A SIMPLE PROOF OF INVARIANT TORI THEOREMS

JACQUES FÉJOZ

Abstract. If $\mathbb{T}^n \times \{0\}$ is a Diophantine invariant torus of a real analytic Hamiltonian $K^o : \{(\theta, r)\} \subset \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$, any real analytic Hamiltonians H close to K^o has a unique normal form $H(\theta, r) = K \circ G(\theta, r) + \beta \cdot r$, where $\mathbb{T}^n \times \{0\}$ is an invariant torus of some Hamiltonian K with the same frequency as K^o, G is a Hamiltonian diffeomorphism of a particular type and $\beta \in \mathbb{R}^n$ is a frequency offset. The existence and uniqueness of this normal form, which we call a *twisted conjugacy*, is proved here with a simple abstract inverse function theorem, relying on the Newton algorithm. The normal form is then shown to be a gateway to celebrated invariant tori theorems of Kolmogorov, Arnold, Rüssmann and Herman.

Contents

Date: July 2010.

1. Introduction

Let $\mathcal{H}$ be the space of germs along $\mathrm{T}^n_0 := \mathbb{T}^n \times \{0\}$ of real analytic Hamiltonians in $\mathbb{T}^n \times \mathbb{R}^n = \{(\theta, r)\}$ ($\mathbb{T}^n = \mathbb{R}^n / 2\pi\mathbb{Z}^n$). The vector field associated with $H \in \mathcal{H}$ is

$$\vec{H} : \quad \dot\theta = \partial_r H, \quad \dot r = -\partial_\theta H.$$

For $\alpha \in \mathbb{R}^n$, let $\mathcal{K}$ be the affine subspace of Hamiltonians $K \in \mathcal{H}$ such that $K|_{\mathrm{T}^n_0}$ is constant (i.e. T^n_0 is invariant) and $\vec{K}|_{\mathrm{T}^n_0} = \alpha$. Those Hamiltonians are characterized by their first order expansion along T^n_0, of the form $c + \alpha \cdot r$ for some $c \in \mathbb{R}$, that is, their expansion is constant with respect to θ and the coefficient of r is α. The germ space $\mathcal{H}$ is endowed with the usual, inductive limit topology [16, Section 6.3, example (3)]; see section 2.

We will focus on Diophantine frequency vectors:

$$\mathrm{D}_{\gamma,\tau} = \{\alpha \in \mathbb{R}^n,\ \forall k \in \mathbb{Z}^n \setminus \{0\}\ |k \cdot \alpha| \geq \gamma |k|^{-\tau}\}, \quad |k| := |k_1| + \cdots + |k_n|.$$

If $\tau > n - 1$, the set $\cup_{\gamma>0} \mathrm{D}_{\gamma,\tau}$ has full Lebesgue measure [3, p. 83], [40]. If additionally $\gamma \ll 1$, $\mathrm{D}_{\gamma,\tau}$ has positive Lebesgue measure, which we will assume in the sequel. Also, see appendix D.

Let also $\mathcal{G}$ be the space of germs along T^n_0 of real analytic exact symplectomorphisms G in $\mathbb{T}^n \times \mathbb{R}^n$ of the following form:

$$G(\theta, r) = (\varphi(\theta), {}^t\varphi'(\theta)^{-1}(r + \rho(\theta))),$$

where φ is an isomorphism of $\mathbb{T}^n$ fixing the origin, and ρ is an exact 1-form on $\mathbb{T}^n$.

Theorem 1 (Twisted conjugacy)**.** *Let $\alpha \in \mathrm{D}_{\gamma,\tau}$ and $K^o \in \mathcal{K}$. For every $H \in \mathcal{H}$ close enough to K^o, there exists a unique $(K, G, \beta) \in \mathcal{K} \times \mathcal{G} \times \mathbb{R}^n$ close to $(K^o, \mathrm{id}, 0)$ such that*

$$H = K \circ G + \beta \cdot r$$

in some neighborhood of T^n_0 which is locally uniform with respect to H.

The frequency being a conjugacy invariant of quasi-periodic flows, the frequency offset $\beta \cdot r$ 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. This normal form thus is of geometrical nature and we will call it a *twisted conjugacy.* Advertised by Herman in the 1990's [22], it is the Hamiltonian analogue of the normal form of vector fields on the torus of Arnold and Moser [1, 31]. It appears in a more general setting in Moser [32]. Its formal analogue actually goes back to Poincaré's proof of existence of Lindstedt series [35, Vol. II, § 126].

Several facts prove or hint that H is generally not of the form $(K + \beta \cdot r) \circ G$: having an invariant torus is not an open property, i.e. the operator $(K, G, \beta) \mapsto (K + \beta \cdot r) \circ G$ is not open, although it has the same invertible derivative as $(K, G, \beta) \mapsto K \circ G + \beta \cdot r$ at $(K^o, \mathrm{id}, 0)$; the corresponding linearized equation

$$\delta H = (\delta K + \delta\beta \cdot r) \circ G + (K' + \beta \cdot dr) \circ G \cdot \delta G$$

(dots emphasizing linear operations) has formal or analytic obstructions as soon as $\alpha+\beta$ is a resonant or Liouville vector; more generally, being invertible is usually not an open property among linear maps of topological vector spaces.

Yet, in Dynamics the goal is to show that the parameter $\beta \in \mathbb{R}^n$ vanishes under adequate hypotheses. The initial conjugacy problem has thus been reduced to a problem of finite dimension. Poincaré, in his case, could use an argument of exactness from symplectic geometry; see also Eliasson [17]. The use of rotating frames in [35, Chap. III] and [19, Proposition 82] in order to break some degeneracies relies on a similar idea in a more specific setting. In sections 4 and 6, we will use respectively the usual implicit function theorem in finite dimension and a result from Arnold-Pyartli in the theory of Diophantine approximations on manifolds, making theorem 2 the common gateway to several celebrated theorems.

However, when the frequency α is varied, a key point is that β depends Whitney-smoothly on $\alpha \in \mathrm{D}_{\gamma,\tau}$. This fact was noticed by Lazutkin for twist maps [27, 28]. Pöschel in finite differentiability, Herman, Rüssmann, Broer-Huitema-Takens with their unfolding theory, and Sevryuk, considerably clarified the whole strategy [9, 22, 36, 41, 42, 45, 47].

Eventually, the seeming detour through twisted conjugacies splits the proof of invariant tori theorems into a functionally well posed inversion problem in infinite dimension, and an argument in finite dimension depending on the non-degeneracy hypothesis (compare with [53, 54]; see also [51]). Moreover, the functional setting chosen here adapts to limit degeneracies (lower dimensional tori), in a straightforward manner [19].

In sections 2 and 3, the main theorem of the paper is proved (theorem 2, a more precise version of theorem 1), by applying the inverse function theorem 17. Theorem 17 itself, a very simple version of the Nash-Moser theorem, can be applied to a variety of operators involving small denominators or maps compositions; it is proved in appendix A using the most standard Newton algorithm. Sections 4 and 6 are devoted to inducing invariant tori theorems of Kolmogorov, Arnold, Rüssmann and Herman. Section 5 introduces an intermediate step between twisted and true conjugacies, which we call a *hypothetical conjugacy*. Appendices B and C allow to improve the quantitative bounds of theorems 2 and 17: with elementary estimates on the inverse of real analytic isomorphisms of the phase space, and with Hadamard-like, interpolation inequalities, which do not seem much spread in the literature. Appendix D weakens the arithmetics conditions. Appendix E treats of quasi-periodic time-dependent perturbations. Some further comments are in appendix F.

2. Functional setting

In this section we will define appropriate source and target spaces for the operator

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

to be defined and, when α is Diophantine, invertible.

For various sets U and V, $\mathcal{A}(U,V)$ will denote the set of continuous maps $U \to V$ which are real analytic on the interior $\mathring{U}$, and $\mathcal{A}(U) := \mathcal{A}(U,\mathbb{C})$.

Recall notations for the abstract torus and its embedding in the phase space:

$$\mathbb{T}^n = \mathbb{R}^n/2\pi\mathbb{Z}^n \quad \text{and} \quad \mathrm{T}^n_0 = \mathbb{T}^n \times \{0\} \subset \mathbb{T}^n \times \mathbb{R}^n.$$

Define complex extensions

$$\mathbb{T}^n_{\mathbb{C}} = \mathbb{C}^n/2\pi\mathbb{Z}^n \quad \text{and} \quad \mathrm{T}^n_{\mathbb{C}} = \mathbb{T}^n_{\mathbb{C}} \times \mathbb{C}^n$$

as well as bases of neighborhoods

$$\mathbb{T}^n_s = \{\theta \in \mathbb{T}^n_{\mathbb{C}},\ \max_{1\le j\le n} |\mathrm{Im}\,\theta_j| \le s\} \quad \text{and} \quad \mathrm{T}^n_s = \{(\theta,r) \in \mathrm{T}^n_{\mathbb{C}},\ |(\theta,r)| \le s\},$$

with $|(\theta,r)| := \max_{1\le j\le n} \max\left(|\mathrm{Im}\,\theta_j|, |r_j|\right)$.

2.1. **Spaces of Hamiltonians.** — Let $\mathcal{H}_s = \mathcal{A}(\mathrm{T}^n_s)$, endowed with the Banach norm

$$|H|_s := \sup_{(\theta,r)\in \mathrm{T}^n_s} |H(\theta,r)|,$$

so that $\mathcal{H}$ be the inductive limit of the spaces $\mathcal{H}_s$.

— For $\alpha \in \mathbb{R}^n$, let $\mathcal{K}^\alpha_s = \mathcal{K}_s$ be the affine subspace consisting of those $K \in \mathcal{H}_s$ such that

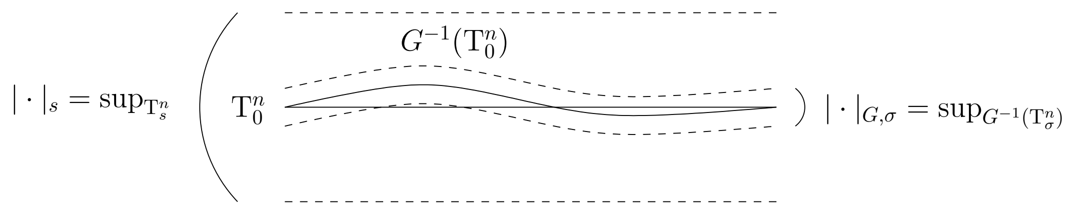
$$K(\theta,r) = c + \alpha\cdot r + O(r^2)$$

for some $c \in \mathbb{R}$.

— If G is a real analytic isomorphism on some open set of $\mathrm{T}^n_{\mathbb{C}}$ and if G is transverse to T^n_s, let $G^*\mathcal{A}(\mathrm{T}^n_s) := \mathcal{A}(G^{-1}(\mathrm{T}^n_s))$ be endowed with the Banach norm

$$|H|_{G,s} := |H \circ G^{-1}|_s.$$

By the principle of analytic continuation, $|\cdot|_{G,s}$ is a Banach norm on $\mathcal{H}_{s-|G-\mathrm{id}|_s}$.



2.2. **Spaces of conjugacies.**

2.2.1. *Diffeomorphisms of the torus.* Let $\mathcal{D}_s$ be the space of maps $\varphi \in \mathcal{A}(\mathbb{T}^n_s, \mathbb{T}^n_{\mathbb{C}})$ which are real analytic isomorphisms from $\mathring{\mathbb{T}}^n_s$ to their image and which fix the origin.

Let also

$$\chi_s := \{v \in \mathcal{A}(\mathbb{T}^n_s)^n,\ v(0) = 0\}$$

be the space of vector fields on $\mathbb{T}^n_s$ which vanish at 0, endowed with the Banach norm

$$|v|_s := \max_{\theta\in\mathbb{T}^n_s} \max_{1\le j\le n} |v_j(\theta)|.$$

According to corollary 24, the map

$$\sigma \mathring{B}^{\chi}_{s+\sigma} := \{v \in \chi_{s+\sigma},\ |v|_s < \sigma\} \to \mathcal{D}_s, \quad v \mapsto \mathrm{id} + v$$

is defined and locally bijective. It endows $\mathcal{D}_s$ with a local structure of Banach manifold in the neighborhood of the identity.

We will consider the contragredient action of $\mathcal{D}_s$ on T^n_s (with values in $\mathrm{T}^n_{\mathbb{C}}$) :

$$\varphi(\theta, r) := (\varphi(\theta), {}^t\varphi'(\theta)^{-1} \cdot r),$$

in order to linearize the dynamics on the alleged invariant tori.

2.2.2. *Straightening tori.* Let $\mathcal{B}_s$ be the space of exact one-forms over $\mathbb{T}^n_s$, with

$$|\rho|_s = \max_{\theta \in \mathbb{T}^n_s} \max_{1 \le j \le n} |\rho_j(\theta)|, \quad \rho = (\rho_1, ..., \rho_n).$$

We will consider its action on T^n_s by translation of the actions:

$$\rho(\theta, r) := (\theta, r + \rho(\theta)),$$

in order to straighten the perturbed invariant tori.

2.2.3. *Our space of conjugacies.* Let $\mathcal{G}_s = \mathcal{D}_s \times \mathcal{B}_s$, identified with a space of Hamiltonian symplectomorphisms by

$$(\varphi, \rho)(\theta, r) := \varphi \circ \rho\,(\theta, r) = (\varphi(\theta), {}^t\varphi'(\theta)^{-1}(r + \rho(\theta))).$$

Endow its tangent space at the identity $T_{\mathrm{id}}\mathcal{G}_s = \mathfrak{g}_s := \chi_s \times \mathcal{B}_s$ with the norm

$$|\dot{G}|_s = |(v, \rho)|_s := \max(|v|_s, |\rho|_s),$$

and its tangent space at $G = (\varphi, \rho)$ with the norm

$$|\delta G|_s := |\delta G \circ G^{-1}|_s, \quad \delta G \in T_G\mathcal{G}.$$

Here and elsewhere, the notation δG, as well as similar ones, should be taken as a whole; there is no separate $\delta \in \mathbb{R}$ in the present paper.

Also consider the following neighborhoods of the identity:

$$\mathcal{G}^{\sigma}_s = \left\{ G \in \mathcal{G}_s,\ \max_{(\theta, r) \in T^n_s} |(\Theta - \theta, R - r)| \le \sigma,\ (\Theta, R) = G(\theta, r) \right\}, \quad \sigma > 0.$$

The operators (commuting with inclusions of source and target spaces)

$$\phi : \mathcal{K}_{s+\sigma} \times \mathcal{G}^{\sigma}_s \times \mathbb{R}^n \to \mathcal{H}_s, \quad (K, G, \beta) \mapsto K \circ G + \beta \cdot r$$

are now defined. In particular, if we let $E_s := \mathcal{K}_s \times \mathcal{G}_s \times \mathbb{R}^n$ and $\sigma B^E_s := \{x \in E_s,\ |x|_s \le \sigma\}$, by restriction ϕ defines the operators

$$\phi : \sigma B^E_{s+\sigma} \to \mathcal{H}_s$$

to which the inverse function theorem of appendix A will apply. Since these operators commute with inclusions, we will speak of them in the singular.

3. Twisted conjugacy

The local existence and uniqueness of the twisted conjugacy (theorem 1) can now be phrased in terms of the operator ϕ.

Theorem 2 (Twisted conjugacy). *Let $\alpha \in \mathrm{D}_{\gamma,\tau}$. The operator ϕ has a unique local inverse in the following sense: for all $0 < s < s+\sigma < 1$, if $K^o \in \mathcal{H}_{s+\sigma}$ and $|H - K^o|_{s+\sigma}$ is small, there is a unique $(K, G, \beta) \in \sigma B_s^E$, $|\cdot|_s$-close to $(K^o, \mathrm{id}, 0)$ such that*

$$H = K \circ G + \beta \cdot r$$

over T^n_s. Moreover $\beta \circ \phi^{-1}$ is a C^1-function locally in the $|\cdot|_{s+\sigma}$-neighborhood of K^o.

Geometrically: the orbits of Hamiltonians $K \in \mathcal{K}$ under the action of symplectomorphisms of $\mathcal{G}$ locally form a subspace of finite codimension n. (The conclusion will also hold in the neighborhood of $(K^o, \mathrm{id}, \beta^o)$ if $\beta^o \in \mathbb{R}^n$.)

The theorem will follow from the inverse function theorem of appendix A applied to ϕ, lemma 20 (for the uniqueness of the inverse) and corollary 22 (for the smoothness of $\beta \circ \phi^{-1}$). Let us now check the two main hypotheses of appendix A, one on ϕ'^{-1} and one on ϕ''.

Let $\mathcal{L}_\alpha$ be the Lie derivative operator in the direction of the constant vector field $\alpha \in \mathrm{D}_{\gamma,\tau}$:

$$\mathcal{L}_\alpha : \mathcal{A}(\mathbb{T}^n_s) \to \mathcal{A}(\mathbb{T}^n_s), \quad f \mapsto f' \cdot \alpha = \sum_{1 \le j \le n} \alpha_j \frac{\partial f}{\partial \theta_j}.$$

We will need the following classical lemma in two instances in the proof of lemma 4.

Lemma 3 (Cohomological equation). *If $g \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})$ has 0-average ($\int_{\mathbb{T}} g \, d\theta = 0$), there exists a unique function $f \in \mathcal{A}(\mathbb{T}^n_s)$ of 0-average such that $\mathcal{L}_\alpha f = g$, and there exists $C_0 = C_0(n, \tau) > 0$ such that, for any σ:*

$$|f|_s \le C_0 \gamma^{-1} \sigma^{-\tau-n} |g|_{s+\sigma}.$$

Proof. Let $g(\theta) = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} g_k \, e^{ik\cdot\theta}$ be the Fourier expansion of g. The unique formal solution to the equation $\mathcal{L}_\alpha f = g$ is given by $f(\theta) = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} \frac{g_k}{i\,k\cdot\alpha} \, e^{i\,k\cdot\theta}$.

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 $\operatorname{Im} \theta_j = \pm(s+\sigma)$.

Using this estimate and replacing the small denominators $k \cdot \alpha$ by the estimate defining the Diophantine property of α, 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 the latter sum is bounded by

$$\begin{aligned}\int_1^\infty (\ell+n-1)^{\tau+n-1}e^{-(\ell-1)\sigma}\,d\ell &= \sigma^{-\tau-n}e^{n\sigma}\int_{n\sigma}^\infty \ell^{\tau+n-1}e^{-\ell}\,d\ell \\ &< \sigma^{-\tau-n}e^{n\sigma}\int_0^\infty \ell^{\tau+n-1}e^{-\ell}\,d\ell = \sigma^{-\tau-n}e^{n\sigma}\Gamma(\tau+n).\end{aligned}$$

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

Fix $0<s<s+\sigma<1$. We will write $x=(K,G,\beta)$ and $\delta x=(\delta K,\delta G,\delta\beta)$.

Lemma 4. *There exists $C'>0$ which is locally uniform with respect to $x\in E_{s+\sigma}$ in the neighborhood of $G=\mathrm{id}$ such that the linear map $\phi'(x)$ has an inverse $\phi'(x)^{-1}$ satisfying*

$$\left|\phi'(x)^{-1}\cdot\delta H\right|_s \le \sigma^{-\tau-n-1}C'\,|\delta H|_{G,s+\sigma}\,.$$

Proof. A function $\delta H\in G^*\mathcal{A}(T_{s+\sigma})$ being given, we want to solve the equation

$$\phi'(x)\cdot\delta x=\delta K\circ G+K'\circ G\cdot\delta G+\delta\beta\cdot r=\delta H,$$

for the unknowns $\delta K\in T_K\mathcal{K}_s\subset\mathcal{A}(\mathrm{T}^n_s)$, $\delta G\in T_G\mathcal{G}_s$, and $\delta\beta\in\mathbb{R}^n$, or, equivalently, after composing with G^{-1} to the right,

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

where we have set $\delta K=\delta c+\dot K$ with $\delta c\in\mathbb{R}$ and $\dot K=O(r^2)$, $\dot G:=\delta G\circ G^{-1}\in\mathfrak{g}_s$ and $\dot H:=\delta H\circ G^{-1}\in\mathcal{H}_{s+\sigma}$.

More specifically, G^{-1} and $\dot G$ are of the form

$$G^{-1}(\theta,r)=(\varphi^{-1}(\theta),{}^t\varphi'\circ\varphi^{-1}(\theta)\cdot r-\rho\circ\varphi^{-1}(\theta)),\quad \dot G=(\dot\varphi,\dot\rho-r\cdot\dot\varphi'),$$

where $\dot\varphi\in\chi_{s+\sigma}$ and $\dot\rho\in\mathcal{B}_{s+\sigma}$, and we can expand

$$K=c+\alpha\cdot r+K_2(\theta)\cdot r^2+O(r^3)\quad\text{and}\quad \dot H=\dot H_0(\theta)+\dot H_1(\theta)\cdot r+O(r^2).$$

The equation becomes

$$\begin{aligned}(1)\quad \left[\dot\rho\cdot\alpha+\delta c-\rho\circ\varphi^{-1}\cdot\delta\beta\right]+r\cdot\left[-\dot\varphi'\cdot\alpha+\varphi'\circ\varphi^{-1}\cdot\delta\beta+2K_2\cdot\dot\rho\right]+ \\ \dot K=\dot H+O(r^2),\end{aligned}$$

where the term $O(r^2)$ in the right hand side depends only on K and $\dot G$, and not on $\dot K$. The equation turns out to be triangular in the five unknowns. The existence and uniqueness of a solution with the wanted estimate follows from repeated applications of lemma 3 and Cauchy's inequality:

— The average over T^n_0 of the first order terms with respect to r in equation (1) yields

$$\delta\beta=\left(\int_{\mathbb{T}^n}\varphi'\circ\varphi^{-1}\,d\theta\right)^{-1}\cdot\int_{\mathrm{T}^n_0}\dot H_1\,d\theta,$$

which does exist if φ is close to the identity (proposition 24).

— Similarly, the average of the restriction to T_0^n of (1) yields:

$$\delta c = \int_{\mathrm{T}_0^n} \dot{H}_0 \, d\theta + \int_{\mathrm{T}_0^n} \rho \circ \varphi^{-1} \, d\theta \cdot \delta\beta.$$

— Next, the restriction to T_0^n of (1) can be solved uniquely with respect to $\dot{\rho}$ according to lemma 3 (applied with $\dot{\rho} = f'$).

— The part of degree one can then be solved for $\dot{\varphi}$ similarly.

— Terms of order ≥ 2 in r determine $\dot{K}$. □

Lemma 5. *There exists a constant $C'' > 0$ which is locally uniform with respect to $x \in E_{s+\sigma}$ in the neighborhood of $G = \mathrm{id}$ such that the bilinear map $\phi''(x)$ satisfies*

$$\left|\phi''(x) \cdot \delta x^{\otimes 2}\right|_{G,s} \leq \sigma^{-2} C'' \, |\delta x|_{s+\sigma}^2.$$

Proof. Differentiating ϕ twice yields

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

hence

$$\left(\phi''(x) \cdot \delta x^{\otimes 2}\right) \circ G^{-1} = 2\,\delta K' \cdot (\delta G \circ G^{-1}) + K'' \cdot (\delta G \circ G^{-1})^{\otimes 2},$$

whence the estimate. □

Exercise 6 (Arnold-Moser normal form) Use theorem 2 to show that, for every vector field $v \in \chi(\mathbb{T}^n)$ close to $\alpha \in \mathrm{D}_{\gamma,\tau}$, there is a unique $\varphi \in \mathcal{D}$ and a unique $\beta \in \mathbb{R}^n$ such that $v = \varphi_* \alpha + \beta$. Hint: apply theorem 2 to the Hamiltonian $v(\theta) \cdot r$; cf. [19, Section 4.3].

4. Strong non-degeneracy

Theorem 7 (Kolmogorov [12, 26])**.** *Let $\alpha \in \mathrm{D}_{\gamma,\tau}$ and $K^o \in \mathcal{K}$ such that the averaged hessian $\int_{\mathbb{T}^n} \frac{\partial^2 K^o}{\partial r^2}(\theta, 0)\, d\theta$ is non degenerate. For every $H \in \mathcal{H}$ close to K^o, there is a unique $(K, G, R) \in \mathcal{K} \times \mathcal{G} \times \mathbb{R}^n$ close to $(K^o, \mathrm{id}, 0)$ such that*

$$H(\theta, r + R) = K \circ G(\theta, r)$$

in a neighborhood of T_0^n which is locally uniform with respect to H; in particular, H possesses an α-quasi-periodic invariant torus.

This theorem has far reaching consequences ; see [7, 8, 11, 14, 34, 37, 46, 48] for references and background. In particular it has led to a partial answer to the long standing question of the stability of the Solar system [4, 10, 19].

Proof. Let $K_2^o(\theta) := \frac{1}{2}\frac{\partial^2 K^o}{\partial r^2}(\theta, 0)$. Let F be the analytic function taking values among symmetric bilinear forms, which solves the cohomological equation

$$\mathcal{L}_\alpha F(\theta, 0) = \frac{1}{2}\frac{\partial^2 K^o}{\partial r^2}(\theta, 0) - \int_{\mathrm{T}_0^n} \frac{1}{2}\frac{\partial^2 K^o}{\partial r^2}(\theta, 0)\, d\theta$$

on T_0^n (see lemma 3), and φ be the germ along T_0^n of the (well defined) time-one map of the flow of the Hamiltonian $F(\theta) \cdot r^2$. The map φ is symplectic and restricts to the identity on T_0^n. At the expense of substituting $K^o \circ \varphi$ and $H \circ \varphi$ for K^o and H respectively, one can thus assume that

$$K^o = c + \alpha \cdot r + Q \cdot r^2 + O(r^3), \quad Q := \int_{\mathrm{T}_0^n} \frac{1}{2} \frac{\partial^2 K^o}{\partial r^2}(\theta, 0)\, d\theta.$$

The germs so obtained are close to one another.

Consider the family of 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^n from its terms $O(R^2)$ which possibly depend on θ:

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

For the Hamiltonian $\hat{K}^o_R$, T_0^n is invariant and quasi-periodic of frequency $\alpha + 2Q \cdot R$. Hence the normal form of $\hat{K}^o_R$ with respect to the frequency α is

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

By hypothesis the matrix $\left.\frac{\partial \hat{\beta}^o}{\partial R}\right|_{R=0} = 2Q$ is invertible, so the map $R \mapsto \hat{\beta}^o(R)$ is a local diffeomorphism.

Now, theorem 1 asserts the existence of an analogous map $R \to \beta(R)$ for H_R, which is a small C^1-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 $G^{-1}(\mathrm{T}_0^n)$ by R in the direction of actions is invariant and α-quasi-periodic for H. □

Exercise 8 Simplify this proof when $K^o = K^o(r)$ is integrable.

Remark 9 (L. Chierchia) Assume the hypotheses of theorem 7 are verified. Applying the theorem to each Hamiltonian $H - \beta \cdot r$, $\beta \in \mathbb{R}^n$, $\|\beta\| \ll 1$, shows that there is a map $\beta \mapsto (K_\beta, G_\beta, R_\beta)$ such that

$$H(\theta, r + R_\beta) - \beta \cdot (r + R_\beta) \equiv K_\beta \circ G_\beta(\theta, r),$$

and $\beta \mapsto R_\beta$ is a local diffeomorphism. Hence there is a unique β such that $R_\beta = 0$:

$$H = K_\beta \circ G_\beta + \beta \cdot r$$

i.e., the twisted conjugacy follows from Kolmogorov's theorem.

5. Hypothetical conjugacy

We now move to a "hypothetical" normal form, the common ground of invariant tori theorems of section 6.

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 T^n_s for which T^n_0 is invariant and quasi-periodic, with unprescribed frequency.

Theorem 10 (Hypothetical conjugacy)**.** *For every $K^o \in \mathcal{K}^{\alpha^o}_{s+\sigma}$ with $\alpha^o \in D_{\gamma,\tau}$, there is a germ of diffeomorphism*

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

at $K^o \mapsto (K^o, \mathrm{id})$ such that for every H with $\alpha_H \in \mathrm{D}_{\gamma,\tau}$,

$$H = K_H \circ G_H$$

and K_H and G_H are unique.

The pair (K_H, G_H) can rightfully be called a *hypothetical conjugacy* of H because the property $H = K_H \circ G_H$ depends on an arithmetical condition involving the unknown frequency α_H.

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

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

locally in the neighborhood of (α^o, K^o), $K^o \in \mathcal{K}^{\alpha^o}_{s+\sigma}$. Since ϕ is infinitely differentiable, by proposition 23 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 $\Theta(H) = \hat{\Theta}(\hat{\alpha}, H)$. □

One can infer Kolmogorov's theorem 7, or the following variation, from theorem 10.

Corollary 11 (Arnold [2], Pöschel [36])**.** *Let $K^o = K^o(r)$ be a germ of completely integrable Hamiltonian. If H is close to K^o, there are C^∞-germs G, c and α along T^n_0 of diffeomorphism, function and frequency, close to* id*, K^o and $K^{o\prime}$, such that whenever $\alpha_R \in \mathrm{D}_{\gamma,\tau}$, the infinite jet along $G^{-1}(\mathrm{T}^n_0 + (0, R))$ of $H \circ G^{-1}$ is*

$$j^\infty_{G^{-1}(\mathrm{T}^n_0 + (0,R))} H \circ G^{-1} = c_R + \alpha_R \cdot (r - R) + O((r - R)^2).$$

Moreover, if the hessian $\partial^2_r K^o(0)$ is non degenerate, the set $\{R,\ \alpha_R \in \mathrm{D}_{\gamma,\tau}$ has positive Lebesgue measure.

Proof. As in the proof of Kolmogorov's theorem, let

$$K_R^0(\theta,R)=K^o(\theta,R+r)\quad\text{and}\quad H_R(\theta,r)=H(\theta,R+r).$$

According to theorem 10, if R is small, there exist $(K_R,G_R)\in\mathcal{K}_s\times\mathcal{G}_s$ for some $s>0$ such that $K_R=c_R+\alpha_R\cdot r+O(r^2)$, the map $R\mapsto\alpha_R$ is C^1-close to $R\mapsto\alpha_R^o:=\partial_r K^o(R)$, and $H_R=K_R\circ G_R$ as soon as $\alpha_R\in\mathrm{D}_{\gamma,\tau}$, i.e.

$$H(\theta,r)=K_R\circ G_R(\theta,r-R).$$

(Beware that $K_R(\theta,r)\neq K(\theta,r+R)$ in general.)

In order to glue the constructed K_R and G_R together, define the germs K of C^∞-Hamiltonians and G of C^∞-map by

$$K(\theta,r)=K_r(\theta,0)\quad\text{and}\quad G(\theta,r)=G_r(\theta,0)+(0,r)=(G_\theta(\theta,r),G_r(\theta,r)).$$

As soon as $\alpha_r\in\mathrm{D}_{\gamma,\tau}$, the torus $\mathrm{T}_0^n+(0,r)$ is K-invariant and quasi-periodic with frequency α_r. Besides, if H is close to K^o, $G_r(\theta,0)$ is C^∞-close to $(\theta,0)$ and thus the map G is a germ of diffeomorphism.

From the trivial the equality

$$K_R\circ G_R(\theta,r-R)=K_R(G_R(\theta,r-R)+(0,R)-(0,R)),$$

we see that the following equality

$$H(\theta,r)=K(G_\theta(\theta,r),R)$$

holds between infinite jets along $G_R^{-1}(\mathrm{T}_0^n)+(0,R)=G^{-1}(\mathrm{T}_0^n+(0,R))$, as soon as $\alpha_r:=\partial_r K(\theta,r)\in\mathrm{D}_{\gamma,\tau}$, whence the first assertion.

If the hessian $\partial_r^2K^o(0)$ is non degenerate, the frequency map $r\mapsto\alpha_r$ is a local diffeomorphism and the preimage by α of the set $\mathrm{D}_{\gamma,\tau}$ thus has positive Lebesgue measure. □

6. Weak non-degeneracy

In theorem 7, the frequency is fixed. We will now deal with more degenerate cases, where the map from actions or more general types of parameters, to the frequencies, might not be a local diffeomorphism. As opposed to the strongly non degenerate case, we will not be able to follow quasi-periodic invariant tori individually.

Assume that the perturbed Hamiltonian $H=H_t$ depends smoothly on some parameter $t\in\mathbb{B}^T$ ($\mathbb{B}^T=$ the closed unit ball of $\mathbb{R}^T$); if H is close to some completely integrable Hamiltonian, t may be the action coordinate r and, in the case of Arnold's theorem, t represents the semi major axes.

Definition 12 A smooth frequency map $\alpha:\mathbb{B}^T\to\mathbb{R}^n$ is *weakly non degenerate* if its local image nowhere lies in a proper vector subspace of $\mathbb{R}^n$.

That this weak non-degeneracy property is relevant in general averaging theory, was discovered by Arnold [5].

Theorem 13 (Rüssmann). *Suppose the family $K^o_t \in \mathcal{K}_{s+\sigma}$, $t \in \mathbb{B}^T$, has a non-planar frequency map $t \mapsto \alpha^o_t$. If γ and $1/\tau$ are small and if H_t is close to K^o_t for all $t \in \mathbb{B}^T$, there exist $\alpha_t \in \mathbb{R}^n$, $K_t \in \mathcal{K}^{\alpha_t}_s$, $G_t \in \mathcal{G}_s$ (all varying smoothly with t) and a subset $\mathfrak{D}_{\gamma,\tau} \subset \mathbb{B}^T$ of positive Lebesgue measure over which*

$$H_t = K_t \circ G_t.$$

Theorem 13 follows from theorems 10 above and 14 below, with $\mathfrak{D}_{\gamma,\tau} := \{t \in \mathbb{B}^T, \alpha_t \in \mathrm{D}_{\gamma,\tau}\}$ and the remark that being weakly-non degenerate is an open property among C^{n-1}-maps (see [19, Lemma 44]).

Theorem 14 (Pyartli [39]). *If $t \in \mathbb{B}^T \mapsto \alpha_t \in \mathbb{R}^n$ is non planar, the Lebesgue measure of $\mathfrak{D}_{\gamma,\tau}$ is positive provided that γ is small enough and τ large enough.*

Remark 15 Theorem 13 provides a convenient setting to check the persistence of invariant tori for parameters varying in any submanifold of $\mathbb{B}^T$. For instance, in celestial mechanics it is often crucial to prove the persistence of invariant tori on some given energy level [18, Section 3.3]. In order to get such an “iso-energetic” statement, it suffices to check that the restriction of the frequency map α^o to the submanifold of $\mathbb{B}^T$ of equation $K^o_t(\theta, 0) = \text{cst}$ is weakly non degenerate.

Exercise 16 (Herman's stability theorem [19]) Let

$$H(\theta, r, z) = H_0(r) + \epsilon \frac{1}{2} Q(r) \cdot z^{\otimes 2} + O(z^{\otimes 3}; r, \epsilon) + \epsilon H_1(\theta, r, z)$$

be a germ along $\mathbb{T}^n \times \{0\} \times \{0\}$ of real analytic Hamiltonian in $\{(\theta, r, z)\} = \mathbb{T}^n \times \mathbb{R}^n \times \mathbb{C}^p$, where

$$Q(r) \cdot z^{\otimes 2} = \sum_{1 \le j \le p} \hat{\alpha}^o_j(r) \mathrm{Re}\,(z_j \bar{z}_j), \quad \hat{\alpha}^o_j(r) \ge 0,$$

and

$$\int_{\mathbb{T}^n} H_1(\theta, r, z)\, d\theta = 0.$$

Assume that the frequency map

$$r \mapsto (\alpha^o(r), \hat{\alpha}^o(r)) \in \mathbb{R}^{n+p}, \quad \alpha(r) := \partial_r H_0(r),$$

is weakly non degenerate. Then there exists $\epsilon > 0$ such that H has an invariant set of positive Lebesgue measure, consisting of Diophantine invariant tori. Hint: eliminate fast angles θ from H_1, then build a Birkhoff normal form to eliminate slow angles $\operatorname{Arg} z_j$, switch to symplectic polar coordinates in the z-direction, and apply Rüssmann's theorem.

The proof of Arnold's theorem on the stability of the planetary system reduces to a similar, quantitative stability result, where H_0 is the Hamiltonian describing the Keplerian motion of non-interacting planets, and Q is the quadratic part of the secular Hamiltonian, describing the slow deformations of the Keplerian ellipses at the first order in the eccentricities and inclinations, ϵ is the small order of masses of planets, r is a function of the major semi-axes, θ is the mean longitudes, and z the secular coordinates.

A. An inverse function 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 $\phi : \sigma B_{s+\sigma}^E \to F_s$, $s < s+\sigma$, $\phi(0) = 0$, 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 such that, if $F_{x,s}$ denotes the normed vector space $(F_s, |\cdot|_{x,s})$,

$$x \in sB_{s+\sigma}^E \Rightarrow \phi(x) \in F_{x,s}.$$

In other words, ϕ is a section of the trivial Banach vector bundle of base $sB_{s+\sigma}^E$ and fiber $F_{x,s}$ over x. 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 one. 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.

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

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

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

Theorem 17. *ϕ is locally surjective and, more precisely, for any s, η and σ with $\eta < s$,*

$$\epsilon B_{s+\sigma}^F \subset \phi\left(\eta B_s^E\right), \quad \epsilon := 2^{-8\tau}C^{-2}\sigma^{2\tau}\eta.$$

In other words, ϕ has a right-inverse $\psi : \epsilon B_{s+\sigma}^F \to \eta B_s^E$.

Proof. Some numbers s, η and σ and $y \in B_{s+\eta}^F$ being given, let

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

and

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

Lemma 18. *The function Q satisfies, for $x, \hat{x} \in sB_{s+2\sigma}^E$:*

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

Proof of the lemma. Let $\hat{x}_t := (1-t)x + t\hat{x}$. Taylor's formula yields

$$Q(x,\hat{x}) = \int_0^1 (1-t)\,\phi''(\hat{x}_t)\,(\hat{x}-x)^2\,dt.$$

Using the asumption on deformed norms,

$$|Q(x,\hat{x})|_{x,s} \le \int_0^1 (1-t)\left|\phi''(\hat{x}_t)(\hat{x}-x)^2\right|_{x,s} dt \le \int_0^1 (1-t)\left|\phi''(\hat{x}_t)(\hat{x}-x)^2\right|_{\hat{x}_t,s+|\hat{x}_t-x|_s} dt,$$

whence the estimate, using the asumption on ϕ''. □

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 $n \ge 0$ and converges towards some preimage $x \in \eta B^E_s$ of y by ϕ.

Let $(\sigma_n)_{n\ge 0}$ be a sequence of positive real numbers such that $3\sum \sigma_n = \sigma$, and $(s_n)_{n\ge 0}$ 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

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

Further assuming that $|x_{k+1}-x_k|_{s_k} \le \sigma_k$, the estimate of the right inverse and lemma 18 entail that

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

The estimate

$$|x_1|_{s_1} \le C'(3\sigma_0)^{-\tau'}|y|_{s_0} \le 2^{-1}C\sigma_0^{-\tau}\epsilon = c_0\epsilon$$

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

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

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

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

Maximizing the upper bound of ϵ under the constraint $3\sum_{n\ge 0}\sigma_n = \sigma$ yields $\sigma_k := \frac{\sigma}{6}2^{-k}$. A posteriori it is straightforward that $|x_{n+1} - x_n|_{s_n} \le \sigma_n$ (as earlier assumed to apply lemma 18) and $c_n \ge 1$ for all $n \ge 0$. Besides, using that $\sum k2^{-k} = \sum 2^{-k} = 2$ we get

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

whence the theorem. □

Exercise 19 The domain of ψ contains ϵB^F_S, $\epsilon = 2^{-12\tau}\tau^{-1}C^{-2}S^{3\tau}$, for any S.

Proof. The above function $\epsilon(\eta,\sigma) = 2^{-8\tau}C^{-2}\sigma^{2\tau}\eta$ attains is maximum with respect to $\eta < s$ for $\eta = s$. Besides, under the constraint $s + \sigma = S$ the function $\epsilon(s,\sigma)$ attains its maximum when $\sigma = 2\tau s$ and $s = \frac{S}{1+2\tau}$. Hence, S being fixed, the domain of ψ contains ϵB^F_S if

$$\epsilon < 2^{-8\tau} C^{-2} \frac{S}{1+2\tau} \left(\frac{2\tau S}{12(1+2\tau)} \right)^{2\tau}.$$

Given that $S < 1 < \tau$ by hypothesis, it suffices that ϵ be equal to the stated value. □

A.1. **Smoothness.** In the proof of theorem 17 we have built right inverses $\psi : \epsilon B^F_{s+\eta+\sigma} \to \eta B^E_{s+\eta}$, of ϕ, commuting with inclusions. The estimate given in the statement 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 2 we have restricted to an adequate class of symplectomorphisms);

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

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

(according to the remark after corollary 26, this estimate is satisfied in the case of interest to us, since $\sigma + \log(1+\sigma/s) \le \tilde\sigma$).

Lemma 20 (Lipschitz regularity)**.** *If $\sigma < s$ and $y, \hat y \in \epsilon B^F_{s+\sigma}$ with $\epsilon = 2^{-14\tau} C^{-3} \sigma^{3\tau}$,*

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

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 17, $x := \psi(y)$ 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

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

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 18,

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

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

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

yields

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

Now, we want to choose η small enough so that

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

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

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

Proposition 21 (Smoothness). *For every $\sigma < s$, there exists ϵ, C_1 such that for every $y, \hat{y} \in \epsilon B^F_{s+\sigma}$,*

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

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, $2 \le k \le \infty$, 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,

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

for some σ' tending to 0 when σ itself tends to 0, and for some $C_1 > 0$ depending on σ. Up 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'. □

Corollary 22. *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 $(\pi \circ \psi)' = \pi \cdot \phi' \circ \psi$.*

This corollary is used with $\pi : (K, G, \beta) \mapsto \beta$ in the proof of theorem 2.

A.2. **Whitney-smoothness with respect to finitely many parameters.** In section 5, it is convenient to extend ϕ^{-1} to non-Diophantine vectors α. Whitney-smoothness is a criterion for such an extension to exist [50, 52].

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

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

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

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

Proposition 23 (Whitney-smoothness)**.** *If s, σ and ϵ are chosen like in proposition 21, 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, $1 \leq k \leq \infty$, 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 20),

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

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 [20]); ψ is defined on a Banach space, but the extension directions are of finite dimension. □

B. Analytic isomorphisms

Here we include an elementary inversion theorem for real analytic isomorphisms on $\mathbb{T}^n_s$. The qualitative part is used in section 2, to parameterize locally $\mathcal{D}_s$ by vector fields, and, in lemma 3, to solve the cohomological equation for the frequency offset $\delta\beta$. The estimates are needed only for an explicit bound in the invariant tori theorems of the paper.

Recall that we have set $\mathbb{T}^n_s := \{\theta \in \mathbb{C}^n/2\pi\mathbb{Z}^n, \quad \max_{1\leq j\leq n} |\mathrm{Im}\,\theta_j| \leq s\}$. We will denote by $p : \mathbb{R}^n_s := \mathbb{R}^n \times i[-s,s]^n \to \mathbb{T}^n_s$ its universal covering.

Proposition 24. *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 $\psi : \mathbb{T}^n_s \to \mathbb{T}^n_{s+\sigma}$:*

$$\begin{tikzcd} \mathbb{T}^n_{s+\sigma} \arrow[r, hook, "\varphi"] & \mathbb{T}^n_{s+2\sigma} \\ & \mathbb{T}^n_s \arrow[u, hook] \arrow[ul, "\psi"] \end{tikzcd} .$$

Furthermore,

$$|\psi - \mathrm{id}|_s \leq |v|_{s+\sigma}$$

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

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

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 $k \in M_n(\mathbb{Z})$, $k(l) := \Phi(x+l) - \Phi(x)$.

(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,

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

and, by Cauchy's inequality,

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

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 is in the sense of $\Psi := \Phi^{-1} : \mathbb{R}^n_s \to \mathbb{R}^n_{s+\sigma}$. If $y \in \mathbb{R}^n_s$,

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

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

(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 $2\sigma^{-1}|v|_{s+2\sigma} \leq 1$,

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

□

C. Interpolation inequalities

In this section we prove some Hadamard interpolation inequalities, which are used in sections A.1 and A.2.

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

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

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

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

Proposition 25. *If $f \in \mathcal{A}(\mathbb{T}^n_{s_1} \times \mathbb{D}^n_{t_1})$,*

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

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 $(\operatorname{Im}\theta, |r|) = cst$, 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 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 $\operatorname{Re} z$ and, due to the relations imposed on the norm indices, if $|\operatorname{Im} z| \le s_1$ then $|e^{-(iz+s)} t| \le e^{s_1 - s} t = t_1$.

The estimate

$$\log \varphi(z) \le \frac{s_1 - \operatorname{Im} z}{s_1 - s_0} \varphi(s_0 i) + \frac{\operatorname{Im} z - s_0}{s_1 - s_0} \varphi(s_1 i)$$

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

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.

Corollary 26. *If $\sigma_1 = -\log\left(1 - \frac{\sigma_0}{s}\right)$ and $f \in \mathcal{A}(\mathrm{T}^n_{s+\sigma_1})$,*

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

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

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

Proof. In proposition 25, 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 $s + \sigma - se^{\sigma}$ 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. □

D. Weaker arithmetic conditions

In this section, we look more carefully to the arithmetic conditions needed for the induction to converge, in the proof of the inverse function theorem 17 applied to the operator ϕ of section 3.

A function $\Delta : \mathbb{N}_* \to [1, +\infty[$ being given, define D_Δ as the subset of vectors $\alpha \in \mathbb{R}^n$ such that
$$|k \cdot \alpha| \geq \frac{(|k| + n - 1)^{n-1}}{\Delta(|k|)} \quad (\forall k \in \mathbb{Z}^n \setminus \{0\}).$$
(The function Δ is just some other normalization of what is an approximation function in [40] or a zone function in [15].) For D_Δ to be non empty, trivially we need $\lim_{+\infty} \Delta = +\infty$.

Proposition 27. *The conclusions of theorems 7 and 2 hold if the Diophantine condition is replaced by the condition that there exist $c > 0$ and $\varsigma \in]0, 1[$ such that*
$$\sum_{\ell \geq 1} \Delta(\ell) e^{-\ell/j^2} \leq \exp\left(c\, 2^{\varsigma j}\right) \quad \textit{as } j \to +\infty.$$

Example 28 The Diophantine set $D_{\gamma,\tau}$ corresponds to a polynomially growing function Δ, and to a polynomially (at most) growing function $\sum_{\ell \geq 1} \Delta(\ell) e^{-\ell\, 2^{-j}}$. A fortiori, $\sum_{\ell \geq 1} \Delta(\ell) e^{-\ell/j^2}$ is less than polynomially growing.

Proof of the proposition. Call L the discrete Laplace transform of Δ:
$$L(\sigma) = \sum_{\ell \geq 1} \Delta(\ell) e^{-\ell\sigma},$$

and assume it is finite for all $\sigma > 0$. Patterning the proof of lemma 3, we get the following generalization.

Lemma 29. *Let $g \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})$ having 0-average. There is a unique function $f \in \mathcal{A}(\mathbb{T}^n_s)$ of zero average such that $\mathcal{L}_\alpha f = g$. This function satisfies*

$$|f|_s \leq C\, L(\sigma)\, |g|_{s+\sigma}, \quad C = \frac{2^n e}{(n-1)!}.$$

(Again, see [40] for improved estimates. But such an improvement is not the crux of our purpose here.)

Taking up the proof of the inverse function theorem of appendix A with our new estimates (see in particular equation (2)), we see that the Newton algorithm converges provided

$$\sum_{j \geq 0} 2^{-j} \log L(\sigma_j) < \infty,$$

for some choice of the converging series $\sum \sigma_j$. Choosing $\sum \sigma_j = \sum j^{-2}$, we see that it is enough that $\log L(\sigma_j) \leq c\, 2^{\varsigma j}$ for some $c > 0$ and $\varsigma \in]0, 1[$, whence the given criterion. ☐

E. Quasi-periodic time-dependent perturbations

As a variation and a second use of the inverse function theorem 17, we briefly treat of Hamiltonians which are quasi-periodically time-dependent. Such Hamiltonians are commonplace in celestial mechanics, e.g. in restricted many-body problems (for a definition of these problems, see [6, Section 2.5]), where the primary bodies have a given quasi-periodic motion, which influences without being influced by, the zero-mass.

Suppose $n = \hat{n} + \check{n}$ with $\hat{n}, \check{n} > 0$. We split variables accordingly: $\alpha = (\hat{\alpha}, \check{\alpha})$, $\theta = (\hat{\theta}, \check{\theta})$, etc. Variables with a caron ˇ are related to the time of the perturbation. Let

$$\bar{\mathcal{H}} = \{H \in \mathcal{H},\ \partial_{\check{r}} H \equiv \check{\alpha}\}.$$

Let $\bar{\mathcal{K}} = \bar{\mathcal{H}} \cap \mathcal{K}$. For Hamiltonians in $\bar{\mathcal{H}}$, the frequency $\dot{\check{\theta}} \equiv \check{\alpha}$ is fixed. Let also $\bar{\mathcal{G}}$ be the subset of $\mathcal{G}$ consisting of symplectomorphisms $G = (\varphi, \rho) \in \mathcal{G}$ such that φ is of the form $\varphi(\theta) = (\hat{\varphi}(\theta), \check{\theta})$, i.e. $\check{\varphi}(\theta) = \check{\theta}$. If $H \in \bar{\mathcal{H}}$ and if $G \in \bar{\mathcal{G}}$ is close enough to the identity for $H \circ G$ to be well defined over $\mathbb{T}^n_s$ for some $s > 0$, then $H \circ G \in \bar{\mathcal{H}}_s$. So, by restriction the operator ϕ (see section 2) defines a map

$$\bar{\phi} : \bar{\mathcal{K}}_{s+\sigma} \times \bar{\mathcal{G}}^\sigma_{s+\sigma} \times \mathbb{R}^{\hat{n}} \to \bar{\mathcal{H}}_s, \quad (K, G, \hat{\beta}) \mapsto K \circ G + \hat{\beta} \cdot \hat{r}.$$

Corollary 30 (Twisted conjugacy for quasi-periodic time-dependent perturbations)**.** *The operator*

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

is a local diffeomorphism.

Proof. We pattern the proof of theorem 2, and additionaly impose that $\delta H \in T_H \bar{\mathcal{H}}$. One only needs to check that the solution δx of equation (1) lies in the tangent space of the source space of ϕ. Indeed, we have

— $\delta\check{\beta} = 0$ because $\check{\dot{H}}_1 = 0$ and $\partial_{\hat{\theta}}\hat{\varphi} = 0$.

— $\check{\dot{\varphi}} = 0$ because $\check{\dot{H}}_1 = 0$ (since $\partial_{\check{r}} H(\theta, 0) \equiv \check{\alpha}$).

— $\partial_{\check{r}} \dot{K} = 0$ because $\dot{H}$ does not either. □

We can now state the following analogue of Kolmogorov's theorem. It does not follow from Kolmogorov's theorem directly because the unperturbed Hamiltonian is degenerate in the direction of the action $\check{r}$. What saves the result is that the perturbation is picked in the particular class $\bar{\mathcal{H}}$.

Theorem 31. *Let $\alpha \in \mathrm{D}_{\gamma,\tau}$ and $K^o \in \bar{\mathcal{K}}$ such that the hessian $\int_{\mathbb{T}^n} \frac{\partial^2 K^o}{\partial \check{r}^2}(\theta, 0)\, d\theta$ is non degenerate. For every $H \in \bar{\mathcal{H}}$ close to K^o, there exist a unique $(K, G, R) \in \bar{\mathcal{K}} \times \bar{\mathcal{G}} \times \mathbb{R}^{\hat{n}}$ such that*

$$H(\hat{\theta}, \check{\theta}, \hat{r} + \hat{R}, \check{r}) = K \circ G(\theta, r)$$

in a neighborhood of T_0^n which is locally uniform with respect to H; in particular, H possesses an α-quaspieriodic invariant torus.

The proof consists in patterning the proof of theorem 7, and using corollary 30 instead of theorem 2.

Let us now focus on weakly non-degenerate Hamiltonians. The following result is the analogue of theorem 10.

Corollary 32 (Hypothetical conjugacy for quasi-periodic time-dependent perturbations)**.** *For every $K^o \in \bar{\mathcal{K}}^{\alpha^o}_{s+\sigma}$ with $\alpha^o \in \mathrm{D}_{\gamma,\tau}$, there is a germ of diffeomorphism*

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

at $K^o \mapsto (K^o, \mathrm{id})$ with $\alpha_H = (\hat{\alpha}_H, \check{\alpha}^o)$, such that for every H with $\alpha_H \in \mathrm{D}_{\gamma,\tau}$,

$$H = K_H \circ G_H$$

and K_H and G_H are unique.

This corollary follows from corollary 30, but not directly from theorem 10, because in theorem 10 there is no uniqueness and it is not obvious that $\check{\alpha} = \check{\alpha}^o$.

The application of corollary 32 to various cases is left to the reader. It often relies on the remark that if $\check{\alpha}^o \in \mathrm{D}_{\check{\gamma},\tau} \subset \mathbb{R}^{\check{n}}$ for some
check$\gamma, \tau > 0$, then $\{\alpha \in \mathrm{D}_{\gamma,\tau},\ \check{\alpha} = \check{\alpha}^o\}$ has positive Lebesgue measure for some $\gamma > 0$.

F. Comments

Section 1. The proof of invariant tori theorems presented here differs from others chiefly for the following reasons:

— The emphasis on twisted conjugacy, which allows to separate the inversion problems from the use of the non-degeneracy hypotheses (see the introduction).

— Classical perturbation series (or some modification of these) have been directly shown to converge in some cases (see [49] for the convergence of Schröder series in the Siegel problem; see [17], or [13], for Lindstedt series of Hamiltonians). These direct proofs of convergence are involved because, as J. Moser noticed in [32, p. 149], the series do not converge absolutely, and thus the proof of conditional convergence must take into account compensations or the precise accumulation of small denominators through a subtle combinatorial analysis. On the other hand, as Kolmogorov discovered, the perturbation series yielded by the Newton algorithm are absolutely convergent, provided that one adequately decreases the width of the analytic extension at each step of the induction. The magics is that compensations are taken into account without any further care, yet without explaining "the whole truth".

— We encapsulate the Newton algorithm in an abstract inverse function theorem à la Nash-Moser. The idea for KAM theory goes back to [53, 54]; see also [51]. The algorithm indeed converges without any specific hypothesis on the internal structure of the variables, as in Nash's construction of a solution to the isometric embedding of Riemannian manifolds ([29, 24]). At the expense of some optimality, ignoring this structure allows for a simple control of the bounds and for solving a whole class of analogous problems with the same toolbox (quasi-periodic time dependent perturbations as in appendix E, lower dimensional tori, codimension-one tori, Rüssmann's translated curve theorem, Siegel problem, as well as a number of problems in singularity theory, etc.).

— The analytic (or Gevrey) category is simpler, in Nash-Moser theory, than Hölder or Sobolev categories because the Newton algorithm can be carried out without intercalating smoothing operators (cf. [44, 7]).

— Hadamard interpolation inequalities are optimal and simple for analytic norms because, again, they do not depend on regularizing operators, as it is shown in appendix C (cf. [23, Theorem A.5]).

— The use of auxiliary norms ($|\cdot|_{G,s}$ in lemmas 3 and 5, $|\cdot|_{x,s}$ in appendix A) prevents from artificially loosing, due to compositions, a fixed width of analyticity at each step of the Newton algorithm —the domains of analyticity being deformed rather than shrunk. As a pitfall, the argument of [24, Sections 5 and 6] to deduce an inverse function theorem in the smooth category abstractly from the theorem in the analytic category, does not apply directly here.

Section 3. Lemma 3. The estimate is 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 allows to find the optimal exponent of σ, making it independent of the dimension; see [30, 40].

Lemmas 4 and 5. The small denominators and the composition operators have the same effet, in the estimates, of reducing the width of analyticity.

Section 6. Definition 12. This property can be expressed in terms of the rank of the matrix of partial derivatives of α at all orders (see [47] for instance). Outside the neglectible set where one needs to take into account some larger order derivatives, it is equivalent to being *essentially non planar* in the terminology of [39], through the following remark: if α is weakly non degenerate, in the neighborhood of every $t \in \mathbb{B}^T$, α passes through points of $\mathbb{R}^n$ which do not lie in any proper vector subspace, so there is a curve drawn on $\alpha(\mathbb{B}^T)$ which is essentially non planar, hence α itself is essentially non planar.

Theorem 13. For the dramatic history of the five proofs found independently in the 80's and 90's, see [46].

Appendix A. Theorem 17. — The two competing small parameters η and σ being fixed, our choice of the sequence (σ_k) maximizes ϵ for the Newton algorithm. It does not modify the sequence (x_k) but only the information we retain from (x_k).

— In the expression of ϵ, the square exponent of C is inherent in the quadratic convergence of Newton's algorithm. From this follows the dependence, in KAM theory, of the size ϵ of the allowed perturbation with respect to the small Diophantine constant γ: $\epsilon = O(\gamma^2)$.

— The method of Jacobowitz [24] (see Moser [30] 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 characterize 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 inversion function theorem there is an interplay between the initial and modified norms of F_s, and the analytic approximations y_j do not belong to the initial domain of definition of ϕ. Such a difficulty is inherent in the presence of composition operators, and did not occur in the problem of isometric embeddings. It is probably simpler to intercalate smoothing operators within the Newton algorithm, as Sergeraert [44], or later, Hamilton [21].

Appendix A.1. One can prove that ψ is C^1 without additional assumptions, just by patterning [44, p. 626]). 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 to ignore the way ψ was built.

Appendix B. See similar statements in [37, 38].

Appendix C. In this paragraph, the obtained inequalities are analogues for complex extensions of tori in their cotangent bundle, of the standard Hadamard convexity inequalities for infinite strips in $\mathbb{C}$. They are optimal and show that analytic norms are not quite convex with respect to the width of the complex extensions, due to the geometry of the phase space. See [33, Chap. 8] for more general but non-optimal inequalities.

Appendix D. Proposition 27. There are reasons to believe that the so-obtained 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 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 allows, 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 the Brjuno condition; see [38, 43].

Thank you to A. Albouy, V. Arnold, P. Bernard, A. Chenciner, L. Chierchia, A. Knauf [25], R. Krikorian, I. Kupka, D. Sauzin, M. Sevryuk and J.-C. Yoccoz, for illuminating comments.

References

[1] V. I. Arnold. Small denominators. I. Mapping the circle onto itself. *Izv. Akad. Nauk SSSR Ser. Mat.*, 25:21–86, 1961.

[2] V. I. Arnold. Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. *Uspehi Mat. Nauk*, 18(5 (113)):13–40, 1963.

[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. Instability of dynamical systems with many degrees of freedom. *Dokl. Akad. Nauk SSSR*, 156:9–12, 1964.

[5] V. I. Arnold. Applicability conditions and an error bound for the averaging method for systems in the process of evolution through a resonance. *Dokl. Akad. Nauk SSSR*, 161:9–12, 1965.

[6] 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. [Dynamical systems. III], Translated from the Russian original by E. Khukhro.

[7] 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.

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

[9] H. W. Broer, G. B. Huitema, F. Takens, and B. L. J. Braaksma. Unfoldings and bifurcations of quasi-periodic tori. *Mem. Amer. Math. Soc.*, 83(421):viii+175, 1990.

[10] A. Celletti and L. Chierchia. KAM stability and celestial mechanics. *Mem. Amer. Math. Soc.*, 187(878):viii+134, 2007.

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

[12] L. Chierchia. A. N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. A comment on: "On conservation of conditionally periodic motions for a

small change in Hamilton's function" [Dokl. Akad. Nauk SSSR (N.S.) **98** (1954), 527–530; mr0068687]. *Regul. Chaotic Dyn.*, 13(2):130–139, 2008.
[13] L. Chierchia and C. Falcolini. A direct proof of a theorem by Kolmogorov in Hamiltonian systems. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 21(4):541–593, 1994.
[14] R. de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.
[15] H. Scott Dumas, J. A. Ellison, and M. Vogt. First-order averaging principles for maps with applications to accelerator beam dynamics. *SIAM J. Appl. Dyn. Syst.*, 3(4):409–432 (electronic), 2004.
[16] R. E. Edwards. *Functional analysis.* Dover Publications Inc., New York, 1995. Theory and applications, Corrected reprint of the 1965 original.
[17] L. H. Eliasson. Absolutely convergent series expansions for quasi periodic motions. *Math. Phys. Electron. J.*, 2:Paper 4, 33 pp. (electronic), 1996.
[18] J. Féjoz. Quasiperiodic motions in the planar three-body problem. *J. Differential Equations*, 183(2):303–341, 2002.
[19] 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.
[20] G. Glaeser. Étude de quelques algèbres tayloriennes. *J. Analyse Math.*, 6:1–124; erratum, insert to 6 (1958), no. 2, 1958.
[21] R. S. Hamilton. The inverse function theorem of Nash and Moser. *Bull. Amer. Math. Soc. (N.S.)*, 7(1):65–222, 1982.
[22] M. Herman. Séminaire de systèmes dynamiques, 1997.
[23] L. Hörmander. The boundary problems of physical geodesy. *Arch. Rational Mech. Anal.*, 62(1):1–52, 1976.
[24] H. Jacobowitz. Implicit function theorems and isometric embeddings. *Ann. of Math. (2)*, 95:191–225, 1972.
[25] A. Knauf. Klassische mechanik. To appear, 500 pages, 2010.
[26] 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.
[27] V. F. Lazutkin. Existence of a continuum of closed invariant curves for a convex billiard. *Uspehi Mat. Nauk*, 27(3(165)):201–202, 1972.
[28] V. F. Lazutkin. Existence of caustics for the billiard problem in a convex domain. *Izv. Akad. Nauk SSSR Ser. Mat.*, 37:186–216, 1973.
[29] 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.
[30] J. Moser. A rapidly convergent iteration method and non-linear differential equations. II. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:499–535, 1966.
[31] J. Moser. A rapidly convergent iteration method and non-linear partial differential equations. I. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:265–315, 1966.
[32] J. Moser. Convergent series expansions for quasi-periodic motions. *Math. Ann.*, 169:136–176, 1967.

[33] R. Narasimhan. *Several complex variables.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1995. Reprint of the 1971 original.
[34] R. Pérez-Marco. KAM techniques in PDE. *Astérisque*, 290:Exp. No. 908, ix, 307–317, 2003. Séminaire Bourbaki. Vol. 2001/2002.
[35] H. Poincaré. *Les méthodes nouvelles de la mécanique céleste [microform] / par H. Poincaré.* Gauthier-Villars et fils, Paris :, 1892.
[36] J. Pöschel. Integrability of Hamiltonian systems on Cantor sets. *Comm. Pure Appl. Math.*, 35(5):653–696, 1982.
[37] 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.
[38] J. Pöschel. KAM à la R. Preprint, 2010.
[39] A. S. Pyartli. Diophantine approximations of submanifolds of a Euclidean space. *Funkcional. Anal. i Priložen.*, 3(4):59–62, 1969.
[40] H. Rüssmann. On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages 598–624.Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975.
[41] 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.
[42] 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.
[43] H. Rüssmann. Kam-iteration with nearly infinitely small steps in dynamical systems of polynomial character. *Discrete Contin. Dyn. Syst. Ser. S*, 2010. to appear.
[44] 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.
[45] 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.
[46] 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 Vladimir Igorevich Arnold on the occasion of his 65th birthday}.
[47] M. B. Sevryuk. Invariant tori in quasi-periodic non-autonomous dynamical systems via Herman's method. *Discrete Contin. Dyn. Syst.*, 18(2-3):569–595, 2007.
[48] M. B. Sevryuk. KAM tori: persistence and smoothness. *Nonlinearity*, 21(10):T177–T185, 2008.
[49] C. L. Siegel. Iteration of analytic functions. *Ann. of Math. (2)*, 43:607–612, 1942.
[50] E. M. Stein. *Singular integrals and differentiability properties of functions.* Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[51] J. A. Vano. *A Nash-Moser implicit function theorem with Whitney regularity and applications*. PhD thesis, University of Texas at Austin, 2002.
[52] H. Whitney. Analytic extensions of differentiable functions defined in closed sets. *Trans. Amer. Math. Soc.*, 36(1):63–89, 1934.
[53] E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. I. *Comm. Pure Appl. Math.*, 28:91–140, 1975.
[54] E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. II. *Comm. Pure Appl. Math.*, 29(1):49–111, 1976.