeneracy.

An immediate consequence is the following theorem.

30 Theorem. *If the frequency map α^o is skew, there is a subset $\mathcal{T} \subset \mathbb{B}^\kappa$ of positive Lebesgue measure such that, for all $t \in \mathcal{T}$, H_t has a Diophantine quasiperiodic invariant torus.*

[8] Poincaré calls this square matrix the "bordered Hessian" of H [77].

31 Remark (Size of the allowed perturbation). In many applications indeed, that there are several time scales. For example, in the planetary 3-body problem the dynamics splits into the fast Keplerian dynamics and the slow secular dynamics. If one wants to apply KAM theory, it is then crucial to know the size of the allowed perturbation in terms of these time scales. The relevant estimates may be established along the following lines.

Consider for example the case of a frequency curve $\alpha = (\hat{\alpha}, \check{\alpha}) : I \mapsto \mathbb{R}^n = \mathbb{R}^{\hat{n}} \times \mathbb{R}^{\check{n}}$, $t \mapsto (\hat{\alpha}(t), \check{\alpha}(t))$, assumed skew at some $t_0 \in I$. Then, after corollary 29, if we want to have some measure estimates which are uniform with respect to small ϵ, we need to choose $\gamma = O(\epsilon^N)$ for some N large enough. Last, due to the estimate of the inverse function theorem 5, if $\gamma \ll 1$, the allowed size of $|H - K^o|_s$ (for some $s > 0$) is polynomial in γ, hence in ϵ. (One can show that $|H - K^o|_s = O(\gamma^2)$ is enough for the conclusion to hold [79].) Hence, it usually suffices to apply theorem 30 to a normal form of high order, whose remainder is in $O(\epsilon^{2N})$.

The analogue of Kolmogorov's theorem for the weak transversality condition of being skew is the following. Consider one Hamiltonian $K \in \mathcal{K}(\alpha^o)$ for some $\alpha^o \in DH_{\gamma,\tau}$ (with γ small enough and τ large enough) and one Hamiltonian $H \in \mathcal{H}$ close to $\mathcal{K}$. Upon putting K^o under normal form at some high enough order, theorem 15 gives the existence of a frequency map $r \mapsto \alpha^o_r$ of K.

32 Theorem (Rüssmann)**.** *If the frequency map α^o is skew, the invariant tori of H form a set of positive Lebesgue measure in the phase space.*

The proof mimicks the second part of the proof of Kolmogorov's theorem.

Bibliographical comments. The theory of Diophantine approximations on manifolds was initiated by the works of Arnold and his students; see [9, 55, 81, 94]. It has later been used in dynamical systems, e.g. in [5, 21, 8, 31, 74, 75, 85, 86, 87].

8 Symmetries

This section consists in a remark regarding Hamiltonian systems invariant under a Hamiltonian group action. The natural way to find invariant tori is to apply KAM theory to the symplectically reduced system. Here, we explain how to take advantage of the symmetries "upstairs", avoiding to carry out explicit computations on the quotient.

Let (X, ω) be an exact symplectic real analytic manifold of dimension $2n$ and G a compact group, acting analytically on X in a Hamiltonian way, freely and properly. Call $2m$ the corank of G. Let us briefly recall the argument why the corank is even. Consider the adjoint action of a maximal torus $\mathcal{T}$ on the Lie algebra $\mathfrak{g}$ of G. Since $\mathcal{T}$ is Abelian, $\mathfrak{g}$ splits into irreducible components of real dimensions 1 or 2, on which $\mathcal{T}$ respectively acts trivially or by rotations. The component with trivial representation V contains the Lie algebra $\mathfrak{t}$ of $\mathcal{T}$. But, if $\mathcal{T}$ is indeed maximal, one must have $V = \mathfrak{t}$ [1].

Let T_0 be a Lagrangian embedded real analytic torus of X and $K^o_t : X \mapsto \mathbb{R}$, $t \in \mathbb{B}^\kappa$, be a smooth family of G-invariant real analytic functions (Hamiltonians) for which T_0 is invariant, quasiperiodic of frequency vector $\alpha^o_t \in \mathbb{R}^n$.

The main example is a rotation-invariant mechanical system. The condition of being skew is always violated, because one frequency (corresponding in the phase space to the two directions of non trivial rotations of the angular momentum vector) vanishes identically. One can get rid of this degeneracy by fixing the direction of the angular momentum (see [103]). The remaining invariance by rotations around the direction of the angular momentum adds some flexibility for checking the transversality condition, since the harmonics which are not invariant have zero Fourier coefficient. What follows is an abstraction of this situation.

33 Lemma. *The image of the frequency map $t \mapsto \alpha^o_t$ lies in a subspace of $\mathbb{R}^n$ of codimension m.*

Proof. Let $\mathcal{T}$ be a maximal torus of G; its codimension is $2m$. Let μ be the moment map, thought of as a map $X \to \mathfrak{g}$, and $\mathfrak{t}_+$ be the positive Weyl chamber of $\mathcal{T}$ (see [48]). As Guillemin-Sternberg have noticed (see [46] for details) $X_+ = \mu^{-1}(\mathfrak{t}_+)$ is

- a codimension-$2m$, real analytic submanifold of X (μ is transverse to $\mathfrak{t}_+$, a connected component of the regular values of μ)
- symplectic (TX_+ intersects its symplectic orthogonal along the space of vectors generated by $\mathfrak{t}_+$ only, hence along the zero section only)
- a section of the G-action (because maximal tori are conjugate to each other and, under our asumptions, the moment map is equivariant).

The velocity vector on T_0 is tangent to X_+, so $\mathrm{T}_0 \cap X_+$ is an invariant torus, whose ergodic components are isotropic (see appendix A), hence of dimension at most $n - m$. By invariance of X_+, the frequencies in the directions which are symplectically orthogonal to X_+ vanish. There are m of them, whence the claim. □

Let $\mathcal{T}$ be a maximal torus as in the proof above, of Lie algebra $\mathfrak{t} = \mathbb{R}^k$ (k thus being the rank of G). Let $\tau : X \to \mathbb{R}^k$ be its moment map (a projection of the full moment map μ). By a classical theorem of Weinstein, we may identify a neighborhood of the Lagrangian torus $\mathcal{T}$ with a neighborhood of the zero section T_0 in the cotangent bundle of $\mathcal{T}$. Let (θ, r) be coordinates. Let us expand τ near T_0:

$$\tau = \tau_0(\theta) + \tau_1(\theta) \cdot r + O(r^2),$$

where $\tau_0(\theta) \in \mathbb{R}^k$ and $\tau_1(\theta) \in M_{k,n}(\mathbb{R})$. The torus T_0 is invariant and quasiperiodic for each component of τ (because those components commute with K^o_0). So τ_0 and τ_1 actually do not depend on θ.

Consider the amended Hamiltonian

$$\hat{K}^o_{t,u} = K^o_t + u \cdot \tau, \tag{16}$$

depending on parameters $t \in \mathbb{B}^\kappa$ and $u \in \mathbb{R}^k$. By Lagrangian intersection theory, it has the same ergodic Lagrangian invariant tori as K_t^o, and the frequency vector of T_0 is changed into

$$\hat{\alpha}_{t,u}^o = \alpha_t^o + u \cdot \tau_1,$$

Call $\mathrm{Vect}\,\tau_1$ the subspace of $\mathbb{R}^n$ spanned by the k row-vectors of τ_1. This is the subspace of frequencies which may be attained by tuning the parameter u.

Rather then repeating the whole theory in the G-invariant setting, we merely adapt four chief statements, according to the following array of hypotheses, where the *partially reduced system* refers to the restriction of the Hamiltonian system to the invariant symplectic manifold X_+, of dimension $2(n-m)$.

	Submersive frequency	Skew frequency
Partially reduced system	1	2
Fully reduced system	3	4

34 Theorem (Partially reduced viewpoint)**.** *1. If the frequency map $\alpha^o : \mathbb{R}^\kappa \to \mathbb{R}^n$ has rank $\geq n-m$ at 0 and $\alpha_0^o \in D_{\gamma,\tau}$, there exists t such that H_t has an invariant torus of frequency α_0^o. Besides, the subset formed by the values of $t \in \mathbb{B}^n$ for which H_t has an invariant torus has positive Lebegue measure.*

2. If the image of the frequency map does not lie in any plane of codimension $> m$ in $\mathbb{R}^n$, for a subset of $t \in \mathbb{B}^\kappa$ of positive Lebesgue measure, H_t has a rank-$(n-m)$ quasiperiodic invariant torus.

35 Theorem (Totally reduced viewpoint)**.** *3. If the amended frequency map $\hat{\alpha}^o : \mathbb{R}^\kappa \times \mathbb{R}^k \to \mathbb{R}^n$, has rank $\geq n-m$ at $(0,0)$, there exists t such that H_t has an invariant torus of frequency $\alpha_0^o \pmod{\mathrm{Vect}\,\tau_1}$. Besides, the subset formed by the values of $t \in \mathbb{B}^n$ for which H_t has an invariant torus has positive Lebegue measure.*

4. If the image of the amended frequency map does not lie in any plane of codimension $> m$ in $\mathbb{R}^n$, for a subset of $t \in \mathbb{B}^\kappa$ of positive Lebesgue measure, H_t has a rank-$(n-m)$ (posibly non minimal) quasiperiodic invariant torus.

The hypotheses in the two statements amount to assuming that the induced frequency map $\hat{\alpha}^o : \mathbb{R}^\kappa \times \mathbb{R}^k \to \mathbb{R}^n/\mathrm{Vect}\,\tau_1$ (of the totally reduced system) respectively has rank $\geq n-m$ or is skew.

If the parameter is the translation in the direction of the action variable r, one could further infer the existence of a subset of the phase space and of positive Lebesgue measure, consisting of invariant tori, as in section 6, using an argument which we will not repeat here.

Items 1 and 2 yield minimal tori. Items 3 and 4 yield strictly more tori, foliated into minimal invariant subtori of codimenion from 0 to k. Determining this codimension requires to compute the frequencies of the lift of the $\mathcal{T}$-action, which boils down to a quadrature, along the lines of the standard theory of symplectic reduction.

Proof. First restrict to the symplectic manifold X_+, which has dimension $2(n-m)$ (partial reduction). Items 1 and 2 of the statement follow from theorems 16 and 30 respectively. Now, restrict to a regular level of μ and quotient by $\mathcal{T}$. The reduced

Hamiltonian system of K_t^o has frequency the equivalence class of α_t^o modulo Vect τ_1. So, resonance hyperplanes in the partially reduced phase space which are broken by $u \cdot \tau_1$ project to zero in the reduced system. Assertions 3 and 4 thus follow from theorems 16 and 30, this time applied to the fully reduced system. □

All the four items of this theorem will be used in our study of the three-body problem.

Bibliographical comments. The idea of amending the Hamiltonian goes back to Poincaré when he would look to the three-body problem in a rotating frame of reference in order to break some degeneracies in his search for periodic orbits [77]. The role of partial reduction (consisting in fixing only the direction of the angular momentum) was brought forward in [65].

9 Lower dimensional tori

In this section, we sketch the theory for lower dimensional invariant tori. Some additional details may be found in [38].

Two integers $n \geq 1$ and $m \geq 0$ being fixed, let $\mathcal{H}$ be the set of germs along $\mathrm{T}_0 = \mathbb{T}^n \times \{0\} \times \{0\}$ of real analytic functions (Hamiltonians) in the phase phase

$$\mathbb{T}^n \times \mathbb{R}^n \times \mathbb{C}^m = \{(\theta, r, z = (x, y))\}.$$

A Hamiltonian $H \in \mathcal{H}$ defines a germ of vector field

$$\begin{cases} \dot{\theta} = \partial_r H \\ \dot{r} = -\partial_\theta H, \end{cases} \qquad \begin{cases} \dot{x} = \partial_y H \\ \dot{y} = -\partial_x H. \end{cases}$$

Let $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^m$. Split the integer m into $m = m' + m''$ (m' and m'' will respectively be the numbers of hyperbolic and elliptic directions), and let Q_β be the matrix

$$Q_\beta = 2\pi \operatorname{Diag}(\beta_1, ..., \beta_m, -\beta_1, ..., -\beta_{m'}, \beta_{m'+1}, ..., \beta_m) \in \mathrm{M}_{2m}(\mathbb{R}).$$

Define $\mathcal{K}(\alpha, \beta)$ as subset of $\mathcal{H}$ of Hamiltonians of the form

$$\begin{aligned} K &= c + \alpha \cdot r + \frac{1}{2} Q_\beta \cdot z^2 + O(r^2, rz, z^3) \\ &= \sum_{j=1}^{n} \alpha_j r_j + \pi \sum_{j=1}^{m'} \beta_j (x_j^2 - y_j^2) + \pi \sum_{j=m'+1}^{m} \beta_j (x_j^2 + y_j^2) + O(r^2, rz, z^3) \end{aligned}$$

where c is some (non-fixed) real number.

In the following definitions, maps are all real analytic. Let $\mathrm{B}^1(\mathbb{T}^n)$ be the group of exact 1-forms on $\mathbb{T}^n$, $\mathcal{D}$ be the group of isomorphisms of $\mathbb{T}^n$ fixing the origin,

$$Sp_{2m} = \left\{ \psi \in M_{2m}(\mathbb{R}),\ {}^t\psi J \psi = J \right\}, \quad J = \begin{pmatrix} 0 & -\mathrm{id}_{\mathbb{R}^m} \\ \mathrm{id}_{\mathbb{R}^m} & 0 \end{pmatrix},$$

be the symplectic group,

$$\mathcal{A}_*(\mathbb{T}^n, Sp(2m)) = \{\exp \Delta\psi \in \mathcal{A}(\mathbb{T}^n, Sp(2n)), \Delta\psi \in \mathcal{A}_*(\mathbb{T}^n, sp_{2m})\}$$

be the image by the exponential of the subspace

$$\mathcal{A}_*(\mathbb{T}^n, sp_{2m}(\mathbb{R})) =$$

$$\left\{\psi \in \mathcal{A}(\mathbb{T}^n, sp_{2m}(\mathbb{R})),\ {}^t\psi = \psi \text{ and } \int_{\mathbb{T}^p} \psi_{jj}(\theta)\, d\theta = 0,\ j = 1, ..., 2m\right\}. \tag{17}$$

Let now

$$\mathcal{G} = B^1(\mathbb{T}^n) \times \mathcal{A}(\mathbb{T}^n, \mathbb{R}^{2m}) \times \mathcal{D}_* \times \mathcal{A}_*(\mathbb{T}^n, Sp(2m)).$$

Let $a = (\theta, r, z) = (\theta, r, x, y) \in \mathbb{T}^p \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$ and $G = (\rho, \zeta, \varphi, \psi) \in \mathcal{G}$. If ψ is C^0-close to the constant map $\theta \mapsto \mathrm{id}_{\mathbb{R}^{2q}}$, there exists a unique $\dot{\psi} \in \mathrm{C}^\infty_*(\mathbb{T}^p, \mathrm{sp}(2q))$ such that $\psi = \exp \dot{\psi}$. Let

$$\begin{cases} \rho(a) & = (\theta, r + \rho, z) \\ \zeta(a) & = (\theta, r + R_\zeta, z + \zeta(\theta)) \\ \varphi(a) & = (\varphi(\theta), {}^t D\varphi(\theta)^{-1} \cdot r, z) \\ \psi(a) & = (\theta, r + S_\psi \cdot z^2, \psi(\theta) \cdot z), \end{cases}$$

with

$$\begin{cases} R_\zeta & = -J \cdot ((z + \zeta/2) D\zeta) \\ S_\psi \cdot z^2 = & \frac{1}{2} \int_0^1 \left(\exp(t\dot{\psi}) \cdot z\right)^2 \cdot \left(J \cdot D\dot{\psi}\right)\, dt, \end{cases} \tag{18}$$

and then

$$G(a) = \psi(\varphi(\zeta(\rho(a)))). \tag{19}$$

This defines an exact symplectomorphism [38].

The generalized twisted conjugacy theorem is as follows.

Assume (α, β) is Diophantine in this sense: for every $k \in \mathbb{Z}^n$, $l' \in \mathbb{Z}^{m'}$ and $l'' \in \mathbb{Z}^{m''}$ such that $|l'|, |l''| = 1$ or 2,

$$\begin{cases} |k \cdot \alpha| \geq \frac{\gamma}{|k|^\tau} & \text{(if } k \neq 0) \\ |l' \cdot \beta'| \geq \gamma & \\ |k \cdot \alpha + l'' \cdot \beta''| \geq \frac{\gamma}{(|k|+1)^\tau} & \textit{(Melnikov condition)} \end{cases} \tag{20}$$

36 Theorem. *If $H \in \mathcal{H}$ is close enough to some $K^o \in \mathcal{K}(\alpha, \beta)$, there exists a unique $(K, G, \hat{\alpha}, \hat{\beta}) \in \mathcal{K}(\alpha, \beta) \times \mathcal{G} \times \mathbb{R}^n \times \mathbb{R}^m$ close to $(K^o, \mathrm{id}, 0, 0)$ such that*

$$H = K \circ G + \hat{\alpha} \cdot r + \frac{1}{2} Q_{\hat{\beta}} \cdot z^2.$$

We will skip the proof here. It only combines the same formal ideas as in the smooth category [38] and the inverse function theorem of section 3.

The theory for lower dimensional tori unwinds as in the Lagrangian case. Of course, there is no direct analogue of Kolmogorov's theorem if $m'' > 0$, since there are not

enough action variables to control all of the tangent and normal frequencies. Let us merely give one statement, corresponding to theorem 30.
Consider a smooth family $(K_t^o)_{t\in\mathbb{B}^\kappa}$ of Hamiltonians in some $\mathcal{K}_s$, defining a frequency map

$$(\alpha^o, \beta^o) : \mathbb{B}^\kappa \mapsto \mathbb{R}^n, \quad t \mapsto (\alpha_t^o, \beta_t^o).$$

Let (H_t) be a smooth family of Hamiltonians in $\mathcal{H}_s$ such that, for each t, H_t is close enough to K_t.

37 Theorem. *If the frequency map (α^o, β^o) is skew, there is a subset $\mathcal{T} \subset \mathbb{B}^\kappa$ of positive Lebesgue measure such that, for all $t \in \mathcal{T}$, H_t has a Diophantine quasiperiodic invariant torus.*

Bibliographical comments. – The existence of normally hyperbolic tori has been acknowledged early, since hyperbolic normal directions do not interfere with the tangent quasiperiodic dynamics [54].
– It was a surprise when H. Eliasson proved an invariant torus theorem for normaly elliptic tori [31], due to the problem of the lack of parameters.
– Bourgain later proved that it suffices to assume $|l| = 1$ in the Melnikov condition. The proof is more difficult since one cannot straighten the normal dynamics of the torus, so the linearized equations are not diagonal anymore in Fourier space [76].

10 Example in the spatial three-body problem

The Hamiltonian of the three-body problem is

$$H = \sum_{0\le j\le 2} \frac{\|p\|^2}{2m_j} - \sum_{0\le j<k\le 2} \frac{m_j m_k}{\|q_j - q_k\|},$$

where $q_j \in \mathbb{R}^3$ is the position of the j-th body and $p_j \in \mathbb{R}^3$ is its impulsion. Periodic solutions have been advertised by Poincaré as the only breach through which to enter the impregnable fortress of the three-body problem. Conjecturally they are dense in the phase space, but also of zero measure. In contrast, we will prove the existence of quasiperiodic motions, at least here in the *hierarchical* (or *lunar*) *problem*, where two bodies (say, q_0 and q_1) revolve around each other while the third body revolves, far away, around the center of mass of the two primaries. Another classical perturbative regime would have been the *planetary problem*, where there is no assumption on the distances of the bodies, but two masses (planets) are assumed small with respect to the remaining one (Sun).

38 Theorem. *There exist a set of initial conditions of positive Lebesgue measure leading to quasiperiodic solutions, arbitrarily close to Keplerian, coplanar, circular motions, with semi major axis ratio arbitrarily small.*

The hurried reader may simplify the following discussion by focusing on the plane invariant subproblem.

Let $(Q_0, Q_1, Q_2, P_0, P_1, P_2)$ be the Jacobi coordinates, defined by:

$$\begin{cases} Q_0 = q_0 \\ Q_1 = q_1 - q_0 \\ Q_2 = q_2 - \sigma_0 q_0 - \sigma_1 q_1, \end{cases} \qquad \begin{cases} P_0 = p_0 + p_1 + p_2 \\ P_1 = p_1 + \sigma_1 p_2 \\ P_2 = p_2, \end{cases}$$

where $1/\sigma_0 = 1 + m_1/m_0$ and $1/\sigma_1 = 1 + m_0/m_1$. P_0 is the total linear momentum, which can be assumed equal to 0 without loss of generality. Besides, H does not depend on Q_0. So, (Q_1, Q_2, P_1, P_2) is a symplectic coordinate system on the phase space reduced by the symmetry of translation, and the equations read

$$\begin{cases} \dot{Q}_i = \partial_{P_i} H \quad (i = 1, 2) \\ \dot{P}_i = -\partial_{Q_i} H. \end{cases}$$

A direct computation shows that

$$H = \sum_{1 \le i \le 2} \frac{\|P_i\|^2}{2\mu_i} - \sum_{0 \le i < j \le 2} \frac{m_i m_j}{\|q_i - q_j\|},$$

with

$$M_0 = m_0, \quad M_1 = m_0 + m_1 \quad \text{and} \quad \frac{1}{\mu_i} = \frac{1}{M_{i-1}} + \frac{1}{m_i}.$$

One can split H into two parts

$$H = \mathrm{Kep} + \mathrm{Rem}$$

where

$$\mathrm{Kep} = \sum_{1 \le i \le 2} \left(\frac{\|P_i\|^2}{2\mu_i} - \frac{\mu_i M_i}{\|Q_i\|} \right)$$

is a sum of two uncoupled Kepler problems, and

$$\mathrm{Rem} = \frac{\mu_2 M_2}{\|Q_2\|} - \frac{m_0 m_2}{\|q_2 - q_0\|} - \frac{m_1 m_2}{\|q_2 - q_1\|}$$

is the remainder.

Let us assume that the two terms of Kep are negative so that each body Q_i under the flow of Kep describes a Keplerian ellipse. Let $(\ell_i, L_i, g_i, G_i, \theta_i, \Theta_i)_{i=1,2}$ be the associated Delaunay coordinates. These coordinates are symplectic and analytic over the open set where motions are non-circular and non-horizontal [41]; since we will precisely be interested in a neighborhood of circular coplanar motions, these variables are only intermediate coordinates for computations. One shows that

$$\mathrm{Kep} = -\sum_{i=1,2} \frac{\mu_i^3 M_i^2}{2L_i^2}.$$

The Keplerian frequencies[9] are

$$\kappa_i = \frac{\partial\, \mathrm{Kep}}{\partial L_i} = \frac{\mu_i^3 M_i^2}{L_i^3} = \frac{\sqrt{M_i}}{a_i^{3/2}},$$

[9] Traditionally given the ununderstandable name of *mean motions*.

so that the *Keplerian frequency map*

$$\kappa : (L_1, L_2) \mapsto (\kappa_1, \kappa_2)$$

is a diffeomorphim $(\mathbb{R}_+^*)^2 \to (\mathbb{R}_+^*)^2$. Due to the fact that the Keplerian part depends only on 2 of the action variables, solutions of the Keplerian approximation are quasiperiodic with atmost 2 independent frequencies. This degeneracy has been interpreted as a hidden $SO(4)$-symmetry for each planet, whose momentum map is given partly by the eccentricity vector. How the Keplerian ellipses slowly rotate and deform will be determined by mutual attractions. This degeneracy is specific to the Newtonian and elastic potentials, as Bertrand's theorem asserts [11].

In the hierarchical regime ($a_1 \ll a_2$) the dominating term of the remainder is

$$\text{Main} := -\mu_1 m_2 P_2(\cos\theta) \frac{\|X_1\|^2}{\|X_1\|^3}, \tag{21}$$

with $P_2(c) = \frac{1}{2}(3c^2 - 1)$ (second Legendre polynomial) and $\theta = \widehat{Q_1, Q_2}$. Since the Keplerian frequencies satisfy $\kappa_1 \gg \kappa_2$, we may average out the fast, Keplerian angles ℓ_1 and ℓ_2 successively, thus without small denominators [37, 54]. The *quadrupolar Hamiltonian* is

$$\text{Quad} = \int_{\mathbb{T}^2} \text{Main} \, \frac{d\ell_1 \, d\ell_2}{4\pi^2}; \tag{22}$$

It is the dominating interaction term which rules the slow deformations of the Keplerian ellipses. It naturally defines a Hamiltonian on the space of pairs of Keplerian ellipses with fixed semi major axes. This space, called the *secular space*, is locally diffeomorphic to $\mathbb{R}^8$, whose origin corresponds to circular horizontal ellipses.

After reduction by the symmetry of rotations (e.g. with Jacobi's reduction of the nodes, which consists in fixing the angular momentum vector, say, vertically, and quotienting the so-obtained codimension-3 Poisson submanifold by rotations around the angular momentum), the secular space has 4 dimensions, with coordinates (g_1, G_1, g_2, G_2) outside coplanar or circular motions.

39 Lemma. *The quadrupolar system* Quad *is integrable.*

Indeed, it happens that Quad does not depend on the argument g_2 of the pericenter of the outer ellipse (but the next higher order term, the "octupolar term", does), thus proving its integrability:

$$\text{Quad} = -\frac{\mu_1 m_2 a_1^2}{8a_2^3 \, (1 - e_2^2)^{3/2}} \begin{bmatrix} (15e_1^2 \cos^2 g_1 - 12e_1^2 - 3) \sin^2(i_2 - i_1) \\ +3e_1^2 + 2 \end{bmatrix}, \tag{23}$$

where i_j is the inclination of the ellipse of Q_j with respect to the Laplace plane (e.g. [63]); the Hamiltonian in the plane problem is simply obtained by letting $i_1 = i_2 = 0$.

We now need to estimate the frequencies and the torsion of the quadrupolar system, somewhere in the secular space. Lidov-Ziglin [63] have established the bifurcation diagram of the system, and proved the existence of 5 regimes in the parameter space, according to the number of equilibrium points of the reduced quadrupolar system. Here, for the sake of simplicity we will localize our study in some regular region (i.e. a region with a uniform action-angle coordinate system) and, more specifically, on a neighborhood of the origin of the secular space, i.e. circular horizontal Keplerian ellipses. See [60] for more details on the computations.

40 Lemma (Lagrange, Laplace)**.** *The first quadrupolar system has a degenerate elliptic singularity at the origin of the secular space, whose normal frequency vector is*

$$\alpha_{\mathrm{Quad}}(0) = -\frac{3a_1^2}{4a_2^3}\begin{pmatrix} \Lambda_1^{-1} \\ \Lambda_2^{-1} \\ -\Lambda_1^{-1} - \Lambda_2^{-1} \\ 0 \end{pmatrix}.$$

Proof. The following steps lead to the desired expansion of Quad:

- Using elementary geometry, express $\cos\theta_{12}$ in terms of the elliptic elements and the true anomalies. Then substitute the variable u_1 for v_1, using the relations

$$\cos(v_1) = \frac{a_1}{\|X_1\|}(\cos u_1 - e_1) \quad \text{and} \quad \sin v_1 = \frac{a_1}{\|X_1\|}\sqrt{1-e_1^2}\sin u_1.$$

- Multiply Main by the Jacobian of the change of angles

$$\frac{d\ell_1\, d\ell_2}{du_1\, dv_2} = \frac{\|X_1\|}{a_1}\,\frac{\|X_2\|^2}{a_2\sqrt{1-e_2^2}}.$$

- In the integrand of (22) with $i = 2$, express the distances to the Sun in terms of the inner eccentric anomaly u_1 and outer true anomaly v_2:

$$\|X_1\| = a_1(1 - e_1\cos u_1) \quad \text{and} \quad \|X_2\| = \frac{a_2(1-e_2^2)}{1+e_2\cos v_2},$$

  and expand at the second order with respect to eccentricitites and inclinations (odd powers vanish; the fourth order yields the second Birkhoff invariant and will be needed later).

- The obtained expression is trigonometric polynomial in u_1 and v_2. Average it.

- Switch to the Poincaré coordinates $(\xi_j, \eta_j, p_j, q_j)$, which are symplectic and analytic in the neighborhood of circular horizontal ellipses, and are defined by the relations

$$\begin{cases} \xi_j + i\eta_j = \sqrt{2L_j}\sqrt{1-\sqrt{1-e_j^2}}\, e^{-i(g_j+\theta_j)} \\ p_j + iq_j = \sqrt{2L_j}\sqrt{\sqrt{1-e_j^2}(1-\cos i_j)}\, e^{-i\theta_j}; \end{cases}$$

  here we use the notation $\Lambda_j = L_j = \mu_j\sqrt{M_j a_j}$.

The invariance of Quad by horizontal rotations entails that, as proved by Lagrange and Laplace, there exist two quadratic forms Q_h and Q_v (indices h and v here stand for "horizontal" and "vertical") on $\mathbb{R}^2$ such that

$$\mathrm{Quad} = -\frac{3a_1^2}{8a_2^3}\left(Q_h(\xi) + Q_h(\eta) + Q_v(p) + Q_v(q) + O_4(\xi, \eta, p, q)\right).$$

The computation shows that

$$\begin{cases} Q_h(\xi) = \frac{\xi_1^2}{L_1} + \frac{\xi_2^2}{L_2} \\ Q_v(p) = -\frac{p_1^2}{L_1} - \frac{p_2^2}{L_2} + \frac{2p_1p_2}{\sqrt{L_1L_2}} \end{cases}$$

The horizontal part is already in diagonal form. The vertical part Q_v is diagonalized by the orthogonal operator of $\mathbb{R}^2$

$$\rho = \frac{1}{\sqrt{\Lambda_1 + \Lambda_2}} \begin{pmatrix} \sqrt{\Lambda_2} & \sqrt{\Lambda_1} \\ -\sqrt{\Lambda_1} & \sqrt{\Lambda_2} \end{pmatrix}.$$

This operator of $\mathbb{R}^2$ lifts to a symplectic operator

$$\begin{aligned} \tilde{\rho}\ &: (x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4) \\ &\mapsto (\xi_1, \eta_1, \xi_2, \eta_2, p_1, q_1, p_2, q_2) = (x_1, y_1, x_2, y_2, p_1, q_1, p_2, q_2) \end{aligned}$$

with

$$(p_1, p_2) = \rho \cdot (x_3, x_4) \quad \text{and} \quad (q_1, q_2) = \rho \cdot (y_3, y_4).$$

In the new coordinates,

$$\mathrm{Quad} = -\frac{a_1^2}{4a_2^3}\begin{pmatrix} 1 + \frac{3}{2\Lambda_1}(x_1^2 + y_1^2) + \frac{3}{2\Lambda_2}(x_2^2 + y_2^2) - \\ \frac{3}{2}\left(\frac{1}{\Lambda_1} + \frac{1}{\Lambda_2}\right)(x_3^2 + y_3^2) + O_4(x, y) \end{pmatrix}, \tag{24}$$

thus showing that the origin is elliptic, and degenerate (since there is no term in $x_4^2 + y_4^2$).
Switching (outside the origin) to symplectic polar coordinates $(\tilde{\varphi}_j, \tilde{r}_j)_{j=1,\dots,4}$ defined by

$$x_j + iy_j = \sqrt{2\tilde{r}_j}\, e^{-i\tilde{\varphi}_j},$$

one gets the wanted expression of $\alpha_{\mathrm{Quad}}(0) = \frac{\partial\,\mathrm{Quad}}{\partial \tilde{r}}(0)$. □

It is an exercise (e.g. using generating functions) to check that all the changes of coordinates we have made on the secular space lift to changes of coordinates in the full phase space, up to adequately modifying the mean longitude. This does not change the Keplerian frequencies.
The quadrupolar frequency vector $\alpha_{\mathrm{Quad}}(0)$ calls for some comments:

- Due to the $SO(3)$-symmetry, rotations of the two inner ellipses around a horizontal axis leave Quad invariant. Hence the infinitesimal generators of such rotations (last two columns of the matrix $\tilde{\rho}$ in the proof of lemma 40) span an eigenplane of the quadratic part of (24), with eigenvalue 0. This explains for the vanishing last component of the normal frequency vector (for all r's for that matter):

$$\alpha_{\mathrm{Quad}}(0)_4 = 0 \qquad (\forall \Lambda_1, \Lambda_2). \tag{25}$$

- Unexpectedly, the sum of the frequencies vanishes:

$$\sum_{1\leq j\leq 4} \alpha_{\text{Quad}}(0)_j = 0 \qquad (\forall \Lambda_1, \Lambda_2). \tag{26}$$

- The local image of the map $(\Lambda_1, \Lambda_2) \mapsto \alpha_{\text{Quad}}(0)$ thus lies in a 2-plane of $\mathbb{R}^4$ but in no line, since the map

$$(\Lambda_1, \Lambda_2) \mapsto -\frac{3a_1^2}{4a_2^3}\begin{pmatrix}\Lambda_1^{-1}\\ \Lambda_2^{-1}\end{pmatrix} = -\frac{3}{4}\frac{M_2^3}{M_1^2}\frac{\mu_2^6}{\mu_1^4}\begin{pmatrix}\Lambda_1^3\Lambda_2^{-6}\\ \Lambda_1^4\Lambda_2^{-7}\end{pmatrix}$$

is a diffeomorphism. Hence, additional resonances may always be removed by slightly shifting Λ_1 and Λ_2.

41 Proposition. *The local image of the frequency map*

$$(\mathbb{R}_*^+)^2 \to \mathbb{R}^6, \quad (a_1, a_2) \mapsto \alpha = (\kappa_1, \kappa_2, \alpha_{\text{Quad}}(0))$$

is contained in the codimension-2 subspace

$$\alpha_6 = 0, \quad \alpha_3 + \alpha_4 + \alpha_5 = 0 \tag{27}$$

but in no subspace of larger codimension.

Proof. What remains to be checked is the second, negative assertion, i.e. that the frequency map

$$\begin{aligned}\tilde{\alpha} : (a_1, a_2) \mapsto \tilde{\alpha} &= (\kappa_1, \kappa_2, \alpha_{\text{Quad}}(0)_1, \alpha_{\text{Quad}}(0)_2)\\ &= \left(c_1 a_1^{-3/2}, c_2 a_2^{-3/2}, c_3 a_1^{3/2} a_2^{-3}, c_4 a_1^2 a_2^{-7/2}\right)\end{aligned}$$

is skew, where the c_i's depend only on the masses. Restricting for example to the curve $a_2 = a_1^3$, one gets a frequency vector whose components are Laurent monomials in $\sqrt{a_1}$, with components of pairwise distinct degrees. Such a curve is skew according to example 25 (using the fact that extracting components of the monomial curve preserves the skew property). □

Resonances (25) and (26) a priori prevent from eliminating all terms in the Lindstedt (or Birkhoff) normal form of Quad, and from applying theorem 30. And resonant terms will not disappear by adjusting the Λ_j's. But, as the following lemma shows, there are no resonant terms at the second order in $\tilde{r}$.
Let

$$\mathcal{L}(2) = \left\{(\Lambda_1, \Lambda_2) \in \mathbb{R}^2,\ \forall |k| \leq 4,\ k_1\Lambda_1^{-1} + k_2\Lambda_2^{-1} \neq 0\right\}$$

be the open set of values of (Λ_1, Λ_2) for which the horizontal first quadrupolar frequency vector satisfies no resonance of order ≤ 4. Here we will restrict to $\mathcal{L}(2)$ for the sake of precision, although, when we let a_1/a_2 tend to 0 later in the lunar problem, this restriction will become an empty constraint.

42 Lemma. *If the parameters* (Λ_1, Λ_2) *belong to* $\mathcal{L}(2)$*,* Quad *has a non-resonant Lindstedt normal form at order* 2 *i.e., there exist coordinates* $(\varphi_j, r_j)_{j=1,\dots,4}$*, tangent to* $(\tilde\varphi_j, \tilde r_i)_{j=1,\dots,4}$*, such that*

$$\mathrm{Quad} = \mathrm{cst} + \alpha_{\mathrm{Quad}}(0) \cdot r + \frac{1}{2}\tau_{\mathrm{Quad}} \cdot r^2 + O(r^3).$$

Besides, the torsion is

$$\tau_{\mathrm{Quad}} = \left(\begin{array}{ccc|c} & & & 0 \\ & \bar\tau_{\mathrm{Quad}} & & 0 \\ & & & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right)$$

with $\bar\tau_{\mathrm{Quad}} =$

$$-\frac{a_1^2}{8a_2^3}\frac{1}{\Lambda_1^2\Lambda_2^2}\begin{pmatrix} -6\,\Lambda_2^2 & 18\,\Lambda_1\,\Lambda_2 & -24\,\Lambda_2^2 - 18\,\Lambda_1\,\Lambda_2 \\ 18\,\Lambda_1\,\Lambda_2 & 24\,\Lambda_1^2 & -18\,\Lambda_1\,\Lambda_2 - 24\,\Lambda_1^2 \\ -24\,\Lambda_2^2 - 18\,\Lambda_1\,\Lambda_2 & -18\,\Lambda_1\,\Lambda_2 - 24\,\Lambda_1^2 & 6\,\Lambda_2^2 + 18\,\Lambda_1\,\Lambda_2 + 6\,\Lambda_1^2 \end{pmatrix}.$$

Proof. We carry out the same computation as in the proof of lemma 40, now up to the order 2 in the $\tilde r_j$'s. The truncated expression is a trigonometric polynomial in the angles $\tilde\varphi_j$, of degree ≤ 4. Eliminating non-resonant monomials, i.e. functions of $k \cdot \varphi$ with $k \cdot \alpha_{\mathrm{Quad}}(0)$, is a classical matter. Two kinds of terms cannot be eliminated by averaging:

- Monomials in the angle $4\tilde\varphi_4$.
- Monomials in $\tilde\varphi_1 + \tilde\varphi_2 + \tilde\varphi_3 + \tilde\varphi_4$.

Such monomials actually cannot occur in the expansion, due to the invariance by rotations (they would not satisfy the d'Alembert relation [25, 65]).[10] A direct computation leads to the given expression of the torsion τ_{Quad}. □

Note that the torsion τ_{Quad}, as a function of Λ_1 and Λ_2, extends analytically outside $\mathcal{L}(2)$ (as often do first order normal forms). This allows us to define the *quadrupolar frequency map*

$$\alpha_{\mathrm{Quad}} : r \mapsto \alpha_{\mathrm{Quad}} + \frac{1}{2}\tau_{\mathrm{Quad}} \cdot r,$$

a first order approximation of the normal frequencies.

43 Proposition. *The first quadrupolar frequency map has constant rank* 3 *and, in restriction to the symplectic submanifold obtained by fixing vertically the direction of the angular momentum, is a local diffeomorphism.*

Proof. For this lemma, we denote by $C = (C_x, C_y, C_z) \in \mathbb{R}^3$ the angular momentum of the first two planets. The submanifold $\mathcal{V}$ of vertical angular momentum,

[10] Thank you to Gabriella Pinzari for pointing out to me that the second kind of monomials too, is ruled out by d'Alembert's relations.

has equation $C_x = C_y = 0$. It is a symplectic, codimension-2 submanifold, transverse to the Hamiltonian vector fields X_{C_x} and X_{C_y} of C_x and C_y. Since it is invariant by the flow of Quad, its tangent space has equations, in the coordinates $(x_j, y_j)_{j=1,\dots,4}$ of the proof of lemma 40, $x_4 = y_4 = 0$. So, the upper left 3×3 submatrix $\tilde{\tau}_{\text{Quad}}$ of τ_{Quad} is the Hessian of the restriction of Quad to $\mathcal{V}$. In order to conclude, one merely needs to notice that the determinant of the torsion $\tilde{\tau}_{\text{Quad}}$:

$$\det \tilde{\tau}_{\text{Quad}} = -\left(\frac{a_1^2}{a_2^3}\right)^3 \frac{27}{64} \frac{1}{\Lambda_1^2 \Lambda_2^2} \left(39\Lambda_1^2 + 39\Lambda_1\Lambda_2 + 4\Lambda_2^2\right)$$

is non-zero. □

So, Quad (adequately truncated) has a non-degenerate quasiperiodic dynamics in the three degrees of freedom corresponding to coordinates $(\psi_j, s_i)_{j=1,2,3}$.

End of the proof of theorem 38. We would like to prove the persistence of some of the invariant tori of our normal form, which have frequencies of the following order (assuming $a_1 = O(1)$ and $a_2 \to \infty$):

$$\alpha = O(1, a_2^{-3/2}, a_2^{-3}, a_2^{-7/2}, a_2^{-3}, 0).$$

The conclusion thus follows from any of the four arguments below:

- the first item of theorem 34 (using proposition 43)
- the second item of theorem 34 (using again proposition 43), which yields not only the precedingly found Diophantine tori but also resonant tori (which induce Diophantine tori after reduction by the symmetry of rotation)
- the third item of theorem 34 (using proposition 41, for which the computation of the torsion is not needed, at the expense of deteriorating measure estimates)
- the fourth item of theorem 34 (using again proposition 41).

In three cases, the existence of Diophantine invariant tori of Kep + Quad is proved, either at the partially reduced level or at the fully reduced one. Locally they will have positive measure provided $\gamma = O(a_2^{-7/2})$. Theorem 34 applies with a perturbation of the size $|H - K^o| = O(\gamma^N)$ for some N (remark 31). Thus the theorem really applies to the perturbation of the normal form of the Hamiltonian of order $\sim 7N/2$ in $1/a_2$. □

Bibliographical comments. – The discovery of the eccentricity vector is often wrongly attributed to Runge and Lenz [2].
– For an anachronistic proof of Bertrand's theorem using Kolmogorov's theorem, see [42].
– Lemma 39 is obvious in the plane problem, where the analoguous reduction leads to a 2-dimensional reduced secular space, with coordinates (g_1, G_1) (see [36, 64]). This is less so in space. Harrington noticed it only after having carried out the

computation [50]. Lidov-Ziglin [63] called this a "happy coincidence", and indeed this invariance allowed them to study the bifurcation diagram of the quadrupolar Hamiltonian Quad. This was also crucial in various studies [54, 104, 105].

– Among the many accounts of the work of Lagrange and Laplace (comprising lemma 40), we refer to [97, 59, 43].

– Resonance (26) of order 3 was known to Clairaut, noticed by Delaunay as *un résultat singulier* [27], and discovered by Herman in the general n-planet problem.

– In the proof of lemma 42, it is a happy coincidence that resonant terms associated with the second resonance actually do not occur at our order of truncation. Malige has computed that higher degree resonant monomials occur, starting at order 10 [65].

– The strategy in Fejoz [38] for proving Arnold's theorem corresponds to the third argument given above at the end of the proof of theorem 38. The strategy of Chierchia-Pinzari [26] corresponds to the first and second arguments.

A Isotropy of invariant tori

Let (X, ω) be a symplectic manifold, (φ_t) a symplectic flow and T be a minimal quasiperiodic invariant embedded torus for (φ_t).

44 Lemma. *If ω is exact, T is isotropic.*

Proof. We may assume that $\varphi_t(\theta) = \theta + t\alpha$ ($t \in \mathbb{R}$, $\theta \in \mathbb{T}^n$) for some non-resonant vector α. Let

$$\nu = \sum_{i<j} \nu_{ij}(\theta)\, d\theta_i \wedge d\theta_2$$

be the 2-form induced by ω on T. Since (φ_t) preserves ω, for all t we have

$$\nu_{ij}(\theta + t\alpha) = \nu(\theta).$$

Since the flow on T is minimal, ν_{ij} is constant. By integrating with respect to θ_i and θ_j, this constant must be zero (more learnedly: according to the Hodge theorem, the zero 2-form is the unique harmonic representative of the cohomology class). □

If ω is not exact, the conclusion may be wrong. M. Herman has even constructed codimension-2 minimal invariant tori, in such a robust manner that this disproved the quasi-ergodic hypothesis [100].

B Two basic estimates

The following lemma is used in two instances in the proof of lemma 3, as well as in the proof of Kolmogorov's theorem 20.

45 Lemma (Cohomological equation). *Let s and σ be given in $]0,1]$. If $g \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})$, there exists a unique function $f \in \mathcal{A}(\mathbb{T}^n_s)$ of 0-average such that*

$$L_\alpha f = g - \int_{\mathbb{T}^n} g(\theta)\, d\theta,$$

and there exists a $C_c = C_c(n,\tau)$ such that, for any s, σ:

$$|f|_s \le C_c \gamma^{-1} \sigma^{-\tau_c} |g|_{s+\sigma}, \quad \tau_c = \tau + n.$$

Proof. Up to substituting $g - \int_{\mathbb{T}^n} g$, we may assume that g has zero average. Then, let $g(\theta) = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} g_k\, e^{i\,2\pi k\cdot\theta}$ be the Fourier expansion of g. The unique formal solution to the equation $L_\alpha f = g$ is given by $f(\theta) = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} \frac{g_k}{i\,2\pi\, k\cdot\alpha}\, e^{i\,2\pi\, k\cdot\theta}$. Since g is analytic, its Fourier coefficients decay exponentially: we find

$$|g_k| = \left| \int_{\mathbb{T}^n} g(\theta)\, e^{-ik\cdot\theta}\, \frac{d\theta}{2\pi} \right| \le |g|_{s+\sigma} e^{-|k|(s+\sigma)}$$

by shifting the torus of integration to a torus $\operatorname{Im}\theta_j = \pm(s+\sigma)$ (the sign depending on the sign of k_j). Using this estimate and replacing the small denominators $k\cdot\alpha$ by its Diophantine lower bound, we get

$$\begin{aligned} |f|_s &\le \frac{|g|_{s+\sigma}}{\gamma} \sum_k |k|^\tau\, e^{-|k|\,\sigma} \\ &\le \frac{2^n |g|_{s+\sigma}}{\gamma} \sum_{\ell\ge 1} \binom{\ell+n-1}{\ell} \ell^\tau\, e^{-\ell\,\sigma} \le \frac{4^n |g|_{s+\sigma}}{\gamma\,(n-1)!} \sum_\ell (\ell+n-1)^{\tau+n-1}\, e^{-\ell\,\sigma}, \end{aligned}$$

where, as a change of variable and a rough approximation show, the latter sum is bounded by

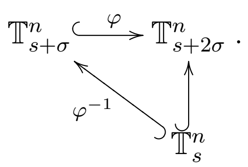
$$\int_1^\infty (\ell+n-1)^{\tau+n-1} e^{-(\ell-1)\sigma}\, d\ell < \sigma^{-\tau-n} e^{n\sigma} \int_0^\infty \ell^{\tau+n-1} e^{-\ell}\, d\ell$$

Hence f belongs to $\mathcal{A}(\mathbb{T}^n_s)$ and satisfies the wanted estimate. □

Bibliographical comments. The estimate has been obtained by bounding the terms of Fourier series one by one. In a more careful estimate, one should take into account the fact that if $|k\cdot\alpha|$ is small, then $k'\cdot\alpha$ is not so small for neighboring k''s. This makes it possible to find the optimal exponent of σ, uniformly with respect to the dimension [70, 82].

We have also used the following inverse function theorem. Recall that we have set $\mathbb{T}^n_s := \{\theta \in \mathbb{C}^n/2\pi\mathbb{Z}^n,\ \max_{1\le j\le n} |\operatorname{Im}\theta_j| \le s\}$.

46 Proposition. *Let $v \in \mathcal{A}(\mathbb{T}^n_{s+2\sigma}, \mathbb{C}^n)$, $|v|_{s+2\sigma} < \sigma$. The map $\mathrm{id} + v : \mathbb{T}^n_{s+2\sigma} \to \mathbb{R}^n_{s+3\sigma}$ induces a map $\varphi : \mathbb{T}^n_{s+2\sigma} \to \mathbb{T}^n_{s+3\sigma}$ whose restriction $\varphi : \mathbb{T}^n_{s+\sigma} \to \mathbb{T}^n_{s+2\sigma}$ has a unique right inverse $\varphi^{-1} : \mathbb{T}^n_s \to \mathbb{T}^n_{s+\sigma}$:*



Furthermore,

$$|\varphi^{-1} - \mathrm{id}\,|_s \le |v|_{s+\sigma}$$

and, provided $2\sigma^{-1}|v|_{s+2\sigma} \le 1$*,*

$$|(\varphi^{-1})' - \mathrm{id}\,| \le 2\sigma^{-1}|v|_{s+2\sigma}.$$

Proof. Let $\Phi : \mathbb{R}^n_{s+2\sigma} \to \mathbb{R}^n_{s+3\sigma}$ be a continuous lift of $\mathrm{id} + v$ and $k \in M_n(\mathbb{Z})$, $k(l) := \Phi(x+l) - \Phi(x)$. Denote by $p : \mathbb{R}^n_s := \mathbb{R}^n \times i[-s,s]^n \to \mathbb{T}^n_s$ the universal covering of $\mathbb{T}^n_s$.

1. *Injectivity of* $\Phi : \mathbb{R}^n_{s+\sigma} \to \mathbb{R}^n_{s+2\sigma}$. Suppose that $x, \hat{x} \in \mathbb{R}^n_{s+\sigma}$ and $\Phi(x) = \Phi(\hat{x})$. By the mean value theorem,

   $$|x - \hat{x}| = |v(p\hat{x}) - v(px)| \le |v'|_{s+\sigma}|x - \hat{x}|,$$

   and, by Cauchy's inequality,

   $$|x - \hat{x}| \le \frac{|v|_{s+2\sigma}}{\sigma}|x - \hat{x}| < |\hat{x} - x|,$$

   hence $x = \hat{x}$.

2. *Surjectivity of* Φ*:* $\mathbb{R}^n_s \subset \Phi(\mathbb{R}^n_{s+\sigma})$. For any given $y \in \mathbb{R}^n_s$, the contraction

   $$f : \mathbb{R}^n_{s+\sigma} \to \mathbb{R}^n_{s+\sigma}, \quad x \mapsto y - v(x)$$

   has a unique fixed point, which is a pre-image of y by Φ.

3. *Injectivity of* $\varphi : \mathbb{T}^n_{s+\sigma} \to \mathbb{T}^n_{s+2\sigma}$. Suppose that $px,\, p\hat{x} \in \mathbb{R}^n_{s+\sigma}$ and $\varphi(px) = \varphi(p\hat{x})$, i.e. $\Phi(x) = \Phi(\hat{x}) + \kappa$ for some $\kappa \in \mathbb{Z}^n$. That k be in $GL(n,\mathbb{Z})$, follows from the invertibility of Φ. Hence, $\Phi\left(x - k^{-1}(\kappa)\right) = \Phi(\hat{x})$, and, due to the injectivity of Φ, $px = p\hat{x}$.

4. *Surjectivity of* $\varphi : \mathbb{T}^n_s \subset \varphi(\mathbb{T}^n_{s+\sigma})$. This is a trivial consequence of that of Φ.

5. *Estimate on* $\psi := \varphi^{-1} : \mathbb{T}^n_s \to \mathbb{T}^n_{s+\sigma}$. Note that the wanted estimate on ψ follows from the corresponding estimate for the lifted map $\Psi := \Phi^{-1} : \mathbb{R}^n_s \to \mathbb{R}^n_{s+\sigma}$. But, if $y \in \mathbb{R}^n_s$,

   $$\Psi(y) - y = -v(p\Psi(y)),$$

   hence $|\Psi - \mathrm{id}\,|_s \le |v|_{s+\sigma}$.

6. *Estimate on* ψ'. We have $\psi' = \varphi'^{-1} \circ \varphi$, where $\varphi'^{-1}(x)$ stands for the inverse of the map $\xi \mapsto \varphi'(x) \cdot \xi$. Hence

   $$\psi' - \mathrm{id} = \varphi'^{-1} \circ \varphi - \mathrm{id},$$

   and, under the assumption that $2\sigma^{-1}|v|_{s+2\sigma} \le 1$,

   $$|\psi' - \mathrm{id}\,|_s \le |\varphi'^{-1} - \mathrm{id}\,|_{s+\sigma} \le \frac{|v'|_{s+\sigma}}{1 - |v'|_{s+\sigma}} \le \frac{\sigma^{-1}|v|_{s+2\sigma}}{1 - \sigma^{-1}|v|_{s+2\sigma}} \le 2\sigma^{-1}|v|_{s+2\sigma}.$$

□

C Interpolation of spaces of analytic functions

In this section we prove some Hadamard interpolation inequalities, which are used in section 4.

Recall that we denote by $\mathbb{T}^n_{\mathbb{C}}$ the infinite annulus $\mathbb{C}^n/2\pi\mathbb{Z}^n$, by $\mathbb{T}^n_s$, $s > 0$, the bounded sub-annulus $\{\theta \in \mathbb{T}^n_{\mathbb{C}},\ |\mathrm{Im}\,\theta_j| \leq s,\ j = 1...n\}$ and by $\mathbb{D}^n_t$, $t > 0$, the polydisc $\{r \in \mathbb{C}^n,\ |r_j| \leq t,\ j = 1...n\}$. The supremum norm of a function $f \in \mathcal{A}(\mathbb{T}^n_s \times \mathbb{D}^n_t)$ will be denoted by $|f|_{s,t}$.

Let $0 < s_0 \leq s_1$ and $0 < t_0 \leq t_1$ be such that

$$\log \frac{t_1}{t_0} = s_1 - s_0.$$

Let also $0 \leq \rho \leq 1$ and

$$s = (1-\rho)s_0 + \rho s_1 \quad \text{and} \quad t = t_0^{1-\rho} t_1^{\rho}.$$

47 Proposition. *If* $f \in \mathcal{A}(\mathbb{T}^n_{s_1} \times \mathbb{D}^n_{t_1})$,

$$|f|_{s,t} \leq |f|_{s_0,t_0}^{1-\rho}\, |f|_{s_1,t_1}^{\rho}.$$

Proof. Let $\tilde f$ be the function on $\mathbb{T}^n_{s_1} \times \mathbb{D}^n_{t_1}$, constant on $2n$-tori of equations $(\mathrm{Im}\,\theta, |r|) = cst$, defined by

$$\tilde f(\theta, r) = \max_{\mu,\nu \in \mathbb{T}^n} \left| f\left((\pm\theta_1 + \mu_1, ..., \pm\theta_n + \mu_n), \left(r_1\, e^{i\nu_1}, ..., r_n\, e^{i\nu_n}\right)\right)\right|$$

(with all possible combinations of signs). Since $\log|f|$ is subharmonic (hence upper semi-continuous) and $\mathbb{T}^{2n}$ is compact, $\log \tilde f$ too is upper semi-continuous. Besides, $\log \tilde f$ satisfies the mean inequality, hence is plurisubharmonic.

By the maximum principle, the restriction of $|f|$ to $\mathbb{T}^n_s \times \mathbb{D}^n_t$ attains its maximum on the distinguished boundary of $\mathbb{T}^n_s \times \mathbb{D}^n_t$. Due to the symmetry of $\tilde f$:

$$|f|_{s,t} = \tilde f(is\epsilon, t\epsilon), \quad \epsilon = (1, ..., 1).$$

Now, the function

$$\varphi(z) := \tilde f(z\epsilon, e^{-(iz+s)} t\epsilon)$$

is well defined on $\mathbb{T}_{s_1}$, for it is constant with respect to $\mathrm{Re}\,z$ and, due to the relations imposed on the norm indices, if $|\mathrm{Im}\,z| \leq s_1$ then $|e^{-(iz+s)}t| \leq e^{s_1-s}t = t_1$.

The estimate

$$\log\varphi(z) \leq \frac{s_1 - \mathrm{Im}\,z}{s_1 - s_0}\varphi(s_0 i) + \frac{\mathrm{Im}\,z - s_0}{s_1 - s_0}\varphi(s_1 i)$$

trivially holds if $\mathrm{Im}\,z = s_0$ or s_1, for, as noted above for $j = 1$, $e^{s_j - s}t = t_j$, $j = 0, 1$. But the left and right hand sides respectively are suharmonic and harmonic. Hence the estimate holds whenever $s_0 \leq \mathrm{Im}\,z \leq s_1$, whence the claim for $z = is$. □

Recall that we have let $\mathrm{T}^n_s := \mathbb{T}^n_s \times \mathbb{D}^n_s$, $s > 0$, and, for a function $f \in \mathcal{A}(\mathrm{T}^n_s)$, let $|f|_s = |f|_{s,s}$ denote its supremum norm on T^n_s. As in the rest of the paper, we now restrict the discussion to widths of analyticity ≤ 1.

48 Corollary. *If* $\sigma_1 = -\log\left(1 - \frac{\sigma_0}{s}\right)$ *and* $f \in \mathcal{A}(\mathrm{T}^n_{s+\sigma_1})$,

$$|f|_s^2 \leq |f|_{s-\sigma_0} |f|_{s+\sigma_1}.$$

In section 4, we use the equivalent fact that, if $\tilde\sigma = s + \log\left(1 + \frac{\sigma}{s}\right)$ and $f \in \mathcal{A}(\mathrm{T}^n_{s+\tilde\sigma})$,

$$|f|_{s+\sigma}^2 \leq |f|_s |f|_{s+\tilde\sigma}.$$

Proof. In proposition 47, consider the following particular case :

- $\rho = 1/2$. Hence
$$s = \frac{s_0 + s_1}{2} \quad \text{and} \quad t = \sqrt{t_0 t_1}.$$
- $s = t$. Hence in particular $t_0 = s\, e^{s_0 - s}$ and $t_1 = s\, e^{s_1 - s}$.

Then

$$|f|_s^2 = |f|_{s,s}^2 \leq |f|_{s_0,t_0} |f|_{s_1,t_1}.$$

We want to determine $\max(s_0, t_0)$ and $\max(s_1, t_1)$. Let $\sigma_1 := s - s_0 = s_1 - s$. Then $t_0 = s\, e^{-\sigma_1}$ and $t_1 = s\, e^{\sigma_1}$. The expression $s + \sigma - s e^{\sigma}$ has the sign of σ (in the relevant region $0 \leq s + \sigma \leq 1$, $0 \leq s \leq 1$); by evaluating it at $\sigma = \pm\sigma_1$, we see that $s_0 \leq t_0$ and $s_1 \geq t_1$.
Therefore, since the norm $|\cdot|_{s,t}$ is non-decreasing with respect to both s and t,

$$|f|_s^2 \leq |f|_{t_0,t_0} |f|_{s_1,s_1} = |f|_{t_0} |f|_{s_1}$$

(thus giving up estimates uniform with respect to small values of s). By further setting $\sigma_0 = s - t_0 = s\,(1 - e^{-\sigma_1})$, we get the wanted estimate, and the asserted relation between σ_0 and σ_1 is readily verified. □

Bibliographical comments. – The obtained inequalities generalize the standard Hadamard inequalities. They are optimal and show that the convexity of analytic norms is twisted by the geometry of the phase space. See [73, Chap. 8] for more general but coarser inequalities.
– Interpolation inequalities in the analytic category do not depend on regularizing operators as they do in the Hölder or Sobolev cases. See, e.g. [52, Theorem A.5] or [49].

Acknowledgments. These notes are the expanded version of a chapter of the Habilitation memoir [39] and of a subsequent short course given at the workshop *Geometric control and related fields*, organized by J.-B. Caillau and T. Haberkorn in RICAM (Linz, November 2014). I thank J.-B. Caillau, A. Chenciner, J. Mather, G. Pinzari and J.-C. Yoccoz for their interest or suggestions. Paradoxycally, this work has been partially funded by the ANR project *Beyond KAM theory* (ANR-15-CE40-0001).

References

[1] J.F. Adams. *Lectures on Lie Groups*. Midway reprints. University of Chicago Press, 1969.

[2] A. Albouy. Lectures on the two-body problem. In *Classical and celestial mechanics (Recife, 1993/1999)*, pages 63–116. Princeton Univ. Press, Princeton, NJ, 2002.

[3] V. I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat. Nauk*, 18(6 (114)):91–192, 1963.

[4] V. I. Arnold. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1989. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[5] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt. *Mathematical aspects of classical and celestial mechanics*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, third edition, 2006. Translated from the Russian original by E. Khukhro.

[6] A. Avila, B. Fayad, and R. Krikorian. A KAM scheme for $\mathrm{SL}(2,\mathbb{R})$ cocycles with Liouvillean frequencies. *Geom. Funct. Anal.*, 21(5):1001–1019, 2011.

[7] A. Avila and R. Krikorian. Reducibility or nonuniform hyperbolicity for quasiperiodic Schrödinger cocycles. *Ann. of Math. (2)*, 164(3):911–940, 2006.

[8] V. I. Bakhtin. Averaging in multifrequency systems. *Funktsional. Anal. i Prilozhen.*, 20(2):1–7, 96, 1986.

[9] V. Bernik, D. Kleinbock, and G. A. Margulis. Khintchine-type theorems on manifolds: the convergence case for standard and multiplicative versions. *Internat. Math. Res. Notices*, 9:453–486, 2001.

[10] M. Berti. *Nonlinear oscillations of Hamiltonian PDEs*. Progress in Nonlinear Differential Equations and their Applications, 74. Birkhäuser Boston Inc., Boston, MA, 2007.

[11] J. Bertrand. Théorème relatif au mouvement d'un point attiré vers un centre fixe. *C. R. Math. Acad. Sci. Paris*, 73:849–853, 1873.

[12] J.-B. Bost. Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel, ...). *Astérisque*, 1984/85(133-134):113–157, 1986. Séminaire Bourbaki.

[13] A. Bounemoura and S. Fischler. The classical KAM theorem for Hamiltonian systems via rational approximations. *Regular and Chaotic Dynamics*, 19:251–265, March 2014.

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

[15] R. C. Calleja, A. Celletti, and R. de la Llave. A KAM theory for conformally symplectic systems: efficient algorithms and their validation. *J. Differential Equations*, 255(5):978–1049, 2013.

[16] T. Castan. *Complex singularities in the three-body problem and applications to stability and instability.* PhD thesis, Observatoire de Paris, 2016. In preparation.

[17] A. Chenciner. Bifurcations de points fixes elliptiques. I. Courbes invariantes. *Inst. Hautes Études Sci. Publ. Math.*, 61:67–127, 1985.

[18] A. Chenciner. Bifurcations de points fixes elliptiques. II. Orbites periodiques et ensembles de Cantor invariants. *Invent. Math.*, 80(1):81–106, 1985.

[19] A. Chenciner. Bifurcations de points fixes elliptiques. III. Orbites périodiques de "petites" périodes et élimination résonnante des couples de courbes invariantes. *Inst. Hautes Études Sci. Publ. Math.*, 66:5–91, 1988.

[20] A. Chenciner. Poincaré and the three-body problem. In *Henri Poincaré, 1912–2012*, volume 67 of *Prog. Math. Phys.*, pages 51–149. Birkhäuser/Springer, Basel, 2015.

[21] C. Q. Cheng and Y. S. Sun. Existence of KAM tori in degenerate Hamiltonian systems. *J. Differential Equations*, 114(1):288–335, 1994.

[22] L. Chierchia. KAM lectures. In *Dynamical Systems. Part I*, Pubbl. Cent. Ric. Mat. Ennio Giorgi, pages 1–55. Scuola Norm. Sup., Pisa, 2003.

[23] 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]. *Regul. Chaotic Dyn.*, 13(2):130–139, 2008.

[24] L. Chierchia and G. Pinzari. Properly-degenerate KAM theory (following V. I. Arnold). *Discrete Contin. Dyn. Syst. Ser. S*, 3(4):545–578, 2010.

[25] L. Chierchia and G. Pinzari. Planetary Birkhoff normal forms. *J. Mod. Dyn.*, 5(4):623–664, 2011.

[26] L. Chierchia and G. Pinzari. The planetary N-body problem: symplectic foliation, reductions and invariant tori. *Invent. Math.*, 186(1):1–77, 2011.

[27] C. Delaunay. Note sur les mouvements du périgée et du noeud de la lune. *Comptes Rendus hebdomadaires des Séances de l'Académie des Sciences*, 74:17, 1872.

[28] J. Lopes Dias. KAM theory as a limit of renormalization. In *Dynamics, games and science. I*, volume 1 of *Springer Proc. Math.*, pages 253–270. Springer, Heidelberg, 2011.

[29] H. S. Dumas. *The KAM story.* World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014.

[30] I. Ekeland and É. Séré. Nash-moser without regularity. unpublished, 2012.

[31] L. H. Eliasson. Perturbations of stable invariant tori for Hamiltonian systems. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 15(1):115–147 (1989), 1988.

[32] L. H. Eliasson. Absolutely convergent series expansions for quasi periodic motions. *Math. Phys. Electron. J.*, 2:Paper 4, 33 pp. (electronic), 1996.

[33] L. H. Eliasson, B. Fayad, and R. Krikorian. Kam-tori near an analytic elliptic fixed point. *Regular and Chaotic Dynamics*, 18(6):801–831, 2013.

[34] L. H. Eliasson, B. Fayad, and R. Krikorian. Around the stability of KAM tori. *Duke Math. J.*, 164(9):1733–1775, 2015.

[35] A. Fathi. *Weak Kam Theorem in Lagrangian Dynamics.* Cambridge Univ. Press, 2007.

[36] J. Féjoz. Global secular dynamics in the planar three-body problem. *Celestial Mech. Dynam. Astronom.*, 84(2):159–195, 2002.

[37] J. Féjoz. Quasiperiodic motions in the planar three-body problem. *J. Differential Equations*, 183(2):303–341, 2002.

[38] J. Féjoz. Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman). *Ergodic Theory Dynam. Systems*, 24(5):1521–1582, 2004.

[39] J. Féjoz. Mouvements périodiques et quasi-périodiques dans le problème des n corps. Mémoire d'Habilitation de l'Université P. et M. Curie, 2010.

[40] J. Féjoz. A proof of the invariant torus theorem of Kolmogorov. *Regular and Chaotic Dynamics*, 17:1–5, 2012.

[41] J. Féjoz. On action-angle coordinates and the Poincaré coordinates. *Regul. Chaotic Dyn.*, 18(6):703–718, 2013.

[42] J. Féjoz and L. Kaczmarek. Sur le théorème de Bertrand (d'après Michael Herman). *Ergodic Theory Dynam. Systems*, 24(5):1583–1589, 2004.

[43] J. Féjoz and S. Serfaty. *Deux cents ans après Lagrange.* Journée Annuelle, Paris, le 28 juin 2013. Société Mathématique de France, 2013.

[44] G. Forni. Solutions of the cohomological equation for area-preserving flows on compact surfaces of higher genus. *Ann. of Math. (2)*, 146(2):295–344, 1997.

[45] G. Glaeser. Étude de quelques algèbres tayloriennes. *J. Analyse Math.*, 6:1–124; erratum, insert to 6 (1958), no. 2, 1958.

[46] V. Guillemin and S. Sternberg. Convexity properties of the moment mapping. *Invent. Math.*, 67:491–513, 1982.

[47] V. Guillemin and S. Sternberg. *Symplectic techniques in physics*. Cambridge University Press, Cambridge, second edition, 1990.

[48] B. Hall. *Lie groups, Lie algebras, and representations*, volume 222 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2015. An elementary introduction.

[49] R. S. Hamilton. The inverse function theorem of Nash and Moser. *Bull. Amer. Math. Soc. (N.S.)*, 7(1):65–222, 1982.

[50] R. S. Harrington. Dynamical evolution of triple stars. *Astronom. J.*, pages 190–194, 1968.

[51] M. R. Herman. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publications Mathématiques de l'IHÉS*, 49:5–233, 1979.

[52] L. Hörmander. The boundary problems of physical geodesy. *Arch. Rational Mech. Anal.*, 62(1):1–52, 1976.

[53] H. Jacobowitz. Implicit function theorems and isometric embeddings. *Ann. of Math. (2)*, 95:191–225, 1972.

[54] W. H. Jefferys and J. Moser. Quasi-periodic solutions for the three-body problem. *Astronom. J.*, 71:568–578, 1966.

[55] D. Y. Kleinbock and G. A. Margulis. Flows on homogeneous spaces and Diophantine approximation on manifolds. *Ann. of Math. (2)*, 148(1):339–360, 1998.

[56] A. Knauf. *Mathematische Physik: Klassische Mechanik*. Springer-Verlag, 2012.

[57] A. N. Kolmogorov. On the conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR (N.S.)*, 98:527–530, 1954.

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

[59] J. Laskar. *Sfogliando La Méchanique analitique, Giornata di studio su Louis Lagrange*, chapter Lagrange et la stabilité du système solaire. Edizioni Universitarie di Lettere Economia Diritto, Milano, 2006.

[60] J. Laskar and P. Robutel. Stability of the planetary three-body problem. I. Expansion of the planetary Hamiltonian. *Celestial Mech. Dynam. Astronom.*, 62(3):193–217, 1995.

[61] V. F. Lazutkin. Existence of a continuum of closed invariant curves for a convex billiard. *Uspehi Mat. Nauk*, 27(3(165)):201–202, 1972.

[62] V. F. Lazutkin. Existence of caustics for the billiard problem in a convex domain. *Izv. Akad. Nauk SSSR Ser. Mat.*, 37:186–216, 1973.

[63] M. L. Lidov and S. L. Ziglin. Non-restricted double-averaged three body problem in Hill's case. *Celestial Mech.*, 13(4):471–489, 1976.

[64] B. B. Lieberman. Quasi-periodic solutions of Hamiltonian systems. *J. Differential Equations*, 11:109–137, 1972.

[65] F. Malige, P. Robutel, and J. Laskar. Partial reduction in the n-body planetary problem using the angular momentum integral. *Celestial Mechanics and Dynamical Astronomy*, 84:283–316, 2002. 10.1023/A:1020392219443.

[66] S. Marmi, P. Moussa, and J.-C. Yoccoz. Linearization of generalized interval exchange maps. *Ann. of Math. (2)*, 176(3):1583–1646, 2012.

[67] J. E. Massetti. *Quasi-periodicity and quasi-conservativity.* PhD thesis, Observatoire de Paris, 2015.

[68] K.R. Meyer and G.R. Hall. *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem.* Number vol. 90 in Applied Mathematical Sciences. Springer-Verlag, 1992.

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

[70] J. Moser. A rapidly convergent iteration method and non-linear differential equations. II. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:499–535, 1966.

[71] J. Moser. A rapidly convergent iteration method and non-linear partial differential equations. I. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:265–315, 1966.

[72] J. Moser. Convergent series expansions for quasi-periodic motions. *Math. Ann.*, 169:136–176, 1967.

[73] R. Narasimhan. *Several complex variables.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1995. Reprint of the 1971 original.

[74] I. O. Parasyuk. Conservation of quasiperiodic motions of reversible multifrequency systems. *Dokl. Akad. Nauk Ukrain. SSR Ser. A*, 9:19–22, 85, 1982.

[75] I. O. Parasyuk. Preservation of multidimensional invariant tori of Hamiltonian systems. *Ukrain. Mat. Zh.*, 36(4):467–473, 1984.

[76] R. Pérez-Marco. KAM techniques in PDE. *Astérisque*, 290:Exp. No. 908, ix, 307–317, 2003. Séminaire Bourbaki. Vol. 2001/2002.

[77] H. Poincaré. *Les méthodes nouvelles de la mécanique céleste*. Gauthier-Villars, 1892.

[78] G. Popov. KAM theorem for Gevrey Hamiltonians. *Ergodic Theory Dynam. Systems*, 24(5):1753–1786, 2004.

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

[80] J. Pöschel. KAM à la R. *Regul. Chaotic Dyn.*, 16(1-2):17–23, 2011.

[81] A. S. Pyartli. Diophantine approximations of submanifolds of a Euclidean space. *Funkcional. Anal. i Priložen.*, 3(4):59–62, 1969.

[82] H. Rüssmann. On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In *Dynamical systems, theory and applications*, pages 598–624. Lecture Notes in Phys., Vol. 38. Springer, 1975.

[83] H. Rüssmann. On the existence of invariant curves of twist mappings of an annulus. In *Geometric dynamics (Rio de Janeiro, 1981)*, volume 1007 of *Lecture Notes in Math.*, pages 677–718. Springer, Berlin, 1983.

[84] H. Rüssmann. Nondegeneracy in the perturbation theory of integrable dynamical systems. In *Number theory and dynamical systems (York, 1987)*, volume 134 of *London Math. Soc. Lecture Note Ser.*, pages 5–18. Cambridge Univ. Press, Cambridge, 1989.

[85] H. Rüssmann. Nondegeneracy in the perturbation theory of integrable dynamical systems. In *Stochastics, algebra and analysis in classical and quantum dynamics (Marseille, 1988)*, volume 59 of *Math. Appl.*, pages 211–223. Kluwer Acad. Publ., Dordrecht, 1990.

[86] H. Rüssmann. On the frequencies of quasi-periodic solutions of analytic nearly integrable Hamiltonian systems. In *Seminar on Dynamical Systems (St. Petersburg, 1991)*, volume 12 of *Progr. Nonlinear Differential Equations Appl.*, pages 160–183. Birkhäuser, Basel, 1994.

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

[88] F. Sergeraert. Un théorème de fonctions implicites sur certains espaces de Fréchet et quelques applications. *Ann. Sci. École Norm. Sup. (4)*, 5:599–660, 1972.

[89] M. B. Sevryuk. The lack-of-parameters problem in the KAM theory revisited. In *Hamiltonian systems with three or more degrees of freedom (S'Agaró, 1995)*, volume 533 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 568–572. Kluwer Acad. Publ., Dordrecht, 1999.

[90] M. B. Sevryuk. The classical KAM theory at the dawn of the twenty-first century. *Mosc. Math. J.*, 3(3):1113–1144, 1201–1202, 2003. Dedicated to V. I. Arnold on the occasion of his 65th birthday.

[91] M. B. Sevryuk. The reversible context 2 in kam theory: the first steps. *Regular and Chaotic Dynamics*, 16(1):24–38, 2010.

[92] M. B.. Sevryuk. *Celestial Mechanics*, chapter Classical Hamiltonian perturbation theory. Encyclopedia of life support systems. Unesco-EOLSS, 2015.

[93] C. L. Siegel and J. K. Moser. *Lectures on celestial mechanics.* Classics in Mathematics. Springer-Verlag, Berlin, 1995. Translated from the German by C. I. Kalme, Reprint of the 1971 translation.

[94] V. G. Sprindžuk. Achievements and problems of the theory of Diophantine approximations. *Uspekhi Mat. Nauk*, 35(4(214)):3–68, 248, 1980.

[95] E. M. Stein. *Singular integrals and differentiability properties of functions.* Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[96] S. Sternberg. *Celestial mechanics.* Number v. 1 in Mathematics lecture note series. W.A. Benjamin, 1969.

[97] F. Tisserand. *Traité de mécanique céleste.* Gauthier-Villars, 1896.

[98] F. Wagener. A parametrised version of Moser's modifying terms theorem. *Discrete Contin. Dyn. Syst. Ser. S*, 3(4):719–768, 2010.

[99] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Trans. Amer. Math. Soc.*, 36(1):63–89, 1934.

[100] J.-C. Yoccoz. Travaux de Herman sur les tores invariants. *Astérisque*, 206:Exp. No. 754, 4, 311–344, 1992. Séminaire Bourbaki, Vol. 1991/92.

[101] E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. I. *Comm. Pure Appl. Math.*, 28:91–140, 1975.

[102] E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. II. *Comm. Pure Appl. Math.*, 29(1):49–111, 1976.

[103] L. Zhao. Partial reduction and Delaunay/Deprit variables. *Celestial Mech. Dynam. Astronom.*, 120:423–432, December 2014.

[104] L. Zhao. Quasi-periodic almost-collision orbits in the spatial three-body problem. *Com. Pure App. Math.*, 2014.

[105] L. Zhao. Quasi-periodic solutions of the spatial lunar three-body problem. *Celestial Mech. Dynam. Astronom.*, 119(1):91–118, 2014.