Available online at www.sciencedirect.com
SCIENCE DIRECT®

J. Differential Equations 206 (2004) 55–93

Journal of Differential Equations

www.elsevier.com/locate/jde

Moser's theorem for lower dimensional tori☆

Luigi Chierchia[a,*], Dingbian Qian[b]

[a] *Dipartimento di Matematica, Università "Roma Tre", Largo S. L. Murialdo 1, 00146 Roma, Italy*
[b] *Department of Mathematics, Suzhou University, Suzhou 215006, PR China*

Received 4 March 2003; revised 13 May 2004

Available online 18 August 2004

Abstract

Moser's C^ℓ-version of Kolmogorov's theorem on the persistence of maximal quasi-periodic solutions for nearly-integrable Hamiltonian system is extended to the persistence of *non*-maximal quasi-periodic solutions corresponding to lower-dimensional *elliptic* tori of any dimension n between one and the number of degrees of freedom. The theorem is proved for Hamiltonian functions of class C^ℓ for any $\ell > 6n+5$ and the quasi-periodic solutions are proved to be of class C^p for any p with $2 < p < p_*$ for a suitable $p_* = p_*(n, \ell) > 2$ (which tends to infinity when $\ell \to \infty$).

MSC: 70H08; 37J40; 34C27

Keywords: Nearly-integrable Hamiltonian systems; KAM theory; Kolmogorov's theorem; Smooth invariant tori; Lower dimensional tori; Small divisors; Fast convergent methods; Smoothing techniques

1. Introduction and results

1.1. Moser's main contribution to the so-called KAM theory was to extend Kolmogorov's invariant-tori-theorem [9] to smooth category. Kolmogorov's celebrated

☆ Supported by MIUR *Variational Methods and Nonlinear Differential Equations*. Dingbian Qian was supported by the "Distinguished Visiting Scholar Program" of China and NNSF of China (No.10071055, No.10271085).

* Corresponding author.
E-mail addresses: luigi@mat.uniroma3.it (L. Chierchia), dbqian@suda.edu.cn (D. Qian).

0022-0396/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jde.2004.06.014

theorem deals, as well known, with the persistence under small, real-analytic perturbations of maximal quasi-periodic solutions (associated to maximal invariant tori) for nearly-integrable Hamiltonian systems. The basic technical tool exploited by Moser in his extension was closely related to ideas of Nash [17] and consisted in using a Newton (quadratic) iteration method, re-inserting at each step enough regularity into the problem so as to beat (together with the so-called "small divisor problem", already overcome by Kolmogorov and Arnold) the loss of regularity due to the inversion of certain (non-elliptic) differential operators. In the original work of Moser [14], which was dealing with twist area-preserving maps (corresponding to the Hamiltonian system case in "one and a half" degrees of freedom), the perturbation was assumed to be C^{333}. The regularity assumption (in the twist map case) was later brought down to five by Rüssmann [22]; for the Hamiltonian case we refer to [16,29], and, especially, [19], where Kolmogorov's theorem is proved under the hypothesis that the perturbation is C^ℓ with $\ell > 2d$, d being the number of degrees of freedom. We recall also that Herman [8] gave a counterexample in the twist map case with $\ell = 3 - \varepsilon$, $\varepsilon > 0$ (corresponding to $\ell = 4 - \varepsilon$ in the Hamiltonian case with two degrees of freedom).

1.2. Right after KAM theory for maximal tori was established, it appeared clear that an important direction of further investigations was that of the existence of *lower dimensional quasi-periodic solutions* corresponding to lower dimensional invariant tori, i.e., tori of dimension[1] $n < d$ (as above, d stands for the number of degrees of freedom). In 1965 Melnikov stated a precise result concerning the persistence of *stable* (or "elliptic") lower-dimensional tori in [13]; the hypotheses of such result are, now, commonly referred to as "Melnikov conditions". However, a proof of Melnikov's theorem was given only later by Moser [15] for the case $n = d - 1$ and, in the general case, by Eliasson in [6] and, independently, by Kuksin [10]; see also [20]. The *unstable* (or "hyperbolic") case (i.e., the case for which the lower dimensional tori are linearly unstable and lie in the intersection of stable and unstable Lagrangian manifolds) is simpler[2] and a complete perturbation theory was worked out in [15,7,29]. Various technical progresses have been recently performed in, e.g., [21,2,28,27,25]. Incidentally we mention that lower dimensional quasi-periodic solutions are particularly relevant in connection with extensions to PDE's; see, e.g., [5,11,12,21,3] and references therein.

1.3. All the above mentioned results concerning the extension of Kolmogorov's theorem to lower dimensional tori *deal only with the real-analytic case*. It is the purpose of this paper to extend Moser's theorem to lower dimensional quasi-periodic solutions proving, under suitable generic assumptions, the persistence and the regularity of lower n-dimensional *elliptic* tori (corresponding to lower dimensional quasi-periodic solutions) for C^ℓ perturbations of nearly-integrable systems with $\ell > 6n + 5$.

[1] Equilibria and periodic orbits, corresponding, respectively, to $n = 0$ and 1, are the simplest examples; in such cases there are no small-divisor problems and existence was already established by Poincaré by means of the standard Implicit Function Theorem: see [18, Volume I, Chapter III].

[2] On a technical level: the normal frequencies to the torus do not resonate with the inner (or "proper") frequencies associated to the quasi-periodic motion.

Before stating in a more precise way our results, let us mention that it was already remarked by Graff in[3] [7] that combining "soft" tools of invariant manifold theory (based on the standard Implicit Function Theorem) and KAM theory for maximal tori one can conclude that lower dimensional unstable tori persist under small perturbations (but regularity of the continued manifolds may be, in general, quite low). As well known, however, such "partially hyperbolic techniques" do not carry over to the elliptic situation.

1.4. We proceed, now, to formulate the main result proved in this paper. Consider a (smooth) Hamiltonian system with $n+m$ degrees of freedom, governed by a Hamiltonian function of the form

$$H(x, y, u, v; \xi) := N(y, u, v; \xi) + P(x, y, u, v; \xi), \tag{1.1}$$

where $(x, y) \in \mathbf{T}^n \times \mathbf{R}^n$ and $(u, v) \in \mathbf{R}^{2m}$ are pairs of standard symplectic coordinates[4] and ξ is a real parameter running over a compact set $\Pi \subset \mathbf{R}^n$ of positive Lebesgue measure[5]; N is in "normal (integrable) form":

$$N = e(\xi) + \sum_{j=1}^{n} \omega_j(\xi) y_j + \frac{1}{2} \sum_{j=1}^{m} \Omega_j(\xi)(u_j^2 + v_j^2), \tag{1.2}$$

P is a small perturbation. The motions generated by N decouple in a Kronecker flow $x \in \mathbf{T}^n \to x + \omega(\xi)t$ times the motion of m (decoupled) harmonic oscillators with characteristic frequencies $\Omega_j(\xi)$ (sometimes referred to as *normal frequencies*); in particular, the n-parameter family (parameterized by ξ) of n dimensional tori

$$T_0^n(\xi) := \mathbf{T}^n \times \{y = 0\} \times \{u = v = 0\}, \quad \xi \in \Pi,$$

are *linearly stable* (*elliptic*) *invariant tori of dimension n carrying quasi-periodic motions with frequency* $\omega(\xi) \in \mathbf{R}^n$.

[3] Compare point **b** of the introduction in [7, p. 6]. Graff's remark has been recently re-considered by Huang, D. and Liu, Z.: *On the persistence of lower dimensional invariant hyperbolic tori for smooth Hamiltonian systems*, Nonlinearity, 13 (2000) 189–202.

[4] Hence the equation of motion are $\dot{x} = H_y$, $\dot{y} = -H_x$, $\dot{u} = H_v$, $\dot{v} = -H_u$, where $H_y := (H_{y_1}, \ldots, H_{y_n})$, etc.; $\mathbf{T}^n := \mathbf{R}^n/(2\pi\mathbf{Z}^n)$.

[5] Typically, ξ may indicate an initial datum y_0 and y the distance from such point or (equivalently, if the system is non-degenerate in the classical Kolmogorov sense) $\xi \to \omega(\xi)$ might be simply the identity, which amounts to consider the unperturbed frequencies as parameter.

Theorem 1.1. *Let $\ell > 6n+5$ and let H in (1.1) be C^ℓ in a neighborhood of $\mathbf{T}^n \times \{y = 0\} \times \{u = v = 0\}$ and (uniformly) Lipschitz continuous in*[6] $\xi \in \Pi$. *Assume that*

$$\Omega_i(\xi) > 0, \quad \Omega_i(\xi) \neq \Omega_j(\xi), \quad \forall \xi \in \Pi, \ \forall\ i \neq j. \tag{1.3}$$

Assume, also, that $\xi \in \Pi \to \omega(\xi) \in \mathbf{R}^n$ is a Lipschitz homeomorphism of Π onto its image and that[7]

$$\begin{aligned} &\operatorname{meas}\{\xi \in \Pi : \langle \omega(\xi), k\rangle + \langle \Omega(\xi), l\rangle = 0\} = 0, \\ &\forall\ k \in \mathbf{Z}^n \backslash \{0\}, \quad \forall\ l \in \mathbf{Z}^m \ : \ |l| \le 2. \end{aligned} \tag{1.4}$$

Then, if the gradient of P, together with its Lipschitz semi-norm in ξ, is small enough, there exists a set $\Pi_\infty \subset \Pi$ of positive Lebesgue measure and a family of n-dimensional linearly stable H-invariant tori $T^n(\xi)$ parameterized by (and Lipschitz continuous in) $\xi \in \Pi_\infty$. The tori $T^n(\xi)$ are C^p-smooth for any $2 < p < p_*$ for a suitable $p_* = p_*(n, \ell) > 2$. On $T^n(\xi)$ the H-flow is C^p-conjugated to the Kronecker flow $x \to x + \omega_\infty(\xi)t$ where ω_∞ is a Lipschitz homeomorphysm on Π_∞ close to ω; for all $\xi \in \Pi_\infty$, $\omega_\infty(\xi)$ is a "Diophantine vector".

1.5. Let us collect, here, a few remarks on the above statements.

1.5.1. Conditions (1.3)–(1.4) are a generalized version [21] of Melnikov's conditions and represent a rather weak independence requirement between ω and Ω (obviously satisfied if, for example, Ω is independent of ξ). Notice that, if ω and Ω are C^1, (1.4) is satisfied whenever[8] (taking ω as independent variable)

$$\partial_\omega \langle \Omega, l\rangle \neq k, \quad \forall\ k \in \mathbf{Z}^n \backslash \{0\}, \quad \forall\ l \in \mathbf{Z}^m \ : \ |l| \le 2, \tag{1.5}$$

in which case the level sets $\{\omega : \langle k, \omega\rangle + \langle l, \Omega(\xi(\omega))\rangle = 0\}$ are $(n-1)$-dimensional C^1 hypersurfaces (and hence of vanishing n-dimensional measure).

1.5.2. Condition (1.3) requires the normal frequencies to be bounded away from zero *and* to be "simple". Recently, in the KAM method of [28], the simplicity of the

[6] A function g is uniformly Lipschitz continuous on Π if $|g|^{\mathrm{Lip}} := \sup \dfrac{|g(\xi) - g(\xi')|}{|\xi - \xi'|}$ is finite, the supremum being taken over all $\xi \neq \xi'$ in Π (and usually, we shall not indicate explicitly the domain Π in the notations since it will be clear from context).

[7] Here, "meas" denotes Lebesgue measure; $\langle \cdot, \cdot \rangle$ denotes the standard inner product; for integer vectors $l = (l_1, \ldots, l_m)$, $|l| = \sum_i |l_i|$. Obviously, $\omega = (\omega_1, \ldots, \omega_n)$ and $\Omega = (\Omega_1, \ldots, \Omega_n)$; later, however, Ω will also be identified with the diagonal matrix $\operatorname{diag}(\Omega_1, \ldots, \Omega_n)$.

[8] Actually, it is sufficient to require (1.5) for a finite number of vectors k; compare (2.89) below.

normal frequencies has been relaxed allowing, in [4], to establish the existence (and the linear stability) of quasi-periodic solutions for the one-dimensional wave equation with periodic boundary conditions. It is conceivable (but not obvious) that methods taken from [28] might lead to remove the second condition in (1.3).

1.5.3. The tori $T^n(\xi)$ are a C^p-embedding of the standard flat n-torus $\mathbf{T}^n$ into the $2(n+m)$-dimensional phase space. In fact, the embedding is C^p-close to the identity for any $2 < p < p_*$. The number p_* may be taken as follows. Pick

$$6n + 5 < \ell_* < \ell \tag{1.6}$$

and let $\theta \in (0, 1/3)$ be such that

$$\frac{(1+\theta)^2}{1-3\theta} = \frac{\ell \ \ -2}{\ell_* - 2}. \tag{1.7}$$

Then (compare (2.67) below),

$$p_* := 2 + a(\ell - 2), \quad \text{with} \quad a := \frac{2}{3}\frac{\theta}{(1+\theta)^2} \ . \tag{1.8}$$

In particular, if P is C^∞, so are the tori $T^n(\xi)$ and the associated quasi-periodic solutions.

1.5.4. The invariant tori $T^n(\xi)$, $\xi \in \Pi_\infty$, correspond to *non-maximal quasi-periodic solutions* with n rationally independent uniformly Diophantine frequencies $\omega_{\infty 1}, \ldots, \omega_{\infty n}$ satisfying

$$|\langle \omega_\infty(\xi), k\rangle| \geq \frac{\gamma_\infty}{1+|k|^\tau}, \quad \forall k \in \mathbf{Z}^n \backslash \{0\}, \ \ \forall \ \xi \in \Pi_\infty, \tag{1.9}$$

where

$$\tau := \frac{\ell_* - 11}{6} > n - 1 \tag{1.10}$$

and γ_∞ is a suitable (small enough) positive number. In fact, a slightly stronger Diophantine property holds, since (1.9) holds also replacing $\langle \omega_\infty(\xi), k\rangle$ with $\langle \omega_\infty(\xi), k\rangle + \lambda$, where $\lambda := \lambda(\xi)$ denotes "$T^n(\xi)$-normal frequencies" or differences of such normal frequencies.

1.5.5. A detailed and quantitative version of Theorem 1.1 is given in Proposition 2.1 (convergence of the KAM iteration) and in Proposition 2.2 (measure estimates on Π_∞) below.

1.5.6. The "smoothing technique" we shall use is due to Jackson, Moser and Zehnder (compare [26]) and it is rather different from the original strategy introduced by Nash and used by Moser in the context of dynamical systems. The Jackson–Moser–Zehnder technique is based on approximating the C^ℓ perturbation P by real-analytic functions on smaller and smaller complex neighborhoods, solving linearized (analytic) equation to a better and better degree (keeping careful quantitative track of the procedure) and recovering in the limit a smooth (at least C^2 in our case) solution.

We point out that we do not use directly an analytic theorem (as done, for instance, in [26]), nor an analytic theorem can be immediately extracted from our approach.

1.5.7. The assumption $\ell > 6n + 5$ is certainly *not optimal*. It would be interesting to find the optimal value: for example, is it true that Theorem 1.1 holds provided $\ell > 2n$ (as in the maximal case)?

1.5.8. Part of the proof relies on analytic tools elaborated in [21] and we, therefore, follow quite closely the notations introduced in [21]. Another reason for using notations borrowed from [21] is that it might facilitate the extension of our results to infinite ($m = \infty$) dimension. However, we restrain to do so here since we believe that such an extension makes sense only if applied to a real infinite dimensional problem, such as, for example, some "relevant" nonlinear PDE.

1.6. The (normal) form (1.2) of the integrable piece N is rather standard in the present context (compare, e.g., [21,27]). However, we mention briefly how more classical situations may be included in the present formulation. As an example, consider a Hamiltonian

$$h(\varphi, I, q, p; \varepsilon) = h_0(I, q, p) + \varepsilon h_1(\varphi, I, q, p; \varepsilon),$$

where (φ, I) and (q, p) are pairs of standard symplectic coordinates with $\varphi \in \mathbf{T}^n$, $I \in B_1(0) \subset \mathbf{R}^n$ and (q, p) in a small neighborhood of the origin in $\mathbf{R}^{2m}$. Assume that $h_0 \in C^{\ell+3}$ and that $h_1 \in C^\ell$. Fix a point $\bar{I}$, say $\bar{I} = 0$, and assume that $(q, p) = (0, 0)$ is a linearly stable equilibrium for $(q, p) \to h_0(0, q, p)$. If such an equilibrium is non-degenerate (i.e., if the Hessian matrix $\partial^2_{(q,p)} h_0(0, 0, 0)$ is invertible), then, up to a symplectic change of coordinates, we may assume that $(q, p) = (0, 0)$ is a non-degenerate, stable equilibrium for $(q, p) \to h_0(I, q, p)$ *for any* $I \in B_\rho(0)$ for some $0 < \rho < 1$. Assume, also, that the eigenvalues of $J_m \partial^2_{(q,p)} h_0(0, 0, 0)$, ($J_m :=$ standard $(2m \times 2m)$-symplectic matrix), are purely imaginary ("linear stability") and *simple* and are given by $\pm \mathrm{i}\Omega_j$ with $\Omega_j > 0$ and $j = 1, \ldots, m$. Finally, assume that also the Hessian matrix $\partial^2_I h_0(0, 0, 0)$ is invertible; this assumption corresponds to the classical KAM non-degeneracy condition. Then, expanding h_0 in a neighborhood of $(\xi, 0, 0) := (I_0, 0, 0)$, (up to order two in $y = I - \xi$ and three in (q, p) for $\xi \in B_{\rho/2}$) and using a classical result of Weierstrass on the diagonalization of quadratic symplectic forms, one can find a symplectic $(2m \times 2m)$ matrix $S(\xi)$ such that, in the symplectic variables $(x, y) := (\varphi, I - \xi)$, $(u, v) = S(\xi)(q, p)$, the Hamiltonian $h_0 + \varepsilon h_1$ takes the form (1.1) where N is as in (1.2) with $\xi := I_0$, $e := h_0(\xi, 0, 0)$, $\omega := \partial_I h_0(\xi, 0, 0)$, $\Omega_j(0) = \Omega_j$. Furthermore, the perturbation $P := h_0 + \varepsilon h_1 - N$ is C^ℓ and satisfies

$$P = O(|y|^2) + O(|y||(u, v)|) + O(|(u, v)|^3) + O(\varepsilon). \tag{1.11}$$

We shall, therefore, consider P on a real domain of the form

$$\{x \in \mathbf{T}^n, \quad |y| < r^2, \quad |(u, v)| < r\}, \quad \xi \in \Pi := \overline{B}_{\rho/2}(0)$$

for a small enough $0 < r < \rho/2$. Notice that, because of the simplicity of the eigenvalues, the dependence of Ω_j upon ξ (possibly reducing ρ) is of class $C^{\ell+1}$; furthermore, the hypothesis on $\partial_I^2 h_0$ implies that $\omega(\xi)$ is a diffeomorphysm. From Theorem 1.1 (or, more precisely, from its quantitative version given in Propositions 2.1 and 2.2 below), it follows that *if one chooses $r := \varepsilon^{\frac{1}{3}}$ and ε is small enough, then, generically, for any ξ in a Cantor subset of $B_{\rho/2}$ of density $O(1 - \varepsilon^\alpha)$, (for some $0 < \alpha < 1$), the unperturbed n-dimensional tori $y = 0 = u = v$, $x \in \mathbf{T}^n$ may be continued into C^p h-invariant tori*; compare Remark 2.2 below.

1.7. The arguments on which the proof of Theorem 1.1 is based are, as often happens in KAM theory, rather technical and somewhat involved. Therefore, we close this introduction with a "guide to the proof" of Theorem 1.1 (divided into four parts). The actual complete proof is given in Section 2.

1.7.1. *Smoothing and analytic approximants (Sections 2.1 and 2.2)*

First, by standard real analytic tools we extend the perturbation function P to $\mathbf{R}^{2(n+m)}$. Then, we fix (see, also, 1.7.2 below) a sequence of fast decreasing numbers $\sigma_\nu \downarrow 0$ ($\nu \geq 1$) and, using the approximation theory of Jackson, Moser and Zehnder (Lemma 2.1), we construct a sequence of real-analytic function $P^{(\nu)}$ such that the following holds.

(i) $P^{(\nu)}$ is real-analytic on the complex strip Δ_{σ_ν} of width σ_ν around $\mathbf{R}^{2(n+m)}$.
(ii) The $P^{(\nu)}$'s satisfy the bounds: $\sup_{\Delta_{\sigma_\nu}} |\nabla(P^{(\nu)} - P^{(\nu-1)})| \leq c\ |P|_{C^\ell} \sigma_{\nu-1}^{\ell-1}$; compare Lemma 2.1. In this section, "c" denotes (different) constants depending only on n, ℓ and ℓ_*.
(iii) The first approximant $P^{(1)}$ is "small" with the perturbation P:

$$\|\nabla P^{(1)}\|_{r_1, s_1} \leq c|P|_{C^\ell}\ |\nabla P|_{r_1}, \tag{1.12}$$

where: $\|\cdot\|_{r,s}$ is a suitable weighted norm on complex functions, while $|\cdot|_r$ is a corresponding weighted norm on real functions;[9] the domain where the complex

[9] In Section 2 the norm $\|\cdot\|_{r,s}$ is denoted $\|\cdot\|_{r,D(r,s)}$; also, in place of the notation $\|\nabla f\|$, below (following [21]) we use the notation $\|X_f\|$. Furthermore, in Section 2 the norm $|\cdot|_r$ is denoted $|\cdot|_{D_r}$ (see (2.61) and (2.62)).

functions are considered is of the form

$$D(r,s) = \{(x,y,u,v) \in \mathbf{C}^{2(n+m)} : |\mathrm{Im}\, x| < s,\ |y| < r^2,\ \ |u| < r,\ \ |v| < r\}, \quad (1.13)$$

while the domain where the real functions are considered is the projection of $D(r,s)$ on $\mathbf{R}^{2(n+m)}$; the positive numbers r_1, s_1 and σ_1 ("the initial analyticity radii") are chosen so as to meet (1.12). The weighted norms are discussed in Section 2.2; in such section we also introduce—as it is costumary in studying Hamiltonian equilibria—symplectic complex variables $\bar{z}$ and z linearly related to the variables u and v. Estimate (1.12) is discussed particularly in (2.71) and (2.63).

1.7.2. *The KAM scheme (Section 2.3)*

This is the heart of the proof. The idea—as in all KAM methods—consists in a super-convergent (sometimes: Newton or quadratic) iterative procedure apt to reduce, at each step of the scheme, the size of the perturbing function by a fixed power $\kappa > 1$ of the size of the perturbing function at the preceding step; this is done in order to beat the loss of smoothness and the divergences introduced by the small divisors arising in the inversion of non-elliptic differential operators. The scheme we need in our specific problem is non-standard and, from a technical point of view, represent the most novel part of the proof. For these reasons we give, now, a rather detailed description of such scheme.

We want to construct, inductively, real-analytic symplectic transformations Φ_ν, $\nu \geq 1$, so that

$$(N + P^{(\nu)}) \circ \Phi_\nu = N_{\nu+1} + P_{\nu+1}, \tag{1.14}$$

where the sequence of N_ν's is in "normal form",

$$N_\nu(y,u,v;\xi) := e_\nu(\xi) + \sum_{j=1}^{n} \omega_{\nu j}(\xi) y_j + \frac{1}{2}\sum_{j=1}^{m} \Omega_{\nu j}(\xi)(u_j^2 + v_j^2), \tag{1.15}$$

while the sequence of real-analytic functions P_ν's are perturbations of smaller and smaller size:

$$\|\nabla P_{\nu+1}\|_{r_{\nu+1},s_{\nu+1}} \sim \|\nabla P_\nu\|^{\kappa}_{r_\nu,s_\nu}, \tag{1.16}$$

the number $\kappa = \kappa(\ell, \ell_*)$ can be taken to be $\kappa = 1 + \theta$, $\theta \in (0, 1/3)$ being defined in (1.7). The parameter ξ appearing in (1.15) will vary in smaller and smaller compact sets Π_ν (of relatively large Lebesgue measure)

$$\Pi \supset \Pi_1 \supset \cdots \Pi_\nu \supset \Pi_{\nu+1} \supset \cdots \Pi_\infty = \bigcap_{\nu=1}^{\infty} \Pi_\nu.$$

The smallness assumption on the size of $|\nabla P|_{r_1}$ and, hence (by (1.12)), of $\|\nabla P^{(1)}\|_{r_1,s_1}$ will allow to turn on the iteration procedure.

The symplectic map Φ_ν will be seeked of the form

$$\Phi_\nu = \Phi_{\nu-1} \circ \phi_\nu = \phi_1 \circ \cdots \circ \phi_\nu.$$

Thus, by induction (for $\nu \geq 2$), (1.14), takes the form

$$(N_\nu + P_\nu + (P^{(\nu)} - P^{(\nu-1)}) \circ \Phi_{\nu-1}) \circ \phi_\nu = N_{\nu+1} + P_{\nu+1}. \tag{1.17}$$

Recalling (ii) in 1.7.1 above, by choosing

$$\sigma_\nu \sim \|\nabla P_\nu\|^q,$$

with a small positive $q > 0$ (taking also into account the relation (1.16) and that ℓ is large enough), one sees that the term $\|\nabla(P^{(\nu)} - P^{(\nu-1)})\|$ can be bounded by $\|\nabla P_\nu\|$. Whence, Eq. (1.14) may be rewritten as

$$(N_\nu + P'_\nu) \circ \phi_\nu = N_{\nu+1} + P_{\nu+1}, \tag{1.18}$$

with

$$P'_\nu := P_\nu + (P^{(\nu)} - P^{(\nu-1)}) \circ \Phi_{\nu-1} \ . \tag{1.19}$$

Thus, $\|\nabla P'_\nu\| \sim \|\nabla P_\nu\|$ and (1.18) fits now in more standard KAM approaches. In fact, the techniques used in, e.g., [21], allow to equip this scheme with the necessary estimates.

We remark that in order for this approach to work, the map Φ_ν has to verify suitable compatibility relations with respect to the analyticity domains (compare the inductive relation (1.17)). More precisely, if $D_\nu := D(r_\nu, s_\nu)$ denotes the analyticity domain of P_ν, one has to show that

$$\phi_\nu : D_{\nu+1} \to D_\nu, \ (\forall \ \nu \geq 1), \quad \Phi_{\nu-1} : D_\nu \to \Delta_{\sigma_\nu}, \ (\forall \ \nu \geq 2). \tag{1.20}$$

Relation (1.20) is checked in Section 2.4; compare (2.40).

The linearized equation associated to (1.18) is thoroughly discussed in Section 2.3. This is the place where *small divisors* arise. Such small divisors have the form

$$\langle \omega_\nu(\xi), k\rangle + \langle \Omega_\nu(\xi), l\rangle, \tag{1.21}$$

where the Fourier/Taylor indices k and l verify the constraints

$$(k,l) \in Z_{K_\nu} := \left\{(k,l) \in \mathbf{Z}^{n+m} \backslash \{0\}, \quad |k| \leq K_\nu, \quad |l| \leq 2\right\}, \tag{1.22}$$

for a suitable "cut-off" $K_\nu \uparrow \infty$. The limitation on l comes from the fact that, choosing the neighborhood of the y, u and v origin as in (1.13), one may consider only lower order terms in y and (u, v); "lower order terms" meaning, here, terms up to order 1 in y and up to order 2 in (u, v). The limitation on k is reminiscent of the Fourier "cut-off" introduced originally by Arnold [1]; the difference being that, while in Arnold's proof one can take the cut-off K_ν to be proportional to the logarithm of the inverse of the size of the perturbation $\|P_\nu\|$, here we have to take it to be proportional to a (small) inverse power of the size of the perturbation $\|P_\nu\|$, making the treatment of the convergence of the algorithm more delicate.

1.7.3. *Iteration and convergence of the KAM scheme (Sections 2.4 and 2.5)*

Once the iterative step is set up, it has to be equipped with estimates. This technical part, carried out in Section 2.4, is, however, rather straightforward and follows quite closely the corresponding part in [21]. Some care has to be devoted to the choice of all the free parameters involved in the iteration so as to make the algorithm convergent: this is done in Section 2.5; see, in particular, (2.53).

Once all the above has been established, the thesis of Theorem 1.1 (apart for the statement concerning the measure of Π_∞ which is discussed in the 1.7.4) follows easily. In fact, from the definition of $P^{(\nu)}$ it follows that $P^{(\nu)}$ tends to P in, say, the $C^{\ell-1}$-norm. Furthermore, the sequence of diffeomorphysms $x \to \Phi_\nu(x, 0, 0, 0; \xi)$ is easily seen to converge in C^p-norm (for $2 < p < p_*$) to a C^p diffeomorphysm $x \to \psi(x; \xi)$, which is Lipschitz continuous in ξ. Therefore, from (1.14), from the (fast) convergence of N_ν to

$$N_\infty = e_\infty(\xi) + \langle \omega_\infty(\xi), y\rangle + \frac{1}{2}\sum_{j=1}^{n} \Omega_{\infty j}(\xi)(u_j^2 + v_j^2) \tag{1.23}$$

(and from the fact that the size of the analyticity radii measuring D_ν goes to zero much slower than the size of $\|P_\nu\|$), it follows that

$$T^n(\xi) := \psi\left(\mathbf{T}^n; \xi\right), \quad \xi \in \Pi_\infty \tag{1.24}$$

is an invariant torus for $N + P$. On such a torus, the flow is C^p-conjugated to the Kronecker flow $x \to x + \omega_\infty t$, ω_∞ being a Diophantine vector with Diophantine constants $\gamma_\infty > 0$ and $\tau = (\ell_* - 11)/6$. Finally, in view of (1.23), the tori $T^n(\xi)$ are linearly stable. Detailed, quantitative results obtained by iterating the KAM scheme are collected in Proposition 2.1.

1.7.4. *Measure estimates and multiplicity of the solutions (Section 2.6)*

The set Π_ν is iteratively defined as the subset of $\Pi_{\nu-1}$ where the small divisors (1.21) obey a Diophantine condition of the type

$$|\langle \omega_\nu(\xi), k\rangle + \langle \Omega_\nu(\xi), l\rangle| \geq \frac{\gamma_\nu}{1 + |k|^\tau}, \quad \forall\ (k, l) \in Z_{K_\nu}, \ \forall \xi \in \Pi_\nu, \tag{1.25}$$

where γ_ν is a decreasing sequence bounded away from zero and $\tau > n-1$ is defined in (1.10). The non-degeneracy assumptions on ω and Ω (i.e., the assumption that ω is a Lipschitz homeomorphysm together with (1.3)) will guarantee that the set Π_∞ is non-empty, and, in fact, of positive Lebesgue measure. Finally, the map $\xi \in \Pi_\infty \to \omega_\infty(\xi)$ is easily seen to be a Lipschitz homeomorphism so that, in particular, to different ξ correspond different tori $T^n(\xi)$. Theorem 1.1, at this point, is completely proven. A detailed formulation of the measure estimates is given in Proposition 2.2.

2. Proof of Theorem 1.1

2.1. Analytic approximants (smoothing)

We start by recalling a well known and fundamental approximation result.

Lemma 2.1 (Jackson, Moser, Zehnder). *Let $f \in C^p(\mathbf{R}^{\mathbf{k}})$ for some $p > 0$ with finite C^p norm*[10] *over $\mathbf{R}^k$. Let ϕ be a radial-symmetric, C^∞ function, having as support the closure of the unit ball centered at the origin, where ϕ is completely flat and takes value 1; let $K = \hat{\phi}$ be its Fourier transform and for all $\sigma > 0$ define*

$$f_\sigma(x) := K_\sigma * f(x) = \sigma^{-n} \int_{\mathbf{R}^{\mathbf{k}}} K\left(\frac{x-y}{\sigma}\right) f(y)\, \mathrm{d}y.$$

Then, there exist a constant $c \geq 1$ depending only on p and k such that the following holds. For any $\sigma > 0$, the function $f_\sigma(x)$ is a real-analytic function on $\mathbf{C}^k$ such that, if Δ^k_σ denotes the k-dimensional complex strip of width σ

$$\Delta^k_\sigma := \{x \in \mathbf{C}^k : |\mathrm{Im}\, x_j| \leq \sigma, \forall\ j\},$$

then, for all $\alpha \in \mathbf{N}^k$ such that $|\alpha| \leq p$, one has[11]

$$\sup_{x\in\Delta^k_\sigma} \left| \partial^\alpha f_\sigma(x) - \sum_{|\beta|\leq p-|\alpha|} \frac{\partial^{\beta+\alpha} f(\mathrm{Re}\, x)}{\beta!} (\mathrm{i}\, \mathrm{Im}\, x)^\beta \right| \leq c\ |f|_{C^p}\ \sigma^{p-|\alpha|}$$

and, for all $0 \leq s \leq \sigma$,

$$\sup_{x\in\Delta^k_s} |\partial^\alpha f_\sigma - \partial^\alpha f_s| \leq c |f|_{C^p}\ \sigma^{p-|\alpha|}.$$

[10] If p is not integer, the C^p norm $|f|_{C^p}$ denotes the $C^{[p]}$ norm of f plus the $(p-[p])$-Hölder norm of the derivatives of order $[p]$ ($[p]$ denoting, as usual, the integer part of p).

[11] "$\partial^\alpha f$" means $\dfrac{\partial^{\alpha_1+\cdots+\alpha_k} f}{\partial x_1^{\alpha_1} \cdots \partial x_k^{\alpha_k}}$.

Moreover, the Hölder norms of f_σ satisfy, for all $0 \le q \le p \le r$,

$$|f_\sigma - f|_{C^q} \le c\ |f|_{C^p}\ \sigma^{p-q}\ , \quad |f_\sigma|_{C^r} \le c\ \frac{|f|_{C^p}}{\sigma^{r-p}}.$$

The function f_σ preserves periodicity (i.e., if f is T-periodic in any of its variable x_j, so is f_σ). Finally, if f depends on some parameter $\xi \in \Pi \subset \mathbf{R}^n$ and if the Lipschitz semi-norm of f and its x-derivatives are uniformly bounded by $|f|^{\mathrm{Lip}}_{C^\ell}$, then all the above estimates hold with $|\cdot|$ replaced by $|\cdot|^{\mathrm{Lip}}$.

Remark 2.1. (i) As pointed out in [26], Lemma 2.1 yields easily the following classical bounds, valid for any [12] $0 \le r \le p \le q$:

$$|f|^{q-r}_{C^p} \le c\ |f|^{q-p}_{C^r}\ |f|^{p-r}_{C^q} \qquad \text{(convexity estimates)},$$

$$|fg|_{C^p} \le c\ (|f|_{C^p}|g|_{C^0} + |f|_{C^0}|g|_{C^p}).$$

(ii) The proof of this lemma (including the statement on dependence upon parameters) consists in a direct check (based on standard tools from calculus and complex analysis); for details see [26] and references therein.

In order to apply the lemma so as to construct a sequence $\{P^{(\nu)}\}$ of real-analytic approximants of the perturbation P *we first extend P to* $\mathbf{R}^{2(n+m)}$ (recall that P needs only be defined in a neighborhood of $\mathbf{T}^n \times \{y = 0\} \times \{u = v = 0\}$): it is clear that if P is defined on $\mathbf{T}^n \times B_{d_1,d_2} := \mathbf{T}^n \times \{|y| < d_1\} \times \{|u| < d_2, |v| < d_2\}$, then one can easily construct a C^ℓ-extension P_{ext} of $P\ |_{\mathbf{T}^n \times B_{d_1/2,d_2/2}}$ onto $\mathbf{R}^{2(n+m)}$, (maintaining periodicity in the first n variables and sharing the same properties of P with respect to the parameter ξ), and so that [13]

$$|P_{\mathrm{ext}}|_{C^\ell(\mathbf{R}^{2(n+m)})} \le a\ |P|_{C^\ell(\mathbf{T}^n \times B_{d_1,d_2})},$$

where a is a suitable positive constant depending only on ℓ and d_i.

Notational Remark 2.1. *From now on, we shall replace P by such an extension P_{ext}, which, with abuse of notation, we shall again denote P. Also, $\Delta^{2(n+m)}_\sigma$ will henceforth be denoted simply Δ_σ.*

Now, given a decreasing sequence (to be fixed later) $\sigma_\nu \downarrow 0$, $\nu \ge 1$, we define the real-analytic approximant $P^{(\nu)}$ as [14]

$$P^{(\nu)} := P_{2\sigma_\nu} := K_{2\sigma_\nu} * P.$$

[12] Clearly, in the first inequality the constant c depends on r, p, q, while in the second inequality the constant c depend only on p.

[13] In fact, one can take $P_{\mathrm{ext}} = \psi \cdot P$, ψ being a function of y, u, v having value 1 on $B_{d_1/2,d_2/2}$ and vanishing outside B_{d_1,d_2}.

[14] Recall the notation in Lemma 2.1. The (irrelevant) presence of the factor 2 will be explained in Section 2.2.

2.2. Complex variables and weighted norms

To treat the linearized equation associated to (1.18), it is convenient to introduce complex variables in a neighborhood of $u = v = 0$. Consider the following linear change of variable $(u, v) \in \mathbf{C}^{2m} \to (z, \bar{z}) \in \mathbf{C}^{2m}$:

$$z = \frac{1}{\sqrt{2}}(u + \mathrm{i}v), \quad \bar{z} = \frac{1}{\sqrt{2}}(u - \mathrm{i}v)$$

and its inverse map[15]

$$u = \frac{1}{\sqrt{2}}(z + \bar{z}) \ , \qquad v = \frac{1}{\mathrm{i}\sqrt{2}}(z - \bar{z}).$$

This map is not symplectic; however the Poisson bracket, the symplectic form and Hamilton equations transform in a simple way: if (as above) (x, y) and (u, v) are couple of conjugate symplectic variables and if f and g are functions of (x, y, u, v) then, with the obvious meaning of the symbols,[16]

$$\begin{aligned} \{f, g\} &:= \{f, g\}_{x,y,u,v} = \{f, g\}_{x,y} + \{f, g\}_{u,v} \\ &= \{f, g\}_{x,y} - \mathrm{i}\{\tilde{f}, \tilde{g}\}_{z,\bar{z}} =: \{f, g\}^{\sim}. \end{aligned}$$

The symplectic form $dx \wedge dy + du \wedge dv$ reads $dx \wedge dy - \mathrm{i}dz \wedge d\bar{z}$ and the Hamiltonian vector field

$$X_f := (f_y, -f_x, f_v, -f_u)$$

is transformed into[17]

$$\tilde{X}_{\tilde{f}} := (\tilde{f}_y, -\tilde{f}_x, -\mathrm{i}\tilde{f}_{\bar{z}}, \mathrm{i}\tilde{f}_z).$$

In the variables $(x, y, z, \bar{z})$ the function N takes the form

$$\tilde{N} = e + \langle \omega(\xi), \ \ y\rangle + \langle \Omega(\xi) \ z, \bar{z}\rangle,$$

[15] Beware that, as standard in this context, $\bar{z}$ does not denote the complex conjugate of z; rather, z and $\bar{z}$ denote a set of $2m$ independent variables. Of course, when u and v are restricted to the real space then, indeed, $\bar{z}$ and z are complex conjugate. This change of variables is standard, for example, in the theory of Birkhoff normal forms.

[16] $\{f, g\}_{x,y} = \sum_j f_{x_j} g_{y_j} - f_{y_j} g_{x_j}$, etc.; $\tilde{f}(x, y, z, \bar{z}) = f\left(x, y, \frac{1}{\sqrt{2}}(z + \bar{z}), \frac{1}{\mathrm{i}\sqrt{2}}(z - \bar{z})\right)$, etc.

[17] In other words, the Hamilton equation for $f(x, y, u, v)$ are equivalent to the "Hamilton equation" for $\tilde{f}$ given by $\dot{x} = \tilde{f}_y$, $\dot{y} = -\tilde{f}_x$, $\dot{z} = -\mathrm{i}\tilde{f}_{\bar{z}}$, $\dot{\bar{z}} = \mathrm{i}\tilde{f}_z$.

where we identify the vector $\Omega = (\Omega_1, \dots, \Omega_m)$ with the diagonal matrix

$$\mathrm{diag}(\Omega_1, \dots, \Omega_m)$$

still denoted Ω. The Poisson bracket between $\widetilde{N}$ and an analytic function

$$f(x, y, z, \bar{z}) = \sum_{\substack{k \in \mathbf{Z}^n \\ q, \bar{q} \in \mathbf{N}^m}} f_{kq\bar{q}}(y) e^{\mathrm{i}\langle k, x\rangle} z^q \bar{z}^{\bar{q}}$$

is given by

$$\{\widetilde{N}, f\}^{\sim} = -\mathrm{i} \sum_{\substack{k \in \mathbf{Z}^n \\ q, \bar{q} \in \mathbf{N}^m}} \Big(\langle \omega, k\rangle + \langle \Omega, \bar{q} - q\rangle\Big) f_{kq\bar{q}}(y) e^{\mathrm{i}\langle k, x\rangle} z^q \bar{z}^{\bar{q}}.$$

Let us now fix the norms we shall work with. In $\mathbf{C}^N$ we shall use maximum norm: if $a \in \mathbf{C}^N$, $|a| := \max_i |a_i|$; for Fourier indices $k \in \mathbf{Z}^N$ or Taylor indices $k \in \mathbf{N}^N$, $|k|$ denotes, as usual, $\sum_i |k_i|$. As norms on matrices we take the standard operator norm (with respect to the above maximum norms). Following [21], Hamiltonian functions will be measured by the following *weighted sup-norm*. For $r, s > 0$, let $D(r, s)$ be defined as in (1.13) with u, v replaced, by z, $\bar{z}$ and let

$$\|X_f\|_r := |f_y| + \frac{|f_x|}{r^2} + \frac{|f_z|}{r} + \frac{|f_{\bar{z}}|}{r}, \quad \|X_f\|_{r, D(r,s)} := \sup_{D(r,s)} \|X_f\|_r.$$

The Lipschitz semi-norm with respect to the parameter $\xi \in \Pi$ (or in subsets of Π, which will be clear from context) is defined analogously: [18]

$$\|X_f\|_r^{\mathrm{Lip}} := |f_y|^{\mathrm{Lip}} + \frac{1}{r^2}|f_x|^{\mathrm{Lip}} + \frac{1}{r}|f_z|^{\mathrm{Lip}} + \frac{1}{r}|f_{\bar{z}}|^{\mathrm{Lip}},$$
$$\|X_f\|_{r,\ D(r,s)}^{\mathrm{Lip}} := \sup_{D(r,s)} \|X_f\|_r^{\mathrm{Lip}}.$$

Notational Remark 2.2. *The notation "$a \le$ const b" means "there exists a constant c depending only on n, ℓ and ℓ_* such that $a \le cb$" (obviously in such estimates, the constants c's will be, in general, different one from another). The notation $\|\cdot\|^*$ stands for either $\|\cdot\|$ or $\|\cdot\|^{\mathrm{Lip}}$.*

Since

$$|\mathrm{Im}\, z|, |\mathrm{Im}\, \bar{z}| \le \sigma_v \quad \Longrightarrow \quad |\mathrm{Im}\, u|, |\mathrm{Im}\, v| \le \sqrt{2}\sigma_v,$$

[18] Recall footnote 6.

we see that the functions

$$\widetilde{P}^{(v)}(x, y, z, \bar{z}; \xi) := P^{(v)}\Big(x, y, \frac{z+\bar{z}}{\sqrt{2}}, \frac{z-\bar{z}}{\mathrm{i}\sqrt{2}}; \xi\Big)$$

are analytic and bounded on Δ_{σ_v}. In fact, for any $|\alpha| \leq \ell$, one finds immediately

$$\sup_{\Delta_{\sigma_v}} |\partial^\alpha \widetilde{P}^{(v)}|^* \leq \text{const} \sup_{\Delta_{2\sigma_v}} |\partial^\alpha P^{(v)}|^*.$$

From Lemma 2.1 it follows that the difference $P^{(v)} - P^{(v-1)}$ satisfies

$$\sup_{\Delta_{2\sigma_v}} |P^{(v)} - P^{(v-1)}| \leq 2^{\ell+1} c |P|_{C^\ell} \ \sigma_{v-1}^\ell,$$

which yields

$$\sup_{\Delta_{\sigma_v}} |\partial^\alpha (\widetilde{P}^{(v)} - \widetilde{P}^{(v-1)})|^* \leq \text{const} |P|_{C^\ell}^* \ \sigma_{v-1}^{\ell-|\alpha|}, \quad \forall \ |\alpha| \leq \ell.$$

Notational Remark 2.3. *The KAM algorithm described in Step* 2 *of Section* 1.7 *will be described in terms of the* $(x, y, z, \bar{z})$ *variables but for ease of notation we shall drop systematically the tilde from functions, vector fields and Poisson brackets, keeping in mind the actual meaning just discussed. In the convergence argument, however, we will have to resume the* (x, y, u, v) *variables* (*since the original perturbation function P is only defined for real arguments*). *We shall not come back on this* (*mathematically*) *trivial point, hoping that the notation will cause no confusion.*

2.3. KAM step and the linearized homological equation

As discussed in 1.7, we shall iteratively look for a real-analytic symplectic transformation

$$\Phi_v := \Phi_{v-1} \circ \phi_v = \phi_1 \circ \cdots \circ \phi_v$$

such that, for $v \geq 1$,

$$(N + P^{(v)}) \circ \Phi_v = N_{v+1} + P_{v+1}, \tag{2.1}$$

with N_{v+1} in normal form (as in (1.15)) and P_{v+1} "smaller" than P_v.

Let

$$P_1 := P_1' := \widetilde{P}^{(1)} \tag{2.2}$$

and assume that, for $\nu \geq 1$, P_ν and P_ν' have vector fields real-analytic and bounded in a domain

$$D_\nu := D(r_\nu, s_\nu) \subset \Delta_{\sigma_\nu} \tag{2.3}$$

for suitable numbers (to be specified later)

$$0 < r_\nu < s_\nu < \sigma_\nu < 1. \tag{2.4}$$

We notice (compare also 1.7) that, for $\nu \geq 2$, in view of the form of Φ_ν, Eq. (2.1) can be rewritten as

$$(N_\nu + P_\nu') \circ \phi_\nu = N_{\nu+1} + P_{\nu+1}, \tag{2.5}$$

with

$$P_\nu' := P_\nu + (P^{(\nu)} - P^{(\nu-1)}) \circ \Phi_{\nu-1}.$$

Following [21], we, now, describe how to solve (2.5). For ease of notation, we shall drop, in this section, the index ν and replace the index "$\nu+1$" by the index "+". Therefore, N, P, P', ϕ, $r, \ldots$ stand for N_ν, P_ν, P_ν', ϕ_ν, $r_\nu \ldots$ while N_+, P_+, P_+', ϕ_+, $r_+, \ldots$ stand for $N_{\nu+1}$, $P_{\nu+1}$, $P_{\nu+1}'$, $\phi_{\nu+1}$, $r_{\nu+1}, \ldots$.

The symplectic map $\phi(=\phi_\nu)$ will be taken to be the time-one map of a Hamiltonian flow X_F^t associated to a Hamiltonian function F (with $\|X_F\| \sim \|X_P\| \sim \|X_{P'}\|$). In such a case, the left-hand side of (2.5) takes the form:

$$(N + P') \circ X_F^1 = N + (\{N, F\} + P') + O_2, \tag{2.6}$$

where O_2 denotes (loosely) terms of order two in F. Therefore, the "linearized equation" to be solved for F has the form

$$\{N, F\} + P' = \widehat{N} + O_2, \tag{2.7}$$

where $\widehat{N}$ denotes a term in "normal form"[19] (i.e., having the same form of N). Since one is interested in solving (2.7) in a small neighborhood of $\{y = 0, z = \bar{z} = 0\}$, one

[19] Clearly, the equation $\{N, F\} + P = O_2$ might not have a solution since P, in general, will not belong to the range of the operator $\{N, \cdot\}$.

can truncate the Taylor expansion of P' *up to order one in* y *and up to order two in* $(z,\bar z)$. Also, in order to control the small divisors (for a "large" set of parameter), as in [1], one can truncate the Fourier expansion up to order K. Thus the equation to be solved becomes:

$$\{N,\ \ F\}+R=\widehat{N}, \tag{2.8}$$

where

$$R=\sum_{\substack{2|l|+|q+\bar q|\le 2\\ |k|\le K}} P'_{klq\bar q}\mathrm{e}^{\mathrm{i}\langle k,x\rangle}y^l z^q\bar z^{\bar q} \tag{2.9}$$

(recall that the Fourier–Taylor coefficients of P' are Lipschitz-continuous functions of ξ). Thus, R is a second degree polynomial in $(z,\bar z)$ (and first degree polynomial in y) having the form:

$$R:=R^0+R^1+R^2:=R^0(x,y)+R^1(x,z,\bar z)+R^2(x,z,\bar z), \tag{2.10}$$

where (without indicating explicitly the Lipschitz continuous dependence upon ξ)

$$\begin{aligned} &R^0:=R^{000}(x)+\langle R^{001}(x),y\rangle\ ,\quad R^1:=\langle R^{10}(x),z\rangle+\langle R^{01}(x),\bar z\rangle,\\ &R^2:=\langle R^{20}(x)z,z\rangle+\langle R^{11}(x)z,\bar z\rangle+\langle R^{02}(x)\bar z,\bar z\rangle. \end{aligned} \tag{2.11}$$

We notice (for later reference) that from such definitions there follows

$$P'=R+O(|y|^2)+O(|z|\ |y|)+O(|z|^3), \tag{2.12}$$

so that

$$\begin{aligned} R^{000}&=P'(x,0,0,0),\quad R^{001}=\partial_y P'(x,0,0,0),\\ R^{10}&=\partial_z P'(x,0,0,0),\quad R^{01}=\partial_{\bar z}P'(x,0,0,0),\\ R^{20}&=\frac12\partial_z^2 P'(x,0,0,0),\quad R^{11}=\partial_z\partial_{\bar z}P'(x,0,0,0),\\ R^{02}&=\frac12\partial_{\bar z}^2P'(x,0,0,0). \end{aligned} \tag{2.13}$$

The projection of R onto the kernel of $\{N,\cdot\}$ (sometimes referred to as the "mean value of R") is given by

$$[R]\ =\sum_{|l|+|q|\le 1}P'_{0lqq}y^l z^q\bar z^q=P'_{0000}+\sum_{|l|=1}P'_{0l00}y^l+\sum_{|q|=1}P'_{00qq}z^q\bar z^q$$

$$\begin{aligned} &= R_0^{000} + \langle R_0^{001}, y\rangle + \langle R_0^{11} z, \bar{z}\rangle \\ &:= \hat{e} + \langle \hat{\omega}, y\rangle + \langle \widehat{\Omega} z, \bar{z}\rangle. \end{aligned} \tag{2.14}$$

Therefore, $[R]$ *is in normal form* and we can set

$$\widehat{N} := [R]. \tag{2.15}$$

At this point, recalling Section 2.2, we can easily solve (2.8):

$$F = \sum_{\substack{2|l|+|q+\bar{q}|\le 2 \\ |k|\le K \\ (k,\bar{q}-q)\neq(0,0)}} F_{klq\bar{q}} \mathrm{e}^{\mathrm{i}\langle k,x\rangle} y^l z^q \bar{z}^{\bar{q}}, \quad F_{klq\bar{q}} := \frac{-\mathrm{i} R_{klq\bar{q}}}{\langle \omega, k\rangle + \langle \Omega, \bar{q} - q\rangle}. \tag{2.16}$$

Obviously, *F is real for real argument.*

Having thus defined R, $\widehat{N}$ and F, one can rewrite (2.6) as

$$(N + P') \circ X_F^1 = N_+ + P_+, \tag{2.17}$$

with

$$N_+ := N + \widehat{N}, \tag{2.18}$$

$$P_+ := \int_0^1 \{(1-t)\widehat{N} + tR,\ \ F\} \circ X_F^t \, dt + (P' - R) \circ X_F^1.$$

2.4. Iteration and recursive estimates

In this section, we describe the estimates associated to one step of the KAM iteration described above.

We start by discussing estimates associated to the solution $F_\nu := F$ given in (2.16) (we re-insert the dependence upon the iteration step ν).

Assume the Diophantine condition (1.25) and *assume that*

$$|\omega_\nu|^{\mathrm{Lip}} + |\Omega_\nu|^{\mathrm{Lip}} \le M_\nu, \quad |\omega_\nu^{-1}|^{\mathrm{Lip}} \le L_\nu \tag{2.19}$$

for some positive numbers L_ν, M_ν such that $L_\nu M_\nu \ge 1$ (the Lipschitz semi-norms are taken, respectively, on Π_ν and on $\omega_\nu(\Pi_\nu)$). Then, by classical KAM estimating techniques—mainly based on Cauchy estimates [20] —one finds the following bounds; for details, compare with Section 2 and, in particular with Lemmas 1 and 2 of [21]:

$$\begin{aligned}
\|X_{\widehat{N_\nu}}\|^*_{r_\nu,\ D(r_\nu,s_\nu)} &\le \mathrm{const}\|X_{R_\nu}\|^*_{r_\nu,\ D(r_\nu,s_\nu)}, \qquad (2.20)\\
\|X_{F_\nu}\|_{r_\nu,\ D(r_\nu/2,s_\nu)} &\le \mathrm{const}\,\frac{B_{s_\nu}}{\gamma_\nu}\|X_{R_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)},\\
\|X_{F_\nu}\|^{\mathrm{Lip}}_{r_\nu,\ D(r_\nu/2,s_\nu)} &\le \mathrm{const}\,\frac{B_{s_\nu}}{\gamma_\nu}\Big(\|X_{R_\nu}\|^{\mathrm{Lip}}_{r_\nu,\ D(r_\nu,s_\nu)} + \frac{M}{\gamma_\nu}\|X_{R_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)}\Big),
\end{aligned}$$

where, changing slightly notation with respect to [21] and using Rüssmann's subtle arguments to give optimal estimates of small divisor series (see [23,24]),

$$\begin{aligned}
B_{s_\nu} &:= {\gamma_\nu}^2\sqrt{\sum_{\substack{(k,l)\in\mathbf{Z}^{n+m}\setminus\{0\}\\ |l|\le 2}} \frac{|k|^2}{|\langle\omega_\nu,k\rangle+\langle\Omega_\nu,l\rangle|^4}\ \mathrm{e}^{-2|k|s_\nu}} \le \mathrm{const}\, {s_\nu}^{-\tau_1},\\
\tau_1 &:= 2\tau+1. \qquad (2.21)
\end{aligned}$$

As in [21], we observe that, setting

$$\|\cdot\|^\lambda_r := \|\cdot\|_r + \lambda\|\cdot\|^{\mathrm{Lip}}_r,$$

the second and the third inequality in (2.20) are equivalent to the inequality

$$\|X_{F_\nu}\|^\lambda_{r_\nu,\ D(r_\nu/2,s_\nu)} \le \mathrm{const}\,\frac{B_{s_\nu}}{\gamma_\nu}\|X_{R_\nu}\|^\lambda_{r_\nu,\ D(r_\nu,s_\nu)}, \quad \forall\ 0\le\lambda\le\frac{\gamma_\nu}{M_\nu}.$$

Thus, in view of (2.21), (2.20) may me rewritten more compactly as

$$\|X_{\widehat{N_\nu}}\|^*_{r_\nu,\ D(r_\nu,s_\nu)} \le \mathrm{const}\,\|X_{R_\nu}\|^*_{r_\nu,\ D(r_\nu,s_\nu)}, \qquad (2.22)$$

$$\|X_{F_\nu}\|^\lambda_{r_\nu,\ D(r_\nu/2,s_\nu)} \le \mathrm{const}\,\frac{1}{\gamma_\nu\ {s_\nu}^{\tau_1}}\|X_{R_\nu}\|^\lambda_{r_\nu,\ D(r_\nu,s_\nu)} \qquad (2.23)$$

for any $0\le\lambda\le\gamma_\nu/M_\nu$.

[20] Cauchy estimates give a bound of derivatives of an analytic function on complex domains in terms of the maximum norm of the function in larger domains: if f is analytic on a domain $D\subset\mathbf{C}^k$, then $\sup_{D_{-\delta}}|\partial^\alpha f|\le\alpha!\ \delta^{-|\alpha|}\ \sup_D|f|$, where $D_{-\delta}$ denotes the set of δ-inner points of D (i.e., those points x for which a ball of center x and radius δ is contained in D). For a generalized version, see, e.g., Lemma A4, p. 147, of [21].

To carry on the KAM step *we shall make inductive hypotheses that will be checked in the next section*, where the convergence of the KAM algorithm is discussed.

We assume that P_ν *and* P'_ν *are such that*

$$\|X_{P_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)} + \frac{\gamma_\nu}{M_\nu}\|X_{P_\nu}\|^{\text{Lip}}_{r_\nu,\ D(r_\nu,s_\nu)} \le \frac{\varepsilon_\nu}{2}, \tag{2.24}$$

$$\|X_{P'_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)} + \frac{\gamma_\nu}{M_\nu}\|X_{P'_\nu}\|^{\text{Lip}}_{r_\nu,\ D(r_\nu,s_\nu)} \le \varepsilon_\nu \le \frac{\gamma_\nu {s_\nu}^{\tau_2}\eta_\nu^2}{c_0}, \tag{2.25}$$

where

$$\tau_2 := \tau_1 + 2, \tag{2.26}$$

$c_0 > 1$ *is a suitable constant depending only on* n *and* ℓ_* (*through* τ), $0 < \eta_\nu < 1/16$ *will be a small number* (*to be fixed later*). The role of η_ν will be that of rescaling the y and $z, \bar{z}$-neighborhood of the origin so that terms of order two in y or three in $(z, \bar{z})$ may be "disregarded"(compare with (2.50) below). In the following estimates we shall make repeated use of Cauchy estimates on smaller domains that we shall denotes here, for short, $D^j_\nu := D(2^{-j}r_\nu, 2^{-j}s_\nu),\ j = 1, 2, 3, 4$. Indeed, *we shall take*

$$s_{\nu+1} \le \frac{s_\nu}{16}, \quad r_{\nu+1} := \eta_\nu r_\nu < \frac{r_\nu}{16}, \quad \sigma_{\nu+1} = 2s_{\nu+1}. \tag{2.27}$$

Estimates on the symplectic transformation $\phi_\nu := X^1_{F_\nu}$: Observing that the gradient of R_ν (appearing in the definition of the norm of X_{R_ν}) is defined in terms of derivatives of P'_ν, one gets (by Cauchy [21])

$$\|X_{R_\nu}\|^*_{r_\nu,\ D^1_\nu} \le \text{const}\,\|X_{P'_\nu}\|^*_{r_\nu,\ D(r_\nu,s_\nu)}. \tag{2.28}$$

Recalling (2.23), we find

$$\|X_{F_\nu}\|^\lambda_{r_\nu,\ D^1_\nu} \le \text{const}\,\frac{1}{\gamma_\nu {s_\nu}^{\tau_1}}\|X_{P'_\nu}\|^\lambda_{r_\nu,\ D(r_\nu,s_\nu)}. \tag{2.29}$$

Then, by (2.29) and by assumption (2.24),

$$\|X_{F_\nu}\|_{r_\nu,\ D^1_\nu} \le \text{const}\, s_\nu^2\eta_\nu^2. \tag{2.30}$$

[21] Recall (2.10), (2.11) and (2.13). Then, $|R_{\nu,y}| \le |P'_{\nu,y}|$ on $D(r_\nu, s_\nu)$. Next, $R_{\nu,x}$ is the second order in $(z,\bar{z})$ truncation of $P'_{\nu,x}$- and the estimates on $D(r_\nu, s_\nu/2)$ follows from Cauchy estimates on $(z,\bar{z})$-coefficients. Notice that, in fact, there is no need for such estimates of reducing the x-domain.

Thus, $X^t_{F_\nu} : D^2_\nu \to D^1_\nu$ for all $-1 \le t \le 1$ and, by a standard ODE result,[22] we get

$$\|X^t_{F_\nu} - \mathrm{id}\|_{r_\nu,\ D^2_\nu} \le \mathrm{const}\, s^2_\nu \eta^2_\nu. \tag{2.31}$$

To estimate the derivatives of X_{F_ν}, we recall that, because of the particular structure of F_ν, the x-component of X_{F_ν} is independent of $y,\ u,\ v$, while the $(z,\bar z)$-components are independent of y. Thus, by Cauchy estimates (and recalling that $r_\nu < s_\nu < 1$), we get

$$|\partial X_{F_\nu}|_{D^2_\nu} \le \mathrm{const}\, s_\nu^{-1} \|X_{F_\nu}\|_{r_\nu, D^1_\nu}. \tag{2.32}$$

By the above cited standard ODE result, (2.30) and (2.32) we obtain

$$|X^t_{F_\nu}|^{\mathrm{Lip}}_{D^3_\nu} = |X^t_{F_\nu} - \mathrm{id}|^{\mathrm{Lip}}_{D^3_\nu} \le \mathrm{const}\, s^2_\nu \eta^2_\nu. \tag{2.33}$$

Moreover, by Cauchy estimates, for any $-1 \le t \le 1$ and for $p = 1, 2$,

$$|\partial^p X^t_{F_\nu}|^{\mathrm{Lip}}_{D^4_\nu}\ ,\ \ |\partial^p (X^t_{F_\nu} - \mathrm{id})|^*_{D^4_\nu} \le \mathrm{const}\, s_\nu^{2-p} \eta^2. \tag{2.34}$$

Assuming

$$\mathrm{const}\, \eta^2_\nu < \frac{1}{8} \tag{2.35}$$

(a fact which shall be verified in next section), (2.34) implies that the Jacobian matrix of $X^t_{F_\nu}$, $\partial X^t_{F_\nu}$, is invertible and close to the identity:[23]

$$|(\partial X^t_{F_\nu})^{-1} - I|_{D^4_\nu} \le \mathrm{const}\, s_\nu \eta^2_\nu, \quad \forall -1 \le t \le 1. \tag{2.36}$$

[22] I.e., essentially Gronwall lemma, which, to fit our purposes may be reformulated as follows: *Let V be an open domain in a real Banach space E with norm $\|\cdot\|$, Π a subset of another real Banach space, and $X : V \times \Pi \to E$ a parameter-dependent vector field on V, which is C^1 on V and Lipschitz on Π. Let ϕ^t be its flow. Suppose there is a subdomain $U \subset V$ such that $\phi^t : U \times \Pi \to V$ for $-1 \le t \le 1$. Then*

$$\|\phi^t - \mathrm{id}\|_U \le \|X\|_V, \quad \|\phi^t\|^{\mathrm{Lip}}_U \le \exp(\|\partial X\|_V)\|X\|^{\mathrm{Lip}}_V$$

for $-1 \le t \le 1$, where all norms are understood to be taken also over Π. Notice that $\|\phi^t - \mathrm{id}\|^{\mathrm{Lip}}_U = \|\phi^t\|^{\mathrm{Lip}}_U$. For a (standard) proof, see [21, p. 147]

[23] Use Neumann identity $A^{-1} - I = \sum_{j=1}^\infty (I - A)^j$ valid for any matrix A such that $|I - A| < 1$, ($|\cdot|$ being any operator norm).

Also, from such relation, (2.34) and (2.35), one obtains the following bound on the Lipschitz semi-norm:

$$|(\partial X^t_{F_\nu})^{-1}|^{\mathrm{Lip}}_{D^4_\nu} \le (1 - \mathrm{const}\, s_\nu \eta_\nu^2)^{-2}(1 + \mathrm{const}\, s_\nu \eta_\nu^2) \le 2. \tag{2.37}$$

As already observed above (after (2.30)), $\phi := X^1_{F_\nu} : D^2_\nu \to D^1_\nu$ and, therefore (compare (1.20) and (2.27)),

$$\phi_\nu : D_{\nu+1} \to D_\nu. \tag{2.38}$$

We, now, *make the following inductive assumption* (which shall be easily verified in the next section):

$$|\partial \Phi_\nu|^*_{D_{\nu+1}} \le 2. \tag{2.39}$$

From this assumption it follows immediately that

$$\Phi_\nu(D(r_{\nu+1}, s_{\nu+1})) \subset \Delta_{\sigma_{\nu+1}}, \tag{2.40}$$

completing the proof of (1.20). In fact, suppose that $w = \Phi_\nu(\varsigma)$ with $\varsigma \in D(r_{\nu+1}, s_{\nu+1})$. Since Φ_ν is real for real argument,[24] we have

$$\begin{aligned} |\mathrm{Im}\, w| &= |\mathrm{Im}\, \Phi_\nu(\varsigma)| = |\mathrm{Im}\, \Phi_\nu(\varsigma) - \mathrm{Im}\, \Phi_\nu(\mathrm{Re}\, \varsigma)| \le |\Phi_\nu(\varsigma) - \Phi_\nu(\mathrm{Re}\, \varsigma)| \\ &\le |\partial \Phi_\nu|_{D(r_{\nu+1}, s_{\nu+1})}\ |\mathrm{Im}\, \varsigma| \le 2|\mathrm{Im}\, \varsigma|. \end{aligned}$$

Estimates on $\omega_{\nu+1}$, $\Omega_{\nu+1}$: Recalling (2.15), (2.14) and (2.13), by Cauchy estimates, one finds

$$\begin{aligned} |\hat{e}_\nu|^* &\le \mathrm{const}\, r_\nu^2 \|X_{P'_\nu}\|^*_{r_\nu, D(r_\nu, s_\nu)} \le \mathrm{const}\, \|X_{P'_\nu}\|^*_{r_\nu, D(r_\nu, s_\nu)}, \\ |\widehat{\omega}_\nu|^* &\le \mathrm{const}\, r_\nu \|X_{P'_\nu}\|^*_{r_\nu, D(r_\nu, s_\nu)} \le \mathrm{const}\, \|X_{P'_\nu}\|^*_{r_\nu, D(r_\nu, s_\nu)}, \\ |\widehat{\Omega}_\nu|^* &\le \mathrm{const}\, \|X_{P'_\nu}\|^*_{r_\nu, D(r_\nu, s_\nu)}. \end{aligned} \tag{2.41}$$

[24] Φ_ν is composition of ϕ_ν's $= X^1_{F_\nu}$'s and F_ν is real for real argument (recall (2.16) and the remark after it).

Definition of $\Pi_{\nu+1}$ and small divisor estimates: Recall that on Π_ν the small divisor bound (1.25) holds and *define*

$$\Pi_{\nu+1} := \Pi_\nu \backslash \bigcup_{\substack{(k,l)\in \mathbf{Z}^{n+m}\backslash\{0\} \\ |l|\le 2,\quad |k|>K_\nu}} R^\nu_{kl}(\gamma_\nu), \tag{2.42}$$

where

$$R^\nu_{kl}(\gamma_\nu) := \left\{ \xi \in \Pi_\nu :\ \ |\langle \omega_\nu(\xi), k\rangle + \langle \Omega_\nu(\xi), l\rangle| < \frac{\gamma_\nu}{1+|k|^\tau} \right\}. \tag{2.43}$$

For a given $K_{\nu+1} > K_\nu$ (to be specified later), let $\gamma_{\nu+1}$ be such that [25]

$$\gamma_{\nu+1} \le \gamma_\nu \left(1 - \mathrm{const} \frac{\varepsilon_\nu K^{\tau+1}_{\nu+1}}{\gamma_\nu} \right). \tag{2.44}$$

Then, *for $\xi \in \Pi_{\nu+1}$ the small divisor bound* (1.25) *with ν replaced by $(\nu+1)$ holds*: by (1.25), the definition of $\Pi_{\nu+1}$, (2.41) and (2.44), for all $(k,l) \in \mathbf{Z}^{n+m}\backslash\{0\}$ such that $|l| \le 2$ and $|k| \le K_{\nu+1}$, one has

$$\begin{aligned} &|\langle \omega_{\nu+1}(\xi), k\rangle + \langle \Omega_{\nu+1}(\xi), l\rangle| \\ &\quad \ge |\langle \omega_\nu(\xi), k\rangle + \langle \Omega_\nu(\xi), l\rangle| \left(1 - \frac{|\langle \widehat{\omega}_\nu, k\rangle| + |\langle \widehat{\Omega}_\nu, l\rangle|}{|\langle \omega_\nu, k\rangle| + |\langle \Omega_\nu, l\rangle|} \right) \\ &\quad \ge \frac{\gamma_\nu}{1+|k|^\tau} \left(1 - \mathrm{const} \frac{\varepsilon_\nu K^{\tau+1}_{\nu+1}}{\gamma_\nu} \right) \\ &\quad \ge \frac{\gamma_{\nu+1}}{1+|k|^\tau}. \end{aligned} \tag{2.45}$$

Estimates on $P_{\nu+1}$ and $P'_{\nu+1}$: Recall the definition of the new "perturbation function" $P_{\nu+1}$ given in (2.18). Let us first discuss the term $(P'_\nu - R_\nu) \circ \phi_\nu$ and, in particular, the norm of the "tail" $Q_\nu := P'_\nu - R_\nu$ on a domain slightly larger than $D_{\nu+1}$, namely, $D(r_\nu/2, 4s_{\nu+1})$ (recall (2.27)). First observe that Q_ν has the form

$$Q_\nu := P'_\nu - R_\nu$$

[25] By (2.24), ε_ν is an upper bound on $2\left(\|X_{P_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)} + \frac{\gamma_\nu}{M_\nu} \|X_{P_\nu}\|^{\mathrm{Lip}}_{r_\nu,\ D(r_\nu,s_\nu)} \right)$.

$$= \sum_{2|l|+|q+\bar{q}|>2} P'_{\nu,lq\bar{q}}(x) y^l z^q \bar{z}^{\bar{q}} + \sum_{\substack{|k|>K \\ 2|l|+|q+\bar{q}|\le 2}} P'_{\nu,klq\bar{q}} e^{i\langle k,x\rangle} y^l z^q \bar{z}^{\bar{q}}$$

$$=: Q_\nu^1 + Q_\nu^2. \tag{2.46}$$

Taking into account the dependence on r_ν of the norm $\|\cdot\|_{r_\nu}$, one sees easily that [26]

$$\|X_{Q_\nu^1}\|^*_{\eta_\nu r_\nu,\ D(r_\nu,4\eta_\nu s_\nu)} \le \text{const}\,\eta_\nu \|X_{P_\nu}\|^*_{r_\nu,\ D(r_\nu,s_\nu)}. \tag{2.47}$$

The estimate for Q_ν^2 brings in the dependence upon K_ν (as in [1]) and one finds

$$|\partial Q_\nu^2|^*_{\eta_\nu r_\nu,\ D(r_\nu/2,4\eta_\nu s_\nu)} \le \text{const}\frac{\|X_{P_\nu}\|^*_{r_\nu,\ D(r_\nu,s_\nu)}}{\eta_\nu^2}\ \frac{e^{-(K_\nu s_\nu)/4}}{s_\nu^n}. \tag{2.48}$$

Thus, *assuming*

$$K_\nu \ge \frac{c_1}{s_\nu}\log(\eta_\nu s_\nu)^{-1}, \tag{2.49}$$

with a suitable $c_1 := c_1(n)$, from (2.47) and (2.48) there follows

$$\|X_{P'_\nu - R_\nu}\|^*_{\eta_\nu r_\nu,\ D(r_\nu/2,4\eta_\nu s_\nu)} \le \text{const}\,\eta_\nu \|X_{P'_\nu}\|^*_{r_\nu,\ D(r_\nu,s_\nu)}. \tag{2.50}$$

Now, it is a general fact that, for any functions f and g and for any symplectic map ϕ, the following relations hold: [27]

$$X_{\{f,g\}} = [X_f, X_g] := Jf'' X_g - Jg'' X_f,$$
$$X_{f\circ\phi} = \phi^* X_f := (\partial\phi)^{-1} X_f \circ \phi. \tag{2.51}$$

At this point one has all the ingredients to estimate $\|X_{P_{\nu+1}}\|_{r_{\nu+1},D_{\nu+1}}$, arriving to the following bound holding for any [28] $0 \le \lambda \le \gamma_\nu / M_\nu$

$$\|X_{P_{\nu+1}}\|^\lambda_{r_{\nu+1},D_{\nu+1}} \le \text{const}\left(\frac{1}{\gamma_\nu s_\nu^{\tau_2}\eta_\nu^2}(\|X_{P_\nu}\|^\lambda_{r_\nu})^2 + \eta_\nu \|X_{P_\nu}\|^\lambda_{r_\nu}\right). \tag{2.52}$$

---

[26] $|\partial_x P_{\nu,lq\bar{q}}| \le \|X_{P'_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)}\ r_\nu^{2-(2|l|+|q+\bar{q}|)} \le 2\|X_{P_\nu}\|_{r_\nu,\ D(r_\nu,s_\nu)}\ r_\nu^{2-(2|l|+|q+\bar{q}|)}$.

[27] J denotes the standard symplectic matrix and f'' the Hessian of f.

[28] For full details, see [21, pp. 130–132].

2.5. Convergence

In this section, we iterate the KAM algorithm presented above and show its convergence. Let us introduce the following *recursive parameters* for $\nu \geq 1$. Let $1 < \kappa < 2$, $0 < q < 1$ be suitable constants (to be chosen later); let $c_2 := c_2(n, \ell, \ell_*)$ be a positive large enough constant.[29] Then, for some $0 < \varepsilon_1 \ll 1$ and $r_1 \geq \varepsilon_1^\kappa$ (to be specified later), we set

$$\begin{aligned} M_\nu &:= M(2 - 2^{-\nu+1}), \quad L_\nu := L(2 - 2^{-\nu+1}), \quad \gamma_\nu := \frac{\gamma}{2}(1 + 2^{-\nu+1}), \\ \varepsilon_{\nu+1} &:= \frac{c_2 \varepsilon_\nu^\kappa}{\gamma_\nu^{1/3}}, \quad \sigma_\nu := \varepsilon_\nu^q, \quad s_\nu := \frac{\sigma_\nu}{2}, \quad \eta_\nu := \frac{\varepsilon_\nu^{\kappa-1}}{\gamma_\nu^{1/3}}, \\ r_{\nu+1} &:= \eta_\nu r_\nu, \quad K := c_2 \log \varepsilon_1^{-1}, \quad K_\nu := K \frac{\kappa^\nu}{\sigma_\nu}. \end{aligned} \tag{2.53}$$

Observe that:

$$\begin{aligned} &M := M_1 \leq M_\nu \uparrow 2M, \quad L := L_1 \leq L_\nu \uparrow 2L, \\ &1 > \gamma := \gamma_1 \geq \gamma_\nu \downarrow \gamma_\infty := \frac{\gamma}{2}. \end{aligned} \tag{2.54}$$

Notational Remark 2.4. *In this section the constant c_i will denote suitable constants depending on n, κ, q, ℓ and ℓ_*.*

We shall need some simple relations among the above parameters:

Lemma 2.2. *For any $\nu \geq 1$*

$$\varepsilon_\nu \leq \frac{(A\varepsilon_1)^{\kappa^{\nu-1}}}{A}, \quad A := \frac{c_3}{\gamma^{a_1}}, \tag{2.55}$$

with $a_1 := \frac{1}{(\kappa-1)} > 1$ and $c_3 := (2^{\frac{1}{3}}\ c_2)^{\frac{1}{\kappa-1}}$. Furthermore, if ε_1 is small enough, i.e., if, for a suitable $c_4 \geq c_3$,

$$c_4 \frac{\varepsilon_1}{\gamma^{a_2}} < 1, \quad a_2 := \max\left\{a_1,\ \frac{\kappa}{3(\kappa-1)^2},\ \frac{1}{1 - q\tau_2}\right\}, \tag{2.56}$$

[29] In particular, one can take $c_2 = 16c$ where c denotes here the largest among all constants "const" appearing in the preceding sections.

then, for any $\nu \geq 1$,

$$r_\nu \geq \varepsilon_\nu^\kappa, \qquad \frac{\varepsilon_{\nu+1}}{\varepsilon_\nu} < \frac{1}{16^{\frac{1}{q}}\ 2^{\nu+1}}. \tag{2.57}$$

Proof. From (2.53) and (2.54), it follows that

$$\varepsilon_{\nu+1} \leq \frac{2^{\frac{1}{3}}\ c_2}{\gamma}\ \varepsilon_\nu^\kappa.$$

Iterating such relation one gets (2.55) with $c_3 := (2^{\frac{1}{3}}\ c_2)^{\frac{1}{\kappa-1}}$. As for (2.57), observe that from definitions (2.53) there follows

$$\varepsilon_{\nu+1} = c_2\eta_\nu\varepsilon_\nu = c_2^\nu\left(\prod_{j=1}^{\nu}\eta_j\right)\varepsilon_1, \quad r_{\nu+1} = \left(\prod_{j=1}^{\nu}\eta_j\right)r_1,$$

hence

$$r_{\nu+1} = \varepsilon_{\nu+1}\ \frac{1}{c_2^\nu}\ \frac{r_1}{\varepsilon_1} \tag{2.58}$$

and the first relation in (2.57) is seen to be equivalent to

$$\varepsilon_\nu\ c_2^{\frac{\nu-1}{\kappa-1}} \leq \left(\frac{r_1}{\varepsilon_1}\right)^{\frac{1}{\kappa-1}},$$

which, since $\left(\frac{r_1}{\varepsilon_1}\right)^{\frac{1}{\kappa-1}} \geq \varepsilon_1$, follows from (2.55) and (2.56). From (2.55), choosing c_4 big enough (and since $a_2 > \frac{1}{3(\kappa-1)}$), there follows

$$\begin{aligned} 2^{\nu+1}\frac{\varepsilon_{\nu+1}}{\varepsilon_\nu} &= 2^{\nu+1}c_2\frac{\varepsilon_\nu^{\kappa-1}}{\gamma_\nu^{\frac{1}{3}}} \\ &\leq \left(\frac{1}{16^{\frac{1}{q(\kappa-1)}}}\frac{c_4}{\gamma^{\frac{1}{3(\kappa-1)}}}\ \varepsilon_1\right)^{\kappa-1} \leq \frac{1}{16^{\frac{1}{q}}}. \qquad \Box \end{aligned}$$

Next proposition is a detailed version of the main Theorem 1.1 apart from the claim concerning the measure of Π_∞, which shall be discussed in the next section. To state such proposition we need some definitions. Given γ and M we introduce two numbers,

β and δ, measuring the regularity and certain geometric properties of the perturbation P. Let $\beta > 0$ be such that

$$\max\left\{1,\ \frac{\gamma}{M},\ |P|_{C^\ell},\ \frac{\gamma}{M}\ |P|^{\rm Lip}_{C^\ell}\right\} \le \beta. \tag{2.59}$$

Now, let

$$A_1 := \left\{\alpha:\ |\alpha| = 1 \text{ and } |\partial^\alpha P|_{C^0} \neq 0\right\},$$
$$A_1^{\rm Lip} := \left\{\alpha:\ |\alpha| = 1 \text{ and } |\partial^\alpha P|^{\rm Lip}_{C^0} \neq 0\right\},$$

let, then, $\delta > 0$ be such that

$$\delta := \inf\ \left\{\frac{\inf_{\alpha\in A_1} |\partial^\alpha P|_{C^0}}{\sup_{\alpha\in A_1} |\partial^\alpha P|_{C^0}},\ \frac{\inf_{\alpha\in A_1} |\partial^\alpha P|^{\rm Lip}_{C^0}}{\sup_{\alpha\in A_1} |\partial^\alpha P|^{\rm Lip}_{C^0}}\right\}. \tag{2.60}$$

Finally, let $R_1 := 2r_1$ and

$$D_{R_1} := \left\{(x, y, u, v) \in \mathbf{R}^{2(n+m)}:\ |y| < R_1^2,\ |u| < R_1,\ |v| < R_1\right\} \tag{2.61}$$

and define[30]

$$\varepsilon_0 := \|X_P\|_{R_1, D_{R_1}} + \frac{\gamma}{M}\|X_P\|^{\rm Lip}_{R_1, D_{R_1}}, \quad \hat{\varepsilon}_0 := |X_P|_{D_{R_1}} + \frac{\gamma}{M}|X_P|^{\rm Lip}_{D_{R_1}}. \tag{2.62}$$

Proposition 2.1. *Let $\ell > \ell_* > 6n + 5$; let $\tau := (\ell_* - 11)/6$ and $\tau_2 := (\ell_* - 2)/3$. Let $\theta := \theta(\ell, \ell_*) > 0$ be defined by the relation*[31]

$$\frac{(1+\theta)^2}{1 - 3\theta} = \frac{\ell\ \ - 2}{\ell_* - 2}$$

and define

$$q := \frac{1 - 3\theta}{\tau_2}, \quad \kappa := 1 + \theta.$$

Let[32] *$\omega_1 := \omega$, $\Omega_1 := \Omega$, $L_1 := L$, $M_1 := M$, $\gamma_1 := \gamma$. Assume (2.19) for $v = 1$, let Π_1 such that (1.25) holds for $v = 1$. There exist a constant $c_5 > c_4 > 1$, depending*

[30] Recall the definitions given in Section 2.2 and replace $D(r, s)$ with the *real* set D_{R_1}.
[31] Whence, $\theta \in (0, \frac{1}{3})$.
[32] Beware, instead, that $\Pi_1 \neq \Pi$ and $P \neq P_1$, $K_1 \neq K$.

on n, ℓ and ℓ_* and constants $C_1, C_2 > 1$, depending upon n, ℓ, ℓ_*, (LM), γ, β and δ, such that, if

$$\varepsilon_1 := c_5\ \beta\ \varepsilon_0, \quad C_1\ \varepsilon_0 \leq 1, \quad C_2\ \hat{\varepsilon}_0^{1/(2+\frac{1}{\kappa})} \leq r_1 \leq 1, \tag{2.63}$$

then the following holds. Let M_ν, L_ν, γ_ν, ε_ν, r_ν, s_ν, σ_ν, K_ν be as in (2.53) with ε_1 and r_1 as in (2.63); let D_ν be as in (2.3); let $P^{(\nu)}$ be as in[33] *Section 2.2; let P_1 be as in (2.2). Then, for $\nu \geq 1$, one can iteratively construct, as described in Section 2.3, a sequence of real-analytic symplectic transformations ϕ_ν (and $\Phi_\nu := \phi_1 \circ \cdots \circ \phi_\nu$) satisfying (1.20), and a sequence of functions N_ν, P_ν, P'_ν real-analytic on D_ν satisfying (2.5). The functions indexed by ν are Lipschitz continuous in $\xi \in \Pi_\nu$, where Π_ν is iteratively defined in (2.42). The following conditions hold for any*[34] *ν:*

$$\begin{aligned} &|\omega_\nu|^{\mathrm{Lip}} + |\Omega_\nu|^{\mathrm{Lip}} \leq M_\nu, \quad |\omega_\nu^{-1}|^{\mathrm{Lip}} \leq L_\nu, \\ &\|X_{P_\nu}\|_{r_\nu,\ D_\nu} + \frac{\gamma_\nu}{M_\nu}\|X_{P_\nu}\|^{\mathrm{Lip}}_{r_\nu,\ D_\nu} \leq \frac{\varepsilon_\nu}{2}, \end{aligned} \tag{2.64}$$

$$\|X_{P'_\nu}\|_{r_\nu,\ D_\nu} + \frac{\gamma_\nu}{M_\nu}\|X_{P'_\nu}\|^{\mathrm{Lip}}_{r_\nu,\ D_\nu} \leq \varepsilon_\nu \leq \frac{2\gamma_\nu s_\nu^{\tau_2}\eta_\nu^2}{c_0}, \tag{2.65}$$

as well as conditions (2.4), (2.27), (2.35), (2.39), (2.44) and (2.49). Furthermore, e_ν ($e_1 := 0$), ω_ν and Ω_ν converge (super-exponentially fast) to functions e_∞, ω_∞ and Ω_∞, which are Lipschitz continuous on $\Pi_\infty := \cap \Pi_\nu$ and obey the bounds

$$|\omega_\infty|^{\mathrm{Lip}} + |\Omega_\infty|^{\mathrm{Lip}} \leq 2M, \quad |\omega_\infty^{-1}|^{\mathrm{Lip}} \leq 2L. \tag{2.66}$$

For any

$$2 < p < p_* := 2 + a(\ell - 2), \quad a := \frac{2}{3}\frac{\theta}{(1+\theta)^2}, \tag{2.67}$$

the diffeomorhysms $x \in \mathbf{T}^n \to \Phi_\nu(x, 0, 0, 0; \xi)$ converge in C^p-norm to a C^p-diffeomorphysm $\psi(x; \xi)$, which is Lipschitz continuous in $\xi \in \Pi_\infty$. In fact, for a suitable $c_6 > 1$:

$$|\psi(x;\xi) - x|_{C^p} \leq \frac{c_6}{\gamma^{\frac{2}{3}}}\ \varepsilon_1^{2\theta\frac{p_*-p}{p_*-2}}, \quad \forall\ \xi \in \Pi_\infty; \ \ |\psi|^{\mathrm{Lip}} \leq c_2\frac{\varepsilon_1^{2(q+\theta)}}{\gamma^{\frac{2}{3}}}. \tag{2.68}$$

[33] Recall the Notational Remark 2.3.

[34] Recall (2.24).

Finally, the tori $T^n(\xi)$ defined in (1.24) are invariant tori for $N+P$ and, on such tori, the flow is C^p-conjugated to the Kronecker flow $x \to x+\omega_\infty t$ where ω_∞ verifies the Diophantine relation

$$\begin{aligned} &|\langle \omega_\infty(\xi), k\rangle + \langle \Omega_\infty(\xi), l\rangle| \geq \frac{\gamma}{2(1+|k|^\tau)}, \\ &\forall \ (k,l) \in \mathbf{Z}^{n+m}\backslash\{0\}, \ \ |l| \leq 2, \quad \forall \ \xi \in \Pi_\infty. \end{aligned} \tag{2.69}$$

Proof. As a first step, let us check that the relation between $\hat{\varepsilon}_0$ and r_1 in (2.63), namely

$$C_2 \ \hat{\varepsilon}_0^{1/(2+\frac{1}{\kappa})} \leq r_1 \tag{2.70}$$

implies that:[35]

$$\|X_{\widetilde{P}_1}\|_{r_1,\ D_1} + \frac{\gamma}{M}\|X_{\widetilde{P}_1}\|^{\mathrm{Lip}}_{r_1,\ D_1} \leq \varepsilon_1. \tag{2.71}$$

Notice that, by the definition of norms and complex variables in Section 2.2, it follows that

$$\|X_{\widetilde{P}_1}\|_{r_1,\ D_1} \leq 2\|X_{P_1}\|_{R_1, D(R_1,s_1)}, \tag{2.72}$$

so that, in the following argument, we may use directly the (x,y,u,v) variables. Introduce, also, for the purpose of this check, the short-hand notation "$|\cdot|^\bullet$" to denote either "$|\cdot|$" or "$(\gamma/M)|\cdot|^{\mathrm{Lip}}$" and observe that from the definitions of δ ((2.60)) and $\hat{\varepsilon}_0$ ((2.62)), it follows that

$$\delta \ \hat{\varepsilon}_0 \leq |\partial^\alpha P|^\bullet_{C^0}, \quad \forall \ \alpha \in A_1. \tag{2.73}$$

Observe, also, that, if

$$C_2 \geq \mathrm{const}\frac{\beta^{\frac{1}{2}}}{\delta^{\frac{1}{2\bar{q}}}}, \quad \bar{q} := q(\ell-1) := 3(1+\theta)^2\frac{\ell-1}{\ell-2}$$

(for a suitable const), then, since (as it easy to check)

$$\frac{\bar{q}-1}{2\bar{q}} > \frac{1}{2+\frac{1}{\kappa}},$$

[35] Only for the purpose of this check we re-introduce tildas to distinguish between functions of $(x,y,z,\bar{z})$ and functions of (x,y,u,v); recall the Notational Remark 2.3.

Eq. (2.70) yields

$$\text{const}\,\frac{\beta^{\bar q}}{\delta}\ \hat\varepsilon_0^{\bar q-1}\le r_1^{2\bar q}. \tag{2.74}$$

Thus, taking into account the weight of the norm $\|\cdot\|_{R_1}$ appearing in the definition of ε_0, recalling the definitions of $\sigma_1=\varepsilon_1^q$, ε_1, $\bar q$, (2.74) and (2.73), we find, for all $\alpha\in A_1$,

$$\begin{aligned}\sigma_1^{\ell-1} &:= \varepsilon_1^{\bar q}=\text{const}\,\beta^{\bar q}\varepsilon_0^{\bar q}\le\text{const}\,\beta^{\bar q}\ \frac{\hat\varepsilon_0^{\bar q}}{r_1^{2\bar q}}\\ &\le\ \text{const}\,\hat\varepsilon_0\ \delta\le\text{const}|\partial^\alpha P|^\bullet_{C^0}.\end{aligned} \tag{2.75}$$

Now, if $\zeta:=(x,y,u,v)\in\Delta_{\sigma_1}$, by Lemma 2.1, the definition of β in (2.59), the convexity estimates in Remark 2.1 and (2.75) we find, for any $\alpha\in A_1$:

$$\begin{aligned}|\partial^\alpha P_1(\zeta)|^\bullet &\le \left|\partial^\alpha P_1(\zeta)-\sum_{|\beta|\le\ell-1}\frac{\partial^{\beta+\alpha}P(\operatorname{Re}\zeta)}{\beta!}\ (\mathrm{i}\operatorname{Im}\zeta)^\beta\right|^\bullet\\ &\quad+\left|\sum_{|\beta|\le\ell-1}\frac{\partial^{\beta+\alpha}P(\operatorname{Re}\zeta)}{\beta!}\ (\mathrm{i}\operatorname{Im}\zeta)^\beta\right|^\bullet\\ &\le c\beta\sigma_1^{\ell-1}+c\sum_{m=0}^{\ell-1}|\partial^\alpha P|^\bullet_{C^m}\ \sigma_1^m\\ &\le c\beta\sigma_1^{\ell-1}+\text{const}\sum_{m=0}^{\ell-1}\left(|\partial^\alpha P|^\bullet_{C^0}\right)^{\frac{\ell-1-m}{\ell-1}}\left(|\partial^\alpha P|^\bullet_{C^{\ell-1}}\right)^{\frac{m}{\ell-1}}\sigma_1^m\\ &\le\text{const}\,\beta\sum_{m=0}^{\ell-1}\left(|\partial^\alpha P|^\bullet_{C^0}\right)^{\frac{\ell-1-m}{\ell-1}}\ \sigma_1^m\\ &\le\text{const}\,\beta\sum_{m=0}^{\ell-1}\left(|\partial^\alpha P|^\bullet_{C^0}\right)^{\frac{\ell-1-m}{\ell-1}}\ \left(|\partial^\alpha P|^\bullet_{C^0}\right)^{\frac{m}{\ell-1}}\\ &\le\text{const}\,\beta\ |\partial^\alpha P|^\bullet_{C^0}.\end{aligned}$$

From this relation, (2.72) and the definition of ε_1, we find immediately [36]

$$\|X_{\widetilde P_1}\|_{r_1,\ D_1}\le\text{const}\,\beta\varepsilon_0:=\varepsilon_1. \tag{2.76}$$

[36] From the definition of P_1 it follows that if $|\partial^\alpha P|^*_{C^0}=0$, for some α, then also $|\partial^\alpha P_1|^*_{\Delta_{\sigma_1}}=0$.

Recall (compare sentence before (2.53)) that we have to check that $r_1 \geq \varepsilon_1^\kappa$ or, equivalently, $\varepsilon_1 r_1^{-\frac{1}{\kappa}} \leq 1$: in fact, from the definition of ε_1 and from (2.63), there follows

$$\frac{\varepsilon_1}{r_1^{\frac{1}{\kappa}}} := \text{const}\,\beta\,\frac{\hat{\varepsilon}_0}{r_1^{2+\frac{1}{\kappa}}} = \text{const}\frac{\beta^{\frac{1}{2+\frac{1}{\kappa}}}\ \hat{\varepsilon}_0^{\frac{1}{2+\frac{1}{\kappa}}}}{r_1} \leq 1$$

provided

$$C_2 \geq \text{const}\ \beta^{\frac{1}{2+\frac{1}{\kappa}}}.$$

To proceed, it is convenient to reformulate the smallness condition, $C_1\varepsilon_0 \leq 1$, on ε_0 (which will not appear any more in the sequel) in terms of ε_1. It is easily seen that $C_1\varepsilon_0 \leq 1$ implies that[37]

$$\begin{aligned} &c_7\ \beta^{\frac{1}{\theta}}\ (LM)\ \frac{\varepsilon_1\ (\log \varepsilon_1^{-1})^{2(\tau+1)}}{\gamma^{a_3}} < 1, \\ &a_3 := \max\left\{a_2,\ \frac{2}{3\theta},\ \frac{3\kappa-1}{\kappa(\kappa-1)}\right\} \end{aligned} \tag{2.77}$$

for a suitable $c_7 > c_5$. Notice that (2.77), in turn, implies (2.56). Next, the inequality

$$\varepsilon_1 \leq \frac{\gamma s_1^{\tau_2}\eta_1^2}{c_0}$$

is equivalent to

$$\frac{2^{\tau_2}c_0}{\gamma^{\frac{2}{3}}}\ \varepsilon_1^{1-q\tau_2-2(\kappa-1)} := \frac{2^{\tau_2}c_0}{\gamma^{\frac{2}{3}}}\ \varepsilon_1^{\theta} \leq 1,$$

which follows from the smallness condition (2.77) (and the fact that $a_3 \geq 2/(3\theta)$). Thus (2.24) and (2.25) are satisfied for $v = 1$ and the KAM iterative procedure, discussed in the previous sections, can be turned on.

[37] For example, one can take $C_1 > \bar{C}_1$, where $\bar{C}_1 > \text{const}\, c_5\beta^{1+\frac{1}{\theta}}\frac{LM}{\gamma^{a_3}}$ and the constant $C_1 \geq \exp(2(\tau+1))$ solves $\log C_1 = (C_1/\bar{C}_1)^{\frac{1}{2}}(\tau+1)$.

We shall, now, proceed to check all iterative conditions claimed in the thesis of the proposition.

(2.4): First notice that (2.4) (the only nontrivial part of which is $r_\nu < s_\nu$) for $\nu = 1$ holds because $\kappa > 1 > q$; to check (2.4) for $\nu > 1$ use (2.58).

(2.27) and (2.35): $s_{\nu+1} \le s_\nu/16$ is equivalent to $\varepsilon_{\nu+1} \le \varepsilon_\nu/16$, which is implied by (2.57). Also, from the definition of $\varepsilon_{\nu+1}$, η_ν and (2.57) it follows that

$$\eta_\nu = \frac{\gamma_\nu^{\frac{1}{3}}}{c_2} \; \frac{\varepsilon_{\nu+1}}{\varepsilon_\nu} < \frac{1}{2^{\nu+1} \; c_2}, \tag{2.78}$$

which implies (2.27) and (2.35) because of the definition of c_2.

(2.39) is consequence of[38] (2.34) and (2.78):

$$\begin{aligned} |\partial \Phi_\nu|_{D_{\nu+1}} &= |(\partial\phi_1 \circ \phi_2 \circ \cdots \circ \phi_\nu) \; (\partial\phi_2 \circ \phi_3 \circ \cdots \circ \phi_\nu) \; \cdots (\partial\phi_\nu)|_{D_{\nu+1}} \\ &\le \prod_{j=1}^{\nu} (1 + \mathrm{const}\, s_j \eta_j^2) < \prod_{j=1}^{\nu} \left(1 + \frac{1}{2^{j+1}}\right) < 2. \end{aligned}$$

Similarly one obtains[39]

$$\begin{aligned} |\partial^2 \Phi_\nu|_{D_{\nu+1}} &= \Bigg| \sum_{j=1}^{\nu} (\partial^2 \phi_j \circ \phi_{j+1} \circ \cdots \circ \phi_\nu)(\partial\phi_{j+1} \circ \phi_{j+2} \circ \cdots \circ \phi_\nu) \; \cdots (\partial\phi_\nu) \\ &\quad \times \prod_{i \ne j} (\partial\phi_i \circ \phi_{i+1} \circ \cdots \circ \phi_\nu) \Bigg|_{D_{\nu+1}} \\ &\le 4\nu. \end{aligned} \tag{2.79}$$

Now, assume, by induction up to $j = \nu - 1$, that $|\partial \Phi_j|^{\mathrm{Lip}}_{D_{j+1}} \le 1 + \alpha_j := 2 - \frac{1}{2^j}$ (which, for $j = 1$ is certainly true, in view of (2.34), since $\Phi_1 := \phi_1$). Then (shortening, here, "const" with "c", using again (2.34), the smallness of η_ν, (2.78) and (2.79)),

$$\begin{aligned} &\frac{|\partial \Phi_\nu(\cdot, \xi') - \partial \Phi_\nu(\cdot, \xi)|}{|\xi' - \xi|} \\ &\quad = \frac{|\partial \Phi_{\nu-1}(\phi_\nu(\cdot, \xi'), \xi')\partial\phi_\nu(\cdot, \xi') - \partial \Phi_{\nu-1}(\phi_\nu(\cdot, \xi), \xi)\partial\phi_\nu(\cdot, \xi)|}{|\xi - \xi'|} \end{aligned}$$

[38] Recall that $\phi_\nu = X^1_{F_\nu}$ and (2.34), one sees that $|\partial\phi_\nu|_{D_{\nu+1}} \le 1 + \mathrm{const}\, s_\nu \eta_\nu^2$.

[39] Use that from (2.34) with $p = 2$ there follows that $|\partial^2 \Phi_j|_{D_{j+1}} \le \mathrm{const}\, \eta_j^2$.

$$
\begin{aligned}
&\leq \frac{|\partial\Phi_{v-1}(\phi_v(\cdot,\xi'),\xi') - \partial\Phi_{v-1}(\phi_v(\cdot,\xi'),\xi)|}{|\xi'-\xi|}\ |\partial\phi_v(\cdot,\xi')| \\
&\quad + \frac{|\partial\Phi_{v-1}(\phi_v(\cdot,\xi'),\xi) - \partial\Phi_{v-1}(\phi_v(\cdot,\xi),\xi)|}{|\xi'-\xi|}\ |\partial\phi_v(\cdot,\xi')| \\
&\quad + |\partial\Phi_{v-1}(\phi_v(\cdot,\xi),\xi)|\ \frac{|\partial\phi_v(\cdot,\xi') - \partial\phi_v(\cdot,\xi)|}{|\xi'-\xi|} \\
&\leq (1+\alpha_{v-1})(1+cs_v\eta_v^2) + |\partial^2\Phi_{v-1}|\ |\phi_v|^{\mathrm{Lip}}\ |\partial\phi_v| + 2c\ s_v\eta_v^2 \\
&\leq (1+\alpha_{v-1})(1+cs_v\eta_v^2) + 4vc^2 s_v^2\eta_v^2\ (1+cs_v\eta_v^2) + 2cs_v\eta_v^2 \\
&\leq 1+\alpha_v,
\end{aligned}
$$

last inequality follows easily by the smallness of η_v.

(1.20): recall from Section 2.5 that (2.38) holds because of (2.35) and that (2.40) is consequence of (2.39) (and recall also that (1.20) is (2.38) plus (2.40)).

(2.44): Since $\frac{\gamma_{v+1}}{\gamma_v} = 1 - \frac{1}{2+2^v}$, (2.44) is implied by

$$
\text{const}\,\frac{\varepsilon_v K_{v+1}^{\tau+1}}{\gamma_v} \leq \frac{1}{2^{v+1}},
$$

which is seen to hold because of the definition of K_v, (2.55), the fact that $1-q(\tau+1)\kappa > 1/2$ (recall the definition of q and κ in Lemma 2.1) and (2.77).

(2.49) follows from the definition of K_v, the fact that $\varepsilon_{v+1} \geq \varepsilon_v^{\kappa} \geq \varepsilon_1^{\kappa^v}$ and the explicit definition of K (*used only here*), $K = c_2 \log \varepsilon_1^{-1}$.

Second inequality in (2.65): Using the definitions of s_v, η_v and the fact that $1 - q\tau_2 - 2\kappa = \theta$, one sees that the claim follows from

$$
\text{const}\,\frac{\varepsilon_v}{\gamma^{\frac{1}{3}}} < 1,
$$

which in turn (using (2.55) and the fact that $a_3 \geq \frac{1}{3\theta}$) is implied by (2.77).

(2.64) for $v > 1$ is proven by induction: Assume it holds up to v. Then observing that $M_{v+1} - M_v = M/2^v$, using the bounds (2.41), the fact that $\|X_{P_v'}\|_{r_v,D_v}^{\mathrm{Lip}} \leq M_v\varepsilon_v/\gamma_v$ (see (2.65)) and the fact that $a_3 \geq (3\kappa-1)/(\kappa(\kappa-1))$, the first of (2.64) is seen to follow from (2.77). To check the second inequality in (2.64), observe that

$$
|\omega_{v+1}^{-1}|^{\mathrm{Lip}} \leq \frac{L_v}{1 - L_v|\widehat{\omega}_v|^{\mathrm{Lip}}} \leq \frac{L_v}{1 - \text{const}\, L_v M_v \varepsilon_v/\gamma_v}.
$$

Thus, the claim follows from the smallness assumption (2.77) (*it is only here that the presence of the term* (LM) *in* (2.77) *is used*), since $a_3 \geq (3\kappa-1)/(\kappa(\kappa-1)) >$

$a_1 + \frac{1}{\kappa} = (2\kappa - 1)/(\kappa(\kappa - 1))$. We turn to the third relation in (2.65). By (2.52) and using the fact that $2 - q\tau_2 - 2(\kappa - 1) = \kappa$ one sees that the claim follows from the definition of ε_{v+1}.

First inequality in (2.65) (for $v \geq 2$): For the purpose of this check call

$$\widehat{P}_v := (P^{(v)} - P^{(v-1)}) \circ \Phi_{v-1}.$$

In view of the already verified bound (2.64), the claim is implied by

$$\|X_{\widehat{P}_v}\|_{r_v, D_v} + \frac{\gamma_v}{M_v}\|X_{\widehat{P}_v}\|^{\mathrm{Lip}}_{r_v, D_v} \leq \frac{\varepsilon_v}{2}. \tag{2.80}$$

Observe, as above, that, by definition of Hamiltonian vector field and of our weighted norms, $\|X_{\widehat{P}_v}\|^*_{r_v, D_v} \leq r_v^{-2}\ |\partial \widehat{P}_v|^*_{D_v}$. Now, by (2.39) and Section 2.2, (on the proper domains),

$$|\partial \widehat{P}_v| \leq |\partial(P^v - P^{v-1})|\ |\partial \Phi_{v-1}| \leq \mathrm{const}\,|P|_{C^\ell}\sigma_{v-1}^{\ell-1}. \tag{2.81}$$

To bound the Lipschitz part, first observe that, by the chain rule, by (2.33), the fact that $q + \kappa > 1$ and (2.77),

$$\begin{aligned}
|\Phi_v|^{\mathrm{Lip}} &= |\Phi_{v-1}(\phi_v, \xi)|^{\mathrm{Lip}} \leq |\partial \Phi_{v-1}|\ |\phi_v|^{\mathrm{Lip}} + |\Phi_{v-1}|^{\mathrm{Lip}} \\
&\leq |\Phi_{v-1}|^{\mathrm{Lip}} + \mathrm{const}\, s_v^2 \eta_v^2 \\
&\leq |\Phi_1|^{\mathrm{Lip}} + \mathrm{const} \sum_{j=2}^{v-1} s_v^2 \eta_v^2 \leq \mathrm{const} \sum_{j=1}^{v-1} s_v^2 \eta_v^2 \\
&\leq \mathrm{const}\, \varepsilon_1^{2(q+\kappa)} \leq \varepsilon_1.
\end{aligned} \tag{2.82}$$

Now, by the chain rule, (2.39), Section 2.2, (2.82), (on the proper domains),

$$\begin{aligned}
|\partial \widehat{P}_v|^{\mathrm{Lip}} &= |\partial(P^v - P^{v-1}) \cdot \partial \Phi_{v-1}|^{\mathrm{Lip}} \\
&\leq \left| \Big(\partial(P^{(v)} - P^{(v-1)}) \Big) \circ \Phi_{v-1} \right|^{\mathrm{Lip}} |\partial \Phi_{v-1}| \\
&\quad + \left| \Big(\partial(P^{(v)} - P^{(v-1)}) \Big) \circ \Phi_{v-1} \right| |\partial \Phi_{v-1}|^{\mathrm{Lip}} \\
&\leq \mathrm{const}\ \Big(|P|_{C^\ell}\sigma_{v-1}^{\ell-1}| + |\Big(\partial(P^{(v)} - P^{(v-1)})\Big) \circ \Phi_{v-1}|^{\mathrm{Lip}} \Big)
\end{aligned}$$

$$\begin{aligned}
&\leq \text{ const } \left(|P|_{C^\ell} \sigma_{v-1}^{\ell-1} + |\partial^2 (P^{(v)} - P^{(v-1)})| \; |\Phi_{v-1}|^{\text{Lip}} \right) \\
&\quad + \partial (P^{(v)} - P^{(v-1)})|^{\text{Lip}} \\
&\leq \text{ const } \left((|P|_{C^\ell} + |P|_{C^\ell}^{\text{Lip}}) \sigma_{v-1}^{\ell-1} + |P|_{C^\ell} \sigma_{v-1}^{\ell-2} \varepsilon_1 \right). \qquad (2.83)
\end{aligned}$$

Putting together (2.81) and (2.83), and using (2.77), the first inequality in (2.57), the relation $\varepsilon_{v-1}^{\kappa} \leq \varepsilon_v$ and the fact that $\frac{q(\ell-2)}{\kappa} - 2\kappa = (1+\theta) > 1$, one gets

$$\|X_{\widehat{P}_v}\|_{r_v, D_v} + \frac{\gamma_v}{M_v} \|X_{\widehat{P}_v}\|_{r_v, D_v}^{\text{Lip}} \leq (\beta \varepsilon_1^q) \; \frac{\sigma_{v-1}^{\ell-2}}{r_v^2} \leq \varepsilon_v^{1+\theta} < \frac{\varepsilon_v}{2},$$

which is (2.80).

The convergence of[40] e_v, ω_v and Ω_v to e_∞, ω_∞ and Ω_∞ is, at this point, proved, as well as the bounds (2.66), which follows at once from (2.64).

First estimate in (2.68):

Write $\Phi_v = \phi_1 + \sum_{j=2}^{v} (\Phi_j - \Phi_{j-1})$ and introduce, here, the short-hand notation $\Phi_j^0(x;\xi) := \Phi_j(x,0,0,0;\xi)$ and $\phi_j^0(x;\xi) := \phi_j(x,0,0,0;\xi)$ so that $\psi(x;\xi) = \lim_{v\to\infty} \Phi_v^0(x;\xi)$. Notice that, for $|\text{Im}\, x| \leq s_j$, by (2.39) and (2.31), one has

$$|\Phi_{j-1}(\phi_j^0(x;\xi)) - \Phi_{j-1}^0(x;\xi)| \leq \sup_{|\text{Im}\, x| \leq s_j} |\partial \Phi_{j-1}| \; |\phi_j^0(x) - x| \leq \text{const}\, \sigma_j^2 \eta_j^2.$$

Then, for any $x \in \mathbf{T}^n$ and $\xi \in \Pi_\infty$, for any $\alpha \in \mathbf{N}^n$ with $|\alpha| \leq p$, by Cauchy estimates, by the definitions of s_j, η_j, q and[41] κ, we have

$$\begin{aligned}
&\left| \partial_x^\alpha \left(\psi(x;\xi) - x \right) \right| \\
&\leq \left| \partial_x^\alpha \left(\phi_1(x;\xi) - x \right) \right| + \sum_{j=2}^{\infty} \left| \partial_x^\alpha \left(\Phi_{j-1}(\phi_j^0(x;\xi)) - \Phi_{j-1}^0(x;\xi) \right) \right| \\
&\leq \text{const} \sum_{j=1}^{\infty} s_j^{2-|\alpha|} \eta_j^2 \leq \frac{\text{const}}{\gamma^{\frac{2}{3}}} \sum_{j=1}^{\infty} \varepsilon_j^{q(2-p)+2(\kappa-1)} \\
&= \frac{\text{const}}{\gamma^{\frac{2}{3}}} \sum_{j=1}^{\infty} \varepsilon_j^{2\theta + q(2-p)} = \frac{\text{const}}{\gamma^{\frac{2}{3}}} \sum_{j=1}^{\infty} \varepsilon_j^{2\theta \, \frac{p_*-p}{p_*-2}} \leq \frac{\text{const}}{\gamma^{\frac{2}{3}}} \varepsilon_1^{2\theta \, \frac{p_*-p}{p_*-2}}.
\end{aligned}$$

[40] Observe that e_v obey the same bound of $\widehat{\omega}_v$ so that its convergence follows from the above discussion; in any case e_∞ has no dynamical relevance.

[41] Note: $a(\ell-2) = 2\tau_2 \frac{\theta}{1-3\theta} = 2\frac{\theta}{q}$, $2\theta + q(2-p) = 2\theta \frac{p_*-p}{p-2}$.

For the bound on the Lipschitz semi-norm just take the limit in (2.82). Finally, the Diophantine relation (2.69) is obtained as the limiting case of (1.25). $\square$

2.6. Measure estimates (multiplicity of solutions)

In this section, *assuming the notations and hypotheses of Proposition* 2.1, we shall prove and make quantitative the claims in Theorem 1.1 concerning the measure of Π_∞, hence establishing multiplicity results for the lower-dimensional quasi-periodic solutions found in Proposition 2.1.

Following [21], we note that if $|k|$ is large, then the discarted "resonant set" $R^\nu_{kl}(\gamma_\nu)$ defined in (2.43) is small: [42]

Lemma 2.3. *If $|k| \geq K_0 := 16LM$, then, for any $\nu \geq 1$ and any $|l| \leq 2$,*

$$\mathrm{meas}(R^\nu_{kl}(\gamma_\nu)) \leq \frac{\lambda}{|k|^{\tau+1}}, \quad \lambda := \mathrm{const}(LM)^n \frac{\gamma}{M} (\mathrm{diam}\, \Pi)^{n-1}. \tag{2.84}$$

This lemma is essentially Lemma 5, p. 136, in [21], to which we refer for the simple proof. [43]

Proposition 2.2. *Assume that ε_1 satisfies also*

$$\varepsilon_1 (LM)^a < 1, \quad a := \max\left\{\frac{1}{\theta}, \frac{1}{q(\tau - n + 1)}\right\} \tag{2.85}$$

and that

$$0 < \gamma < \min_{\substack{\xi\in\Pi \\ i\neq j}} \{|\Omega_i(\xi)|, \ |\Omega_i(\xi) - \Omega_j(\xi)|\}. \tag{2.86}$$

Then,

$$\mathrm{meas}\, \Pi_\infty \geq \mathrm{meas}\, \Pi_0 - \mathrm{const}\, \frac{\gamma}{M} (LM\, \mathrm{diam}\, \Pi)^{n-1}, \tag{2.87}$$

[42] Recall that $\tau > n - 1$.

[43] Just for completeness we sketch here an alternative argument: ω_ν and Ω_ν are Lipschitz in ξ and in fact ω_ν is a Lipschitz diffeomorphysm. Thus, such function have derivatives in L^1 and the standard formula for the change of variables in integrations holds. Using $\omega = \omega_\nu(\xi)$ as independent variable, *up to a suitable k-dependent rotation*, we see that it is enough to estimate sets of the form $\{\omega \in \omega_\nu(\Pi_\nu) : \ |\omega_1 + g_k(\omega)| < \gamma_k/|k|^{\tau+1}\}$ where g_k is a Lipschitz function that because of the assumption on $|k|$ is smaller than, say, 1/2. Now, make a further change of variables setting $\omega_1' = \omega_1 + g_k(\omega)$, $\omega_2' = \omega_2, \ldots, \omega_n' = \omega_n$, etc.

where the set $\Pi_0 := \Pi_0(\gamma)$ *is defined as*

$$\Pi_0 := \left\{ |\langle \omega(\xi), k\rangle + \langle \Omega(\xi), l\rangle| \geq \frac{\gamma}{1+|k|^\tau}, \ \forall\ 0 < |k| \leq K_0,\ \ |l| \leq 2 \right\}.$$

Furthermore,

$$\lim_{\gamma\downarrow 0} \text{meas}\,(\Pi\backslash\Pi_0(\gamma)) = 0, \tag{2.88}$$

showing that meas $\Pi_\infty > 0$ *provided* γ *is small enough. Finally, if* ω *and* Ω *are* $C^1(\Pi)$ *and if* (*taking* ω *as independent variable*[44])

$$\begin{aligned} \mu &:= \min_{\substack{0<|k|\leq K_0,|l|\leq 2\\ \omega\in S_{kl}}} \left(|k|^{-1} \left| k + \frac{\partial\langle\Omega, l\rangle}{\partial\omega} \right| \right) > 0, \\ S_{kl} &:= \{\omega \in \omega(\Pi) :\ \ \langle\omega, k\rangle + \langle\Omega(\omega), l\rangle = 0\}. \end{aligned} \tag{2.89}$$

then

$$\text{meas}\ (\Pi\backslash\Pi_0(\gamma)) \leq \text{const}\ \frac{\gamma}{M\mu}\ (LM\,\text{diam}\,\Pi)^{n-1}. \tag{2.90}$$

Remark 2.2. Recall point 1.6 in Section 1 and especially (1.11) and let $r := r_1$. Notice that, in such a case, $\hat\varepsilon_0 \sim r_1^3 + \varepsilon$ and $\varepsilon_0 \sim r_1 + \frac{\varepsilon}{r_1}$. Thus, choosing $r := r_1 := \varepsilon^{\frac{1}{3}}$, we see that $\varepsilon_0 \sim \varepsilon_1 \sim \varepsilon^{\frac{1}{3}}$ and that hypotheses (2.63) and (2.85) are satisfied and the claim in 1.6 follows; "genericity" refers to conditions (1.3)–(1.4).

Proof. Notice that by definition of K_ν in (2.53) and (2.85), there follows that $K_\nu \geq K_1 > K_0 := 16(LM)$. Thus, by Lemma 2.3 and the definition of $\Pi_{\nu+1}$,

$$\begin{aligned} \text{meas}(\Pi_{\nu+1}) &\geq \text{meas}(\Pi_\nu) - \sum_{\substack{|l|\leq 2\\ |k|>K_\nu}} \text{meas}(R^\nu_{kl}(\gamma_\nu)) \\ &\geq \text{meas}(\Pi_\nu) - \text{const}\,\lambda \sum_{|k|>K_\nu} |k|^{-\tau+1} \\ &\geq \text{meas}(\Pi_\nu) - \text{const}\,\lambda \frac{1}{K_\nu^{\tau-n+1}}. \end{aligned}$$

[44] I.e., $\Omega(\omega)$ is, by definition, $\Omega(\xi(\omega))$ where $\omega \to \xi(\omega)$ is the C^1 inverse function of $\xi \to \omega(\xi)$.

Iterating this relation, using the definition of K_ν and (2.85), we get

$$\begin{aligned}\mathrm{meas}(\Pi_{\nu+1}) &\geq \mathrm{meas}\,(\Pi_1) - \mathrm{const}\,\lambda\varepsilon_1^{q(\tau-n+1)}\\ &\geq \mathrm{meas}(\Pi_1) - \mathrm{const}\,\frac{\gamma}{M}\,(LM\,\mathrm{diam}\,\Pi)^{n-1},\end{aligned}$$

which implies

$$\mathrm{meas}(\Pi_\infty) \geq \mathrm{meas}(\Pi_1) - \mathrm{const}\,\frac{\gamma}{M}\,(LM\,\mathrm{diam}\,\Pi)^{n-1}. \tag{2.91}$$

From (2.86) it follows that

$$\Pi_1 = \Pi_0 \Big\backslash \bigcup_{\substack{K_0<|k|\leq K_1\\ |l|\leq 2}} R^1_{kl}(\gamma)$$

and we see, again by Lemma 2.3, that

$$\mathrm{meas}(\Pi_1) \geq \mathrm{meas}(\Pi_0) - \mathrm{const}\,\lambda(LM)^{-1},$$

which, together with (2.91), implies (2.87).

The claim in (2.88) follows immediately from the compactness of Π, assumption (1.4) and the “monotonicity” of the sets $R^\nu_{kl}(\gamma)$ in γ (i.e., $R^\nu_{kl}(\gamma) \subset R^\nu_{kl}(\gamma')$ if $\gamma < \gamma'$).

The claim in (2.90) follows easily by noting that (2.89) implies that S_{kl} are C^1 hyper-surfaces in $\omega(\Pi)$ and observing that μ is a lower bound on the norm of the gradient of the function $\langle\omega, k\rangle + \langle\Omega(\omega), l\rangle$. $\square$

References

[1] 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)) (1963) 13–40 (Russian).

[2] J. Bourgain, On Melnikov’s persistency problem, Math. Res. Lett. 4 (4) (1997) 445–458.

[3] J. Bourgain, Quasi-periodic solutions for Hamiltonian perturbations for 2D linear Schrödinger equations, Ann. of Math. 148 (1998) 363–439.

[4] L. Chierchia, J. You, KAM tori for 1D nonlinear wave equations with periodic boundary conditions, Comm. Math. Phys. 211 (2000) 497–525.

[5] W. Craig, C. Wayne, Newton’s methods and periodic solutions of nonlinear wave equation, Comm. Pure Appl. Math. 46 (1993) 1409–1501.

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

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

[8] M.-R. Herman, Sur les courbes invariantes par les difféomorphismes de l'anneau. Vol. 1. (French) [On the curves invariant under diffeomorphisms of the annulus. Vol. 1] With an appendix by Albert Fathi. With an English summary. Astérisque, 103-104. Société Mathématique de France, Paris, 1983, pp. i+221.
[9] A.N. Kolmogorov, On conservation of conditionally periodic motions for a small change in Hamilton's function, Dokl. Akad. Nauk SSSR (N.S.) 98 (1954) 527–530 (Russian).
[10] S.B. Kuksin, Perturbation theory of conditionally periodic solutions of infinite-dimensional Hamiltonian systems and its applications to the Korteweg-de Vries equation, Mat. Sb. (N.S.) 136(178)(3) (1988) 396–412, 431 (Russian) (transl. in Math. USSR-Sb. 64(2) (1989) 397–413.
[11] S.B. Kuksin, Nearly Integrable Infinite-Dimensional Hamiltonian Systems, Lecture Notices in Mathematics, vol. 1556, Springer, New York, 1993.
[12] S.B. Kuksin, J. Pöschel, Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation, Ann. of Math. (2) 143 (1) (1996) 149–179.
[13] V.K. Melnikov, On certain cases of conservation of almost periodic motions with a small change of the Hamiltonian function, Dokl. Akad. Nauk SSSR 165 (1965) 1245–1248 (Russian).
[14] J. Moser, On invariant curves of area-preserving mappings of an annulus, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II (1962) 1–20.
[15] J. Moser, Convergent series expansions for quasi-periodic motions, Math. Ann. 169 (1967) 136–176.
[16] J. Moser, On the construction of almost periodic solutions for ordinary differential equations, 1970 Proceedings of the International Conference on Functional Analysis and Related Topics, Tokyo, 1969, University of Tokyo Press, Tokyo, pp. 60–67.
[17] J. Nash, The imbedding problem for Riemannian manifolds, Ann. of Math. (2) 63 (1956) 20–63.
[18] H. Poincarè, Les Methodes Nouvelles de la Mechanique Celeste, Gauthier Villars, Paris, 1892.
[19] J. Pöschel, Integrability of Hamiltonian systems on Cantor sets, Comm. Pure Appl. Math. 35 (1982) 653–695.
[20] J. Pöschel, On elliptic lower dimensional tori of Hamiltonian systems, Math. Z. 202 (1989) 559–608.
[21] J. Pöschel, A KAM-theorem for some nonlinear PDEs, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 23 (1996) 119–148.
[22] H. Rüssmann, Kleine Nenner. I. Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II (1970) 67–105 (German).
[23] H. Rüssmann, On optimal estimates for the solutions of linear partial differential equations with first order coefficients on the torus, J. Moser (Ed.), Dynamical Systems Theory and Applications, Lecture Notes in Physics, vol. 38, Springer, New York, 1975, pp. 598–624.
[24] H. Rüssmann, Konvergente Reihenentwicklungen in der Störungstheorie der Himmelsmechanik, (German) Selecta Mathematica, V (German), pp. 93–260, Heidelberger Taschenbücher, Vol. 201, Springer, Berlin, New York, 1979.
[25] H. Rüssmann, Invariant tori in non-degenerate nearly integrable Hamiltonian systems, Regul. Chaotic Dynamics 6 (2) (2001) 119–204.
[26] D. Salamon, E. Zehnder, KAM theory in configuration space, Comm. Math. Helv. 64 (1989) 84–132.
[27] J. Xu, J. You, Persistence of lower-dimensional tori under the first Melnikov's non-resonance condition, J. Math. Pures Appl. (9) 80 (10) (2001) 1045–1067.
[28] J. You, Perturbations of lower dimensional tori for Hamiltonian systems, J. Differential Equations 152 (1999) 1–29.
[29] E. Zehnder, Generalized implicit function theorems with applications to small divisor problems I II, Comm. Pure Appl. Math. 28 (1975) 91–140 29 (1976) 49–113.