IOP Publishing | London Mathematical Society Nonlinearity

Nonlinearity **33** (2020) 3526–3567 https://doi.org/10.1088/1361-6544/ab7a01

On the topology of nearly-integrable Hamiltonians at simple resonances

L Biasco and L Chierchia[1]

Dipartimento di Matematica e Fisica, Università degli Studi Roma, Tre Largo San L. Murialdo 1, 00146 Roma, Italy

E-mail: biasco@mat.uniroma3.it and luigi@mat.uniroma3.it

Received 4 February 2019, revised 20 January 2020
Accepted for publication 25 February 2020
Published 29 May 2020

Abstract

We show that, in general, averaging at simple resonances a real-analytic, nearly-integrable Hamiltonian, one obtains a one-dimensional system with a cosine-like potential; 'in general' means for a generic class of holomorphic perturbations and apart from a finite number of simple resonances with small Fourier modes; 'cosine-like' means that the potential depends only on the resonant angle, with respect to which it is a Morse function with one maximum and one minimum. Furthermore, the (full) transformed Hamiltonian is the sum of an effective one-dimensional Hamiltonian (which is, in turn, the sum of the unperturbed Hamiltonian plus the cosine-like potential) and a perturbation, which is uniformly exponentially small. As a corollary, under the above hypotheses, if the unperturbed Hamiltonian is also strictly convex, the effective Hamiltonian at any simple resonance (apart a finite number of low-mode resonances) has the phase portrait of a pendulum. The results presented in this paper are an essential step in the proof (in the 'mechanical' case) of a conjecture by Arnold–Kozlov–Neishdadt [Arnold V I, Kozlov V V and Neishtadt A I 2006 *Mathematical aspects of classical and celestial mechanics Encyclopaedia of Mathematical Sciences* 3rd edn vol 3 (Berlin: Springer), remark 6.8, p 285], claiming that the measure of the 'non-torus set' in general nearly-integrable Hamiltonian systems has the same size of the perturbation; compare [Biasco L and Chierchia L 2015 On the measure of Lagrangian invariant tori in nearly-integrable mechanical systems *Rendiconti Lincei. Mat. Appl.* **26** 1–10 and Biasco L and Chierchia L KAM Theory for Secondary Tori (arXiv:1702.06480v1 [math.DS])].

Keywords: nearly-integrable Hamiltonian systems, averaging theory, simple resonances
Mathematics Subject Classification numbers: 37J40, 70H08.

[1] Author to whom any correspondence should be addressed.

1. Introduction

1.1 Consider a real-analytic, nearly-integrable Hamiltonian given, in action-angle variables, by

$$H_\varepsilon(y,x) = h(y) + \varepsilon f(y,x), \quad (y,x) \in \mathcal{M} := D \times \mathbb{T}^n, \tag{1}$$

where D is a bounded domain in $\mathbb{R}^n$, $\mathbb{T}^n = \mathbb{R}^n/(2\pi\mathbb{Z}^n)$ is the usual flat n dimensional torus and ε is a small parameter measuring the size of the perturbation εf. The phase space $\mathcal{M}$ is endowed with the standard symplectic form $\mathrm{d}y \wedge \mathrm{d}x$ so that the Hamiltonian flow $\phi^t_{H_\varepsilon}(y_0,x_0) =: (y(t),x(t))$ governed by H_ε is the solution of the standard Hamiltonian equations

$$\begin{cases} \dot{y} = -\partial_x H_\varepsilon(y,x), \\ \dot{x} = \partial_y H_\varepsilon(y,x), \end{cases} \qquad \begin{cases} y(0) = y_0, \\ x(0) = x_0, \end{cases}$$

(where t is time and dot is time derivative).
It is well known that, in general, the $\phi^t_{H_\varepsilon}$-dynamics is strongly influenced by resonances of the (unperturbed) frequencies $\omega(y) := h'(y) = \partial_y h(y)$, i.e., by rational relations

$$\omega(y)\cdot k = \sum_{j=1}^n \omega_j(y)k_j = 0,$$

with $k \in \mathbb{Z}^n \setminus \{0\}$; for general information, compare, e.g., [1].

Indeed, assuming a standard non-degeneracy assumption on h, e.g., that the frequency map $y \in D \to \omega(y)$ is a real-analytic diffeomorphism of D onto the 'frequency space' $\Omega := h'(D)$, then the action space D can be covered by three open sets

$$D \subseteq D^0 \cup D^1 \cup D^2$$

so that the following holds. Roughly speaking, $D^0 \times \mathbb{T}^n$ is a completely non-resonant set which is filled, up to an exponentially (in $1/\varepsilon^c$) small set, by primary KAM tori, namely, Lagrangian tori $\phi^t_{H_\varepsilon}$-invariant on which the flow is analytically conjugated to the linear flow

$$\theta \in \mathbb{T}^n \mapsto \theta + \omega t$$

with ω satisfying a Diophantine condition

$$|\omega\cdot k| \geqslant \frac{\gamma}{|k|_1^\tau} \quad \forall\, k \in \mathbb{Z}^n \setminus \{0\},$$

(for some $\gamma, \tau > 0$); $\omega \cdot k$ denoting the standard inner product $\sum \omega_i k_i$ and $|k|_1 := \sum |k_i|$. Morevover, such tori are graphs over $\mathbb{T}^n$ and are analytic deformation of integrable tori. D^1 is an open $O(\sqrt{\varepsilon})$-neighbourhood of simple resonances (i.e., of regions where exactly one independent resonance $\omega(y)\cdot k = 0$ holds) and D^2 is a set of measure $O(\varepsilon)$.

Indeed, such description (up to logarithmic corrections) follows easily by the covering lemma below (proposition 2.1), by choosing suitably parameters (e.g., the 'small divisor constant' $\alpha \sim \sqrt{\varepsilon}$ and 'Fourier cutoffs' $K \sim |\log \varepsilon|^a$).

The region D^2 contains double (and higher) resonances and, in general, in $D^2 \times \mathbb{T}^n$ there are $O(\varepsilon)$ regions where the dynamics is non-perturbative, being 'essentially' governed (after suitable rescalings) by an ε-independent Hamiltonian, as the following quotation from [1] clarifies:

'It is natural to expect that in a generic system with three or more degrees of freedom the measure of the 'non-torus' set has order ε. Indeed, the $O(\sqrt{\varepsilon})$-neighbourhoods of two resonant surfaces intersect in a domain of measure $\sim \varepsilon$. In this domain, after the partial averaging taking into account the resonances under consideration, normalizing the deviations of the "actions" from the resonant values by the quantity $\sqrt{\varepsilon}$, normalizing time, and discarding the terms of higher order, we obtain a Hamiltonian of the form $\frac{1}{2}(Ap, p) + V(q_1, q_2)$, which does not involve a small parameter (see the definition of the quantity p above). Generally speaking, for this Hamiltonian there is a set of measure ~ 1 that does not contain points of invariant tori. Returning to the original variables we obtain a 'non-torus' set of measure $\sim \varepsilon$.' ([1, remark 6.8, p 285]).

Incidentally, we mention that, from the above description, it follows, as it is well known, that the measure of the complement of the KAM primary tori in phase space is in general of $O(\sqrt{\varepsilon})$ ([2–4]). In fact, this result is optimal, as the following trivial example shows: the phase region inside the separatrix of the pendulum $\frac{1}{2}y^2 + \varepsilon \cos x$, with $y \in \mathbb{R}$ and $x \in \mathbb{T}^1$, does not contain any primary invariant torus (circle) which is a global graph over the angle x, and such a region has measure $4\sqrt{2\varepsilon}$; (of course, in this integrable case, such a region is filled up by secondary tori, corresponding to oscillations of the pendulum).

1.2 The dynamics in the simple-resonance region $D^1 \times \mathbb{T}^n$ is particularly relevant and interesting. For example, it plays a major rôle in Arnold diffusion, as shown by Arnold himself [5], who based his famous instability argument on shadowing partially hyperbolic trajectories arising near simple resonances.

On the other hand, in $D^1 \times \mathbb{T}^n$ there appear secondary KAM tori, namely n-dimensional KAM tori with different topologies, which depend upon specific characteristics of the perturbation εf. The appearance of secondary tori is a genuine non-integrable effect, since such tori do not exist in the integrable regime.

In the announcement [6] it is claimed that, in the case of mechanical systems, namely systems governed by Hamiltonians of the form $H(y, x) = |y|^2/2 + \varepsilon f(x)$, and for generic potentials f, primary and secondary tori fill the region $D^1 \times \mathbb{T}^n$ up to a set of measure nearly exponentially small, showing that the 'non-torus set' is essentially $O(\varepsilon)$ as conjectured in [1] and studied in [7].

Indeed, the main results presented here, namely theorem 2.1, is one of the building blocks of the proof (in the mechanical case) of Arnold–Kozlov–Neishdadt conjecture. The idea of such proof is the following (compare, also, [8]).

As mentioned above D^0 is filled, up to an exponentially small set, by KAM invariant primary tori. On the other hand, the region D^2 (which, according to Arnold, Kozlov and Neishtadt, contains the main part of the non torus-set) has measure $O(\varepsilon)$. It remains to investigate D^1, which is union of $O(\sqrt{\varepsilon})$ neighbourhoods of simple resonances with Fourier modes k, with $|k| \leqslant K$, where the Fourier cut-off K goes to infinity as ε goes to zero (which is a standard fact in any normal form or averaging theory). In a neighbourhood of any fixed resonance, by averaging theory, one can symplectically conjugate H_ε to a Hamiltonian $\bar{H}^k_\varepsilon$ of the form

$$\bar{H}^k_\varepsilon := h^k(y) + \varepsilon G^k(y, k \cdot x) + \varepsilon R^k(y, x) \tag{2}$$

with R^k a small remainder and $\theta \to G^k(y, \theta)$ a periodic function of one single angle θ (and zero average), making the 'secular Hamiltonian' $h^k + \varepsilon G^k$ an integrable systems (parameterized by $n-1$ action variables). The natural strategy is then to put $\bar{H}_\varepsilon$ in the action-angle variables of the integrable Hamiltonian $h^k + \varepsilon G^k$ and to apply KAM theory. However, in order to do this one has to control an essentially infinite number of singular change of coordinates in a quantitative way (recall that $K \to \infty$ as $\varepsilon \to 0$), which is clearly a technically difficult task unless one has a uniform way to deal with infinitely many resonances. This is where theorem 2.1 below comes in, as it shows that, under genericity assumptions and up to a finite and fixed number of low

Fourier modes, all simple resonances lead to essentially the same pendulum-like system. The proof of the conjecture will then follow by showing KAM nondegeneracy of $\bar{H}_\varepsilon$ (this is a non trivial task, which will be addressed elsewhere; compare [8]; for a related stronger conjecture in two degrees of freedom, see [9]).

1.3 In the rest of this introduction we briefly discuss the main ideas needed to make a uniform quantitative analysis of nearly-integrable systems at simple resonances. For simplicity we focus on the case of purely positional potentials; precise statements are given in section 2 below, and, for the general (but more technical and implicit) case, in section 8. D^1 is the union of suitable regions $D^{1,k}$, which are $O(\sqrt{\varepsilon})$-close to exact simple resonances $\{y \in D | \omega(y) \cdot k = 0\}$, and which are labelled by generators k of one dimensional, maximal sublattices of $\mathbb{Z}^n$ (see (16) below); 'exact' meaning that $\omega(y)$ does not verify double or higher resonant relations.

In averaging/normal form theory (see, e.g., [1, section 6] and references therein), one typically considers a finite but large (typically, ε-dependent) number of simple resonances. More precisely, one considers generators k with $|k|_1 \leqslant K$, and K can be chosen according to the application one has in mind. Typically, one chooses $K \sim 1/\varepsilon^a$ for a suitable $a > 0$ (as in Nekhoroshev theorem [10, 11] or in the KAM theory for secondary tori in two degrees of freedom [9]) or $K \sim |\log \varepsilon|^a$ (as in the general KAM theory for secondary tori of [6, 8]).

Furthermore, in any fixed simple resonant region $D^{1,k}$, one can remove the non-resonant angle dependence, so as to symplectically conjugate H_ε, for ε small enough, to the Hamiltonian $\bar{H}^k_\varepsilon$ in (2) where: h^k is ε-close to h, $\theta \to G^k(y, \theta)$ is a function of one angle and R^k is a 'very small' remainder, with mth Fourier coefficient $R^k_m(y)$ vanishing when $m = jk$ for some $j \in \mathbb{Z}$. Thus, up to the remainder R^k, the Hamiltonian depends effectively only on the 'resonant angle' $\theta := k \cdot x$ and therefore the secular Hamiltonian $h^k + \varepsilon G^k$ is integrable: this is the starting point for ('*a priori* stable') Arnold diffusion (compare, e.g., [5, 12–19]) or for the KAM theory for secondary tori of [6, 8].

There are here two main issues to be clarified:

(a) What is the actual 'generic form' of G^k?

(b) How small (and compared to what) is the remainder R^k?

(a) According to averaging (or normal form) theory G^k is 'close' to the projection of the potential f on the Fourier modes of the resonant maximal sublattice $\mathbb{Z}k$:

$$\mathrm{p}_{\mathbb{Z}k} f(x) := \sum_{j \in \mathbb{Z}} f_{jk}\, \mathrm{e}^{\mathrm{i} jk \cdot x}.$$

Now, since f is real-analytic on $\mathbb{T}^n$, it is holomorphic in a complex strip $\mathbb{T}^n_s$ around $\mathbb{T}^n$ of width $s > 0$ and its Fourier coefficients decay exponentially fast, behaving typically as

$$|f_k| \sim \|f\|\, \mathrm{e}^{-|k|_1 s}. \tag{3}$$

Hence, in general, one will have

$$\sum_{j \in \mathbb{Z}} f_{jk}\, \mathrm{e}^{\mathrm{i} jk \cdot x} = f_k\, \mathrm{e}^{\mathrm{i}k \cdot x} + f_{-k}\, \mathrm{e}^{-\mathrm{i}k \cdot x} + O\left(\|f\|\, \mathrm{e}^{-2|k|_1 s}\right)$$

which, by the reality condition $f_{-k} = \bar{f}_k$, can be written as

$$\sum_{j \in \mathbb{Z}} f_{jk}\, \mathrm{e}^{\mathrm{i} jk \cdot x} = 2|f_k| \cos\left(k \cdot x + \theta^{(k)}\right) + O\left(\|f\|\, \mathrm{e}^{-2|k|_1 s}\right) \tag{4}$$

for a suitable $\theta^{(k)} \in [0, 2\pi)$. Thus,

$$\mathrm{p}_{\mathbb{Z}k} f(x) = \sum_{j \in \mathbb{Z}} f_{jk}\, \mathrm{e}^{\mathrm{i}jk\cdot x} = 2|f_k| \left(\cos\left(k \cdot x + \theta^{(k)}\right) + o(1)\right) \tag{5}$$

provided the higher Fourier modes are smaller compared to $|f_k|$, i.e.:

$$|k|_1 \gtrapprox \frac{1}{s}. \tag{6}$$

(precise statements, norms, etc, will be given in section 2). In other words, one expects (5) to hold for generic real-analytic potentials and for all generators k's satisfying (6). Indeed, this is the case: we shall introduce certain classes of analytic potentials $\mathcal{H}_{s,\tau}$ (definition 2.2), for which (choosing suitably the 'tail' function τ) (5) holds for generators k satisfying (6). The class $\mathcal{H}_{s,\tau}$ is 'generic' in several ways (compare definition 2.2 and theorem 3.1 below):

(a) It contains an open dense set in the class of real-analytic functions having holomorphic extension on a complex neighbourhood of size s of $\mathbb{T}^n$ (in the topology induced by a suitably weighted Fourier norm);
(b) Its unit ball is of measure 1 (with respect to a natural probability measure);
(c) It is a 'prevalent set' (compare [20] or [21]).

Next, in order for G^k to be close to $\mathrm{p}_{\mathbb{Z}k} f$ in (5), one needs to have a bound of the type

$$\sup_{D^{1,k} \times \mathbb{T}^n} |G^k - \mathrm{p}_{\mathbb{Z}k} f| \ll |f_k| \overset{(3)}{\sim} \|f\|\, \mathrm{e}^{-|k|_1 s}. \tag{7}$$

As well known, analytic averaging methods involve an analyticity loss in complex domains. In particular, the Hamiltonian in (2) and, therefore, G^k, can be analytically defined only in a smaller complex strip $\mathbb{T}^n_{s_\star}$ with $s_\star < s$. Therefore, by analyticity arguments, the best one can hope for is an estimate of the type

$$\sup_{D^{1,k} \times \mathbb{T}^n} |G^k - \mathrm{p}_{\mathbb{Z}k} f| \leqslant c \cdot \|f\|\, \mathrm{e}^{-|k|_1 s_\star}, \tag{8}$$

for a suitable constant c that can be taken to be smaller than any prefixed positive number. But then, for (8) and (7) to be compatible one sees that one must 'essentially' have $s_\star \sim s$ and that standard averaging theory is not enough (compare, e.g., [11], where $s_\star = s/6$; for a more detailed comparison with the averaging lemma of [11], see also remark 4.1(d) below; for other refined normal forms, see also [22–25]). To overcome this problem, we provide (section 4) a normal form lemma with small analyticity loss, 'small' meaning that one can take

$$s_\star = s(1 - 1/K) \tag{9}$$

(compare, in particular, (70)). The value (9) is compatible with (7) for $|k|_1 \leqslant K$, showing that, generically, one has that $G^k(y, k \cdot x)$ has the same form of the right hand side of (5). In particular we prove that: the 'effective Hamiltonians' $h^k + \varepsilon G^k$, as k varies ($c/s \leqslant |k|_1 \leqslant K$), have essentially the same cosine-like potential (up to a phase-shift) and, hence, the same topological features (compare remark 2.2(a) below).

Notice also that for low modes this last property, in general, does not hold, as one immediately sees by considering $k = e_1 = (1, 0, \ldots, 0)$ and a potential f such that

$$\mathrm{p}_{\mathbb{Z}e_1} f(x) = \cos x_1 + \cos 2x_1,$$

which is a Morse function with two maxima and two minima in $\mathbb{T}^1$. In fact, in the above mentioned proof of Arnold–Kozlov–Neishtadt conjecture ([8]), low modes need a ad hoc analysis and, in particular, fine analytic properties of the singular action-angle variables for general Morse potential have to be used (compare, [26]).

(b) What we have just discussed gives also an indication of how small R^k needs to be in order to be negligible. In fact, if (5) is the leading behaviour, one should have $\|R^k\| \ll |f_k|$. But in order to perform averaging procedures, one has, typically, control on small divisors up to the truncation order K, so that the remainder will be (by analyticity estimates) of order e^{-Ks} even using a refined normal fomal lemma as the one mentioned above. But this is not enough since $|f_k|$ can get as small as e^{-Ks} for large mode k (see (3)).

To overcome this problem, we introduce in section 5, at difference with standard geometry of resonances in Nekhoroshev's theory (compare, [10, 11, 27–31]), two Fourier cutoffs K_1 and K_2 (with $K_2 \geqslant 3K_1$) in such a way that on the simple resonant regions $D^{1,k}$ one has non-resonance conditions for double and higher resonances up to order K_2, while K_1, which is introduced to make averaging theory in D^0, is also the maximum value of the size of the generators k associated to simple resonances. Therefore, we will get an estimate of the remainder R^k of the type

$$\|R^k\| \leqslant \mathrm{const}\, K_2^a\, \mathrm{e}^{-K_2 s/2} \leqslant \mathrm{const}\, \mathrm{e}^{-K_1 s} \cdot \mathrm{e}^{-K_2 s/8} \overset{(3)}{\leqslant} \mathrm{const} |f_k|\; \mathrm{e}^{-K_2 s/8}\, ; \tag{10}$$

(where in the second inequality we used $K_2 \geqslant 3K_1$ and $\|f\|$ has been absorbed in the constants). Notice that in the last inequality we are also using the genericity of the potential f, which is assumed to have a typical behaviour with respect to the Fourier modes corresponding to the generators k (see, in particular, (32) below). The final upshot is that one can construct, via a near-identity symplectic transformation Ψ_k, a complete normal form

$$H_\varepsilon \circ \Psi_k =: h^k(y) + 2|f_k|\varepsilon \left(\cos(k\cdot x + \theta^{(k)}) + \tilde{G}^k(y, k\cdot x) + \tilde{R}^k(y,x)\right) \tag{11}$$

with $\|\tilde{G}^k\| \ll 1$ and $\|\tilde{R}^k\| \leqslant C'\, \mathrm{e}^{-K_2 s/8}$; compare theorem 2.1 below and, in particular, formula (42).

Summarising: generically, for all k large enough, G^k in (2) is 'cosine-like', i.e., a Morse function with one maximum and one minimum (compare remark 2.2(a) below) and R^k is exponentially small with respect to the oscillations of G^k (see (42) and (44) below).

As a consequence we get that, if $h(y)$ is strictly convex, the effective Hamiltonian has a phase portrait of a pendulum (compare remark 2.2(b)) below).

The paper is organized as follows.

In section 2 (definitions and main theorem) the principal definitions, assumptions and results are given (KAM non-degeneracy, generators of 1D maximal lattices $\mathcal{G}_1^n$, covering lemma and the sets D^i, the classes of generic potentials $\mathcal{H}_{s,\tau}$, and theorem 2.1, which is the main result of the paper).

In section 3 (the generic class $\mathcal{H}_{s,\tau}$) the properties of the class of potentials $\mathcal{H}_{s,\tau}$ are discussed and proved.

Section 4 (normal form lemma) is a technical section devoted to a refined normal form lemma, allowing for minimal analyticity loss in angle variables.

In section 5 (geometry of resonances and proof of the covering lemma) the proof of proposition 2.1 (the 'covering lemma') on the geometry of resonances and their properties is given.

In section 6 (averaging theory) the covering lemma and the normal form lemma are put together so as to get the high-order averaging theory in the non-resonant region $D^0 \times \mathbb{T}^n$ and, especially, in the simple-resonant regions $D^{1,k} \times \mathbb{T}^n$; here a definite choice for the Fourier cut-offs and the parameters controlling small divisors is done (assumption A).

In section 7 (proof of theorem 2.1) the proof of the main theorem is given.

In section 8 (general (y-dependent) potentials) the more general case of potentials depending also on action variables is briefly discussed.

Appendix A contains the proof of an elementary result in linear algebra needed in a linear symplectic change of variables, clarifying the universal pendulum structure after normalization in the case of mechanical systems.

2. Definitions and main theorem

Let $n \geqslant 2$ and D be a bounded domain in $\mathbb{R}^n$. Assume that H_ε in (1) admits, for some $r, s > 0$, a holomorphic extension on the complex domain $D_r \times \mathbb{T}^n_s \subset \mathbb{C}^{2n}$, where

$$D_r := \bigcup_{y\in D} \{z \in \mathbb{C}^n : |z - y| < r\},$$

$$\mathbb{T}^n_s := \{x = (x_1, \ldots, x_n) \in \mathbb{C}^n : |\mathrm{Im}\, x_j| < s\}/(2\pi\mathbb{Z}^n),$$

$|\cdot|$ denoting the standard Euclidean norm.

The integrable Hamiltonian h is supposed to be 'KAM non-degenerate' in the following sense.

Definition 2.1 (KAM non-degeneracy). A real-analytic function

$$h : y \in D \subset \mathbb{R}^n \mapsto h(y) \in \mathbb{R}, \quad (n \geqslant 2), \tag{12}$$

D being a bounded domain of $\mathbb{R}^n$, is said to be KAM non-degenerate if the frequency map

$$y \in D \mapsto \omega(y) := \partial_y h(y) \in \Omega := \omega(D) \subseteq B_M(0) \subset \mathbb{R}^n, \quad M := \sup_D |\omega(y)|, \tag{13}$$

is a global diffeomorphism of D onto Ω with Lipschitz constants given by

$$|y - y_0|\bar{L}^{-1} \leqslant |\omega(y) - \omega(y_0)| \leqslant L|y - y_0|, \quad (\forall\, y, y_0 \in D); \tag{14}$$

$B_M(0)$ denotes the real Euclidean ball of radius M centred in the origin.

Let $\mathbb{Z}^n_\star$ denote the set of integer vectors $k \neq 0$ in $\mathbb{Z}^n$ such that the first non-null component is positive:

$$\mathbb{Z}^n_\star := \{k \in \mathbb{Z}^n : k \neq 0 \quad \text{and} \quad k_j > 0 \quad \text{where } j = \min\{i : k_i \neq 0\}\}, \tag{15}$$

and denote by $\mathcal{G}^n_1$ the generators of 1D maximal lattices, namely, the set of vectors $k \in \mathbb{Z}^n_\star$ such that the greater common divisor (gcd) of their components is 1:

$$\mathcal{G}^n_1 := \{k \in \mathbb{Z}^n_\star : \gcd(k_1, \ldots, k_n) = 1\}. \tag{16}$$

Then, the list of one-dimensional maximal lattices is given by the sets $\mathbb{Z}k$ with $k \in \mathcal{G}^n_1$. Given $K > 0$ we set

$$\mathcal{G}^n_{1,K} := \mathcal{G}^n_1 \cap \{|k|_1 \leqslant K\}. \tag{17}$$

Proposition 2.1 (Covering lemma). *Let h be KAM non-degenerate and let ω denote its gradient. Fix $K_2 \geqslant K_1 \geqslant 2$ and $\alpha > 0$. Then, the domain D can be covered by three sets $D^i \subseteq D$,*

$$D = D^0 \cup D^1 \cup D^2\,, \tag{18}$$

so that the following holds.

(*a*) *D^0 is $(\alpha/2, K_1)$ completely non-resonant (i.e., non-resonant modulus $\{0\}$), namely,*

$$y \in D^0 \Longrightarrow |\omega(y)\cdot k| \geqslant \alpha/2, \quad \forall\, 0 < |k|_1 := \sum |k_j| \leqslant K_1. \tag{19}$$

(*b*) *$D^1 = \bigcup_{k\in\mathcal{G}^n_{1,K_1}} D^{1,k}$, where, for each $k \in \mathcal{G}^n_{1,K_1}$, $D^{1,k}$ is a neighbourhood of a simple resonance $\{y \in D : \omega(y)\cdot k = 0\}$, which is $(2\alpha K_2/|k|, K_2)$ non-resonant modulo $\mathbb{Z}k$, namely,*

$$y \in D^{1,k} \Longrightarrow |\omega(y)\cdot \ell| \geqslant 2\alpha K_2/|k|, \quad \forall \ell \in \mathbb{Z}^n,\ \ell \notin \mathbb{Z}k, |\ell|_1 \leqslant K_2. \tag{20}$$

(*c*) *D^2 contains all the resonances of order two or more and has Lebesgue measure small with α^2: more precisely, there exists a constant $c > 0$ depending only on n such that*

$$meas\,(D^2) \leqslant c\,\bar{L}^n M^{n-2}\,\alpha^2\,K_2^{n+1} K_1^{n-1}; \tag{21}$$

($\bar{L}$ and M being as in definition 2.1).

The proof is given in section 5.
The covering $\{D^i\}$ in (18) is the pull back of a covering in frequency space:

$$D^i := \{y \in D : \omega(y) \in \Omega^i\}, \tag{22}$$

where the Ω^i's are defined as follows.

$$\Omega^0 := \{\omega \in B_M(0) : \min_{k\in\mathcal{G}^n_{1,K_1}} |\omega\cdot k| > \alpha/2\}\,. \tag{23}$$

To define Ω^1 and $D^{1,k}$ for $k \in \mathcal{G}^n_{1,K_1}$, let $\mathrm{p}_k^\perp$ denote the orthogonal projection on the subspace perpendicular to k, i.e.,

$$\mathrm{p}_k^\perp \omega := \omega - \frac{1}{|k|^2}(\omega\cdot k)k.$$

Then,

$$\Omega^{1,k} := \left\{\omega \in \mathbb{R}^n : |\omega\cdot k| < \alpha,\ |\,\mathrm{p}_k^\perp\omega| < M, \quad \text{and} \quad |\,\mathrm{p}_k^\perp\omega\cdot\ell| > \frac{3\alpha K_2}{|k|}, \quad \forall \ell \in \mathcal{G}^n_{1,K_2}\setminus\mathbb{Z}k\right\}$$

$$D^{1,k} := \{y \in D : \omega(y) \in \Omega^{1,k}\}, \tag{24}$$

$$\Omega^1 := \bigcup_{k\in\mathcal{G}^n_{1,K_1}} \Omega^{1,k}. \tag{25}$$

Finally, the set Ω^2 is the union of neighbourhoods of exact double resonances: if

$$R_{k,\ell} := \{\omega\cdot k = \omega\cdot\ell = 0\}\,, \quad k \in \mathcal{G}^n_{1,K_1},\ \ell \in \mathcal{G}^n_{1,K_2},\ \ell \notin \mathbb{Z}k\,, \tag{26}$$

then

$$\Omega^2_{k,\ell} := \{|\omega\cdot k| < \alpha\} \cap \{|\,\mathtt{p}_k^{\perp}\omega| < M\} \cap \{|\,\mathtt{p}_k^{\perp}\omega\cdot\ell| \leqslant 3\alpha K_2/|k|\}, \tag{27}$$

$$\Omega^2 = \bigcup_{k\in\mathcal{G}^n_{1,K_1}} \bigcup_{\substack{\ell\in\mathcal{G}^n_{1,K_2}\\ \ell\notin\mathbb{Z}k}} \Omega^2_{k,\ell}. \tag{28}$$

Remark 2.1.

(a) The simply resonant regions $D^{1,k}$ in (24) are labelled by generators of 1D maximal lattices $k\in\mathcal{G}^n_1$ up to size $|k|_1 \leqslant K_1$, however, the non-resonance condition (20) holds for integer vectors ℓ with $|\ell|_1$ up to a (possibly) larger order K_2. This generalization (with respect to having $K_2 = K_1$ as, e.g., in [11]) is technical but important if one wants to have sharp control over the averaged Hamiltonian in a normal form near simple resonances; in particular in order to obtain (10) (compare, also, (150), (172), and (44)).

(b) The non-resonance relations (19) and (20) allow to apply averaging theory and to remove the dependence upon the 'non-resonant angle variables' up to exponential order; for precise statements, see theorem 6.1 in section 6.

We proceed, now, to describe the generic non-degeneracy assumption on periodic holomorphic functions, which will allow us to state the main theorem (in the case of positional potentials).

If $s > 0$, we denote by $\mathbb{B}^n_s$ the Banach space of real-analytic functions on $\mathbb{T}^n_s$ having zero average and finite ℓ^∞ weighted Fourier norm:

$$\mathbb{B}^n_s := \left\{ f = \sum_{\substack{k\in\mathbb{Z}^n\\ k\neq 0}} f_k\, e^{ik\cdot x} : \|f\|_s < \infty \right\}, \quad \text{where } \|f\|_s := \sup_{k\in\mathbb{Z}^n} |f_k|\, e^{|k|_1 s}. \tag{29}$$

Note that $f \in \mathbb{B}^n_s$ can be uniquely written as:

$$f(x) = \sum_{k\in\mathcal{G}^n_1} \sum_{j\in\mathbb{Z}\setminus\{0\}} f_{jk}\, e^{i\,jk\cdot x}. \tag{30}$$

For functions (not necessarily holomorphic in y) $f : D_r \times \mathbb{T}^n_s \to \mathbb{C}$ we will also use the (stronger) norm

$$|f|_{D,r,s} = |f|_{r,s} := \sup_{y\in D_r} \sum_{k\in\mathbb{Z}^n} |f_k(y)|\, e^{|k|_1 s}. \tag{31}$$

Definition 2.2 (Non degenerate potentials)**.** A tail function τ is, by definition, a non-increasing, non-negative continuous function

$$\tau : \delta \in (0,1] \mapsto \tau(\delta) \geqslant 0.$$

Given $s > 0$ and a (possibly s-dependent) tail function τ, we define, for $\delta\in(0,1]$, $\mathcal{H}_{s,\tau}(\delta)$ as the set of functions in $\mathbb{B}^n_s$ such that, for any generator $k \in \mathcal{G}^n_1$, the following holds

$$\text{if } |k|_1 > \tau(\delta), \quad \text{then } |f_k| \geqslant \delta |k|_1^{-n}\, e^{-|k|_1 s}, \tag{32}$$

The class $\mathcal{H}_{s,\tau}$ is the union over δ of the classes $\mathcal{H}_{s,\tau}(\delta)$:

$$\mathcal{H}_{s,\tau} := \bigcup_{0<\delta\leqslant 1} \mathcal{H}_{s,\tau}(\delta). \tag{33}$$

In fact, one could substitute n in (32) with every $\bar{n} > n/2$; compare remark 3.2 below. The 'weight' $|k|_1^{-n}$ is necessary in order to show that $\mathcal{H}_{s,\tau}(\delta)$ in (33) has positive measure in a suitable probability space; compare theorem 3.1(b) below.

The classes $\mathcal{H}_{s,\tau}$ contain (if the tail is choosen properly) the non degenerate potentials for which theorem 2.1 below holds and, as mentioned in the introduction, satisfy three genericity properties, as shown in theorem 3.1 below (compare, also, remark 3.1). We remark that such properties hold for any tail τ, which can be chosen differently according to the particular problem at hand.

We are ready to the state the main result of this paper.

Theorem 2.1. *Let $n \geqslant 2$, $s > 0$, $0 < \delta, \gamma \leqslant 1$ such that*

$$\gamma\delta < \frac{2^9}{s^n}\,\mathrm{e}^{-n^2/2}. \tag{34}$$

consider a Hamiltonian $H_\varepsilon(y,x) = h(y) + \varepsilon f(x)$ as in (1) with h KAM non-degenerate (definition 2.1) and f purely positional (i.e., independent of the y-variable) with

$$\|f\|_s = 1. \tag{35}$$

Assume that the potential is non-degenerate in the sense that

$$f \in \mathcal{H}_{s,\tau_o}(\delta) \tag{36}$$

with tail function

$$\tau_o(\delta;\gamma) := \frac{4}{s}\,\log\left(e + \frac{2^9}{s^n\gamma\delta}\right). \tag{37}$$

Let $K_2 \geqslant 3K_1 \geqslant 6$ satisfy

$$K_2^{2\nu-3n-3} \geqslant e^{s+5}2^{2n+11}n^{2n}\frac{L}{s^{2n+1}}\frac{1}{\gamma\delta}, \quad \textit{for some } \nu \geqslant \frac{3}{2}n+2, \tag{38}$$

L being defined in (14). Set

$$r_k := \sqrt{\varepsilon}\frac{K_2^\nu}{L|k|}. \tag{39}$$

Finally, assume that

$$\varepsilon \leqslant \frac{(Lr)^2}{K_2^{2\nu}}. \tag{40}$$

Then, for any $k \in \mathcal{G}^n_{1,K_1}$ with $\tau_o(\delta;\gamma) \leqslant |k|_1 \leqslant K_1$, there exists $\theta^{(k)} \in [0, 2\pi)$ and a symplectic change of variables defined in the neighbourhood of the simple resonance $D^{1,k} \times \mathbb{T}^n$, with $D^{1,k}$ defined in (24), such that the following holds:

$$\Psi_k : D^{1,k}_{r_k/2} \times \mathbb{T}^n_{s(1-1/K_2)^2} \to D^{1,k}_{r_k} \times \mathbb{T}^n_{s(1-1/K_2)}, \tag{41}$$

and

$$\boxed{H_\varepsilon \circ \Psi_k =: h^k(y) + 2|f_k|\varepsilon\,\left(\cos(k\cdot x + \theta^{(k)}) + \mathtt{G}^k(y, k\cdot x) + \mathtt{f}^k(y,x)\right)} \tag{42}$$

where $\mathtt{G}^k(y,\cdot) \in \mathbb{B}^1_2$ *for every* $y \in D^{1,k}_{r_k/2}$ *and*

$$\sup_{D^{1,k}_{r_k/2}} |h^k - h| \leqslant \gamma\delta\varepsilon, \quad |\mathtt{G}^k|_{D^{1,k},r_k/2,2} \leqslant \gamma\,. \tag{43}$$

Finally,

$$\mathtt{p}_{\mathbb{Z}k}\mathtt{f}^k = 0 \quad \textit{and} \quad |\mathtt{f}^k|_{D^{1,k},r_k/2,s(1-1/K_2)/2} \leqslant \frac{2^{10n}n^{3n}}{s^{3n}\delta}\,\mathrm{e}^{-K_2 s/8}. \tag{44}$$

The proof of this theorem is given in section 7.

Remark 2.2.

(a) Morse structure of the 'secular' potentials.

Recalling (31), estimate (43) means

$$\sup_{y\in D^{1,k}_{r_k/2}} \sum_{j\in\mathbb{Z}} |\mathtt{G}^k_j(y)|\,\mathrm{e}^{2|j|} \leqslant \gamma\,. \tag{45}$$

This implies that for every $y \in D^{1,k}$ the 2π-periodic real function

$$\theta \mapsto \cos(\theta + \theta^{(k)}) + \mathtt{G}^k(y,\theta) \tag{46}$$

behaves like a cosine in the sense that it is a Morse function with only one maximum and one minimum and no other critical points. To prove this, notice that by (45) we have

$$\sup_{y\in D^{1,k}, x\in\mathbb{T}^1} |\partial_\theta \mathtt{G}^k(y,\theta)| \leqslant \gamma/\mathrm{e}^2, \qquad \sup_{y\in D^{1,k}, x\in\mathbb{T}^1} |\partial^2_{\theta\theta} \mathtt{G}^k(y,\theta)| \leqslant \gamma/\mathrm{e}^2.$$

Therefore, denoting by $\psi(\theta)$ the derivative of the function in (46), we have that $\psi(\theta) > 0$ for $\theta \in (-\theta^{(k)} + \theta_*, \pi - \theta^{(k)} - \theta_*)$ and $\psi(\theta) < 0$ for $\theta \in (\pi - \theta^{(k)} + \theta_*, 2\pi - \theta^{(k)} - \theta_*)$, where $\theta_* := \arcsin(\gamma/e^2)$. Moreover in the interval $(-\theta^{(k)} - \theta_*, -\theta^{(k)} + \theta_*)$ the function $\psi(\theta)$ has a zero and is strictly increasing since $\psi'(\theta) \geqslant \sqrt{1-\gamma/e^2} - \gamma/e^2 =: c > 0$. Finally in the interval $(\pi - \theta^{(k)} - \theta_*, \pi - \theta^{(k)} + \theta_*)$ it has a zero and is strictly decreasing since $\psi'(\theta) \leqslant -c < 0$.

As a consequence the phase portrait of the effective Hamiltonian

$$h^k(y) + 2|f_k|\varepsilon\left(\cos(k\cdot x + \theta^{(k)}) + \mathtt{G}^k(y, k\cdot x)\right) \tag{47}$$

and that of the Hamiltonian $h^k(y) + 2|f_k|\varepsilon\cos(k\cdot x + \theta^{(k)})$ are topologically equivalent.

(b) The 'pendulum structure' of the secular Hamiltonian.

The secular Hamiltonian (47) is an integrable system as it depends only on one angle. Fix $k \in \mathbb{Z}^n\backslash\{0\}$ with $\gcd(k_1,\ldots,k_n) = 1$, then, there exists a matrix $A_k \in \mathrm{Mat}_{n\times n}(\mathbb{Z})$ such that

$$A_k = \begin{pmatrix} \hat{A}_k \\ k \end{pmatrix} \in \mathrm{Mat}_{n\times n}(\mathbb{Z}), \quad \hat{A}_k \in \mathrm{Mat}_{(n-1)\times n}(\mathbb{Z}), \quad \det A_k = 1, \quad |\hat{A}_k|_\infty \leqslant K_\infty, \tag{48}$$

where $|\cdot|_\infty$, as usual, denotes sup-norm (of matrices or vectors). The existence of such a matrix is guaranteed by an elementary result of linear algebra based on Bezout's lemma (for completeness, we include the proof in appendix A; see lemma A.1).

Let us perform the linear symplectic change of variables

$$\Phi_k : (Y,X) \mapsto (y,x) := (A_k^{\mathrm{T}}Y, A_k^{-1}X), \tag{49}$$

which is generated by the generating function $S(Y,x) := Y \cdot A_k x$. Note that Φ_k does not mix actions with angles, its projection on the angles is a diffeomorphism of $\mathbb{T}^n$ onto $\mathbb{T}^n$, and, most relevantly, $X_n = k \cdot x$ is the 'secular angle'.
In the (Y, X)-variables, the secular Hamiltonian in (47) takes the form

$$\mathsf{h}^k(Y) + 2|f_k|\varepsilon \left(\cos(X_n + \theta^{(k)}) + \mathsf{G}^k(A_k^{\mathrm{T}}Y, X_n)\right), \quad \text{with } \mathsf{h}^k(Y) := h^k(A_k^{\mathrm{T}}Y). \tag{50}$$

Fix $y_0 \in D^{1,k}$ on the exact resonance, namely $\partial_y h^k(y_0) \cdot k = 0$. Let Y_0 be such that $y_0 = A_k^{\mathrm{T}} Y_0$. We have

$$\partial_{Y_n} \mathsf{h}^k(Y_0) \overset{(48)}{=} \partial_y h^k(A_k^{\mathrm{T}} Y_0) \cdot k = \partial_y h^k(y_0) \cdot k = 0, \quad \partial^2_{Y_n Y_n} \mathsf{h}^k(Y_0) = \partial^2_{yy} h^k(y_0) k \cdot k,$$

where $\partial^2_{yy} h^k$ is the Hessian matrix of h^k. By Taylor expansion the secular Hamiltonian in (50) takes the form (up to an addictive constant)

$$\frac{1}{2}\left(\partial^2_{yy} h^k(y_0) k \cdot k\right)(Y_n - Y_{0n})^2 + O\left((Y_n - Y_{0n})^3\right) + 2|f_k|\varepsilon \left(\cos(X_n + \theta^{(k)}) + \mathsf{G}^k(A_k^{\mathrm{T}} Y, X_n)\right). \tag{51}$$

In particular if the Hamiltonian h is convex the coefficient $\partial^2_{yy} h^k(y_0) k \cdot k = : m_k$ is bounded away from zero (for ε small enough independently of k) and the phase portrait of the secular Hamiltonian in (51) is topologically equivalent, for $|Y_n - Y_{0n}|$ small (in the region $D^{1,k} \times \mathbb{T}^n$ in the original variables), to that of the pendulum

$$\frac{1}{2} m_k (Y_n - Y_{0n})^2 + 2|f_k|\varepsilon \cos(X_n + \theta^{(k)}).$$

(c) The y-dependent case.

Let us briefly discuss the case in which f depends explicitly also on the actions y. First we note that it can happen that, even if the potential $f(y, x)$ satisfies the non-degeneracy condition given in definition 2.2 at some point y_0, there is no neighborhoud of y_0 on which the non-degeneracy condition holds. For example consider the potential

$$f(y,x) = f(y_1, x) := \frac{1}{2}\sum_{k \neq 0} \left(|k|_1^{-n} - \frac{y_1}{r}\right) \mathrm{e}^{-|k|_1 s}\, \mathrm{e}^{\mathrm{i}k\cdot x}.$$

We have that $\|f\|_{D,r,s} := \sup_{\mathrm{D_r}} |\mathrm{f}|_{\mathrm{s}} = 1$ with $D = \{0\}$ and

$$f(0, \cdot) \in \mathcal{H}_{s,0}(1/2).$$

However, $f_k(rj^{-n}) = 0$ for every $|k| = j$; in particular for every odd number $j = 2h + 1$, $h \geqslant 1$ and $k := (h + 1, h, 0, \ldots, 0) \in \mathcal{G}_1^n$. Then for every $\delta > 0$, tail function $\tau > 0$ and odd $j \geqslant 3$, we have

$$f(rj^{-n}, \cdot) \notin \mathcal{H}_{s,\tau}(\delta).$$

However, one can prove that the non-degeneracy condition holds in a set of large measure. In particular, we will prove that given $\mu > 0$, if for a certain point $y_0 \in D$ the potential $f(y_0, \cdot) \in \mathcal{H}_{s,\tau_*}(\delta)$ for a suitable $\tau_* = \tau_*(\mu)$ and $|k| \geqslant \tau_*(\mu)$, then (42)–(44) holds (with $f_k = f_k(y)$ and for a suitable phase $\theta^{(k)} = \theta^{(k)}(y)$) for every $y \in B_{r/2e}(y_0)$ up to a set of relative measure smaller that μ. The precise statement is given in theorem 8.2, section 8.

(d) On the choice of norms.

The space of functions $f : \mathbb{T}^n_s \to \mathbb{C}^m$ endowed with the ℓ^1-Fourier norm $|\cdot|_s$ is a Banach algebra, while $\{f : \mathbb{T}^n_s \to \mathbb{C}^m \text{ s.t. } \|f\|_s < \infty\}$ is just a Banach space (not a Banach algebra). However, the norm $\|\cdot\|_s$ is particularly suited to describe the space of potentials as a probability space (see next section).

Of course, the norms $\|\cdot\|_s$ and $|\cdot|_s$ are not equivalent and, in general, as it is easy to check, one has:

$$\|f\|_{r,s} \leqslant |f|_{r,s} \leqslant (\coth^n(\sigma/2) - 1)\|f\|_{r,s+\sigma} \leqslant (2n/\sigma)^n \|f\|_{r,s+\sigma}. \tag{52}$$

3. The generic class $\mathcal{H}_{s,\tau}$

Here we discuss some properties of the classes $\mathcal{H}_{s,\tau}$ of non-degenerate potentials introduced in definition 2.2.

To a given $f \in \mathbb{B}^n_s$ and $k \in \mathcal{G}^n_1$ we associate a periodic function F^k as follows

$$f \in \mathbb{B}^n_s \mapsto F^k \in \mathbb{B}^1_{|k|_1 s} \quad \text{where } F^k(\theta) := \sum_{j \in \mathbb{Z}\setminus\{0\}} f_{jk}\, e^{\mathrm{i} j\theta}\,, \tag{53}$$

(f_{jk} being the Fourier coefficient of f with Fourier index $jk \in \mathbb{Z}^n$), and notice that any $f \in \mathbb{B}^n_s$ can be uniquely written as:

$$\boxed{f(x) = \sum_{k \in \mathcal{G}^n_1} F^k(k \cdot x)} \tag{54}$$

Notice also that, if $k \in \mathcal{G}^n_1$ and $|f|_{r,s} < \infty$, then

$$|F^k|_{r,|k|_1 s} \leqslant |f|_{r,s}. \tag{55}$$

Next, we introduce a probability measure on the unit ball in $\mathbb{B}^n_s$. Denote by $\ell^\infty(\mathbb{Z}^n_\star)$ the Banach space of complex sequences (over $\mathbb{Z}^n_\star$) given by

$$\ell^\infty(\mathbb{Z}^n_\star) := \left\{ z \in \mathbb{C}^{\mathbb{Z}^n_\star} \text{ s.t. } z_k \neq 0 \quad \text{and} \quad |z|_\infty := \sup_{k \in \mathbb{Z}^n_\star} |z_k| < +\infty \right\}. \tag{56}$$

Then, the map

$$j : f \in \mathbb{B}^n_s \to \left\{ f_k\, e^{|k|_1 s} \right\}_{k \in \mathbb{Z}^n_\star} \in \ell^\infty(\mathbb{Z}^n_\star) \tag{57}$$

is an isomorphism of Banach spaces, which allows to identify functions in $\mathbb{B}^n_s$ with points in $\ell^\infty(\mathbb{Z}^n_\star)$ and the Borellians of $\mathbb{B}^n_s$ with those of $\ell^\infty(\mathbb{Z}^n_\star)$; recall that since the functions in $\mathbb{B}^n_s$ are real-analytic one has the reality condition $f_k = \bar{f}_{-k}$..

Denote by $\mathbf{B}_1$ the closed ball of radius one in $\mathbb{B}^n_s$ and by $\mathcal{B}$ the Borellians in $\mathbf{B}_1$.

On $\mathbf{B}_1$ we can introduce the following natural (product) probability measure. Consider, first, the probability measure given by the normalized Lebesgue-product measure on the unit closed ball of $\ell^\infty(\mathbb{Z}^n_\star)$, namely, the unique probability measure μ on the Borellians of $\{z \in \ell^\infty(\mathbb{Z}^n_\star) : |z|_\infty \leqslant 1\}$ such that, given Lebesgue measurable sets A_k in the unit complex disk $\mathrm{A}_k \subseteq D := \{w \in \mathbb{C} : |w| \leqslant 1\}$ with $\mathrm{A}_k \neq D$ only for finitely many k, one has

$$\mu\left(\prod_{k\in\mathbb{Z}^n_\star} \mathrm{A}_k\right) = \prod_{\{k\in\mathbb{Z}^n_\star : \mathrm{A}_k\neq D\}} \frac{1}{\pi}\mathrm{meas}(\mathrm{A}_k)$$

where 'meas' denotes the Lebesgue measure on the unit complex disk D. Then, the isometry j in (57) naturally induces a probability measure μ_s on the Borellians $\mathcal{B}$, so that $\mu_s(\mathbf{B}_1) = 1$.

Remark 3.1.

(a) Since $f \in \mathbb{B}^n_s$, one has that $|f_k| \leqslant \|f\|_s\, \mathrm{e}^{-|k|_1 s}$ for all k's and (32) says that, when k is a generator of maximal 1d-lattices (later corresponding to simple resonances), the k-Fourier coefficient does not vanish and is controlled in a quantitive way from below: $|k|_1^{-n}$ is a suitable weight (needed in the proof of theorem 3.1 below), while δ is any number satisfiying

$$\inf_{|k|_1>\tau(\delta)} |f_k||k|_1^n\, \mathrm{e}^{|k|_1 s} \geqslant \delta > 0\,. \tag{58}$$

(b) It is easy to construct functions in $\mathcal{H}_{s,\tau}(\delta)$. For example let

$$f(x) := 2\delta \sum_{k\in\mathcal{G}^n_1} |k|_1^{-n}\, \mathrm{e}^{-|k|_1 s} \cos(k\cdot x)\,, \tag{59}$$

which has Fourier coefficients

$$f_k = \begin{cases} \delta|k|_1^{-n}\, \mathrm{e}^{-|k|_1 s}, & \text{if } \pm k \in \mathcal{G}^n_1 \\ 0, & \text{otherwise} \end{cases}$$

and 1d-Fourier projections

$$F^k(\theta) = \delta|k|_1^{-n}\, \mathrm{e}^{-|k|_1 s} \cos\theta.$$

Then, $f \in \mathcal{H}_{s,0}(\delta)$ and, also, $f \in \mathcal{H}_{s,\tau}(\delta)$ for any choice of tail function $\tau(\delta)$.

Recall that a Borel set P of a Banach space X is called prevalent if there exists a compactly supported probability measure ν on the Borellians of X such that $\nu(x + P) = 1$ for all $x \in X$; compare, e.g., [20] or [21].

Theorem 3.1 (Properties of $\mathcal{H}_{s,\tau}$). *Let $s > 0$ and τ be a tail function. Then:*

(*a*) *The set $\mathcal{H}_{s,\tau} \subseteq \mathbb{B}^n_s$ contains an open dense set.*
(*b*) *$\mathcal{H}_{s,\tau} \cap \mathbf{B}_1 \in \mathcal{B}$ and $\mu_s(\mathcal{H}_{s,\tau} \cap \mathbf{B}_1) = 1$.*
(*c*) *$\mathcal{H}_{s,\tau}$ is a prevalent set.*

Proof.

(a) $\mathcal{H}_{s,\tau}$ contains an open subset $\mathcal{H}'_{s,\tau}$ which is dense in the unit ball of $\mathbb{B}^n_s$.

Let us define $\mathcal{H}'_{s,\tau}$ as $\mathcal{H}_{s,\tau}$ but with the difference that (32) is replaced by the stronger condition

$$\exists\delta > 0\ \text{s.t.}\ |f_k| \geqslant \delta\, \mathrm{e}^{-|k|_1 s}, \quad \forall\, k \in \mathcal{G}^n_1,\ |k|_1 > \tau(\delta)\,; \tag{60}$$

(note, however, that $\mu_s(\mathcal{H}'_s) = 0$).

Let us first prove that $\mathcal{H}'_{s,\tau}$ is open. Let $f \in \mathcal{H}'_{s,\tau}$. We have to show that there exists $\rho > 0$ such that if $\|g\|_s < \rho$, then $f + g \in \mathcal{H}'_{s,\tau}$. Fix $\delta > 0$ such that (60) holds and, by continuity of $\tau(\delta)$, choose $\rho < \delta$ small enough such that $[\tau(\delta)] > \tau(\delta') - 1$, where $\delta' := \delta - \rho$ and $[\cdot]$ denotes integer part. Then, since $\tau(\delta)$ is not increasing, it is immediate to verify that $|k|_1 > \tau(\delta) \Longleftrightarrow |k|_1 > \tau(\delta')$. Moreover

$$|f_k + g_k|e^{|k|_1 s} \geqslant |f_k|e^{|k|_1 s} - \|g\|_s \geqslant \delta - \rho = \delta', \quad \forall\, k \in \mathcal{G}_1^n,\ |k|_1 > \tau(\delta'),$$

namely $f + g$ satisfies (60) (with δ' instead of δ).

Let us now show that $\mathcal{H}'_{s,\tau}$ is dense in the unit ball of $\mathbb{B}_s^n$. Take f in the unit ball of $\mathbb{B}_s^n$ and $0 < \lambda < 1$. We have to find $\tilde{f} \in \mathcal{H}'_{s,\tau}$ with $\|\tilde{f} - f\|_s \leqslant \lambda$. Let $\delta := \lambda/4$ and denote by f_k and $\tilde{f}_k$ (to be defined) be the Fourier coefficients of, respectively, f and $\tilde{f}$. We, then, let $\tilde{f}_k = f_k$ unless $k \in \mathcal{G}_1^n$, $|k|_1 > \tau(\delta)$ and $|f_k|e^{|k|_1 s} < \delta$, in which case, $\tilde{f}_k = \delta\, e^{-|k|_1 s}$. It is, now, easy to check that $\tilde{f} \in \mathcal{H}'_s$ and is λ-close to f.

(b) $\mathcal{H}_{s,\tau} \cap \mathbf{B}_1 \in \mathcal{B}$ and $\mu_s(\mathcal{H}_{s,\tau} \cap \mathbf{B}_1) = 1$

We shall prove that, for every $\delta > 0$, the measure of the sets of potentials f that do not satisfy (32) is $O(\delta^2)$, the result will follow letting $\delta \to 0$.

By the identification (57), the measure of the set of potentials f that do not satisfy (32) with a given δ is bounded by

$$\delta^2 \sum_{k \in \mathbb{Z}^n} |k|_1^{-2n}. \tag{61}$$

(c) $\mathcal{H}_{s,\tau}$ is prevalent.

Consider the following compact subset of $\ell^\infty(\mathbb{Z}_\star^n)$: let $\mathcal{K} := \left\{z = \{z_k\}_{k \in \mathbb{Z}_\star^n} : z_k \in D_{1/|k|_1}\right\}$, where $D_{1/|k|_1} := \{w \in \mathbb{C} : |w| \leqslant 1/|k|_1\}$, and let ν be the unique probability measure supported on $\mathcal{K}$ such that, given Lebesgue measurable sets $\mathrm{A}_k \subseteq D_{1/|k|_1}$, with $\mathrm{A}_k \neq D_{1/|k|_1}$ only for finitely many k, one has

$$\nu\left(\prod_{k \in \mathbb{Z}_\star^n} \mathrm{A}_k\right) := \prod_{\left\{k \in \mathbb{Z}_\star^n : \mathrm{A}_k \neq D_{1/|k|_1}\right\}} \frac{|k|_1^2}{\pi} \mathrm{meas}(\mathrm{A}_k).$$

The isometry j_s in (57) naturally induces a probability measure ν_s on $\mathbb{B}_s^n$ with support in the compact set $\mathcal{K}_s := j_s^{-1}\mathcal{K}$. Reasoning as in the proof of $\mu_s(\mathcal{H}_{s,\tau}) = 1$, one can show that $\nu_s(\mathcal{H}_{s,\delta}) \geqslant 1 - \mathrm{const}\,\delta^2$. It is also easy to check that, for every $g \in \mathbb{B}_s^n$, the translated set $\mathcal{H}_{s,\delta} + g$ satisfies $\nu_s(\mathcal{H}_{s,\delta} + g) \geqslant \nu_s(\mathcal{H}_{s,\delta})$. Thus, one gets $\nu_s(\mathcal{H}_{s,\tau} + g) = \nu_s(\mathcal{H}_{s,\tau}) = 1$, $\forall\, g \in \mathbb{B}_s^n$, which means that $\mathcal{H}_{s,\tau}$ is prevalent. □

Remark 3.2. As mentioned above, one could impose the condition $|f_k| \geqslant \delta |k|_1^{-\bar{n}}\, e^{-|k|_1 s}$ in (32). Then (61) would become $\delta^2 \sum_{k \in \mathbb{Z}^n} |k|_1^{-2\bar{n}}$, which is still fine if $\bar{n} > n/2$.

4. Normal form lemma

In this section we describe an analytic normal form lemma for nearly-integrable Hamiltonians $H(y, x) = h(y) + f(y, x)$, which allows to average out non-resonant Fourier modes of the perturbation f on suitable non-resonant regions, and allows for 'very small' analyticity loss in the angle variables, a fact, which will be crucial in our applications. In fact, the main point of the following normal form lemma is that the new averaged Hamiltonian is defined, in the fast variable (angle) domain, in a region almost equal to the original domain, 'almost equal' meaning a complex strip of width $s(1 - 1/K)$, if s is the width of the initial angle analyticity.

We recall ([11, 29]) that, given an integrable Hamiltonian $h(y)$, positive numbers α, K and a lattice $\Lambda \subset \mathbb{Z}^n$, a (real or complex) domain U is (α, K) non-resonant modulo Λ (with respect to h) if

$$|h'(y) \cdot k| \geqslant \alpha, \quad \forall\, y \in U, \forall\, k \in \mathbb{Z}^n \setminus \Lambda,\ |k|_1 \leqslant K. \tag{62}$$

We need also the following notations. Given $f(y,x) = \sum_{k\in\mathbb{Z}^n} f_k(y)\mathrm{e}^{\mathrm{i}k\cdot x}$ and a sublattice Λ of $\mathbb{Z}^n$, we denote

$$\mathrm{p}_\Lambda f := \sum_{k\in\Lambda} f_k(y)\mathrm{e}^{\mathrm{i}k\cdot x}, \quad \mathrm{p}_\Lambda^\perp f := \sum_{k\notin\Lambda} f_k(y)\mathrm{e}^{\mathrm{i}k\cdot x}. \tag{63}$$

Notice that:

$$|\,\mathrm{p}_\Lambda f|_{r,s},\ |\,\mathrm{p}_\Lambda^\perp f|_{r,s} \leqslant |f|_{r,s}. \tag{64}$$

Given $N > 0$, we let:

$$T_N f(y,x) := \sum_{|k|_1\leqslant N} f_k(y)\mathrm{e}^{\mathrm{i}k\cdot x}, \quad T_N^\perp f(y,x) := \sum_{|k|_1 > N} f_k(y)\mathrm{e}^{\mathrm{i}k\cdot x}. \tag{65}$$

Note that p_Λ and T_N commute and that

$$|T_N f|_{r,s},\ |T_N^\perp f|_{r,s} \leqslant |f|_{r,s}, \tag{66}$$

$$|T_N^\perp f|_{r,s-\sigma} \leqslant \mathrm{e}^{-(N+1)\sigma}|f|_{r,s}, \quad 0 < \sigma < s. \tag{67}$$

Proposition 4.1 (Normal form with 'small' analyticity loss)**.** *Let $r, s, \alpha > 0$, $K \in \mathbb{N}$, $K \geqslant 2$, $D \subseteq \mathbb{R}^n$, and let Λ be a lattice of $\mathbb{Z}^n$. Let*

$$H(y,x) = h(y) + f(y,x) \tag{68}$$

be real-analytic on $D_r \times \mathbb{T}^n_s$ with $|f|_{r,s} < \infty$. Assume that D_r is (α, K) non-resonant modulo Λ and that

$$\vartheta_\star := \frac{2^{11}K^2}{\alpha r s}|f|_{r,s} < 1. \tag{69}$$

Then, there exists a real-analytic symplectic change of variables

$$\Psi : (y', x') \in D_{r_\star} \times \mathbb{T}^n_{s_\star} \mapsto (y,x) \in D_r \times \mathbb{T}^n_s \quad \text{with} \quad r_\star := r/2,\ s_\star := s(1 - 1/K) \tag{70}$$

satisfying

$$|y - y'|_1 \leqslant \frac{\vartheta_\star}{2^7 K} r, \quad \max_{1\leqslant i\leqslant n}|x_i - x_i'| \leqslant \frac{\vartheta_\star}{16K^2}s, \tag{71}$$

and such that

$$H \circ \Psi = h + f^\flat + f_\star, \quad f^\flat := \mathrm{p}_\Lambda f + T_K^\perp\, \mathrm{p}_\Lambda^\perp f \tag{72}$$

with

$$|f_\star|_{r_\star,s_\star} \leqslant \frac{1}{K}\vartheta_\star |f|_{r,s}, \quad |T_K\, \mathrm{p}_\Lambda^\perp f_\star|_{r_\star,s_\star} \leqslant (\vartheta_\star/8)^K \frac{8}{eK}|f|_{r,s}. \tag{73}$$

Moreover, re-writing (72) as

$$H \circ \Psi = h + g + f_{\star\star} \quad \textit{where } \mathrm{p}_\Lambda g = g,\ \mathrm{p}_\Lambda f_{\star\star} = 0, \tag{74}$$

one has

$$|g - \mathrm{p}_\Lambda f|_{r_\star,s_\star} \leqslant \frac{1}{K}\vartheta_\star |f|_{r,s}, \quad |f_{\star\star}|_{r_\star,s/2} \leqslant 2\mathrm{e}^{-(K-2)\bar{s}}|f|_{r,s}, \tag{75}$$

where

$$\bar{s} := \min\left\{\frac{s}{2}, \log\frac{8}{\vartheta_\star}\right\}. \tag{76}$$

Remark 4.1.

(a) The 'novelty' of this lemma is that the bounds in (73) and the first one in (75) hold on the large angle domain $\mathbb{T}^n_{s_\star}$ with $s_\star = s(1 - 1/K)$. In particular the first estimate in (73) (or, equivalently, in (75)) will be important in our analysis in order to obtain (150), (153) and, therefore, (159), (166) and finally (171), which is the key to prove (43) in theorem 2.1. The drawback of the gain in angle-analyticity strip is that the power of K in the smallness condition (69) is not optimal: for example in [11] the power of K is one (but $s_\star = s/6$, which would not work in our applications).

(b) Having information on non-resonant Fourier modes up to order K, the best one can do is to average out the non-resonant Fourier modes up to order K, namely, to 'kill' the term $T_K\, \mathrm{p}^\perp_\Lambda f$ of the Fourier expansion of the perturbation. This explains the 'flat' term $f^\flat = \mathrm{p}_\Lambda f + T^\perp_K\, \mathrm{p}^\perp_\Lambda f$ surviving in (72) and which cannot be removed in general. Now, think of the remainder term $f_\star$ as

$$f_\star = \mathrm{p}_\Lambda f_\star + \left(T^\perp_K\, \mathrm{p}^\perp_\Lambda f_\star + T_K\, \mathrm{p}^\perp_\Lambda f_\star\right);$$

then, $\mathrm{p}_\Lambda f_\star$ is a $(\vartheta_\star |f|_{r,s}/K)$-perturbation of the part in normal form (i.e., with Fourier modes in Λ), while $T^\perp_K\, \mathrm{p}^\perp_\Lambda f_\star$ is, by (67), a term exponentially small with K (see also below) and $T_K\, \mathrm{p}^\perp_\Lambda f_\star$ is a very small remainder bounded by $8(\vartheta_\star/8)^K |f|_{r,s}/eK$.

(c) We note that (74) follows from (72). Indeed we take

$$g = \mathrm{p}_\Lambda f + \mathrm{p}_\Lambda f_\star, \quad f_{\star\star} = T^\perp_K\, \mathrm{p}^\perp_\Lambda f + \mathrm{p}^\perp_\Lambda f_\star = T_K\, \mathrm{p}^\perp_\Lambda f_\star + T^\perp_K\, \mathrm{p}^\perp_\Lambda (f_\star + f).$$

Then the first estimate in (75) follows by the first bound in (73) and (64). Regarding the second estimate in (75), we first note by (73) and (67), (used with f, r, s, σ and N, replaced, respectively, by $f_\star, r_\star, s_\star, \frac{s}{2} - \frac{s}{K}$ and K, so that $s_\star - \sigma = s/2$ and $\mathrm{e}^{-(K+1)\sigma} \leqslant \mathrm{e}^{-(K-2)s/2}$)

$$|T^\perp_K f_\star|_{r_\star, s/2} = |T^\perp_K f_\star|_{r_\star, s_\star - \sigma} \leqslant \mathrm{e}^{-(K+1)\sigma} |f_\star|_{r_\star, s_\star} \leqslant \mathrm{e}^{-(K-2)s/2} \vartheta_\star |f|_{r,s}/K.$$

By (64), (73) and (67) we get

$$\begin{aligned} |f_{\star\star}|_{r_\star, s/2} &\leqslant |T_K\, \mathrm{p}^\perp_\Lambda f_\star|_{r_\star, s/2} + |T^\perp_K f_\star|_{r_\star, s/2} + |T^\perp_K f|_{r_\star, s/2} \\ &\leqslant (\vartheta_\star/8)^K \frac{8}{eK} |f|_{r,s} + \mathrm{e}^{-(K-2)s/2} (\vartheta_\star/K + \mathrm{e}^{-3s/2}) |f|_{r,s} \\ &\leqslant 2\mathrm{e}^{-(K-2)\bar{s}} |f|_{r,s}. \end{aligned}$$

(d) Let us compare our results with more standard formulations, such as the normal form lemma in section 2 of [11]. In that formulation, imposing the weaker smallness condition $|f|_{r,s} \leqslant \text{const}\ \alpha r/K$, the normal form Hamiltonian writes $h + \mathtt{G} + \mathtt{f}$ with $\mathtt{f}$ exponentially small (of order $|f|_{r,s} \mathrm{e}^{-Ks/6}$) and, regarding $\mathtt{G}$ one knows that

$$|\mathtt{g} - T_K\, \mathrm{p}_\Lambda f|_{r/2, s/6} \leqslant \text{const}.\frac{K}{\alpha r} |f|^2_{r,s}. \tag{77}$$

For our purposes we need to prove that, when $k \in \mathbb{Z}^n_\star$, $|k|_1 \leqslant K_1 \leqslant K$ ($k \in \mathbb{Z}^n_{K_1}$ indexes the simple resonance we want to consider while $l \in \mathbb{Z}^n_K$ indexes the second order resonance beyond k) and $|f_k|/|f|_{r,s} \geqslant \delta|k|_1^{-n}\,\mathrm{e}^{-|k|_1 s}$, the quantity

$$\frac{1}{|f_k|}\sup_{y\in D_{r/2}}|g_k(y)-f_k|$$

is small. Indeed by (75) we have

$$\frac{1}{|f_k|}\sup_{y\in D_{r/2}}|g_k(y)-f_k| \leqslant \frac{\vartheta_\star}{K}|f|_{r,s}\frac{\mathrm{e}^{-|k|_1 s_\star}}{|f_k|} \leqslant \frac{\vartheta_\star}{K}\frac{|k|_1^n\,\mathrm{e}^{(s-s_\star)|k|_1}}{\delta}$$
$$= \frac{\vartheta_\star}{K}\frac{|k|_1^n\,\mathrm{e}^{s|k|_1/K}}{\delta} \leqslant \frac{\vartheta_\star}{K}\,\mathrm{e}^s\frac{K_1^n}{\delta}, \tag{78}$$

which is small when

$$K_1 \ll \left(\frac{\alpha r s\delta}{K|f|_{r,s}}\right)^{1/n}. \tag{79}$$

Consider, for example, the function $f=\varepsilon\hat{f}$ with ε small and $\hat{f}$ defined in (59). We have that $|f|_{r,s}=c\delta\varepsilon$, for a suitable constant $c>0$. In this case (79) writes

$$K_1 \ll \left(\frac{\alpha r s}{K\varepsilon}\right)^{1/n}. \tag{80}$$

On the other hand by estimate (77) one only have

$$\frac{1}{|f_k|}\sup_{y\in D_{r/2}}|\mathsf{g}_k(y)-f_k| \leqslant \text{const}.\frac{K\varepsilon}{\alpha r}|k|_1^n\,\mathrm{e}^{|k|_1 s}\,\mathrm{e}^{-|k|_1 s/6} \leqslant \text{const}.\frac{\varepsilon K}{\alpha r}K_0^n\,\mathrm{e}^{\frac{5}{6}K_0 s},$$

which is small only for

$$K_1 \ll \frac{6}{5s}\log\frac{\alpha r}{K\varepsilon}, \tag{81}$$

that is a considerably stronger bound than the one in (80).

Since we are considering simple resonances indexed by $|k|_1 \leqslant K_1$, the non resonant region will be non-resonant only up to order K_1; therefore we have that the perturbation, after normal form in the non-resonant region, will be of magnitude

$$\varepsilon\mathrm{e}^{-K_1 s/6} \gg \varepsilon(K\varepsilon/\alpha r)^{1/5},$$

when the bound (81) applies. This estimate is very bad. On the other hand, in our case, the weaker bound (80) applies and we obtain that the perturbation is exponentially small.

Next, let us recall a techincal lemma from [11]. Given a function ϕ we denote by X^t_ϕ the Hamiltonian flow at time t generated by ϕ and by 'ad' the linear operator

$$u \mapsto \mathrm{ad}_\phi u = \{u,\phi\} := \sum_{i=1}^n (u_{x_i}\phi_{y_i} - u_{y_i}\phi_{x_i})$$

and ad^ℓ its iterates:

$$\mathrm{ad}^0_\phi u := u, \quad \mathrm{ad}^\ell_\phi u := \{\mathrm{ad}^{\ell-1}_\phi u,\phi\}, \quad \ell \geqslant 1,$$

(as standard, $\{\cdot,\cdot\}$ denotes Poisson bracket).

Recall the identity ('Lie series expansion')

$$u\circ X^1_\phi=\sum_{\ell\geqslant 0}\frac{1}{\ell!}\mathrm{ad}^\ell_\phi u=\sum_{\ell=0}^{\infty}\frac{\partial_t^\ell(u\circ X^t_\phi)}{\ell!}|_{t=0}, \tag{82}$$

valid for analytic functions and small ϕ.

Lemma 4.1 (Lemma B.3 of [11]). *For $0<\rho<r$, $0<\sigma<s$, $D\subseteq\mathbb{R}^n$*

$$\sup_{y\in D_r}\sum_{1\leqslant i\leqslant n}|\partial_{x_i}\phi(y,\cdot)|_{s-\sigma}\leqslant\frac{1}{e\sigma}|\phi|_{r,s},\qquad \sup_{y\in D_{r-\rho}}\max_{1\leqslant i\leqslant n}|\partial_{y_i}\phi(y,\cdot)|_s\leqslant\frac{1}{\rho}|\phi|_{r,s},$$

By lemma 4.1 we get (see also lemma B.4 of [11])

Lemma 4.2. *For $0<\rho<\bar r:=\{r_0,r\}$, $0<\sigma<\bar s:=\{s_0,s\}$,*

$$|\{f,g\}|_{\bar r-\rho,\bar s-\sigma}\leqslant\frac{1}{e}\left(\frac{1}{(r_0-\bar r+\rho)(s-\bar s+\sigma)}+\frac{1}{(r-\bar r+\rho)(s_0-\bar s+\sigma)}\right)|f|_{r_0,s_0}|g|_{r,s}. \tag{83}$$

Summing the Lie series in (82) (see lemma B5 of [11]) we get, also,

Lemma 4.3. *Let $0<\rho<r_0$ and $0<\sigma<s_0$. Assume that*

$$\hat\vartheta:=\frac{4e|\phi|_{r_0,s_0}}{\rho\sigma}\leqslant 1. \tag{84}$$

Then for every $\rho<r'\leqslant r_0$, $\sigma<s'\leqslant s_0$, the time-1-flow X^1_ϕ of vector field X_ϕ define a good canonical transformation

$$X^1_\phi: D_{r'-\rho}\times\mathbb{T}^n_{s'-\sigma}\to D_{r'-\rho/2}\times\mathbb{T}^n_{s'-\sigma/2} \tag{85}$$

satisfying

$$|y-y'|_1\leqslant\hat\vartheta\frac{\rho}{4e},\qquad \max_{1\leqslant i\leqslant n}|x_i-x_i'|\leqslant\hat\vartheta\frac{\sigma}{4} \tag{86}$$

Moreover let $r>\rho, s>\sigma$ and set

$$\bar r:=\min\{r_0,r\},\quad \bar s:=\min\{s_0,s\}.$$

Then for any $j\geqslant 0$

$$\begin{aligned}|u\circ X^1_\phi-\sum_{h\leqslant j}\mathrm{ad}^h_\phi u|_{\bar r-\rho,\bar s-\sigma}&\leqslant\sum_{h>j}\frac{1}{h!}|\mathrm{ad}^h_\phi u|_{\bar r-\rho,\bar s-\sigma}\\ &\leqslant 2(\hat\vartheta/2)^j|\{u,\phi\}|_{\bar r-\rho/2,\bar s-\sigma/2}\end{aligned} \tag{87}$$

for every function u with $|u|_{r,s} < \infty$.
In particular when $r \leqslant r_0, s \leqslant s_0$

$$|u \circ X^1_\phi - u|_{r-\rho,s-\sigma} \leqslant \sum_{h\geqslant 1} \frac{1}{h!}|\mathrm{ad}^h_\phi u|_{r-\rho,s-\sigma} \leqslant 2\hat{\vartheta}|u|_{r,s}, \tag{88}$$

$$|u \circ X^1_\phi - u - \{u,\phi\}|_{r-\rho,s-\sigma} \leqslant \hat{\vartheta}^2|u|_{r,s}, \tag{89}$$

Proof. We first note that by lemma 4.1 (applied with r and s replaced, respectively, by r_0 and s_0) for every $(y,x) \in D_{r_0-\rho} \times \mathbb{T}^n_{s_0-\sigma}$ we have

$$|\partial_x \phi(y,x)|_1 \leqslant \frac{1}{e\sigma}|\phi|_{r_0,s_0} = \frac{\hat{\vartheta}\rho}{4e} \leqslant \frac{\rho}{4e}, \quad \max_{1\leqslant i\leqslant n} |\partial_{y_i}\phi(y,x)| \leqslant \frac{1}{\rho}|\phi|_{r_0,s_0} = \frac{\hat{\vartheta}\sigma}{4} \leqslant \frac{\sigma}{4}.$$

Then (85) holds.
For $h \geqslant 1$, set for brevity

$$|\cdot|_i := |\cdot|_{\bar{r}-\frac{\rho}{2}-i\tilde{\rho},\bar{s}-\frac{\sigma}{2}-i\tilde{\sigma}}, \quad 0 \leqslant i \leqslant h,\ \tilde{\rho} := \frac{\rho}{2h},\ \tilde{\sigma} := \frac{\sigma}{2h}.$$

We get

$$\begin{aligned} &|\mathrm{ad}^i_\phi\{u,\phi\}|_i \\ &\overset{(83)}{\leqslant} \frac{1}{e}\left(\frac{1}{\tilde{\rho}(s_0-\bar{s}+i\tilde{\sigma}+\sigma/2)} + \frac{1}{\tilde{\sigma}(r_0-\bar{r}+i\tilde{\rho}+\rho/2)}\right)|\phi|_{r_0,s_0}|\mathrm{ad}^{i-1}_\phi\{u,\phi\}|_{i-1} \\ &\leqslant \frac{8h^2}{e\rho\sigma}\frac{1}{h+i}|\phi|_{r_0,s_0}|\mathrm{ad}^{i-1}_\phi\{u,\phi\}|_{i-1}, \end{aligned}$$

and, iterating,

$$|\mathrm{ad}^h_\phi\{u,\phi\}|_h \leqslant \frac{8h^2}{e\rho\sigma}\frac{h!}{(2h)!}|\phi|_{r_0,s_0}|\{u,\phi\}|_{r-\rho/2,s-\sigma/2} \leqslant h!(\hat{\vartheta}/2)^h|\{u,\phi\}|_{r-\rho/2,s-\sigma/2}$$

by Stirling's formula. Then

$$\sum_{h\geqslant j}\frac{1}{(h+1)!}|\mathrm{ad}^{h+1}_\phi u|_{\bar{r}-\rho,\bar{s}-\sigma} \leqslant \sum_{h\geqslant j}\frac{1}{h+1}(\hat{\vartheta}/2)^h|\{u,\phi\}|_{r-\rho/2,s-\sigma/2}$$

proving (87) in view of (84).
Finally (88) and (89) follows by (87) and since $|\{u,\phi\}|_{\bar{r}-\rho/2,\bar{s}-\sigma/2} \leqslant 2\,\mathrm{e}^{-1}\hat{\vartheta}|u|_{r,s}$ by (83). □

Given $K \geqslant 2$ and a lattice Λ, recall the definition of $f^\flat$ in (72) and define

$$f^K := f - f^\flat = T_K\,\mathrm{p}^\perp_\Lambda f,$$

so that we have the decomposition (valid for any f):

$$f = f^\flat + f^K, \quad f^\flat := P_\Lambda f + T^\perp_K\,\mathrm{p}^\perp_\Lambda f, \quad f^K := T_K\,\mathrm{p}^\perp_\Lambda f. \tag{90}$$

Lemma 4.4. *Let* $0 < \rho < r$ *and* $0 < \sigma < s$. *Consider a real-analytic Hamiltonian*

$$H = H(y,x) = h(y) + f(y,x) \quad \textit{analytic on } D_r \times \mathbb{T}^n_s. \tag{91}$$

Suppose that D_r is (α, K) non-resonant modulo Λ for h (with $K \geqslant 2$). Assume that

$$\check{\vartheta} := \frac{4e}{\alpha\rho\sigma}|f^K|_{r,s} \leqslant 1. \tag{92}$$

Then there exists a real-analytic symplectic change of coordinates

$$\Psi := X^1_\phi : D_{r_+} \times \mathbb{T}^n_{s_+} \ni (y', x') \to (y, x) \in D_r \times \mathbb{T}^n_s, \quad r_+ := r - \rho,\ s_+ := s - \sigma,$$

generated by a function $\phi = \phi^K = T_K\, \mathtt{p}^\perp_\Lambda \phi$ with

$$|\phi|_{r,s} \leqslant |f^K|_{r,s}/\alpha, \tag{93}$$

satisfying

$$|y - y'|_1 \leqslant \check{\vartheta}\frac{\rho}{4e}, \quad \max_{1\leqslant i\leqslant n}|x_i - x_i'| \leqslant \check{\vartheta}\frac{\sigma}{4}, \tag{94}$$

such that

$$H \circ \Psi = h(y') + f_+(y', x'), \quad f_+ := f^\flat + f_\star \tag{95}$$

with

$$|f_\star|_{r_+,s_+} \leqslant 4\check{\vartheta}|f|_{r,s}. \tag{96}$$

Notice that, by (90) and (96), one has

$$f^K_+ = f^K_\star, \quad |f_+|_{r_+,s_+} \leqslant |f_\star|_{r_+,s_+} + |f|_{r,s} \leqslant (1 + 4\check{\vartheta})|f|_{r,s}\,. \tag{97}$$

Notice also that

$$f^\flat_+ - f^\flat \overset{(95)}{=} f^\flat_\star \quad \Rightarrow \quad |f^\flat_+ - f^\flat|_{r_+,s_+} \leqslant |f_\star|_{r_+,s_+} \overset{(96)}{\leqslant} 4\check{\vartheta}|f|_{r,s}\,. \tag{98}$$

Proof. Let us define

$$\phi = \phi(y, x) := \sum_{|m|\leqslant K, m\notin\Lambda} \frac{f_m(y)}{\mathrm{i}h'(y)\cdot m}\,\mathrm{e}^{\mathrm{i}m\cdot x},$$

and note that ϕ solves the homological equation

$$\{h, \phi\} + f^K = 0. \tag{99}$$

Since D_r is (α, K) non-resonant modulo Λ the estimate (93) holds. We now use lemma 4.3 with r_0 and s_0 replaced, respectively, by r and s. With these choices it is $\hat{\vartheta} = \check{\vartheta}$, and, by (92) $\check{\vartheta} \leqslant 1$. Thus, (84) holds and lemma 4.3 applies. (94) follows by (86). We have

$$H \circ \Psi = h + f^\flat + f_\star$$

with

$$f_\star = (h \circ \Psi - h - \{h, \phi\}) + (f \circ \Psi - f).$$

Since

$$h \circ \Psi - h - \{h, \phi\} = \sum_{\ell\geqslant 2} \frac{1}{\ell!}\mathrm{ad}^\ell_\phi h = \sum_{\ell\geqslant 1} \frac{1}{(\ell+1)!}\mathrm{ad}^\ell_\phi \{h, \phi\} \overset{(99)}{=} -\sum_{\ell\geqslant 1} \frac{1}{(\ell+1)!}\mathrm{ad}^\ell_\phi f^K,$$

we have

$$|h\circ\Psi - h - \{h,\phi\}|_{r_+,s_+} \leqslant \sum_{\ell\geqslant 1}\frac{1}{\ell!}|\mathrm{ad}_\phi^\ell f^K|_{r_+,s_+} \overset{(88)}{\leqslant} 2\check{\vartheta}|f^K|_{r,s} \leqslant 2\check{\vartheta}|f|_{r,s}.$$

Finally, applying again lemma 4.3 with $u=f$, by (88), we get $|f\circ\Psi - f|_{r_+,s_+} \leqslant 2\check{\vartheta}|f|_{r,s}$, proving (96) and concluding the proof of lemma 4.4. □

As a preliminary step we apply lemma 4.4 to the Hamiltonian $H=h+f$ in (68) with $\rho=r/4$ and $\sigma=s/2K$. By (64), (66), (90) and (69) hypothesis (92) holds, namely

$$\vartheta_{-1} := \frac{2^5 eK}{\alpha rs}|f^K|_{r,s} \leqslant 1. \tag{100}$$

Then there exists a real-analytic symplectic change of coordinates

$$\Psi_{-1}: D_{r_0}\times\mathbb{T}^n_{s_0} \ni (y^{(0)},x^{(0)}) \to (y,x)\in D_r\times\mathbb{T}^n_s, \quad r_0 := \frac{3}{4}r,\ s_0 := \left(1-\frac{1}{2K}\right)s,$$

satisfying

$$|y-y^{(0)}|_1 \leqslant \vartheta_{-1}\frac{r}{16e}, \quad \max_{1\leqslant i\leqslant n}|x_i - x_i^{(0)}| \leqslant \vartheta_{-1}\frac{s}{8K}, \tag{101}$$

such that

$$H\circ\Psi_{-1} =: H_0 = h(y^{(0)}) + f_0(y^{(0)},x^{(0)}), \quad f_0 = f^\flat + f_\star,\ f^\flat := P_\Lambda f + T_K^\perp \mathrm{p}_\Lambda^\perp f, \tag{102}$$

with

$$|f_\star|_{r_0,s_0} \leqslant 4\vartheta_{-1}|f|_{r,s}. \tag{103}$$

Recalling (90) and (102) we get

$$f_0^K = f_\star^K$$

and, by (103) and (100),

$$|f_0^K|_{r_0,s_0} \leqslant 4\vartheta_{-1}|f|_{r,s} \leqslant \frac{2^7 eK}{\alpha rs}|f|^2_{r,s}. \tag{104}$$

Then, setting

$$\vartheta_0 := \delta|f_0^K|_{r_0,s_0} \quad \text{with}\ \delta := \frac{2^5 eK^3}{\alpha rs}, \tag{105}$$

we have

$$\vartheta_0 \leqslant \left(\frac{2^6 eK^2}{\alpha rs}|f|_{r,s}\right)^2 \overset{(69)}{\leqslant} (\vartheta_\star/8)^2 \leqslant \frac{1}{2^6}. \tag{106}$$

Finally, since $f_0^\flat - f^\flat = f_\star^\flat$ by (98) we get

$$|f_0^\flat - f^\flat|_{r_0,s_0} \leqslant 4\vartheta_{-1}|f|_{r,s} \overset{(100)}{\leqslant} \frac{2^7 eK}{\alpha rs}|f|^2_{r,s} \overset{(69)}{\leqslant} \frac{1}{4K}\vartheta_\star|f|_{r,s}. \tag{107}$$

The idea is to construct Ψ by applying K times lemma 4.4.

Let

$$\rho := \frac{r}{4K}, \quad \sigma := \frac{s}{2K^2},$$
$$r_i := \frac{3}{4}r - i\rho, \quad s_i := \left(1 - \frac{1}{2K}\right)s - i\sigma, \quad |\cdot|_i := |\cdot|_{r_i,s_i}, \tag{108}$$

Fix $1 \leqslant j \leqslant K$ and make the following inductive assumptions:

Assume that there exist, for $1 \leqslant i \leqslant j$, real-analytic symplectic transformations

$$\Psi_{i-1} := X^1_{\phi_{i-1}} : D_{r_i} \times \mathbb{T}^n_{s_i} \ni (y^{(i)}, x^{(i)}) \to (y^{(i-1)}, x^{(i-1)}) \in D_{r_{i-1}} \times \mathbb{T}^n_{s_{i-1}},$$

generated by a function $\phi_{i-1} = \phi^K_{i-1}$ with

$$|\phi_{i-1}|_{i-1} \leqslant |f^K_{i-1}|_{i-1}/\alpha, \tag{109}$$

satisfying

$$|y^{(i-1)} - y^{(i)}|_1 \leqslant \vartheta_{i-1}\frac{r}{16eK}, \quad \max_{1\leqslant\ell\leqslant n}|x^{(i-1)}_\ell - x^{(i)}_\ell| \leqslant \vartheta_{i-1}\frac{s}{8K^2}, \tag{110}$$

such that

$$H_i := H_{i-1} \circ \Psi_{i-1} =: h + f_i = h + f^K_i + f^\flat_i \tag{111}$$

satisfies, for $1 \leqslant i \leqslant j$, the estimates

$$\vartheta_i \leqslant \left(\frac{2^8K^2|f|_{r,s}}{\alpha rs}\right)^{i+1} \overset{(69)}{=} \left(\frac{\vartheta_\star}{8}\right)^{i+1}, \quad |f^\flat_i - f^\flat_{i-1}|_i \leqslant \frac{1}{\delta}\left(\frac{\vartheta_\star}{8}\right)^{i+1}, \tag{112}$$

where

$$\vartheta_i := \delta|f^K_i|_i. \tag{113}$$

Let us first show that the inductive hypothesis is true for $j = 1$ (which implies $i = 1$). Indeed by (106) we see that we can apply lemma 4.4 with f and $\check{\vartheta}$ replaced, respectively, by f^K_0 and $\vartheta_0 = \delta|f^K_0|_0$. Thus, we obtain the existence of $\Psi_0 = X^1_{\phi_0}$, generated by a function $\phi_0 = \phi^K_0$ with

$$|\phi_0|_{r_0,s_0} \leqslant \frac{1}{\alpha}|f^K_0|_{r_0,s_0} \overset{(104)}{\leqslant} \frac{2^7eK}{\alpha^2 rs}|f|^2_{r,s}, \tag{114}$$

satisfying (109) and (110), so that $(h + f^K_0) \circ \Psi_0 =: h + \tilde{f}_1$ and, by (95) and (96),

$$|\tilde{f}_1|_1 \leqslant 4\vartheta_0|f^K_0|_0 \overset{(106)}{\leqslant} \frac{1}{4}|f^K_0|_0 \overset{(104)}{\leqslant} \frac{2^5eK}{\alpha rs}|f|^2_{r,s}; \tag{115}$$

note that $(f^K_0)^\flat = 0$. We have that $f_1 = \tilde{f}_1 + f^\flat_0 \circ \Psi_0$. Then, since $(f^\flat_0)^K = 0$ and $(f^\flat_0)^\flat = f^\flat_0$, one finds

$$f^K_1 = \tilde{f}^K_1 + (f^\flat_0 \circ \Psi_0 - f^\flat_0)^K, \quad f^\flat_1 - f^\flat_0 = \tilde{f}^\flat_1 + (f^\flat_0 \circ \Psi_0 - f^\flat_0)^\flat. \tag{116}$$

Write

$$f^\flat_0 \circ \Psi_0 - f^\flat_0 = (f^\flat_0 - f^\flat) \circ \Psi_0 - (f^\flat_0 - f^\flat) + (f^\flat \circ \Psi_0 - f^\flat - \{f^\flat, \phi_0\}) + \{f^\flat, \phi_0\}.$$

By (88) (with u, r and s replaced by $f_0^\flat - f^\flat$, r_0 and s_0) we have

$$|(f_0^\flat - f^\flat)\circ\Psi_0 - (f_0^\flat - f^\flat)|_1 \leqslant 2\vartheta_0|f_0^\flat - f^\flat|_0 \leqslant \frac{2^4 eK}{\alpha rs}|f|_{r,s}^2$$

by (106) and (107). By (87) (with u, ϕ, j, $\bar r$ and $\bar s$ replaced, respectively, by $f^\flat$, ϕ_0, 1, r_0 and s_0)

$$|f^\flat\circ\Psi_0 - f^\flat - \{f^\flat,\phi_0\}|_1 \leqslant 2\vartheta_0|\{f^\flat,\phi_0\}|_{r_0-\rho/2,s_0-\sigma/2} \leqslant \frac{2^9K^3}{\alpha^2r^2s^2}|f|_{r,s}^3 \overset{(69)}{\leqslant} \frac{K}{4\alpha rs}|f|_{r,s}^2,$$

by (106), (114) and (83) (with f and g replaced by ϕ_0 and $f^\flat$). Analogously by (83) we get

$$|\{f^\flat,\phi_0\}|_1 \leqslant \frac{2^4K^2}{ers}|\phi_0|_0\,|f|_{r,s} \overset{(114)}{\leqslant} \frac{2^{11}K^3}{\alpha^2r^2s^2}|f|_{r,s}^3 \overset{(69)}{\leqslant} \frac{K}{\alpha rs}|f|_{r,s}^2.$$

Summarising:

$$|f_0^\flat\circ\Psi_0 - f_0^\flat|_1 \leqslant \frac{2^6K}{\alpha rs}|f|_{r,s}^2.$$

Then, by (115) and (116) we get

$$|f_1^K|_1\,,\,|f_1^\flat - f_0^\flat|_1 \;\leqslant\; \frac{2^7K}{\alpha rs}|f|_{r,s}^2 \tag{117}$$

checking (112) in the case $j = i = 1$.

Now take $2 \leqslant j \leqslant K$ and assume that the inductive hypothesis holds true for $1 \leqslant i \leqslant j$ and let us prove that it holds also for $i = j+1$. By (112) and (69) we can apply lemma 4.4 (with f and $\check\vartheta$ replaced by f_j^K and ϑ_j). Thus, we obtain the existence of $\Psi_j = X^1_{\phi_j}$, generated by a function $\phi_j = \phi_j^K$ with

$$|\phi_j|_j \overset{(109)}{\leqslant} \frac{1}{\alpha}|f_j^K|_j \overset{(113)}{=} \frac{\vartheta_j}{\alpha\delta}, \tag{118}$$

so that $(h + f_j^K)\circ\Psi_j =: h + \tilde f_{j+1}$ and, by (95) and (96),

$$|\tilde f_{j+1}|_{j+1} \leqslant 4\vartheta_j|f_j^K|_j \overset{(113)}{=} \frac{4}{\delta}\vartheta_j^2 \overset{(112)}{\leqslant} \frac{4}{\delta}(\vartheta_\star/8)^{2j+2} \overset{(69)}{\leqslant} \frac{1}{2^{3j-2}\delta}(\vartheta_\star/8)^{j+2} \leqslant \frac{1}{2^4\delta}(\vartheta_\star/8)^{j+2}, \tag{119}$$

since $j \geqslant 2$. We have that $f_{j+1} = \tilde f_{j+1} + f_j^\flat\circ\Psi_j$. Then (as above, $(f_j^\flat)^K = 0$ and $(f_j^\flat)^\flat = f_j^\flat$)

$$f_{j+1}^K = \tilde f_{j+1}^K + (f_j^\flat\circ\Psi_j - f_j^\flat)^K, \quad f_{j+1}^\flat - f_j^\flat = \tilde f_{j+1}^\flat + (f_j^\flat\circ\Psi_j - f_j^\flat)^\flat. \tag{120}$$

Writing

$$f_j^\flat = f^\flat + (f_0^\flat - f^\flat) + \sum_{h=1}^{j} f_h^\flat - f_{h-1}^\flat$$

we have

$$
\begin{aligned}
f_j^\flat \circ \Psi_j - f_j^\flat = \{f^\flat, \phi_j\} \\
&+ f^\flat \circ \Psi_j - f^\flat - \{f^\flat, \phi_j\} \\
&+ (f_0^\flat - f^\flat) \circ \Psi_j - (f_0^\flat - f^\flat) \\
&+ \sum_{i=1}^{j} \Big((f_i^\flat - f_{i-1}^\flat) \circ \Psi_j - (f_i^\flat - f_{i-1}^\flat) \Big)
\end{aligned}
\tag{121}
$$

where $\Psi_j = X^1_{\phi_j}$. By (83) (with f, g, r_0 and s_0 replaced, respectively, by ϕ_j, $f^\flat$, r_j and s_j) we get, by (109) and (113),

$$
|\{f^\flat, \phi_j\}|_{j+1} \leqslant \frac{2^4 K^2}{ers} |\phi_j|_j |f|_{r,s} \leqslant \frac{2^4 K^2 \vartheta_j}{e\alpha rs\delta} |f|_{r,s} \overset{(69)}{=} \frac{1}{e 2^4 \delta} (\vartheta_\star/8) \vartheta_j \overset{(112)}{\leqslant} \frac{1}{e 2^4 \delta} (\vartheta_\star/8)^{j+2}
$$

By (87) (with u, ϕ, j, $\bar{r}$ and $\bar{s}$ replaced, respectively, by $f^\flat$, ϕ_j, 1, r_j and s_j) reasoning as above we get

$$
|f^\flat \circ \Psi_j - f^\flat - \{f^\flat, \phi_j\}|_{j+1} \leqslant \vartheta_j |\{f^\flat, \phi_j\}|_{r_j - \rho/2, s_j - \sigma/2} \leqslant \frac{\vartheta_j}{4e\delta} (\vartheta_\star/8)^{j+2} \leqslant \frac{1}{2^6 e\delta} (\vartheta_\star/8)^{j+2}
$$

by (112) and (69). By (88) (with u, r and s replaced, respectively, by $f_0^\flat - f^\flat$, r_j, s_j) we have

$$
|(f_0^\flat - f^\flat) \circ \Psi_j - (f_0^\flat - f^\flat)|_{j+1} \leqslant 2\vartheta_j |f_0^\flat - f^\flat|_j \leqslant \frac{2^8 eK}{\alpha rs} |f|^2_{r,s} \vartheta_j \leqslant \frac{1}{4\delta} (\vartheta_\star/8)^{j+2}
$$

by (107), (112), (105) and (69). Analogously, for $1 \leqslant i \leqslant j$, by (88) (with u replaced by $f_i^\flat - f_{i-1}^\flat$)

$$
|(f_i^\flat - f_{i-1}^\flat) \circ \Psi_j - (f_i^\flat - f_{i-1}^\flat)|_{j+1} \leqslant 2\vartheta_j |f_i^\flat - f_{i-1}^\flat|_j \leqslant \frac{2}{\delta} (\vartheta_\star/8)^{j+i+2}
$$

by (112). Then by (69)

$$
\Big|\sum_{i=1}^{j} \Big((f_i^\flat - f_{i-1}^\flat) \circ \Psi_j - (f_i^\flat - f_{i-1}^\flat) \Big) \Big|_{j+1} \leqslant \frac{2}{7\delta} (\vartheta_\star/8)^{j+2}.
$$

Whence:

$$
|f_j^\flat \circ \Psi_j - f_j^\flat|_{j+1} \leqslant \frac{4}{7\delta} (\vartheta_\star/8)^{j+2}.
$$

Then by (119) we get

$$
|\tilde{f}_{j+1}|_{j+1} + |f_j^\flat \circ \Psi_j - f_j^\flat|_{j+1} \leqslant \frac{1}{\delta} (\vartheta_\star/8)^{j+2}.
$$

By (120) we get (112) with $i = j+1$. This completes the proof of the induction.

Now, we can conclude the proof of proposition 4.1. Set

$$
\Psi := \Psi_{-1} \circ \Psi_0 \circ \cdots \circ \Psi_{K-1}.
$$

Notice that, by (108), $r_K = r/2 = r_\star$ and $s_K = s(1-1/K) = s_\star$. By the induction, it is

$$H \circ \Psi = H_{K-1} \circ \Psi_{K-1} \overset{(111)_K}{=} h + f_K =: h + f^\flat + f_\star, \tag{122}$$

with $f^\flat = \mathtt{p}_\Lambda f + T_K^\perp \mathtt{p}_\Lambda^\perp f$ (recall (72)). Note that by (112) and (106)

$$\sum_{i=1}^K \vartheta_{i-1} \leqslant \sum_{i=1}^K (\vartheta_\star/8)^i \leqslant \vartheta_\star/7. \tag{123}$$

Since $(y', x') = (y^{(K)}, x^{(K)})$ by (101), (110) and triangular inequality we get

$$|y'-y|_1 \leqslant |y-y^{(0)}|_1 + \sum_{i=1}^K |y^{(i)} - y^{(i-1)}|_1 \leqslant \frac{r\vartheta_{-1}}{16e} + \frac{r}{16eK}\sum_{i=1}^K \vartheta_{i-1}$$
$$\overset{(123)}{\leqslant} \frac{r}{16e}\left(\vartheta_{-1} + \frac{\vartheta_\star}{7K}\right) \overset{(100)}{\leqslant} \frac{r}{16e}\left(\frac{\vartheta_\star}{8K} + \frac{\vartheta_\star}{7K}\right),$$

then (71) follows (the estimate on the angle being analogous).

Since $T_K P_\Lambda^\perp f^\flat = (f^\flat)^K = 0$ (for any f, recall (90)) we have

$$|T_K P_\Lambda^\perp f_\star|_{r_\star, s_\star} = |f_K^K|_K \overset{(113)}{=} \delta^{-1}\vartheta_K \overset{(112)}{\leqslant} \delta^{-1}(\vartheta_\star/8)^{K+1} = (\vartheta_\star/8)^K \frac{8}{eK}|f|_{r,s}, \tag{124}$$

proving the second estimates in (73).

Finally, (using that $K \geqslant 2$ and that $\vartheta_\star \leqslant 1$)

$$|f_\star|_{r_\star,s_\star} \overset{(122)}{=} |f_K - f^\flat|_K \overset{(90)}{=} |f_K^K + f_K^\flat - f^\flat|_K \leqslant |f_K^K|_K + |f_0^\flat - f^\flat|_0 + \sum_{i=1}^K |f_i^\flat - f_{i-1}^\flat|_i$$
$$\overset{(107),(112)}{\leqslant} |f_K^K|_K + \frac{1}{4K}\vartheta_\star |f|_{r,s} + \frac{1}{\delta}\sum_{i=1}^K (\vartheta_\star/8)^{i+1}$$
$$\overset{(124),(123)}{\leqslant} (\vartheta_\star/8)^K \frac{8}{eK}|f|_{r,s} + \frac{1}{4K}\vartheta_\star|f|_{r,s} + \frac{\vartheta_\star^2}{56\delta} \leqslant \frac{1}{K}\vartheta_\star |f|_{r,s},$$

which proves also the first estimate in (73). □

5. Geometry of resonances and proof of the covering lemma

We start by observing that from the definitions given in section 2 (in particular, (23) ÷ (28)), it follows at once that

$$\Omega^0 \cup \Omega^1 \cup \Omega^2 \supset B_M(0)\,. \tag{125}$$

Next, let us point out the non-resonance properties satisfied by the frequencies in Ω^i.

(a) If $0 \neq |k|_1 \leqslant K_1$, then there exists $\bar{k} \in \mathcal{G}^n_{1,K_1}$ and a $0 \neq j \in \mathbb{Z}$ such that $k = j\bar{k}$ and, therefore,

$$\omega \in \Omega^0 \implies |\omega \cdot k| = |j||\omega \cdot \bar{k}| \geqslant |\omega \cdot \bar{k}| \geqslant \min_{k \in \mathcal{G}^n_{1,K_1}} |\omega \cdot k| > \alpha/2\,. \tag{126}$$

(b) Let $\omega \in \Omega^{1,k}$ with $k \in \mathcal{G}^n_{1,K_1}$ and let $\ell \notin \mathbb{Z}k$, $|\ell|_1 \leqslant K_2$. Then, there exist $j \in \mathbb{Z}\backslash\{0\}$ and $\ell' \in \mathcal{G}^n_{1,K_2}$ such that $\ell = j\ell'$. Hence,

$$|\omega \cdot \ell| = |j||\omega \cdot \ell'| \geqslant |\omega \cdot \ell'| = \left| \frac{(\omega \cdot k)(k \cdot \ell')}{|k|^2} + \mathrm{p}_k^\perp \omega \cdot \ell' \right|$$
$$\geqslant |\,\mathrm{p}_k^\perp \omega \cdot \ell'| - \frac{\alpha K_2}{|k|} > \frac{3\alpha K_2}{|k|} - \frac{\alpha K_2}{|k|} = \frac{2\alpha K_2}{|k|}. \tag{127}$$

(c) It remains to evaluate the measure of Ω^2. To do this, we first prove the following

Lemma 5.1. *If $\omega \in \Omega^2_{k,\ell}$ with $k \in \mathcal{G}^n_{1,K_1}$, $\ell \in \mathcal{G}^n_{1,K_2}$, $\ell \notin \mathbb{Z}k$, then*

$$\mathrm{dist}(\omega, R_{k,\ell}) \leqslant \sqrt{10}\alpha K_2 |k||\ell|. \tag{128}$$

Moreover,

$$meas\,(\Omega^2_{k,\ell}) \leqslant 3 \cdot 2^n M^{n-2} \alpha^2 \frac{K_2}{|k|}. \tag{129}$$

Proof. Let $v \in \mathbb{R}^n$ be the projection of ω onto $R^\perp_{k,\ell}$, which is the plane generated by k and ℓ (recall that, by hypothesis, k and ℓ are not parallel). Then,

$$\mathrm{dist}\,(\omega, R_{k,\ell}) = \mathrm{dist}\,(v, R_{k,\ell}) = |v| \tag{130}$$

and

$$|v \cdot k| = |\omega \cdot k| < \alpha, \quad |\,\mathrm{p}_k^\perp v \cdot \ell| = |\,\mathrm{p}_k^\perp \omega \cdot \ell| \leqslant 3\alpha K_2/|k|. \tag{131}$$

Set

$$h := \mathrm{p}_k^\perp \ell = \ell - \frac{\ell \cdot k}{|k|^2} k. \tag{132}$$

Then, v decomposes in a unique way as

$$v = ak + bh$$

for suitable $a, b \in \mathbb{R}$. By (131),

$$|a| < \frac{\alpha}{|k|^2}, \quad |\,\mathrm{p}_k^\perp v \cdot \ell| = |bh \cdot \ell| \leqslant 3\alpha K_2/|k|, \tag{133}$$

and

$$|h \cdot \ell| \overset{(132)}{=} \frac{|\ell|^2|k|^2 - (\ell \cdot k)^2}{|k|^2} \geqslant \frac{1}{|k|^2},$$

since $|\ell|^2|k|^2 - (\ell \cdot k)^2$ is a positive integer (recall, that k and ℓ are integer vectors not parallel). Hence,

$$|b| \leqslant 3\alpha K_2 |k|, \tag{134}$$

and (128) follows since $|h| \leqslant |\ell|$ and $|v| = \sqrt{a^2|k|^2 + b^2|h|^2} \leqslant \sqrt{10}\alpha K_2 |k||\ell|$.

To estimate the measure of $\Omega^2_{k,\ell}$ we write $\omega \in R_{k,\ell}$ as $\omega = v + v^\perp$ with $v^\perp$ in the orthogonal complement of the plane generated by k and ℓ. Since $|v^\perp| \leqslant |\omega| < M$ and v lies in a rectangle of sizes of length $2\alpha/|k|^2$ and $6\alpha K_2|k|$ (compare (133) and (134)) we find

$$\text{meas}\,(\Omega^2_{k,\ell}) \leqslant \frac{2\alpha}{|k|^2}(6\alpha K_2|k|)(2M)^{n-2} = 3\cdot 2^n M^{n-2}\alpha^2\frac{K_2}{|k|}\,, \tag{135}$$

finishing the proof of lemma 5.1. □

From (28) and (135) it follows immediately (recall that $n \geqslant 2$) that

$$\text{meas}\,(\Omega^2) \leqslant cM^{n-2}\alpha^2K_2^{n+1}K_1^{n-1}\,, \tag{136}$$

for a suitable constant c depending only on n.

At this point the proof of proposition 2.1 follows at once: recalling the definition of $\bar{L}$ in definition 2.1, (125) implies (18), while (126), (127) and (136) imply immediately (19), (20) and (21) respectively, proving proposition 2.1. □

6. Averaging theory

Putting together the normal form lemma (proposition 4.1) and the covering lemma (proposition 2.1) there follows easily the following averaging theorem for non-resonant and simply resonant zones.

Assumption A. Let $r, s > 0$ and let h be KAM non-degenerate (definition 2.1).

Let $f : D_r \times \mathbb{T}^n_s \to \mathbb{C}$ be a holomorphic function with

$$\|f\|_{D,r,s} = 1 \tag{137}$$

and define

$$H_\varepsilon(y,x) := h(y) + \varepsilon f(y,x)\,, \qquad (y,x) \in D_r \times \mathbb{T}^n_s\,,\ \varepsilon > 0\,. \tag{138}$$

Let K_2, K_1, ν and α be such that

$$K_2 \geqslant 3K_1 \geqslant 6, \qquad \nu \geqslant n+2, \qquad \alpha := \sqrt{\varepsilon}K_2^\nu. \tag{139}$$

For $k \in \mathcal{G}^n_{1,K_1}$, define

$$r_0 := \frac{\alpha}{4LK_1} = \sqrt{\varepsilon}\frac{K_2^\nu}{4LK_1}; \qquad r_k := \frac{\alpha}{L|k|} = \sqrt{\varepsilon}\frac{K_2^\nu}{L|k|}, \tag{140}$$

$$\bar{\vartheta} := 2^{14}n^{2n}\frac{L}{s^{2n+1}}\frac{1}{K_2^{2\nu-2n-3}}; \qquad \vartheta := 2^{2n+10}n^{2n}\frac{L}{s^{2n+1}}\frac{1}{K_2^{2\nu-2n-3}}. \tag{141}$$

For later use, we observe that (38) implies:

$$\vartheta \leqslant \gamma\delta. \tag{142}$$

Theorem 6.1. *Let assumption A hold and assume that ε satisfies (40) and*

$$K_2^{2\nu-n-4} \geqslant 2^{13+n}n^n\frac{L\,\mathrm{e}^{s/2}}{s^{n+1}}\,. \tag{143}$$

Then the following holds.

(*a*) *There exists a symplectic change of variables*

$$\Psi_0 : D^0_{r_0/2} \times \mathbb{T}^n_{s(1-1/K_1)^2} \to D^0_{r_0} \times \mathbb{T}^n_{s(1-1/K_1)}, \tag{144}$$

such that

$$H_\varepsilon \circ \Psi_0 = h(y) + \varepsilon g^{\mathrm{o}}(y) + \varepsilon f^o_{\star\star}(y,x), \quad \langle f^o_{\star\star}\rangle = 0, \tag{145}$$

where $\langle\cdot\rangle = \mathrm{p}_{\{0\}}$ *denotes the average with respect to the angles x and*

$$\sup_{D^0_{r_0/2}} |g^{\mathrm{o}} - \langle f\rangle| \leqslant \bar{\vartheta}, \quad |f^{\mathrm{o}}_{\star\star}|_{D^0,r_0/2,s(1-1/K_1)/2} \leqslant 2\left(\frac{2nK_1}{s}\right)^n \mathrm{e}^{-(K_1-3)s/2}. \tag{146}$$

(*b*) $D^1 = \bigcup_{k\in\mathcal{G}^n_{1,K_1}} D^{1,k}$ *and for any* $k \in \mathcal{G}^n_{1,K_1}$ *there exists a symplectic change of variables*

$$\Psi_k : D^{1,k}_{r_k/2} \times \mathbb{T}^n_{s_\star} \to D^{1,k}_{r_k} \times \mathbb{T}^n_{s(1-1/K_2)}, \quad s_\star := s(1-1/K_2)^2, \tag{147}$$

such that

$$H_\varepsilon \circ \Psi_k = h(y) + \varepsilon g^k(y,x) + \varepsilon f^k_{\star\star}(y,x) \tag{148}$$

where

$$g^k = \mathrm{p}_{\mathbb{Z}k} g^k, \quad \mathrm{p}_{\mathbb{Z}k} f^k_{\star\star} = 0, \tag{149}$$

and

$$|g^k - \mathrm{p}_{\mathbb{Z}k} f|_{D^{1,k},r_k/2,s_\star} \leqslant \vartheta, \quad |f^k_{\star\star}|_{D^{1,k},r_k/2,s(1-1/K_2)/2} < 2\left(\frac{2nK_2}{s}\right)^n \mathrm{e}^{-(K_2-3)s/2}. \tag{150}$$

Remark 6.1.

(a) The functions g^k and $\mathrm{p}_{\mathbb{Z}k} f$ depend, effectively, only on one angle $\theta \in \mathbb{T}^1$: more precisely, setting

$$\begin{cases} F^k_j(y) := f_{jk}(y) \\ G^k_j(y) := g^k_{jk}(y) \end{cases} \qquad \begin{cases} F^k(y,\theta) := \sum_{j\in\mathbb{Z}, j\neq 0} F^k_j(y)\mathrm{e}^{\mathrm{i}j\theta} \\ G^k(y,\theta) := \sum_{j\in\mathbb{Z}, j\neq 0} G^k_j(y)\mathrm{e}^{\mathrm{i}j\theta} \end{cases} \tag{151}$$

we have (recall (54))

$$(\mathrm{p}_{\mathbb{Z}k} f)(y,x) = F^k(y, k\cdot x), \quad g^k(y,x) = g^k_0(y) + G^k(y, k\cdot x). \tag{152}$$

From (150) and (55) (and recalling that f has zero average) it follows

$$\sup_{D^{1,k}_{r_k/2}} |g^k_0| \leqslant \vartheta, \quad |G^k - F^k|_{D^{1,k},r_k/2,|k|_1 s_\star} \leqslant \vartheta. \tag{153}$$

The function $\theta \in \mathbb{T}^1_{|k|_1 s_\star} \to G^k(y,\theta)$ is the 'effective potential' at the simple resonance k.

(b) We have assumed that $\|f\|_{r,s} = 1$ (see (137)), since this is the natural assumption in term of genericity properties, however the normal form lemma is formulated in term of the stronger norm $|\cdot|$. We need therefore to restrict slightly the angle-analyticity domain in order to pass to the norm $|\cdot|$. This can be done through (52), which yields (for $r = r_0$ or $r = r_k$ and $K = K_1$ or K_2)

$$|f|_{r,s(1-1/K)} \overset{(52),(137)}{\leqslant} \left(\frac{2nK}{s}\right)^n . \tag{154}$$

(c) The choice of α in (139) is not restrictive (since it is done through the introduction of ν, a new parameter) and it has the effect of making disappear ε from the smallness conditions and from the definition of the smallness parameters $\bar{\vartheta}$ and ϑ.
According to the choice of K_1 and K_2 one will get different kind of statements.

Remark 6.2.

(a) Observe that $r_0 \leqslant r_k \leqslant \sqrt{\varepsilon}K_2^\nu/L$ so that assumption (40) ensures the necessary condition:

$$r_0 \leqslant r_k \leqslant \frac{\sqrt{\varepsilon}K_2^\nu}{L} \leqslant r . \tag{155}$$

(b) The hypotheses of the normal form lemma (proposition 4.1) concern a complex domain D_r, while the non-resonance properties of the covering lemma (proposition 2.1) hold on real domains. The following simple observation allows to use directly the covering lemma:
If a set $D \subseteq \mathbb{R}^n$ is (α, K) non-resonant modulo Λ for h, then the complex domain D_r is $(\alpha - LrK, K)$ non-resonant modulo Λ, provided $LrK < \alpha$, where L is the Lipschitz constant of ω on the complex domain D_r.
Indeed, if $y \in D_r$ there exists $y_0 \in D$ such that $|y - y_0| < r$ and $|\omega(y_0)\cdot k| \geqslant \alpha$ for all $k \in \mathbb{Z}^n \setminus \Lambda$, $|k|_1 \leqslant K$. Thus, for such k's, one has

$$|\omega(y)\cdot k| = |\omega(y_0)\cdot k - (\omega(y_0) - \omega(y))\cdot k| \geqslant |\omega(y_0)\cdot k| - LrK \geqslant \alpha - LrK .$$

□

Proof of theorem 6.1.

(a) By remark 6.2(b), (19) and the choice of r_0 in (140), the domain $D^0_{r_0}$ is $(\alpha/4, K_1)$ completely non-resonant (or non-resonant modulo the trivial lattice $\{0\}$) and, in view of (154) and (143), one can apply proposition 4.1 to H_ε in (138) with $f, D, r, \Lambda, \alpha, K$ and s replaced, respectively, by εf, D^0, r_0, $\{0\}$, $\alpha/4$, K_1 and $s(1-1/K_1)$. Thus, using (140) and that $K_1 \geqslant 2$, one sees that

$$\begin{aligned}
\vartheta_0 &:= 2^{15}\,\frac{LK_1^3|f|_{r_0,s(1-1/K_1)}}{K_2^{2\nu}s(1-1/K_1)} \\
&\overset{(154),(139)}{<} 2^{16}\,\frac{LK_1^3}{sK_2^{2\nu}}\left(\frac{2nK_1}{s}\right)^n \\
&\overset{(139)}{\leqslant} 2^{13}n^n\,\frac{L}{s^{n+1}}\,\frac{1}{K_2^{2\nu-n-3}} \overset{(143)}{\leqslant} e^{-s/2} \leqslant 1,
\end{aligned} \tag{156}$$

showing that (69) holds (with $\vartheta_\star$ replaced by ϑ_0), together with (76) with $\bar{s}$ replaced by $s(1-1/K_1)/2$. Then, by (75) and (154), one has:

$$\sup_{D^0_{r_0/2}} |g^{\rm o} - \langle f\rangle| \leqslant \vartheta_0 \left(\frac{2nK_1}{s}\right)^n \overset{(139)}{\leqslant} \left(\frac{nK_2}{s}\right)^n \vartheta_0 \overset{(156),(141)}{\leqslant} \bar{\vartheta},$$

$$|f^{\rm o}_{\star\star}|_{D^0,r_0/2,s(1-1/K_1)/2} \leqslant 2\,{\rm e}^{-(K_1-2)s(1-1/K_1)/2}\left(\frac{2nK_1}{s}\right)^n \leqslant 2\left(\frac{2nK_1}{s}\right)^n {\rm e}^{-(K_1-3)s/2},$$

from which (146) follows.

(b) By remark 6.2(b), the definition of r_k in (140) and (20), the domain $D^{1,k}_{r_k}$ is

$$(2\alpha K_2/|k| - r_k L K_2, K_2) = (\alpha K_2/|k|, K_2)$$

non-resonant modulo $\mathbb{Z}k$.
Using again (154), we can apply proposition 4.1 with f, D, r, Λ, α, K and s replaced, respectively, by εf, $D^{1,k}$, r_k, $\mathbb{Z}k$, $\alpha K_2/|k|$, K_2 and $s(1-1/K_2)$. Then, since $|k| \leqslant K_1$, one finds

$$\begin{aligned}\vartheta_k &:= 2^{11}\frac{LK_2|k|^2\varepsilon|f|_{r_k,s(1-1/K_2)}}{\alpha^2 s(1-1/K_2)}\\ &\overset{(139),(154)}{\leqslant} 2^{n+12}n^n\frac{L}{s^{n+1}}\frac{1}{K_2^{2\nu-n-3}} \overset{(143)}{\leqslant} {\rm e}^{-s/2} \leqslant 1,\end{aligned} \tag{157}$$

showing that (69) holds with $\vartheta_\star$ replaced by ϑ_k, together with (76) with $\bar{s}$ replaced by $s(1-1/K_2)/2$. From (75), (157) and (154) there follows (150); indeed

$$|g^k - {\rm p}_{\mathbb{Z}k} f|_{D^{1,k},r_k/2,s_\star} \leqslant \frac{1}{K_2}\left(\frac{2nK_2}{s}\right)^n \vartheta_k \overset{(157),(141)}{\leqslant} \vartheta,$$

$$|f^k_{\star\star}|_{D^{1,k},r_k/2,s(1-1/K_2)/2} \leqslant 2\,{\rm e}^{-(K_2-2)s(1-1/K_2)/2}\left(\frac{2nK_2}{s}\right)^n \leqslant 2\left(\frac{2nK_2}{s}\right)^n {\rm e}^{-(K_2-3)s/2}. \qquad \square$$

7. Proof of theorem 2.1

Under the above hypotheses, apart from a finite number of simple resonances the effective potential G^k at simple resonances is close to a (shifted) cosine:

Proposition 7.1. *Let the assumptions of theorem 6.1 hold, let $k \in \mathcal{G}^n_{1,K_1}$ and let G^k be as in (151), (149). Then, if*

$$|k|_1 > 3/s\,, \tag{158}$$

one has that

$$|G^k - T_1 F^k|_{D^{1,k},r_k/2,2} \leqslant \vartheta\,{\rm e}^{s+5}\,{\rm e}^{-|k|_1 s} + 2^8\,{\rm e}^{-2|k|_1 s}\,. \tag{159}$$

Proof. Recall (65) and observe that by definition of T_N and $T_N^\perp$,

$$G^k - T_1F^k = T_1G^k - T_1F^k + T_1^\perp G^k\,. \tag{160}$$

Now, since $3/s < |k|_1 \leqslant K_1 \leqslant K_2/3$,

$$\sup_{D^{1,k}_{r_k/2}} |G^k_{\pm 1} - F^k_{\pm 1}| \overset{(153)}{\leqslant} \vartheta \mathrm{e}^{-|k|_1 s_\star} \overset{(147)}{\leqslant} \vartheta \mathrm{e}^{-|k|_1 s(1-2/K_2)} \leqslant \vartheta \mathrm{e}^{s} \mathrm{e}^{-|k|_1 s}$$

so that

$$|T_1 G^k - T_1 F^k|_{D^{1,k}, r_k/2, 2} = |G^k_1 - F^k_1| e^2 + |G^k_{-1} - F^k_{-1}|\, \mathrm{e}^2 < 2\, \mathrm{e}^2\, \mathrm{e}^s \vartheta\, \mathrm{e}^{-|k|_1 s}\,. \tag{161}$$

Next, recalling (151), we have that

$$|T_1^\perp G^k|_{D^{1,k}, r_k/2, 2} = \sum_{\substack{|j|\geqslant 2\\ j\in\mathbb{Z}}} |g_{jk}|\, \mathrm{e}^{2|j|} \leqslant \sum_{|j|\geqslant 2} |f_{jk}|\, \mathrm{e}^{2|j|} + \sum_{|j|\geqslant 2} |g_{jk} - f_{jk}|\, \mathrm{e}^{2|j|}\,. \tag{162}$$

Let us estimate the two sums separately. Since $\|f\|_s = 1$, $|f_\ell| \leqslant \mathrm{e}^{-|\ell|_1 s}$ so that $|f_{jk}| \leqslant \mathrm{e}^{-|j||k|_1 s}$ and:

$$\sum_{|j|\geqslant 2} |f_{jk}| \mathrm{e}^{2|j|} \leqslant \sum_{|j|\geqslant 2} \mathrm{e}^{-|j||k|_1 s}\, \mathrm{e}^{2|j|} = 2 \frac{\mathrm{e}^{-2(|k|_1 s - 2)}}{1 - \mathrm{e}^{-(|k|_1 s - 2)}} \leqslant 4 \mathrm{e}^4\, \mathrm{e}^{-2(|k|_1 s)}\,, \tag{163}$$

where in the last inequality we used the assumption $|k|_1 s > 3 > 2 + \log 2$.

Then (again, because $|k|_1 s > 3$), we see that

$$\begin{aligned}\sum_{|j|\geqslant 2} |g_{jk} - f_{jk}| \mathrm{e}^{2|j|} &= \sum_{|j|\geqslant 2} |g_{jk} - f_{jk}| \mathrm{e}^{|j||k|_1 s_\star}\, \mathrm{e}^{-|j||k|_1 s_\star + 2|j|} \\ &\overset{(153)}{\leqslant} \sup_{j\geqslant 2} \left(\mathrm{e}^{-j(|k|_1 s_\star - 2)}\right) \vartheta \leqslant \mathrm{e}^{-2(|k|_1 s_\star - 2)}\, \vartheta \\ &\leqslant \mathrm{e}^4 \mathrm{e}^{-2|k|_1 s(1-2/K_2)} \vartheta \leqslant \vartheta\, \mathrm{e}^{2s+4} \mathrm{e}^{-2|k|_1 s}.\end{aligned} \tag{164}$$

Putting (163) and (164) together, by (160) and (158), (159) follows. □

Proposition 7.2. *Let the assumptions of theorem 6.1 hold; let $s > 0$, $0 < \delta \leqslant 1$ and fix any $0 < \gamma \leqslant 1$. Assume (38) and (34). If $k \in \mathcal{G}^n_{1,K_1}$ satisfies*

$$|k|_1 > \tau_o(\delta; \gamma)\,, \tag{165}$$

(with $\tau_o(\delta;\gamma)$ defined in (37)) then,

$$|G^k - T_1 F^k|_{D^{1,k}, r_k/2, 2} \leqslant \gamma \delta_k\,, \tag{166}$$

where

$$\delta_k := \delta |k|_1^{-n}\, \mathrm{e}^{-|k|_1 s}\,. \tag{167}$$

Remark 7.1. Conditions (38) and (165) are stronger than the ones on ν in (139), (143) and (158). In particular the assumptions of proposition 7.1 hold.

Proof of proposition 7.2. As mentioned in the above remark, proposition 7.1 holds. Let us estimate the two terms in (159) separately. Recalling the definition of ϑ in (141) (and that $|k|_1 \leqslant K_1 \leqslant K_2/3$), we find:

$$\begin{aligned}\vartheta\, \mathrm{e}^{s+5}\, \mathrm{e}^{-|k|_1 s} &\overset{(141)}{=} \mathrm{e}^{s+5}\, \mathrm{e}^{-|k|_1 s} 2^{2n+10} n^{2n} \frac{L}{s^{2n+1}} \frac{1}{K_2^{2\nu-2n-3}} \\ &= \mathrm{e}^{s+5} 2^{2n+11} n^{2n} \frac{L}{s^{2n+1}} \frac{|k|_1^n}{K_2^{2\nu-2n-3}} \frac{1}{\gamma\delta} \frac{\gamma\delta_k}{2} \\ &\leqslant \mathrm{e}^{s+5} 2^{n+11} n^{2n} \frac{L}{s^{2n+1}} \frac{1}{K_2^{2\nu-3n-3}} \frac{1}{\gamma\delta} \frac{\gamma\delta_k}{2} \\ &\overset{(38)}{\leqslant} \frac{\gamma\delta_k}{2}.\end{aligned} \tag{168}$$

To control second term in (159), one may use the following elementary calculus estimate (whose simple proof is left to the reader):

$$a > 2\log 2, \quad 0 < \lambda < \mathrm{e}^{-a^2/2}, \quad t > 4\log \lambda^{-1} \Rightarrow \mathrm{e}^{-t} t^a < \lambda\,. \tag{169}$$

Then, by (169) (with a, t and λ replaced, respectively, by n, $s|k|_1$ and $s^n\gamma\delta/2^9$), and by (34) and (165), one sees that

$$2^8\, \mathrm{e}^{-2|k|_1 s} = (s|k|_1)^n\, \mathrm{e}^{-|k|_1 s} \frac{2^9}{s^n\gamma\delta} \cdot \frac{\gamma\delta_k}{2} < \frac{\gamma\delta_k}{2}. \tag{170}$$

The claim, now, follows from (168) and (170). □

The quantity δ_k defined in (167) is a 'Fourier-measure' for the non-degeneracy of analytic potentials f holomorphic on $\mathbb{T}^n_s$: indeed, such potentials will have, in general, Fourier coefficients $f_k \sim \mathrm{e}^{-s|k|_1}$.

We still need a lemma.

Lemma 7.1. *Let $s > 0$, $0 < \delta \leqslant 1$ and fix any $0 < \gamma \leqslant 1$. Let the assumptions of theorem 6.1 hold; assume (38), (34) and that the positional potential $f \in \mathcal{H}_{s,\tau_o}(\delta)$ with the tail function τ_o defined in (165). Then, for $\tau_o(\delta;\gamma) \leqslant |k|_1 \leqslant K_1 \leqslant K_2/3$, one has*

$$\sup_{y \in D^{1,k}_{r_k/2}} \frac{|G^k(y,\cdot) - T_1 F^k(\cdot)|_2}{|f_k|} \leqslant \gamma \tag{171}$$

and

$$\frac{1}{|f_k|} |f^k_{\star\star}|_{D^{1,k}, r_k/2, s(1-1/K_2)/2} \leqslant \frac{2^{10n} n^{3n}}{s^{3n}\delta} \mathrm{e}^{-K_2 s/8}. \tag{172}$$

Proof. Since proposition 7.2 holds, by (166) and since $f \in \mathcal{H}_{s,\tau_o}(\delta)$ we get (171). By (150) we get

$$\begin{aligned}\frac{1}{|f_k|} |f^k_{\star\star}|_{D^{1,k}, r_k/2, s(1-1/K_2)/2} &< \frac{|k|_1^n\, \mathrm{e}^{|k|_1 s}}{\delta} 2 \left(\frac{2nK_2}{s}\right)^n \mathrm{e}^{-(K_2-3)s/2} \\ &\leqslant \frac{2^{n+1} n^n}{s^n\delta} K_1^n K_2^n\, \mathrm{e}^{-(K_2-2K_1)s/2} \overset{(139)}{\leqslant} \frac{n^n}{s^n\delta} K_2^{2n}\, \mathrm{e}^{-K_2 s/6}.\end{aligned}$$

Then, observing that for $\alpha > 0$ we have $\max_{x>0} x^\alpha e^{-x} = (\alpha/e)^\alpha$ and

$$K_2^{2n}\, e^{-K_2 s/24} \leqslant \left(\frac{48n}{se}\right)^{2n} \leqslant \frac{2^{10n} n^{2n}}{s^n},$$

(172) follows. □

Finally, we may conclude the

Proof of theorem 2.1. Recalling the definition of T_N given in (65), we have that

$$T_1 F^k(\theta) = f_k e^{i\theta} + f_{-k} e^{-i\theta} = 2|f_k| \cos(\theta + \theta^{(k)})$$

for a suitable constant $\theta^{(k)}$. Recalling (148), (152) and setting

$$h^k(y) := h(y) + \varepsilon g_0^k(y),$$
$$\mathtt{G}^k(y,\theta) := \frac{G^k(y,\theta) - \cos(\theta + \theta^{(k)})}{2|f_k|},$$
$$\mathtt{f}^k(y,x) := \frac{f^k_{\star\star}(y,\theta)}{2|f_k|},$$

we get (42). Finally theorem 2.1 follows from lemma 7.1, in particular (43) and (44) follow from (153), (142), (171) and (172), respectively. □

8. General (y-dependent) potentials

In this section we discuss briefly the generalization of the above analysis to the case of potentials which depend also on the action variables y.

For $k \in \mathbb{Z}^n \backslash \{0\}$ let $b_k > 0$ such that

$$\sum_{k\neq 0} b_k < \infty.$$

For $\mathcal{Z} \subseteq \mathbb{Z}^n \backslash \{0\}$ we set

$$b_{\mathcal{Z}} := \sum_{k\in\mathcal{Z}} b_k.$$

For definiteness we will fix

$$b_k := |k|^{-\frac{n}{2}},$$

but every other possible choice is fine. As usual, we denote by S^{n-1} the real unit ball in $\mathbb{R}^n$: $S^{n-1} = \{y \in \mathbb{R}^n |\, |y| = 1\}$.

Proposition 8.1. *Let $r, \mu > 0$ and $\mathcal{Z} \subseteq \mathbb{Z}^n \backslash \{0\}$. For any $k \in \mathcal{Z}$ let $\varphi_k(y)$ be holomorphic functions on the complex ball $\{y \in \mathbb{C}^n : |y| < r\}$ with*

$$\sup_{|y|<r} |\varphi_k(y)| \leqslant 1, \quad \textit{and} \quad |\varphi_k(0)| \geqslant \hat{\delta}_k > 0.$$

Then, for every $y \in \mathbb{R}^n$ with $|y| < r/2e$, up, at most, to a set of measure

$$\frac{1}{2}\, \mathrm{meas}_{\,n-1}(S^{n-1}) b_{\mathcal{Z}} \left(\frac{r}{2e}\right)^n \mu,$$

we have

$$|\varphi_k(y)| \geqslant \hat{\delta}_k \left(\frac{\mu b_k}{30\,\mathrm{e}^3} \right)^{\log\, 1/\hat{\delta}_k}, \quad \forall k \in \mathcal{Z}. \tag{173}$$

The proof relies on the following classical result in function theory (see, e.g., [32]):

Lemma 8.1 (Cartan's estimate). *Assume that* $\mathtt{f} : \mathbb{C} \to \mathbb{C}$ *is holomorphic and bounded by* $M > 0$ *on the complex ball* $|z| < 2eR$. *If* $|\mathtt{f}(0)| = 1$ *then, for* $0 < \eta < 1$

$$|\mathtt{f}(z)| \geqslant \left(\frac{\eta}{15\mathrm{e}^3} \right)^{\log M} \tag{174}$$

for any $z \in \mathbb{C}$, $|z| < R$ *up to a set of balls of radii* r_j *satisfying*

$$\sum_j r_j \leqslant \eta R.$$

Remark 8.1. Note that (174) holds in the complex ball $|z| < R$ up to a set of measure smaller than $\pi\eta^2 R^2$. Moreover it holds on the real interval $(-R, R)$ up to a set of (real) measure $2\eta R$.

Proof of proposition 8.1. Fix $k \in \mathcal{Z}$. Fix $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$ with $|\xi| = 1$ and $\xi_1 \geqslant 0$. We apply Cartan's estimates simultaneously for every $k \in \mathcal{Z}$ with $\mathtt{f}(z)$, R, M and η replaced, respectively, by

$$\frac{\varphi_k(z\xi)}{\varphi_k(0)}, \quad \frac{r}{2e}, \quad \frac{1}{\hat{\delta}_k} \quad \text{and} \quad \frac{\mu}{2} b_k.$$

Recall remark 8.1 and observe that, by (174), for fixed k, estimate (173) holds on the segment $\{y\xi : y \in (-r/2e, r/2e)\}$, up to a set of measure at most $\mu b_k r/(2e)$. Integrating on the half-sphere $|\xi| = 1$, $\xi_1 \geqslant 0$, we get that (173) for the fixed k holds on the ball $|y| < r/2e$ up, at most, to a set of measure

$$\frac{1}{2}\,\mathrm{meas}_{\,n-1}(S^{n-1}) \left(\frac{r}{2e} \right)^n \mu b_k.$$

Summing on all $k \in \mathcal{Z}$ we get that (173) holds for all $k \in \mathcal{Z}$. □

Fix $0 < \mu, \gamma < 1$. Define the following tail function

$$\tau_*(\delta; \gamma, \mu) := \tag{175}$$
$$\frac{2^6 n^2}{\tilde{s}} \max \left\{ \log^3 \frac{2^6 n^2}{\tilde{s}}, \left(\log \frac{30\mathrm{e}^3}{\delta\mu} \right) \log^2 \left(\frac{4}{\tilde{s}} \log \frac{30\mathrm{e}^3}{\delta\mu} \right), \log \frac{30\mathrm{e}^3}{\mu} \log \frac{1}{\delta}, \log \frac{2^{10}}{\delta\gamma} \right\},$$

where

$$\tilde{s} := \min\{s, 1\}.$$

Fix $y_0 \in D$ and assume that

$$f(y_0, \cdot) \in \mathcal{H}_{s,\tau_*}(\delta).$$

Set

$$\varphi_k(y) := f_k(y)\mathrm{e}^{|k|_1 s}. \tag{176}$$

We have that

$$\sup_{y\in\mathbb{C}^n,|y-y_0|<r} |\varphi_k(y)| \overset{(137)}{\leqslant} 1, \quad |\varphi_k(y_0)| \geqslant \hat{\delta}_k := \delta/|k|_1^n, \quad \forall k \in \mathcal{G}_1^n,\ |k|_1 > \tau_*(\delta). \tag{177}$$

Let $\mu > 0$. Then by proposition 8.1 (with $\varphi_k(y) \rightsquigarrow \varphi_k(y+y_0)$) there exists real sets

$$\mathcal{D} \subseteq B_{r/2e}(y_0) \quad \text{satisfying meas}\,(B_{r/2e}(y_0)\backslash\mathcal{D}) \leqslant \frac{b}{2}\,\text{meas}_{\,n-1}(S^{n-1})\Big(\frac{r}{2e}\Big)^n \mu, \tag{178}$$

with

$$b := \sum_{|k|_1>\tau_*(\delta;\gamma,\mu)} |k|_1^{-n/2} \leqslant \sum_{k\neq 0} |k|_1^{-n/2},$$

such that

$$|f_k(y)|\mathrm{e}^{|k|_1 s} = |\varphi_k(y)| \geqslant \delta_k(\mu) := \hat{\delta}_k \left(\frac{\mu}{30\,\mathrm{e}^3|k|_1^{n/2}}\right)^{\log\ 1/\hat{\delta}_k},$$

$$\forall y \in \mathcal{D},\ k \in \mathcal{G}_1^n, \quad |k|_1 > \tau_*(\delta). \tag{179}$$

□

Theorem 8.1. *Let the assumption of theorem 6.1 hold. Fix* $0 < \mu, \delta < 1/e^8$ *and* $0 < \gamma < 1$. *Assume that for some* $y_0 \in D$ *we have* $f(y_0,\cdot) \in \mathcal{H}_{s,\tau_*}(\delta)$. *Set*

$$\tilde{\mu} := \mu/30\,\mathrm{e}^3, \quad \tilde{n} := 2\nu - 2n - 3, \quad \kappa := 2^{2n+10}n^{2n}\frac{L}{s^{2n+1}}. \tag{180}$$

Assume that

$$K_2 \geqslant \max\left\{K_1^{\frac{2n^2}{\tilde{n}}\log K_1},\ K_1^{\frac{9}{\tilde{n}}\log\frac{1}{\delta\tilde{\mu}}},\ \mathrm{e}^{\frac{4}{\tilde{n}}\log\frac{1}{\delta}\log\frac{1}{\tilde{\mu}}},\ \left(\frac{4\mathrm{e}^{s+5}\kappa}{\delta\gamma}\right)^{\frac{4}{\tilde{n}}},\ \frac{2^5}{s}\log^2\frac{1}{\delta\mu},\ \frac{2^{14}n^4}{s^2}\right\}. \tag{181}$$

If $k \in \mathcal{G}_{1,K_1}^n$ *with* $|k|_1 > \tau_*(\delta;\gamma,\mu)$ *then*

$$\sup_{D^{1,k}_{r_k/2}} |h^k - h| \leqslant \gamma\delta\varepsilon, \tag{182}$$

$$\sup_{y\in(D^{1,k}\cap\mathcal{D})_{\hat{r}_k}} \frac{|G^k(y,\cdot) - T_1F^k(y,\cdot)|_2}{|f_k(y)|} \leqslant \gamma, \tag{183}$$

$$\sup_{y\in(D^{1,k}\cap\mathcal{D})_{\hat{r}_k}} \frac{|f^k_{\star\star}(y,\cdot)|_{s(1-1/K_2)/2}}{|f_k(y)|} < \frac{4\,\mathrm{e}^{3s/2}n^n}{\delta s^n}\,\mathrm{e}^{-K_2 s/8}, \tag{184}$$

where $\mathcal{D}$ was defined in (178) *and*

$$\hat{r}_k := \frac{1}{2}\min\{r_k, \delta_k(\mu)\}. \tag{185}$$

Proof. First, note that by (155) $\hat{r}_k \leqslant r/2$. Observe, also, that, by (179), (177), (185) and Cauchy estimates

$$|\varphi_k(y)| \geqslant \frac{1}{2}\delta_k(\mu), \quad \forall y \in \mathcal{D}_{\hat{r}_k},\ k \in \mathcal{G}^n_1,\ |k|_1 > \tau_*(\delta). \tag{186}$$

By (159), (176) and (186) we have that for every $y \in \mathcal{D}$, $k \in \mathcal{G}^n_{1,K_1}$

$$\frac{|G^k(y,\cdot) - T_1 F^k(y,\cdot)|_2}{|f_k(y)|} \leqslant \frac{2\,\mathrm{e}^{s+5}\,\vartheta\, + 2^9\, e^{-|k|_1 s}}{\delta_k(\mu)}\,. \tag{187}$$

Then, in order to prove (182), it is enough to show that

$$\frac{4\,\mathrm{e}^{s+5}\vartheta}{\delta_k(\mu)} \leqslant \gamma, \quad \frac{2^{10}\,\mathrm{e}^{-|k|_1 s}}{\delta_k(\mu)} \leqslant \gamma. \tag{188}$$

Let us consider the first inequality in (188). Since by (180) and recalling (141) we have $\vartheta = \kappa/K_2^{\tilde{n}}$, then, recalling (177) and (179), for $|k|_1 \leqslant K_1$

$$\frac{4\,\mathrm{e}^{s+5}\vartheta}{\delta_k(\mu)} \leqslant \frac{4\,\mathrm{e}^{s+5}\kappa K_1^n}{\delta K_2^{\tilde{n}}}\left(\frac{K_1^{n/2}}{\tilde{\mu}}\right)^{\log\, K_1^n \delta^{-1}} = \frac{4\,\mathrm{e}^{s+5}\kappa}{\delta}\,\mathrm{e}^A,$$

where

$$A := \left(\frac{n}{2}\,\log\,K_1 + \log\,\frac{1}{\tilde{\mu}}\right)\left(n\,\log\,K_1 + \log\,\frac{1}{\delta}\right) + n\,\log\,K_1 - \tilde{n}\,\log\,K_2.$$

Since

$$A \leqslant -\frac{\tilde{n}}{4}\,\log\,K_2$$

by (181), we obtain that

$$\frac{4\,\mathrm{e}^{s+5}\vartheta}{\delta_k(\mu)} \leqslant \frac{4\,\mathrm{e}^{s+5}\kappa}{\delta}\,\mathrm{e}^{-\frac{\tilde{n}}{4}\log K_2} \overset{(181)}{\leqslant} \gamma,$$

proving the first estimate in (188).
Regarding the second inequality in (188) we have

$$\frac{2^{10}\,\mathrm{e}^{-|k|_1 s}}{\delta_k(\mu)} = \frac{2^{10}}{\delta}\,\mathrm{e}^B,$$

with

$$B := \left(\frac{n}{2}\,\log|k|_1 + \log\,\frac{1}{\tilde{\mu}}\right)\left(n\,\log|k|_1 + \log\,\frac{1}{\delta}\right) + n\,\log|k|_1 - |k|_1 s.$$

We note that

$$B \leqslant -\frac{1}{4}|k|_1 s$$

by (175), indeed

$$\frac{2}{s}n^2 \leqslant \frac{|k|_1}{\log^2|k|_1}, \quad \frac{4n}{s}\left(\frac{1}{2}\log\frac{1}{\delta}+\log\frac{1}{\tilde{\mu}}+1\right) \leqslant \frac{|k|_1}{\log|k|_1},$$

where we have used the elementary estimates: $x/\log x \geqslant \alpha$ if $x \geqslant \alpha\log^2\alpha$ and $\alpha \geqslant 5$, and $x/\log^2 x \geqslant \alpha$ if $x \geqslant \alpha\log^3\alpha$ and $\alpha \geqslant 2^6$. Then

$$\frac{2^{10}}{\delta}\,\mathrm{e}^{-\frac{1}{4}|k|_1 s} \overset{(175)}{\leqslant} \gamma.$$

This proves (188) and, therefore, completes the proof of (182).
Let us now show (184). By (150) and (186) we get

$$\begin{aligned}
\sup_{y\in(D^{1,k}\cap\mathcal{D})_{\hat{r}_k}} \frac{|f^k_{\star\star}(y,\cdot)|_{s(1-1/K_2)/2}}{|f_k(y)|} &< \frac{2\,\mathrm{e}^{|k|_1 s}}{\delta_k(\mu)}2\left(\frac{2nK_2}{s}\right)^n \mathrm{e}^{-(K_2-3)s/2}\\
&\overset{(179)}{=} \frac{2\,\mathrm{e}^{|k|_1 s}|k|_1^n}{\delta}\left(\frac{30\,\mathrm{e}^3|k|_1^{n/2}}{\mu}\right)^{\log(|k|_1^n/\delta)} 2\left(\frac{2nK_2}{s}\right)^n \mathrm{e}^{-(K_2-3)s/2}\\
&\overset{(139)}{\leqslant} \frac{4\,\mathrm{e}^{3s/2}n^n}{\delta s^n}\left(\frac{30\,\mathrm{e}^3K_2^{n/2}}{3^{n/2}\mu}\right)^{\log(K_2^n/3^n\delta)} K_2^{2n}\,\mathrm{e}^{-K_2 s/6}\\
&= \frac{4\,\mathrm{e}^{3s/2}n^n}{\delta s^n}\,\mathrm{e}^{-K_2 s/8}\,\mathrm{e}^{-Q}
\end{aligned}$$

where

$$Q := \frac{1}{8}K_2 s - \left(n\log\frac{K_2}{3}+\log\frac{1}{\delta}\right)\left(\frac{n}{2}\log\frac{K_2}{3}+\log\frac{1}{\mu}+\log 30+3\right) - 2n\log K_2.$$

Then (184) follows if we prove that $Q \geqslant 0$. Recalling (139) and observing that $\log^2 x \leqslant \sqrt{x}$ for $x \geqslant 2^{13}$, we get

$$Q \geqslant \frac{1}{8}K_2 s - 8n^2\log^2 K_2 + 2\log^2\frac{1}{\delta\mu} \geqslant 0$$

by (181). □

We rewrite theorem 8.1 in the fashion of theorem 2.1.

Theorem 8.2. *Let $n \geqslant 2$, $0 < s \leqslant 1$. Fix $0 < \mu, \delta < 1/e^8$ and $0 < \gamma < 1$. Consider a Hamiltonian $H_\varepsilon(y,x) = h(y) + \varepsilon f(y,x)$ as in (1) with h KAM non-degenerate (definition 2.1) and f with norm one: $\|f\|_{D,r,s} = 1$. Assume that for some $y_0 \in D$ we have $f(y_0,\cdot) \in \mathcal{H}_{s,\tau_*}(\delta)$, with $\tau_* = \tau_*(\delta;\gamma,\mu)$ defined in (175). Let $K_2 \geqslant 3K_1 \geqslant 6$ with K_1 satisfying (181) and (143). Let $\hat{r}_k$ as in (185) and $\mathcal{D}$ as in (178). Finally assume that ε satisfies (40).*

Then, for any $k \in \mathcal{G}^n_{1,K_1}$ with $\tau_*(\delta;\gamma,\mu) \leqslant |k|_1 \leqslant K_1$, there exists a symplectic change of variables Ψ_k as in (41) such that the following holds.

For every $y \in (D^{1,k} \cap \mathcal{D})_{\hat{r}_k}$ there exist a phase $\theta^{(k)}(y)$ and functions $\mathtt{G}^k(y,\cdot) \in \mathbb{B}^1_2$ and $\mathtt{f}^k(y,\cdot) \in \mathbb{B}^n_{s(1-1/K_2)^2}$ satisfying

$$\boxed{H_\varepsilon \circ \Psi_k = h(y) + 2\varepsilon\,|f_k(y)|\Big(\cos(k\cdot x + \theta^{(k)}(y)) + \mathtt{G}^k(y, k\cdot x) + \mathtt{f}^k(y,x)\Big)} \tag{189}$$

with

$$\sup_{y \in (D^{1,k}\cap\mathcal{D})_{\hat{r}_k}} |\mathtt{G}^k(y,\cdot)|_2 \leqslant \gamma \tag{190}$$

and

$$\sup_{y \in (D^{1,k}\cap\mathcal{D})_{\hat{r}_k}} |\mathtt{f}^k|_{s(1-1/K_2)/2} \leqslant \frac{4\,\mathrm{e}^{3s/2} n^n}{\delta s^n}\,\mathrm{e}^{-K_2 s/8}. \tag{191}$$

Proof. The claim follows directly from theorem 8.1. We only note that $\theta^{(k)}(y)$ is defined so that

$$|f_k(y)|\cos(k\cdot x + \theta^{(k)}(y)) = T_1 F^k(y,x),$$

while

$$\mathtt{G}^k(y,x) := \frac{G^k(y,\cdot) - T_1 F^k(y,\cdot)}{|f_k(y)|},$$

$$\mathtt{f}^k(y,x) := \frac{f^k_{\star\star}(y,x)}{|f_k(y)|}.$$

Note that $\theta^{(k)}(y)$, $\mathtt{G}^k(y,x)$ and $\mathtt{f}^k(y,x)$ are not analytic in y (due to the presence of $|f_k(y)|$), but, obviously, $H_\varepsilon \circ \Psi_k$ is real-analytic in x and y. □

Acknowledgments

We are grateful to V Kaloshin and A Sorrentino for useful discussions.

Appendix A. An elementary result in linear algebra

First, let us recall the classical Bezout's lemma (whose proof can be found in any elementary book on number theory):

Bezout's lemma. *Given two integers a and b not both zero, there exist two integers x and y such that* $ax + by = d := \gcd(a, b)$*, and such that* $\max\{|x|, |y|\} \leqslant \max\{|a|/d, |b|/d\}$.

Then, the following simple result holds.

Lemma A.1. *Given* $k \in \mathbb{Z}^n$, $k \neq 0$ *there exists a matrix* $A = (A_{ij})_{1\leqslant i,j\leqslant n}$ *with integer entries such that* $A_{nj} = k_j\ \forall 1 \leqslant j \leqslant n$, $\det A = d := \gcd(k_1, \ldots, k_n)$*, and* $|A|_\infty = |k|_\infty$.

Notice that the estimates on x and y are easily deduced from the well known fact that given a solution x_0 and y_0 of the equation $ax + by = d$, all other solutions have the form $x = x_0 + k(b/d)$ and $y = y_0 - k(a/d)$ with $k \in \mathbb{Z}$ and by choosing k so as to minimize $|x|$.

Proof. The argument is by induction over n. For $n = 1$ the lemma is obviously true. For $n = 2$, it follows at once from Bezout's lemma: indeed, if x and y are as in Bezout's lemma with $a = k_1$ and $b = k_2$ one can take $A = \begin{pmatrix} y & -x \\ k_1 & k_2 \end{pmatrix}$. Now, assume, by induction for $n \geqslant 3$ that the claim holds true for $(n-1)$ and let us prove it for n. Let $\bar{k} = (k_1, \ldots, k_{n-1})$ and $\bar{d} = \gcd(k_1, \ldots, k_{n-1})$ and notice that $\gcd(\bar{d}, k_n) = d$. By the inductive assumption, there exists a matrix $\bar{A} = \begin{pmatrix} \tilde{A} \\ \bar{k} \end{pmatrix} \in \mathrm{Mat}_{(n-1)\times(n-1)}(\mathbb{Z})$ with $\tilde{A} \in \mathrm{Mat}_{(n-2)\times(n-1)}(\mathbb{Z})$, such that $\det \bar{A} = \bar{d}$ and $|\bar{A}|_\infty = |\bar{k}|_\infty$. Now, let x and y be as in Bezout's lemma with $a = \bar{d}$, and $b = k_n$. We claim that A can be defined as follows:

$$A = \begin{pmatrix} \tilde{k} & \tilde{x} \\ \bar{A} & \begin{pmatrix} 0 \\ \vdots \\ 0 \\ k_n \end{pmatrix} \end{pmatrix}, \quad \tilde{k} = (-1)^n y \frac{\bar{k}}{\bar{d}}, \quad \tilde{x} := (-1)^{n+1} x. \tag{192}$$

First, observe that since $\bar{d}$ divides k_j for $j \leqslant (n-1)$, $\tilde{k} \in \mathbb{Z}^{n-1}$. Then, expanding the determinant of A from last column, we get

$$\begin{aligned} \det A &= (-1)^{n+1}\tilde{x}\ \det \bar{A} + k_n\ \det \begin{pmatrix} \tilde{k} \\ \tilde{A} \end{pmatrix} \\ &= (-1)^{n+1}\tilde{x}\bar{d} + k_n(-1)^{n-2}\ \det \begin{pmatrix} \tilde{A} \\ \tilde{k} \end{pmatrix} \\ &= (-1)^{n+1}\tilde{x}\bar{d} + k_n(-1)^{n-2}(-1)^n \frac{y}{\bar{d}}\ \det \bar{A} \\ &= x\bar{d} + k_n y = d. \end{aligned}$$

Finally, by Bezout's lemma, we have that $\max\{|x|, |y|\} \leqslant \max\{\bar{d}/d, |k_n|/d\}$, so that

$$|\tilde{k}|_\infty = |y| \frac{|\bar{k}|_\infty}{\bar{d}} \leqslant \frac{|\bar{k}|_\infty}{d} \leqslant |k|_\infty, \quad |\tilde{x}| = |x| \leqslant \frac{|k_n|}{d} \leqslant K_\infty,$$

which, together with $|\bar{A}|_\infty = |\bar{k}|_\infty$, shows that $|A|_\infty = K_\infty$. □

ORCID iDs

L Chierchia https://orcid.org/0000-0002-4356-2092

References

[1] Arnold V I, Kozlov V V and Neishtadt A I 2006 Mathematical aspects of classical and celestial mechanics *Encyclopaedia of Mathematical Sciences* 3rd edn vol 3 (Berlin: Springer)
[2] Lazutkin V F 1974 Concerning a theorem of Moser on invariant curves (in Russian) *Vopr. Din. Teor. Rasprostr. Seism. Voln* **14** 109–20
[3] Neishtadt A I 1981 Estimates in the Kolmogorov theorem on conservation of condition- ally periodic motions *J. Appl. Math. Mech.* **45** 766–72
[4] Pöschel J 1982 Integrability of Hamiltonian systems on Cantor sets *Commun. Pure Appl. Math.* **35** 653–95
[5] Arnold V I 1964 Instability of dynamical systems with several degrees of freedom *Dokl. Akad. Nauk SSSR* **156** 9–12
Arnold V I 1964 Instability of dynamical systems with several degrees of freedom. *Sov. Math., Dokl.* **5** 581–5 (Engl. transl.)
[6] Biasco L and Chierchia L 2015 On the measure of Lagrangian invariant tori in nearly-integrable mechanical systems *Rendiconti Lincei. Mat. Appl.* **26** 1–10
[7] Medvedev A G, Neishtadt A I and Treschev D V 2015 Lagrangian tori near resonances of near-integrable Hamiltonian systems *Nonlinearity* **28** 2105–30
[8] Biasco L and Chierchia L KAM 2017 Theory for Secondary Tori (arXiv:1702.06480v1)
[9] Biasco L and Chierchia L 2020 On the measure of KAM tori in two degrees of freedom Discrete Continuous Dyn. Syst. - Ser. A. (https://doi.org/10.3934/dcds.2020134)
[10] Guzzo M, Chierchia L and Benettin G 2016 The steep Nekhoroshev theorem *Commun. Math. Phys.* **342** 569–601
[11] Pöschel J 1993 Nekhoroshev estimates for quasi-convex Hamiltonian systems *Math. Z.* **213** 187
[12] Bernard P, Kaloshin V and Zhang K 2016 Arnold diffusion in arbitrary degrees of freedom and normally hyperbolic invariant cylinders *Acta Math.* **217** 1–79
[13] Berti M, Biasco L and Bolle P 2003 Drift in phase space: a new variational mechanism with optimal diffusion time *J. Pure Appl. Math.* **82** 613–64
[14] Bessi U, Chierchia L and Valdinoci E 2001 Upper bounds on Arnold diffusion times via Mather theory *J. Pure Appl. Math.* **80** 105–29
[15] Cheng C Q and Yan J 2004 Existence of diffusion orbits in a priori unstable Hamiltonian systems *J. Differ. Geom.* **67** 457–517
[16] Chierchia L and Gallavotti G 1994 Drift and diffusion in phase space *Ann. Inst. Henri Poincare* A **60** 1–144
Chierchia L and Gallavotti G 1998 Drift and diffusion in phase space *Ann. Inst. Henri Poincare* A **68** 135 (erratum)
[17] Delshams A, de la Llave R and Seara T M 2006 A geometric mechanism for diffusion in Hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model *Memoir. Am. Math. Soc.* **179** 141
[18] Treschev D 2012 Arnold diffusion far from strong resonances in multidimensional a priori unstable Hamiltonian systems *Nonlinearity* **25** 2717–57
[19] Zhang K 2011 Speed of Arnold diffusion for analytic Hamiltonian systems *Invent Math.* **186** 255–90
[20] Hunt B R, Sauer T and Yorke J A 1992 Prevalence: a translation-invariant 'almost every' on infinite-dimensional spaces *Bull. Am. Math. Soc.* **27** 217–38
[21] Hunt B R and Kaloshin V Y 2010 Prevalence *Handbook in Dynamical Systems* ed H Broer, F Takens and B Hasselblatt vol 3 ch 2 (Amsterdam: Elsevier) pp 43–87
[22] Delshams A and Gutiérrez P 1996 Effective stability and KAM theory *J. Differ. Equ.* **128** 415–90
[23] Giorgilli A and Galgani L 1985 Rigorous estimates for the series expansions of Hamiltonian perturbation theory *Celest. Mech.* **37** 95–112
[24] Neishtadt A I 1975 On passage through resonances in the two-frequency problem *Sov. Phys. - Dokl.* **20** 189–91
[25] Neishtadt A I 2014 Averaging, passage through resonances, and capture into resonance in two-frequency systems *Russ. Math. Surv.* **69** 771–843
[26] Biasco L and Chierchia L 2020 Action–angle Variables for Generic 1D Mechanical Systems arXiv:2003.05211
[27] Benettin G, Galgani L and Giorgilli A 1985 A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems *Celest. Mech.* **37** 1–25

[28] Guzzo M, Lega E and Froeschlé C 2006 Diffusion and stability in perturbed non-convex integrable systems *Nonlinearity* **19** 1049–67
[29] Nekhoroshev N N 1977 An exponential estimate of the time of stability of nearly- integrable Hamiltonian systems I *Uspekhi Mat. Nauk.* **32** 5–66
Nekhoroshev N N 1977 An exponential estimate of the time of stability of nearly- integrable Hamiltonian systems I *Russ. Math. Surv.* **32** 1–65
[30] Nekhoroshev N N 1979 Exponential estimates of the stability time of near-integrable Hamiltonian systems 2 *Trudy Sem. Petrovs.* **5** 5–50
[31] Niederman L 2007 Prevalence of exponential stability among nearly integrable Hamiltonian systems *Ergod. Theor. Dyn. Syst.* **27** 905–28
[32] Levin B Y 1996 *Lectures on Entire Functions* (Providence, RI: American Mathematical Society)