Singular KAM Theory

L. Biasco & L. Chierchia
Dipartimento di Matematica e Fisica
Università degli Studi Roma Tre
Via della Vasca Navale, 84 - 00146 Roma, Italy
luca.biasco@uniroma3.it, luigi.chierchia@uniroma3.it

September 29, 2023

Dedicated to the memory of our friend and colleague Walter L. Craig

Abstract

The question of the total measure of invariant tori in analytic, nearly–integrable Hamiltonian systems is considered. In 1985, Arnol'd, Kozlov and Neishtadt, in the Encyclopaedia of Mathematical Sciences [4], and in subsequent editions, conjectured that in $n = 2$ degrees of freedom the measure of the non torus set of general analytic nearly–integrable systems away from critical points is exponentially small with the size ε of the perturbation, and that for $n \geqslant 3$ the measure is, in general, of order ε (rather than $\sqrt{\varepsilon}$ as predicted by classical KAM Theory).
In the case of generic natural Hamiltonian systems, we prove lower bounds on the measure of primary and secondary invariant tori, which are in agreement, up to a logarithmic correction, with the above conjectures.
The proof is based on a new *singular* KAM theory, particularly designed to study analytic properties in neighborhoods of the secular separatrices generated by the perturbation at simple resonances.

MSC2010 numbers: 37J05, 37J35, 37J40, 70H05, 70H08, 70H15
Keywords: Nearly–integrable systems. Natural Hamiltonian systems. Singular KAM Theory. Measure of invariant tori. Primary and secondary tori. Simple resonances. Hamiltonian perturbation Theory. Kolmogorov's non–degeneracy. Measure of the non–torus set.

Introduction

Classical KAM Theory[1] deals with the persistence of Lagrangian invariant tori of integrable Hamiltonian systems under the effect of small perturbations. In the early 1980's it was clarified that an analytic integrable system, which is Kolmogorov non–degenerate (i.e., such that the action–to–frequency map is a local diffeomorphism), preserves, under a perturbation of size $\varepsilon > 0$, all its Diophantine Lagrangian invariant tori in a bounded domain *up to a set of measure proportional to*[2] $\sqrt{\varepsilon}$. In fact, this estimate cannot be improved, since trivial examples – such as a classical pendulum with Hamiltonian $\frac{p^2}{2} + \varepsilon \cos q$ – show that, in bounded domains, *the measure of the complement of persistent primary*

[1] [27], [1], [32], [2], [33]; for a divulgative account, see [22].
[2] [28], [34], [35], [38]

tori (i.e., tori which are a deformation of integrable ones) *is exactly proportional to the square root of the perturbing function* – the rest of the phase space being filled, in the case of the pendulum, by *secondary tori* (curves) enclosed by the pendulum separatrix. In fact, positive measure sets of secondary Lagrangian tori (i.e., tori, which are not a smooth deformation of integrable ones) appear in general nearly–integrable systems, for example, near elliptic equilibria ([31]).

The natural question is therefore: *What is the measure of all Lagrangian tori in general nearly–integrable analytic Hamiltonian systems?*

In 1985 Arnold, Kozlov and Neishtadt, motivated by the exponentially small splitting of separatrices in general systems with two degrees of freedom, conjectured that[3]

"It is natural to expect that in a generic (analytic) system with two degrees of freedom and with frequencies that do not vanish simultaneously the total measure of the 'non-torus' set corresponding to all the resonances is exponentially small."

In [5], again Arnold, Kozlov and Neishtadt, arguing on the basis of a simple rescaling argument in neighbourhoods of double resonances[4], conjectured that

"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."

In this paper, we develop a 'singular KAM theory' for generic analytic *nearly–integrable natural systems*, apt to deal, in particular, with the construction of maximal KAM tori that live exponentially close to the separatrices appearing near simple resonances, which are singularities of the action–angle variables of the integrable secular (averaged) systems. As a consequence, we can prove lower bounds on the total measure of KAM tori, which are in agreement with the above conjectures up to a logarithmic correction $|\log\varepsilon|^\gamma$. We announced these results in 2015 in [9] (see also [13]), and it goes without saying that to complete proofs took *much* longer than we thought.

The reason for dealing with the special class of nearly–integrable natural systems, namely, Hamiltonian systems on $\mathbb{R}^n \times \mathbb{T}^n$ (endowed with the standard symplectic form $dy \wedge dx$) with Hamiltonian given by

$$\mathtt{H}(y,x;\varepsilon) := \tfrac{1}{2}|y|^2 + \varepsilon f(x)\,, \quad (y,x) \in \mathbb{R}^n \times \mathbb{T}^n\,,\ 0 < \varepsilon < 1\,, \quad \big(|y|^2 := y\cdot y := \textstyle\sum_j |y_j|^2\big)\,, \tag{1}$$

is twofold. On one side, this choice allows to avoid technical unessential details, which would make even heavier the already highly technical methods. On the other hand, and more importantly, it allows to formulate the genericity condition (whose definition is part of the problem) in a simple way, singling out a suitable class of *generic analytic* potentials f's (Definition 1.1), which guarantees, in particular, a uniform behaviour of the *secondary nearly–integrable structure* at simple resonances with high modes (compare Open Problems, (i) in § 1).

Let us informally discuss the overall picture.

Analogously to what is done in Nekhoroshev theory (compare, e.g., [36], and [5] for general information), fixed a maximal size of resonances $\mathtt{K}$ to be taken into account[5], one covers the action space

[3] Compare [4, p. 189] and [5, Remark 6.17, p. 285].

[4] From p. 285 of [5]: "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 $1/2(Ap,p) + V(q_1,q_2)$, which does not involve a small parameter. 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 $\sim\varepsilon$."

[5] Actually, we will need to consider *two* orders of resonances $\mathtt{K}_{\rm o}$ and $\mathtt{K} > \mathtt{K}_{\rm o}$; but for the purpose of this introduction we call them both $\mathtt{K}$.

with three sets: a non–resonant set $\mathcal{R}^0$, a $\sqrt{\varepsilon}\mathtt{K}^c$–neighbourhood[6] $\mathcal{R}^1$ of simple resonances, and a neighborhood $\mathcal{R}^2$ of double (and higher) resonances. Eventually, the number of resonances $\mathtt{K}$ is taken as a suitable function of ε tending to $+\infty$ as $\varepsilon \to 0$ (e.g., $\mathtt{K} \sim 1/\varepsilon^c$, or $\mathtt{K} \sim |\log \varepsilon|$).
The set $\mathcal{R}^2$, which has measure proportional to $\varepsilon\mathtt{K}^c$, is a *non perturbative set* in the sense that the dynamics ruled by $\mathtt{H}(y, x; \varepsilon)$ on $\mathcal{R}^2 \times \mathbb{T}^n$ is essentially equivalent to the dynamics of the parameter free Hamiltonian $\frac{1}{2}|y|^2 + f(x)$ (compare the argument given by Arnold, Kozlov and Neishtadt, reproduced in footnote 4 above). Therefore, no further perturbative analysis on the set $\mathcal{R}^2 \times \mathbb{T}^n$ is possible.
On the non resonant phase space $\mathcal{R}^0 \times \mathbb{T}^n$, after high order averaging, classical KAM theory yields the existence of primary maximal KAM tori up to a set of measure $O(\sqrt{\varepsilon}e^{-c\mathtt{K}})$.

The main game has then to be played on the simple–resonance neighborhood $\mathcal{R}^1 \times \mathbb{T}^n$.
$\mathcal{R}^1$ is defined as union of sets $\mathcal{R}^{1,k}$, which are $\sqrt{\varepsilon}\mathtt{K}^c$–neighborhoods of simple resonances $\{y \big| \, y \cdot k = 0\}$ with $k \in \mathbb{Z}^n$ and co–prime entries. On $\mathcal{R}^{1,k}$ high–order averaging theory can be applied so as to remove, up to order $\varepsilon e^{-c\mathtt{K}}$, the angle dependence, apart from the resonant combination $k \cdot x$, obtaining a symplectically conjugated real analytic Hamiltonian of the form

$$\mathtt{H}_k(y, x) = \frac{|y|^2}{2} + \varepsilon\big(g^k_{\rm o}(y) + g^k(y, k \cdot x) + f^k(y, x)\big)\,, \qquad f^k \sim e^{-c\mathtt{K}}\,. \tag{2}$$

Now, all these Hamiltonian systems labelled by the simple–resonance index k ($|k| \leqslant \mathtt{K}$), have a *secondary* (secular) *near–integrability structure*, as, disregarding the exponentially small terms f^k, they are Arnol'd–Liouville integrable, depending effectively only on one resonant angle $\mathtt{x}_1 = k \cdot x \in \mathbb{T}^1$.
Then, the plan is obvious: Put all these systems into their Arnol'd–Liouville action–angle variables, check twist (i.e., Kolmogorov's non–degeneracy), and apply KAM so as to obtain Lagrangian primary and secondary tori (with different topologies; compare Remark 5.3).
However, a considerable series of problems arise in trying to carry out such a plan. Let us try to highlight the most important points.

First of all, as already mentioned, the resonance cut–off $\mathtt{K}$ will go to $+\infty$ as $\varepsilon \to 0$ and therefore one has to deal, *de facto*, with infinitely many Hamiltonian systems and unless there is some uniform way of treating them, there is no hope. The idea, here, has been suggested in [11] and refined in [15]: The *secular Hamiltonians* $\bar{\mathtt{H}}_k(\mathtt{y}, \mathtt{x}_1)$, i.e., the integrable Hamiltonians in (2) obtained disregarding f^k and setting $\mathtt{x}_1 = k \cdot x$, are one–degree–of–freedom Hamiltonians, with external parameters, and with potentials $g^k(y, \mathtt{x}_1)$, which are close to the projections $(\pi_{\mathbb{Z}k} f)(\mathtt{x}_1)$ over the Fourier modes proportional to k of the potential $f(x)$; compare (7) below. Now, one can show that for high Fourier modes $|k| > \mathtt{N}$ ($\mathtt{N}$ suitable but independent of ε), $(\pi_{\mathbb{Z}k} f)(\mathtt{x}_1)$ behaves generically as a shifted cosine[7]

$$2|f_k|\varepsilon \cos(\mathtt{x}_1 + \theta_k)$$

for a suitable $\theta_k \in \mathbb{R}$; where 'generically' means that f belongs to a suitable class of generic real–analytic potential, whose Fourier coefficients, for large k's with co–prime entries, behave as $e^{-|k|s}|k|^{-n}$ for a suitable $s > 0$. Incidentally, to obtain such a result, one has to use a non–standard averaging theory, allowing for essentially *no analyticity loss* in the angle variables; for more information on this point, see the Introduction in [11].

[6]In this introduction, we indicate with 'c' various different constants, which are independent of ε. In general, keeping track of the quantities, on which the various constants appearing in singular KAM Theory depend, is a somewhat important matter (for example, from the constructive point of view) and we try to devote some care to it; compare, e.g., Remark (R_5) in § 1.
[7]Compare item (iii) in Theorem 2.1 below.

Analytic properties of the action–angle variables for the pendulum are quite well known, and this is encouraging (and it was also the basis for the optimistic 2015 announcement [9]).
However, for low modes $|k| \leqslant \mathtt{N}$, the secular leading potentials $(\pi_{\mathbb{Z}k} f)(\mathtt{x}_1)$ are, in general, quite arbitrary functions, and one needs, therefore, a general holomorphic, quantitative theory of action–angle variable for one–degree–of freedom systems containing parameters. Such a theory is discussed in [14] for a special class of real analytic Hamiltonians – called there *Generic Standard Form Hamiltonians* – given by

$$\mathtt{H}_\flat(p, q_1) = \big(1 + \mathsf{v}(p, q_1)\big)p_1^2 + \mathtt{G}(\hat{p}, q_1)\,, \tag{3}$$

where $p = (p_1, \hat{p})$, p_1 is the momentum conjugated to the angle q_1 and $\hat{p} = (p_2, ..., p_n)$ are the 'external parameters'; see Definition 2.1 for specifications. In particular, the properties of the energy–to–action functions are discussed in the limit as the energy approaches the critical values (i.e., the energy levels of the hyperbolic points and the associated separatrices): It turns out that such functions have the form

$$I_1(E_{\rm crit} \pm \epsilon z) = a(z) + b(z)\, z \log z \tag{4}$$

where ϵ is a suitable reference energy, $E_{\rm crit}$ is a fixed critical energy level of some equilibrium of the secular system, a and b are analytic functions of z (and, of course, everything depends on other $(n-1)$ dumb action; compare Theorem 2.3 below). This representations will play a crucial rôle in studying the twist of the secular Hamiltonians at simple resonances in their Arnold-Liouville action–angle variables.
Now, one can prove ([15]) that *all secular Hamiltonians* $\overline{\mathtt{H}}_k$ *can be put into standard form as in* (3), so that the main rescaling properties are controlled by one single parameter κ, which is independent of ε *and* k (compare Theorem 2.2 below). The draw back of this uniformization is that the symplectic transformations performing the task *are not well defined in the fast angle–variables* $(\mathtt{x}_2, ..., \mathtt{x}_n)$, and preserves periodicity only in the resonant angle $\mathtt{x}_1$.

This is the starting point of this paper.

In § 3 we show how to overcome the homotopy problem of the uniformization of [15]: Exploiting the particular group structure of the various symplectic transformations involved, we show that, introducing a special *ad hoc* symplectic 'semi–conjugacy', one can indeed obtain, for all $|k| \leqslant \mathtt{K}$, well defined symplectic action–angle maps ϕ_k^i, which conjugate the original Hamiltonian $\mathtt{H}$ in (1) on $\mathcal{R}^{1,k} \times \mathbb{T}^n$ to the nearly–integrable form

$$\mathcal{H}_k^i := \mathtt{H} \circ \phi_k^i(I, \varphi) = h_k^i(I) + \varepsilon f_k^i(I, \varphi)\,, \qquad f_k^i \sim e^{-c\mathtt{K}}\,,$$

where i labels the various regions in which the phase spaces of the secular Hamiltonians $\overline{\mathtt{H}}_k$ are split by their separatrices (compare Theorem 3.1 below).

As in all KAM applications, the main problem is to prove (a suitable) *non–degeneracy* of the frequency map $I \to \omega = \partial_I h_k^i$.
It should be clear from the context, that the original non–degeneracy of $\mathtt{H}|_{\varepsilon=0}$ plays a little rôle here, as the action–structure depends on analytic properties of the secular potentials. Indeed, it is a fact that *in the phase space of standard Hamiltonians* (3) *there are, in general, points where the twist vanishes*. For instance, points of vanishing twist appear *always* in regions bounded by two separatrices (with different energy); but they appear also in very simple examples with only one separatrix near the elliptic equilibrium enclosed by the separatrix, like, e.g., in the case of the Hamiltonian

$$p_1^2 + \cos q_1 - \tfrac{1}{8}\cos(2q_1)\,;$$

compare Remark 4.1 below. Furthermore (and more seriously), when the distance in energy from the separatrices goes to zero, the problem becomes a *singular perturbation problem* with dramatic singularities.

Therefore, entirely new methods have to be developed in order to prove that *the measure where the twist vanishes is actually exponentially small in the whole phase space of all Hamiltonians* $\mathcal{H}^i_k$. This is the main result of the paper; compare the Twist Theorem 4.1 in § 4.

The proof of the Twist Theorem is based on two different approaches according to whether one considers regions *far* from separatrices or regions *close* to separatrices.

In regions far from separatrices the analysis is significantly simpler, since it is partly perturbative. In such a case, one fist proves that the (normalized) second derivative of the action–to–energy functions are non–degenerate (i.e., at each point of their domains, some derivative is different from zero); then, uniform estimates can be worked out and, using standard tools from the theory of Diophantine approximations ([37], [23]), one can show that η–sub–levels of the twist determinant have measure smaller than η^c, which easily yields the claim.

The real heart of the matter is the analysis of the twist in regions close to separatrices, where no perturbative arguments can be used, nor uniform estimates hold. The proof, in this case, rests on the construction of a suitable differential operator with non–constant coefficients, which, exploiting in a subtle way the analytic structure (4), can be shown not to vanish on a suitable regularization of the twist determinant. This is good enough to prove that the twist determinant is non–degenerate also near separatrices, and to conclude the proof of the Twist Theorem.

At this point, choosing carefully the various free parameters of the game, a suitable KAM Theorem (Theorem 5.1 below) yields the existence of maximal primary and secondary KAM tori, which fill the complementary phase set of $\mathcal{R}^2 \times \mathbb{T}^n$ up to a very small set $\mathcal{A}$. How small is $\mathcal{A}$ – which dynamically is very rich and where, e.g., Arnol'd diffusion can take place – depends on how big is chosen the order $\mathtt{K}$ of resonances considered. For example, if $\mathtt{K}$ is chosen as $|\log \varepsilon|^2$, then meas($\mathcal{A}$) will be almost–exponentially small (i.e., smaller than any power of ε), while meas($\mathcal{A}$) is actually exponentially small in $1/\varepsilon^c$, if $\mathtt{K}$ is taken to be an inverse power of ε; compare Remark (R_2) in § 1.

Acknowledgments During the decade needed to complete this paper the authors benefited from conversations with A. Delshams, M. Guardia, V. Kaloshin, S. Kuksin, C. Liverani, A. Neishtadt, L. Niederman, M. Procesi, V. Rom Kedar, T. Seara, A. Sorrentino and D. Treshev.

The authors acknowledge the encouragement of Caterina Biasco.

This work is supported by the Italian grant PRIN 2022FPZEES.

Contents

1 Results, Remarks, and Open Problems

In order to state the main results of this paper, we recall a few standard definitions.

• *Maximal KAM tori*: A set $\mathcal{T} \subset \mathcal{M} = B \times \mathbb{R}^n$ is called a *maximal KAM torus* for a real analytic Hamiltonian $H : \mathcal{M} \to \mathbb{R}$ if there exist a real analytic embedding $\phi : \mathbb{T}^n \to \mathcal{M}$ and a Diophantine frequency vector[8] $\omega \in \mathbb{R}^n$ such that $\mathcal{T} = \phi(\mathbb{T}^n)$, and for each $z \in \mathcal{T}$, $\Phi_H^t(z) = \phi(x + \omega t)$, where[9] $x = \phi^{-1}(z)$ and $t \to \Phi_H^t(z)$ denotes the standard Hamiltonian flow governed by H starting at $z \in \mathcal{M}$.

• *Generators of 1d maximal lattices*: Let $\mathbb{Z}^n_\star$ be the set of integer vectors $k \neq 0$ in $\mathbb{Z}^n$ such that the first non–null component is positive:

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

$\mathcal{G}^n$ denotes the set of *generators of 1d maximal lattices* in $\mathbb{Z}^n$, namely, the set of vectors $k \in \mathbb{Z}^n_\star$ such that the greater common divisor (gcd) of their components is 1:

$$\mathcal{G}^n := \{k \in \mathbb{Z}^n_\star : \ \gcd(k_1, \ldots, k_n) = 1\}\,; \tag{5}$$

for $K \geqslant 1$, we set[10]:

$$\mathcal{G}^n_K := \mathcal{G}^n \cap \{|k|_1 \leqslant K\}\,. \tag{6}$$

• *1d Fourier projectors*: Given a zero–average real analytic periodic function

$$f : x \in \mathbb{T}^n := \mathbb{R}^n/(2\pi\mathbb{Z}^n) \mapsto f(x) := \sum_{\mathbb{Z}^n\setminus\{0\}} f_k e^{\mathrm{i}k\cdot x}$$

and fixed a vector $k \in \mathbb{Z}^n \setminus \{0\}$, we denote by $\pi_{\mathbb{Z}k} f$ the (real analytic) periodic function of *one variable* $\theta \in \mathbb{T}$ given by

$$\theta \in \mathbb{T} \mapsto \pi_{\mathbb{Z}k} f(\theta) := \sum_{j\in\mathbb{Z}} f_{jk} e^{\mathrm{i}j\theta}\,. \tag{7}$$

Notice that one has the following (unique) decomposition:

$$f(x) = \sum_{k\in\mathcal{G}^n} \pi_{\mathbb{Z}k} f(k\cdot x)\,.$$

• *Resonances*: Given $k \in \mathcal{G}^n$, a resonance $\mathcal{R}_k$ with respect to the free Hamiltonian $\frac{1}{2}|y|^2$ is the set $\{y \in \mathbb{R}^n : y\cdot k = 0\}$. We call $\mathcal{R}_{k,\ell}$ a *double resonance* if $\mathcal{R}_{k,\ell} = \mathcal{R}_k \cap \mathcal{R}_\ell$ with k and ℓ in $\mathcal{G}^n$ linearly independent; the order of a double resonance is given by $\max\{|k|_1, |\ell|_1\}$.

[8]A vector $\omega \in \mathbb{R}^n$ is called *Diophantine* if there exist $\alpha > 0$ and $\tau \geqslant n-1$ such that $|\omega\cdot k| \geqslant \alpha/|k|_1^\tau$, for any non vanishing integer vector $k \in \mathbb{Z}^n$, where $|k|_1 := \sum |k_j|$.

[9]In particular, maximal KAM tori are minimal invariant invariant sets for Φ_H^t.

[10]As usual $|k|_1 := \sum_{j=1}^n |k_j|$.

• *Morse functions with distinct critical values*: A C^2–function of one variable $\theta \to F(\theta)$ is a Morse function if its critical points are non–degenerate, i.e., $F'(\theta_0) = 0 \implies F''(\theta_0) \neq 0$; 'distinct critical values' means that if $\theta_1 \neq \theta_2$ are distinct critical points, then $F(\theta_1) \neq F(\theta_2)$.

• *Banach spaces of real analytic periodic functions*: For $s > 0$ and $n \in \mathbb{N} = \{1, 2, 3...\}$, consider the Banach space of zero–average real analytic periodic functions on $\mathbb{T}^n$ with finite norm

$$\|f\|_s := \sup_{k\in\mathbb{Z}^n} |f_k| e^{|k|_1 s}\,, \tag{8}$$

and denote by $\mathbb{B}^n_s$ *its closed unit ball.*

Now we are ready to introduce a suitable generic class of potentials f and state the main result of this paper concerning the typical dynamics of nearly–integrable natural systems with Hamiltonian $\mathtt{H}$ as in (1).

Definition 1.1 (The class of potentials $\mathbb{G}^n_s$) *We denote by $\mathbb{G}^n_s$ the subset of functions $f \in \mathbb{B}^n_s$ such that the following two properties hold:*

$$\varliminf_{\substack{|k|_1\to+\infty\\ k\in\mathcal{G}^n}} |f_k| e^{|k|_1 s} |k|_1^n > 0\,, \tag{9}$$

$$\forall\ k \in \mathcal{G}^n\,,\ \ \pi_{\mathbb{Z}k} f \ \text{ is a Morse function with distinct critical values}\,. \tag{10}$$

Remark 1.1 (i) For natural systems – whose natural phase space is $\mathbb{R}^n \times \mathbb{T}^n$ – any ball in action space can be transformed (by translation and rescaling) to the unit ball; thus, rescaling time and renaming the smallness parameter one can always restrict oneself to study the Hamiltonian $\mathtt{H}$ in (1) on the unit ball in $\mathbb{R}^n$ and with $\|f\| = 1$.

(ii) The requirement in (10) that the projections $\pi_{\mathbb{Z}k} f$ have different critical values is not really necessary, but it is generic and it simplifies the proof.

Main Results

Theorem 1.1 *Let $n \geqslant 2$, $s > 0$, $\mathrm{B} := \{y \in \mathbb{R}^n :\ |y| < 1\}$, $\gamma := 11n + 4$, $f \in \mathbb{G}^n_s$ with $\|f\|_s = 1$. Then, there exist a constant $\mathtt{c} > 1$, such that for all $\mathtt{K}$ and $\varepsilon > 0$ satisfying*

$$\mathtt{K} \geqslant \mathtt{c}\,, \qquad\qquad \varepsilon\, \mathtt{K}^\gamma \leqslant 1\,, \tag{11}$$

the following holds. There exist three sets $\mathcal{R}^2 \subseteq \mathrm{B}$, $\mathcal{A} \subseteq \mathrm{B} \times \mathbb{T}^n$, $\mathcal{T} \subseteq \mathbb{R}^n \times \mathbb{T}^n$ such that:

(i) $\mathrm{B} \times \mathbb{T}^n \subseteq (\mathcal{R}^2 \times \mathbb{T}^n) \cup \mathcal{A} \cup \mathcal{T}$;

(ii) $\mathcal{R}^2$ *is a neighborhood of double resonances of order smaller than $\mathtt{K}$ satisfying the measure estimate*

$$\operatorname{meas} \mathcal{R}^2 \leqslant c_\star\, \varepsilon\, \mathtt{K}^\gamma\,,$$

where $c_\star$ is a suitable constant depending only on n;

(iii) $\mathcal{A}$ *is exponentially small with respect to $\mathtt{K}$:*

$$\operatorname{meas} \mathcal{A} \leqslant e^{-\mathtt{K}/\mathtt{c}}\,;$$

(iv) $\mathcal{T}$ *is union of maximal KAM tori for the natural Hamiltonian* $\mathtt{H}(y,x;\varepsilon) := \frac{1}{2}|y|^2 + \varepsilon f(x)$.

An immediate corollary of this theorem is that the measure of the 'non–torus set' for $\mathtt{H}$ does not exceed $O(\varepsilon|\log\varepsilon|^\gamma)$:

Corollary 1.1 *Under the assumptions of Theorem 1.1, there exists* $0<\varepsilon_{\rm o}<1$ *such that for* $\varepsilon<\varepsilon_{\rm o}$*, all points in* $\mathrm{B}\times\mathbb{T}^n$ *lie on a maximal KAM torus for* $\mathtt{H}$*, except for a subset whose measure is bounded by* $\bar{\mathtt{c}}\ \varepsilon|\log\varepsilon|^\gamma$ *where* $\bar{\mathtt{c}}\ = 1+(2\pi)^n c_\star \mathfrak{c}^\gamma$.

In particular, this corollary implies the theorem announced in [9, p. 426].
Indeed, from items (i), (ii) and (iii) of Theorem 1.1, there follows

$$\mathrm{meas}\,\big((\mathrm{B}\times\mathbb{T}^n)\setminus\mathcal{T}\big) \leqslant (2\pi)^n c_\star\,\varepsilon\mathtt{K}^\gamma + e^{-\mathtt{K}/\mathfrak{c}}\,, \tag{12}$$

and Corollary 1.1 follows immediately by choosing $\mathtt{K} := \mathfrak{c}|\log\varepsilon|$ (and $\varepsilon_{\rm o}$ small enough).
The two degrees of freedom is special: in this case the only double resonance is the origin and one can take as $\mathcal{R}^2$ a disk of measure ε^a with any $0<a<1$ getting a set of KAM tori of exponential density in the complementary of $\mathcal{R}^2\times\mathbb{T}^2$. This is the content of next corollary (compare, also, [12]).

Corollary 1.2 *Let the assumptions of Theorem 1.1 hold and let* $n=2$*. Then, there exists* $0<\varepsilon_{\rm o}<1$*, such that for* $\varepsilon<\varepsilon_{\rm o}$ *and* $0<a<1$*, all points in the set* $\{y\in\mathrm{B} : |y|>\varepsilon^{a/2}\}\times\mathbb{T}^2$ *lie on a maximal KAM torus for* $\mathtt{H}$ *in (1), except for an exponentially small set of measure bounded by* $e^{-1/(2\mathfrak{c}\varepsilon^{\hat a})}$*, with* $\hat a := (1-a)/24$.

Remarks and Open Problems

First, we briefly discuss the class of potentials $\mathbb{G}^n_s$, for which the above results hold; for more information on $\mathbb{G}^n_s$ and complete proofs, see [15, Sect. 2].
It is very simple to give explicit examples of functions in $\mathbb{G}^n_s$, a prototype being[11]

$$f(x) := 2\sum_{k\in\mathcal{G}^n} e^{-|k|_1 s}\cos k\cdot x\,, \tag{13}$$

which (as it is trivial to verify) satysfies

$$\|f\|_s = 1\,,\qquad \lim_{\substack{|k|_1\to+\infty\\ k\in\mathcal{G}^n}} |f_k|e^{|k|_1 s}|k|_1^n = +\infty\,,\qquad \pi_{\mathbb{Z}k} f(\theta) = 2e^{-|k|_1 s}\cos\theta\,.$$

The class of potentials $\mathbb{G}^n_s$ is quite general from various points of view. For example, the following result (proven in [15, Sect. 1]) holds:

Proposition 1.1 (i) *The class* $\mathbb{G}^n_s$ *contains an open and dense set in* $\mathbb{B}^n_s$.
(ii) *Let* F *denote the weighted Fourier isometry* $\mathrm{F} : f\in\mathbb{B}^n_s \to \big\{f_k e^{|k|_1 s}\big\}_{k\in\mathbb{Z}^n_\star} \in \ell^\infty(\mathbb{Z}^n_\star)$*. Then* $\mathrm{F}(\mathbb{G}^n_s)$ *is a set of probability 1 with respect to the standard product probability measure on* $\mathrm{F}(\mathbb{B}^n_s)$.

[11] Recall the definitions given in (5), (8) and (7).

We shall not use this proposition here, however, we shall often use a suitable quantitative characterization of $\mathbb{G}^n_s$.
To state such a characterization one needs to make quantitative the notion of Morse functions with different critical values and to introduce a uniform Fourier cut–off function, depending on the dimension n and on two parameters $0 < \delta \leqslant 1$ and $s > 0$.

Definition 1.2 **(β–Morse functions)** *Let $\beta > 0$. $F \in C^2(\mathbb{T}, \mathbb{R})$ is called β–Morse, if*

$$\min_{\theta \in \mathbb{T}} \big(|F'(\theta)| + |F''(\theta)|\big) \geqslant \beta\,, \quad \min_{i \neq j} |F(\theta_i) - F(\theta_j)| \geqslant \beta\,, \tag{14}$$

where $\theta_i \in \mathbb{T}$ are the critical points of F.

Definition 1.3 **(The cut–off function $\mathtt{N}$)** *Given $0 < \delta \leqslant 1$ and $n, s > 0$ define the following 'Fourier cut–off function':*

$$\mathtt{N} = \mathtt{N}(\delta; s, n) := 2 \max\left\{1\,, \frac{1}{s} \log \frac{c_{_n}}{s^n\, \delta}\right\}, \qquad c_{_n} := 2^{44}\ (2n/e)^n\,. \tag{15}$$

Then, the following elementary result holds:

Lemma 1.1 *Let $n, s > 0$. Then, $f \in \mathbb{G}^n_s$ if and only if $f \in \mathbb{B}^n_s$ and there exist $0 < \delta \leqslant 1$ and $\beta > 0$ such that*

$$|f_k| \geqslant \delta |k|_{_1}^{-n}\, e^{-|k|_1 s}\,, \qquad \forall\ k \in \mathcal{G}^n\,,\ |k|_{_1} \geqslant \mathtt{N}\,, \tag{16}$$

$$\pi_{_{\mathbb{Z}k}} f \ \text{ is } \beta\text{–Morse}\,, \qquad \forall\ k \in \mathcal{G}^n\,,\ |k|_{_1} \leqslant \mathtt{N}\,. \tag{17}$$

The proof of this lemma is given in Appendix.

Remarks

(R_1) By item (i) in Theorem 1.1, we see that the phase space of a generic nearly–integrable natural system can be covered by two 'small' sets, namely, $\mathcal{R}^2 \times \mathbb{T}^n$ and $\mathcal{A}$, and one 'large' set, namely, $\mathcal{T}$. Such sets exhibit quite different dynamics and satisfy the following properties.

- $\mathcal{R}^2 \times \mathbb{T}^n$ is contained in a tubolar neighborhood of double resonances $\mathcal{R}_{k,\ell} \cap \mathrm{B}$ of order not exceeding $\mathtt{K}$; compare (25) and (26) below. As mentioned in the Introduction, the set $\mathcal{R}^2 \times \mathbb{T}^n$ contains a set of measure ε where *the dynamics is not perturbative* in the sense that, as long as trajectories lie in this set, the dynamics is ruled by an effective Hamiltonian having no small parameters.

- The set $\mathcal{A}$, which has measure $\sim e^{-\mathtt{K}/\mathtt{c}}$, is dynamically very interesting. For example, it is where the asymptotic manifolds of lower dimensional tori break up ('exponentially small splitting of separatrices') giving rise, e.g., to local horse shoe dynamics and, most likely, to Arnol'd diffusion[12].

- In the complement of the two above small sets, namely in $\mathcal{T} \cap (\mathrm{B} \times \mathbb{T}^n)$, *all trajectories lie on maximal KAM tori* for $\mathtt{H}$ and the dynamics is quasi–periodic (with possibly exponentially small Diophantine constant).

[12]This statement has not yet been proven in any generality in the analytic class; for references to Arnol'd diffusion see, e.g., [3], [18], [8], [39], [20], [40], [30], [7], [21], [17], [25], [26].

(R_2) The sets $\mathcal{R}^2$ and $\mathcal{A}$ depend, in particular, upon ε and $\mathtt{K}$ and choosing $\mathtt{K}$ as a suitable function of ε so as to obtain larger neighborhoods of double resonances, leads to different coverings of the phase space with improved measure estimates on $\mathcal{A}$. For example:

• *Almost–exponential density outside a region of measure $O(\varepsilon|\log\varepsilon|^{2\gamma})$:*
Letting $\mathtt{K} = \log^2\varepsilon$ in Theorem 1.1 with ε small enough (so that (11) is met), the sets $\mathcal{R}^2$ and $\mathcal{A}$ satisfy the estimates

$$\mathrm{meas}(\mathcal{R}^2 \times \mathbb{T}^n) \leqslant c_\star\, \varepsilon\, |\log\varepsilon|^{2\gamma}\,, \qquad \mathrm{meas}\,\mathcal{A} \leqslant \varepsilon^{(\log|\varepsilon|)/\mathtt{c}}\,.$$

In other terms, *outside a neighborhood of $O(\varepsilon\,|\log\varepsilon|^{2\gamma})$ of double resonances, the 'non–torus set' is almost exponentially small* (i.e., smaller than any power of ε).

If we allow for a neighborhood of double resonances of size ε^a with $a < 1$, we get a pure exponential density of KAM tori outside $\mathcal{R}^2 \times \mathbb{T}^n$:

• *Exponential density outside outside a region of measure $O(\varepsilon^a)$:*
Let $0 < a < 1$ and choose $\mathtt{K} = 1/\varepsilon^{\bar a}$ with $\bar a := (1-a)/\gamma$ in Theorem 1.1 and let ε be small enough. Then, the sets $\mathcal{R}^2$ and $\mathcal{A}$ satisfy the estimates

$$\mathrm{meas}(\mathcal{R}^2 \times \mathbb{T}^n) \leqslant c_\star\, \varepsilon^a\,, \qquad \mathrm{meas}\,\mathcal{A} \leqslant e^{-1/\mathtt{c}\varepsilon^{\bar a}}\,. \tag{18}$$

In other terms, *outside a neighborhood of $O(\varepsilon^a)$ of double resonances, the 'non–torus set' is exponentially small in $1/\varepsilon$.*

(R_3) As anticipated in the Introduction, Corollary 1.2 and Corollary 1.1 prove – or, more precisely, are in agreement with – the conjectures made by Arnol'd, Kozlov and Neishtadt mentioned in the Introduction. Notice, however, that the argument sketched by Arnol'd, Kozlov and Neishtadt in [5] to support their conjecture for $n \geqslant 3$ (reported in footnote 4 above) only suggests a *lower* bound on the measure of the non–torus set, while, here, we provide a rigorous *upper* bound on it.
We also mention that that Corollary 1.2 is a particular case (with slightly better constants) of (18), but, since we are in two action–dimensions, it is possible to take $\mathcal{R}^2$ simply as a small ball around the origin (while for $n \geqslant 3$ it is a more complicate set).

(R_4) The 'Kolmogorov's set' $\mathcal{T}$ if formed by primary and secondary tori: Such secondary tori are *not* deformation of integrable tori and, in particular, they are never graphs over $\mathbb{T}^n$.
We remark also that the set $\mathcal{T}$ is *not* contained in $\mathrm{B} \times \mathbb{T}^n$, and indeed many of the invariant tori in $\mathcal{T}$ (corresponding to a set of measure $\sim \sqrt{\varepsilon}$) have oscillations outside $\mathrm{B} \times \mathbb{T}^n$; this fact is unavoidable, as near the boundary tori do oscillate by a quantity of order $\sqrt{\varepsilon}$.

(R_5) In Theorem 1.1 there appear two constants $\mathtt{c}$ and $c_\star$ (the other constants appearing in the corollaries of Theorem 1.1 are simply related to such two constants). The constant $c_\star$ depends only on the action–dimension n (compare Lemma 2.1 below). More relevant for measure estimates in phase space is the constant $\mathtt{c}$.
The constant $\mathtt{c}$ can be calculated in terms of a few analytic properties of the potential f. In fact, $\mathtt{c}$ depends on six parameters: $n \geqslant 2$; $s > 0$; a positive number δ quantifying property (9) (compare (16)); a positive number β quantifying property (10) (compare (17)); and two more positive parameters $\mathtt{\xi}$ and $\mathtt{m}$ (introduced in Definition 4.3), which, in turn, are suitable non–degeneracy parameters associated to the (normalized) second derivative of the action–to–energy maps associated to the integrable 1–degree–of–freedom Hamiltonians $p^2 + \pi_{\mathbb{Z}k} f(q)$ with $|k|_1 \leqslant \mathtt{N}$ where $\mathtt{N}$ is as in Definition 1.3. For more details on how the constant $\mathtt{c}$ depends on the various parameters, see (243) and (244) below.

(R_6) The actual effective hypothesis of Theorem 1.1 is the condition $\mathtt{K} \geqslant \mathtt{c}$ in (11), which is used almost in all the proofs given in this paper.

The second condition, $\varepsilon\mathtt{K}^\gamma \leqslant 1$, strictly speaking, is not really necessary, and it is used (for simplicity) only in Proposition 3.2 below. However, as it obvious even from the statement of the Theorem 1.1 (compare with the measure estimate in item (ii)), if $\varepsilon\mathtt{K}^\gamma$ is not small, some statements might be empty.

(R_7) One of the main issues in singular KAM theory in analytic class is the identification of a suitable generic class of analytic potentials. We stress that the choice of the class $\mathbb{G}^n_s$ is tailored on the simple structure of natural Hamiltonian systems.

Open problems

(i) It might not be difficult to prove that Theorem 1.1 can be generalized to natural systems with Hamiltonians of the form $h(y)+\varepsilon f(y,x)$ with h Kolmogorov non–degenerate and f verifying (16) and (17) *uniformly in* $y \in \mathrm{B}$. However, in such a case, the class of perturbations f would not be generic, since, in general, one expects that the Fourier coefficients $f_k(y)$ may vanish at some points $y \in \mathrm{B}$. Selecting an analytic generic class of perturbations, to which Theorem 1.1 extends, is a non–trivial issue.

(ii) The results in this paper hold for generic potentials $f \in \mathbb{G}^n_s$, and, do not cover special cases such as, e.g., the case of f trigonometric polynomial, or other cases with special symmetries, as they arise, e.g., in Celestial Mechanics.

(iii) In view of our techniques, the logarithm appearing in Corollary 1.1 appears to be unavoidable, and one may wonder if it is possible to get rid of it.

(iv)* The argument sketched by Arnol'd, Kozlov and Neishtadt for the lower bound on the measure of the non–torus set rests on the claim that a general real analytic Hamiltonian system with no small parameters has a positive measure set free of invariant Lagrangian tori, however this is not been proved[13].

(v)* *Generic Arnol'd diffusion in analytic class:* It is natural to expect that, for $n \geqslant 3$ and for generic potentials $f \in \mathbb{G}^n_s$, almost every non–empty energy level of $\mathtt{H} = \frac{1}{2}|y|^2 + \varepsilon f(x)$ is orbit–connected, i.e., arbitrary neighborhoods of two points on such levels intersect an orbit of $\phi^t_{\mathtt{H}}$ no matter how small ε is.

2 Prerequisites

In this section we recall a few prerequisites, which are needed to discuss the 'secondary' nearly–integrable structure that appears near simple resonances.

We begin by recalling the averaging theory for nearly–integrable real analytic Hamiltonian systems as discussed in [11] and [15], especially designed for neighborhoods of simple resonances.

On one hand, apart from a finite (although arbitrarily large) number of simple resonances of order less than $\mathtt{N}$, the secular (averaged) Hamiltonians have a uniform normal form with a potential close to a shifted cosine (§ 2.1). On the other hand, the secular Hamiltonians at simple resonances of order less or equal than $\mathtt{N}$ admit a simple normal form called 'Generic Standard Form' (§ 2.2). Such Generic Standard Form Hamiltonians can, then, be put into action–angle variables that can be analytically and *uniformly* controlled thanks to the theory developed in [14] (§ 2.3).

[13] For related results in smooth category, see [29].

2.1 Averaging

Non–resonant and simply resonant sets

First, we need to introduce suitable non/simply–resonant sets, which depend upon some quantitative parameters measuring Fourier cut–offs and small divisors. In particular, we define the 'non–perturbative' set $\mathcal{R}^2$ – which is a neighborhood of double resonances up to order $\mathtt{K}$ – and prove item (ii) of Theorem 1.1.

For $k \neq 0$, denote by π_k and $\pi_k^\perp$, the standard orthogonal projections

$$\pi_k y := \big(y \cdot \tfrac{k}{|k|}\big)\, \tfrac{k}{|k|}\,, \qquad \pi_k^\perp y := y - \pi_k y\,, \tag{19}$$

and let $\mathtt{K}_{\rm o}$, $\mathtt{K}$ and α be positive numbers such that

$$\mathtt{K}_{\rm o} \geqslant 2\,, \qquad \mathtt{K} \geqslant 6\mathtt{K}_{\rm o}\,, \qquad \alpha := \sqrt{\varepsilon}\mathtt{K}^\nu\,, \qquad \nu := \tfrac{9}{2}n + 2\,. \tag{20}$$

Recall the definition of $\mathcal{G}^n_K$ in (6) and define the following real subsets of $\mathrm{B} = \{y \in \mathbb{R}^n : \ |y| < 1\}$:

$$\mathcal{R}^0 := \{y \in \mathrm{B} : \ |y \cdot k| > \tfrac{\alpha}{2}\,, \ \forall k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}\}\,, \tag{21}$$

$$\begin{cases} \mathcal{R}^{1,k} := \big\{y \in \mathrm{B} : |y \cdot k| < \alpha;\ |\pi_k^\perp y \cdot \ell| > \frac{3\alpha\mathtt{K}}{|k|}, \forall \ell \in \mathcal{G}^n_{\mathtt{K}} \setminus \mathbb{Z}k\big\}\,, \quad (k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}); \\ \mathcal{R}^1 := \bigcup_{k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}} \mathcal{R}^{1,k}\,; \end{cases} \tag{22}$$

$$\mathcal{R}^2 := \mathrm{B} \backslash (\mathcal{R}^0 \cup \mathcal{R}^1)\,. \tag{23}$$

The first key remark is that *the measure of $\mathcal{R}^2$ is proportional to ε*, as shown in the following lemma[14], which proves item (ii) of Theorem 1.1,

Lemma 2.1 *There exists a constant $c_\star = c_\star(n) > 1$ such that*

$$\mathrm{meas}\,\mathcal{R}^2 \leqslant c_\star\, \alpha^2\, \mathtt{K}^{2n} = c_\star\, \varepsilon\, \mathtt{K}^\gamma\,, \qquad \gamma := 11n + 4\,. \tag{24}$$

Proof First observe that from the definitions of $\mathcal{R}^0$, $\mathcal{R}^{1,k}$ and $\mathcal{R}^2$ in (21), (22) and (23), it follows immediately that

$$\mathcal{R}^2 \subseteq \bigcup_{k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}} \bigcup_{\substack{\ell \in \mathcal{G}^n_{\mathtt{K}} \\ \ell \notin \mathbb{Z}k}} \mathcal{R}^2_{k,\ell}\,, \tag{25}$$

with

$$\mathcal{R}^2_{k\ell} := \big\{y \in \mathrm{B} : |y \cdot k| < \alpha;\ |\pi_k^\perp y \cdot \ell| \leqslant \tfrac{3\alpha\mathtt{K}}{|k|}\big\}\,, \qquad (k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}\,,\ \ell \in \mathcal{G}^n_{\mathtt{K}} \setminus \mathbb{Z}k)\,. \tag{26}$$

Let us, then, estimate the measure of $\mathcal{R}^2_{k,\ell}$ in (26). Denote by $v \in \mathbb{R}^n$ the projection of y onto the plane generated by k and ℓ (recall that, by hypothesis, k and ℓ are not parallel); then,

$$|v \cdot k| = |y \cdot k| < \alpha\,, \qquad |\pi_k^\perp v \cdot \ell| = |\pi_k^\perp y \cdot \ell| \leqslant 3\alpha\mathtt{K}/|k|\,. \tag{27}$$

Set

$$h := \pi_k^\perp \ell = \ell - \tfrac{\ell \cdot k}{|k|^2} k\,. \tag{28}$$

Then, v decomposes in a unique way as $v = ak + bh$ for suitable $a, b \in \mathbb{R}$. By (27),

$$|a| < \frac{\alpha}{|k|^2}\,, \qquad |\pi_k^\perp v \cdot \ell| = |bh \cdot \ell| \leqslant 3\alpha\mathtt{K}/|k|\,, \tag{29}$$

[14] Compare, also, Lemma 2.5 in [15].

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

$$|h\cdot\ell| \overset{(28)}{=} \frac{|\ell|^2|k|^2-(\ell\cdot k)^2}{|k|^2} \geqslant \frac{1}{|k|^2}\,.$$

Hence,

$$|b| \leqslant 3\alpha \mathtt{K}|k|\,. \tag{30}$$

Then, write $y\in\mathcal{R}^2_{k,\ell}$ as $y=v+v^\perp$ with $v^\perp$ in the orthogonal complement of the plane generated by k and ℓ. Since $|v^\perp|\leqslant|y|<1$ and v lies in the plane spanned by k and ℓ inside a rectangle of sizes of length $2\alpha/|k|^2$ and $6\alpha\mathtt{K}|k|$ (compare (29) and (30)), we find

$$\mathrm{meas}(\mathcal{R}^2_{k,\ell}) \leqslant \tfrac{2\alpha}{|k|^2}\,(6\alpha\mathtt{K}|k|)\;2^{n-2} = 3\cdot 2^n\,\alpha^2\,\tfrac{\mathtt{K}}{|k|}\,,\qquad \forall \begin{cases} k\in\mathcal{G}^n_{\mathtt{K}_{\rm o}}\,,\\ \ell\in\mathcal{G}^n_{\mathtt{K}}\setminus\mathbb{Z}k\,.\end{cases}$$

Thus, since $\displaystyle\sum_{k\in\mathcal{G}^n_{\mathtt{K}_{\rm o}}}|k|^{-1}\leqslant c\,\mathtt{K}_{\rm o}^{n-1}$ for a suitable $c=c(n)$, (24) follows immediately taking $c_\star=18c$. ∎

Remark 2.1 (i) By the second relation in (11) it follows that $\alpha<1/\mathtt{K}^n$.
(ii) $\mathcal{R}^0$ is a non resonant set up to order $\mathtt{K}_{\rm o}$; $\mathcal{R}^{1,k}$ is a simply resonant set around $\mathcal{R}_k$ but far away from any $\mathcal{R}_\ell$ with $\ell\in\mathcal{G}^n_{\mathtt{K}}$ $(\ell\neq k)$; $\mathcal{R}^2$ is contained in a neighborhood of double resonances of order $\mathtt{K}$ (compare relation (25) below). According to the terminology in [36], $\mathcal{R}^0$ is $(\alpha/2,\mathtt{K}_{\rm o})$ completely non–resonant, while, for each $k\in\mathcal{G}^n_{\mathtt{K}_{\rm o}}$, the set $\mathcal{R}^{1,k}$ is $(2\alpha\mathtt{K}/|k|)$–non resonant modulo $\mathbb{Z}k$ up to order $\mathtt{K}$.
(iii) From the definition of $\mathcal{R}^2$ in (23) it follows trivially that $\{\mathcal{R}^i\}$ is a covering of B.
(iv) Having two different Fourier cut–offs $\mathtt{K}_{\rm o}$ and $\mathtt{K}$ is necessary in order to obtain high order 'cosine–like' normal forms as described in point (iii) of the averaging Theorem 2.1 below; compare also [11].

Notations
Given $m\geqslant 1$, $D\subseteq\mathbb{R}^m$, and $r>0$, let us denote D_r the complex neighborhood of D given by

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

For $s>0$, let $\mathbb{T}^m_s$ denote the complex neighborhood of width $2s$ of $\mathbb{T}^m$ given by

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

We shall also use the notation $\mathrm{Re}\,(V_r)$ to denote the *real* r–neighbourhood of $V\subseteq\mathbb{R}^n$, namely,

$$\mathrm{Re}\,(V_r) := V_r\cap\mathbb{R}^n = \bigcup_{z\in V}\ \{y\in\mathbb{R}^n:\ |y-z|<r\}\,. \tag{31}$$

Given $D\subseteq\mathbb{R}^n$ and a function f defined, respectively, on D_r, $\mathbb{T}^m_s$, $D_r\times\mathbb{T}^m_s$, we denote its sup norm, respectively, by

$$|f|_{D,r} := \sup_{y\in D_r}|f(y)|\,,\quad |f|_s := \sup_{x\in\mathbb{T}^m_s}|f(x)|\,,\quad |f|_{D,r,s} := \sup_{(y,x)\in D_r\times\mathbb{T}^m_s}|f(x,y)|\,.$$

Analytcity parameters To formulate properly next (normal form) theorem, we need to introduce a few 'analyticity parameters' related to the analyticity width s and the numbers α, $\mathtt{K}_{\rm o}$ and $\mathtt{K}$ in (20). We define:

$$\begin{aligned} & r_{\rm o} := \tfrac{\alpha}{16\mathtt{K}_{\rm o}}\,, \quad r'_{\rm o} := \tfrac{r_{\rm o}}{2}\,, \quad s_{\rm o} := s\big(1-\tfrac{1}{\mathtt{K}_{\rm o}}\big)\,, \quad s'_{\rm o} := s_{\rm o}\big(1-\tfrac{1}{\mathtt{K}_{\rm o}}\big)\,, \\ & s_\star := s\big(1-\tfrac{1}{\mathtt{K}}\big)\,, \quad s'_\star := s_\star\big(1-\tfrac{1}{\mathtt{K}}\big)\,, \quad s'_k := |k|_1 s'_\star\,, \\ & r_k := \tfrac{\alpha}{|k|}\,, \quad r'_k := \tfrac{r_k}{2}\,, \\ & \tilde r_k := \tfrac{r_k}{c_1|k|}\,, \qquad \tilde s_k := \tfrac{s}{c_1|k|^{n-1}}\,, \quad \text{where} \quad c_1 := 5n(n-1)^{\frac{n-1}{2}}\,. \end{aligned} \tag{32}$$

We also need the following consequence of Bezout's Lemma, which will allow to define the effective 'resonant angle' near simple resonances:

Lemma 2.2 *For any $k \in \mathcal{G}^n$ there exists a matrix $\hat{\rm A} \in \mathbb{Z}^{(n-1)\times n}$ such that*[15]

$$\begin{aligned} & {\rm A} := \begin{pmatrix} k \\ \hat{\rm A} \end{pmatrix} = \begin{pmatrix} k_1 \cdots k_n \\ \hat{\rm A} \end{pmatrix} \in \ {\rm SL}(n,\mathbb{Z})\,, \\ & |\hat{\rm A}|_\infty \leqslant |k|_\infty\,, \quad |{\rm A}|_\infty = |k|_\infty\,, \quad |{\rm A}^{-1}|_\infty \leqslant (n-1)^{\frac{n-1}{2}}|k|_\infty^{n-1}\,. \end{aligned} \tag{33}$$

Proof From Bézout's lemma it follows easily that[16]:
Given $k \in \mathbb{Z}^n$, $k \neq 0$ there exists a matrix ${\rm A} = ({\rm A}_{ij})_{1\leqslant i,j\leqslant n}$ with integer entries such that $A_{nj} = k_j$ $\forall$ $1 \leqslant j \leqslant n$, $\det {\rm A} = \gcd(k_1, ..., k_1)$, and $|{\rm A}|_\infty = |k|_\infty$.
Since $k \in \mathcal{G}^n$, it is $\gcd(k_1, ..., k_1) = 1$ and, therefore, $\det {\rm A} = 1$.
The first two relations in (33) are consequence of the above statement.
Observing that for any $m \times m$ matrix M, one has $|\det {\rm M}| \leqslant m^{m/2}|{\rm M}|_\infty^m$, the bound on $|{\rm A}^{-1}|_\infty$ follows from D'Alembert expansion of determinants. ∎

The following normal form result – proven in[17] [15] – holds:

Theorem 2.1 (Normal Form Theorem) *Fix $n \geqslant 2$, $s > 0$. Let $\mathtt{H}$ be as in (1) with $f \in \mathbb{B}^n_s$ satisfying (16) with $\mathtt{N}$ as in (15); let (20)÷(22) and (32) hold. For $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, let ${\rm A}$ be the matrix in Lemma 2.2 and define the following real sets:*

$$\widetilde{\mathcal{R}}^0 := \ {\rm Re}\,(\mathcal{R}^0_{r'_{\rm o}/2})\,, \quad \widetilde{\mathcal{R}}^{1,k} := \ {\rm Re}\,(\mathcal{R}^{1,k}_{r'_k/2})\,, \quad \mathscr{D}^k := {\rm A}^{-T}\widetilde{\mathcal{R}}^{1,k}\,, \qquad (k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}})\,. \tag{34}$$

Then, there exists a constant $\mathbf{c_o} = \mathbf{c_o}(n, s, \delta) \geqslant \mathtt{N}$ such that if $\mathtt{K}_{\rm o} \geqslant \mathbf{c_o}$, there exist real analytic symplectic maps

$$\Psi_{\rm o} : \mathcal{R}^0_{r'_{\rm o}} \times \mathbb{T}^n_{s'_{\rm o}} \to \mathcal{R}^0_{r_{\rm o}} \times \mathbb{T}^n_{s_{\rm o}}\,, \qquad \Psi^k : \mathscr{D}^k_{\tilde r_k} \times \mathbb{T}^n_{\tilde s_k} \to \mathcal{R}^{1,k}_{r_k} \times \mathbb{T}^n_{s_\star} \tag{35}$$

having the following properties.
(i) *In the symplectic variables $(y, x) \in \mathcal{R}^0_{r_{\rm o}} \times \mathbb{T}^n_{s_{\rm o}}$, $\mathtt{H}$ takes the form:*

$$\mathtt{H}_{\rm o}(y, x) := \big(\mathtt{H} \circ \Psi_{\rm o}\big)(y, x) = \frac{|y|^2}{2} + \varepsilon\big(g^{\rm o}(y) + f^{\rm o}(y, x)\big)\,, \qquad \langle f^{\rm o}\rangle = 0\,,$$

[15] $|M|_\infty$, with M matrix (or vector), denotes the maximum norm $\max_{ij}|M_{ij}|$ (or $\max_i |M_i|$).
[16] See Lemma A.1 in [11] for a proof.
[17] See Theorem 2.1 and the Covering Lemma 2.3 in [15].

with $g^{\rm o}$ and $f^{\rm o}$ satisfying

$$|g^{\rm o}|_{r'_{\rm o}} \leqslant \vartheta_{\rm o} := \frac{1}{\mathtt{K}^{6n+1}}\,, \qquad |f^{\rm o}|_{r'_{\rm o},s'_{\rm o}} \leqslant e^{-\mathtt{K}_{\rm o}s/3}\,. \tag{36}$$

(ii) *Let $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$. In the symplectic variables $(\mathtt{y},\mathtt{x}) = \big(\mathtt{y},(\mathtt{x}_1,\hat{\mathtt{x}})\big) \in \mathscr{D}^k_{\tilde r_k} \times \mathbb{T}^n_{\tilde s_k}$, $\mathtt{H}$ takes the form:*

$$\mathcal{H}_k(\mathtt{y},\mathtt{x}) := \mathtt{H}\circ\Psi^k(\mathtt{y},\mathtt{x}) = \overline{\mathtt{H}}_k(\mathtt{y},\mathtt{x}_1) + \varepsilon \bar f^k(\mathtt{y},\mathtt{x})\,, \quad (\mathtt{y},\mathtt{x}) \in \mathscr{D}^k_{\tilde r_k} \times \mathbb{T}^n_{\tilde s_k}\,, \tag{37}$$

where

$$\overline{\mathtt{H}}_k(\mathtt{y},\mathtt{x}_1) := \frac{1}{2}|\mathrm{A}^T\mathtt{y}|^2 + \varepsilon \mathtt{g}^k_{\rm o}(\mathtt{y}) + \varepsilon \mathtt{g}^k(\mathtt{y},\mathtt{x}_1) \tag{38}$$

is real analytic in $\mathtt{y} \in \mathscr{D}^k_{\tilde r_k}$ and $\mathtt{x}_1 \in \mathbb{T}_{s'_k}$. In particular $\mathtt{g}^k(y,\cdot) \in \mathbb{B}^1_{s'_k}$ for every $y \in \mathscr{D}^k_{\tilde r_k}$. Furthermore, the following estimates hold:

$$|\mathtt{g}^k_{\rm o}|_{\tilde r_k} \leqslant \vartheta_{\rm o}\,, \qquad |\mathtt{g}^k - \pi_{\mathbb{Z}k} f|_{\tilde r_k, s'_k} \leqslant \vartheta_{\rm o}\,, \qquad |\bar f^k|_{\tilde r_k,\tilde s_k} \leqslant e^{-\mathtt{K}s/3}\,. \tag{39}$$

(iii) *If $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$ satisfies $|k|_1 \geqslant \mathtt{N}$, then there exists $\theta_k \in [0,2\pi)$ such that*

$$\mathcal{H}_k = \frac{1}{2}|\mathrm{A}^T\mathtt{y}|^2 + \varepsilon \mathtt{g}^k_{\rm o}(\mathtt{y}) + 2|f_k|\varepsilon\,\big[\cos(\mathtt{x}_1+\theta_k) + F^k_\star(\mathtt{x}_1) + \mathtt{g}^k_\star(\mathtt{y},\mathtt{x}_1) + \mathtt{f}^k_\star(\mathtt{y},\mathtt{x})\big]\,,$$

where

$$F^k_\star(\theta) := \frac{1}{2|f_k|}\sum_{|j|\geqslant 2} f_{jk}e^{\mathrm{i}j\theta} \in \mathbb{B}^1_1\,, \qquad |F^k_\star|_1 \leqslant 2^{-40}\,.$$

Moreover, $\mathtt{g}^k_\star(y,\cdot) \in \mathbb{B}^1_1$ (for every $y \in \mathscr{D}^k_{\tilde r_k}$), $\pi_{\mathbb{Z}k}\mathtt{f}^k_\star = 0$, and one has

$$|\mathtt{g}^k_\star|_{\tilde r_k,1} \leqslant \tfrac{1}{\mathtt{K}^{5n}}\,, \qquad |\mathtt{f}^k_\star|_{\tilde r_k,\tilde s_k} \leqslant e^{-\mathtt{K}s/7}\,.$$

(iv) *Finally, the following 'coverings' holds:*

$$\Psi_{\rm o}\big(\widetilde{\mathcal{R}}^0 \times \mathbb{T}^n\big) \supseteq \mathcal{R}^0 \times \mathbb{T}^n\,, \qquad \Psi^k\big(\mathscr{D}^k \times \mathbb{T}^n\big) \supseteq \mathcal{R}^{1,k} \times \mathbb{T}^n\,. \tag{40}$$

Remark 2.2 (i) Beware that, while $\Psi_{\rm o}$ is a map close to the identity, Ψ^k is not, as it is the composition of a linear transformation[18] with a near–to–identity map.

(ii) The larger covering in (34) is introduced so that (40) holds: Such a property will be essential in covering also boundary regions by KAM tori without leaving out (as it happens in standard KAM theory) regions of size of order $\sqrt{\varepsilon}$, a fact that, for our purposes, would be clearly not acceptable.

(iii) Point (iii) in Theorem 2.1 shows that the secular Hamiltonian $\overline{\mathtt{H}}_k$ (obtained disregarding the exponentially small perturbation $\mathtt{f}^k_\star$) has a potential, which is $O(1/\mathtt{K}^{5n})$–perturbation of the 'cosine–like function'

$$\cos(\mathtt{x}_1+\theta_k) + F^k_\star(\mathtt{x}_1)\,, \qquad \text{where} \quad |F^k_\star|_1 \leqslant 2^{-40}\,.$$

This means that for $|k|_1 \geqslant \mathtt{N}$, *the secular Hamiltonians at simple resonances all look the same*, allowing, in particular, for a uniform analysis in terms of action–angle variables (compare § 2.3 below).

Notice also that the perturbation $\mathtt{f}^k_\star$, which is bounded by $e^{-\mathtt{K}s/7}$ has a factor $|f_k|$ in front of it and that such a factor, in turn, may be exponentially small (since $|f_k| \sim e^{-|k|_1 s}$ for large $|k|_1$).

(iv) For later use we observe that[19]

$$\mathtt{K}_{\rm o} \geqslant \mathtt{N} \geqslant 2\mathtt{c}_s\,, \qquad \text{where} \qquad \mathtt{c}_s := \max\{1, 1/s\}\,. \tag{41}$$

[18] Namely, the symplectic transformation, the generating function of which is given by $\mathtt{y}\cdot \mathrm{A}x$, and which maps the resonant combination $k\cdot x$ to the 'resonant' angle $\mathtt{x}_1$.

[19] If $s \geqslant 1$ then $\mathtt{N} \geqslant 2 \geqslant 2/s$, while if $s < 1$ then the logarithm in (15) is larger than one, so that $\mathtt{N} \geqslant 2/s$ also in this case.

2.2 Generic Standard Form at simple resonances

It turns out that the secular Hamiltonians $\bar{\mathtt{H}}_k$ in (37)–(38) in the Normal Form Theorem 2.1 for $k \in \mathcal{G}^n_{\mathtt{K}_o}$, have a common uniform analytic structure: They can be put into a standard form, which has uniform (in $k \in \mathcal{G}^n_{\mathtt{K}_o}$) analytic characteristics. The precise formulation of this fact is the main theorem in [15], whose statement needs some preparation.

Definition 2.1 (1D Hamiltonians in standard form) *Let $\hat{D} \subseteq \mathbb{R}^{n-1}$ be a bounded domain, $\mathtt{R} > 0$ and $D := (-\mathtt{R}, \mathtt{R}) \times \hat{D}$. We say that a real analytic Hamiltonian $\mathtt{H}_\flat$ is in Generic Standard Form (in short, 'standard form') with respect to the symplectic variables $(p_1, q_1) \in (-\mathtt{R}, \mathtt{R}) \times \mathbb{T}$ and 'external actions' $\hat{p} = (p_2, ..., p_n) \in \hat{D}$, if $\mathtt{H}_\flat$ has the form*

$$\mathtt{H}_\flat(p, q_1) = \big(1 + \nu(p, q_1)\big)p_1^2 + \mathtt{G}(\hat{p}, q_1)\,, \tag{42}$$

where $p = (p_1, \hat{p}) = (p_1, p_2, ..., p_n)$, and the following specifications hold.

- *ν and $\mathtt{G}$ are real analytic functions defined on, respectively, $D_{\mathtt{r}} \times \mathbb{T}_{\mathtt{s}}$ and $\hat{D}_{\mathtt{r}} \times \mathbb{T}_{\mathtt{s}}$ for some $0 < \mathtt{r} \leqslant \mathtt{R}$ and $\mathtt{s} > 0$;*
- *$\mathtt{G}$ has zero–average and there exists a zero–average function $\bar{\mathtt{G}}$ (the 'reference potential') depending only on q_1 such that, for some $\beta > 0$, $\bar{\mathtt{G}}$ is β–Morse*[20]*;*
- *the following estimates hold:*

$$\begin{cases} \sup\limits_{\mathbb{T}^1_{\mathtt{s}}} |\bar{\mathtt{G}}| \leqslant \epsilon\,, \\ \sup\limits_{\hat{D}_{\mathtt{r}} \times \mathbb{T}^1_{\mathtt{s}}} |\mathtt{G} - \bar{\mathtt{G}}| \leqslant \epsilon\mu\,, \quad \text{for some} \quad 0 < \epsilon \leqslant \mathtt{r}^2/2^{16}\,, \quad 0 \leqslant \mu < 1\,, \\ \sup\limits_{D_{\mathtt{r}} \times \mathbb{T}^1_{\mathtt{s}}} |\nu| \leqslant \mu\,. \end{cases} \tag{43}$$

We shall call $(\hat{D}, \mathtt{R}, \mathtt{r}, \mathtt{s}, \beta, \epsilon, \mu)$ *the 'analyticity characteristics' of $\mathtt{H}_\flat$ with respect to the reference potential $\bar{\mathtt{G}}$.*

Remark 2.3 (i) A Hamiltonian in standard form $\mathtt{H}_\flat$ has the analytic features of its reference natural Hamiltonian

$$\bar{\mathtt{H}}_\flat := p_1^2 + \bar{\mathtt{G}}(q_1)\,.$$

In particular, for μ small with respect to $1/\kappa$, $\mathtt{H}_\flat$ has the same finite (because of analyticity) number of equilibria (which lie on the q_1 axis) of $\bar{\mathtt{G}}$ and in the same relative order, which is also preserved by the corresponding critical energies; compare Lemma 2.4 below.

(ii) If $\mathtt{H}_\flat$ is in standard form, then β and ϵ satisfy the relation[21] $\epsilon/\beta \geqslant 1/2$. Furthermore, one can always fix a number $\kappa \geqslant 4$ so that:

$$1/\kappa \leqslant \mathtt{s} \leqslant 1\,, \qquad 1 \leqslant \mathtt{R}/\mathtt{r} \leqslant \kappa\,, \qquad 1/2 \leqslant \epsilon/\beta \leqslant \kappa\,. \tag{44}$$

Such a parameter κ rules the main scaling properties of these Hamiltonians.

[20] Recall Definition 1.2.

[21] By (43), $\beta \leqslant |\bar{\mathtt{G}}(\theta_i) - \bar{\mathtt{G}}(\theta_i)| \leqslant 2\max_{\mathbb{T}} |\bar{\mathtt{G}}| \leqslant 2\epsilon$.

(iii) Hamiltonians in standard form are particularly suited for the analytic theory of action–angle variables (in neighborhoods of separatrices) as developed in [14], where the notion of Generic Standard Form has been introduced. Such action–angle variables will be reviewed in § 2.3 below.

(iv) The smallness of the 'adimensional ratio' $\epsilon/\mathtt{r}^2$ in (43) is needed in the analytic theory of action-angle variables for Hamiltonians in standard form developed in [14], however the factor $1/2^{16}$ is rather arbitrary and not optimal.

Notation If w is a vector with n or $2n$ components, $\hat{w} = (w)\hat{}$ denotes the last $(n-1)$ components; if w is vector with $2n$ components, $\check{w} = (w)\check{}$ denotes the first $n+1$ components. Explicitly:

$$w = (y,x) = \big((y_1, ..., y_n), (x_1, ...x_n)\big) \quad \Longrightarrow \quad \begin{cases} \hat{w} = (w)\hat{} = (x_2, ..., x_n) = \hat{x}\,, \\ \hat{y} = (y)\hat{} = (y_2, ..., y_n)\,, \\ \check{w} = (w)\check{} = (y, x_1)\,, \\ w = (\check{w}, \hat{w})\,. \end{cases} \tag{45}$$

Next, we introduce a special simple group of symplectic transformations, which will appear in Theorem 2.2 below.

Definition 2.2 *Given a domain* $\hat{\mathrm{D}} \subseteq \mathbb{R}^{n-1}$*, we denote by* $\mathfrak{G}_\dagger$ *the abelian group of symplectic diffeomorphisms* Ψ_{g} *of* $(\mathbb{R} \times \hat{\mathrm{D}}) \times \mathbb{R}^n$ *given by*

$$(p,q) \in (\mathbb{R} \times \hat{\mathrm{D}}) \times \mathbb{R}^n \overset{\Psi_{\mathsf{g}}}{\mapsto} (P,Q) = (p_1 + \mathsf{g}(\hat{p}), \hat{p}, q_1, \hat{q} - q_1 \partial_{\hat{p}} \mathsf{g}(\hat{p})) \in (\mathbb{R} \times \hat{\mathrm{D}}) \times \mathbb{R}^n\,, \tag{46}$$

with $\mathsf{g} : \hat{\mathrm{D}} \to \mathbb{R}$ *smooth.*

Remark 2.4 The group properties of $\mathfrak{G}_\dagger$ are trivial:

$$\mathrm{id}_{\mathfrak{G}_\dagger} = \Psi_0\,, \qquad \Psi_{\mathsf{g}}^{-1} = \Psi_{-\mathsf{g}}\,, \qquad \Psi_{\mathsf{g}} \circ \Psi_{\mathsf{g}'} = \Psi_{\mathsf{g}+\mathsf{g}'}\,. \tag{47}$$

Notice that, unless $\partial_{\hat{p}}\mathsf{g} \in \mathbb{Z}^{n-1}$, maps $\Psi_{\mathsf{g}} \in \mathfrak{G}_\dagger$ *do not induce well defined maps*[22]

$$q \in \mathbb{T}^n \mapsto (q_1, \hat{q} - q_1 \partial_{\hat{p}} \mathsf{g}(\hat{p})) \in \mathbb{T}^n\,,$$

a fact that will create a problem in applying the theory of this and next section to the normalized Hamiltonians $\mathcal{H}_k$ of Theorem 2.1; compare Remark 2.6–(ii) below.

Let us now spell out all the *assumptions and definitions, which, from now, will be part of the hypotheses of all statements regarding the natural system with Hamiltonian* $\mathtt{H}$ *as in* (1).

Assumptions 2.1 *Fix* $n \geqslant 2$*,* $s > 0$*, and let* $\mathtt{H}$ *be as in* (1) *with* $f \in \mathbb{G}^n_s$ *(Definition 1.1) satisfying* (16) *and* (17) *for some* $0 < \delta \leqslant 1$ *and* $\beta > 0$ *with*[23] $\mathtt{N}$ *as in* (15).

[22] In general, given $A \in \mathrm{SL}(n, \mathbb{Z})$ and a 2π–multi–periodic function $f : \mathbb{R}^n \to \mathbb{R}^n$, we identify the $\mathbb{R}^n$–map $x \in \mathbb{R}^n \to f(x) = Ax + g(x) \in \mathbb{R}^n$ with the $\mathbb{T}^n$–map given by $\theta \in \mathbb{T}^n \to F(\theta) = \pi_{\mathbb{T}^n}\big(Ax + f(x)\big) \in \mathbb{T}^n$ where $\theta = x + 2\pi\mathbb{Z}^n$ and $x \to \pi_{\mathbb{T}^n}(x) = x + 2\pi\mathbb{Z}^n$ is the projection of $\mathbb{R}^n$ onto $\mathbb{T}^n$.

[23] By Lemma 1.1 such δ and β always exist.

Definition 2.3 *Given* $\mathtt{H}$ *as in Assumption 2.1, we define the following sets and parameters.*
• *Let* $\mathtt{K_o}$, $\mathtt{K}$ *and* α *be as in (20); let* $\mathcal{R}^i$*'s be the domains defined in (21)÷(22); let the definitions in (32) hold ('analiticity parameters').*
• *For* $k \in \mathcal{G}^n_{\mathtt{K_o}}$*, let* A *be the matrix in Lemma 2.2. Let* $\widetilde{\mathcal{R}}^0$*,* $\widetilde{\mathcal{R}}^{1,k}$ *and* $\mathscr{D}^k$ *be the real domains defined in (34).*
• *Define the following parameters*[24]*:*

$$\begin{aligned}
&\mathtt{R} = \alpha/|k|^2 = \sqrt{\varepsilon}\mathtt{K}^\nu/|k|^2\,, \quad c_{_2} = 4\,n^{\frac{3}{2}}c_{_1}\,, \quad \mathtt{r} = \mathtt{R}/c_{_2}\,, \quad \varepsilon_k = \tfrac{2\varepsilon}{|k|^2}\,,\\
&\hat{D} = \big\{\hat{I} \in \mathbb{R}^{n-1} : \ |\pi_k^\perp \hat{\mathrm{A}}^T \hat{I}| < 1\,, \ \min_{\substack{\ell \in \mathcal{G}^n_{\mathtt{K}}\\ \ell \notin \mathbb{Z}k}} \big|\big(\pi_k^\perp \hat{\mathrm{A}}^T \hat{I}\big)\cdot \ell\big| \geqslant \tfrac{3\alpha\mathtt{K}}{|k|}\big\}\,, \quad D = (-\mathtt{R},\mathtt{R}) \times \hat{D}\,,\\
&\upbeta = \begin{cases} \varepsilon_k \beta, & \text{if } |k|_{_1} < \mathtt{N}\\ \varepsilon_k |f_k|, & \text{if } |k|_{_1} \geqslant \mathtt{N}\end{cases}, \qquad \chi_{_k} = \begin{cases} 1\,, & \text{if } |k|_{_1} < \mathtt{N}\\ |f_k|\,, & \text{if } |k|_{_1} \geqslant \mathtt{N}\end{cases}, \quad \epsilon = \mathtt{c}_s \varepsilon_k\, \chi_{_k}\,,\\
&\mathtt{s} = \begin{cases} \min\{\frac{s}{2}, 1\}\,, & \text{if } |k|_{_1} < \mathtt{N}\\ 1\,, & \text{if } |k|_{_1} \geqslant \mathtt{N}\end{cases}, \quad \check{\mathtt{s}} = \begin{cases} s'_k\,, & \text{if } |k|_{_1} < \mathtt{N}\,,\\ 1\,, & \text{if } |k|_{_1} \geqslant \mathtt{N}\end{cases}, \quad \upmu = \frac{1}{\mathtt{K}^{5n}}\,.
\end{aligned} \tag{48}$$

Remark 2.5 (i) Since $|f_k| \leqslant 1$ one has:

$$|\chi_{_k}| \leqslant 1\,. \tag{49}$$

Furthermore, by the definitions in (48) and (20), by (49) and (41), one has

$$\sqrt{\epsilon} < \mathtt{c}_s \mathtt{R}/\mathtt{K}^{\nu-1} < \mathtt{R}/\mathtt{K}^{\frac{9}{2}n}\,. \tag{50}$$

(ii) Since $(1 - \frac{1}{\mathtt{K}})^{-2} < 2$, by definition of s'_k in (32), one has

$$\mathtt{s} \leqslant 2\check{\mathtt{s}}\,. \tag{51}$$

We can, now, state the main result of[25] [15]:

Theorem 2.2 (Generic Standard Form at simple resonances)
Let Assumptions 2.1 and Definitions 2.3 hold, let $\mathbf{c_o}$ *be the constant defined in Theorem 2.1, and assume that* $\mathtt{K_o} \geqslant \max\{c_{_2}, \mathbf{c_o}\}$*. Then, for all* $k \in \mathcal{G}^n_{\mathtt{K_o}}$*, the following holds.*
(i) *There exists a real analytic symplectic transformation*

$$\Phi_\star : (\mathtt{p},\mathtt{q}) \ \in D \times \mathbb{R}^n \to (\mathtt{y},\mathtt{x}) = \Phi_\star(\mathtt{p},\mathtt{q}) \in \mathbb{R}^{2n}\,, \tag{52}$$

such that: $\Phi_\star$ *fixes* $\hat{\mathtt{p}}$ *and*[26] $\mathtt{q}_1$*; for every* $\hat{\mathtt{p}} \in \hat{D}$ *the map* $(\mathtt{p}_1,\mathtt{q}_1) \mapsto (\mathtt{y}_1,\mathtt{x}_1)$ *is symplectic; the* $(n+1)$*–dimensional map*[27] $\check{\Phi}_\star$ *depends only on the first* $n+1$ *coordinates* $(\mathtt{p},\mathtt{q}_1)$*, is* 2π*–periodic in* $\mathtt{q}_1$ *and, if* $\mathscr{D}^k = \mathrm{A}^{-T}\mathcal{R}^{1,k}$ *and* $\overline{\mathtt{H}}_k$ *are as in Theorem 2.1, one has*[28]

$$\begin{aligned}
&\check{\Phi}_\star : D_{\mathtt{r}} \times \mathbb{T}_{\check{\mathtt{s}}} \ \to \mathscr{D}^k_{\tilde{r}_k} \times \mathbb{T}_{\check{\mathtt{s}}}\,,\\
&\overline{\mathtt{H}}_k \circ \check{\Phi}_\star(\mathtt{p},\mathtt{q}) =: \tfrac{|k|^2}{2}\big(\mathtt{H}_k(\mathtt{p},\mathtt{q}_1) + \hat{h}_{_k}(\hat{\mathtt{p}})\big)\,,\\
&\sup_{\hat{\mathtt{p}} \in \hat{D}_{2\mathtt{r}}} \big|\hat{h}_{_k}(\hat{\mathtt{p}}) - \hat{\mathcal{Q}}_k(\hat{\mathtt{p}})\big| \leqslant \tfrac{12}{|k|^2}\varepsilon\upmu\,, \qquad \hat{\mathcal{Q}}_k(\hat{\mathtt{p}}) := \tfrac{1}{|k|^2}|\,\pi_k^\perp \hat{\mathrm{A}}^T \hat{\mathtt{p}}|^2\,.
\end{aligned} \tag{53}$$

[24]Here and in what follows we shall not always indicate explicitly the dependence upon k. Recall the definitions of $c_{_1}$, $\hat{\mathrm{A}}$ and $\mathtt{c}_s$ in, respectively, (32), Lemma 2.2 and (41).
[25]Compare Theorem 3.1 in [15].
[26]I.e., in (52) it is $\mathtt{y} = \hat{\mathtt{p}}, \mathtt{x}_1 = \mathtt{q}_1$.
[27]Recall the notation in (45).
[28] r_k, $\tilde{r}_k$ and s'_k are defined in (32).

(ii) $\mathtt{H}_k$ *in* (53) *is in Generic Universal Form according to Definition 2.1:*

$$\mathtt{H}_k(\mathtt{p},\mathtt{q}_1)=\big(1+\nu_{_k}(\mathtt{p},\mathtt{q}_1)\big)\,\mathtt{p}_1^2+\mathtt{G}_{_k}(\hat{\mathtt{p}},\mathtt{q}_1)\,, \tag{54}$$

having reference potential

$$\bar{\mathtt{G}}=\bar{\mathtt{G}}_{_k}:=\varepsilon_k\,\pi_{\mathbb{Z}k}f\,, \tag{55}$$

analyticity characteristics given in (48)*, and* κ *verifying* (44) *with*[29]

$$\kappa=\kappa(n,s,\beta):=\max\big\{c_{_2}\,,4\mathtt{c}_s\,,\mathtt{c}_s/\beta\big\}\,. \tag{56}$$

(iii) *The map* $\Phi_\star$ *is obtained as composition of three symplectic maps:*

$$\Phi_\star=\Phi_{_1}\circ\Phi_{_2}\circ\Phi_{_3}\,, \tag{57}$$

where[30]*:*

- $\Phi_{_1}:=\Psi_{\mathtt{g}_{_1}}\in\mathfrak{G}_\dagger$ *with* $\mathtt{g}_{_1}(\hat{\mathtt{p}}):=-\frac{1}{|k|^2}(\hat{\mathrm{A}}k)\cdot\hat{\mathtt{p}}$*;*
- $\Phi_{_2}(\mathtt{p},\mathtt{q})=(\mathtt{p}_1+\eta_{_2},\hat{\mathtt{p}},\mathtt{q}_1,\hat{\mathtt{q}}+\chi_{_2})$ *for suitable real analytic functions* $\eta_{_2}=\eta_{_2}(\hat{\mathtt{p}},\mathtt{q}_1)$ *and* $\chi_{_2}=\chi_{_2}(\hat{\mathtt{p}},\mathtt{q}_1)$ *satisfying*
$$|\eta_{_2}|_{4\mathtt{r},\check{\mathtt{s}}}<\tfrac{\varepsilon_k\chi_k}{\mathtt{r}}\,\mu\,,\qquad|\chi_{_2}|_{2\mathtt{r},\check{\mathtt{s}}}<\tfrac{4\varepsilon_k\chi_k}{\mathtt{r}^2}\,\mu\,; \tag{58}$$
- $\Phi_{_3}:=\Psi_{\mathtt{g}_{_3}}\in\mathfrak{G}_\dagger$ *for a suitable real analytic function* $\mathtt{g}_{_3}(\hat{\mathtt{p}})$ *satisfying*
$$|\mathtt{g}_{_3}|_{4\mathtt{r}}<\tfrac{\varepsilon_k\chi_k}{\mathtt{r}}\,\mu\,. \tag{59}$$

Remark 2.6 (i) The main point of the above theorem is item (ii), which shows that the 'simply–resonant Hamiltonians' $\mathtt{H}_k$ in (53) are in *uniform* Generic Standard Form. The word 'uniform' refers to the fact that the parameter κ (defined in (56) and satisfying (44)) – which rules the scaling properties of the normalized Hamiltonians $\mathtt{H}_k$ – *does not depend upon* k, allowing, e.g., for a uniform (in $k\in\mathcal{G}^n_{\mathtt{K}_{\mathtt{o}}}$) treatment of action–angle variables (compare next Section 2.3).

(ii) There is, however, a drawback in the construction of the above normal forms, namely, that the maps $\Phi_{_1}$ and $\Phi_{_3}$ appearing in the definition of $\Phi_\star$ (item (iii) in the above theorem), *do not induce well defined maps on* $\mathbb{T}^n$; compare Remark 2.4. Therefore, a non trivial homotopy issue will have to be faced in considering the global secondary nearly–integrable structure of the system near simple resonances. On the other hand, the map $\Phi_{_2}$ is well defined also on $\mathbb{T}^n$. This matter will be discussed in details in Section 3.

The following remark explains the individual purpose of the three symplectic transformations $\Phi_{_j}$ whose composition forms $\Phi_\star$.

Remark 2.7 (i) The map $\Phi_{_1}$ in the definition of $\Phi_\star$ is a linear map that has the purpose of block–diagonalize the quadratic part $|\mathrm{A}^T\mathtt{y}|^2$ appearing in (38), so as *to obtain a kinetic part which is the sum of a quadratic part in* p_1 *and a quadratic* $(n-1)$*–dimensional part in* $\hat{p}$. Indeed, rewriting $\Phi_{_1}$ as

$$(\mathtt{y},\mathtt{x})=\Phi_{_1}(p,q):=\big(\mathrm{U}p,\mathrm{U}^{-T}q\big)\,,\qquad\text{where}\qquad\mathrm{U}:=\begin{pmatrix}1&-\frac{1}{|k|^2}(\hat{\mathrm{A}}k)^T\\0&\mathrm{id}_{_{n-1}}\end{pmatrix}\,, \tag{60}$$

[29] $\mathtt{c}_s$ is defined in (41).
[30] Recall Definition 3.1.

and observing that

$$\mathrm{A}^T \mathtt{y} = \mathrm{A}^T \mathrm{U} p = p_1 k + \pi_k^\perp \hat{\mathrm{A}}^T \hat{p}\,,$$

one sees that

$$|\mathrm{A}^T \mathrm{U} p|^2 = |k|^2 p_1^2 + |\,\pi_k^\perp \hat{\mathrm{A}}^T \hat{p}|^2 = |k|^2 (p_1^2 + \hat{\mathcal{Q}}_k(\hat{p}))\,, \tag{61}$$

$\hat{\mathcal{Q}}_k$ being the positive definite quadratic form in $\hat{\mathtt{p}} = \hat{p}$ defined in (53).

Furthermore, $y = (\mathrm{A}^T\mathrm{U})p$ if and only if $y \cdot k = p_1|k|^2$ and $\pi_k^\perp y = \pi_k^\perp \mathrm{A}^T \hat{p}$, which, recalling the definition of $\widetilde{\mathcal{R}}^{1,k}$ in (22), shows that

$$\mathrm{A}^T \mathrm{U} D = \widetilde{\mathcal{R}}^{1,k} \qquad \Longrightarrow \qquad \mathrm{meas}\, D = \mathrm{meas}\, \widetilde{\mathcal{R}}^{1,k}\,. \tag{62}$$

Notice also that, from (48), the definitions of Φ_1 and U, and the definition of $\mathscr{D}^k$ in Theorem 2.1–(i), it follows that

$$\check{\Phi}_1(D \times \mathbb{T}) = \mathrm{U} D \times \mathbb{T} = \mathscr{D}^k \times \mathbb{T}\,. \tag{63}$$

Incidentally, observe that from (33) it follows that the norms of U and its inverse satisfy the bounds[31]

$$|\mathrm{U}|, |\mathrm{U}^{-1}| \leqslant n\sqrt{n}\,. \tag{64}$$

(ii) The second map Φ_2 is a near–to–identity symplectic (globally well–defined) transformation, which is introduced so as to transform $\mathtt{H}_k$ into a Hamiltonian with a potential *independent of* p_1.

(iii) Φ_3 is a near–to–identity symplectic map, which sets *all critical points on the line* $p_1 = 0$.

2.3 Action–angle variables for 1D standard Hamiltonians

In this subsection we review the general theory of action–angle variables for Hamiltonian systems in standard form as developed in [14], where complete proofs may be found.

This subsection is independent from the previous ones; in particular the analytic characteristics $\hat{D}$, $\mathtt{R}$, $\mathtt{r}$, etc., are arbitrary (and do not refer to the definitions given in (48) in the specific case of the secular Hamiltonians $\overline{\mathtt{H}}_k$).

Topology of the phase space of 1D Hamiltonians in standard form

We begin by describing the topological structure of the $\hat{p}$–dependent phase space of a givern Hamiltonian $(p_1, q_1) \mapsto \mathtt{H}_\flat(p_1, \hat{p}, q_1)$ in generic standard form according to Definition 2.1.

For a fixed $\hat{p} \in \hat{D}$, we take as phase space of $\mathtt{H}_\flat$ the subset of $\mathbb{R} \times \mathbb{T}$ given by

$$\mathcal{M} = \mathcal{M}(\hat{p}) := \{(p_1, q_1) \in \mathbb{R} \times \mathbb{T} \,\big|\, \mathtt{H}_\flat(p_1, \hat{p}, q_1) < \mathtt{E}_\flat\}\,, \quad \mathtt{E}_\flat := \mathtt{R}^2 + \mathtt{R}\mathtt{r}\,, \tag{65}$$

where $\mathtt{R}$ and $\mathtt{r}$ are as in Definition 2.1. Although such sets depend on the parameter $\hat{p} \in \hat{D}$, for μ small enough, they are close to a box:

[31] As usual, for a matrix M we denote by $|M| = \sup_{u \neq 0} |Mu|/|u|$ the standard operator norm.

Lemma 2.3 *Let $\mathtt{H}_\flat$ be as in Definition 2.1 and $\mathcal{M}$ be as in (65), and assume that*[32]

$$\mu \leqslant 1/(4\kappa)^2 \,. \tag{66}$$

Then, for all $\hat{p} \in \hat{D}$, one has

$$\big(-\mathtt{R} - \tfrac{\mathtt{r}}{3}, \mathtt{R} + \tfrac{\mathtt{r}}{3} \big) \times \mathbb{T} \subseteq \mathcal{M}(\hat{p}) \subseteq \big(-\mathtt{R} - \tfrac{\mathtt{r}}{2}, \mathtt{R} + \tfrac{\mathtt{r}}{2} \big) \times \mathbb{T} \,. \tag{67}$$

The simole proof is given in Appendix.

Since the reference potential $\bar{\mathtt{G}}$ is a β–Morse function, it has $2N$ critical points, for some $N \in \mathbb{N}$, with different critical values. Let $\bar{\theta}_0 \in [0, 2\pi)$ be the unique point of absolute maximum of the reference potential $\bar{\mathtt{G}}$ of $\mathtt{H}_\flat$. Then, the relative strict non–degenerate maximum and minimum points of $\bar{\mathtt{G}}$, $\bar{\theta}_i \in [\bar{\theta}_0, \bar{\theta}_0 + 2\pi]$, $(0 \leqslant i \leqslant 2N)$ follow in alternating order, $\bar{\theta}_0 < \bar{\theta}_1 < \bar{\theta}_2 < \ldots < \bar{\theta}_{2N} := \bar{\theta}_0 + 2\pi$, in particular, *$\bar{\theta}_i$ are relative maxima/minima points for i even/odd*. The corresponding distinct critical energies will be denoted by

$$\bar{E}_i := \bar{\mathtt{G}}(\bar{\theta}_i) \,, \quad \bar{E}_{2N} = \bar{E}_0 \ \text{ being the unique global maximum of } \bar{\mathtt{G}} \,. \tag{68}$$

By the Implicit Function Theorem, for μ small enough with respect to κ, one can continue the $2N$ critical points $\bar{\theta}_i$ of $\bar{\mathtt{G}}$ obtaining $2N$ critical points $\theta_i = \theta_i(\hat{p})$ of $\mathtt{G}(\hat{p}, \cdot)$ for $\hat{p} \in \hat{D}$. The corresponding distinct critical energies become

$$E_i = E_i(\hat{p}) := \mathtt{G}(\hat{p}, \theta_i(\hat{p})) \,. \tag{69}$$

Furthermore, for μ small , the functions $\theta_i(\hat{p})$ and $E_i(\hat{p})$ preserve the same order of $\bar{\theta}_i$ and $\bar{E}_i$. Indeed, from Definition 1.2 and the Implicit Function Theorem, the following result proven in [14] holds[33]:

Lemma 2.4 *Let $\mathtt{H}_\flat$ be as in Definition 2.1 and assume that*[34]

$$\mu \leqslant 1/(2\kappa)^6 \,. \tag{70}$$

Then, the functions $\theta_i(\hat{p})$ and $E_i(\hat{p})$ defined above are real analytic in $\hat{p} \in \hat{D}_{\mathtt{r}}$ and

$$\sup_{\hat{p} \in \hat{D}_{\mathtt{r}}} |\theta_i(\hat{p}) - \bar{\theta}_i| \leqslant \tfrac{2\epsilon\mu}{\beta\mathtt{s}} \,, \qquad \sup_{\hat{p} \in \hat{D}_{\mathtt{r}}} |E_i(\hat{p}) - \bar{E}_i| \leqslant 3\kappa^3\epsilon\mu \,. \tag{71}$$

Furthermore, the relative order of $\theta_i(\hat{p})$ and $E_i(\hat{p})$ is, for every $\hat{p} \in \hat{D}_{\mathtt{r}}$, the same as that of, respectively, $\bar{\theta}_i$ and $\bar{E}_i$.

Therefore, under the assumption (70), we see that the phase space $\mathcal{M}$ is disconnected by the separatrices[35] into exactly $2N+1$ open connected components $\mathcal{M}^i = \mathcal{M}^i(\hat{p})$, for $0 \leqslant i \leqslant 2N$, which can be labelled so that:

- the *odd regions* $\mathcal{M}^{2j-1}$ (for $1 \leqslant j \leqslant N$) contain the elliptic points $(0, \theta_{2j-1})$ and have as boundary parts of separatrices; topologically, such regions are discs;
- the *outer even regions* $\mathcal{M}^0$ and $\mathcal{M}^{2N}$ are homotopically *non trivial* annuli bounded by the most external separatrices and one of the two curves $\mathtt{H}_\flat^{-1}(\mathtt{E}_\flat)$;

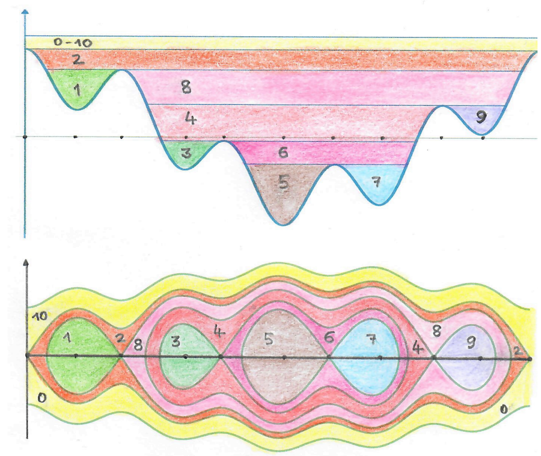
[32] Recall the definition of κ in (44).

[33] See Lemma 3.1 in [14].

[34] Notice that condition (70) is stronger than (66).

[35] I.e., the stable manifolds (curves) of the hyperbolic points $(0, \theta_{2j})$.

- when $N > 1$, the *inner even regions* $\mathcal{M}^{2j}$ (for $1 \leqslant j \leqslant N-1$) are homotopically trivial annuli[36] whose boundary is given by two pieces of separatrices (with different energies).



The Morse potential $\bar{\mathtt{G}}(q_1) = \sin q_1 + \frac{1}{2}\cos(5q_1)$ with 10 critical points (top)
and the phase portrait of $\mathtt{H}_\flat := p_1^2 + \bar{\mathtt{G}}(q_1)$ (bottom)
Labels of corresponding regions are as in Definition 2.4

More formally, we can define the $2N+1$ regions $\mathcal{M}^i$ in terms of suitable energy intervals $(E_-^{(i)}, E_+^{(i)})$ as follows.

Let E_i be the critical energies defined in (69), and let $\mathtt{E}_\flat$ the reference energy defined in (65).

Definition 2.4 (i) (Outer regions) *For $i = 0, 2N$, let $E_-^{(0)} = E_-^{(2N)} := E_0$, and $E_+^{(0)} = E_+^{(2N)} := \mathtt{E}_\flat$. Then, the 'lower outer region' $\mathcal{M}^{(0)}$ is the connected component of $\mathtt{H}_\flat^{-1}\big((E_-^{(0)}, E_+^{(0)})\big)$ contained in $\{p_1 < 0\}$, while the 'upper outer region' $\mathcal{M}^{(2N)}$ is the connected component of $\mathtt{H}_\flat^{-1}\big((E_-^{(2N)}, E_+^{(2N)})\big)$ contained in $\{p_1 > 0\}$.*
(ii) (Inner region, $N = 1$) *When $N = 1$, $\mathcal{M}^{(1)}$ is just the region enclosed by the unique separatrix $\mathtt{H}_\flat^{-1}(E_0)$; the orbits in $\mathcal{M}^{(1)}$ have energies ranging in the critical interval $[E_-^{(1)}, E_+^{(1)}) := [E_1, E_0)$.*
(ii) (Inner regions, $N > 1$) *Define $E_-^{(i)} := E_i$.*
For i odd, let $E_+^{(i)} := \min\{E_{i-1}, E_{i+1}\}$ and define $\mathcal{M}^{(i)}$ as the connected component of $\mathtt{H}_\flat^{-1}\big([E_-^{(i)}, E_+^{(i)})\big)$ containing the elliptic equilibrium $(0, \theta_i)$.
Finally, for $0 < i = 2j < 2N$ even, define

$$j_- := \max\{\ell < j \big| \; E_{2\ell} > E_{2j}\}\,, \; j_+ := \min\{\ell > j \big| \; E_{2\ell} > E_{2j}\}\,, \; E_+^{(i)} := \min\{E_{2j_-}, E_{2j_+}\}\,;$$

and define $\mathcal{M}^{(i)}$ as the connected component of $\mathtt{H}_\flat^{-1}\big((E_-^{(i)}, E_+^{(i)})\big)$ whose boundary contains the hyperbolic point $(0, \theta_i)$.

[36] I.e., annuli in the cylinder $\mathbb{R} \times \mathbb{T}$ which are contractible.

Notice that the phase space $\mathcal{M}$ is the union of the regions $\mathcal{M}^{(i)}$ and the singular zero–measure set $S = S(\hat{p})$ formed by the N separatrices:

$$\mathcal{M} = \mathcal{M}(\hat{p}) = \bigcup_{i=0}^{2N} \mathcal{M}^i \ \cup \ S = \bigcup_{i=0}^{2N} \mathcal{M}^i(\hat{p}) \ \cup \ S(\hat{p}) \,. \tag{72}$$

Below we shall also consider the following $(n+1)$–dimensional domains:

$$\begin{aligned} \check{\mathcal{M}} \ &:= \ \{(p, q_1) \ \text{ s.t. } \ \hat{p} \in \hat{D}, \ (p_1, q_1) \in \mathcal{M}(\hat{p})\} \,, \\ \check{\mathcal{M}}^i \ &:= \ \{(p, q_1) \ \text{ s.t. } \ \hat{p} \in \hat{D}, \ (p_1, q_1) \in \mathcal{M}^i(\hat{p})\} \,. \end{aligned} \tag{73}$$

Notice that $\bigcup_{0 \leqslant i \leqslant 2N} \check{\mathcal{M}}^i$ covers $\check{\mathcal{M}}$ up to a set of measure zero.

Arnol'd–Liouville's action/energy functions

Let $E \in [E^i_-(\hat{p}), E^i_+(\hat{p})]$ and let γ^i be the (possibly, piece–wise) smooth closed curve in the clusure of $\mathcal{M}^i(\hat{p})$ given by

$$\gamma^i = \gamma^i(E, \hat{p}) := \{(p_1, q_1) \in \overline{\mathcal{M}^i(\hat{p})} \ \text{ s.t. } \ \mathtt{H}_\flat(p_1, \hat{p}, q_1) = E\} \,,$$

oriented clockwise[37]; for $2 \leqslant j \leqslant N$ consider also the trivial curves $\gamma^i_j = \{(p_j, s) : s \in \mathbb{T}\}$.
Then, the classical *Arnol'd–Liouville's action functions* are given by

$$\begin{aligned} I_1^{(i)}(E) &= I_1^{(i)}(E, \hat{p}) := \frac{1}{2\pi} \oint_{\gamma^i} p_1 dq_1 \,, \\ I_j &= \frac{1}{2\pi} \oint_{\gamma^i_j} p_j dq_j = \frac{p_j}{2\pi} \int_{\mathbb{T}} dq_j = p_j \,, \qquad \forall \ 2 \leqslant j \leqslant N \,. \end{aligned}$$

The action function $E \to I^i_1(E, \hat{I})$ is strictly monotone and its inverse is, by definition, the *energy function* $I_1 \to \mathtt{E}^i(I_1, \hat{I})$. We also define $\bar{I}^i_1 := I^i_1|_{\mu=0}$ and its inverse function[38] $\bar{\mathtt{E}}^i := \mathtt{E}^i|_{\mu=0}$.
We can now describe the fine analytic properties of the action/energy functions.

Critical holomorphic behaviour and action estimates

The first result describes the exact behaviour of the action functions as the energy approaches the critical energy of separatrices and contains estimates on the derivatives of the action functions that will play a central rôle in the discussion on the twist Hessian matrix in § 4. The following theorem has been proven in [14, Theorem 3.1].

Theorem 2.3 *Let $\mathtt{H}_\flat$ be a Hamiltonian in standard form as in Definition 2.1, let $\kappa \geqslant 4$ be such that (44) holds and let $2N$ be the number of critical points of the reference potential $\bar{\mathtt{G}}$. Then, there exists a suitable constant $\mathbf{c} = \mathbf{c}(n, \kappa) \geqslant 2^8 \kappa^3$ such that, if* [39]

$$\mu \leqslant 1/\mathbf{c}^2 \leqslant 1/(2^{16}\kappa^6) \,, \tag{74}$$

[37] For the non contractible curves ($i = 0, 2N$) the orientation is 'to the right' on $\mathcal{M}^{2N}$, 'to the left' on $\mathcal{M}^0$.
[38] Note that when $\mu = 0$, $\mathtt{H}_\flat$ becomes simply $\bar{\mathtt{H}}_\flat = p_1^2 + \bar{\mathtt{G}}(q_1)$.
[39] Note that (74) implies the hypothesis of Lemma 2.4. Thus, in particular also $\mathtt{H}_\flat$ has $2N$ critical points.

then, for all $0 \leqslant i \leqslant 2N$ *and* $\hat{I} \in \hat{D}$, *the action functions* $E \in (E^i_-(\hat{I}), E^i_+(\hat{I})) \mapsto I^i_1(E, \hat{I})$ *verify the following properties.*

(i) **(Universal behaviour at critical energies)** *There exist functions* $\phi^i_-(z, \hat{I})$, $\psi^i_-(z, \hat{I})$ *for* $0 \leqslant i \leqslant 2N$, *and, functions* $\phi^i_+(z, \hat{I})$, $\psi^i_+(z, \hat{I})$, *for* $0 < i < 2N$, *which are real analytic in a complex neighborhood of the set* $\{z = 0\} \times \hat{D}$ *and satisfy*

$$I^i_1\big(E^i_\mp(\hat{I}) \pm \epsilon z,\ \hat{I}\big) = \phi^i_\mp(z, \hat{I}) + \psi^i_\mp(z, \hat{I})\ z \log z\ , \quad \forall\ 0 < z < 1/\mathbf{c}\,,\ \hat{I} \in \hat{D}\,. \tag{75}$$

the functions $\phi^i_\pm(z, \hat{I})$, $\psi^i_\pm(z, \hat{I})$ *are real analytic on* $\{z \in \mathbb{C} : |z| < 1/\mathbf{c}\} \times \hat{D}_{\mathbf{r}}$, *where satisfy:*

$$\begin{aligned}
&\sup_{|z|<1/\mathbf{c},\, \hat{I}\in\hat{D}_{\mathbf{r}}} \big(|\phi^i_\pm| + |\psi^i_\pm|\big) \leqslant \mathbf{c}\sqrt{\epsilon}\,,\\
&\sup_{|z|<1/\mathbf{c},\, \hat{I}\in\hat{D}_{\mathbf{r}/2}} \big(|\partial_{\hat{I}}\phi^i_\pm| + |\partial_{\hat{I}}\psi^i_\pm|\big) \leqslant \mathbf{c}\mu_{\mathrm{o}}\,, \qquad \mu_{\mathrm{o}} := \frac{\sqrt{\epsilon}}{\mathbf{r}}\mu \overset{(43)}{\leqslant} 2^{-8}\mu\,.
\end{aligned} \tag{76}$$

Moreover,

$$|\phi^i_\pm - \bar{\phi}^i_\pm|\,,\ |\psi^i_\pm - \bar{\psi}^i_\pm|\ \leqslant\ \mathbf{c}\sqrt{\epsilon}\mu\,, \tag{77}$$

where $\bar{\phi}^i_\pm := \phi^i_\pm|_{\mu=0}$ *and* $\bar{\psi}^i_\pm := \psi^i_\pm|_{\mu=0}$.

(ii) **(Limiting critical values)** *The following bounds at the limiting critical energy values hold:*

$$\begin{array}{lll}
|\psi^i_+(0, \hat{I})| \geqslant \sqrt{\epsilon}/\mathbf{c}\,, & 0 < i < 2N\,, & \forall\ \hat{I} \in \hat{D}_{\mathbf{r}}\,,\\
|\psi^{2j}_-(0, \hat{I})| \geqslant \sqrt{\epsilon}/\mathbf{c}\,, & 0 \leqslant j \leqslant N\,, & \forall\ \hat{I} \in \hat{D}_{\mathbf{r}}\,,\\
\psi^i_+(0, \hat{I}) > 0\,, & 0 < i < 2N\,, & \forall\ \hat{I} \in \hat{D}\,,\\
\psi^{2j}_-(0, \hat{I}) < 0\,, & 0 \leqslant j \leqslant N\,, & \forall\ \hat{I} \in \hat{D}\,,
\end{array} \tag{78}$$

while, in the case of relative minimal critical energies, one has, $\forall\ \hat{I} \in \hat{D}$, $0 < z < 1/\mathbf{c}$,

$$\phi^{2j-1}_-(0, \hat{I}) = 0\,, \qquad \psi^{2j-1}_-(z, \hat{I}) = 0\,, \qquad \forall\ 1 \leqslant j \leqslant N\,. \tag{79}$$

(iii) **(Estimates on derivatives of actions on real domains)** *The derivatives of the actions with respect to energy verify, on real domains, the following estimates:*

$$\inf_{(E^i_-, E^i_+)} \partial_E I^i_1 \geqslant \frac{1}{\mathbf{c}\sqrt{\epsilon}}\,, \qquad \forall\ \hat{I} \in \hat{D}\,,\ \forall\ 0 < i < 2N\,; \tag{80}$$

$$\min\big\{\partial_E I^{2N}_1\ ,\ \partial_E I^0_1\big\} \geqslant \frac{1}{\mathbf{c}\sqrt{E + \epsilon}}\,, \quad \forall\, E > E_{2N}\,,\ \forall\ \hat{I} \in \hat{D}\,.$$

(iv) **(Estimates on derivatives of actions on complex domains and perturbative bounds)** *For* $\lambda > 0$ *satisfying*

$$\mathbf{c}\mu \leqslant \lambda \leqslant 1/\mathbf{c}\,, \tag{81}$$

define the following complex energy–domains:

$$\mathcal{E}^i_\lambda := \left\{ \begin{array}{ll}
\{z \in \mathbb{C} : \bar{E}^i_- - \epsilon/\mathbf{c}\ < \mathrm{Re}\, z < \bar{E}^i_+ - \lambda\epsilon\,,\ |\mathrm{Im}\, z| < \epsilon/\mathbf{c}\}\,, & i \text{ odd}\,,\\
\{z \in \mathbb{C} : \bar{E}^i_- + \lambda\epsilon < \mathrm{Re}\, z < \bar{E}^i_+ - \lambda\epsilon\,,\ |\mathrm{Im}\, z| < \epsilon/\mathbf{c}\}\,, & i \text{ even}\,, i \neq 0, 2N\,,\\
\{z \in \mathbb{C} : \bar{E}^i_- + \lambda\epsilon < \mathrm{Re}\, z < \bar{E}^i_+\,,\ |\mathrm{Im}\, z| < \epsilon/\mathbf{c}\}\,, & i = 0, 2N\,.
\end{array}\right. \tag{82}$$

Then, for $0 \leqslant i \leqslant 2N$, the functions I_1^i and $\bar{I}_1^i$ are holomorphic on the domains $\mathcal{E}^i_\lambda \times \hat{D}_{\mathtt{r}}$, and satisfy the following estimates:

$$\sup_{\mathcal{E}^i_\lambda \times \hat{D}_{\mathtt{r}/2}} |\partial_{\hat{I}} I_1^i| \leqslant \mathbf{c}^2\, \mu_{\mathrm{o}}\,, \quad \sup_{\mathcal{E}^i_\lambda} \big|\partial_E \bar{I}_1^i\big| \leqslant \mathbf{c}^2\, \frac{|\log\lambda|}{\sqrt{\epsilon}}\,, \quad \sup_{\mathcal{E}^i_\lambda \times \hat{D}_{\mathtt{r}}} \big|\partial_E I_1^i - \partial_E \bar{I}_1^i\big| \leqslant \frac{\mathbf{c}^2 \mu}{\lambda\sqrt{\epsilon}}\,. \tag{83}$$

Remark 2.8 (i) Eq. (79) confirms the known analyticity at minima of actions as function of energy. (ii) A formula similar to (75) is given in [6] (compare Eq. (5.8) of Theorem 5.2 there).

We finally report a remarkable property of standard Hamiltonians $\mathtt{H}_\flat$, whose reference potential $\bar{\mathtt{G}}$ is close enough to a cosine. In such a case, in fact, one has uniform concavity of the second derivative of the energy function:

Proposition 2.1 *Assume that, for some $\theta_0 \in \mathbb{R}$, $\bar{\mathtt{G}}$ satisfies*

$$|\bar{\mathtt{G}}(\theta) - \cos(\theta+\theta_0)|_1 := \sup_{\mathbb{T}_1} |\bar{\mathtt{G}}(\theta) - \cos(\theta+\theta_0)| \leqslant 2^{-40}\,. \tag{84}$$

Then $N = 1$ and

$$\partial^2_{I_1} \bar{\mathtt{E}}^1(\bar{I}_1^1(E)) \leqslant -\frac{1}{27}\,, \qquad \forall E \in (\bar{E}_1, \bar{E}_2)\,.$$

Also this result is proven in [14]; compare Proposition 5.12 there.

Arnol'd–Liouville's action–angle variables in n d.o.f.

Let us now discuss the Arnol'd–Liouville's action–angle variables for the Hamiltonian $\mathtt{H}_\flat$ viewed as a n degrees of freedom Hamiltonian on the $2n$–dimensional phase space $\check{\mathcal{M}}^i \times \mathbb{T}^{n-1}$.

For every fixed $\hat{p} \equiv \hat{I} \in \hat{D}$, the map $(p_1, q_1) \to I_1^{(i)}\big(\mathtt{H}_\flat(p_1, \hat{I}, q_1), \hat{I}\big)$ can be symplectically completed with the angular term[40] $(p_1, q_1) \to \varphi_1^{(i)}(p_1, q_1; \hat{I}) = \varphi_1^{(i)}(p_1, \hat{I}, q_1)$.
Defining the normal domains[41]

$$\mathcal{B}^i := \big\{ I = (I_1, \hat{I}) \mid \hat{I} \in \hat{D},\ \ I_1^{(i)}(E^i_-(\hat{I}), \hat{I}) < I_1 < I_1^{(i)}(E^i_+(\hat{I}), \hat{I}) \big\}\,, \tag{85}$$

we see that, by construction, the map[42]

$$(p, q_1) \in \check{\mathcal{M}}^i \to (I, \varphi_1) = \big(I_1^{(i)}(\mathtt{H}_\flat(p, q_1), \hat{I}), \hat{I}, \varphi_1^{(i)}(p, q_1)\big) \in \mathcal{B}^i \times \mathbb{T}$$

is surjective and invertible; let us denote by

$$\check{\Phi}^i : (I, \varphi_1) \in \mathcal{B}^i \times \mathbb{T} \ \to\ (p, q_1) \in \check{\mathcal{M}}^i\,, \qquad (\hat{p} = \hat{I})\,,$$

its inverse map. Note that such 'Arnol'd-Liouville suspended' transformation $\check{\Phi}^i$ integrates $\mathtt{H}_\flat$, i.e.,

$$\mathtt{H}_\flat \circ \check{\Phi}^i(I, \varphi_1) = \mathtt{E}^{(i)}(I)\,, \qquad dp_1 \wedge dq_1|_{\hat{I} = \mathrm{const}} = dI_1 \wedge d\varphi_1\,. \tag{86}$$

[40] Such completion is unique if one fixes, e.g., $\varphi_1^{(i)}(p_1, 0; \hat{I}) = 0$.
[41] Recall Definition 2.4. For i odd, $I_1^{(i)}(E^i_-(\hat{I}), \hat{I}) = 0$, which is the action of the elliptic point.
[42] Recall the definition of $\check{\mathcal{M}}^i$ in (73).

By the standard Arnol'd–Liouville construction of the angle variables, one sees easily that the complete symplectic action–angle map $\Phi^i : (I, \varphi) \mapsto (p, q)$ has the form

$$\Phi^i(I,\varphi) = \begin{cases} (\eta^i, \hat{I}, \psi^i, \hat{\varphi} + \chi^i)\,, & \text{if } 0 < i < 2N\,, \\ (\eta^i, \hat{I}, \varphi_1 + \psi^i, \hat{\varphi} + \chi^i)\,, & \text{if } i = 0, 2N\,, \end{cases} \tag{87}$$

where η^i, χ^i, ψ^i are function of (I, φ_1) only and are 2π–periodic in φ_1, and, in the case $i = 0, 2N$, $\sup |\partial_{\varphi_1} \psi^i| < 1$.

By construction, $\Phi^i : \mathcal{B}^i \times \mathbb{T}^n \xrightarrow{\text{onto}} \check{\mathcal{M}}^i \times \mathbb{T}^{n-1}$ *is a global symplectomorphism*, and by (86), one has

$$(\mathtt{H}_\flat \circ \Phi^i)(I,\varphi) = (\mathtt{H}_\flat \circ \check{\Phi}^i)(I, \varphi_1) = \mathtt{E}^{(i)}(I)\,, \qquad \forall\; 0 \leqslant i \leqslant 2N\,. \tag{88}$$

Next, we introduce suitable decreasing subdomains $\mathcal{B}^i(\lambda)$ of $\mathcal{B}^i$ depending on a non negative parameter λ so that $\mathcal{B}^i(0) = \mathcal{B}^i$ and such that the map Φ^i has, for positive λ, a holomorphic extension on a suitable complex neighborhood of $\mathcal{B}^i(\lambda) \times \mathbb{T}^n$.

Define

$$\lambda_{\text{max}} = \lambda_{\text{max}}(\hat{I}) := \big(E_+(\hat{I}) - E_-(\hat{I})\big)/\epsilon\,, \qquad \bar{\lambda}_{\text{max}} := \big(\bar{E}_+ - \bar{E}_-\big)/\epsilon\,. \tag{89}$$

Notice that, by (44), Definitions 1.2, 2.1, and (43) one has

$$1/\kappa \leqslant \beta/\epsilon \leqslant \bar{\lambda}_{\text{max}} \leqslant 2\,; \tag{90}$$

notice also that, by (71), we have[43]

$$|\lambda_{\text{max}} - \bar{\lambda}_{\text{max}}| \leqslant 6\kappa^3\mu\,, \qquad \lambda_{\text{max}} \geqslant 1/2\kappa\,. \tag{91}$$

Then, for $0 \leqslant \lambda \leqslant \lambda_{\text{max}}$ define[44]:

$$\begin{aligned} a^i_\lambda(\hat{I}) &:= I^i_1(E^i_-(\hat{I}) + \lambda\epsilon, \hat{I})\,, \qquad \forall\, 0 \leqslant i \leqslant 2N\,, \\ b^i_\lambda(\hat{I}) &:= \begin{cases} I^i_1(E^i_+(\hat{I}) - \lambda\epsilon, \hat{I})\,, & \forall\; 0 < i < 2N \\ I^i_1\big(\mathtt{E}_\flat, \hat{I}\big)\,, & i = 0, 2N\,. \end{cases} \\ a^i(\hat{I}) &:= a^i_0(\hat{I})\,,\; b^i(\hat{I}) := b^i_0(\hat{I})\,, \quad \forall\; 0 \leqslant i \leqslant 2N\,, \\ \mathcal{B}^i(\lambda) &:= \{I = (I_1, \hat{I}) :\; \hat{I} \in \hat{D}\,,\; a^i_\lambda(\hat{I}) < I_1 < b^i_\lambda(\hat{I})\}\,, \quad 0 \leqslant \lambda \leqslant \lambda_{\text{max}}\,. \end{aligned} \tag{92}$$

Remark 2.9 (i) By the above definitions one has that

$$a^{2j-1}(\hat{I}) := a^{2j-1}_0(\hat{I}) = I^{2j-1}_1(E^{2j-1}_-(\hat{I}), \hat{I}) \equiv 0\,, \tag{93}$$

reflecting the analyticity at the elliptic points; compare Remark 2.8–(i) above.
(ii) By (85) and (92) one sees that $\mathcal{B}^i = \mathcal{B}^i(0) = \bigcup_{0<\lambda<\lambda_{\text{max}}} \mathcal{B}^i(\lambda)$.

The holomorphic properties of the Arnol'd–Liouville symplectic maps are described in following theorem, proven in [14, Theorem 4.1]. Recall the definition of the constant $\mathbf{c}$ in Theorem 2.3.

[43] Recall that $\mu \leqslant 1/\mathbf{c}^2$ and $\mathbf{c} \geqslant 2^8\kappa^3$ (compare Theorem 2.3).
[44] Recall the definition of $\mathtt{E}_\flat$ in (65).

Theorem 2.4 *Under the hypotheses of Theorem 2.3 there exists a constant* $\hat{\mathbf{c}} = \hat{\mathbf{c}}(n, \kappa) \geqslant 4\mathbf{c}^2$ *depending only on* n *and* κ *such that, taking*

$$\mu \leqslant 1/\hat{\mathbf{c}}\,, \tag{94}$$

the symplectic transformation $\check{\Phi}^i$ *extends, for any* $0 \leqslant i \leqslant 2N$ *and* $0 < \lambda \leqslant 1/\hat{\mathbf{c}}$*, to a real analytic map*

$$\Phi^i : \big(\mathcal{B}^i(\lambda)\big)_{\rho_\lambda} \times \mathbb{T}^n_{\sigma_\lambda} \;\to\; D_{\mathtt{r}} \times \mathbb{T}^n_{\mathtt{s}/4}\,, \qquad \forall\, 0 < \lambda \leqslant 1/\hat{\mathbf{c}}\,, \tag{95}$$

where

$$\rho_\lambda := \tfrac{\sqrt{\epsilon}}{\hat{\mathbf{c}}}\,\lambda|\log\lambda|\,, \qquad \sigma_\lambda := \tfrac{1}{\hat{\mathbf{c}}|\log\lambda|}\,. \tag{96}$$

Now, let $0 < \lambda \leqslant 1/\hat{\mathbf{c}}$*, then the function* $\mathtt{E}^i$ *admits a holomorphic extension on* $\big(\mathcal{B}^i(\lambda)\big)_{\rho_\lambda}$*, where, setting* $\hat{\lambda} := \lambda|\log\lambda|^3$*, one has*

$$\big|\partial_{I_1}\mathtt{E}^i\big| \leqslant \hat{\mathbf{c}}\sqrt{\epsilon + |\mathtt{E}^i|}\,, \quad \big|\partial^2_{I_1}\mathtt{E}^i\big| \leqslant \tfrac{\hat{\mathbf{c}}}{\hat{\lambda}}\,, \quad \big|\partial^2_{I_1\hat{I}}\mathtt{E}^i\big| \leqslant \hat{\mathbf{c}}\,\tfrac{\mu_{\mathrm{o}}}{\hat{\lambda}}\,, \quad \big|\partial^2_{\hat{I}}\mathtt{E}^i\big| \leqslant \hat{\mathbf{c}}\big(\tfrac{\sqrt{\epsilon}}{\mathtt{r}} I_1^i + \tfrac{\mu_{\mathrm{o}}}{\hat{\lambda}}\big)\mu_{\mathrm{o}}\,; \tag{97}$$

furthermore, defining

$$D^\flat := (-\mathtt{R} - \mathtt{r}/3, \mathtt{R} + \mathtt{r}/3) \times \hat{D}\,, \qquad \check{\mathcal{M}}^i(\lambda) := \check{\Phi}^i(\mathcal{B}^i(\lambda) \times \mathbb{T})\,, \tag{98}$$

one has

$$\mathrm{meas}\Big((D^\flat \times \mathbb{T}) \setminus \bigcup_{0\leqslant i\leqslant 2N} \check{\mathcal{M}}^i(\lambda)\Big) \leqslant \hat{\mathbf{c}}\,\sqrt{\epsilon}\,\mathrm{meas}(\hat{D})\;\lambda|\log\lambda|\,. \tag{99}$$

Remark 2.10 Observe that, by (48), (49), (20), (74), (32), (53) and (56), it is[45]

$$1/\kappa < \check{\mathtt{s}}/4\,, \qquad \tfrac{\varepsilon_k \chi_k}{\mathtt{r}}\,\mu < \mathtt{r}/6\,, \qquad \tfrac{4\varepsilon_k\chi_k}{\mathtt{r}^2}\,\mu < \tfrac{\check{\mathtt{s}}}{2^{20}\kappa^3} < \check{\mathtt{s}}/2^{20}\,. \tag{100}$$

Thus, since[46] $\lambda \leqslant 1/\hat{\mathbf{c}}$, by (74), σ_λ in (96) satisfies

$$\sigma_\lambda < \check{\mathtt{s}}/2^{20}\,. \tag{101}$$

3 Secondary nearly–integrable structure at simple resonances

Now we go back to the original system in the simply–resonant zones governed by the Hamiltonians $\mathcal{H}_k(\mathtt{y}, \mathtt{x})$ in (37) and discuss their global nearly–integrable structure with exponential small perturbations (compare Theorem 3.1 below).

As mentioned above (see item (ii) in Remark 2.6), the problem here is that the symplectic transformations of Theorem 2.2, which put the simply–resonant Hamiltonians $\mathtt{H}_k$ in (53) in standard form, are, in general, *not well defined in the fast angles* $\hat{\mathtt{q}} = (\mathtt{q}_2, ..., \mathtt{q}_n)$, making the construction of *global* action–angle variables for the full Hamiltonians $\mathcal{H}_k(\mathtt{y}, \mathtt{x})$ in (37) not straightforward.

To overcome such homotopy problems, we shall exploit the particular group structure of the various symplectic transformations involved, and show that, introducing a special *ad hoc* conjugacy, one can indeed obtain globally well defined symplectic maps; see, in particular, (122) below.

[45] Recall that $\varepsilon < 1$; see (1).

[46] Recall the hypotheses of Theorem 2.4.

Special sets of symplectic transformations

Besides the group $\mathfrak{G}_\dagger$ introduced in Definition 2.2 above, we shall introduce two new special classes of symplectic transformations, which will be used in the proof of Theorem 3.1. Recall the notation in (45).

Definition 3.1 (a) *Given a domain $\hat{\mathrm{D}} \subseteq \mathbb{R}^{n-1}$, $\mathfrak{G}$ denotes the formal*[47] *group of symplectic transformations of the form*

$$(p, q) \in \mathrm{D} \times \mathbb{T}^n \overset{\Phi}{\mapsto} (P, Q) = (\eta, \hat{p}, q_1 + \psi, \hat{q} + \chi) \in \mathbb{R}^n \times \mathbb{T}^n\,,$$

where: $\mathrm{D} \subseteq \mathbb{R}^n$ is a normal smooth domain[48] *over $\hat{\mathrm{D}}$, the functions η, ψ, χ depend on (p, q_1), are 2π–periodic in q_1 and the $(n+1)$–dimensional the map*

$$(p, q_1) \mapsto \check{\Phi}(p, q_1) = (\eta, \hat{p}, q_1 + \psi)$$

is injective.

(b) *Given a domain $\hat{\mathrm{D}} \subseteq \mathbb{R}^{n-1}$, $\mathfrak{G}_0$ denotes the set of smooth symplectic transformations of the form*

$$(p, q) \in \mathrm{D} \times \mathbb{T}^n \overset{\Phi}{\mapsto} (P, Q) = (\eta, \hat{p}, \psi, \hat{q} + \chi) \in \mathbb{R}^{n+1} \times \mathbb{T}^{n-1}\,,$$

where $\mathrm{D} \subseteq \mathbb{R}^n$ is a normal smooth domain over $\hat{\mathrm{D}}$; the functions η, ψ, χ depend only on (p, q_1) and are 2π–periodic in q_1.

Let us collect a few observations and discuss the main properties of such classes, but, first of all, notice that all the above maps leave fixed the variable $\hat{p} \in \hat{\mathrm{D}} \subseteq \mathbb{R}^n$ and the set $\hat{\mathrm{D}}$. Thus, in the following discussion, the domain $\hat{\mathrm{D}}$ is fixed once and for all.

Remark 3.1 (i) The Arnol'd–Liouville map Φ^i in the outer cases (87) $(i = 0, 2N)$ belongs to $\mathfrak{G}$ (since $\sup |\partial_{q_1} \psi| < 1$), while Φ^i in the inner case (87) $(0 < i < 2N)$ belongs to $\mathfrak{G}_0$.
Notice also that Φ_2 in Theorem 2.2–(iii) is a near–to–the–identity symplectic map belonging to $\mathfrak{G}$.

(ii) In the definition of $\mathfrak{G}$ and $\mathfrak{G}_0$, the functions η and ψ are scalar functions, while χ has $(n-1)$ components. Notice that, since Φ is assumed to be symplectic, these maps are such that

$$\begin{aligned} d\eta \wedge dq_1 + d\eta \wedge d\psi + d\hat{p} \wedge d\chi &= dp_1 \wedge dq_1\,, && (\Phi \in \mathfrak{G})\,,\\ d\eta \wedge d\psi + d\hat{p} \wedge d\chi &= dp_1 \wedge dq_1\,, && (\Phi \in \mathfrak{G}_0). \end{aligned}$$

(iii) All maps in the group $\mathfrak{G}_\dagger$ in Definition 2.2 have a common domain of definition, i.e., $(\mathbb{R} \times \hat{\mathrm{D}}) \times \mathbb{R}^n$. On the other hand, every map $\Psi \in \mathfrak{G}$ has its own domain of definition D. Thus, the composition $\Psi_1 \circ \Psi_2$ of two maps in $\mathfrak{G}$

$$\Psi_1 : \mathrm{D}_1 \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n\,, \qquad \Psi_2 : \mathrm{D}_2 \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n$$

is well defined *only when the compatibility condition $\Psi_2\big(\mathrm{D}_2 \times \mathbb{T}^n\big) \subseteq \mathrm{D}_1 \times \mathbb{T}^n$ is satisfied*. This is the reason why the cautionary word 'formal' appears in the definition of $\mathfrak{G}$. However, as already noticed, all maps in $\mathfrak{G}$ verify $\pi_{\hat{p}}(\mathrm{D}) = \hat{\mathrm{D}}$, which is fixed *a priori*.

[47] See Remark 3.1–(iii) below.

[48] I.e., $D = \{(p_1, \hat{p}) :\ \alpha(\hat{p}) < p_1 < \hat{\beta}(\hat{p})\,,\ \hat{p} \in \hat{\mathrm{D}}\}$ where α and β are smooth function on $\hat{\mathrm{D}}$.

(iv) If $\Phi \in \mathfrak{G}$, by definition $\check{\Phi}$ is injective, so that also Φ itself is injective. Furthermore, for any fixed p, the map $q_1 \to Q_1 = q_1 + \psi$ is a continuous injective map on the circle $\mathbb{T}^1$, hence it is surjective, and, therefore, it is a smooth (orientation preserving) circle diffeomorphism. Thus, $q \to Q = (q_1 + \psi, \hat{q} + \chi)$ *is a global diffeomorphism of* $\mathbb{T}^n$, and $\Phi : \mathrm{D} \times \mathbb{T}^n \to \Phi(\mathrm{D} \times \mathbb{T}^n) \subseteq \mathbb{R}^n \times \mathbb{T}^n$ *is a global symplectomorphism.* Notice also that *if* $\Phi, \Phi' \in \mathfrak{G}$ *and the composition* $\Phi \circ \Phi'$ *is well defined, then* $\Phi \circ \Phi' \in \mathfrak{G}$.

(v) The definition of the first $(n+1)$ component of any member of the above families depends only on the first $(n+1)$ variables (p, q_1). Therefore, any finite compositions of maps $\Psi_i \in \mathfrak{G}_\dagger \cup \mathfrak{G} \cup \mathfrak{G}_0$, $1 \leqslant i \leqslant m$, *whenever the composition is well defined*, satisfies

$$\Psi_i \in \mathfrak{G}_\dagger \cup \mathfrak{G} \cup \mathfrak{G}_0 \quad \Longrightarrow \quad (\Psi_1 \circ \cdots \circ \Psi_m)\check{} = (\check{\Psi}_1 \circ \cdots \circ \check{\Psi}_m) \, . \tag{102}$$

(vi) Finally, one readily verifies that the following property holds:

$$\Phi \in \mathfrak{G}_0 \quad \text{and} \quad \Psi \in \mathfrak{G}_\dagger \cup \mathfrak{G} \qquad \Longrightarrow \qquad \Psi \circ \Phi \in \mathfrak{G}_0 \, . \tag{103}$$

Action–angles variables for the secular standard Hamiltonians $\mathtt{H}_k$ at simple resonances

For each $k \in \mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$, we may apply the theory of § 2.3 to the secular Hamiltonians described in Theorem 2.2 in standard form $\mathtt{H}_\flat = \mathtt{H}_k$; see (54), (55), and (56).

By (88), we get that, for every $k \in \mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$ and $0 \leqslant i \leqslant 2N_k$, the Arnol'd–Liouville map

$$\Phi^i \ : \ \mathcal{B}^i_k \times \mathbb{T}^n \ \overset{\text{onto}}{\longrightarrow} \ \check{\mathcal{M}}^i_k \times \mathbb{T}^{n-1} \tag{104}$$

integrates $\mathtt{H}_k$, i.e.:

$$(\mathtt{H}_k \circ \Phi^i)(I, \varphi) = (\mathtt{H}_k \circ \check{\Phi}^i)(I, \varphi_1) = \mathtt{E}^{(i)}_k(I) \, , \qquad \forall \ 0 \leqslant i \leqslant 2N_k \, , \tag{105}$$

where $\mathcal{B}^i_k$, $\check{\mathcal{M}}^i_k$ and $\mathtt{E}^{(i)}_k$ correspond to $\mathcal{B}^i$, $\check{\mathcal{M}}^i$ and $\mathtt{E}^{(i)}$ in § 2.3 in the case[49] $\mathtt{H}_\flat = \mathtt{H}_k$.

Beware that, even if sometimes, for ease of notation, we do not report the dependence upon the resonance label $k \in \mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$, we are treating different Hamiltonians in the neighbourhoods of simple resonances labelled by $k \in \mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$.

Finally, we shall use the following notations: Given a function $\mathtt{g} : \hat{\mathrm{D}} \to \mathbb{R}$, we shall denote by $\jmath_{\mathtt{g}}$ the translation

$$\jmath_{\mathtt{g}}(p) := (p_1 + \mathtt{g}(\hat{p}), \hat{p}) \, . \tag{106}$$

Notice that, by the definition of $\Psi_{\mathtt{g}}$ in (46), one has

$$\check{\Psi}_{\mathtt{g}}(p, q_1) = (\jmath_{\mathtt{g}}(p), q_1) \, . \tag{107}$$

Global action–angle variables at simple resonances

We are now ready to state and prove the first step of the proof of Theorem 1.1, which consists in showing how to construct symplectic action–angle maps which put a generic nearly–integrable natural systems, near simple resonances, *for all* $k \in \mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$, into uniform analytic nearly–integrable form *with exponentially small perturbations*:

[49] Compare, in particular, (72) and (73) for the definitions of $\mathcal{M}^i_k$ and $\check{\mathcal{M}}^i_k$; (85) for the definition of $\mathcal{B}^i_k$; (92) and (89) for the definition of $\mathcal{B}^i_k(\lambda)$; the definition of $\check{\mathcal{M}}^i_k(\lambda)$ is given (98).

Let Assumptions 2.1 and Definitions 2.3 hold; let $\mathbf{c}_{\mathrm{o}}$ be as in Theorem 2.1, and $\hat{\mathbf{c}}$ as in Theorem 2.4 with κ as in (56). Let $\mathtt{g}_1$ and $\mathtt{g}_3$ be as in (ii) of Theorem 2.2, and define

$$B_k^i := \begin{cases} \mathcal{B}_k^i\,, & \text{if } \ 0 < i < 2N_k\,, \\ j_{-\mathtt{g}_*}(\mathcal{B}_k^i)\,, & \text{if } i = 0, 2N_k\,, \end{cases} \qquad \mathtt{g}_* := -(\mathtt{g}_1 + \mathtt{g}_3)\,. \tag{108}$$

Then, the following result holds.

Theorem 3.1 (Secondary nearly–integrable structure at simple resonances) *There exists* $\mathbf{c}_* = \mathbf{c}_*(n, s, \beta, \delta) \geqslant \max\{c_2, \mathbf{c}_{\mathrm{o}}, \hat{\mathbf{c}}\}$ *such that if* $\mathtt{K}_{\mathrm{o}} \geqslant \mathbf{c}_*$, *then for any* $k \in \mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$, $0 \leqslant i \leqslant 2N_k$, *there exist real analytic symplectomorphisms*[50]

$$\phi_k^i : B_k^i \times \mathbb{T}^n \to \mathrm{Re}\,(\mathcal{R}^{1,k}_{r_k}) \times \mathbb{T}^n\,, \tag{109}$$

such that, if $\mathtt{E}_k^i = \mathtt{E}_k^i(I)$ *is the integrable Hamiltonian* $\mathtt{H}_k$ *of Theorem 2.2 in its Arnol'd–Liouville action variables,* $\widetilde{\mathtt{E}}_k^i := \mathtt{E}_k^i \circ j_{\mathtt{g}_*}$, *and* $\hat{h}_k$ *is as in Theorem 2.2, then*[51]

$$\begin{aligned} &\mathcal{H}_k^i := \mathtt{H} \circ \phi_k^i(I, \varphi) = h_k^i(I) + \varepsilon f_k^i(I, \varphi)\,, \quad \textit{with:} \\ &h_k^i := \tfrac{|k|^2}{2}\mathtt{h}_k^i\,, \ \mathtt{h}_k^i := \begin{cases} \mathtt{E}_k^i + \hat{h}_k\,, & \text{if } 0 < i < 2N_k\,, \\ \widetilde{\mathtt{E}}_k^i + \hat{h}_k & \text{if } i = 0, 2N_k\,. \end{cases} \end{aligned} \tag{110}$$

Furthermore, for $0 < \lambda \leqslant 1/\mathbf{c}_*$ *define:*

$$\begin{aligned} &\rho_* := \tfrac{\sqrt{\epsilon}}{\mathbf{c}_* \mathtt{K}_{\mathrm{o}}^n}\, \lambda|\log\lambda|\,, \quad \sigma_* := \tfrac{1}{\mathbf{c}_* \mathtt{K}_{\mathrm{o}}^n |\log\lambda|}\,, \\ &B_k^i(\lambda) := \begin{cases} \mathcal{B}_k^i(\lambda)\,, & \text{if } \ 0 < i < 2N_k\,, \\ j_{-\mathtt{g}_*}(\mathcal{B}_k^i(\lambda))\,, & \text{if } \ i = 0, 2N_k\,, \end{cases} \qquad \forall\ 0 \leqslant \lambda < 1/\mathbf{c}_*\,. \end{aligned} \tag{111}$$

Then, ϕ_k^i *admits a holomorphic extension*

$$\phi_k^i : (B_k^i(\lambda))_{\rho_*} \times \mathbb{T}^n_{\sigma_*} \to \mathcal{R}^{1,k}_{r_k} \times \mathbb{T}^n_{s_*} \tag{112}$$

and the perturbation f_k^i *in* (110) *satisfies the exponential estimate*

$$\sup_{(B_k^i(\lambda))_{\rho_*} \times \mathbb{T}^n_{\sigma_*}} |f_k^i| \leqslant e^{-\mathtt{K}s/3}\,. \tag{113}$$

Remark 3.2 (i) Notice that, since $\mu = 1/\mathtt{K}^{5n}$ (see (48)), and since[52]

$$\mathtt{K} > \mathtt{K}_{\mathrm{o}} \geqslant \mathbf{c}_* > \mathbf{c}\,,$$

condition (94) – which is stronger than condition (74) – is implied by the assumption $\mathtt{K}_{\mathrm{o}} \geqslant \mathbf{c}_*$.
Observe also that from the definitions of the constants in Theorem 3.1, Theorem 2.3 and from (91) it follows that

$$\mathbf{c}_* \geqslant \mathbf{c} \geqslant 2^8\kappa^3 \geqslant 2^{14}\,, \qquad \lambda_{\max} \geqslant 2^{12}/\mathbf{c}_*\,. \tag{114}$$

Finally, we remark that, recalling the definitions of ρ_λ and σ_λ in (96), since $\mathbf{c}_* \geqslant \hat{\mathbf{c}}$, one has

$$\rho_* < \rho_\lambda\,, \qquad \sigma_* < \sigma_\lambda\,. \tag{115}$$

(ii) In the proof of the theorem the maps ϕ_k^i are explicitly given; compare (120) and (126) below.

[50] Recall the notation (31).
[51] Recall (22) and (35).
[52] The constant $\mathbf{c}$ is defined in Theorem 2.3.

The following simple lemma will be one of the key points of the proof of Theorem 3.1. Recall Definition 3.1.

Lemma 3.1 *Let* $\Phi : (p,q) \in \mathrm{D} \times \mathbb{T}^n \mapsto (\eta, \hat{p}, q_1 + \psi, \hat{q} + \chi) \in \mathbb{R}^n \times \mathbb{T}^n$ *be in* $\mathfrak{G}$, $\Psi_{\mathrm{g}} \in \mathfrak{G}_{\dagger}$, *and denote by* $\mathcal{T}_{\mathrm{g}}\Phi$ *the map*

$$\mathcal{T}_{\mathrm{g}}\Phi := \mathcal{T}_{\mathrm{g}}\Phi(p,q) := \big(\eta_{\mathrm{g}} + \mathrm{g}, \hat{p}, q_1 + \psi_{\mathrm{g}}, \hat{q} + \chi_{\mathrm{g}} - \psi_{\mathrm{g}} \partial_{\hat{p}} \mathrm{g}\big)\,, \tag{116}$$

where for a function $u : D \times \mathbb{T} \to \mathbb{R}^m$, u_{g} *denotes te map*

$$u_{\mathrm{g}} := u \circ \check{\Psi}_{-g} : \jmath_{\mathrm{g}}(D) \times \mathbb{T} \to \mathbb{R}^m\,. \tag{117}$$

Then, $\mathcal{T}_{\mathrm{g}}\Phi$ *belongs to* $\mathfrak{G}$ *and it is a symplectomorphism satisfying*

$$\mathcal{T}_{\mathrm{g}}\Phi : \jmath_{\mathrm{g}}(\mathrm{D}) \times \mathbb{T}^n \xrightarrow{\text{onto}} \big(\check{\Psi}_{\mathrm{g}} \circ \Phi(\mathrm{D} \times \mathbb{T}^n)\big) \times \mathbb{T}^{n-1}\,, \tag{118}$$

and

$$(\mathcal{T}_{\mathrm{g}}\Phi)\check{} = (\eta_{\mathrm{g}} + \mathrm{g}, \hat{p}, q_1 + \psi_{\mathrm{g}}) = \check{\Psi}_{\mathrm{g}} \circ \check{\Phi} \circ \check{\Psi}_{-\mathrm{g}}\,. \tag{119}$$

Proof First observe that since $\eta_{\mathrm{g}}, \psi_{\mathrm{g}}, \chi_{\mathrm{g}}$ are 2π-periodic in q_1, the map

$$q \in \mathbb{T}^n \mapsto \pi_Q \mathcal{T}_{\mathrm{g}}\Phi(p,q) = \big(q_1 + \psi_{\mathrm{g}}, \hat{q} + \chi_{\mathrm{g}} - \psi_{\mathrm{g}} \partial_{\hat{p}} \mathrm{g}\big) \in \mathbb{T}^n$$

is a well defined $\mathbb{T}^n$–map and (119) follows immediately by direct computation. Thus, $(\mathcal{T}_{\mathrm{g}}\Phi)\check{}$ is injective being the composition of three injective maps, and, therefore, the whole map $\mathcal{T}_{\mathrm{g}}\Phi$ is injective, and (118) follows. To check symplecticity, just note that, locally, on the universal cover $\mathbb{R}^{2n}$, $\mathcal{T}_{\mathrm{g}}\Phi$ coincides (as it is immediate to check) with the composition $\Psi_{\mathrm{g}} \circ \Phi \circ \Psi_{-\mathrm{g}}$ of three symplectic maps. Hence $\mathcal{T}_{\mathrm{g}}\Phi$ is symplectic and the claim follows. ∎

Proof of Theorem 3.1 We start by defining the maps ϕ^i_k.
Consider, first, the inner case $0 < i < 2N_k$. Recall Definition 3.1. By Theorem 2.2–(iii), $\Phi_\star$ is the composition of maps in $\mathfrak{G}_{\dagger}$ and $\mathfrak{G}$ while, for $0 < i < 2N_k$, $\Phi^i \in \mathfrak{G}_0$ (Remark 3.1–(i)). Hence, by (103), it follows that $\Phi_\star \circ \Phi^i \in \mathfrak{G}_0$ and we may define[53]

$$\Phi^i_\star := \Phi_\star \circ \Phi^i\,, \quad \phi^i_k := \Psi^k \circ \Phi^i_\star : B^i_k \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n\,, \qquad (0 < i < 2N_k)\,, \tag{120}$$

provided the composition is well defined. To check that this is the case, we observe that by (102), (95), (51), (53), (115) for $0 < \lambda \leqslant 1/\hat{\mathbf{c}}$, we get

$$\check{\Phi}^i_\star = (\Phi_\star \circ \Phi^i)\check{} = \check{\Phi}_\star \circ \check{\Phi}^i : (\mathcal{B}^i_k(\lambda))_{\rho_\lambda} \times \mathbb{T}_{\sigma_\lambda} \to \mathscr{D}^k_{\tilde{r}_k} \times \mathbb{T}_{\check{\mathtt{s}}}\,, \quad (0 < i < 2N_k)\,, \tag{121}$$

thus the composition is well defined and (120) is well posed.
Let us now consider the outer case $i = 0, 2N_k$. In this case $\Phi^i \in \mathfrak{G}$ (Remark 3.1–(i)). Recalling the definition in (116)–(117), by Lemma 3.1, we may define

$$\Phi^i_{23} := \Phi_2 \circ \mathcal{T}_{\mathrm{g}_3}\Phi^i\,, \qquad \text{and} \qquad \Phi^i_\star := \mathcal{T}_{\mathrm{g}_1}\Phi^i_{23}\,, \qquad (i = 0, 2N_k)\,. \tag{122}$$

Recalling that $\Phi_2 \in \mathfrak{G}$, by Lemma 3.1 and Remark 3.1–(iv), $\Phi^i_{23} \in \mathfrak{G}$ and, again by Lemma 3.1, $\Phi^i_\star \in \mathfrak{G}$, provided the compositions are well defined. To check that this is the case, as above, it is enough to

[53] Ψ^k appears in Theorem 2.1. Recall that, when $0 < i < 2N_k$, $B^i_k := \mathcal{B}^i_k$.

control the complex domains of the first $(n+1)$ components. By (119) (used twice), (47), (57), and (102), one finds[54]

$$\check{\Phi}^i_* = \check{\Phi}_* \circ \check{\Phi}^i \circ \check{\Psi}_{\mathtt{g}_*}\,, \qquad\qquad (i=0,2N_k)\,. \tag{123}$$

Then, by (142), we get,

$$\jmath_{\mathtt{g}_*}\big((B^i_k(\lambda))_{\rho'_\lambda}\big) \subseteq \big(\jmath_{\mathtt{g}_*}\big(B^i_k(\lambda)\big)\big)_{\rho_\lambda} \overset{(111)}{=} \big(\mathcal{B}^i_k(\lambda)\big)_{\rho_\lambda}\,, \quad \text{where } \rho'_\lambda := \tfrac{\rho_\lambda}{n+2}\,, \qquad (i=0,2N_k)\,. \tag{124}$$

Observing that $\check{\Psi}_{\mathtt{g}_*}(p,q_1) = \big(\jmath_{\mathtt{g}_*}(p), q_1\big)$, by (123), (124), (95), (51) and (53), we get, for $0<\lambda\leqslant 1/\hat{\mathbf{c}}$,

$$\check{\Phi}^i_*:\ (B^i_k(\lambda))_{\rho'_\lambda} \times \mathbb{T}_{\sigma_\lambda} \ \to\ \mathscr{D}^k_{\bar{r}_k} \times \mathbb{T}_{\check{\mathtt{s}}}\,, \qquad (i=0,2N_k)\,. \tag{125}$$

Thus, the composition is well defined and (122) is well posed. So, we may define:

$$\phi^i_k := \Psi^k \circ \Phi^i_* : B^i_k \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n\,, \qquad \Phi^i_* \text{ as in } (122)\,, \qquad (i=0,2N_k)\,. \tag{126}$$

We can, now, prove (110). Recall the definition of $\bar{f}^k$ in Theorem 2.1 and define

$$f^i_k := f\circ\phi^i_k \overset{(37)}{=} \bar{f}^k \circ \Phi^i_*\,, \qquad\qquad (0\leqslant i\leqslant 2N_k)\,. \tag{127}$$

Then, by definition of ϕ^i_k in (120) and (126), we have, for $0\leqslant i\leqslant 2N_k$,

$$\mathcal{H}^i_k := \mathtt{H}\circ\phi^i_k(I,\varphi) := \mathtt{H}\circ\Psi^k\circ\Phi^i_* \overset{(37,127)}{=} \overline{\mathtt{H}}_k\circ\Phi^i_* + \varepsilon f^i_k\,. \tag{128}$$

Since $\overline{\mathtt{H}}_k$ in (38) depends only on the first $(n+1)$ variables, by (121) and (123), we find

$$\overline{\mathtt{H}}_k\circ\Phi^i_* = \overline{\mathtt{H}}_k\circ\check{\Phi}^i_* = \left\{\begin{array}{ll} \overline{\mathtt{H}}_k\circ\check{\Phi}_*\circ\check{\Phi}^i\,, & \text{if } 0<i<2N_k \\ \overline{\mathtt{H}}_k\circ\check{\Phi}_*\circ\check{\Phi}^i\circ\check{\Psi}_{\mathtt{g}_*}\,, & \text{if } i=0,2N_k\,, \end{array}\right. \tag{129}$$

and, by (53) and (105),

$$\overline{\mathtt{H}}_k\circ\check{\Phi}_*\circ\check{\Phi}^i = \tfrac{|k|^2}{2}\big(\mathtt{E}^{(i)}_k + \hat{h}_{{}_k}\big)\,. \tag{130}$$

Thus, (110) follows from (128), (129), (130) and (107).

Next, we show that ϕ^i_k has, for $0<\lambda\leqslant 1/\mathbf{c}_*$, a holomorphic extension satisfying (112). To do this we have to consider the last $n-1$ components of Φ^i_*, namely[55] $\pi_{\hat{\varphi}}\Phi^i_* = \hat{\Phi}^i_*$. By definition of Φ^i_* in (120) and (122) it follows that

$$\hat{\Phi}^i_*(I,\varphi) = \left\{\begin{array}{ll} \hat{\varphi} + \chi^i_*(I,\varphi_1)\,, & \text{if } 0<i<2N_k\,, \\ \hat{\varphi} + \chi^i_*\big(\jmath_{\mathtt{g}_*}(I),\varphi_1\big)\,, & \text{if } i=0,2N_k\,, \end{array}\right. \tag{131}$$

with[56]

$$\chi^i_* := \chi^i + \chi^\flat_2 + \psi^i\,\partial_{\hat{I}}\mathtt{g}_*\,, \qquad \chi^\flat_2(I,\varphi_1) := \left\{\begin{array}{ll} \chi_2(\hat{I},\psi^i)\,, & \text{if } 0<i<2N_k\,, \\ \chi_2(\hat{I},\varphi_1+\psi^i)\,, & \text{if } i=0,2N_k\,. \end{array}\right. \tag{132}$$

[54] Recall that $\mathtt{g}_* = -(\mathtt{g}_1+\mathtt{g}_3)$; compare (108).
[55] Recall the notation in (45).
[56] Recall the form of Φ^i in (87); $\mathtt{g}_*$ is defined in (108); χ_2 is as in Theorem 2.2–(iii).

Now, we claim that

$$|\psi^i|_{\rho_\lambda,\sigma_\lambda} < \frac{3}{4}\check{\mathtt{s}}\,, \qquad \forall\ 0 \leqslant i \leqslant 2N_k\,. \tag{133}$$

Indeed, if $0 < i < 2N_k$, (133) follows directly from (95) and (51); in the case $i = 0, 2N_k$, (133) follows again from (95) and (51) observing that

$$|\psi^i|_{\rho_\lambda,\sigma_\lambda} = |(\varphi_1 + \psi^i) - \varphi_1|_{\rho_\lambda,\sigma_\lambda} \leqslant \frac{\mathtt{s}}{4} + \sigma_\lambda < \frac{3}{4}\check{\mathtt{s}}\,.$$

Next, since $\rho'_\lambda = \rho_\lambda/(n+2)$, by (124), (132), (95), (51), (58), (100), (101), (133), (142), we find, for every $0 \leqslant i \leqslant 2N_k$, and for every $2 \leqslant \ell \leqslant n$,

$$\begin{aligned} |\operatorname{Im}\hat{\Phi}^i_{*\ell}|_{\rho'_\lambda,\sigma_\lambda} &\leqslant |\operatorname{Im}(\varphi_\ell + \chi^i_{*\ell})|_{\rho_\lambda,\sigma_\lambda} \\ &\leqslant |\operatorname{Im}(\varphi_\ell + \chi^i_\ell)|_{\rho_\lambda,\sigma_\lambda} + |\chi^\flat_2|_{\rho_\lambda,\sigma_\lambda} + |\psi^i|_{\rho_\lambda,\sigma_\lambda}\,|\partial_{\hat{I}}\mathtt{g}_*|_{\rho_\lambda} \\ &\leqslant \frac{\check{\mathtt{s}}}{2} + \frac{\check{\mathtt{s}}}{2^{20}} + \frac{3}{4}(n+1)\check{\mathtt{s}} < 2n\check{\mathtt{s}}\,. \end{aligned} \tag{134}$$

Thus, by (121), (125) and (134), we get

$$\Phi^i_*\,:\,(B^i_k(\lambda))_{\rho'_\lambda} \times \mathbb{T}^n_{\sigma_\lambda} \ \to\ \mathscr{D}^k_{\tilde{r}_k} \times \mathbb{T}^n_{2n\check{\mathtt{s}}}\,, \qquad (0 \leqslant i \leqslant 2N_k)\,.$$

We need, now, an elementary result on real analytic functions, whose proof is given in Appendix:

Lemma 3.2 *Let $g : D_r \times \mathbb{T}^n_s \to \mathbb{C}$ be a real analytic function satisfying $|\operatorname{Im} g| \leqslant \xi$. Then, for every $0 < \zeta \leqslant 1/2$, one has*

$$\sup_{D_{\zeta r}\times\mathbb{T}^n_{\zeta s}} |\operatorname{Im} g| \leqslant 8\zeta\xi\,.$$

Now, define

$$\zeta := \frac{1}{16n\,c_1\,\mathtt{c}_s\,\mathtt{K}_{\mathtt{o}}^n}\,. \tag{135}$$

Then, since $|k| \leqslant \mathtt{K}_{\mathtt{o}}$, by (53), (48), we find

$$8\zeta(2n\check{\mathtt{s}}) < 16n\,\zeta\,\mathtt{K}_{\mathtt{o}}\max\{1,s\} \overset{(135)}{=} \frac{\max\{1,s\}}{c_1\ \mathtt{c}_s\,\mathtt{K}_{\mathtt{o}}^{n-1}} = \frac{s}{c_1\mathtt{K}_{\mathtt{o}}^{n-1}} \overset{(32)}{=} \tilde{s}_k\,.$$

Thus, by Lemma 3.2 (applied with $g = \hat{\Phi}^i_{*\ell}$ for $2 \leqslant \ell \leqslant n$, ζ as in (135) and $\xi = 2n\check{\mathtt{s}}$), it follows that

$$\Phi^i_*\,:\,(B^i_k(\lambda))_{\rho_*} \times \mathbb{T}^n_{\sigma_*} \ \to\ \mathscr{D}^k_{\tilde{r}_k} \times \mathbb{T}^n_{\tilde{s}_k}\,, \qquad (0 \leqslant i \leqslant 2N_k)\,,$$

with ρ_* and σ_* as in (111), provided

$$\mathbf{c}_* := \max\{c_2\,, \mathbf{c_o}\,, \hat{\mathbf{c}}\,c_1\,\mathtt{c}_s\,16n\,(n+2)\}\,.$$

In conclusion, (112) follows by the definition of ϕ^i_k in (120), (126) and by (35).

Finally, estimate (113) follows at once from (127), (112) and (39). The proof is complete. ∎

The following measure estimate will play a crucial rôle in the proof of Theorem 1.1.

Proposition 3.1 *For every $0 \leqslant \lambda < 1/\mathbf{c}_*$, the following measure estimate holds*[57]*:*

$$\mathrm{meas}\left((\mathcal{R}^{1,k} \times \mathbb{T}^n) \setminus \bigcup_{0 \leqslant i \leqslant 2N_k} \phi_k^i\big(B_k^i(\lambda) \times \mathbb{T}^n\big) \right) \leqslant \mathbf{c}_* \,\mathrm{meas}\left(\tilde{\mathcal{R}}^{1,k} \times \mathbb{T}^n\right) \lambda |\log \lambda| \,. \tag{136}$$

Proof Since $\check{\Phi}_*^i$ depends only on the first $(n+1)$ variables, by (131), (120), (123) and the definitions of $B_k^i(\lambda)$ in (111) and $\check{\mathcal{M}}_k^i(\lambda)$ in (104), one has

$$\begin{aligned} \Phi_*^i\big(B_k^i(\lambda) \times \mathbb{T}^n\big) &= \check{\Phi}_*^i\big(B_k^i(\lambda) \times \mathbb{T}\big) \times \mathbb{T}^{n-1} = \big(\check{\Phi}_* \circ \check{\Phi}^i(\mathcal{B}_k^i(\lambda) \times \mathbb{T})\big) \times \mathbb{T}^{n-1} \\ &\overset{(99)}{=} \big(\check{\Phi}_* \circ \check{\mathcal{M}}_k^i(\lambda)\big) \times \mathbb{T}^{n-1} \,. \end{aligned} \tag{137}$$

Analogously, one has

$$\Phi_*^{-1}(\mathscr{D}^k \times \mathbb{T}^n) = \check{\Phi}_*^{-1}(\mathscr{D}^k \times \mathbb{T}) \times \mathbb{T}^{n-1} \,. \tag{138}$$

Observe also that, by (59), (58) and (the second estimate in) (100) it follows that[58]

$$\check{\Phi}_3^{-1} \circ \check{\Phi}_2^{-1}(D \times \mathbb{T}) \subseteq \big((-\mathtt{R} - \mathtt{r}/3, \mathtt{R} + \mathtt{r}/3) \times \hat{D}\big) \times \mathbb{T} = (D^\flat \times \mathbb{T}) \,. \tag{139}$$

Then[59], recalling Theorem 2.1, using the fact that $(\Psi^k)^{-1}$ and Φ_*^{-1} are diffeomorphysms preserving Liouville measure, we find

$$\begin{aligned} &\mathrm{meas}(\mathcal{R}^{1,k} \times \mathbb{T}^n \setminus \textstyle\bigcup \phi_k^i(B_k^i(\lambda) \times \mathbb{T}^n)) \\ &\overset{(120,126)}{=} \mathrm{meas}((\Psi^k)^{-1}(\mathcal{R}^{1,k} \times \mathbb{T}^n) \setminus \textstyle\bigcup \Phi_*^i(B_k^i(\lambda) \times \mathbb{T}^n)) \\ &\overset{(40)}{\leqslant} \mathrm{meas}((\mathscr{D}^k \times \mathbb{T}^n) \setminus \textstyle\bigcup \Phi_*^i(B_k^i(\lambda) \times \mathbb{T}^n)) \\ &= \mathrm{meas}(\Phi_*^{-1}(\mathscr{D}^k \times \mathbb{T}^n) \setminus \textstyle\bigcup \Phi_*^{-1}\Phi_*^i(B_k^i(\lambda) \times \mathbb{T}^n)) \\ &\overset{(137,138)}{=} (2\pi)^{n-1}\,\mathrm{meas}(\check{\Phi}_*^{-1}(\mathscr{D}^k \times \mathbb{T}) \setminus \textstyle\bigcup \check{\mathcal{M}}_k^i(\lambda)) \\ &\overset{(57,63)}{=} (2\pi)^{n-1}\,\mathrm{meas}(\check{\Phi}_3^{-1} \circ \check{\Phi}_2^{-1}(D \times \mathbb{T}) \setminus \textstyle\bigcup \check{\mathcal{M}}_k^i(\lambda)) \\ &\overset{(139)}{\leqslant} (2\pi)^{n-1}\,\mathrm{meas}(D^\flat \times \mathbb{T} \setminus \textstyle\bigcup \check{\mathcal{M}}_k^i(\lambda)) \\ &\overset{(99)}{\leqslant} (2\pi)^{n-1} \hat{\mathbf{c}}\, \sqrt{\epsilon}\, \mathrm{meas}(\hat{D})\ \lambda |\log \lambda| \\ &\overset{(50)}{<} (2\pi)^{n-1} \hat{\mathbf{c}}\ \mathtt{R}\ \mathrm{meas}(\hat{D})\ \lambda |\log \lambda| \\ &\overset{(62)}{=} \frac{\hat{\mathbf{c}}}{2\pi}\ \mathrm{meas}(\tilde{\mathcal{R}}^{1,k} \times \mathbb{T}^n)\ \lambda |\log \lambda| \,, \end{aligned}$$

which yields (136) since $\mathbf{c}_* \geqslant \hat{\mathbf{c}}$. ∎

Remark 3.3 The measure estimate (136) holds in view of the covering property (40), which takes care of the deformations near the boundaries.
The logarithmic correction is unavoidable and is related to the Lyapunov exponents of the hyperbolic equilibria issuing the separatrices of the secondary integrable systems at simple resonances.

---

[57]The sets $\tilde{\mathcal{R}}^{1,k}$ are defined in (34).

[58]Observe that $\check{\Phi}_3^{-1}(p, q) = (p_1 - \mathtt{g}_3(\hat{p}), \hat{p}, q_1)$ and $\check{\Phi}_2^{-1}(p, q) = (p_1 - \eta_2(\hat{p}, q_1), \hat{p}, q_1)$. Recall the definition of $D^\flat$ in (98).

[59]The unions are over $0 \leqslant i \leqslant 2N_k$.

The final result of this section deals with the size of the domains B^i_k, which depends on k and actually *grows* with k. It is therefore important to control such a growth.

Proposition 3.2 *Assume that*[60] $\alpha < 1$. *Then, there exists a constant* $c_* = c_*(n) > 1$ *such that*

$$\operatorname{diam} B^i_k \leqslant c_* |k|^{n-1}\,, \qquad \operatorname{meas} B^i_k \leqslant c_*\,. \tag{140}$$

Proof For the purpose of this proof, we denote by 'c' suitable (possibly different) constants greater than one and depending only on n.
Since $\alpha < 1$, by the definition of $\mathcal{B}^i_k$ in (85), by (44) and the definition of $\mathtt{R}$ in (48), we have, for every $0 \leqslant i \leqslant 2N_k$,

$$\operatorname{diam} \mathcal{B}^i_k \leqslant c\big(\mathtt{R} + \operatorname{diam} \hat{D}\big) \leqslant c\big(\tfrac{\alpha}{|k|^2} + \operatorname{diam} \hat{D}\big) < c(1 + \operatorname{diam} \hat{D})\,.$$

By (61), (33) and (64) it follows that

$$|\,\pi^\perp_k \hat{\mathrm{A}}^T \hat{I}| = |\mathrm{A}^T \mathrm{U}\tbinom{0}{\hat{I}}| \geqslant \tfrac{|\hat{I}|}{\|\mathrm{A}^{-1}\|\,\|\mathrm{U}^{-1}\|} \geqslant \tfrac{|\hat{I}|}{c|k|^{n-1}}\,,$$

and since by (48) $\hat{D} \subseteq \{\hat{I} \in \mathbb{R}^{n-1} : \ |\,\pi^\perp_k \hat{\mathrm{A}}^T \hat{I}| < 1\}$, it follows that $\operatorname{diam} \mathcal{B}^i_k \leqslant c|k|^{n-1}$, proving the first relation in (140) in the case $0 < i < 2N_k$.
In the case $i = 0, 2N_k$, we need to estimate the Lipschitz constant of[61] $\mathtt{g}_i$. The map $\mathtt{g}_1$ is linear and its gradient is given by $\hat{\mathrm{A}}k/|k|^2$, thus, by (33) one gets

$$|\partial_{\hat{I}} \mathtt{g}_1| = \big|\tfrac{\hat{\mathrm{A}}k}{|k|^2}\big| \leqslant n\,.$$

By (59), recalling that $|\chi_k| \leqslant 1$, the definitions in (48), and by Cauchy estimates[62], one sees that

$$|\mathtt{g}_3|_{4\mathtt{r}} < \tfrac{2\varepsilon}{|k|^2}\,\tfrac{\mu}{\mathtt{r}} < \tfrac{c_2 \mathbf{c_0}}{2}\,\tfrac{\sqrt{\varepsilon}}{\mathtt{K}^{14n+5}}\,, \qquad |\partial_{\hat{I}} \mathtt{g}_3|_{3\mathtt{r}} \leqslant \tfrac{2\varepsilon\mu}{\mathtt{r}|k|^2\mathtt{r}^2} \leqslant c\,\tfrac{\mathbf{c_0}}{\mathtt{K}^{14n+3}} < \tfrac14\,, \tag{141}$$

by taking $\mathtt{K}_{\rm o}$ big enough (recall that $\mathtt{K} \geqslant 6\mathtt{K}_{\rm o}$). Hence[63],

$$|\partial_{\hat{I}} \mathtt{g}_*|_{3\mathtt{r}} \leqslant n+1\,, \qquad \operatorname{Lip}_{D_{3\mathtt{r}}}(j_{\mathtt{g}_*}) \leqslant n+2\,, \tag{142}$$

and, choosing c_* suitably, the first relation in (140) follows also in this case.
Let us check the second relation in (140). Since ϕ^i_k in (109) is symplectic, we have

$$\begin{aligned}\operatorname{meas} B^i_k \quad &= \quad \frac{1}{(2\pi)^n}\operatorname{meas}(B^i_k \times \mathbb{T}^n) = \frac{1}{(2\pi)^n}\operatorname{meas}\big(\phi^i_k(B^i_k \times \mathbb{T}^n)\big)\\ &\overset{(109)}{\leqslant} \quad \operatorname{meas}\big(\operatorname{Re}(\mathcal{R}^{1,k}_{r_k})\big)\,.\end{aligned}$$

Now, since $\mathcal{R}^{1,k} \subseteq \mathrm{B}$ and $r_k \leqslant \alpha < 1$, choosing c_* suitably, also the second relation in (140) follows, and claim (i) has been proved. ∎

---

[60] Notice that, since $\gamma = 2(\nu + n)$, the hypothesis $\alpha = \varepsilon \mathtt{K}^\nu < 1$ is implied by the second condition in (11).
[61] $\mathtt{g}_*$ is defined (108).
[62] Compare, e.g., [16].
[63] $\operatorname{Lip}_B(g)$ denotes absolute value of the Lipschitz constant of a function g over a domain B.

4 Twist at simple resonances

In this section – which is the heart of the paper – we discuss the main issue in singular KAM theory, namely, the twist of the integrable (rescaled) secular Hamiltonians $\mathtt{h}_k^i$ in (110) near simple resonances and, in particular, in neighborhoods of secular separatrices, where the action become singular.

In general, it has to be expected that *there are points where the twist of the secular Hamiltonians* $\mathtt{h}_k^i$ *vanishes*; compare Remark 4.1 below. Furthermore, and more importantly, when approaching separatrices, the evaluation of the twist becomes a *singular perturbation problem*, where no standard tools can be applied and a new strategy is needed.

Our approach – which exploits in an essential way the fine analytic structure of the action functions described in Theorem 2.3 – roughly speaking, consists in constructing a suitable differential operator with non–constant coefficients, which does not vanish on (a suitable regularization of) the Kolmogorov's twist determinant. This will be enough to prove that *the Liouville measure of the set where the twist is smaller than a positive quantity* η *may be bounded, uniformly in* k*, by a power of* η. This is the content of the Twist Theorem 4.1 below, from which the proof of the results described in § 1 will follow easily.

Remark 4.1 (Points where the twist vanishes) First, let us consider a region bounded by separatrices, i.e., (in the above setting) the case when i is even and different from 0 and $2N$. From (92), (75) and (78) there follows that $\partial_E I_1^i \to +\infty$ as E approaches $E_\pm^i$. Thus, since $\partial_E I_1^i > 0$ (always), $E \to \partial_E^2 I_1^i$ must have at least one zero in (E_-^i, E_+^i). Since, by the chain rule,

$$\partial_{I_1}^2 \mathtt{E}^i(I_1) = -\left.\frac{\partial_E^2 I_1^i}{(\partial_E I_1^i)^3}\right|_{\mathtt{E}^i(I_1)}, \tag{143}$$

we see that $\partial_{I_1}^2 \mathtt{E}^i$ must vanish at some points in the interval (a^i, b^i) defined in (92).

Let us next consider the case i odd, i.e., regions whose closure contains an elliptic point. Let us first consider the case $\mu = 0$, and let us denote $\bar a^i = a^i|_{\mu=0}$ and $\bar b^i = b^i|_{\mu=0}$. As above, by (78), the function $E \to \partial_E^2 \bar I_1^i$ tends to $+\infty$ when $E \to \bar E_+^i$. Thus, by (143), $\partial_{I_1}^2 \bar{\mathtt{E}}^i(I_1)$ is negative when I_1 is close to $\bar b^i$. Now, $\bar{\mathtt{E}}^i(I_1)$ is analytic at $I_1 = \bar a^i = 0$, and, evaluating the Birkhoff normal form of $p_1^2 + \bar{\mathtt{G}}(q_1)$ at order 4 close to the elliptic point $(p_1, q_1) = (0, \bar\theta_i)$, one sees that

$$\bar{\mathtt{E}}^i(I_1) := \omega_0 I_1 + \frac{1}{2} c I_1^2 + O(I_1^3), \quad \text{with} \quad \omega_0 = \sqrt{2d_2}, \quad c = \frac{1}{4}\left(\frac{d_4}{d_2} - \frac{5d_3^2}{3d_2^2}\right),$$

where d_j are the j-th order derivatives of the reference potential $\bar{\mathtt{G}}$ evaluated at the minimum $\bar\theta_i$. Thus, $\partial_{I_1}^2 \bar{\mathtt{E}}^i(0) > 0$ whenever the condition

$$\delta := 3d_2 d_4 - 5d_3^2 > 0, \qquad d_j := (\partial_{q_1}^j \bar{\mathtt{G}})(\bar\theta_i), \tag{144}$$

is satisfied, in which case $\partial_{I_1}^2 \bar{\mathtt{E}}^i$ must vanish at some point in $\big(0, \bar b^i\big)$. By (53), $\hat h_k = \hat{\mathcal{Q}}_k$, so that $\mathtt{h}_k^i|_{\mu=0} = \bar{\mathtt{E}}_k^i(I_1) + \hat{\mathcal{Q}}_k(\hat I)$, which implies

$$\det \partial_I^2 \mathtt{h}_k^i(I)|_{\mu=0} = \partial_{I_1}^2 \bar{\mathtt{E}}_k^i(I_1) \cdot \det \partial_{\hat I}^2 \hat{\mathcal{Q}}_k(\hat I).$$

Thus, by continuity, for μ small enough it follows that the Hessian matrix $\partial_I^2 \mathtt{h}^i(I)$ is singular at some point.

Condition (144) is easily satisfied. For example, if $\bar{\mathtt{G}}(\theta) = \cos\theta - \frac{1}{8}\cos(2\theta)$, one finds that $\updelta = 3/2$, so that, in this very simple cases, inside the (unique) region enclosed by the main separatrices, there are points where the twist vanishes. However, this is not the case if the potential is close enough to a cosine, compare Proposition 2.1.

Twist Theorem near simple resonances (statement)

To state the Twist Theorem we need to introduce two parameters ($\upxi > 0$, $\mathtt{m} \geqslant 1$) which measure the non–degeneracy (in a suitable sense to be specified below) of the energy as function of actions in the inner regions $0 < i < 2N_k$. This requires some preparation.

Non–degenerate functions and theirs sub–levels

First, let us recall a standard quantitative definition of non–degenerate functions.

Definition 4.1 *Given $\xi > 0$, an open set $A \subseteq \mathbb{R}$ and $f \in C^m(A, \mathbb{R})$, we say that f is ξ–non–degenerate at order $m \geqslant 1$ on A (or, in short, (ξ, m)–non–degenerate), if*

$$\inf_{x \in A} \max_{1 \leqslant j \leqslant m} |f^{(j)}(x)| \geqslant \xi \,. \tag{145}$$

An important property of non–degenerate functions is that one can easily estimate the measure of their sub–levels:

Lemma 4.1 *Let f be a (ξ, m)–non–degenerate function on a bounded interval (a, b) and let*[64] *$M := \|f\|_{C^{m+1}(a,b)}$. Then, there exist a constants $c_m > 1$ depending only on m such that, for all $\eta > 0$, one has*

$$\mathrm{meas}\{x \in (a, b) : |f(x)| \leqslant \eta\} \leqslant \frac{c_m}{\xi^{1/m}} \big(\tfrac{M}{\xi}(b - a) + 1\big)\, \eta^{1/m} \,.$$

The proof of this lemma can be found, e.g., in [23, Lemma B.1]; compare, also, [37].

Non–degeneracy of the rescaled reference potentials for $|k|_1 \leqslant \mathtt{N}$

Consider a general Hamiltonian (42) in standard form, recall Definition 2.4, recall (92), and define also, for $0 \leqslant \uplambda \leqslant \bar{\uplambda}_{\max}$ (defined in (89)),

$$\bar{a}^i := a^i|_{\mu=0} \,, \quad \bar{b}^i := b^i|_{\mu=0} \,, \quad \bar{a}^i_{\uplambda} := a^i_{\uplambda}|_{\mu=0} \,, \quad \bar{b}^i_{\uplambda} := b^i_{\uplambda}|_{\mu=0} \,, \quad \forall\, 0 \leqslant i \leqslant 2N_k \,. \tag{146}$$

In the following, we shall explicitly indicate the dependence upon the reference potential $\bar{\mathtt{G}}$ and write, e.g, $\bar{I}^i_{1,\bar{\mathtt{G}}}$, $\bar{\mathtt{E}}^i_{\bar{\mathtt{G}}}$, $\bar{a}^i_{\bar{\mathtt{G}}}$, $\bar{b}^i_{\bar{\mathtt{G}}}$ for $\bar{I}^i_1$, $\bar{\mathtt{E}}^i$, $\bar{a}^i$, $\bar{b}^i$, respectively.

Definition 4.2 *Given $\mathtt{H}_\flat$ in standard form with reference potential $\bar{\mathtt{G}}$, we denote by*

$$\mathtt{F}^i_{\bar{\mathtt{G}}}(x) := (\partial^2_{I_1} \bar{\mathtt{E}}^i_{\bar{\mathtt{G}}})\big(\bar{a}^i_{\bar{\mathtt{G}}} + (\bar{b}^i_{\bar{\mathtt{G}}} - \bar{a}^i_{\bar{\mathtt{G}}})x\big) \,, \quad \forall\ x \in (0, 1) \,, \qquad (0 < i < 2N_k) \,, \tag{147}$$

the 'normalized second derivative of the energy function within separatrices'.

These functions satisfy a remarkable rescaling property:

[64] $\|f\|_{C^{m+1}(a,b)} := \max_{0 \leqslant j \leqslant m+1} \sup_{(a,b)} |f^{(j)}|$.

Lemma 4.2 *If* $\mathtt{F}^i_{\bar{\mathtt{G}}}$ *is as in Definition 4.2, then, for any* $\lambda > 0$*, one has* $\mathtt{F}^i_{\bar{\mathtt{G}}} = \mathtt{F}^i_{\lambda\bar{\mathtt{G}}}$.

Proof Indeed, from the definition of actions, there follows easily that

$$\bar{I}^i_{1,\lambda\bar{\mathtt{G}}}(E) = \sqrt{\lambda}\bar{I}^i_{1,\bar{\mathtt{G}}}(E/\lambda)\,, \qquad \bar{\mathtt{E}}^i_{\lambda\bar{\mathtt{G}}}(I_1) = \lambda\bar{\mathtt{E}}^i_{\bar{\mathtt{G}}}(I_1/\sqrt{\lambda})\,, \qquad \forall\lambda > 0\,. \tag{148}$$

Indeed, considering the case $i = 2N_k$ (the other cases being similar), one has

$$\bar{I}^{2N_k}_{1,\lambda\bar{\mathtt{G}}}(E) \overset{(250)}{=} \frac{1}{2\pi}\int_0^{2\pi}\sqrt{E-\lambda\bar{\mathtt{G}}(x)}dx = \frac{\sqrt{\lambda}}{2\pi}\int_0^{2\pi}\sqrt{\frac{E}{\lambda}-\bar{\mathtt{G}}(x)}dx = \sqrt{\lambda}\bar{I}^i_{1,\bar{\mathtt{G}}}(E/\lambda)\,,$$

which proves the first equality in (148), which, in turns, implies immediately the second inequality. From (148), then , follows that

$$\bar{a}^i_{\lambda\bar{\mathtt{G}}} = \sqrt{\lambda}\bar{a}^i_{\bar{\mathtt{G}}}\,, \qquad \bar{b}^i_{\lambda\bar{\mathtt{G}}} = \sqrt{\lambda}\bar{b}^i_{\bar{\mathtt{G}}}\,, \tag{149}$$

and the claim follows at once from (148) and (149). ∎

Let us go back to the Hamiltonians in standard form $\mathtt{H}_k$ of Theorem 3.1, and let us prove that the functions $\mathtt{F}^i_{\bar{\mathtt{G}}}$ – and hence $\bar{\mathtt{E}}^i_{\lambda\bar{\mathtt{G}}}$ – with $\bar{\mathtt{G}}$ as in (55), are (ξ, m)–non–degenerate.

Lemma 4.3 *For every* $0 < i < 2N_k$*, the function* $\mathtt{F}^i_{\bar{\mathtt{G}}}$ *defined in* (147) *is* (ξ, m)*–non–degenerate for some* $\xi, m > 0$.

Proof We consider only the case i odd, the even case being similar. Deriving (143) we get, for $\mu = 0$,

$$\partial^3_{I_1}\bar{\mathtt{E}}^i(\bar{I}^i_1(E)) = -\frac{\partial^3_E\bar{I}^i_1(E)}{\big(\partial_E\bar{I}^i_1(E)\big)^4} + 3\frac{\big(\partial^2_E\bar{I}^i_1(E)\big)^2}{\big(\partial_E\bar{I}^i_1(E)\big)^5}\,. \tag{150}$$

By (75)–(80) (which hold also for $\bar{I}^i_1$, corresponding to $\mu = 0$), we have that the dominant term in (150) as $z := (\bar{E}^i_+ - E)/\epsilon \to 0^+$ has the form $-1/(\mathtt{c}^3 z^2\log^4 z)$ with $\mathtt{c} := \psi^i_+(0)|_{\mu=0}$. Then,

$$\lim_{E\to(\bar{E}^i_+)^-}\big|\partial^3_{I_1}\bar{\mathtt{E}}^i(\bar{I}^i_1(E))\big| = \lim_{I_1\to(\bar{b}^i)^-}\big|\partial^3_{I_1}\bar{\mathtt{E}}^i(I_1)\big| = +\infty\,.$$

By (147) we obtain

$$\lim_{x\to 1^-}|\partial_x\mathtt{F}^i_{\bar{\mathtt{G}}}(x)| = +\infty\,. \tag{151}$$

Moreover $\partial_x\mathtt{F}^i_{\bar{\mathtt{G}}}(x)$ is analytic in a neighborhood of $x = 0$ (recall in particular (79)). Assume now by contradiction that (145) does not hold, namely that there exists a sequence $x_m \in (0,1)$ such that

$$|\partial^j_x\mathtt{F}^i_{\bar{\mathtt{G}}}(x_m)| < 1/m\,, \qquad \forall\, 1 \leqslant j \leqslant m\,.$$

By (151), up to a subsequence, x_m converges to some $\bar{x} \in [0,1)$ such that $\partial^j_x\mathtt{F}^i_{\bar{\mathtt{G}}}(\bar{x}) = 0$ for every $j \geqslant 1$. By analyticity we would have that $\mathtt{F}^i_{\bar{\mathtt{G}}}$ is constant on $[0,1)$ leading to a contradiction with (151). ∎

This lemma allows us to introduce *uniform* non–degeneracy parameters $\upxi > 0$ and $\mathtt{m} \geqslant 1$ for the function $\mathtt{F}^i_{\bar{\mathtt{G}}}$ in (147) associated to the reference potentials $\bar{\mathtt{G}} \overset{(55)}{=} \frac{2\varepsilon}{|k|^2}\pi_{\mathbb{Z}k}f$, for $k \in \mathcal{G}^n$, $|k|_1 < \mathtt{N}$ and $0 < i < 2N_k$. Indeed, by Lemma 4.2,

$$\mathtt{F}^i_{\bar{\mathtt{G}}} = \mathtt{F}^i_{\frac{2\varepsilon}{|k|^2}\pi_{\mathbb{Z}k}f} = \mathtt{F}^i_{\pi_{\mathbb{Z}k}f}\,, \tag{152}$$

and, by (17), every potential $\pi_{\mathbb{Z}k} f$ is β–Morse. By the above Lemma 4.3, every function in (152) is (ξ, m)–non–degenerate for some $\xi, m > 0$. We therefore can define uniform ε–independent non–degeneracy parameters $\mathtt{\xi}$, $\mathtt{m}$ by setting:

Definition 4.3 *Let $\mathtt{F}^i_{\pi_{\mathbb{Z}k} f}$ be as in Definition 4.2 with rescaled reference potential $\bar{\mathtt{G}} = \pi_{\mathbb{Z}k} f$. We define $\mathtt{\xi} > 0$ and $\mathtt{m} \geqslant 1$ to be, respectively, the largest and smallest number such that all the functions $\mathtt{F}^i_{\pi_{\mathbb{Z}k} f}$, for $0 < i < 2N_k$, $k \in \mathcal{G}^n$ with $|k|_1 \leqslant \mathtt{N}$, are $(\mathtt{\xi}, \mathtt{m})$–non–degenerate (Definition 4.1).*

The Twist Theorem

Let Assumptions 2.1 and Definitions 2.3 hold, let $\mathtt{\kappa}$ be as in (56), let $\mathtt{\xi}, \mathtt{m}$ be as in Definition 4.3, let B^i_k be as in (108), let $\mathtt{h}^i_k$ be as in (110), and define

$$\delta_{\rm o} := |k|^{-2n} \,. \tag{153}$$

Then, the following result holds.

Theorem 4.1 *There exists a constant $\mathtt{c}_{\rm o} = \mathtt{c}_{\rm o}(n, \mathtt{\kappa}, \mathtt{\xi}, \mathtt{m}) > 1$ such that, for $\mathtt{K}_{\rm o} \geqslant \mathtt{c}_{\rm o}$, $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, $0 \leqslant i \leqslant 2N_k$, and $0 < \eta < \delta_{\rm o}/2^5$, one has:*

$$\mathrm{meas}\big(\big\{\, I \in B^i_k : \ \big|\det \partial^2_I \mathtt{h}^i_k(I)\big| \leqslant \eta\big\}\big) \leqslant \mathtt{c}_{\rm o}(|k|^{2n}\eta)^{\mathtt{b}}\, \mathrm{meas}\, B^i_k \,, \qquad \mathtt{b} := \min\{\tfrac{1}{9n^4}\,,\, \tfrac{1}{\mathtt{m}}\} \,. \tag{154}$$

Theorem 4.1 will be proven in several steps:

Step 1: *Preliminaries*
(a) Explicit expressions for the twist matrix in the inner case ($0 < i < 2N_k$) are given;
(b) analogous formulae are given for the outer case ($i = 0, 2N_k$), but, due to the presence of the translation $j_{-\mathtt{g}_\star}$ in (106), the measure estimate is expressed in terms of the domains $\mathcal{B}^i_k$ rather than the domains B^i_k (recall that such domains differ in the outer case; compare (108));
(c) uniform estimates on the sub–matrix $\partial^2_{\hat{I}} \hat{h}_k$ of order $(n-1)$, depending only on the 'trivial actions' $\hat{I}$, are given.

Step 2: *Coverings of the phase space into regions close to separatrices and far from separatrices*
This is a necessary step, since the analysis will be non perturbative near separatrices, while in regions away from separatrices, the analysis will be partly perturbative (and significantly simpler).

Step 3*: *Non–degeneracy of the twist function in neighborhoods of separatrices*
In such regions perturbative arguments do not hold, and, in particular the energy function $\mathtt{E}^i$ is singular at the boundary (corresponding to separatrices) and its derivatives diverge as the boundary is approached. Furthermore, $\mathtt{E}^i$ and $\bar{\mathtt{E}}^i = \mathtt{E}^i|_{\mu=0}$ have singularities in different points. Exploiting the singularity structure described in Theorem 2.3, we will prove that a suitable regularization of the twist determinant is a non–degenerate function allowing to control the measure of its sub–levels. This is the core of the proof.

Step 4: *The Twist Theorem in neighborhoods of separatrices*
By the previous step, measure estimates in regions close to separatrices follow easily, yielding the proof of the Twist Theorem in this case.

Step 5: *The Twist Theorem far from separatrices in the inner case*
It is here (in particular, in the low mode case $|k|_1 \leqslant \mathtt{N}$) that the non–degeneracy condition involving the parameters $\mathtt{\xi}$ and $\mathtt{m}$, is needed.

Step 6: *Uniform twist in outer regions far from separatrices*
In such regions there is uniform twist; the proof rests on a simple argument based on Jensen's inequality.

Step 7: *Conclusion of the proof of the Twist Theorem*

Proof of the Twist Theorem

Fix $k \in \mathcal{G}^n_{\mathrm{K}_\mathrm{o}}$, $0 \leqslant i \leqslant 2N_k$, and $\eta > 0$.

Throughout the proof the $(n-1)$ dimensional domain $\hat{D}$ (defined in (48)) will be kept fixed and often the variables $\hat{I}$ will not be indicated explicitly. Also, the label k will usually be omitted it in the notation, as well as the suffix i (when this does not lead to confusion).

Step 1 *Preliminaries*

(a) We give the analytic expression of the twist determinant inside separatrices, i.e., for $0 < i < 2N$. Recall that in this case, by (110) and (108), one has $\mathtt{h}^i = \mathtt{E}^i + \hat{h}_{\scriptscriptstyle k}$ and $B^i = \mathcal{B}^i_k$. Then[65]

$$\det \partial^2_I \mathtt{h}^i = \det\left(\partial^2_I \mathtt{E}^i + \partial^2_I \hat{h}_{\scriptscriptstyle k}\right) = \det \begin{pmatrix} \partial^2_{I_1}\mathtt{E}^i & \partial^T_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) \\ \partial_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) & \partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{\hat{I}}\hat{h}_{\scriptscriptstyle k} \end{pmatrix} \tag{155}$$
$$= (\partial^2_{I_1}\mathtt{E}^i)\cdot\det(\partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{\hat{I}}\hat{h}_{\scriptscriptstyle k}) + \det \begin{pmatrix} 0 & \partial^T_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) \\ \partial_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) & \partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{\hat{I}}\hat{h}_{\scriptscriptstyle k} \end{pmatrix} .$$

(b) We now consider the case outside the outer separatrices, i.e., $i = 0, 2N$.
The Hamiltonian $\mathtt{h}^i(I)$, in this case, is given by[66] $\mathtt{h}^i(I) = \widetilde{\mathtt{E}}^i(I) + \hat{h}_{\scriptscriptstyle k}(\hat{I})$ for $I \in B^i = \jmath_{-\mathtt{g}_*}(\mathcal{B}^i_k)$. Recalling (53) and (59) we note that in the evaluation of the Hessian of $\mathtt{h}$ involves the *non–small linear term* $\frac{(\hat{A}k)\cdot\hat{I}}{|k|^2}$, a fact that complicates analytic expressions. However, such complications may be avoided, using the following trick.
Let us introduce new action variables I, defined by the relation $I = \mathrm{U}\mathrm{I} = \jmath_{\mathtt{g}_1}(\mathrm{I})$, where U is defined in (60). Then, we observe that, defining

$$h^i(\mathrm{I}) := \mathtt{E}^i_*(\mathrm{I}) + \hat{h}_{\scriptscriptstyle k}(\hat{\mathrm{I}})\,, \qquad \mathtt{E}^i_* := \mathtt{E}^i_k \circ \jmath_{-\mathtt{g}_3}\,, \tag{156}$$

one has that

$$\jmath_{\mathtt{g}_*} = \jmath_{-\mathtt{g}_3} \circ \mathrm{U}^{-1}\,, \qquad \mathtt{h}^i(\mathrm{U}\mathrm{I}) = h^i(\mathrm{I})\,, \qquad (\forall\ \mathrm{I} \in \mathrm{U}^{-1}B^i)\,. \tag{157}$$

Now, since $\det \mathrm{U} = 1$,

$$\det\left[\partial^2_{\mathrm{I}} h^i(\mathrm{I})\right] \overset{(157)}{=} \det\left[\partial^2_{\mathrm{I}}\left(\mathtt{h}^i(\mathrm{U}\mathrm{I})\right)\right] = \det\left[\mathrm{U}^T \partial^2_I \mathtt{h}^i(I)\ \mathrm{U}\right] = \det\left[\partial^2_I \mathtt{h}^i(I)\right].$$

Thus,

$$(\det \partial^2_I \mathtt{h}^i) \circ \mathrm{U} = \det \partial^2_{\mathrm{I}} h^i\,. \tag{158}$$

Recalling (156) we then obtain

$$\det \partial^2_{\mathrm{I}} h^i \quad = \quad \det\left(\partial^2_{\mathrm{I}}\mathtt{E}^i_* + \partial^2_{\mathrm{I}}\hat{h}_{\scriptscriptstyle k}\right) = \det \begin{pmatrix} \partial^2_{\mathrm{I}_1}\mathtt{E}^i_* & \partial^T_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) \\ \partial_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) & \partial^2_{\hat{\mathrm{I}}}\mathtt{E}^i_* + \partial^2_{\hat{\mathrm{I}}}\hat{h}_{\scriptscriptstyle k} \end{pmatrix} \tag{159}$$

[65] Observe that if $\mathrm{S} = (\mathrm{s}_{ij})_{i,j\leqslant n}$ is an $(n\times n)$ matrix and $\hat{\mathrm{S}}$ denotes the $(n-1)\times(n-1)$ sub–matrix $(\mathrm{s}_{ij})_{i,j\geqslant 2}$, and S_0 denotes the matrix obtained by S replacing the entry s_{11} with 0, then $\det \mathrm{S} = \mathrm{s}_{11}\cdot\det\hat{\mathrm{S}} + \det \mathrm{S}_0$.

[66] Recall (110), (108) and (92).

$$= (\partial^2_{\mathrm{I}_1}\mathtt{E}^i_*)\cdot\det(\partial^2_{\hat{\mathrm{I}}}\mathtt{E}^i_* + \partial^2_{\hat{\mathrm{I}}}\hat{h}_{\mathtt{k}}) + \det\begin{pmatrix} 0 & \partial^T_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) \\ \partial_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) & \partial^2_{\hat{\mathrm{I}}}\mathtt{E}^i_* + \partial^2_{\hat{\mathrm{I}}}\hat{h}_{\mathtt{k}}\end{pmatrix}$$

and, by the chain rule,

$$\begin{aligned}(\partial^2_{\mathrm{I}_1}\mathtt{E}^i_*)\circ \jmath_{\mathtt{g}_3} &= \partial^2_{I_1}\mathtt{E}^i\,,\\ (\partial^2_{\mathrm{I}_1\hat{\mathrm{I}}}\mathtt{E}^i_*)\circ \jmath_{\mathtt{g}_3} &= \partial^2_{I_1\hat{I}}\mathtt{E}^i - \partial^2_{I_1}\mathtt{E}^i\partial_{\hat{I}}\mathtt{g}_3 =: \hat{\mathrm{v}}\,,\\ (\partial^2_{\hat{\mathrm{I}}\hat{\mathrm{I}}}\mathtt{E}^i_*)\circ \jmath_{\mathtt{g}_3} &= \partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{I_1}\mathtt{E}^i\partial^T_{\hat{I}}\mathtt{g}_3\partial_{\hat{I}}\mathtt{g}_3 - \partial_{I_1}\mathtt{E}^i\partial^2_{\hat{I}}\mathtt{g}_3 - \partial^T_{\hat{I}}\partial_{I_1}\mathtt{E}^i\partial_{\hat{I}}\mathtt{g}_3 - \partial^T_{\hat{I}}\mathtt{g}_3\partial_{\hat{I}}\partial_{I_1}\mathtt{E}^i =: \hat{\mathrm{M}}\,.\end{aligned} \tag{160}$$

Recalling that by (157) $\jmath_{\mathtt{g}_*} = \jmath_{-\mathtt{g}_3}\circ \mathrm{U}^{-1}$, by (158), (159) and (160) we get

$$\delta^i_\sharp := (\det\partial^2_I\mathtt{h}^i)\circ \jmath_{-\mathtt{g}_*} = \partial^2_{I_1}\mathtt{E}^i\cdot\det(\hat{\mathrm{M}} + \partial^2_{\hat{I}}\hat{h}_{\mathtt{k}}) + \det\begin{pmatrix} 0 & \hat{\mathrm{v}}^T \\ \hat{\mathrm{v}} & \hat{\mathrm{M}} + \partial^2_{\hat{I}}\hat{h}_{\mathtt{k}}\end{pmatrix}\,. \tag{161}$$

Finally, since the map $\jmath_{\mathtt{g}_*} : B^i_k \to \mathcal{B}^i_k$ is volume preserving, it is

$$\operatorname{meas} B^i_k = \operatorname{meas}\mathcal{B}^i_k\,,\qquad i = 0, 2N\,, \tag{162}$$

so that one obtains the following

Lemma 4.4 *Let $i = 0, 2N$ and $\delta^i_\sharp$ as in (161). Then,*

$$\operatorname{meas}\big(\{\, I\in B^i \text{ s.t. } \big|\det\partial^2_I\mathtt{h}^i(I)\big|\leqslant\eta\}\big) = \operatorname{meas}\big(\{\, I\in\mathcal{B}^i_k \text{ s.t. } \big|\delta^i_\sharp(I)\big|\leqslant\eta\}\big)\,.$$

(c) Here we prove the following uniform bound on the Hessian sub–matrix $\partial^2_{\hat{I}}\hat{h}_{\mathtt{k}}$. Recall the definition of δ_0 in (153)

Lemma 4.5 *There exists $c_3 = c_3(n) > 1$ such that if $\mathtt{K}\geqslant c_3$ the following estimates on the sub–matrix $\partial^2_{\hat{I}}\hat{h}_{\mathtt{k}}$ hold:*

$$\sup_{\hat{D}_{\mathtt{r}}}|\partial^2_{\hat{I}}\hat{h}_{\mathtt{k}}|\leqslant 2n^5+1\,,\qquad \inf_{\hat{D}_{\mathtt{r}}\cap\mathbb{R}^{n-1}}\det\partial^2_{\hat{I}}\hat{h}_{\mathtt{k}}\geqslant\delta_0\,. \tag{163}$$

Proof By (19),

$$(\mathrm{A}^T\mathrm{U})I = I_1 k + \hat{\mathrm{A}}^T\hat{I} - \tfrac{(\hat{\mathrm{A}}k)\cdot\hat{I}}{|k|^2}k = I_1 k + \hat{\mathrm{A}}^T\hat{I} - \tfrac{\hat{\mathrm{A}}^T\hat{I}\cdot k}{|k|^2}k = I_1 k + \pi^\perp_k\hat{\mathrm{A}}^T\hat{I}\,. \tag{164}$$

Recalling the definition of $\hat{\mathcal{Q}}_k$ in (53), we have

$$\partial^2_I\big(I_1^2 + \hat{\mathcal{Q}}_k(\hat{I})\big) = \partial^2_I\big(I_1^2 + \tfrac{|\pi^\perp_k\hat{\mathrm{A}}^T\hat{I}|^2}{|k|^2}\big) \overset{(164)}{=} \tfrac{\partial^2_I|\mathrm{A}^T\mathrm{U}I|^2}{|k|^2} = \tfrac{2(\mathrm{A}^T\mathrm{U})^T\mathrm{A}^T\mathrm{U}}{|k|^2} \tag{165}$$

and

$$|\partial^2_{\hat{I}}\hat{\mathcal{Q}}_k|\leqslant 2|k|^{-2}|\mathrm{A}|^2|\mathrm{U}|^2 \overset{(33),(64)}{\leqslant} 2n^5\,. \tag{166}$$

Using that $|k|\leqslant\mathtt{K}/6$, by (20), (48), (53) and Cauchy estimates we get, for a suitable $c' = c'(n)$,

$$\sup_{\hat{D}_{\mathtt{r}}}|\partial^2_{\hat{I}}(\hat{h}_{\mathtt{k}} - \hat{\mathcal{Q}}_k)|\leqslant\frac{c'}{\mathtt{K}^{14n+2}}\,. \tag{167}$$

By (166) and (167), taking $\mathtt{K}$ large enough (depending only on n), we get the first estimate in (163). Let us prove the second estimate in (163). Observe that

$$2\det\partial^2_{\hat I}\hat{\mathcal{Q}}_k = \det\partial^2_I\big(I_1^2+\hat{\mathcal{Q}}_k\big) \overset{(165)}{=} \tfrac{2^n}{|k|^{2n}}\det\big((\mathrm{A}^T\mathrm{U})^T\mathrm{A}^T\mathrm{U}\big) \overset{(33,60)}{=} \tfrac{2^n}{|k|^{2n}} \geqslant \tfrac{4}{|k|^{2n}} = 4\delta_{\mathrm{o}}\,,$$

and that

$$|(\partial^2_{\hat I}\hat{\mathcal{Q}}_k)^{-1}| \leqslant |(\partial^2_I(I_1^2+\hat{\mathcal{Q}}_k))^{-1}| \overset{(165)}{\leqslant} \tfrac{|k|^2}{2}|\mathrm{A}^{-1}|^2|\mathrm{U}^{-1}|^2 \overset{(33),(64)}{\leqslant} \tfrac12 n^5(n-1)^{n-1}|k|^{2n}\,. \tag{168}$$

Then, by (168), (167), using $|k|\leqslant \mathtt{K}/6$, we get, for a suitable constant $c''=c''(n)$,

$$|(\partial^2_{\hat I}\hat{\mathcal{Q}}_k)^{-1}|\cdot|\partial^2_{\hat I}(\hat h_{{}_k}-\hat{\mathcal{Q}}_k)| \leqslant \frac{c''}{\mathtt{K}^{12n+2}}\,, \tag{169}$$

We now need an elementary result on perturbation of positive–definite matrices, whose proof is given in Appendix:

Lemma 4.6 *Let P,Q be $d\times d$ positive–definite matrices and assume that $\lambda := |P^{-1}||Q|$ is strictly smaller than 1. Then $\det(P+Q)\geqslant(1-\lambda)^d\det P$. In particular, if $\lambda\leqslant(2d)^{-1}$, then $\det(P+Q)\geqslant(\det P)/2$.*

Then, observe that since $\partial^2_I\big(I_1^2+\hat{\mathcal{Q}}_k\big)$ is positive–definite (by (165)), so is $\partial^2_{\hat I}\hat{\mathcal{Q}}_k$. Therefore, since $\partial^2_{\hat I}\hat h_{{}_k} = \partial^2_{\hat I}\hat{\mathcal{Q}}_k + \partial^2_{\hat I}(\hat h_{{}_k}-\hat{\mathcal{Q}}_k)$, in view of (169), taking $\mathtt{K}\geqslant c_{{}_3}$ for a suitable $c_{{}_3}=c_{{}_3}(n)>1$, Lemma 4.6 implies also the second estimate in (163) and the proof of Lemma 4.5 is complete. ∎

Step 2 Here we define suitable *coverings of the sets* $\mathcal{B}^i_k$ defined in (92), (48). Such coverings are made up of sets corresponding to zones close to the separatrices and zones away from them.

Recall the definitions given in (92), (89) and (146). For any $\lambda_{\mathrm{o}}\in(0,1/\mathbf{c}_*)$, define the following subsets of[67] $\mathcal{B}^i_k$:

$$\begin{cases}\mathcal{B}^i_{{}_{\rm near}}(\lambda_{\mathrm{o}}) := \{I : b^i_{\lambda_{\mathrm{o}}}(\hat I) < I_1 < b^i(\hat I)\,,\hat I\in\hat D\}\,,\\ \mathcal{B}^i_{{}_{\rm far}}(\lambda_{\mathrm{o}}) := \mathcal{I}^i\times\hat D\,,\quad \mathcal{I}^i := (0,\bar b^i_{\lambda_{\mathrm{o}}/2})\,,\end{cases} \qquad i\ \text{ odd};$$
$$\begin{cases}\mathcal{B}^i_{{}_{\rm near}}(\lambda_{\mathrm{o}}) := \{I : a^i(\hat I) < I_1 < a^i_{\lambda_{\mathrm{o}}}(\hat I)\,,\hat I\in\hat D\}\\ \qquad\qquad\qquad \cup\,\{I : b^i_{\lambda_{\mathrm{o}}}(\hat I) < I_1 < b^i(\hat I)\,,\hat I\in\hat D\}\,,\\ \mathcal{B}^i_{{}_{\rm far}}(\lambda_{\mathrm{o}}) := \mathcal{I}^i\times\hat D\,,\quad \mathcal{I}^i := (\bar a^i_{\lambda_{\mathrm{o}}/2},\bar b^i_{\lambda_{\mathrm{o}}/2})\,,\end{cases} \qquad i\ \text{ even},\quad i\neq 0,2N_k\,; \tag{170}$$
$$\begin{cases}\mathcal{B}^i_{{}_{\rm near}}(\lambda_{\mathrm{o}}) := \{I : a^i(\hat I) < I_1 < a^i_{\lambda_{\mathrm{o}}}(\hat I)\,,\hat I\in\hat D\}\,,\\ \mathcal{B}^i_{{}_{\rm far}}(\lambda_{\mathrm{o}}) := \mathcal{I}^i\times\hat D\,,\quad \mathcal{I}^i := (\bar a^i_{\lambda_{\mathrm{o}}/2},b^i(\hat I))\,,\end{cases} \qquad i=0,2N_k\,.$$

Then, one has:

Lemma 4.7 *Let $0\leqslant i\leqslant 2N$ and assume that*[68]

$$\lambda_{\mathrm{o}} < 1/\mathbf{c}_*\,,\qquad\qquad \mu\leqslant\lambda_{\mathrm{o}}^2/2^8\mathbf{c}^4\,. \tag{171}$$

Then, $\mathcal{B}^i_k = \mathcal{B}^i_{{}_{\rm near}}(\lambda_{\mathrm{o}})\cup\mathcal{B}^i_{{}_{\rm far}}(\lambda_{\mathrm{o}})$.

[67] The constant $\mathbf{c}_*$ satisfies (114). Recall that for i odd $a^i_0\equiv 0$ (see (93)). The number λ_{o} will be fixed in Proposition 4.1 below.

[68] $\mathbf{c}_*$ appears in Theorem 3.1, while $\mathbf{c}\leqslant\mathbf{c}_*$ appears in Theorem 2.3. Recall (114).

Proof We give a detailed proof in the case (170) (i odd), as there is no extra difficulty in extending the proof to the other cases. For ease of notation in this proof we omit the suffix i.
Since the functions $E \to \bar I_1(E)$ and $E \to I_1(E,\hat I)$ are positive and strictly increasing (see (80)), the functions $\lambda \to \bar b_\lambda$ and $\lambda \to b_\lambda(\hat I)$ are positive and strictly decreasing. We claim that

$$b_{\lambda_o}(\hat I) < \bar b_{\lambda_o/2} < b_{\lambda_o/4}(\hat I)\,, \qquad \forall\, \hat I \in \hat D\,. \tag{172}$$

From such relations the claim follows: the fact that $\mathcal{B}_{\rm far}$ is a subset of $\mathcal{B}$ follows from the second inequality in (172), and then the equality $\mathcal{B} = \mathcal{B}_{\rm near} \cup \mathcal{B}_{\rm far}$ follows from the first inequality in (172).
Let us prove in detail the first inequality in (172) (the second one being analogous). By (170) and (89) we have[69]

$$b_{\lambda_o}(\hat I) = \epsilon \int_{\lambda_o}^{\lambda_{\rm max}} \partial_E I_1(E_+(\hat I) - \epsilon z, \hat I)\, dz = \epsilon \int_{\lambda_o + \lambda_\sharp}^{\lambda_{\rm max}+\lambda_\sharp} \partial_E I_1(\bar E_+ - \epsilon z, \hat I)\, dz\,,$$

where $\lambda_\sharp := (\bar E_+ - E_+(\hat I))/\epsilon$. Analogously,

$$\bar b_{\lambda_o/2} = \epsilon \int_{\lambda_o/2}^{\bar\lambda_{\rm max}} \partial_E \bar I_1(\bar E_+ - \epsilon z)\, dz\,,$$

where $\bar\lambda_{\rm max}$ was defined in (89). Note that, by (71) and (114), $|\lambda_\sharp| \leqslant 3\kappa^3\mu \leqslant \mathbf{c}\mu$, and that by (89), (90) we have that $\lambda_o \leqslant 1/\mathbf{c}_* \leqslant \min\{\lambda_{\rm max}/4, \bar\lambda_{\rm max}/8\}$. Then again by (90), (89), (91) and (171) we get that, for every $\hat I \in \hat D$,

$$\tfrac{\lambda_o}{2},\ \lambda_o + \lambda_\sharp,\ \lambda_{\rm max} + \lambda_\sharp,\ \bar\lambda_{\rm max}\ \in\ \big(\tfrac{\lambda_o}{8}, \bar\lambda_{\rm max} + \tfrac{1}{\mathbf{c}}\big)\,.$$

We write

$$\begin{aligned}\frac{\bar b_{\lambda_o/2} - b_{\lambda_o}(\hat I)}{\epsilon} \quad &= \int_{\lambda_o/2}^{\lambda_o+\lambda_\sharp} \partial_E \bar I_1(\bar E_+ - \epsilon z)\, dz + \int_{\lambda_{\rm max}+\lambda_\sharp}^{\bar\lambda_{\rm max}} \partial_E \bar I_1(\bar E_+ - \epsilon z)\, dz \\ &\quad + \int_{\lambda_o+\lambda_\sharp}^{\lambda_{\rm max}+\lambda_\sharp} \big(\partial_E \bar I_1(\bar E_+ - \epsilon z) - \partial_E I_1(\bar E_+ - \epsilon z, \hat I)\big)\, dz\,,\end{aligned}$$

observing that, for every z in the three integration intervals (and for every $\hat I \in \hat D$), the quantity $\bar E_+ - \epsilon z$ belongs to the set[70] $\mathcal{E}_{\lambda_o/8}$. Then, by (80), (83) and (171) we get, for every $\hat I \in \hat D$,

$$\begin{aligned}\frac{\bar b_{\lambda_o/2} - b_{\lambda_o}(\hat I)}{\epsilon} &\geqslant \frac{\lambda_o - 2|\lambda_\sharp|}{2\mathbf{c}\sqrt\epsilon} - \mathbf{c}^2 \frac{|\log\frac{\lambda_o}{8}|}{\sqrt\epsilon}\big(|\bar\lambda_{\rm max} - \lambda_{\rm max}| + |\lambda_\sharp|\big) - 8\mathbf{c}^2\mu\frac{\lambda_{\rm max}}{\lambda_o\sqrt\epsilon} \\ &\overset{(91),(90)}{\geqslant} \frac{1}{2\mathbf{c}\sqrt\epsilon}\big(\lambda_o - 2\mathbf{c}\mu - 2^4\mathbf{c}^4\mu|\log\tfrac{\lambda_o}{8}| - 2^6\mathbf{c}^3\mu/\lambda_o\big) \geqslant \frac{\lambda_o}{4\mathbf{c}\sqrt\epsilon} > 0\,. \quad \blacksquare\end{aligned}$$

Step 3* *Non–degeneracy of the twist function in neighborhoods of separatrices*

Here we show that (a suitable regularization of) the twist determinant $\det \partial^2_I \mathtt{h}^i$ in (155) is a non–degenerate function in the sense of Definition 4.1 in *suitable neighborhoods of separatrices.*

---

[69] Recall that $b_{\lambda_{\rm max}}(\hat I) = I_1(E_-(\hat I), \hat I) = a_0(\hat I) = 0$.

[70] Recall (82). Note that (171) implies (81) with $\lambda = \lambda_o/8$.

Actually, it will be convenient to study the twist directly as a function of the energy, for values $E = E^i_\mp(\hat I) \pm \epsilon z$ close to critical separatrix values $E^i_\mp$. We therefore define:

$$\delta_\mp(z,\hat I) := \det\left[\partial^2_{\hat I}\mathtt{E}\big(I_1(E^i_\mp(\hat I) \pm \epsilon z, \hat I), \hat I\big) + \partial^2_{\hat I}\hat h_{\kappa}(\hat I)\right]. \tag{173}$$

The study of the twist determinant (173) will be based on the analytic properties described in Theorem 2.3. In particular, the properties that we shall use are the same in the plus and the minus case. Hence, we shall consider only the plus case and *consider, henceforth,* $\delta := \delta_+$.

The precise statement on the non–degeneracy of $z \to \delta(z,\hat I)$ (see Proposition 4.1 below) needs some preparation.

First of all, we introduce a suitable 'regularization' function $\zeta = \zeta(z,\hat I)$

$$\zeta(z,\hat I) := z\cdot\big(\sqrt{\epsilon}\,\partial_E I_1(E_+(\hat I) - \epsilon z, \hat I)\big)^3, \tag{174}$$

and define *the regularized twist determinant* $\bar\delta$ by setting[71]

$$\bar\delta = \tilde\delta/\det\partial^2_{\hat I}\hat h_{\kappa}, \quad \text{with} \quad \tilde\delta(z,\hat I) := \zeta(z,\hat I)^n\cdot\delta(z,\hat I). \tag{175}$$

The functions appearing in Theorem 2.3, as well as the functions in (174) and (175) belong to the following ring of functions $\mathcal F$.

Definition 4.4 *We denote by* $\mathcal F$ *the set of functions of the form*

$$f(z,\hat I) = z^h \sum_{j=0}^{\ell} u_j(z,\hat I)\log^j z, \tag{176}$$

where $h,\ell \in \mathbb Z$ *with* $\ell \geqslant 0$ *and the* u_j *are real analytic functions on a (complex) neighborhood of*[72] $\{z=0\}\times\hat D \subset \mathbb C^n$.

We shall also use the following notation: given two functions $f_i \in \mathcal F$ *we say that* $f = f_1 \oplus f_2$ *if there exists two functions* u_i *real analytic on a neighborhood of* $\{z=0\}\times\hat D$ *such that*[73] $f = u_1 f_1 + u_2 g_2$.

We say that $f(z,\hat I) = \mathcal O_\varrho(h,\ell)$ *if* $f \in \mathcal F$ *as in* (176) *and there exists* $\varrho > 0$ *such that*

$$|||f|||_\varrho := \sup_{0\leqslant j\leqslant \ell}\ \sup_{\substack{\{z\in\mathbb C: |z|<\varrho\}\\ \hat I\in\hat D}} |u_j| < +\infty.$$

Remark 4.2 (i) The functions $(z,\hat I) \to f(z,\hat I) = I^i_1\big(E^i_\mp(\hat I) \pm \epsilon z,\ \hat I\big)$ in (75) of Theorem 2.3 belongs to $\mathcal F$ and, by (76),

$$|||f|||_{1/\mathbf c} \leqslant \mathbf c\sqrt{\epsilon};$$

furthermore, the 'algebraic structure' of such function f is given by

$$f = \sqrt{\epsilon}(1 \oplus z\log z).$$

[71] Recall (163).

[72] Recall that the domain $\hat D$ is defined in (48), but, essentially plays no rôle.

[73] E.g., f in (176) can be written as $z^h\big(\bigoplus_{j=0}^{\ell}\log^j z\big)$.

(ii) The following elementary properties (which, in particular, show that $\mathcal{F}$ is a ring) will be often used:

$$\begin{cases} \mathcal{O}_\varrho(h,p)\cdot\mathcal{O}_\varrho(k,q) = \mathcal{O}_\varrho(h+k,p+q)\,, \\ (\mathcal{O}_\varrho(h,p))^j = \mathcal{O}_\varrho(jh,jp)\,, \\ \mathcal{O}_\varrho(h,p)+\mathcal{O}_\varrho(k,q) = \mathcal{O}_\varrho(\min\{h,k\},\max\{p,q\})\,. \end{cases}$$

Finally, define the following linear differential operators:

$$\mathcal{L} := L^{3\bar n}(\partial_z\cdot L^{3\bar n})^{\bar n}\,, \quad \text{where:} \quad L := z\partial_z\,, \quad \bar n := n-1\,.$$

Notice that $\mathcal{L}$ is a linear differential operator of order $\bar{\mathrm{m}} := 3\bar n^2 + 4\bar n = 3n^2 - 2n - 1 \geqslant 7$ and there exist suitable polynomials $a_j(z)$ such that[74]

$$\mathcal{L} = \sum_{j=1}^{\bar{\mathrm{m}}} a_j(z)\partial_z^j\,. \tag{177}$$

Proposition 4.1 *There exists*[75] $\mathbf{c}_1 = \mathbf{c}_1(n,\kappa) > c_3$ *such that if* $\mathtt{K} \geqslant \mathbf{c}_1$, *then the following holds.*

(i) *One has*

$$\mathcal{L}[\bar\delta] = \bar n!^{3\bar n+1}(3\bar n)!\,\gamma^{3\bar n} + \mathcal{O}_\varrho(1,3\bar n+1)\,, \tag{178}$$

where[76]

$$\gamma(\hat I) := -\epsilon^{-1/2}\psi_+(0,\hat I)\,, \qquad \varrho := 1/\mathbf{c}\,,$$

and

$$1/\mathbf{c} \leqslant \inf_{\hat D}|\gamma| \leqslant \sup_{\hat D}|\gamma| \leqslant \mathbf{c}\,. \tag{179}$$

(ii) *There exist suitable positive constants* $\xi_\sharp = \xi_\sharp(n,\kappa) < 1$ *and*[77] $\lambda_{\mathrm{o}} = \lambda_{\mathrm{o}}(n,\kappa) < 1/\mathbf{c}_\star$, *such that, for* $\hat I \in \hat D$, *the function* $z \to \bar\delta(z,\hat I)$ *defined in* (175) *is* $\xi_\sharp$*–non–degenerate at order* $\bar{\mathrm{m}} = 3n^2 - 2n - 1$ *on the interval* $(0,\lambda_{\mathrm{o}})$.

To prove this proposition we need a couple of preparatory lemmata.

Notation 4.1 *In the rest of this section, it is understood that in an expansion* $f = z^h\big(\bigoplus_{j=0}^{\ell}\log^j z\big)$, *one has* $|||f|||_\varrho \leqslant c$ *for a suitable constant* $c = c(n,\kappa)$; *furthermore,* $\mathcal{O}$ *stands for* $\mathcal{O}_\varrho$ *with* $\varrho = 1/\mathbf{c}$.

We shall consider in detail only the *inner odd case* $0 < i < 2N$, since the other cases do not present any new difficulties; for ease of notation, we do not indicate explicitly the labels k and i.

Lemma 4.8 *If* ζ *is as in* (174), $\tilde I = \tilde I(z,\hat I) := (b_z(\hat I),\hat I)$ *and* μ_{o} *is as in* (76), *one has* $(2 \leqslant i,j \leqslant n)$

$$\begin{aligned} \zeta &= z\big(\gamma\log z + (1\oplus z\log z)\big)^3 \\ &= \gamma^3 z\log^3 z + \mathcal{O}(1,2) + \mathcal{O}(2,3) = \mathcal{O}(1,3)\,, \end{aligned}$$

[74] Actually, $\mathcal{L} = \sum_{j=\bar n+1}^{\bar{\mathrm{m}}} a_j z^{j-\bar n}\partial_z^j$, with $a_j \in \mathbb{N}$. For example, if $n=2$, $\bar{\mathrm{m}} = 7$ and $\mathcal{L}$ is given by:
$\mathcal{L} = z^6\partial_z^7 + 18z^5\partial_z^6 + 98z^4\partial_z^5 + 184z^3\partial_z^4 + 100z^2\partial_z^3 + 8z\partial_z^2\,.$

[75] The constant c_3 has been introduced in Lemma 4.5.

[76] Recall (78).

[77] The constant $\mathbf{c}_\star$ has been introduced in Theorem 3.1.

$$
\begin{aligned}
\zeta \cdot \partial^2_{I_1} \mathtt{E}|_{I=\tilde I} &= \gamma + z(1 \oplus \log z) = \gamma + \mathcal{O}(1,1)\,, \\
\zeta \cdot \partial^2_{I_1 \hat I_i} \mathtt{E}|_{I=\tilde I} &= \mu_o (1 \oplus z \log z \oplus z \log^2 z) = \mu_o\, \mathcal{O}(0,2)\,, \\
\zeta \cdot \partial^2_{\hat I_i \hat I_j} \mathtt{E}|_{I=\tilde I} &= \mu_o (1 \oplus z \log z \oplus z \log^2 z \oplus z^2 \log^3 z) = \mu_o\, \mathcal{O}(0,3)\,.
\end{aligned}
\tag{180}
$$

Proof By the chain rule, one has (writing I_1 in place of I_1^i)

$$
\begin{aligned}
&\partial_{I_1} \mathtt{E}^i = \frac{1}{\partial_E I_1}\,, \qquad \partial_{\hat I} \mathtt{E}^i = -\frac{\partial_{\hat I} I_1}{\partial_E I_1}\,, \qquad \partial^2_{I_1} \mathtt{E}^i = -\frac{\partial^2_E I_1}{(\partial_E I_1)^3}\,, \\
&\partial^2_{I_1 \hat I} \mathtt{E}^i = \frac{\partial^2_E I_1 \partial_{\hat I} I_1}{(\partial_E I_1)^3} - \frac{\partial^2_{E \hat I} I_1}{(\partial_E I_1)^2}\,, \\
&\partial^2_{\hat I} \mathtt{E}^i = -\frac{\partial^2_{\hat I} I_1}{\partial_E I_1} + \frac{\partial^T_{\hat I} I_1\ \partial_{\hat I}(\partial_E I_1) + \partial^T_{\hat I}(\partial_E I_1)\ \partial_{\hat I} I_1}{(\partial_E I_1)^2} - \frac{\partial^2_E I_1\ \partial^T_{\hat I} I_1\ \partial_{\hat I} I_1}{(\partial_E I_1)^3}\,,
\end{aligned}
\tag{181}
$$

where the derivatives of $\mathtt{E}^i$ and $I_1 = I_1^i$ are evaluated in $\big(I_1^i(E, \hat I), \hat I\big)$ and $(E, \hat I)$, respectively. Now, by (181) and (75), we have

$$
\begin{aligned}
&\sqrt{\epsilon} \partial_E I_1 = \gamma \log z + (1 \oplus z \log z) = 1 \oplus \log z\,, \quad \partial_{\hat I_i} I_1 = \mu_o (1 \oplus z \log z)\,, \\
&\epsilon^{3/2} \partial^2_E I_1 = -\gamma z^{-1} + (\log z \oplus z^{-1}) = \log z \oplus z^{-1}\,, \\
&\epsilon \partial^2_{E \hat I_i} I_1 = \mu_o (1 \oplus \log z)\,, \quad \partial^2_{\hat I_i \hat I_j} I_1 = \mu_o \mathtt{r}^{-1} (1 \oplus z \log z) \overset{(43)}{=} \mu_o \epsilon^{-1/2} (1 \oplus z \log z)\,.
\end{aligned}
\tag{182}
$$

Finally, by (182), (181), (75) and (76) we get

$$
\begin{aligned}
\sqrt{\epsilon} \partial_E I_1 &= \gamma \log z + (1 \oplus z \log z) = 1 \oplus \log z \\
\zeta \cdot \partial^2_{I_1} \mathtt{E} &= -\epsilon^{3/2} z \partial^2_E I_1 = \gamma + z(1 \oplus \log z) = 1 \oplus z \log z\,, \\
\zeta \cdot \partial^2_{I_1 \hat I_i} \mathtt{E} &= \epsilon^{3/2} z (\partial^2_E I_1 \partial_{\hat I_i} I_1 - \partial^2_{E \hat I_i} I_1 \partial_E I_1) = \mu_o (1 \oplus z \log z \oplus z \log^2 z)\,, \\
\zeta \cdot \partial^2_{\hat I_i \hat I_j} \mathtt{E} &= \epsilon^{3/2} z \big(- (\partial_E I_1)^2 \partial^2_{\hat I_i \hat I_j} I_1 + 2 \partial_E I_1 \partial_{\hat I_i} I_1\ \partial^2_{\hat I_j E} I_1 - \partial^2_E I_1\ \partial_{\hat I_i} I_1\ \partial_{\hat I_j} I_1 \big) \\
&= \mu_o (1 \oplus z \log z \oplus z \log^2 z \oplus z^2 \log^3 z)\,. \quad \blacksquare
\end{aligned}
$$

Lemma 4.9 *One has*

$$
\bar\delta = \bar\delta(z, \hat I) = \gamma^{3\bar n} z^{\bar n} \log^{3\bar n} z + \mathcal{O}(\bar n + 1, 3\bar n + 1) + \mathcal{O}(0, 3\bar n - 1)\,, \quad \bar n := n - 1\,.
\tag{183}
$$

Furthermore, there exists $\mathbf{c}_1 = \mathbf{c}_1(n, \kappa) > c_3$ *such that if* $\mathtt{K} \geqslant \mathbf{c}_1$ *and* $\lambda_o \leqslant 1/\mathbf{c}_1$*, one has:*

$$
|\delta(z, \hat I)| \geqslant \delta_o\ |\bar\delta(z, \hat I)|\,, \qquad \forall\, 0 < z \leqslant \lambda_o\,,\ \hat I \in \hat D\,.
\tag{184}
$$

Proof Recalling (155) we split $\tilde\delta$ in (175) in two terms. The first term is

$$
\zeta^n (\partial^2_{I_1} \mathtt{E})\ \det(\partial^2_{\hat I} \hat h_{_\kappa} + \partial^2_{\hat I} \mathtt{E}) \overset{(180)}{=} \big(\gamma + \mathcal{O}(1,1)\big)\, \zeta^{\bar n} \det(\partial^2_{\hat I} \hat h_{_\kappa} + \partial^2_{\hat I} \mathtt{E})\,,
$$

and, by (180), we have that

$$
\zeta^{\bar n}\ \det(\partial^2_{\hat I} \hat h_{_\kappa} + \partial^2_{\hat I} \mathtt{E})
$$

$$
\begin{aligned}
&= \quad \zeta^{\bar n}\ \det \partial^2_{\hat I} \hat h_{_k} + \sum_{j=1}^{\bar n} \zeta^{\bar n - j} \mu_o^j \big(1 \oplus z \log z \oplus z \log^2 z \oplus z^2 \log^3 z\big)^j \\
&= \quad \big(\gamma^{3\bar n} z^{\bar n} \log^{3\bar n} z + \mathcal{O}(\bar n, 3\bar n - 1) + \mathcal{O}(\bar n + 1, 3\bar n)\big) \det \partial^2_{\hat I} \hat h_{_k} \\
&\quad\ + \mu_o \big(\mathcal{O}(0, 3\bar n - 3) + \mathcal{O}(\bar n, 3\bar n - 1) + \mathcal{O}(\bar n + 1, 3\bar n)\big) \\
&= \quad \big(\gamma^{3\bar n} z^{\bar n} \log^{3\bar n} z + \mathcal{O}(\bar n + 1, 3\bar n) + \mathcal{O}(0, 3\bar n - 1)\big) \det \partial^2_{\hat I} \hat h_{_k}\,,
\end{aligned}
$$

where in the last line we used (recall (76), (48), (163))

$$
\mu_o \leqslant |k|^{-2n} = \delta_o \leqslant \inf_{\hat D} \det \partial^2_{\hat I} \hat h_{_k}\,. \tag{185}
$$

The second term is,

$$
\zeta^n \det \begin{pmatrix} 0 & \partial^T_{\hat I}(\partial_{I_1} \mathtt{E}) \\ \partial_{\hat I}(\partial_{I_1}\mathtt{E}) & \partial^2_{\hat I}\mathtt{E} + \partial^2_{\hat I} \hat h_{_k} \end{pmatrix} = \mu_o \det \begin{pmatrix} 0 & w^T \\ w & \mathrm{N} \end{pmatrix}
$$

$$
\begin{aligned}
&= \quad \mu_o^2 (1 \oplus z \log z \oplus z \log^2 z)^2 (1 \oplus z \log z \oplus z \log^2 z \oplus z \log^3 z)^{n-2} \\
&= \quad \mu_o^2 \mathcal{O}(0, 3n - 4) + \mathcal{O}(n, 3n - 2)\,,
\end{aligned}
$$

where w is a $\bar n$–dimensional vector and N is an $(\bar n \times \bar n)$ matrix satisfying by (180)

$$
w_i = \mathcal{O}(0,0) + \mathcal{O}(1,2)\,, \quad 2 \leqslant i \leqslant n\,, \qquad \mathrm{N}_{ij} = \mathcal{O}(0,0) + \mathcal{O}(1,3)\,, \quad 2 \leqslant i, j \leqslant n\,.
$$

Thus, the second term has the form $\mu_o^2\big(\mathcal{O}(\bar n + 1, 3\bar n + 1) + \mathcal{O}(0, 3\bar n - 1)\big)$. Summing up the two terms and using (185) we get (183).

By the first line in (180), we see that, taking $\mathbf{c}_{_1}$ big enough, one has

$$
|\zeta(z, \hat I)| \leqslant 1\,, \qquad \forall\, 0 < z \leqslant \lambda_o\,,\ \hat I \in \hat D\,.
$$

Thus, by the definitions in (173), (175) and by (163), one obtains (184). ∎

Before giving the proof of Proposition 4.1, we need one more lemma. Define

$$
\mathcal{L}_{m,k} := L^k (\partial_z \cdot L^k)^m\,.
$$

Lemma 4.10 *Let $0 \leqslant \ell < k \leqslant \bar{\mathrm{m}}$, $0 \leqslant m, q \leqslant \bar{\mathrm{m}}$, and $f_1 = \mathcal{O}_\varrho(0, \ell)$, $f_2 = \mathcal{O}_\varrho(m + 1, q)$. Then*

$$
\mathcal{L}_{m,k}[z^m \log^k z + f_1 + f_2] = (m!)^{k+1} k! + f_3\,, \tag{186}
$$

where, for a suitable constant c, which depends only on n, one has

$$
f_3 = \mathcal{O}_{\varrho/2}(1, \max\{k - 1, q\})\,, \quad \text{and} \quad |||f_3|||_{\varrho/2} \leqslant c \max\{|||f_1|||_\varrho\,, |||f_2|||_\varrho\}\,.
$$

Proof Observing that $L z^m = m z^m$, $L \log^{\ell+1} z = (\ell + 1) \log^\ell z$, one easily checks that, for any $0 \leqslant m, \ell \leqslant \bar{\mathrm{m}}$, one has

$$
\begin{aligned}
L[\mathcal{O}_\varrho(m, \ell)] &= \mathcal{O}_{\frac34\varrho}(m, \ell)\,, \\
L\mathcal{O}_\varrho(0, \ell + 1) &= \mathcal{O}_{\frac34\varrho}(1, \ell + 1) + \mathcal{O}_{\frac34\varrho}(0, \ell)\,, \\
L^{\ell+1}[\mathcal{O}_\varrho(0, \ell)] &= \mathcal{O}_{\frac34\varrho}(1, \ell)\,,
\end{aligned}
$$

where the norm $\|\!|\cdot|\!\|_{\frac{3}{4}\varrho}$ of the functions in the right hand sides are bounded by $c' = c'(n)$ times the norm $\|\!|\cdot|\!\|_{\varrho}$ of the functions in the left hand sides[78]. Analogously, from the above relations, it follows that, for any $0 \leqslant \ell < k \leqslant \bar{\mathrm{m}}$ and $0 \leqslant q \leqslant \bar{\mathrm{m}}$, one has

$$\mathcal{L}_{m,k}[z^m \log^k z] = (m!)^{k+1}(k)! + \mathcal{O}_{\frac{1}{2}\varrho}(1, k-1)\,,$$

$$\mathcal{L}_{m,k}[\mathcal{O}_{\frac{3}{4}\varrho}(0,\ell)] = \mathcal{O}_{\frac{1}{2}\varrho}(1,\ell)\,,$$

$$\mathcal{L}_{m,k}[\mathcal{O}_{\frac{3}{4}\varrho}(m+1,q)] = \mathcal{O}_{\frac{1}{2}\varrho}(1,q)\,,$$

where, the norm $\|\!|\cdot|\!\|_{\frac{1}{2}\varrho}$ of the functions in the right hand sides are bounded by $c = c(n) > c'$ times the norm $\|\!|\cdot|\!\|_{\frac{3}{4}\varrho}$ of the functions in square brackets in the left hand side. From such relations, the claim of the lemma follows easily. ∎

Proof of Proposition 4.1 The estimates in (179) follow trivially from (78) and (76).
To check (178), observe that $\mathcal{L} = \mathcal{L}_{\bar{n},3\bar{n}}$, and use (183) in Lemma 4.9 and (186) (with $m = \bar{n}$, $k = 3\bar{n}$, $\ell = 3\bar{n}-1$, and $q = 3\bar{n}+1$).
It remains to prove claim (ii). By (178), and (179), we see that for $\lambda_o < 1/\mathbf{c}_\star$ small enough one has:

$$\frac{1}{\mathbf{c}^{3\bar{n}}} \leqslant \inf_{0<z\leqslant\lambda_o} \inf_{\hat{I}\in\hat{D}} \big|\mathcal{L}[\bar{\delta}]\big| \overset{(177)}{\leqslant} c'' \max_{1\leqslant j\leqslant \bar{\mathrm{m}}} |\partial_z^j \bar{\delta}|\,,$$

where $c'' = c''(n)$. Thus, for $\hat{I} \in \hat{D}$, $z \to \bar{\delta}(z,\hat{I})$ is $\xi_\sharp$–non–degenerate at order $\bar{\mathrm{m}} = 3n^2 - 2n - 1$ on the interval $(0,\lambda_o)$ with $\xi_\sharp = (c''\mathbf{c}^{3\bar{n}})^{-1}$. ∎

Step 4: *The Twist Theorem in neighborhoods of separatrices*
We can now state and prove the Twist Theorem in neighborhoods of separatrices.

Proposition 4.2 *Let $k \in \mathcal{G}^n_{\mathtt{K}_o}$, $0 \leqslant i \leqslant 2N$, $\eta > 0$, and λ_o as in Proposition 4.1–(ii). Then, there exist a positive constant $\mathbf{c}_2 = \mathbf{c}_2(n,\kappa) \geqslant \mathbf{c}$ such that, if $\mathtt{K} \geqslant \mathbf{c}_2$, then*

$$\operatorname{meas}\big(\{I \in \mathcal{B}^i_{\text{near}}(\lambda_o) :\ |\det \partial_I^2 \mathrm{h}^i(I)| \leqslant \eta\}\big) \leqslant \mathbf{c}_2\,(|k|^{2n}\eta)^{1/9n^4} \operatorname{meas} \mathcal{B}^i_k\,. \tag{187}$$

Before the proof, which will be based on two lemmata, we introduce the following

Notation 4.2 *Given two non negative functions f and g we say that $f \lessdot g$ if there exists a constant $c = c(n,\kappa) \geqslant 1$, depending only on n and κ, such that $f \leqslant cg$. Similarly, given a function f and a non negative function g, we say that $f = O(g)$ if there exists a constant $c = c(n,\kappa) \geqslant 1$, such that $|f| \leqslant cg$.*

Lemma 4.11 *There exists a constant $\mathbf{c}_3 = \mathbf{c}_3(n,\kappa) > 1$ such that, for every $\hat{I} \in \hat{D}$ and $\eta > 0$, one has*[79]

$$\operatorname{meas}\{z \in (0,\lambda_o] \text{ s.t. } |\bar{\delta}(z,\hat{I})| \leqslant \eta\} \leqslant \mathbf{c}_3 \eta^{\bar{\mathrm{a}}}\,, \qquad \bar{\mathrm{a}} := \tfrac{1}{\bar{\mathrm{m}}(\bar{\mathrm{m}}+3)}\,. \tag{188}$$

[78] The algebraic relations are just calculus, while the estimates follow easily by (iterated) use of Cauchy estimates.
[79] $\bar{\mathrm{m}} = 3n^2 - 2n - 1$ is defined in Proposition 4.1.

Proof If $z_0 \leqslant 2\eta^{\bar{a}}$ estimate (188) is obvious. Consider the case $z_0 > 2\eta^{\bar{a}}$. Let

$$\lambda_1 := \eta^{\bar{a}} < z_0/2\,.$$

By (183), (163) and (179) we have that[80]

$$\sup_{[\lambda_1,\lambda_o]_{\lambda_1/2}} \sup_{\hat{D}} |\bar{\delta}(z,\hat{I})| \lessdot 1 + |\log^{3n-4}\lambda_1| \lessdot 1/\lambda_1\,.$$

By Cauchy estimates

$$\sup_{\lambda_1 \leqslant z \leqslant \lambda_o} \sup_{\hat{I}\in\hat{D}} \max_{1\leqslant j\leqslant \bar{m}+1} |\partial_z^j \bar{\delta}(z,\hat{I})| \leqslant c_\flat/\lambda_1^{\bar{m}+2} =: M\,, \tag{189}$$

for a suitable $c_\flat \geqslant 1$, depending only on n, κ. Now we want to apply Lemma 4.1 with

$$f = \bar{\delta}\,,\quad m = \bar{m}\,,\quad a = \lambda_1\,,\quad b = \lambda_o\,,\quad \xi = \xi_\sharp\,,\quad M \text{ as in (189)}\,.$$

Then, we get

$$\text{meas}\{z \in (\lambda_1, \lambda_o) : |\bar{\delta}(z,\hat{I})| \leqslant \eta\} \lessdot \eta^{\bar{a}}\,,\quad \forall\,\hat{I}\in\hat{D}\,. \tag{190}$$

Since the interval $(0,\lambda_1)$, has length $\lambda_1 = \eta^{\bar{a}}$, from (190) we obtain the measure estimate (188). ∎

Now, recalling that $\delta_o = |k|^{-2n}$ (see (153)), we have:

Lemma 4.12 *There exists $\mathbf{c}_4 = \mathbf{c}_4(n,\kappa) \geqslant \max\{\mathbf{c}_1, \mathbf{c}_3\}$ such that for $k \in \mathcal{G}^n_{K_o}$, i odd and $\eta > 0$,*

$$\text{meas}\left(\{I \in \mathcal{B}^i_{\text{near}} : |\det \partial_I^2 \mathtt{h}^i(I)| \leqslant \eta\}\right) \leqslant \mathbf{c}_4 \sqrt{\epsilon}\,(\eta/\delta_o)^{\frac{1}{9n^4}}\ \text{meas}\,\hat{D}\,.$$

Proof Let $\mathcal{Z}_\eta(\hat{I}) := \left(\{z \in (0,\lambda_o] : |\delta(z,\hat{I})| \leqslant \eta\}\right.$. By (184) and (188) we get

$$\text{m}_\eta = \text{m}_\eta(\hat{I}) := \text{meas}(\mathcal{Z}_\eta(\hat{I})) \leqslant \mathbf{c}_3(\eta/\delta_o)^{\bar{a}}\,,\qquad \forall\ \hat{I}\in\hat{D}\,. \tag{191}$$

Note that since $\lambda_o \leqslant 1/2$ (see (171)), by definition

$$\text{m}_\eta \leqslant \lambda_o \leqslant 1/2\,. \tag{192}$$

Recalling (92), we define, for $\hat{I}\in\hat{D}$ and $\eta > 0$,

$$\mathcal{I}_\eta(\hat{I}) := \left\{I_1 \in \left[b_{\lambda_o}(\hat{I}), b(\hat{I})\right) : \left|\det\left[\partial_I^2\left(\hat{h}_k(\hat{I}) + \mathtt{E}(I)\right)\right]\right| \leqslant \eta\right\}.$$

We have that

$$\mathcal{I}_\eta(\hat{I}) = b_{\mathcal{Z}_\eta(\hat{I})}(\hat{I}) := \left\{I_1 = b_z(\hat{I}) : z \in \mathcal{Z}_\eta(\hat{I})\right\}, \tag{193}$$

since by definition of $\mathcal{Z}_\eta$, (173) and (170) $\delta(z,\hat{I}) = \det\left[\partial_I^2\hat{h}_k(\hat{I}) + \partial_I^2\mathtt{E}\left(b_z(\hat{I}),\hat{I}\right)\right]$. For every $\hat{I}\in\hat{D}$ and $\eta > 0$, making the change of variable $I_1 = b_z(\hat{I})$, and noticing that $\partial_z b_z(\hat{I}) = -\epsilon\partial_E I_1(E_+(\hat{I}) - \epsilon z, \hat{I})$, we get

$$\text{meas}(\mathcal{I}_\eta(\hat{I})) \quad = \quad \int_{\mathcal{I}_\eta(\hat{I})} dI_1 \overset{(193)}{=} \int_{b_{\mathcal{Z}_\eta(\hat{I})}} dI_1 = \int_{\mathcal{Z}_\eta(\hat{I})} |\partial_z b_z(\hat{I})| dz$$

[80] Denoting, as usual, the $\lambda_1/2$–complex–neighborhood of the real interval $[\lambda_1,\lambda_o]$ by $[\lambda_1,\lambda_o]_{\lambda_1/2}$.

$$\overset{(180)}{\lessdot} \quad \sqrt{\epsilon}\int_{\mathcal{Z}_\eta(\hat{I})} |\log z| dz\,.$$

Moreover, recalling (192),

$$\int_{\mathcal{Z}_\eta(\hat{I})} |\log z| dz \leqslant \int_0^{\mathrm{m}_\eta} |\log z| dz + \int_{\mathcal{Z}_\eta(\hat{I})\cap(\mathrm{m}_\eta,\lambda_o]} |\log z| dz \leqslant 2\mathrm{m}_\eta |\log \mathrm{m}_\eta|\,.$$

Thus, by (192), using $\frac{1}{9n^4} < \bar{\mathrm{a}}$ (see (188) and recall the definition of $\bar{\mathrm{m}} = 3n^2 - 2n - 1$ in Proposition 4.1), we get

$$\mathrm{meas}(\mathcal{I}_\eta(\hat{I})) \lessdot \sqrt{\epsilon}\,\mathrm{m}_\eta |\log \mathrm{m}_\eta| \lessdot \sqrt{\epsilon}\,\mathrm{m}_\eta^{1/(9n^4\,\bar{\mathrm{a}})}\,.$$

By (191), $\mathrm{meas}(\mathcal{I}_\eta(\hat{I})) \lessdot \sqrt{\epsilon}\,(\eta/\delta_o)^{1/9n^4}$, for every $\hat{I} \in \hat{D}$ and $\eta > 0$. By Fubini's Theorem, the claim follows. ∎

Proof of Proposition 4.2 By (92) we get

$$\begin{aligned}
\mathrm{meas}\,\mathcal{B}_k^i &= \int_{\hat{D}} b(\hat{I}) d\hat{I} = \int_{\hat{D}} I_1\big(E_+(\hat{I}),\hat{I}\big) d\hat{I} = \int_{\hat{D}} d\hat{I} \int_{E_-(\hat{I})}^{E_+(\hat{I})} \partial_E I_1\big(E_+(\hat{I}),\hat{I}\big) dE \\
&\overset{(80)}{\geqslant} \frac{E_+(\hat{I}) - E_-(\hat{I})}{\mathbf{c}\sqrt{\epsilon}}\,\mathrm{meas}\,\hat{D} \overset{(89)}{\geqslant} \frac{\sqrt{\epsilon}}{2\kappa\mathbf{c}}\,\mathrm{meas}\,\hat{D}\,.
\end{aligned} \tag{194}$$

Lemma 4.12 and (194) imply at once (187) if one takes $\mathbf{c}_2 \geqslant \mathbf{c}_4$ big enough. The proof of Proposition 4.2 is complete for i odd. The changes for the inner case with i even are straightforward .

Let us indicate the changes one needs to do in order to prove the outer case $i = 0, 2N$. Recalling (141), (48), (20), by Cauchy estimates, we have (recall Notation 4.2)

$$|\partial_{\hat{I}} \mathtt{g}_3|_{\hat{D},3\mathtt{r}} \lessdot 1/\mathtt{K}^{14n+2}\,, \qquad |\partial^2_{\hat{I}} \mathtt{g}_3|_{\hat{D},3\mathtt{r}} \lessdot 1/(\sqrt{\varepsilon}\,\mathtt{K}^{\frac{37}{2}n+2})\,. \tag{195}$$

Note that the term $\partial^2_{\hat{I}}\mathtt{g}_3$ in (195) has a 'big' estimate, containing a $\sqrt{\varepsilon}$ at the denominator. However this does not cause any problem, since, by the first lines in (180), (181) and (182) one has[81]

$$\zeta\,\partial_{I_1}\mathtt{E}\,\partial^2_{\hat{I}_i\hat{I}_j}\mathtt{g}_3 = \mathtt{K}^{-\frac{37}{2}n-2}\,z(1 \oplus \log z)^2\,,$$

where, the function in brackets belongs to $\mathcal{F}$ and has norm $\|\!| \cdot \|\!|_\varrho$ bounded by a constant depending only on n and κ.

At this point, mimicking the proof for the inner case, one gets easily (187) also in the outer case $i = 0, 2N$, if one chooses $\mathbf{c}_2$ big enough. The proof of of Proposition 4.2 is complete. ∎

Step 5: *The Twist Theorem far from separatrices in the inner case*

Proposition 4.3 *Let* $0 < i < 2N$.

(i) *There exists a constant* $\mathbf{c}_5 = \mathbf{c}_5(n,\kappa) > 1$ *such that if* $\mathtt{K} \geqslant \mathbf{c}_5$, $\mathtt{N} \leqslant |k|_1 \leqslant \mathtt{K}_o$, *then, on*[82] $\mathcal{B}^i_{\mathrm{far}}$, $|\det \partial^2_I \mathtt{h}| \geqslant \delta_o/2^5$.

(ii) *There exists a suitable constant* $\hat{\mathfrak{c}}_0 = \hat{\mathfrak{c}}_0(n,\kappa,\xi,\mathtt{m}) \geqslant \mathbf{c}_5$, *such that if* $\mathtt{K} \geqslant \hat{\mathfrak{c}}_0$ *and* $\eta < \delta_o/2^5$, *then*

$$\mathrm{meas}\left(\{I \in \mathcal{B}^i_{\mathrm{far}}\ :\ |\det \partial^2_I \mathtt{h}(I)| \leqslant \eta\}\right) \leqslant \hat{\mathfrak{c}}_0 (|k|^{2n}\eta)^{\frac{1}{\mathtt{m}}}\ \mathrm{meas}\,\mathcal{B}^i_k\,. \tag{196}$$

[81] The regularizing term ζ defined in Lemma 4.8.

[82] Recall the definition of $\mathcal{B}_{\mathrm{far}} = \mathcal{B}^i_{\mathrm{far}}$ in (170).

Remark 4.3 Notice that by point (i), the set $\{I \in \mathcal{B}_{\rm far} : \ |\det \partial_I^2 \mathtt{h}(I)| \leqslant \eta\}$ is, for $\eta < \delta_o/2^5$ empty. Therefore, in proving point (ii) one needs to consider only $|k|_1 < \mathtt{N}$.

For definiteness, *in the proof of Proposition 4.3, we consider only* i *odd*, as the case i even can be treated in a completely analogous way.

First, we prove some perturbative estimates on the derivatives of the energy.

Recalling the definition of $\mathcal{I} = \mathcal{I}^i$ in (170) with λ_o as in Proposition 4.1–(ii), and notice that $\mathcal{I}$ only depends on n, κ. Then, the following estimates hold.

Lemma 4.13 *There exists* $\mathbf{c}_6 = \mathbf{c}_6(n,\kappa) > 1$ *such that, defining* $\mathtt{r}_* := \sqrt{\epsilon}/\mathbf{c}_6$*, one has, for* $I \in \mathcal{I}_{2\mathtt{r}_*} \times \hat{D}$*,*

$$|\partial_{I_1}\mathtt{E}| \leqslant \mathbf{c}_6\sqrt{\epsilon}\,, \qquad |\partial^2_{I_1}\mathtt{E}| \leqslant \mathbf{c}_6\,, \qquad \left|\partial^2_{I_1\hat{I}}\mathtt{E}\right| \leqslant \mathbf{c}_6\mu_o\,, \qquad |\partial^2_{\hat{I}}\mathtt{E}| \leqslant \mathbf{c}_6\mu_o \tag{197}$$

and

$$|\partial_{I_1}\mathtt{E} - \partial_{I_1}\bar{\mathtt{E}}| \leqslant \mathbf{c}_6\sqrt{\epsilon}\mu\,. \tag{198}$$

Proof Recall the definitions in (92) and (146). Then:

$$\begin{aligned}
I_1(E_+(\hat{I}) - \epsilon\tfrac{\lambda_o}{8}, \hat{I}) - \bar{b}_{\lambda_o} \quad &\geqslant \quad \bar{I}_1(\bar{E}_+ - \epsilon\tfrac{\lambda_o}{8}) \\
&\qquad -\bar{I}_1(\bar{E}_+ - \epsilon\tfrac{\lambda_o}{2}) - |I_1(E_+(\hat{I}) - \epsilon\tfrac{\lambda_o}{8}, \hat{I}) - \bar{I}_1(\bar{E}_+ - \epsilon\tfrac{\lambda_o}{8})| \\
&\overset{(80),(75)}{\geqslant} \quad \tfrac{1}{\mathbf{c}\sqrt{\epsilon}}\left(\tfrac{\epsilon\lambda_o}{2} - \tfrac{\epsilon\lambda_o}{8}\right) - |\phi_+(\tfrac{\lambda_o}{8}, \hat{I}) - \bar{\phi}_+(\tfrac{\lambda_o}{8})| \\
&\qquad\qquad - |\psi_+(\tfrac{\lambda_o}{8}, \hat{I}) - \bar{\psi}_+(\tfrac{\lambda_o}{8})| \cdot \tfrac{\lambda_o}{8}|\log\tfrac{\lambda_o}{8}| \\
&\overset{(77)}{\geqslant} \quad \tfrac{3\sqrt{\epsilon}\lambda_o}{8\mathbf{c}} - 2\mathbf{c}\sqrt{\epsilon}\mu \overset{(171)}{\geqslant} \tfrac{\sqrt{\epsilon}\lambda_o}{4\mathbf{c}} > 0\,,
\end{aligned}$$

which imply $\mathcal{I} \times \hat{D} \subseteq \mathcal{B}^i_k(\lambda_o/8)$.

Now, we can take $\mathbf{c}_6 > 1$ big enough so that (recall (96)) $\mathcal{I}_{2\mathtt{r}_*} \times \hat{D} \subseteq \big(\mathcal{B}^i_k(\lambda_o/8)\big)_{\rho_{\lambda_o/8}}$. Thus, by (97) we get[83] (197).

Now, observe that by the definitions in (170) and (146) we get $\bar{\mathtt{E}}(\mathcal{I}) = (\bar{E}_-, \bar{E}_+ - \epsilon\lambda_o/2)$. Then, recalling (82), by the first estimate in (197), we get

$$\bar{\mathtt{E}}(\mathcal{I}_{2\mathtt{r}_*}) \subseteq \mathcal{E}_{\lambda_o/8}\,, \qquad \mathtt{E}(\mathcal{I}_{2\mathtt{r}_*}, \hat{I}) \subseteq \mathcal{E}_{\lambda_o/8}\,, \quad \forall\, \hat{I} \in \hat{D}\,, \tag{199}$$

taking $\mathbf{c}_6$ big enough.

Let us, now, prove (198). Observe that

$$\partial_{I_1}\mathtt{E}(I) - \partial_{I_1}\bar{\mathtt{E}}(I_1) = \big(\partial_E \bar{I}_1(\bar{\mathtt{E}}(I_1)) - \partial_E I_1(\mathtt{E}(I_1), \hat{I})\big)\partial_{I_1}\mathtt{E}(I) \cdot \partial_{I_1}\bar{\mathtt{E}}(I_1)\,,$$

so that

$$\sup_{\mathcal{I}_{2\mathtt{r}_*} \times \hat{D}} |\partial_{I_1}\mathtt{E}(I) - \partial_{I_1}\bar{\mathtt{E}}(I_1)|$$

[83] Obviously the first two estimates holds also for $\bar{\mathtt{E}} = \mathtt{E}|_{\mu=0}$.

$$\overset{(199)}{\leqslant} \sup_{\mathcal{E}_{\lambda_o/8}\times\hat{D}} \left|\partial_E \bar{I}_1(E) - \partial_E I_1(E,\hat{I})\right| \cdot \sup_{\mathcal{I}_{2\mathtt{r}_*}\times\hat{D}} \left|\partial_{I_1}\mathtt{E}(I)\cdot\partial_{I_1}\bar{\mathtt{E}}(I_1)\right| \overset{(83),(197)}{<} \sqrt{\epsilon}\mu\,. \quad \blacksquare$$

Next, we provide perturbative estimates on the twist.
By Cauchy estimates, from (198), there follows

$$\sup_{\mathcal{I}_{\mathtt{r}_*}\times\hat{D}} |\partial^2_{I_1}\mathtt{E} - \partial^2_{I_1}\bar{\mathtt{E}}| < \mu\,. \tag{200}$$

Hence, by (155) and (197), on $\mathcal{I}_{\mathtt{r}_*}\times\hat{D}$, we get[84]

$$\det\partial^2_I \mathtt{h} = (\partial^2_{I_1}\mathtt{E})\cdot\det\partial^2_{\hat{I}}\hat{h}_{_k} + O(\mu_o) = (\partial^2_{I_1}\bar{\mathtt{E}} + O(\mu))\cdot\det\partial^2_{\hat{I}}\hat{h}_{_k} + O(\mu_o)\,. \tag{201}$$

Now, by (163), (76), one has that $\delta_o^{-1} \leqslant \mathtt{K}^{2n}$ and $\mu_o/\delta_o = O(\mathtt{K}^{-3n})$. Finally, since, by (48), $\mu = 1/\mathtt{K}^{5n}$, from (201) one gets at once the following

Lemma 4.14 *Let $\mathtt{r}_*$ be as in Lemma 4.13, $0 \leqslant i \leqslant 2N$, and $\mathcal{I} = \mathcal{I}^i$ as in (170). Then,*

$$\left|\det\partial^2_I\mathtt{h}(I)\right| \geqslant \delta_o|g(I)|\,, \qquad \forall\, I\in\mathcal{I}_{\mathtt{r}_*}\times\hat{D}\,, \tag{202}$$

with

$$g(I) = \partial^2_{I_1}\bar{\mathtt{E}}(I_1) + O(\mathtt{K}^{-3n})\,, \quad \forall\, I\in\mathcal{I}_{\mathtt{r}_*}\times\hat{D}\,. \tag{203}$$

Proof of Proposition 4.3 (i) Since $|k|_1 \geqslant \mathtt{N}$, $\bar{\mathtt{G}} = \frac{2\varepsilon}{|k|^2}\pi_{\mathbb{Z}k} f$ in (55) is close to a cosine, as proved in Lemma A.1 in Appendix. Hence, (84) in Proposition 2.1 holds, so that by (202) and (203), taking $\mathbf{c}_5$ large enough and $\mathtt{K} \geqslant \mathbf{c}_5$, the claimed estimate $|\det\partial^2_I\mathtt{h}| \geqslant \delta_o/2^5$ follows.
(ii) Recall (146). Since $\lambda \to \bar{b}_\lambda$ is a decreasing function, we get $\bar{b}_{\lambda_o/2} \leqslant \bar{b} = \bar{b}_0$. Rescaling we get $\tilde{\mathcal{I}} := (0, \bar{b}_{\lambda_o/2}/\bar{b}) \subseteq (0,1)$ so that $\bar{b}\tilde{\mathcal{I}} = \mathcal{I}$. Recalling (92), by (75)÷(78) we have that $\bar{b} < \sqrt{\epsilon}$. Then, choosing $0 < \tilde{\mathtt{r}} \leqslant 1$ small enough, we have that

$$\bar{b}\tilde{\mathcal{I}}_{2\tilde{\mathtt{r}}} \subseteq \mathcal{I}_{\mathtt{r}_*}\,. \tag{204}$$

By (147) we get

$$\partial^2_{I_1}\bar{\mathtt{E}}(I_1) = \mathtt{F}_{\bar{\mathtt{G}}}(I_1/\bar{b})\,. \tag{205}$$

By (203) and (204) we get

$$g(\bar{b}x,\hat{I}) = \mathtt{F}_{\bar{\mathtt{G}}}(x) + O(\mu) + O(\mu_o/\delta_o) \quad \text{uniformly for} \quad (x,\hat{I})\in\tilde{\mathcal{I}}_{2\tilde{\mathtt{r}}}\times\hat{D}\,. \tag{206}$$

By (197), (204) and (205) we get

$$\sup_{\tilde{\mathcal{I}}_{2\tilde{\mathtt{r}}}} |\mathtt{F}_{\bar{\mathtt{G}}}| < 1\,. \tag{207}$$

For $\hat{I}\in\hat{D}$, set[85]

$$\mathcal{I}'_\eta(\hat{I}) := \left\{I_1\in\mathcal{I} : \left|\det\partial^2_I\mathtt{h}^i(I)\right| \leqslant \eta\right\}, \quad \tilde{\mathcal{I}}'_\eta(\hat{I}) := \{x\in\tilde{\mathcal{I}} : |g(\bar{b}x,\hat{I})| \leqslant \eta/\delta_o\}\,. \tag{208}$$

[84] Recall Notation 4.2.
[85] Recall that $\mathcal{I} := (0, \bar{b}_{\lambda_o/2})$ in (170).

By (208) and (202) we have that, for every $\hat{I} \in \hat{D}$,

$$\operatorname{meas} \mathcal{I}'_\eta(\hat{I}) \leqslant \bar{b} \operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I}) < \sqrt{\epsilon} \operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I}) \,. \tag{209}$$

using $\bar{b} < \sqrt{\epsilon}$. Before estimating $\operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I})$ we need the following bound:

$$1/2 \leqslant \bar{b}_{\lambda_o/2}/\bar{b} \,. \tag{210}$$

Recalling (92) we have

$$\bar{b} = \bar{I}_1(\bar{E}_+) = \int_{\bar{E}_-}^{\bar{E}_+} \partial_E \bar{I}_1 \overset{(80)}{\geqslant} \frac{\bar{E}_+ - \bar{E}_-}{\mathbf{c}\sqrt{\epsilon}} \overset{(90)}{\geqslant} \frac{\sqrt{\epsilon}}{\mathbf{c}\kappa} \,. \tag{211}$$

Recalling (170) we have by[86] (83) that, for $0 < z \leqslant \lambda_o/2$, $\partial_E \bar{I}_1(\bar{E}_+ - \epsilon z) \leqslant \mathbf{c}^2 |\log z|/\sqrt{\epsilon}$. Therefore

$$\begin{aligned} 1 - \frac{\bar{b}_{\lambda_o/2}}{\bar{b}} \quad &= \quad \frac{\bar{I}_1(\bar{E}_+) - \bar{I}_1(\bar{E}_+ - \epsilon\lambda_o/2)}{\bar{b}} = \frac{\epsilon}{\bar{b}} \int_0^{\lambda_o/2} \partial_E \bar{I}_1(\bar{E}_+ - \epsilon z) dz \\ &\overset{(211)}{\leqslant} \quad \mathbf{c}^3 \kappa \int_0^{\lambda_o/2} |\log z| dz \leqslant \mathbf{c}^4 \lambda_o \log|\lambda_o| \overset{(171)}{\leqslant} \frac{1}{2} \,, \end{aligned}$$

proving (210).
Let us come back to the estimate of $\operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I})$. Recalling Definition 4.3 we have that $\mathtt{F}$ is ξ–non–degenerate at order $\mathtt{m}$. By (206), (207) and Cauchy estimates, taking μ and[87] μ_o/δ_o small enough (i.e., $\mathtt{K} \geqslant \hat{\mathfrak{c}}_0$ for a suitable $\hat{\mathfrak{c}}_0$ arge enough) depending only on κ, n, ξ and $\mathtt{m}$ we have that the function $x \mapsto g(\bar{b}x, \hat{I})$ is $(\xi/2)$–non–degenerate at order $\mathtt{m}$. Now we want to apply Lemma 4.1 with η replaced by η/δ_o, and with the following choices:

$$f(x) = g(\bar{b}x, \hat{I}) \,, \quad m = \mathtt{m} \,, \quad a = 0 \,, \quad 1/2 \overset{(210)}{\leqslant} b = \bar{b}_{\lambda_o/2}/\bar{b} < 1 \,, \quad \xi = \xi/2 \,;$$

the constant M controlling the derivatives of f, by (206), (207) and Cauchy estimates can be bounded by $1 \leqslant M \leqslant c_{n,\kappa}/\tilde{\mathtt{r}}^{\mathtt{m}+1}$ for[88] for a suitably large constant $c_{n,\kappa}$ depending only on n and κ. In conclusion, by Lemma 4.1, we get

$$\operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I}) \leqslant c_{\mathtt{m}} \Big(\frac{2c_{n,\kappa}}{\xi \tilde{\mathtt{r}}^{\mathtt{m}+1}} + 1 \Big) \Big(\frac{\eta}{\delta_o \xi} \Big)^{\frac{1}{\mathtt{m}}} \,.$$

Then (196) follows by (163), (209) Fubini's theorem and (194). The proof of Proposition 4.3 is complete. ∎

Step 6: *Uniform twist in outer regions far from separatrices*
Recall the definition of the twist $\delta_\sharp$ in the outer regions in (161), and that $\delta_o = |k|^{-2n}$ (see (153)).

Proposition 4.4 *Let $i = 0, 2N$. Then, there exists a suitable constant $\mathbf{c}_7 = \mathbf{c}_7(n,\kappa) > 1$ such that if $\mathtt{K} \geqslant \mathbf{c}_7$, then on $\mathcal{B}^i_{\rm far}$*

$$|\delta_\sharp| \geqslant \delta_o/2 \,.$$

[86] Note that condition (81) reduces here to $\lambda \leqslant 1/\mathbf{c}$ since we are considering $\bar{I}_1$, namely the case $\mu = 0$. In any case one can prove the estimate also directly by (75) and (76).
[87] δ_o is defined in (153).
[88] Recall that $\tilde{\mathtt{r}} = \tilde{\mathtt{r}}(n,\kappa)$ was chosen in (204) small enough.

Proof Taking $\lambda = \lambda_o/2$ in (97), on $\mathcal{B}^i_{\text{far}}$, we have that[89]

$$|\hat{\mathrm{v}}|,\ |\hat{\mathrm{M}}|\ < 1/\mathtt{K}^{\frac{19}{2}n-1}\,. \tag{212}$$

By (48) and (76), we also have

$$\frac{\sqrt{\epsilon}}{\mathtt{r}} < \frac{1}{\mathtt{K}^{\frac{9}{2}n}}\,, \qquad \mu_o < \frac{1}{\mathtt{K}^{\frac{19}{2}n+1}}\,.$$

Then, recalling Definition 2.4 and (48), we get $|\mathtt{E}| \leqslant 2\mathtt{R}^2 \leqslant 2\varepsilon\mathtt{K}^{9n+4}$.
Now, it is a general fact that, in the outer case, the unperturbed energy function is strictly concave, as it follows from the following simple consequence of Jensen's inequality[90].

Lemma 4.15 *Let $i = 0, 2N$. Then, for every $E > \bar{E}_0 = \bar{E}_{2N}$, $\partial^2_{I_1}\bar{\mathtt{E}}^i(\bar{I}^i_1(E)) \geqslant 2$.*

The proof is given in Appendix.
Now, since estimate (200) still holds in the present case $i = 2N$ we get by Lemma 4.15, $\partial^2_{I_1}\mathtt{E} \geqslant \frac{1}{2}\partial^2_{I_1}\bar{\mathtt{E}} \geqslant 1$, so that the claim follows by (161), (212) and (163). ∎

Step 7: *Conclusion*

Proof of the Twist Theorem 4.1 Let $\lambda_o = \lambda_o(n, \kappa)$ be as in Proposition 4.1–(ii), and let $\mathbf{c}_8 = \mathbf{c}_8(n, \kappa) \geqslant 1$ be such that the second estimate in (171) holds if[91] $\mathtt{K} = 1/\mu^{5n} \geqslant \mathbf{c}_8$. Then, by Lemma 4.7,

$$\mathcal{B}^i_k = \mathcal{B}^i_{\text{near}}(\lambda_o) \cup \mathcal{B}^i_{\text{far}}(\lambda_o)\,, \qquad \forall\ 0 \leqslant i \leqslant 2N\,. \tag{213}$$

Define[92]

$$\mathfrak{c}_0 := 2\max\{\mathbf{c}_2, \mathbf{c}_5, \mathbf{c}_6, \mathbf{c}_7, \mathbf{c}_8, \hat{\mathfrak{c}}_0\}\,. \tag{214}$$

Let us consider first the outer case $i = 0, 2N$. Recall the definition of $\mathtt{b}$ in (154). By Lemma 4.4, Proposition 4.4, Proposition 4.2, and by (214), we find

$$\begin{aligned}
\operatorname{meas}\left(\left\{ I \in B^i_k :\ \left|\det \partial^2_I \mathtt{h}^i_k(I)\right| \leqslant \eta\right\}\right) &= \operatorname{meas}\left(\left\{ I \in \mathcal{B}^i_k :\ \left|\delta^i_\sharp(I)\right| \leqslant \eta\right\}\right)\\
&= \operatorname{meas}\left(\left\{ I \in \mathcal{B}^i_{\text{near}}(\lambda_o) :\ \left|\det \partial^2_I \mathtt{h}^i(I)\right| \leqslant \eta\right\}\right)\\
&\leqslant \mathbf{c}_2\, (|k|^{2n}\eta)^{1/9n^4}\ \operatorname{meas}\mathcal{B}^i_k\\
&\overset{(162)}{=} \mathbf{c}_2\, (|k|^{2n}\eta)^{1/9n^4}\ \operatorname{meas} B^i_k\\
&\overset{(214)}{<} \mathfrak{c}_0\, (|k|^{2n}\eta)^{\mathtt{b}}\ \operatorname{meas} B^i_k\,,
\end{aligned}$$

proving Theorem 4.1 in the outer case $i = 0, 2N$.
In the inner case $0 < i < 2N$, $B^i_k = \mathcal{B}^i_k$ (compare (108)) and, since $\mathtt{K} \geqslant \mathfrak{c}_0$, (154) follows by (213), (187) in Proposition 4.2, and (196) in Proposition (4.3). ∎

---

[89] The quantities $\hat{\mathrm{v}}$ and $\hat{\mathrm{M}}$ are defined in (160).
[90] $\bar{E}_i$ is defined in (68).
[91] Recall (48).
[92] Recall that $\mathbf{c}_2 \geqslant \mathbf{c}_4 \geqslant \max\{\mathbf{c}_1, \mathbf{c}_3\} \geqslant c_3$.

5 Maximal KAM tori and proof of the main results

In this final section we show that primary and secondary maximal KAM tori of $\mathtt{H}$ span the complementary of $\mathcal{R}^2 \times \mathbb{T}^n$ apart from an exponentially small (in $1/\mathtt{K}$) set and prove the results in Section 1. To construct such tori we shall use the following 'KAM theorem'.

Theorem 5.1 **([10])** *Fix $n \geqslant 2$ and let $\mathtt{D}$ be any non–empty, bounded subset of $\mathbb{R}^n$. Let*

$$\mathrm{H}(\mathrm{p},\mathrm{q}) := \mathrm{h}(\mathrm{p}) + \mathrm{f}(\mathrm{p},\mathrm{q})$$

be real analytic on $\mathtt{D}_{\mathfrak{r}} \times \mathbb{T}^n_{\mathfrak{s}}$, for some $\mathfrak{r} > 0$ and $0 < \mathfrak{s} \leqslant 1$, and having finite norms

$$\mathtt{M} := |\partial^2_{\mathrm{p}}\mathrm{h}|_{\mathfrak{r}}\,, \qquad\qquad |\mathrm{f}|_{\mathfrak{r},\mathfrak{s}}\,. \tag{215}$$

Assume that the frequency map $\mathrm{p} \in \mathtt{D} \to \omega = \partial_{\mathrm{p}}\mathrm{h}$ is a local diffeomorphism, namely, assume:

$$\mathrm{d} := \inf_{\mathtt{D}} |\det \partial^2_{\mathrm{p}}\mathrm{h}| > 0\,, \tag{216}$$

and let $\mathrm{d}_* := \mathrm{d}/\mathtt{M}^n$ and $r_* := \mathrm{d}_*^2\mathfrak{r}$. Then, there exists $\mathrm{C}_* = \mathrm{C}_*(n) > 1$ such that, if

$$\epsilon := \frac{|\mathrm{f}|_{\mathfrak{r},\mathfrak{s}}}{\mathtt{M}\mathfrak{r}^2} \leqslant \frac{\mathrm{d}_*^8\ \mathfrak{s}^{4(\mathrm{n}+1)}}{\mathrm{C}_*}\ , \tag{217}$$

there exists a set $\mathcal{T} \subseteq (\mathtt{D}_{\mathrm{r}_*} \cap \mathbb{R}^{\mathrm{n}}) \times \mathbb{T}^{\mathrm{n}}$ formed by primary KAM tori such that[93]

$$\mathrm{meas}\,\big((\mathtt{D}\times\mathbb{T}^n)\setminus\mathcal{T}\big) \leqslant \mathrm{C}\,\sqrt{\epsilon}\,, \qquad \mathrm{C} := \big(\max\big\{\mathrm{d}_*^2\mathfrak{r}\,,\,\mathrm{diam}\,\mathtt{D}\big\}\big)^{\mathrm{n}}\cdot\frac{\mathrm{C}_*}{\mathrm{d}_*^{n+5}\ \mathfrak{s}^{3(\mathrm{n}+1)}}\,. \tag{218}$$

This statement is an immediate corollary of Theorem 1 in[94] [10].

Remark 5.1 (i) Note that in the formulation of Theorem 5.1 the action domain $\mathtt{D}$ is a completely arbitrary bounded set and that the smallness quantitative condition (217) depends on $\mathtt{D}$ only through its diameter, which in our application depends on k. For a similar statement, which takes into account the geometry of $\mathtt{D}$, see [19].
(ii) We point out that the smallness condition (217) can be rewritten as

$$|\mathrm{f}|_{\mathtt{D},\mathfrak{r},\mathfrak{s}} \leqslant \frac{\mathfrak{r}^2\mathrm{d}^8\ \mathfrak{s}^{4n+4}}{\mathrm{C}_*\,\mathtt{M}^{8n-1}}\,. \tag{219}$$

(iii) Finally, observe that, since[95] $\mathrm{d}_* \leqslant 1$, estimate (218) implies

$$\mathrm{meas}\,\big((\mathtt{D}\times\mathbb{T}^n)\setminus\mathcal{T}\big) \leqslant \big(\max\big\{\mathfrak{r}\,,\,\mathrm{diam}\,\mathtt{D}\big\}\big)^n\cdot\frac{\mathrm{C}_*\,\mathtt{M}^{n^2+5n-1/2}}{\mathrm{d}^{n+5}\ \mathfrak{s}^{3n+3}\mathfrak{r}}\sqrt{|\mathrm{f}|_{\mathtt{D},\mathfrak{r},\mathfrak{s}}}\,. \tag{220}$$

[93]Here 'meas' denotes the outer Lebesgue measure.

[94]In Theorem 1 of [10] take $\tau = n$ and substitute λ with its maximal value $2\cdot n!\,\mathrm{d}_*^{-1}$ (see (14) of [10]).

[95]Indeed the absolute value of any eigenvalues of the symmetric matrix $\partial^2_{\mathrm{p}}\mathrm{h}$ is bounded by $\mathtt{M}$, which implies $\mathrm{d} \leqslant \sup_{\mathtt{D}} |\det \partial^2_{\mathrm{p}}\mathrm{h}| \leqslant \mathtt{M}^n$.

KAM tori in the non–resonant region

Proposition 5.1 *Let the assumptions of Theorem 2.1 hold. There exists a constant* $\mathtt{C}_{\rm o} = \mathtt{C}_{\rm o}(n,s) \geqslant \mathbf{c_o}$ *such that, if* $\mathtt{K}_{\rm o} \geqslant \mathtt{C}_{\rm o}$*, then there exists a family of primary maximal KAM tori* $\mathcal{T}^0$ *invariant for the Hamiltonian* $\mathtt{H}$ *in* (1)*, satisfying*

$$\mathrm{meas}\,\big((\mathcal{R}^0 \times \mathbb{T}^n)\setminus \mathcal{T}^0\big) \leqslant \mathtt{C}_{\rm o}\,\sqrt{\varepsilon}\, e^{-\mathtt{K}_{\rm o} s/6}\,. \tag{221}$$

Remark 5.2 The above result is essentially classical, and, in fact, no genericity assumptions on the potentials are needed. However, there is one delicate point related to the KAM tori near the boundary. Indeed, primary tori oscillates, in general, by a quantity of order $\sqrt{\varepsilon}$, and naive applications of classical KAM theorems would leave out regions near the boundary of the phase space of measure $\sim \sqrt{\varepsilon}$. Such a problem is overcome by using the second covering in (34) in Theorem 2.1, which is introduced so that (40) holds; compare, also, Remark 2.2–(ii).

Proof of Theorem 5.1 We apply the KAM Theorem 5.1 to the nearly–integrable Hamiltonian $\mathtt{H}_{\rm o}$ in Theorem 2.1–(ii). More precisely, we let[96]

$$\mathrm{h}(\mathrm{p}) = \frac{|\mathrm{p}|^2}{2} + \varepsilon g^{\rm o}(\mathrm{p})\,,\quad \mathrm{f} = \varepsilon f^{\rm o}\,,\quad \mathtt{D} = \widetilde{\mathcal{R}}^0\,,\quad \mathfrak{r} = \frac{r_{\rm o}'}{2} = \frac{\sqrt{\varepsilon}\mathtt{K}^{\frac{9}{2}n+2}}{16\mathtt{K}_{\rm o}}\,,\quad \mathfrak{s} = \min\{\tfrac{s}{2}\,,1\}\,.$$

By (36) and Cauchy estimates we get

$$\mathtt{M} \leqslant 2\,,\qquad |\mathrm{f}|_{\mathfrak{r},\mathfrak{s}} \leqslant \varepsilon e^{-\mathtt{K}_{\rm o} s/3}\,,\qquad \mathrm{d} \geqslant 1/2\,.$$

If $\mathtt{K}_{\rm o}$ is taken large enough (larger than a constant despending on n and s) the KAM smallness condition (219) is satisfied, and the KAM Theorem 5.1 yields the existence of a set $\widetilde{\mathcal{T}}^0$ of invariant tori for the Hamiltonian $\mathtt{H}_{\rm o}$ in Theorem 2.1–(ii), which, by[97] (220), satisfy

$$\mathrm{meas}\,\big((\widetilde{\mathcal{R}}^0 \times \mathbb{T}^n)\setminus \widetilde{\mathcal{T}}^0\big) \leqslant \mathtt{C}_{\rm o}\sqrt{\varepsilon}\, e^{-\mathtt{K}_{\rm o} s/6}\,, \tag{222}$$

for a suitable constant $\mathtt{C}_{\rm o} = \mathtt{C}_{\rm o}(n,s)$ large enough (so that also the condition on $\mathtt{K}_{\rm o}$ is met). Since the map $\Psi_{\rm o}$ in (40) is symplectic, the family of tori $\mathcal{T}^0 := \Psi_{\rm o}(\widetilde{\mathcal{T}}^0)$ is formed by KAM invariant for $\mathtt{H}$ in (1). The first relation in (40) and the bound (222) imply (221). ∎

KAM tori near simple resonances

Now, we turn to the construction, in all neighbourhoods of simple resonances, of families of primary tori for the nearly–integrable Hamiltonians $\mathcal{H}_k^i$ of Theorem 3.1, for all $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$ and $0 \leqslant i \leqslant 2N_k$. Note that such tori correspond, in the inner case $0 < i < 2N_k$, to *secondary tori for the Hamiltonian* $\mathtt{H}$.

Let us introduce zones $B_k^i(\lambda,\eta) \subseteq B_k^i$, which are λ–away in energy from separatrices and where the twist is bounded away from zero by a quantity $\eta > 0$, namely (recall (111), (108)), let us define:

$$B_k^i(\lambda,\eta) := \{I \in B_k^i(\lambda) \text{ s.t. } |\det \partial_I^2 \mathtt{h}_k^i(I)| > \eta\} \subset B_k^i\,. \tag{223}$$

[96] Recall Theorem 2.1, the definitions (20), (32) and (21).

[97] Notice that the hypothesis $\mathtt{K} < \varepsilon^{-1/(9n+4)}$ implies that $\mathfrak{r} < 1$, so that $\max\{\mathrm{d}^2\mathtt{M}^{-2n}\mathfrak{r}\,,\,\mathrm{diam}\,\mathtt{D}\} = 2$.

Proposition 5.2 (KAM tori for $\mathcal{H}_k^i$) *Let the assumptions of Theorem 3.1 hold. There exist positive constants $\bar{\mathtt{C}}_1 = \bar{\mathtt{C}}_1(n,s,\beta) > 1$ and $\mathtt{C}_1 = \mathtt{C}_1(n,s,\beta,\delta) \geqslant \mathbf{c}_*$ such that the following holds. Let $k \in \mathcal{G}^n_{\mathtt{K}_o}$, $0 \leqslant i \leqslant 2N_k$; $0 < \lambda \leqslant 1/\mathbf{c}_*$ and $0 < \eta < 1/2$. Then, if*

$$\mathtt{K} \geqslant \mathtt{C}_1 \log \frac{1}{\lambda\eta}\,, \tag{224}$$

there exists a set $\mathcal{T}_k^i$ of maximal KAM tori for the Hamiltonian $\mathcal{H}_k^i$ in (110) such that

$$\operatorname{meas}\big((B_k^i(\lambda,\eta)\times\mathbb{T}^n)\setminus\mathcal{T}_k^i\big) \leqslant \bar{\mathtt{C}}_1 e^{-\mathtt{K}s/7}\,. \tag{225}$$

Proof We apply the KAM Theorem 5.1 to the Hamiltonian $\mathcal{H}_k^i$ of Theorem 3.1 with (recall (110) and (111)):

$$\begin{aligned} &\mathrm{h} = h_k^i = \tfrac{|k|^2}{2}\mathtt{h}_k^i\,, \qquad \mathrm{f} = \varepsilon f_k^i\,, \qquad \mathtt{D} = B_k^i(\lambda,\eta)\,,\\ &\mathfrak{r} = \rho_* = \frac{\sqrt{\epsilon}}{\mathbf{c}_*\mathtt{K}_o^n}\lambda|\log\lambda|\,, \qquad \mathfrak{s} = \sigma_* = \frac{1}{\mathbf{c}_*\mathtt{K}_o^n|\log\lambda|}\,. \end{aligned} \tag{226}$$

Note that, by (114) and (56), $0<\lambda\leqslant 1/\mathbf{c}_* \leqslant 1/8c_2$, which implies easily $\mathfrak{r}\leqslant\mathtt{r}$ and $\mathfrak{s}\leqslant 1$. Also, since $\mathbf{c}_*\geqslant\hat{\mathbf{c}}$ (see Theorem 3.1) and $\mathtt{K}_o^n\geqslant 2^n\geqslant n$, one has $\rho_*\leqslant\rho_\lambda/n$.

In the following arguments we denote by $c(\cdot)$ possibly different constants depending only on the quantities inside brackets.

We first have to estimate $\mathtt{M}$ in (215), namely, $\partial_I^2\mathtt{h}_k^i$. By (97), (76) and (74) we get

$$\sup_{(\mathcal{B}_k^i(\lambda))_{\rho_\lambda}} \big|\partial_I^2\mathtt{E}^i\big| \leqslant \frac{n\hat{\mathbf{c}}}{\lambda}\,. \tag{227}$$

In the case $0<i<2N_k$, by (111), we have $\mathcal{B}_k^i(\lambda) = B_k^i(\lambda)$. Therefore, recalling (110), we can bound $|\partial_I^2\mathtt{h}_k^i|$ by $c(n,s,\beta)/\lambda$.

The estimate on $|\partial_I^2\mathtt{h}_k^i|$ in the case $i=0,2N_k$ needs some extra attention. In particular fix $i=2N_k$ (the case $i=0$ being analogous). Recalling the definition of $j_{\mathtt{g}_*}$ in (106), (108) we have that $\partial_I^2 j_{\mathtt{g}_*}$ depends only on $\hat{I}$ and not on I_1. Moreover by (20), (33), (59), (48) and Cauchy estimates we get

$$\sup_{\hat{I}\in\hat{D}_{3\mathtt{r}}} |\partial_I j_{\mathtt{g}_*}| \leqslant c(n)\,, \qquad \sup_{\hat{I}\in\hat{D}_{3\mathtt{r}}} |\partial_I^2 j_{\mathtt{g}_*}| \leqslant \frac{c(n)|k|^2}{\sqrt{\varepsilon}\mathtt{K}^\nu}\,. \tag{228}$$

Recalling Definition 2.4, (97) and (48), we have that

$$\sup_{(\mathcal{B}_k^{2N_k}(\lambda))_{\rho_\lambda}} |\mathtt{E}^{2N_k}| \leqslant 4\mathtt{R}^2 = \frac{4\varepsilon\mathtt{K}^{2\nu}}{|k|^4}\,.$$

Then, by (97) and (48) we get

$$\sup_{(\mathcal{B}_k^{2N_k}(\lambda))_{\rho_\lambda}} |\partial_{I_1}\mathtt{E}^{2N_k}| \leqslant \hat{\mathbf{c}}\sqrt{8\mathtt{c}_s\varepsilon + 4\varepsilon\mathtt{K}^{2\nu}|k|^{-4}} \leqslant 4\hat{\mathbf{c}}\sqrt{\varepsilon}\mathtt{K}^\nu|k|^{-2}$$

(taking $\mathtt{K} \geqslant \mathtt{c}_s$ defined in (48)). Finally, recalling also (111), (124), (227), (228) , we get by the chain rule

$$\sup_{(B_k^{2N_k}(\lambda))_{\mathfrak{r}}} \left|\partial_I^2 \big(\mathtt{E}^{2N_k} \circ I_*\big)\right| \leqslant \frac{c(n,s,\beta)}{\lambda} \, .$$

By (215), (226), (110), (163) (and that $\mathfrak{r} \leqslant \mathtt{r}$), we finally get

$$\mathtt{M} \leqslant |k|^2 \frac{c(n,s,\beta)}{\lambda} \, , \qquad \forall \; 0 \leqslant i \leqslant 2N_k \, . \tag{229}$$

Next, by (226) and (113),

$$|\mathrm{f}|_{\mathfrak{r},\mathfrak{s}} \leqslant \varepsilon \, e^{-\mathtt{K}s/3} \, . \tag{230}$$

By (216), (226) and (223), we get

$$\mathrm{d} \geqslant 2^{-n} |k|^{2n} \eta \qquad \text{and} \qquad \frac{\mathtt{M}^n}{\mathrm{d}} \leqslant \frac{c(n,s,\beta)}{\lambda^n \eta} \, . \tag{231}$$

By (48), (16) and using $\mathtt{K}_{\mathrm{o}} \leqslant 6\mathtt{K}$, we have:

$$\frac{\varepsilon}{\epsilon} \leqslant \frac{\mathtt{K}_{\mathrm{o}}^{n+2}}{8\mathtt{c}_s \delta} \, e^{\mathtt{K}_{\mathrm{o}} s} \leqslant \frac{\mathtt{K}^{n+2}}{6^{n+3} \mathtt{c}_s \delta} \, e^{\mathtt{K}s/6} \, . \tag{232}$$

It is now easy to check, by (229), (230), (231) and (232), that the KAM smallness condition (219) is satisfied taking $\mathtt{K}$ as in (224) with $\mathtt{C}_1$ large enough. By the KAM Theorem 5.1 we, then, obtain a set $\mathcal{T}_k^i$ of invariant tori for the Hamiltonian in (110), which, in view of (220) and by (229), (230), (231) and (232), satisfies (225) with a suitable constant $\bar{\mathtt{C}}_1 = \bar{\mathtt{C}}_1(n,s,\beta)$; in particular, note that, by (140) and (226), the maximum in (220) is estimated by $c(n)\mathtt{K}_{\mathrm{o}}^{n^2}$. ∎

Putting together these KAM statements and the Twist Theorem 4.1, the proof of the results stated in Section 1 follow easily.

Proof of Theorem 1.1 and its corollaries

By Lemma 1.1, since $f \in \mathbb{G}_s^n$, there exist $\delta, \beta > 0$ such that (16) and (17) hold with $\mathtt{N}$ as in (15). Let

$$\mathtt{K}_{\mathrm{o}} := \mathtt{K}/6 \, ,$$

with $\mathtt{K} \geqslant 12$ and let α be as in (20). Then, Assumptions 2.1 hold, and we may let the Definitions 2.3 hold. Let $c_2 = c_2(n)$ and $\mathbf{c_o} = \mathbf{c_o}(n,s,\delta)$ be as in Theorem 2.2, and assume that[98]

$$\mathtt{K} \geqslant 6 \max\{c_2, \mathbf{c_o}\} \, . \tag{233}$$

Then, Theorem 2.2 holds and we may define the parameters $\xi > 0$ and $\mathtt{m} \geqslant 1$ as in Definition 4.3 with respect to standard Hamiltonians $\mathtt{H}_k$ (with $|k|_1 \leqslant \mathtt{K}_{\mathrm{o}}$) of Theorem 2.2–(ii).

We now let $\mathtt{b} < 1$ as in (154), $\mathtt{C}_1 = \mathtt{C}_1(n,s,\beta,\delta)$ be as in Proposition 5.2, and define

$$\eta := e^{-\frac{\mathtt{K}}{\mathtt{C}_1(1+\mathtt{b})}} \, , \qquad \lambda := \eta^{\mathtt{b}} \, . \tag{234}$$

[98]Eq. 233 implies that $\mathtt{K} > 12$.

Notice that, with such definitions, it is

$$\mathtt{K} = \mathtt{C}_1 \log \frac{1}{\lambda\eta}\,, \tag{235}$$

(compare (224)).
With these premises, let us turn to the proof of the claims of Theorem 1.1.
Claim (ii) has already been proven in Lemma 2.1 above.
Next, we define the set of maximal KAM tori $\mathcal{T}$ for $\mathtt{H}$ as it appears in item (iv) of the theorem.
Let $\mathtt{C}_o = \mathtt{C}_o(n,s)$ as in Proposition 5.1.There exists a constant

$$\hat{\mathfrak{c}} = \hat{\mathfrak{c}}(n,s,\beta,\delta,\mathtt{m}) \geqslant \max\{\mathtt{C}_o, 2\mathtt{C}_1\,\mathbf{c}_*/\mathtt{b}\}\,,$$

such that, if $\mathtt{K} \geqslant \hat{\mathfrak{c}}$, then

$$\mathtt{K}^{2n}\eta \overset{(234)}{=} \mathtt{K}^{2n} e^{-\frac{\mathtt{K}}{\mathtt{C}_1(1+\mathtt{b})}} \leqslant 1\,.$$

Assume that

$$\mathtt{K} \geqslant \hat{\mathfrak{c}}\,. \tag{236}$$

Then, $\lambda = \eta^{\mathtt{b}}$ in (234) is smaller than $1/\mathbf{c}_*$ and (recall (153))

$$\eta \leqslant \frac{\delta_o}{2^5} < \frac{1}{2}\,. \tag{237}$$

Thus, in view of (235), by (236) the assumptions of Propositions 5.1 and 5.2 are satisfied, and we can define the following families of tori[99]:

$$\begin{cases} \mathcal{T}_i^{1,k} := \phi_k^i(\mathcal{T}_k^i)\,, \quad \mathcal{T}^{1,k} := \displaystyle\bigcup_{0\leqslant i\leqslant 2N_k} \mathcal{T}_i^{1,k}\,, \quad \mathcal{T}^1 := \bigcup_{k\in\mathcal{G}^n_{\mathtt{K}_o}} \mathcal{T}^{1,k}\,, \\ \mathcal{T} := \mathcal{T}^0 \cup \mathcal{T}^1\,. \end{cases} \tag{238}$$

Observe that $\mathcal{T}_k^i$ are invariant tori for $\mathcal{H}_k^i$ in (110), while $\mathcal{T}_i^{1,k}$, $\mathcal{T}^1$ and $\mathcal{T}^0$ are invariant for the original Hamiltonian $\mathtt{H}$.
Thus, $\mathcal{T}$ is a family of maximal KAM tori for $\mathtt{H}$ as in item (iv) of Theorem 1.1.
Claim (i) follows, now, immediately by (23), setting

$$\mathcal{A} := \big((\mathcal{R}^0 \cup \mathcal{R}^1) \times \mathbb{T}^n\big) \setminus \mathcal{T}\,. \tag{239}$$

It remains to prove claim (iii), namely, the exponential measure estimate on $\mathcal{A}$.
Observe that by (239) and (238)

$$\mathcal{A} \subseteq \big((\mathcal{R}^0 \times \mathbb{T}^n)\setminus \mathcal{T}^0\big) \cup \big((\mathcal{R}^1 \times \mathbb{T}^n)\setminus \mathcal{T}^1\big) \subseteq \big((\mathcal{R}^0 \times \mathbb{T}^n)\setminus \mathcal{T}^0\big) \cup \bigcup_{k\in\mathcal{G}^n_{\mathtt{K}_o}} (\mathcal{R}^{1,k} \times \mathbb{T}^n)\setminus \mathcal{T}^{1,k}\,. \tag{240}$$

We now need the following elementary result, whose proof is given in Appendix.

Lemma 5.1 *If $f \in \mathbb{B}^n_s$ satisfies (17), then, for any $k \in \mathcal{G}^n$, the number $2N_k$ of critical points of $\pi_{\mathbb{Z}k} f$ is bounded by $\bar{\mathtt{c}} := \max\{4, \pi\sqrt{8/\beta}\}$.*

[99] $\mathcal{T}_k^i$ is defined in Proposition 5.2, $\mathcal{T}^0$ in Proposition 5.1 and ϕ_k^i in (112).

Obviously, the hypothesis of this lemma are met by our fixed potential in $\mathbb{G}^n_s$, and the following measure estimate holds.

Lemma 5.2 *Let* λ *as above in* (234) *and* $\bar{\mathtt{c}}$ *as in Lemma* 5.1*. Then, for any* k *in* $\mathcal{G}^n_{\mathtt{K}_{\rm o}}$*, one has*

$$\begin{aligned}&\operatorname{meas}\big((\mathcal{R}^{1,k}\times\mathbb{T}^n)\setminus\mathcal{T}^{1,k}\big)\\ &\leqslant \mathtt{c}_{\ast}\operatorname{meas}\big(\widetilde{\mathcal{R}}^{1,k}\times\mathbb{T}^n\big)\lambda|\log\lambda| + \bar{\mathtt{c}}\max_{0\leqslant i\leqslant 2N_k}\operatorname{meas}\Big(\big(B^i_k(\lambda)\times\mathbb{T}^n\big)\setminus\mathcal{T}^i_k\Big)\,.\end{aligned}\tag{241}$$

Proof Since ϕ^i_k in Theorem 3.1 is a diffeomorphism, one has

$$\begin{aligned}(\mathcal{R}^{1,k}\times\mathbb{T}^n)\setminus\mathcal{T}^{1,k} &\overset{(238)}{=} (\mathcal{R}^{1,k}\times\mathbb{T}^n)\setminus\Big(\bigcup_{0\leqslant i\leqslant 2N_k}\phi^i_k(\mathcal{T}^i_k)\Big)\\ \subseteq\ &\Big(\big(\mathcal{R}^{1,k}\times\mathbb{T}^n\big)\setminus\bigcup_{0\leqslant i\leqslant 2N_k}\phi^i_k\big(B^i_k(\lambda)\times\mathbb{T}^n\big)\Big)\cup\bigcup_{0\leqslant i\leqslant 2N_k}\phi^i_k\Big(\big(B^i_k(\lambda)\times\mathbb{T}^n\big)\setminus\mathcal{T}^i_k\Big)\,,\end{aligned}$$

then, passing to measures, using (136), the fact that ϕ^i_k is symplectic and Lemma 5.1, we get (241). ∎

Now, assume that, together with (233) and (236), it is also $\mathtt{K}\geqslant\mathtt{c}_0$. Then, recalling (237), Theorem 4.1 holds. Thus, recalling (223), observing that

$$B^i_k(\lambda) \;=\; \big\{I\in B^i_k \text{ s.t. } |\det\partial^2_I\mathtt{h}^i_k(I)|\leqslant\eta\big\}\cup B^i_k(\lambda,\eta)\,,$$

by (154) and (225) we get

$$\operatorname{meas}\big((B^i_k(\lambda)\times\mathbb{T}^n)\setminus\mathcal{T}^i_k\big)\;\leqslant\;\mathtt{c}_0(|k|^{2n}\eta)^{\mathtt{b}}\operatorname{meas}B^i_k\;+\;\bar{\mathtt{C}}_1 e^{-\mathtt{K}s/7}\,.\tag{242}$$

Now, by (240), (221), (241), (242), (140), (234) and since $|k|_1\;\leqslant\;\mathtt{K}_{\rm o}\;=\;\mathtt{K}/6$ we get, for a suitable constant[100] $\mathtt{c}_1\;=\mathtt{c}_1\,(n,s,\delta,\beta,\xi,\mathtt{m})$,

$$\operatorname{meas}(\mathcal{A})\leqslant\mathtt{c}_1\;\mathtt{K}^{2n}e^{-\mathtt{K}/\mathtt{c}_\star}\,,\qquad \mathtt{c}_\star:=\max\big\{36/s,2\mathtt{C}_1/\mathtt{b}\big\}\,.\tag{243}$$

Finally, let

$$\mathtt{c}=\mathtt{c}(n,s,\delta,\beta,\xi,\mathtt{m})\geqslant 1+\mathtt{c}_\star\tag{244}$$

be such that, if $\mathtt{K}\geqslant\mathtt{c}$, then $\mathtt{c}_1\,\mathtt{K}^{2n}e^{-\mathtt{K}/\mathtt{c}_\star}\leqslant e^{-\mathtt{K}/(1+\mathtt{c}_\star)}$. Then, if $\mathtt{K}\geqslant\mathtt{c}$, claim (iii) follows, and the proof of Theorem 1.1 is complete. ∎

Remark 5.3 Notice that $\mathcal{T}^0$ is a family of maximal primary tori for $\mathtt{H}$, and so are the families $\mathcal{T}^{1,k}_i$ for all $k\in\mathcal{G}^n_{\mathtt{K}_{\rm o}}$ and $i=0,2N_k$. On the other hand, $\mathcal{T}^{1,k}_i$ for all $k\in\mathcal{G}^n_{\mathtt{K}_{\rm o}}$ and $0<i<2N_k$ are families of *maximal secondary tori* for $\mathtt{H}$. In particular these families do not bifurcate from integrable tori.

[100] To get (243), use the following: $\varepsilon\leqslant\mathtt{K}^{-\gamma}<1$ (compare (11)); $\operatorname{meas}\big(\widetilde{\mathcal{R}}^{1,k}\times\mathbb{T}^n\big)\leqslant c(n)$, as $\widetilde{\mathcal{R}}^{1,k}\subset\{y:|y|\leqslant 2\}$; $|\log\lambda|=\frac{\mathtt{b}}{\mathtt{C}_1(1+\mathtt{b})}\mathtt{K}$; $\operatorname{meas}B^i_k\leqslant c_{\ast}$ by (140); $\#\mathcal{G}^n_{\mathtt{K}_{\rm o}}<(2\mathtt{K}_{\rm o}+1)^n$.

Proofs of Corollaries 1.1 and 1.2

Proof of Corollary 1.1 As already pointed out in § 1, Corollary 1.1 follows trivially from Theorem 1.1 and the measure estimate (12), by taking $\mathtt{K} := \mathfrak{c}|\log\varepsilon|$, $\bar{\mathfrak{c}} := 1 + (2\pi)^n c_\star \mathfrak{c}^\gamma$, and $\varepsilon_{\rm o}$ so that $\varepsilon \mathtt{K}^\gamma < 1$ for $\varepsilon < \varepsilon_{\rm o}$. ∎

Proof of Corollary 1.2 Let $n = 2$. We claim that $\mathcal{R}^2$ in (23) satisfies

$$\mathcal{R}^2 \subseteq \{y \in \mathbb{R}^2 : |y| < \varepsilon^{a/2}\}\,. \tag{245}$$

Fix $y \in \mathcal{R}^2$. Then $|y| < 1$ and there exists $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$ such that $|y \cdot k| \leqslant \alpha/4$ (since $y \notin \mathcal{R}^0$). Moreover, since $y \notin \mathcal{R}^{1,k}$, there exists $\ell \in \mathcal{G}^n_{\mathtt{K}} \setminus \mathbb{Z}k$ such that

$$|\,\pi_k^\perp y \cdot \ell| \leqslant \frac{6\alpha\mathtt{K}}{|k|}\,. \tag{246}$$

Then, $|\,\pi_k y| < \alpha/(4|k|) \leqslant \alpha/4$. Moreover, since $\ell \notin \mathbb{Z}k$,

$$|\,\pi_k^\perp \ell| = \frac{|k_1\ell_2 - k_2\ell_1|}{|k|} \geqslant \frac{1}{|k|}\,, \qquad |\,\pi_k^\perp y \cdot \ell| = |\,\pi_k^\perp y|\,|\,\pi_k^\perp \ell| \geqslant \frac{|\,\pi_k^\perp y|}{|k|}\,,$$

which implies, by (246), $|\,\pi_k^\perp y| \leqslant 6\alpha\mathtt{K}$. In conclusion

$$|y| = |\,\pi_k y + \pi_k^\perp y| < 7\alpha\mathtt{K} \overset{(20)}{=} 7\sqrt{\varepsilon}\mathtt{K}^{12}\,. \tag{247}$$

Now, let $\hat{a} := (1-a)/24$ and $\mathtt{K} := 1/(\sqrt[12]{7}\varepsilon^{\hat{a}})$. Then, (245) follows by (247).
Finally, let $\varepsilon_{\rm o} < 1$ be so small that $\varepsilon\mathtt{K}^\gamma < 1$ is satisfied for any $\varepsilon < \varepsilon_{\rm o}$. Then, by the estimate in Theorem 1.1–(iii), we get

$$\mathrm{meas}(\mathcal{A}) \leqslant \mathrm{meas}\Big(\big(\{\varepsilon^{a/2} < |y| < 1\} \times \mathbb{T}^n\big) \setminus \mathcal{T}\Big) \leqslant e^{-\frac{1}{7^{1/12}\mathfrak{c}\varepsilon^{\hat{a}}}} < e^{-\frac{1}{2\mathfrak{c}\varepsilon^{\hat{a}}}}\,. \qquad ∎.$$

A Proofs of elementary lemmata

Proof of Lemma 1.1

Assume $f \in \mathbb{G}^n_s$ for some $s > 0$ and let $0 < \delta_0 \leqslant 1$ be smaller than

$$\varliminf_{\substack{|k|_1 \to +\infty \\ k \in \mathcal{G}^n}} |f_k| e^{|k|_1 s} |k|_1^n \overset{(9)}{>} 0\,.$$

Then, there exists N_0 such that $|f_k| > \delta_0 |k|_1^{-n} e^{-|k|_1 s}$, for any $|k|_1 \geqslant N_0$, $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$. Since $\lim_{\delta\to 0}\mathtt{N} = +\infty$, there exists $0 < \delta < \delta_0$ such that $\mathtt{N} > N_0$. Hence, if $|k|_1 \geqslant \mathtt{N}$ and $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, (16) holds.
Since $\pi_{\mathbb{Z}k} f$ is, for any $|k|_1 \leqslant \mathtt{N}$, a Morse function with distinct critical values one can, obviously, find a $\beta > 0$ for which (17) holds.

To prove the 'if part', we need two lemmata. The first lemma can also be found in [15] (compare Proposition 1.1 there); for completeness, we reproduce the simple but instructive proof also here.

Lemma A.1 *Let $f \in \mathbb{B}^n_s$ such that (16) holds. Then, for any $k \in \mathcal{G}^n$ with $|k|_1 \geqslant \mathtt{N}$, there exists $\theta_k \in [0, 2\pi)$ so that*

$$\pi_{\mathbb{Z}k} f(\theta) = 2|f_k|\big(\cos(\theta + \theta_k) + F^k_\star(\theta)\big)\,, \quad F^k_\star(\theta) := \frac{1}{2|f_k|} \sum_{|j| \geqslant 2} f_{jk} e^{\mathrm{i}j\theta}\,, \tag{248}$$

with $F^k_\star \in \mathbb{B}^1_1$ and $|F^k_\star|_1 \leqslant 2^{-40}$.

Proof We write $\pi_{\mathbb{Z}k} f$ as

$$\pi_{\mathbb{Z}k} f(\theta) := \sum_{j \in \mathbb{Z} \setminus \{0\}} f_{jk} e^{\mathrm{i}j\theta} = \sum_{|j|=1} f_{jk} e^{\mathrm{i}j\theta} + \sum_{|j| \geqslant 2} f_{jk} e^{\mathrm{i}j\theta}\,,$$

and, defining $\theta_k \in [0, 2\pi)$ so that $e^{\mathrm{i}\theta_k} = f_k/|f_k|$, one has

$$\frac{1}{2|f_k|} \sum_{|j|=1} f_{jk} e^{\mathrm{i}j\theta} = \mathrm{Re}\Big(\frac{f_k}{|f_k|} e^{\mathrm{i}\theta}\Big) = \mathrm{Re}\, e^{\mathrm{i}(\theta+\theta_k)} = \cos(\theta + \theta_k)\,,$$

which is equivalent to (248). Now, since $\|f\|_s \leqslant 1$ (so that $|f_k| \leqslant e^{-|k|_1 s}$), one finds, for $|k|_1 \geqslant \mathtt{N}$,

$$\begin{aligned}
|F^k_\star|_1 &\overset{(248)}{\leqslant} \frac{1}{2|f_k|} \sum_{|j| \geqslant 2} |f_{jk}| e^{|j|} \overset{(16)}{\leqslant} \frac{|k|_1^n e^{|k|_1 s}}{2\delta} \sum_{|j| \geqslant 2} |f_{jk}| e^{|j|} \leqslant \frac{|k|_1^n e^{|k|_1 s}}{2\delta} \sum_{|j| \geqslant 2} e^{-|j|(|k|_1 s - 1)} \\
&\leqslant \frac{2e^2 |k|_1^n}{\delta}\, e^{-|k|_1 s} = \frac{2^{n+1} e^2}{s^n \delta} e^{-\frac{|k|_1 s}{2}} \ \Big(\frac{|k|_1 s}{2}\Big)^n e^{-\frac{|k|_1 s}{2}} \leqslant \Big(\frac{2n}{es}\Big)^n \frac{2e^2}{\delta}\, e^{-\frac{|k|_1 s}{2}} \leqslant 2^{-40}\,,
\end{aligned}$$

where last inequality follows since $|k|_1 \geqslant \mathtt{N}$ (see (15)). ∎

Now, assume that (16) and (17) hold for some $\delta \in (0, 1]$ and $\beta > 0$. Then, from (16) follows immediately (9). It remains to prove that (10) holds for any $k \in \mathcal{G}^n$ with $|k|_1 > \mathtt{N}$. In view of Lemma A.1, the thesis follows from the following elementary

Lemma A.2 *Let $F \in C^2(\mathbb{T}, \mathbb{R})$, $\bar\theta$ and $0 < \mathtt{c} < \frac{1}{2}$ are such that*[101] *$\|F - \cos(\theta + \bar\theta)\|_{C^2} \leqslant \mathtt{c}$. Then, F has only two critical points and it is $(1 - 2\mathtt{c})$–Morse.*

Proof By considering the translated function $\theta \to F(\theta - \bar\theta)$, one can reduce oneself to the case $\theta = 0$ (note that F is β–Morse, if and only if $\theta \to F(\theta - \bar\theta)$ is β–Morse).
Thus, set $\bar\theta = 0$, and note that, by assumption $|F'| = |F' + \sin\theta - \sin\theta| \geqslant |\sin\theta| - \mathtt{c}$, and, analogously, $|F''| \geqslant |\cos\theta| - \mathtt{c}$. Hence, $|F'| + |F''| \geqslant |\sin\theta| + |\cos\theta| - 2\mathtt{c} \geqslant 1 - 2\mathtt{c}$. Next, let us show that F has a unique strict maximum $\theta_0 \in I := (-\pi/6, \pi/6)$ (mod 2π). Writing $F = \cos\theta + g$, with $g := F - \cos\theta$, one has that $F'(-\pi/6) = 1/2 + g'(\pi/6) \geqslant 1/2 - \mathtt{c} > 0$, and, similarly $F'(\pi/6) \leqslant -1/2 + \mathtt{c}$, thus F has a critical point in I, and, since $-F'' = \cos\theta - g'' \geqslant \cos\theta - \mathtt{c} \geqslant \sqrt{3}/2 - \mathtt{c} > 0$, F is strictly concave in I, showing that such critical point is unique and it is a strict local minimum. In fact, similarly one shows that F has a second critical point $\theta_1 \in (\pi - \pi/6, \pi + \pi/6)$ where F is strictly convex, so that θ_1 is a strict local minimum; but, since in the complementary of these intervals F is strictly monotone (as it is easy to check), it follows that F has a unique global strict maximum and a unique global strict minimum. Finally, $F(\theta_0) - F(\theta_1) \geqslant \sqrt{3} - 2\mathtt{c} > 1 - 2\mathtt{c}$ and the claim follows. ∎

[101] $\|F\|_{C^2} := \max_{0 \leqslant k \leqslant 2} \sup |F^{(k)}|$. Note that, by Cauchy estimates, $\|F\|_{C^2} \leqslant 2|F|_1$.

Proof of Lemma 2.3

First note that by (42), (43) and (44)

$$(1-\mu)p_1^2 - (1+\mu)2^{-16}\mathtt{r}^2 \leqslant \mathtt{H}_\flat(p,q_1) \leqslant (1+\mu)(p_1^2 + 2^{-16}\mathtt{r}^2)\,. \tag{249}$$

By the first inequality in (249) we have that if $(p,q_1)\in\mathcal{M}(\hat{p})$ then $p_1^2 \leqslant \frac{\mathtt{E}_\flat + 2^{-16}\mathtt{r}^2(1+\mu)}{1-\mu}$, which is indeed smaller than $(\mathtt{R}+\mathtt{r}/2)^2$ by (66) and (44). This proves the second inclusion in (67).
By the second inequality in (249) we have that if $|p_1|\leqslant \mathtt{R}+\mathtt{r}/3$ then $\mathtt{H}_\flat(p,q_1)\leqslant (1+\mu)\big((\mathtt{R}+\mathtt{r}/3)^2+2^{-16}\mathtt{r}^2\big)$, which is smaller than $\mathtt{E}_\flat$ again by (66) and (44). This proves the first inclusion in (67). ∎

Proof of Lemma 3.2

Let $(J,\psi)\in D_{\zeta r}\times\mathbb{T}^n_{\zeta s}$. Then there exists $J_0\in D$ such that $|J-J_0|<\zeta r$. Set

$$w(z) := \big(J_0 + \tfrac{\zeta}{z}(J-J_0),\ \mathrm{Re}\,\psi + \tfrac{\zeta}{z}\,\mathrm{Im}\,\psi\big)\,,$$

with $\mathrm{Re}\,\psi := (\,\mathrm{Re}\,\psi_1,\ldots,\,\mathrm{Re}\,\psi_n)$, and analogously for $\mathrm{Im}\,\psi$. Note that $w(\zeta)=(J,\psi)$, and that $w(z)\in D_r\times\mathbb{T}^n_s$ for every $|z|<1$. Consider the holomorphic function $G(z):=g(w(z))$ defined for $|z|<1$. Then, $|\,\mathrm{Im}\,G|\leqslant\xi$ for $|z|<1$. Let u and v be real harmonic functions such that $G(z)=u(x,y)+\mathrm{i}v(x,y)$, where $z=x+\mathrm{i}y$. Since by hypothesis $\sup_{|z|<1}|v|\leqslant\xi$, by interior estimate of derivatives of for harmonic functions[102] we have that $\sup_{|z|\leqslant 1/2}|v_x|\leqslant 4\xi$ and analogously for v_y. By Cauchy–Riemann equations, the same estimate holds for u_x. Therefore $\sup_{|z|\leqslant 1/2}|G'| = \sup_{|z|\leqslant 1/2}|u_x+\mathrm{i}v_x|\leqslant 8\xi$. Since $w(0)=(J_0,\,\mathrm{Re}\,\psi)\in D\times T^n$ and g is real analytic, we have that $G(0)=g(J_0,\,\mathrm{Re}\,\psi)\in\mathbb{R}$. Then, for any $0<\zeta\leqslant 1/2$, by the the mean value theorem, we have that

$$|\,\mathrm{Im}\,g(J,\psi)| = |\,\mathrm{Im}\,G(\zeta)| = |\,\mathrm{Im}\,G(\zeta) - \,\mathrm{Im}\,G(0)| \leqslant |G(\zeta)-G(0)| \leqslant 8\zeta\xi\,. \quad ∎$$

Proof of Lemma 4.6

Let us consider first the case $P=\mathrm{I_d}$. Consider the unitary matrix U diagonalizing Q, namely $U^{-1}QU=\Lambda=\mathrm{diag}_{1\leqslant j\leqslant d}\lambda_j$. Note that $|Q|=|\Lambda|=\max_{1\leqslant j\leqslant d}|\lambda_j|=\lambda$. Then $U^{-1}(\mathrm{I_d}+Q)U = I+\Lambda$ and $\det(\mathrm{I_d}+Q) = \det(I+\Lambda)\geqslant(1-\lambda)^d$, proving the case $P=\mathrm{I_d}$.
Consider now the general case. Write $P+Q = P^{1/2}(\mathrm{I_d}+P^{-1/2}QP^{-1/2})P^{1/2}$. Note that, since $P^{-1/2}$ is symmetric, then $P^{-1/2}QP^{-1/2}$ is symmetric too. Since $|P^{-1/2}QP^{-1/2}|\leqslant|P^{-1/2}|^2|Q| = |P^{-1}||Q|$ and $\det P^{1/2} = (\det P)^{1/2}$, from the previous case the general case follows.
The final claim in Lemma 4.6 follows, as $(1-\lambda)^d\geqslant 1-d\lambda$. ∎

Proof of Lemma 4.15

First observe that the cases $i=0$ and $i=2N$ are identical since

$$\bar{I}_1^0(E) = \bar{I}_1^{2N}(E)\,, \qquad \bar{\mathtt{E}}^0(I_1) = \bar{\mathtt{E}}^{2N}(I_1)\,.$$

Let us then consider the case $i=2N$. By definition,

$$\bar{I}_1^{2N}(E) = \frac{1}{2\pi}\int_0^{2\pi}\sqrt{E-\bar{\mathtt{G}}(x)}dx\,, \tag{250}$$

[102] See Theorem 2.10 in [24].

so that, by Jensen's inequality,

$$(2\partial_E \bar{I}_1^i(E))^3 = \Big(\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{\sqrt{E-\bar{\mathtt{G}}(x)}}dx\Big)^3 \leqslant \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{(E-\bar{\mathtt{G}}(x))^{3/2}}dx = -4\partial_E^2\bar{I}_1^i(E)\,,$$

and the claim follows by (143). $\blacksquare$

Proof of Lemma 5.1

Consider first the case $|k|_1 \geqslant \mathtt{N}$. By Lemma A.1, $F_\star^k := \pi_{\mathbb{Z}k} f/2|f_k|$ satisfies

$$|F_\star^k - \cos(\theta+\theta_k)|_1 \leqslant 2^{-40}\,.$$

Thus, by Cauchy estimates we get $\|F_\star^k - \cos(\theta+\theta_k)\|_{C^2} \leqslant 2^{-39}$, so that by Lemma A.2 it follows that $2N_k = 4$. For the case $|k|_1 \leqslant \mathtt{N}$ we need the following elementary observation:

Lemma A.3 *If G is β–Morse, then the number $2N$ of its critical points is bounded by $\pi\sqrt{2\max_{\mathbb{R}}|G''|/\beta}$.*

Proof If θ_i and θ_j are different critical points of G, then, by Taylor expansion at order two and by (14) one has $\beta \leqslant |G(\theta_i)-G(\theta_j)| \leqslant \frac{1}{2}(\max_{\mathbb{R}}|G''|)|\theta_i-\theta_j|^2$, which implies that the minimal distance between two critical points is at least $\sqrt{2\beta/\max_{\mathbb{R}}|G''|}$, from which the claim follows. $\blacksquare$

Now, by (17) we know that $\pi_{\mathbb{Z}k} f$ is β–Morse, and since $\|f\|_s \leqslant 1$ we have $\sup_{\mathbb{R}}|(\pi_{\mathbb{Z}k} f)''| \leqslant \sum_{j\neq 0}|f_{jk}|j^2 \leqslant \sum_{j\neq 0} e^{-|j|}j^2 < 4$. Then, by Lemma A.3, the claim follows also in this case. $\blacksquare$

References

[1] V. I. Arnold, *Proof of A. N. Kolmogorov's theorem on the conservation of conditionally periodic motions with a small variation in the Hamiltonian*, Russian Math. Surv, 1963, vol. 18, no 5, pp. 9–36.

[2] V. I. Arnold, *Small denominators and problems of stability of motion in classical and celestial mechanics.* Usp. Mat. Nauk 18, No.6, 91–192 (1963). Engl. transl.: Russ. Math. Surv. 18, No.6, 85–191 (1963)

[3] V. I. Arnol'd, *Instability of dynamical systems with many degrees of freedom*, Dokl. Akad. Nauk SSSR, 156 (1964), pp. 9–12

[4] V. I. Arnol'd, V. V. Kozlov, and A. I. Neishtadt. Dynamical systems. III. Translated from the 1985 Russian Edition by A. Iacob. Encyclopaedia of Mathematical Sciences, 3. Springer–Verlag, Berlin, 1988. xiv+291 pp.

[5] V. I. Arnol'd, V. V. Kozlov, and A. I. Neishtadt. *Mathematical aspects of classical and celestial mechanics*, volume 3 of Encyclopaedia of Mathematical Sciences. Springer-Verlag, Berlin, third edition, 2006. [Dynamical systems. III], Translated from the 2002 Russian original by E. Khukhro

[6] D. Bambusi, A. Fusè, M. Sansottera. *Exponential stability in the perturbed central motion.* Regular and Chaotic Dynamics, 23, 821–841 (2018)

[7] P. Bernard, V. Kaloshin, and K. Zhang, *Arnol'd diffusion in arbitrary degrees of freedom and 3–dimensional normally hyperbolic invariant cylinders*, Acta Math., 217:1 (2016), pp. 1–79

[8] M. Berti, L. Biasco, and P. Bolle, *Drift in phase space: a new variational mechanism with optimal diffusion time.* J. Math. Pures Appl. (9) 82 (2003), no. 6, 613–664

[9] L. Biasco, and L. Chierchia. *On the measure of Lagrangian invariant tori in nearly–integrable mechanical systems.* Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. **26** (2015), 1–10

[10] L. Biasco, and L. Chierchia. *Explicit estimates on the measure of primary KAM tori.* Ann. Mat. Pura Appl. (4), 2017, 261–281

[11] L. Biasco, and L. Chierchia. *On the topology of nearly-integrable Hamiltonians at simple resonances.* Nonlinearity 33 (2020) 3526–3567

[12] L. Biasco, and L. Chierchia. *On the measure of KAM tori in two degrees of freedom*. Discrete and Continuous Dynamical Systems, Volume 40, Number 12 (2020) 6635-6648

[13] L. Biasco, and L. Chierchia. *Quasi–periodic motions in generic nearly–integrable mechanical systems*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. 33 (2022), no. 3, pp. 575–580, *arXiv:2206.01055*

[14] L. Biasco, and L. Chierchia. *Complex Arnol'd–Liouville maps*, Regular and Chaotic Dynamics, 2023. *doi 10.1134/S1560354723520064*. *arXiv:2306.00875v1*

[15] L. Biasco, and L. Chierchia. *Global properties of generic real–analytic nearly–integrable Hamiltonian systems*, *arXiv:2306.13527*

[16] A. Celletti, and L. Chierchia *A Constructive Theory of Lagrangian Tori and Computer-Assisted Applications*, in Dynamics Reported, Dynam. Report. Expositions Dynam. Systems (N. S.), vol. 4, Berlin: Springer, 1995, pp. 60–129

[17] Q. Chen; R. de la Llave, *Analytic genericity of diffusing orbits in a priori unstable Hamiltonian systems.* Nonlinearity 35 (2022), no. 4, 1986–2019

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

[19] L. Chierchia, and C.E. Koudjinan. *V.I. Arnol'd's "Global" KAM Theorem and geometric measure estimates.* Regular and Chaotic Dynamics, 2021, Vol. 26, no. 1, 61–88

[20] A. Delshams and G. Huguet. *Geography of resonances and Arnol'd diffusion in a priori unstable Hamiltonian systems.* Nonlinearity, 22(8):1997–2077, 2009

[21] A. Delshams, R. de la Llave, and T.M. Seara, *Instability of high dimensional Hamiltonian systems: multiple resonances do not impede diffusion.* Adv. Math. 294 (2016), 689–755

[22] H. S. Dumas, *The KAM Story*, World Scientific, 2014

[23] H. Eliasson, *Perturbation of linear quasi-periodic systems.* Dynamical Systems and Small Divisors (Cetraro, Italy, 1998), 1-60, Lect. Notes Math. 1784, Springer, 2002

[24] D. Gilbarg, and N. S. Trudinger *Elliptic partial differential equations of second order*. Springer–Verlag, Berlin–New York, 1977. Grundlehren der Mathematischen Wissenschaften, Vol. 224

[25] M. Gidea, R. de la Llave, and T. M. Seara, *A general mechanism of diffusion in Hamiltonian systems: qualitative results* Comm. Pure Appl. Math. 73 (2020), no. 1, 150–209

[26] V. Kaloshin and K. Zhang. *Arnol'd diffusion for smooth systems of two and a half degrees of freedom*, volume 208 of Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, 2020

[27] A. N. Kolmogorov, *On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian*, Dokl. Akad. Nauk. SSR 98 (1954), 527-530. English translation by Helen Dahlby in: Stochastic behavior in classical and quantum Hamiltonian systems, Volta Memorial conference, Como, 1977, Lecture Notes in Physics, 93, Springer, 1979.

[28] V.F. Lazutkin, *Concerning a theorem of Moser on invariant curves. Probl. Dyn. Theory Propag. Seism. Waves 14, 109–120. Leningrad: Nauka. (1974) (Russian)*

[29] L. Lazzarini, J-P. Marco, and D. Sauzin, *Measure and capacity of wandering domains in Gevrey near-integrable exact symplectic systems*, Memoirs of the AMS, **257**, (2019), 1-110

[30] J. N. Mather, *Arnol'd diffusion by variational methods.* Essays in mathematics and its applications, 271–285, Springer, Heidelberg, 2012

[31] A.G. Medvedev, A.I. Neishtadt, D.V. Treschev, *Lagrangian tori near resonances of near–integrable Hamiltonian systems*, Nonlinearity, **28**:7 (2015), 2105–2130

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

[33] J. Moser: *Lectures on Hamiltonian Systems*. Mem. Am. Math. Soc. 81. American Mathematical Society, Providence, R I. (1968), 60 pp.

[34] A. I. Neishtadt, *Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions*. J. Appl. Math. Mech. 45 (1981), no. 6, 766–772

[35] J. Pöschel, *Integrability of Hamiltonian systems on Cantor sets*. Comm. Pure Appl. Math. 35 (1982), no. 5, 653–696

[36] J. Pöschel, *Nekhoroshev estimates for quasi–convex Hamiltonian systems*. Math. Z. **213**, pag. 187 (1993)

[37] A. S. Pyartli,*Diophantine approximations on submanifolds of Euclidean space*. Functional Analysis Appl. 3, 303–306 (1970) (Translation from Russian, Funkts. Analys. Prilozh. 3, 59–62, 1969)

[38] A.A. Svanidze, *Small perturbations of an integrable dynamic system with integral invariant*. Proc. Steklov Inst. Math. 2, 127–151 (1981)

[39] D. Treschev. *Arnol'd diffusion far from strong resonances in multidimensional a priori unstable Hamiltonian systems*. Nonlinearity, 25(9):2717–2757, 2012

[40] Zhang, Ke *Speed of Arnol'd diffusion for analytic Hamiltonian systems* Invent. Math. 186 (2011), no. 2, 255–290.