Measure of primary and secondary KAM tori in mechanical systems

L. Biasco & L. Chierchia
Dipartimento di Matematica e Fisica
Università degli Studi Roma Tre
Largo San Leonardo Murialdo, 1 - 00146 Roma, Italy
luca.biasco@uniroma3.it, luigi.chierchia@uniroma3.it

Draft version – May 11, 2026

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 addressed. In 1985, Arnol'd, Kozlov, and Neishtadt conjectured that, for systems with $n = 2$ degrees of freedom, the measure of the non-torus set of general analytic nearly integrable systems away from critical points is exponentially small in the size ε of the perturbation. In 2002, they further conjectured that, for $n \geqslant 3$, such measure is typically of order ε, rather than $\sqrt{\varepsilon}$, as predicted by classical KAM theory. In the case of generic analytic mechanical Hamiltonian systems, we prove lower bounds on the measure of primary and secondary invariant tori that agree, up to a logarithmic correction, with the above conjectures.

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

This work has been supported by grant PNRR-M4C2-I1.1-PRIN2022-PE1-Stability in Hamiltonian dynamics and beyond-F53D23002730006-Finanziato dall'U.E. - NextGenerationEU

Contents

1 Introduction **2**

2 Results, Remarks, and Open Problems **10**

3 Prerequisites **17**
3.1 Coverings and averaging 18
3.2 Standard form at simple resonances 23
3.3 Action-angle variables for 1D standard Hamiltonians 29

4 Secondary nearly integrable structure at simple resonances **38**

5 Twist at simple resonances **49**
5.1 Twist Theorem near simple resonances (statement) 50
5.2 Proof of the Twist Theorem 55

6 Maximal KAM tori and proof of the main results **75**
6.1 KAM tori in the nonresonant region 76
6.2 KAM tori around simple resonances 78
6.3 Proof of the main theorem and its corollaries 80

A Proofs of elementary lemmata **83**

References **88**

1 Introduction

A conjecture by Arnol'd, Kozlov and Neishtadt on the measure of KAM tori

Classical KAM Theory deals with the persistence of Lagrangian invariant tori of integrable Hamiltonian systems under the effect of small perturbations ([30, 1, 36, 2, 35]; for a divulgative account, see [24]).

In the early 1980's ([31, 37, 39, 42]), it was clarified that an analytic integrable system, which is Kolmogorov nondegenerate (i.e., such that the action-to-frequency map is a local diffeomorphism), preserves, under a perturbation of size $\varepsilon > 0$, all its (unperturbed) invariant tori *up to a set of measure proportional to* $\sqrt{\varepsilon}$. In fact, this estimate cannot be improved, as far as *primary tori* (i.e., tori which are a deformation of integrable ones) are concerned.[1] However, positive measure sets of secondary Lagrangian tori (i.e., tori, which are not a deformation of integrable ones) appear in general nearly integrable systems, for example, near elliptic equilibria ([34]).

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

In 1985 Arnol'd, Kozlov and Neishtadt ([4, p. 189]), motivated by the exponentially small splitting of separatrices in general systems with two degrees of freedom, conjectured that:

"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 2002 ([5]), the same authors, arguing on the basis of a simple rescaling argument in neighbourhoods of double resonances,[2] 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."

Main result

In this paper, we develop tools to detect KAM tori in *generic analytic nearly integrable mechanical systems*. In particular, we discuss the construction of maximal KAM tori that live 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 prove lower bounds on the total measure of KAM tori, which are in agreement with the above conjectures up to a logarithmic correction in* $1/\varepsilon$.[3]

The reason for dealing with the special class of nearly integrable mechanical systems, namely, Hamiltonian systems on $\mathbb{R}^n \times \mathbb{T}^n$ (endowed with the standard symplectic

[1] Trivially, the phase space of a pendulum with Hamiltonian $H = \frac{p^2}{2} + \varepsilon \cos q$, is filled by primary tori $\{H = E\}$, $E > \varepsilon$, up to the region enclosed by the separatrix, whose area is proportional to $\sqrt{\varepsilon}$.

[2] 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 $\sim \varepsilon$."

[3] Of course, for $n = 2$ there is no logarithmic correction; for the purpose of this introduction, we will mainly focus on the case $n \geqslant 3$. The results of this paper were announced in the short notes [11, 15].

form $dy \wedge dx$) with Hamiltonian given by

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

(where, $|y|^2 = y \cdot y = \sum_j |y_j|^2$), is twofold. On one side, mechanical systems are the simplest nontrivial models of nearly integrable Hamiltonian systems; 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. In fact, we shall introduce a suitable class of *generic analytic* potentials f's (Definition 2.1), which guarantees a uniform control on the *secondary nearly integrable structure* at simple resonances with high modes (for generalizations of the model and more comments, see Open Problems, (i) in § 2).

General plan

As well known, resonances are the principal hindrance for stability in Hamiltonian systems. Here, the plot will take place *away from double resonances*, and, mainly, *near simple resonances*. The order of the resonances considered, denoted by $\mathtt{K}$, will play a fundamental role in our analysis, and will be taken as a suitable function of ε, larger than $|\log \varepsilon|$.

In the following informal description of the general strategy, "c" will denote different ε-independent constants.

Invariant maximal tori will be detected in a domain of phase space, where unperturbed frequencies are *away* from suitable tubular neighborhoods of double resonances (up to order $\mathtt{K}$) of measure $\sim \varepsilon \mathtt{K}^c$. Indeed, in such neighborhoods, the $\mathtt{H}$-dynamics is "essentially" equivalent to the dynamics of a two–degree-of-freedom system independent of ε (compare the argument given by Arnol'd, Kozlov and Neishtadt, reproduced in footnote 2 above). Our analysis, being perturbative, will therefore be focussed on the complementary of such neighborhoods of double resonances.

We will start by introducing a suitable cover of the action space consisting of three sets: a $\mathtt{K}$-nonresonant[4] set $\mathcal{R}^0$ of relative measure $\sim (1 - \sqrt{\varepsilon}\mathtt{K}^c)$; a $\sqrt{\varepsilon}\,\mathtt{K}^c$-neighbourhood $\mathcal{R}^1$ of simple resonances (up to order $\mathtt{K}$), and a $\sqrt{\varepsilon}\mathtt{K}^c$ neighborhood $\mathcal{R}^2$ of double (and higher) resonances of order $\leqslant \mathtt{K}$, corresponding to the above mentioned exceptional set of measure $\sim (\sqrt{\varepsilon}\mathtt{K}^c)^2 \sim \varepsilon \mathtt{K}^c$; for precise statements, see beginning of § 3.1. We remark that such covering depends on ε *and* on the parameter $\mathtt{K}$. The exceptional set $\mathcal{R}^2 \times \mathbb{T}^n$ is put aside and no further analysis of it will be carried out. Since we want to be in agreement with the conjecture of Arnol'd, Kozlov and Neishtadt up to a logarithmic correction (when $n \geqslant 3$), $\mathtt{K}$ can be taken *at most* as $|\log \varepsilon|^c$, at difference with standard

[4] "$\mathtt{K}$-nonresonant" means that only resonances up to order $\mathtt{K}$ are taken into account.

analytic averaging theory, where, typically, $\mathtt{K}$ is taken as an inverse power of ε. On the other hand, we shall soon see that $\mathtt{K}$ must also be taken $\gtrsim |\log \varepsilon|$.

Next step is to perform averaging up to order $\mathtt{K}$ on the nonresonant and simply resonant regions. In quantitative averaging theory the main parameters are the minimum size of small divisors α and the radii of analyticity in action space r: in our case, *these quantities have to be taken both of order* $\sim \sqrt{\varepsilon}\mathtt{K}^c$; indeed, taking a radius of analyticity larger would have the effect of invading the double resonance neighborhood making meaningless the above covering.

On the non resonant region $\mathcal{R}^0 \times \mathbb{T}^n$, after high order averaging, classical KAM theory yields the existence of primary maximal KAM tori up to a set of measure $O(e^{-c\mathtt{K}})$. Notice that, here, no ε appears explicitly: Indeed, the size of the new (i.e., after averaging) perturbing function is, roughly speaking, $\varepsilon e^{-c\mathtt{K}}$, and the KAM smallness condition becomes:

$$\epsilon_* := \varepsilon e^{-c\mathtt{K}}/r^2 \ \sim e^{-c\mathtt{K}} \ll 1;$$

compare condition (220) in the KAM Theorem 6.1 below.

In the non resonant region $\mathcal{R}^0 \times \mathbb{T}^n$, the measure of the non-torus set can be bounded by $\sqrt{\epsilon_*} = e^{-c\mathtt{K}}$, which becomes smaller than ε, if $\mathtt{K} \geqslant |\log \varepsilon|/c$, as anticipated above.

Much more difficult is the analysis in the neighborhood $\mathcal{R}^1 \times \mathbb{T}^n$ of simple resonances, which is of measure $\sqrt{\varepsilon}\mathtt{K}^c$. *We claim that also this region is filled by primary and secondary maximal tori up to a set of measure* $e^{-c\mathtt{K}}$. A sketch of proof is the following.

$\mathcal{R}^1$ is defined as union of sets $\mathcal{R}^{1,k}$, for $k \in \mathbb{Z}^n$ with $0 < |k| \leqslant \mathtt{K}$. Such sets are $\sqrt{\varepsilon}\,\mathtt{K}^c$-neighborhoods of simple resonances $\{y\colon\ y \cdot k = 0\}$, which are $\sim \sqrt{\varepsilon}\,\mathtt{K}^c$ away from double resonances.[5] On $\mathcal{R}^{1,k} \times \mathbb{T}^n$ high-order partial 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

$$\begin{aligned} &\mathcal{H}_k(\mathtt{y},\mathtt{x}) = \overline{\mathtt{H}}_k(\mathtt{y},\mathtt{x}_1) + \varepsilon \bar{f}^k(\mathtt{y},\mathtt{x}), \quad \bar{f}^k \sim e^{-c\mathtt{K}},\\ &\overline{\mathtt{H}}_k(\mathtt{y},\mathtt{x}_1) := Q^k(\mathtt{y}) + \varepsilon\big(\mathtt{g}^k_{\mathrm{o}}(\mathtt{y}) + \mathtt{g}^k(\mathtt{y},\mathtt{x}_1)\big), \end{aligned} \tag{2}$$

where Q^k is a nondegenerate quadratic form and $\mathtt{x}_1 = k \cdot x \in \mathbb{T}^1$ is the resonant angle (which cannot be averaged out); $\mathtt{g}^k$ has zero average, and, as already mentioned, the action analyticity radius of $\mathcal{H}_k$ is $\sim O(\sqrt{\varepsilon}\mathtt{K}^c)$; compare Theorem 3.3 below, part (ii), and, in particular, (37) and (38).

Now, all these Hamiltonian systems labelled by the simple-resonance index k, have a "secular near integrability structure". Indeed, disregarding the small terms $\varepsilon \bar{f}^k$,

[5] Actually, the distance from exact simple resonances is $\alpha/|k| = \sqrt{\varepsilon}\mathtt{K}^c/|k|$; compare (22).

the secular Hamiltonians $\overline{\mathtt{H}}_k$ in (2) are Arnol'd–Liouville integrable, since they depend effectively only on the (resonant) angle $\mathtt{x}_1$.

Then, the plan is obvious: Put all these systems into their Arnol'd–Liouville action-angle variables, check twist (i.e., Kolmogorov's nondegeneracy), and apply a KAM theorem, so as to obtain Lagrangian primary and secondary tori (with different topologies; compare Remark 6.3 below).

However, implementing such a program poses several challenges. We now proceed to discuss these difficulties and describe the techniques employed to resolve them.

Problems and methodology near simple resonances

In the proof of the main result, the resonance cut-off $\mathtt{K}$ will be taken as a function of ε going to $+\infty$ as $\varepsilon \to 0$ (at least as a logarithm of $1/\varepsilon$). Therefore one has to deal, *de facto*, with infinitely many Hamiltonian systems, and without a uniform way of treating them, there is no hope for a global quantitative analysis.

The idea, here, was suggested in [14] and refined in [17]. The *secular Hamiltonians* $\overline{\mathtt{H}}_k(\mathtt{y}, \mathtt{x}_1)$ in (2) are one-degree-of-freedom Hamiltonians with potentials $\varepsilon \mathtt{g}^k(\mathtt{y}, \mathtt{x}_1)$ and external parameters $\mathtt{y}_2,\dots,\mathtt{y}_n$. Recall, however, that we are in a $\sqrt{\varepsilon}\mathtt{K}^c$ neighborhood of simple resonances (and that the radius of analyticity is $\sim \sqrt{\varepsilon}\mathtt{K}^c$), so that the kinetic part is also of order $\sim \varepsilon \mathtt{K}^c$. As a consequence of averaging theory, the rescaled potential $\mathtt{g}^k(\mathtt{y}, \mathtt{x}_1)$ turns out to be only μ-close, with $\mu = 1/\mathtt{K}^c$, to the partially averaged potential, i.e., to

$$(\pi_{\mathbb{Z}k} f)(\mathtt{x}_1) := \textstyle\sum_{j\in\mathbb{Z}} f_{jk} e^{\mathrm{i}j\theta}.$$

Indeed, since the radius of analyticity r and the small divisor size α are both of order $\sqrt{\varepsilon}\mathtt{K}^c$, one has that (in suitable analytic norm)

$$\|\mathtt{g}^k - \pi_{\mathbb{Z}k} f\| \leqslant c\,\varepsilon/(\alpha r) \leqslant \mu = 1/\mathtt{K}^c;$$

compare (39), (36) and (49) below.

Because of the chosen generic class of potentials, for "high" Fourier modes $|k| > \mathtt{N}$ ($\mathtt{N}$ is independent of ε and depends only on the potential f),[6] one can show that $(\pi_{\mathbb{Z}k} f)(\mathtt{x}_1)$ "behaves as a shifted cosine" $2|f_k|\varepsilon\cos(\mathtt{x}_1 + \theta_k)$, for a suitable $\theta_k \in \mathbb{R}$; compare item (iii) in Theorem 3.3 below. This fact allows for a *uniform* treatment of the high order resonance case when $\mathtt{N} < |k| \leqslant \mathtt{K}$.

Analytic properties of the action-angle variables for the pendulum are quite well known, but, 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. Therefore, one needs a general holomorphic quantitative

[6] As we shall see, $\mathtt{N}$, which is defined in (15) below, depends on the dimension n, the angle analyticity width s of the potential f and a fixed positive number δ strictly less than the liminf in (9); compare also Lemma 2.5.

theory of action-angle variables for one-degree-of-freedom systems containing parameters. Such a theory is discussed in [16] for a special class of real analytic Hamiltonians – called *standard form Hamiltonians* – given by

$$\mathtt{H}_\flat(\mathtt{p},\mathtt{q}_1) = \big(1+\nu(\mathtt{p},\mathtt{q}_1)\big)\mathtt{p}_1^2 + \mathtt{G}(\hat{\mathtt{p}},\mathtt{q}_1), \tag{3}$$

where $\mathtt{p} = (\mathtt{p}_1, \hat{\mathtt{p}}) \in \mathbb{R}^n$, $\mathtt{p}_1$ is the action conjugated to the angle $\mathtt{q}_1$ and $\hat{\mathtt{p}} = (\mathtt{p}_2, ..., \mathtt{p}_n)$ are dumb parameters; ν is small, and $\mathtt{G}$ is close to a fixed reference Morse potential $\bar{\mathtt{G}}(\mathtt{q}_1)$; compare Definition 3.1 below.

An important analytic property of Hamiltonians in standard form is that the energy-to-action map, $E \mapsto I_1(E)$, has a universal behaviour, in a neighbourhood of critical energies. Indeed, one has:

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

where ϵ is the reference energy given by the norm of $\bar{\mathtt{G}}$, and $E_{\rm crit}$ is the critical energy of the secular separatrix; a and b are analytic functions of z in a neighbourhood of $z = 0$ (and, of course, everything depends on the other $(n-1)$ dumb actions; compare Theorem 3.7 below); for related results, see, [38].

The representation (4) will play a crucial role in studying the twist of the secular Hamiltonians at simple resonances in their Arnol'd–Liouville action-angle variables.

Now, one can then prove that: *For all* $|k| \leqslant \mathtt{K}$*, all secular Hamiltonians* $\overline{\mathtt{H}}_k$ *in* (2) *can be put into a standard form* $\mathtt{H}_k(\mathtt{p},\mathtt{q}_1)$ *as in* (3), where:

$$|\mathtt{p}_1| \leqslant \sqrt{\varepsilon}\mathtt{K}^c, \quad \nu \sim 1/\mathtt{K}^c, \quad \bar{\mathtt{G}} = \varepsilon\, \pi_{\mathbb{Z}k} f, \quad (\mathtt{G}-\bar{\mathtt{G}}) \sim \varepsilon/\mathtt{K}^c, \quad \mathtt{G} \sim \varepsilon,$$

and action analyticity radius $\sim \sqrt{\varepsilon}\mathtt{K}^c$; compare Theorem 3.4 below and the definitions in (49). Notice however, that here ε *is not a perturbative parameter*, since, by the same rescaling argument reported in footnote 2 above, ε could be rescaled out and would not appear explicitly any more. Thus, the actual perturbation parameter is $1/\mathtt{K}$.

The standard form $\mathtt{H}_k(\mathtt{p},\mathtt{q}_1)$ can be obtained by means of three k-dependent symplectic transformations Φ_j, which keep fixed the "slow" (or resonant) angle $\mathtt{x}_1$ and the dumb actions $\hat{p}$, where:

(i) Φ_1 is a transformation that decouples the quadratic term Q^k in (2) as a sum of a quadratic part depending only on the dumb actions and a quadratic part in the new action conjugated to the slow angle; compare (60) and (61) below. We remark, however, that the transformation Φ_1 *is not well defined on the torus* $\mathbb{T}^n$ (compare point (ii) of Remark 3.6);

(ii) Φ_2 is a close to identity (globally well defined) transformation, which is introduced so that the (new) secular potential becomes independent of p_1.

(iii) Finally, Φ_3 is a close to identity map, which sets all critical points on the line $p_1 = 0$. As Φ_1, also Φ_3 is not globally well defined on the torus.

We finally remark that the main rescaling properties of the Hamiltonian in standard form $\mathtt{H}_k$ are ruled by a *single uniform* parameter $\mathfrak{g}$ (defined in (56) below), which will depend only on the dimension n, the angle analyticity initial width s, and a uniform parameter β measuring the Morse properties of reference potentials $\pi_{\mathbb{Z}k} f$, for $|k| \leqslant \mathtt{N}$ (compare (17)).

We proceed to sketch the proof of the main theorem, outlining its structure in three main parts.

(i) Uniform action-angle variables at simple resonances

In § 4 we show how to overcome the homotopy problem due to the fact that the above symplectic transformations Φ_1 and Φ_3 are not globally well defined on the torus. Exploiting the particular form of the various symplectic transformations involved, we show that, introducing a special *ad hoc* symplectic "semi-conjugacy", one can obtain, for all $|k| \leqslant \mathtt{K}$, well defined symplectic action-angle maps ϕ_k^i, which conjugate the Hamiltonian $\mathcal{H}_k$ in (2) 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 Hamiltonians $\mathtt{H}_k$ is split by their separatrices compare Theorem 4.1 below; here, $\mathtt{H}_k = \mathtt{H}_k(\mathtt{p}, \mathtt{q}_1)$ denotes the standard form of the secular Hamiltonian $\overline{\mathtt{H}}_k$ in (2).

(ii) Kolmogorov's nondegeneracy and the Twist Theorem

As in all KAM applications, the main problem is then to prove (a suitable) *nondegeneracy* of the frequency map $I \to \omega = \partial_I h_k^i$.

The high mode case when $|k| > \mathtt{N}$ is quite simple, as the secular reference potential $\pi_{\mathbb{Z}k} f$ behaves like a cosine. On the other hand, when $|k| \leqslant \mathtt{N}$, the situation is more complicated, as, in this case, in the phase space of standard Hamiltonians (3) there are, in general, *points where the twist vanishes*. For instance, they *always* appear in regions bounded by two separatrices (with different energies). Furthermore, they appear also in very simple examples with only one separatrix in the region enclosed by the separatrix,[7] compare Remark 5.1 below.

[7]For example, in the case: $p_1^2 + \varepsilon\big(\cos q_1 - \frac{1}{8}\cos(2q_1)\big)$.

Moreover, since, as already noticed, in addressing optimal bounds on the non-torus set, one has to consider regions at a distance, in energy, of order $\varepsilon^{3/2}$ from secular separatrices, we have to face a *singular perturbation problem*.[8]

Therefore, new methods have to be developed in order to prove that the measure where the twist is small is also small in the phase spaces of all Hamiltonians $\mathcal{H}_k^i$; compare the Twist Theorem 5.4 in § 5.

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 first proves that the (normalized) second derivative of the action-to-energy functions are nondegenerate (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 real analysis ([41], [25]), one can show that η-sublevels 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. The proof, in this case, rests on the construction of a suitable differential operator with non-constant coefficients, which, exploiting 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 nondegenerate also near separatrices, and to conclude the proof of the Twist Theorem.

(iii) Conclusion of the proof

At this point, choosing carefully the various free parameters appearing in the construction, taking, in particular, $\mathtt{K} \sim |\log \varepsilon|$, by KAM, we obtain the existence of maximal primary and secondary invariant tori, filling the complementary of the double resonance neighborhood $\mathcal{R}^2 \times \mathbb{T}^n$, up to a set of measure ε, completing the proof of the main result in dimension $n \geqslant 3$.

Taking larger values of $\mathtt{K}$, and, hence, different coverings with larger neighborhoods of double resonances, allows for much better measure estimates away from double resonances. For example, disregarding a neighborhood of double resonances of measure $\sim \varepsilon^a$ with any $a < 1$, one would obtain a family of invariant tori filling $(\mathcal{R}^0 \cup \mathcal{R}^1) \times \mathbb{T}^n$, up to a set of exponentially small measure in $1/\varepsilon^c$; compare remark (R_2) below.

[8] Indeed, the largest neighborhood around secular separatrices that one can disregard has measure at most $\sim \varepsilon$ (up to a logarithmic correction), which corresponds, in energy, to a distance from separatrices of order $\sim \varepsilon^{3/2}$, indicating that standard perturbation theories are bound to fail; compare Remark 4.3, (ii), and Remark 5.2 below.

This observation, in fact, allows to conclude the proof also in the $n = 2$ case. In such a case, the only double resonance is the origin, and taking away a small disc of measure $\sim \varepsilon^a$, one obtains an *exponential density* of KAM tori, as predicted by Arnol'd, Kozlov and Neishtadt in 1985.

2 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 $\mathcal{T} \subset \mathcal{M} = B \times \mathbb{T}^n$ (B open set in $\mathbb{R}^n$) is called a *maximal KAM torus* for a real analytic Hamiltonian $H : \mathcal{M} \to \mathbb{R}$, if there exists a real analytic embedding $\phi : \mathbb{T}^n \to \mathcal{M}$ and a Diophantine frequency vector $\omega \in \mathbb{R}^n$ such that, $\mathcal{T} = \phi(\mathbb{T}^n)$, and for each $z \in \mathcal{T}$, $\Phi^t_H(z) = \phi(x + \omega t)$, where $x = \phi^{-1}(z)$ and $t \to \Phi^t_H(z)$ denotes the standard Hamiltonian flow governed by H starting at $z \in \mathcal{M}$. We recall that 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 $\omega \cdot k = \sum_j \omega_j k_j$ and $|k|_1 := \sum |k_j|$.

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:

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

$\mathcal{G}^n$ will then denote 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 also set:

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

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 a fixed 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

$$\theta \in \mathbb{T} \mapsto \pi_{\mathbb{Z}k} f(\theta) := \sum_{j \in \mathbb{Z} \setminus \{0\}} f_{jk} e^{\mathrm{i} j \theta}. \tag{7}$$

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 $\max\{|k|_1, |\ell|_1\}$.

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 nondegenerate, 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 mechanical systems with Hamiltonian $\mathtt{H}$ as in (1).

Definition 2.1 (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:*

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

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

Remark 2.1 *The requirement in (10) that the projections $\pi_{\mathbb{Z}k} f$ have distinct critical values is not really necessary, but it is generic and it simplifies proofs.*

Main Results

Theorem 2.1 *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 $\mathfrak{c} > 1$, such that for all $\mathtt{K}$ and $\varepsilon > 0$ satisfying*

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

the following holds. There exist three sets

$$\mathcal{R}^2 \subseteq \mathrm{B}, \qquad \mathcal{A} \subseteq \mathrm{B} \times \mathbb{T}^n, \qquad \mathcal{T} \subseteq \mathbb{R}^n \times \mathbb{T}^n,$$

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* $\mathtt{K}$*:*

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

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

Remark 2.2 *The three sets* $\mathcal{R}^2$*,* $\mathcal{A}$ *and* $\mathcal{T}$ *depend on the free parameter* $\mathtt{K}$*, and choosing* $\mathtt{K}$ *as a suitably large function of* ε *one obtains a small bound on the measure of* $\mathcal{A}$*; compare Corollary 2.2 below. On the other hand, one may wonder why, when* $\mathtt{K}$ *is chosen independent of* ε*, the measure estimate in point (iii) above yields such a poor bound. This is due to the following fact. We want to perform averaging theory in the complement of* $\mathcal{R}^2 \times \mathbb{T}^n$*, and, therefore, we must stay at a distance of order* $\sqrt{\varepsilon}$ *from double resonances. This means that the small divisors and the action analyticity radius have to be taken both of order* $\sqrt{\varepsilon}$*. Now, averaging out only an* ε*-independent number* $\mathtt{K}$ *of harmonics, the new perturbation would still be of order* ε*, but with an analyticity radius of order* $\sqrt{\varepsilon}$*, so that one would simply have worsened the initial situation. In particular, in applying KAM measure estimates, one would get the paradoxical result of bounding the non-torus set by an order one quantity (compare also the displayed equation at page 5).*

As immediate consequence of Theorem 2.1, choosing

$$\mathtt{K} = \mathfrak{c} |\log \varepsilon|, \tag{12}$$

there follows

$$\operatorname{meas}\big((\mathrm{B} \times \mathbb{T}^n) \setminus \mathcal{T}\big) \leqslant (2\pi)^n c_\star \, \varepsilon \mathtt{K}^\gamma + e^{-\mathtt{K}/\mathfrak{c}} \leqslant \bar{\mathfrak{c}} \; \varepsilon |\log \varepsilon|^\gamma, \tag{13}$$

with $\bar{\mathfrak{c}} := 1 + (2\pi)^n c_\star \mathfrak{c}^\gamma$ yielding the following

Corollary 2.2 *Under the assumptions of Theorem 2.1, 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* $\bar{\mathtt{c}} = 1 + (2\pi)^n c_\star \mathtt{c}^\gamma$*.*

This corollary implies the theorem announced in [11, p. 426].

For further results obtained by choosing differently the parameter $\mathtt{K}$ (and, hence, the sets $\mathcal{R}^2$ and $\mathcal{A}$), see Remark (R_2) below.

The two degrees of freedom case is special. In this case the only double resonance is the origin, and one can take $\mathcal{R}^2$ to be 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, [13]).

Corollary 2.3 *Let the assumptions of Theorem 2.1 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* $e^{-1/(2\mathtt{c}\varepsilon^{\hat{a}})}$*, with* $\hat{a} := (1-a)/24$*.*

Remark 2.3 *(i) Notice that fixing the norm of the potential* f *to be 1 in the above statements can always be achieved just by renaming the perturbative parameter.*

(ii) The choice of the action domain to be the unit sphere is done for simplicity. Notice, however, that the natural domain of a mechanical Hamiltonian is $\mathbb{R}^n \times \mathbb{T}^n$ *and that any bounded action domain will be contained in some ball centred at the origin, which, after rescaling actions and time, can be mapped onto the unit ball.*

Further remarks and 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 [17, Sect. 2].

It is very simple to give explicit examples of functions in $\mathbb{G}^n_s$, a prototype being (recall the definitions in (5), (8) and (7)):

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

which satisfies

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

as one readily verifies.

The class of potentials $\mathbb{G}^n_s$ is quite general from various points of view. For example, the following result (proven in [17, Sect. 1]) holds.

Proposition 2.4 *(i) $\mathbb{G}^n_s$ 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* $\mathrm{F}(\mathbb{B}^n_s)$*.*

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 distinct 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 2.2 (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 2.3 (The cut-off function $\mathtt{N}$) *Given $0 < \delta \leqslant 1$ and $n, s > 0$, define the following "Fourier cut-off function":*

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

Then, the following elementary result holds:

Lemma 2.5 *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 is given in Appendix A.

Remarks

(R_1) By item (i) in Theorem 2.1, we see that the phase space of a generic nearly integrable mechanical 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 tubular 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 which is not known to be close to integrable.

- The set $\mathcal{A}$, which has measure $\sim e^{-\mathtt{K}/\mathfrak{c}}$, 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; for references on Arnol'd diffusion see, e.g., [3, 20, 10, 43, 23, 45, 33, 9, 22, 19, 27, 29].

- 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 primary or secondary KAM tori* for $\mathtt{H}$ and the dynamics is quasi-periodic.

(R_2) The sets $\mathcal{R}^i$ and $\mathcal{A}$ depend, in particular, upon ε and $\mathtt{K}$. Choosing $\mathtt{K}$ larger than $|\log \varepsilon|$, one obtains larger neighborhoods of double resonances, and a higher density of KAM tori in the complementary region. For example, Letting $\mathtt{K} = \log^2 \varepsilon$ in Theorem 2.1 with ε small enough (so that (11) is met), the sets $\mathcal{R}^2$ and $\mathcal{A}$ satisfy the estimates

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

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 $\mathcal{R}^2 \times \mathbb{T}^n$. Indeed, let $0 < a < 1$, choose $\mathtt{K} = 1/\varepsilon^{\bar a}$ with $\bar a := (1-a)/\gamma$ in Theorem 2.1, and let ε be small enough. Then, the sets $\mathcal{R}^2$ and $\mathcal{A}$ satisfy the estimates

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

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) Corollary 2.3 and Corollary 2.2 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 2 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 Corollary 2.3 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. The secondary tori are *not* deformations 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 2.1 there appear two constants $\mathtt{c}$ and $c_\star$ (the other constants appearing in the corollaries of Theorem 2.1 are simply related to such two constants). The constant $c_\star$ depends only on the action-dimension n (compare Lemma 3.1 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. Indeed, $\mathtt{c}$ depends on six parameters: $n \geqslant 2$; $s > 0$; the positive numbers δ and β in Lemma 2.5 above, and two more positive parameters $\mathfrak{x}$ and $\mathtt{m}$ (introduced in Definition 5.3 below). These latter parameters, in turn, are suitable nondegeneracy 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 2.3. For more details on this point, see (246) and (247) below.

(R_6) One of the main issues in this context 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 mechanical Hamiltonian systems; compare, also open problem (i) below.

(R_7) As it is well known, classical KAM Theory may be used to construct locally KAM tori in various singular situations (e.g., secondary tori near elliptic equilibria; in case of degenerate integrable limits, when the degeneracy is removed by the perturbation; near a fixed separatrix ([34]); etc. (for general discussions, see [5]). The methods developed here (and in [14], [16] and [17]) might be viewed as a *global and uniform* extensions of such ideas: "Global", since one has to deal, simultaneously, with tori in the *whole* phase space (except for a neighbourhood of double resonances); "uniform" since one needs to determine explicit quantitative conditions working for *all* neighbourhoods of simple resonances (whose number tends to infinity as $\varepsilon \to 0$).

Open problems

(i) In [8], using methods described here, Theorem 2.1 has been extended to the case $h(y) + \varepsilon f(x)$ with h convex, while extensions to the case where h is Kolmogorov non-degenerate or steep are not as simple. More relevantly, one would like to consider potentials depending also on actions. In fact, besides the obvious (but non generic) case in which f verifies (16) and (17) *uniformly in* $y \in \mathrm{B}$, in general, the Fourier coefficients $f_k(y)$ will vanish at some points $y \in \mathrm{B}$, introducing more singularities; for a discussion of this point in the planar, restricted, circular three-body problem, see [26]. Selecting an analytic generic class of perturbations $f(y, x)$, to which Theorem 2.1 extends, is a non-trivial issue.[9]

(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 2.2 appears to be unavoidable, and one may wonder if it is really necessary or if it is an artefact of the methods.

(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; for related results in smooth category, see [32].

(v)* *Generic Arnol'd diffusion in analytic class:* One might wonder if, 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.[10]

3 Prerequisites

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

[9] As pointed out by Laurent Niederman, it might be useful to consider the quantitative Morse–Sard theory developed by Yomdin and Comte in [44]; for a first step in this direction, see [7].

[10] The stress, here, is on analyticity. Indeed, after John Mather's seminal work, there are well known generic ("cusp-generic") results in Arnol'd diffusion in C^r regularity and low dimensions; compare [29].

We begin by recalling the averaging theory for nearly integrable real analytic Hamiltonian systems especially designed for neighborhoods of simple resonances; compare [14, 17].

3.1 Coverings and averaging

Coverings

Here, we first introduce a suitable covering of the action space (depending on ε and parameters $\mathtt{K}$ and $\mathtt{K}_{\rm o}$), and then prove item (ii) of Theorem 2.1.

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

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

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 the unit ball $\mathrm{B} = \{y \in \mathbb{R}^n : \ |y| < 1\}$:

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

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

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

The first key remark is that *the measure of* $\mathcal{R}^2$ *is proportional to* $\varepsilon\mathtt{K}^\gamma$, as shown in the following lemma, which proves item (ii) of Theorem 2.1; compare, also, Lemma 2.5 in [17].

Lemma 3.1 *There exists a constant* $c_\star = c_\star(n) > 1$ *such that*

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

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

$$\mathcal{R}^2_{k\ell} := \big\{ y \in \mathrm{B} : |y \cdot k| < \alpha;\ |\pi_k^\perp y \cdot \ell| \leqslant \tfrac{3\alpha \mathtt{K}}{|k|} \big\}. \tag{26}$$

The measure of $\mathcal{R}^2_{k,\ell}$ can be estimated as follows. 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 can be written uniquely as $v = ak + bh$, for suitable $a, b \in \mathbb{R}$. By (27),

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

and, since $|\ell|^2|k|^2 - (\ell \cdot k)^2$ is a positive integer,

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

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 sides $2\alpha/|k|^2$ and $6\alpha\mathtt{K}|k|$ (compare (29) and (30)), we find, for $k \in \mathcal{G}^n_{\mathtt{K}_o}$ and $\ell \in \mathcal{G}^n_{\mathtt{K}} \setminus \mathbb{Z}k$,

$$\mathrm{meas}(\mathcal{R}^2_{k,\ell}) \leqslant \frac{2\alpha}{|k|^2}\,(6\alpha\mathtt{K}|k|)\ 2^{n-2} = 3 \cdot 2^n\, \alpha^2 \frac{\mathtt{K}}{|k|}.$$

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

Remark 3.1 *(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* $\mathtt{K}_o$*;* $\mathcal{R}^{1,k}$ *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) *above). According to the terminology in [40],* $\mathcal{R}^0$ *is* $(\alpha/2, \mathtt{K}_{\rm o})$ *completely nonresonant, 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 3.3 below; compare also [14].*

(v) The size of $\mathcal{R}^{1,k}$ *depends also (mildly) on* $|k|$*, and it is conceivable that changing such dependence, the constant* γ *in* (24) *could be improved.*

(vi) The set $\mathcal{A}$ *in Theorem 2.1 is defined as* $\mathcal{A} := \big((\mathcal{R}^0 \cup \mathcal{R}^1) \times \mathbb{T}^n\big) \setminus \mathcal{T}$*, compare* (242) *below.*

Averaging

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

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

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,[11] by

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

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

[11] The sup-norm (rather than the norm in (8)), is more natural in averaging theory.

(20). We define:

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

We also need the following consequence of Bezout's Lemma, which will allow to define a "resonant angle" near simple resonances. For M matrix (or vector), we denote by $|M|_\infty$ its maximum norm $\max_{ij}|M_{ij}|$ (or $\max_i |M_i|$).

Lemma 3.2 *For any $k \in \mathcal{G}^n$ there exists a matrix $\hat{\rm A} \in \mathbb{Z}^{(n-1)\times n}$ such that:*

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

Proof From Bézout's lemma it follows easily that: *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* $|{\rm A}|_\infty = |k|_\infty$; for a detailed proof, see, Lemma A.1 in [14].

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 the D'Alembert expansion of determinants. ∎

In fact, the linear symplectic transformation generated by the function $\mathtt{y}\cdot {\rm A}x$, namely, the map

$$(y,x) \in \mathbb{R}^n \times \mathbb{T}^n \mapsto (\mathtt{y},\mathtt{x}) = \big({\rm A}^{-T}y, {\rm A}x\big) \in \mathbb{R}^n \times \mathbb{T}^n,$$

transforms the resonant combination $k\cdot x$ into the "resonant" angle $\mathtt{x}_1$.

Thus, the following normal form result holds:

Theorem 3.3 (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. Let $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, let ${\rm A}$ be a matrix as in Lemma 3.2, and define, for $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, the following real sets:*

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

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

$$\Psi_o : \mathcal{R}^0_{r'_o} \times \mathbb{T}^n_{s'_o} \to \mathcal{R}^0_{r_o} \times \mathbb{T}^n_{s_o}, \qquad \Psi^k : \mathcal{D}^k_{\tilde{r}_k} \times \mathbb{T}^n_{\tilde{s}_k} \to \mathcal{R}^{1,k}_{r_k} \times \mathbb{T}^n_{s_\star}, \tag{35}$$

having the following properties.

(i) In the symplectic variables $(y, x) \in \mathcal{R}^0_{r'_o} \times \mathbb{T}^n_{s'_o}$, $\mathtt{H}$ *takes the form:*

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

with g^o *and* f^o *satisfying*

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

(ii) In the symplectic variables $(\mathtt{y}, \mathtt{x}) = \big(\mathtt{y}, (\mathtt{x}_1, \hat{\mathtt{x}})\big) \in \mathcal{D}^k_{\tilde{r}_k} \times \mathbb{T}^n_{\tilde{s}_k}$, $\mathtt{H}$ *takes the form:*

$$\mathcal{H}_k(\mathtt{y}, \mathtt{x}) := \mathtt{H} \circ \Psi^k(\mathtt{y}, \mathtt{x}) = \overline{\mathtt{H}}_k(\mathtt{y}, \mathtt{x}_1) + \varepsilon \bar{f}^k(\mathtt{y}, \mathtt{x}), \tag{37}$$

where

$$\overline{\mathtt{H}}_k(\mathtt{y}, \mathtt{x}_1) := \frac{1}{2}|\mathrm{A}^T \mathtt{y}|^2 + \varepsilon \mathtt{g}^k_o(\mathtt{y}) + \varepsilon \mathtt{g}^k(\mathtt{y}, \mathtt{x}_1), \qquad \langle \mathtt{g}^k(\mathtt{y}, \cdot) \rangle = 0, \tag{38}$$

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

$$|\mathtt{g}^k_o|_{\tilde{r}_k} \leqslant \vartheta_o, \qquad |\mathtt{g}^k - \pi_{\mathbb{Z}k} f|_{\tilde{r}_k, s'_k} \leqslant \vartheta_o, \qquad |\bar{f}^k|_{\tilde{r}_k, \tilde{s}_k} \leqslant e^{-\mathtt{K}s/3}. \tag{39}$$

(iii) If $|k|_1 \geqslant \mathtt{N}$, *then* $\mathtt{g}^k$ *and* $\bar{f}^k$ *in* (38), (37) *take the form*

$$\mathtt{g}^k = 2|f_k|\big(\cos(\mathtt{x}_1 + \theta_k) + F^k_\star(\mathtt{x}_1) + \mathtt{g}^k_\star(\mathtt{y}, \mathtt{x}_1)\big), \quad \bar{f}^k = 2|f_k|\ \mathtt{f}^k_\star(\mathtt{y}, \mathtt{x}), \tag{40}$$

where θ_k *is a suitable number in* $[0, 2\pi)$,

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

$\mathtt{g}^k_\star(y, \cdot) \in \mathbb{B}^1_1$ *(*$\forall\ y \in \mathcal{D}^k_{\tilde{r}_k}$*),* $\pi_{\mathbb{Z}k} \mathtt{f}^k_\star = 0$, *and*

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

(iv) Finally, the following "coverings" holds:

$$\Psi_o\big(\widetilde{\mathcal{R}}^0 \times \mathbb{T}^n\big) \supseteq \mathcal{R}^0 \times \mathbb{T}^n, \qquad \Psi^k\big(\mathcal{D}^k \times \mathbb{T}^n\big) \supseteq \mathcal{R}^{1,k} \times \mathbb{T}^n. \tag{41}$$

For the proof, see [17] (in particular, Theorem 2.1 and the Covering Lemma 2.3).

Remark 3.2 *(i) Note that the transformation $\Psi_{\rm o}$ is close to the identity, while Ψ^k is not, being the composition of the linear transformation defined after Lemma 3.2 with a close to the identity map.*

(ii) The larger covering in (34) is introduced so that (41) 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) Having the common factor $|f_k|$, for $|k|_1 \geqslant \mathtt{N}$, in (40) is non trivial and it depends essentially on the fact that f belongs to the class $\mathbb{G}^n_s$, and, more precisely, that satisfies (16) with $\mathtt{N}$ as in (15). Notice also that the functions in (40) have three different scales as $\mathtt{K} \to +\infty$: a $O(1)$ scale of the cosine-like function $\cos(\mathtt{x}_1+\theta_k)+F^k_\star(\mathtt{x}_1)$, a $O(1/\mathtt{K}^{5n})$ scale of the term $\mathtt{g}^k_\star$, which depends only on the slow angle $\mathtt{x}_1$, and the exponential scale $e^{-\mathtt{K}s/7}$ of the term $\mathtt{f}^k_\star$, which depend on all variables. All these scales will be important in the following analysis.

(iv) In Theorem 3.3, it is assumed that $\mathtt{K}_{\rm o} \geqslant \mathbf{c_o} \geqslant \mathtt{N}$. In fact, one always has

$$\mathtt{K}_{\rm o} \geqslant \mathtt{N} \geqslant 2\mathtt{c}_s, \qquad \text{where} \qquad \mathtt{c}_s := \max\{1, 1/s\}; \tag{42}$$

indeed, 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.

3.2 Standard form at simple resonances

It turns out that the secular Hamiltonians $\overline{\mathtt{H}}_k(\mathtt{y}, \mathtt{x}_1)$ in the Normal Form Theorem 3.3 have a common uniform analytic structure. Indeed, $\overline{\mathtt{H}}_k$, regarded as one-degree-of-freedom systems parameterized by $(\mathtt{y}_2, ..., \mathtt{y}_n)$, can be put, *uniformly in* $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, into a suitable "standard form", close to a generalized pendulum $p_1^2 + \bar{\mathtt{G}}(q_1)$. This procedure has the advantage to make possible an explicit analytic treatment of the corresponding (one-dimensional) action-angle variables (as carried out in details in [16], and reviewed in § 3.3 below). We remark, however, that this construction *cannot* be done, in general, in the full phase space, because of a homotopy problem arising in the rotational regions; compare Remark 3.4 and Remark 3.6-(ii) below.

Definition 3.1 (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 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*

$$\mathtt{H}_\flat(p, q_1) = \big(1 + \nu(p, q_1)\big)p_1^2 + \mathtt{G}(\hat{p}, q_1), \tag{43}$$

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 vanishing* q_1*-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 (see Definition 2.2);*

- *the following estimates hold:*

$$\begin{aligned} &\sup_{\mathbb{T}^1_{\mathtt{s}}} |\bar{\mathtt{G}}| \leqslant \epsilon, \\ &\sup_{\hat{D}_{\mathtt{r}} \times \mathbb{T}^1_{\mathtt{s}}} |\mathtt{G} - \bar{\mathtt{G}}| \leqslant \epsilon\mu, \quad \text{for some} \quad 0 < \epsilon \leqslant \mathtt{r}^2/2^{16}, \quad 0 \leqslant \mu < 1, \\ &\sup_{D_{\mathtt{r}} \times \mathbb{T}^1_{\mathtt{s}}} |\nu| \leqslant \mu. \end{aligned} \tag{44}$$

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 3.3 *(i) If* $\mathtt{H}_\flat$ *is in standard form, then* β *and* ϵ *satisfy the relation* $\epsilon/\beta \geqslant 1/2$*. Indeed, by* (14) *and* (44)*, if* $i \neq j$*,*

$$\beta \leqslant |\bar{\mathtt{G}}(\theta_i) - \bar{\mathtt{G}}(\theta_j)| \leqslant 2 \max_{\mathbb{T}} |\bar{\mathtt{G}}| \leqslant 2\epsilon.$$

Furthermore, one can always fix a number $\mathfrak{g} \geqslant 4$*, so that:*

$$1/\mathfrak{g} \leqslant \mathtt{s} \leqslant 1, \qquad 1 \leqslant \mathtt{R}/\mathtt{r} \leqslant \mathfrak{g}, \qquad 1/2 \leqslant \epsilon/\beta \leqslant \mathfrak{g}. \tag{45}$$

Such a parameter $\mathfrak{g}$ *rules the main scaling properties of Hamiltonians in standard form.*

(ii) A Hamiltonian in standard form $\mathtt{H}_\flat$ *has the analytic features of its reference mechanical Hamiltonian*

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

In particular, for μ *small with respect to* $1/\mathfrak{g}$*,* $\mathtt{H}_\flat$ *has the same finite (because of analyticity) number of equilibria of* $\bar{\mathtt{G}}$*. Furthermore, these equilibria lie on the* q_1 *axis and are in the same relative order of those of* $\bar{\mathtt{G}}$*; the order is preserved also for critical energies; compare Lemma 3.6 below.*

(iii) The smallness of the "adimensional ratio" $\epsilon/\mathtt{r}^2$ *in* (44) *is needed in the analytic theory of action-angle variables for Hamiltonians in standard form developed in [16], however the factor* $1/2^{16}$ *is rather arbitrary and not optimal.*

Notation 1 *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:*

$$\begin{aligned}
&\textit{if}\quad w = (y,x) = \big((y_1,...,y_n),(x_1,...x_n)\big) \quad \textit{then:}\\
&\hat{w} = (w)\hat{} = (x_2,...,x_n) = \hat{x},\\
&\hat{y} = (y)\hat{} = (y_2,...,y_n),\\
&\check{w} = (w)\check{} = (y,x_1),\\
&w = (\check{w},\hat{w}).
\end{aligned} \tag{46}$$

Next, we introduce a special group of symplectic transformations, which will appear in Theorem 3.4 below.

Definition 3.2 *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*

$$\begin{aligned}
&(p,q) \in (\mathbb{R}\times\hat{\mathrm{D}})\times\mathbb{R}^n \overset{\Psi_{\mathsf{g}}}{\mapsto} (P,Q)\\
&(P,Q) := (p_1+\mathsf{g}(\hat{p}),\hat{p},q_1,\hat{q}-q_1\partial_{\hat{p}}\mathsf{g}(\hat{p})) \in (\mathbb{R}\times\hat{\mathrm{D}})\times\mathbb{R}^n,
\end{aligned} \tag{47}$$

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

Remark 3.4 *The group properties of $\mathfrak{G}_{\dagger}$ are trivial:*

$$\Psi_0 = \mathrm{id}_{\mathfrak{G}_{\dagger}}, \qquad \Psi_{\mathsf{g}}^{-1} = \Psi_{-\mathsf{g}}, \qquad \Psi_{\mathsf{g}}\circ\Psi_{\mathsf{g}'} = \Psi_{\mathsf{g}+\mathsf{g}'}. \tag{48}$$

Notice that a map $\Psi_{\mathsf{g}} \in \mathfrak{G}_{\dagger}$ induces on $\mathbb{T}^n$ the map

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

which, unless $\partial_{\hat{p}}\mathsf{g} \in \mathbb{Z}^{n-1}$, is not a well defined map on $\mathbb{T}^n$.[12] *This fact raises a problem in applying the theory of this and next section to the normalized Hamiltonians $\mathcal{H}_k$ of Theorem 3.3; compare Remark 3.6-(ii) below.*

From now on we shall make the following quantitative assumption.

Recall Lemma 2.5.

[12] 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 map $x\in\mathbb{R}^n \to Ax+f(x)\in\mathbb{R}^n$ with the 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 $\mathbb{T}^n$.

Assumption 3.1 *Fix* $n \geqslant 2$*,* $s > 0$*, and let* $\mathtt{H}$ *be as in* (1) *with* $f \in \mathbb{G}^n_s$ *(Definition 2.1), satisfying* (16) *and* (17) *for some* $0 < \delta \leqslant 1$ *and* $\beta > 0$*, with* $\mathtt{N}$ *as in* (15)*.*

Recall, now, the definitions of: $\mathtt{K}_{\rm o}$, $\mathtt{K}$, α and ν in (20); A and $\hat{\text{A}}$ in Lemma 3.2; $c_{_1}$ and s'_k in (32); $\mathtt{c}_s$ in (42). In the following, we shall not always indicate explicitly the dependence upon k.

Definition 3.3 *For* $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$*, let:*

$$
\begin{aligned}
&\mathtt{R} = \alpha/|k|^2 = \sqrt{\varepsilon}\mathtt{K}^\nu/|k|^2, \quad c_{_2} = 4\, n^{\frac{3}{2}} c_{_1}, \quad \mathtt{r} = \mathtt{R}/c_{_2}, \quad \varepsilon_k = \frac{2\varepsilon}{|k|^2},\\
&\upbeta = \begin{cases} \varepsilon_k \beta, & \textit{if } |k|_{_1} < \mathtt{N} \\ \varepsilon_k |f_k|, & \textit{if } |k|_{_1} \geqslant \mathtt{N} \end{cases}, \qquad \chi_{_k} = \begin{cases} 1, & \textit{if } |k|_{_1} < \mathtt{N} \\ |f_k|, & \textit{if } |k|_{_1} \geqslant \mathtt{N} \end{cases},\\
&\mathtt{s} = \begin{cases} \min\{\frac{s}{2}, 1\}, & \textit{if } |k|_{_1} < \mathtt{N} \\ 1, & \textit{if } |k|_{_1} \geqslant \mathtt{N} \end{cases}, \quad \check{\mathtt{s}} = \begin{cases} s'_k, & \textit{if } |k|_{_1} < \mathtt{N}, \\ 1, & \textit{if } |k|_{_1} \geqslant \mathtt{N} \end{cases},\\
&\hat{D} = \Big\{ \hat{I} \in \mathbb{R}^{n-1} : \ |\pi^\perp_k \hat{\text{A}}^T \hat{I}| < 1, \ \min_{\substack{\ell \in \mathcal{G}^n_{\mathtt{K}} \\ \ell \notin \mathbb{Z}k}} \big| \big(\pi^\perp_k \hat{\text{A}}^T \hat{I}\big) \cdot \ell \big| \geqslant 3\alpha \mathtt{K}/|k| \Big\},\\
&D = (-\mathtt{R}, \mathtt{R}) \times \hat{D}, \qquad \upepsilon = \mathtt{c}_s \varepsilon_k \, \chi_{_k}, \qquad \upmu = \frac{1}{\mathtt{K}^{5n}}.
\end{aligned}
\tag{49}
$$

Remark 3.5 *(i) Since* $|f_k| \leqslant 1$ *one has:*

$$
|\chi_{_k}| \leqslant 1. \tag{50}
$$

Furthermore, by the definitions in (49) *and* (20)*, by* (50) *and* (42)*, one has*

$$
\sqrt{\upepsilon} < \mathtt{c}_s \mathtt{R}/\mathtt{K}^{\nu-1} < \mathtt{R}/\mathtt{K}^{\frac{9}{2}n}. \tag{51}
$$

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

$$
\mathtt{s} \leqslant 2\check{\mathtt{s}}. \tag{52}
$$

We can, now, state the main result of [17] (namely, Theorem 3.1 there). Recall the notation in (46) and Definition 3.2; r_k, $\tilde{r}_k$ and s'_k are defined in (32); $\mathtt{c}_s$ is defined in (42).

Theorem 3.4 (Standard form at simple resonances)
Let Assumption 3.1 and Definition 3.3 hold, let $\mathbf{c_o}$ *be the constant defined in Theorem 3.3, and assume that* $\mathtt{K}_{\rm o} \geqslant \max\{c_{_2}, \mathbf{c_o}\}$*. Then, for all* $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$*, the following holds.*

(i) There exists a real analytic symplectic transformation

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

such that: $\Phi_\star$ *fixes* $\hat{\mathtt{p}}$ *and* $\mathtt{q}_1$ *(namely,* $\hat{\mathtt{y}} = \hat{\mathtt{p}}, \mathtt{x}_1 = \mathtt{q}_1$*); 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* $\check{\Phi}_\star$ *depends only on the first* $n+1$ *coordinates* $(\mathtt{p},\mathtt{q}_1)$*, is* 2π*-periodic in* $\mathtt{q}_1$ *and, if* $\mathcal{D}^k = \mathrm{A}^{-T}\mathcal{R}^{1,k}$ *and* $\overline{\mathtt{H}}_k$ *is as in Theorem 3.3, one has*

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

(ii) $\mathtt{H}_k$ *in (53) is in standard form according to Definition 3.1:*

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

having reference potential

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

and analyticity characteristics as in (49). The constant $\mathfrak{g}$ *in (45) can be taken as:*

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

(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:

- $\Phi_{{}_1} := \Psi_{\mathtt{g}_{{}_1}} \in \mathfrak{G}_{{}_\dagger}$ *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*

$$|\mathtt{g}_{{}_3}|_{4\mathtt{r}} < \tfrac{\varepsilon_k}{\mathtt{r}}\mu. \tag{59}$$

Remark 3.6 *(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 standard form. The word "uniform" refers to the fact that the parameter* $\mathfrak{g}$ *(defined in* (56) *and satisfying* (45)*) – which rules the scaling properties of the normalized Hamiltonians* $\mathtt{H}_k$ *– does not depend upon* k*, allowing, e.g., for a uniform (in* $k \in \mathcal{G}^n_{\mathtt{K}_o}$*) treatment of action-angle variables (compare next Section 3.3).*

(ii) There is, however, a drawback in the construction of the above normal forms, namely, that the maps Φ_1 *and* Φ_3 *appearing in the definition of* $\Phi_\star$ *(see item (iii) in the above theorem), do not induce well defined maps on* $\mathbb{T}^n$*; compare Remark 3.4. 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* Φ_2 *is well defined also on* $\mathbb{T}^n$*. This matter will be discussed in details in Section 4.*

The following remark explains the individual purpose of the three symplectic transformations Φ_j, whose composition forms $\Phi_\star$.

Remark 3.7 *(i) The map* Φ_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* Φ_1 *as*

$$(\mathtt{y},\mathtt{x}) = \Phi_1(p,q) := \big(\mathrm{U}p, \mathrm{U}^{-T}q\big), \quad \text{where} \quad \mathrm{U} := \begin{pmatrix} 1 & -\frac{1}{|k|^2}(\hat{\mathrm{A}}k)^T \\ 0 & \mathrm{id}_{n-1} \end{pmatrix}, \tag{60}$$

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 (49)*, the definitions of* Φ_1 *and* U*, and the definition of* $\mathcal{D}^k$ *in Theorem 3.3-(i), it follows that*

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

Incidentally, we observe that from (33) *it follows that the norms of* U *and its inverse satisfy the bounds*

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

where, as usual, for a matrix M, $|M| = \sup_{u \neq 0} |Mu|/|u|$.

(ii) The second map Φ_2 *is a close 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 close to identity symplectic map, which sets all critical points on the line* $p_1 = 0$.

3.3 Action-angle variables for 1D standard Hamiltonians

Here, we review the general theory of action-angle variables for Hamiltonian systems in standard form, as developed in [16], 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 (49) 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 given Hamiltonian $(p_1, q_1) \mapsto \mathtt{H}_\flat(p_1, \hat{p}, q_1)$ in standard form according to Definition 3.1.

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

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

where $\mathtt{R}$ and $\mathtt{r}$ are as in Definition 3.1. Although such sets depend on the parameter $\hat{p} \in \hat{D}$, for μ small enough, they are close to cylinders:

Lemma 3.5 *Let* $\mathtt{H}_\flat$, $\mathcal{M}$, *and* $\mathfrak{g}$ *be, respectively, as in Definition 3.1,* (65), *and* (45). *Assume that*

$$\upmu \leqslant 1/(4\mathfrak{g})^2. \tag{66}$$

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 simple proof is given in Appendix A.

Since the reference potential $\bar{\mathtt{G}}$ is a $\upbeta$-Morse function, it has $2N$ critical points, for some $N \in \mathbb{N}$, and distinct 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 nondegenerate 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 $\mathfrak{g}$, 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 2.2 and the Implicit Function Theorem, the following result proven in [16, Lemma 3.1] holds:

Lemma 3.6 *Let $\mathtt{H}_\flat$ be as in Definition 3.1 and assume that*

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

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

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

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 $\bar{E}_i$.

Notice that condition (70) is stronger than (66). Thus, under the assumption (70), we see that the phase space $\mathcal{M}$ is disconnected by the separatrices (i.e., the stable manifolds of the hyperbolic points $(0, \theta_{2j})$) 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 (see Figure 1):

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

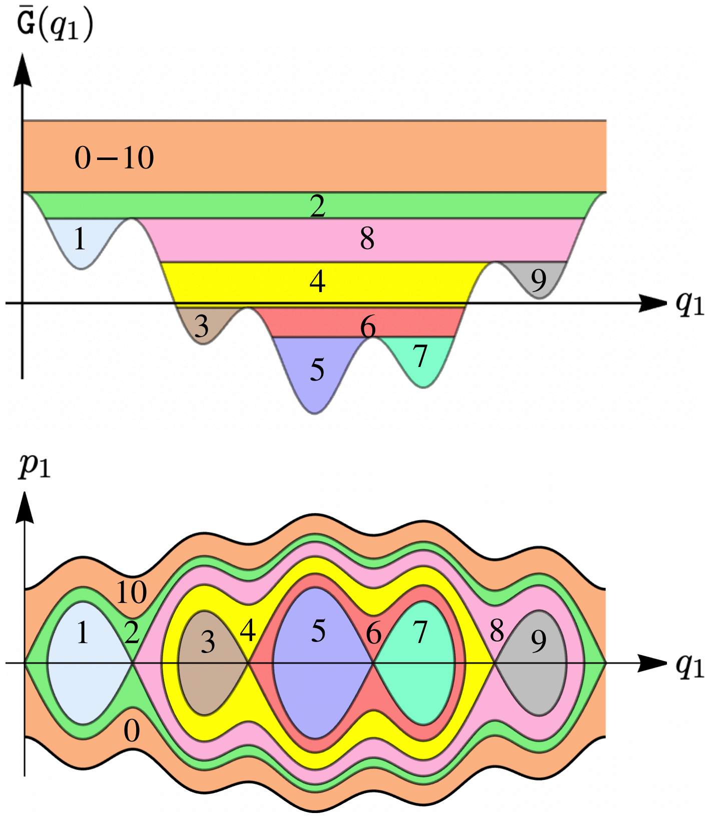

Figure 1: Top: the Morse potential $\bar{\mathtt{G}}(q_1) := f(q_1 + \bar{q}_1)$ where $f(q_1) := \sin q_1 + \frac{1}{2}\cos(5q_1)$ attains its maximum at $\bar{q}_1$and has $2N = 10$ critical points. Bottom: the phase portrait of $\mathtt{H}_\flat := p_1^2 + \bar{\mathtt{G}}(q_1)$. Labels $(0 \leqslant i \leqslant 10)$ in corresponding regions are as in Definition 3.4.

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 3.4 (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*

$$
\begin{aligned}
& j_- := \max\{\ell < j : \ E_{2\ell} > E_{2j}\}, \quad j_+ := \min\{\ell > j : \ E_{2\ell} > E_{2j}\}, \\
& E_+^{(i)} := \min\{E_{2j_-}, E_{2j_+}\};
\end{aligned}
$$

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* $(0, \theta_i)$.

Notice that the phase space $\mathcal{M}$ of $\mathtt{H}_\flat$ 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:

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

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, piecewise) smooth closed curve in the closure of $\mathcal{M}^i(\hat{p})$ given by

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

oriented clockwise; for the non contractible curves ($i = 0, 2N$) the orientation is "to the right" on $\mathcal{M}^{2N}$, "to the left" on $\mathcal{M}^0$. 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

$$I_1^{(i)}(E) = I_1^{(i)}(E, \hat{p}) := \frac{1}{2\pi} \oint\limits_{\gamma^i} p_1 dq_1,$$

$$I_j = \frac{1}{2\pi} \oint\limits_{\gamma^i_j} p_j dq_j = \frac{p_j}{2\pi} \int_{\mathbb{T}} dq_j = p_j, \qquad (2 \leqslant j \leqslant N).$$

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

$$\bar{I}^i_1 := I^i_1|_{\mu=0} \qquad \text{and its inverse function} \qquad \bar{\mathtt{E}}^i := \mathtt{E}^i|_{\mu=0}; \tag{74}$$

note that when $\mu = 0$, $\mathtt{H}_\flat$ becomes simply $\bar{\mathtt{H}}_\flat = p_1^2 + \bar{\mathtt{G}}(q_1)$.

We can now describe the fine analytic properties of the action/energy functions.

Critical holomorphic behaviour and action estimates

Here, we describe the behaviour of the action functions of a standard form Hamiltonian, as the energy approaches the critical energy of separatrices. We also recall estimates on the derivatives of the action functions: such estimates will play a central role in the discussion on the twist Hessian matrix in § 5. The following theorem has been proven in [16, Theorem 3.1].

Theorem 3.7 *Let $\mathtt{H}_\flat$ be a Hamiltonian in standard form as in Definition 3.1, let $\mathfrak{g} \geqslant 4$ be such that (45) holds and let $2N$ be the number of critical points of the reference potential $\bar{\mathtt{G}}$. Then, there exists a suitable constant $\mathbf{c} = \mathbf{c}(n, \mathfrak{g}) \geqslant 2^8\mathfrak{g}^3$ such that, if*

$$\mu \leqslant 1/\mathbf{c}^2 \leqslant 1/(2^{16}\mathfrak{g}^6), \tag{75}$$

then, the action function $E \in (E^i_-(\hat{I}), E^i_+(\hat{I})) \mapsto I^i_1(E, \hat{I})$ verifies, for all $\hat{I} \in \hat{D}$ and $0 \leqslant i \leqslant 2N$, 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, for all* $0 < z < 1/\mathbf{c}$ *and* $\hat{I} \in \hat{D}$,

$$I_1^i\big(E_\mp^i(\hat{I}) \pm \epsilon z,\ \hat{I}\big) = \phi_\mp^i(z,\hat{I}) + \psi_\mp^i(z,\hat{I})\ z\log z; \tag{76}$$

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

$$\begin{aligned} &\sup_{|z|<1/\mathbf{c},\, \hat{I}\in\hat{D}_{\mathtt{r}}} \big(|\phi_\pm^i| + |\psi_\pm^i|\big) \leqslant \mathbf{c}\sqrt{\epsilon},\\ &\sup_{|z|<1/\mathbf{c},\, \hat{I}\in\hat{D}_{\mathtt{r}/2}} \big(|\partial_{\hat{I}}\phi_\pm^i| + |\partial_{\hat{I}}\psi_\pm^i|\big) \leqslant \mathbf{c}\mu_{\rm o}, \qquad \mu_{\rm o} := \frac{\sqrt{\epsilon}}{\mathtt{r}}\mu \overset{(44)}{\leqslant} 2^{-8}\mu. \end{aligned} \tag{77}$$

Moreover, setting $\bar{\phi}_\pm^i := \phi_\pm^i|_{\mu=0}$ *and* $\bar{\psi}_\pm^i := \psi_\pm^i|_{\mu=0}$, *one has*

$$\sup_{|z|<1/\mathbf{c},\, \hat{I}\in\hat{D}_{\mathtt{r}}} \big(|\phi_\pm^i - \bar{\phi}_\pm^i| + |\psi_\pm^i - \bar{\psi}_\pm^i|\big) \leqslant \mathbf{c}\sqrt{\epsilon}\mu. \tag{78}$$

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

$$\begin{array}{lll} |\psi_+^i(0,\hat{I})| \geqslant \sqrt{\epsilon}/\mathbf{c}, & 0 < i < 2N, & \forall\ \hat{I} \in \hat{D}_{\mathtt{r}},\\ |\psi_-^{2j}(0,\hat{I})| \geqslant \sqrt{\epsilon}/\mathbf{c}, & 0 \leqslant j \leqslant N, & \forall\ \hat{I} \in \hat{D}_{\mathtt{r}},\\ \psi_+^i(0,\hat{I}) > 0, & 0 < i < 2N, & \forall\ \hat{I} \in \hat{D},\\ \psi_-^{2j}(0,\hat{I}) < 0, & 0 \leqslant j \leqslant N, & \forall\ \hat{I} \in \hat{D}, \end{array} \tag{79}$$

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

$$\phi_-^{2j-1}(0,\hat{I}) = 0, \qquad \psi_-^{2j-1}(z,\hat{I}) = 0, \qquad \forall\ 1 \leqslant j \leqslant N. \tag{80}$$

(iii) (Estimates on derivatives of actions on real domains) *The derivatives of the actions with respect to energy verify, on real domains, for all* $\hat{I} \in \hat{D}$, *the following estimates:*

$$\begin{aligned} &\inf_{(E_-^i, E_+^i)} \partial_E I_1^i \geqslant \frac{1}{\mathbf{c}\sqrt{\epsilon}}, && \forall\ 0 < i < 2N;\\ &\min\big\{\partial_E I_1^{2N},\ \partial_E I_1^0\big\} \geqslant \frac{1}{\mathbf{c}\sqrt{E+\epsilon}}, && \forall\, E > E_{2N}. \end{aligned} \tag{81}$$

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

$$\mathbf{c}\mu \leqslant \lambda \leqslant 1/\mathbf{c}, \tag{82}$$

define the complex energy domains $\mathcal{E}^i_\lambda$*, respectively, as follows:*

$$\begin{aligned}
&\Big\{z \in \mathbb{C} : \bar{E}^i_- - \frac{\epsilon}{\mathbf{c}} \ < \operatorname{Re} z < \bar{E}^i_+ - \lambda\epsilon,\ |\operatorname{Im} z| < \frac{\epsilon}{\mathbf{c}}\Big\}, && i \text{ odd},\\
&\Big\{z \in \mathbb{C} : \bar{E}^i_- + \lambda\epsilon < \operatorname{Re} z < \bar{E}^i_+ - \lambda\epsilon,\ |\operatorname{Im} z| < \frac{\epsilon}{\mathbf{c}}\Big\}, && i \text{ even}, i \neq 0, 2N,\\
&\Big\{z \in \mathbb{C} : \bar{E}^i_- + \lambda\epsilon < \operatorname{Re} z < \bar{E}^i_+,\ |\operatorname{Im} z| < \frac{\epsilon}{\mathbf{c}}\Big\}, && i = 0, 2N.
\end{aligned} \tag{83}$$

Then, for $0 \leqslant i \leqslant 2N$*, the functions* I^i_1 *and* $\bar{I}^i_1$ *are holomorphic on the domains* $\mathcal{E}^i_\lambda \times \hat{D}_{\mathtt{r}}$*, and satisfy the following estimates:*

$$\begin{aligned}
&\sup_{\mathcal{E}^i_\lambda \times \hat{D}_{\mathtt{r}/2}} |\partial_{\hat{I}} I^i_1| \leqslant \mathbf{c}^2\, \mu_{\mathrm{o}}, \quad \sup_{\mathcal{E}^i_\lambda} \big|\partial_E \bar{I}^i_1\big| \leqslant \mathbf{c}^2\, \frac{|\log \lambda|}{\sqrt{\epsilon}},\\
&\sup_{\mathcal{E}^i_\lambda \times \hat{D}_{\mathtt{r}}} \big|\partial_E I^i_1 - \partial_E \bar{I}^i_1\big| \leqslant \frac{\mathbf{c}^2 \mu}{\lambda\sqrt{\epsilon}}.
\end{aligned} \tag{84}$$

Remark 3.8 *(i) Note that* (75) *implies the hypothesis of Lemma 3.6. Thus, in particular, also* $\mathtt{H}_\flat$ *has* $2N$ *critical points.*

(ii) The identities in (80) *confirm the known analyticity at minima of actions as functions of energy.*

(iii) A formula similar to (76) *is given in [6] (compare Eq. (5.8) in Theorem 5.2 there).*

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

Proposition 3.8 *Assume that, for some* $\theta_0 \in \mathbb{R}$*,* $\bar{\mathtt{G}}$ *satisfies*

$$|\bar{\mathtt{G}}(\theta) - \cos(\theta + \theta_0)|_1 := \sup_{\mathbb{T}_1} |\bar{\mathtt{G}}(\theta) - \cos(\theta + \theta_0)| \leqslant 2^{-40}. \tag{85}$$

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 [16]; compare Proposition 5.12 there.

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

Following [16], we now discuss the Arnol'd–Liouville's action-angle variables for $\mathtt{H}_\flat$ viewed as a n-degree-of-freedom Hamiltonian on the $2n$-dimensional phase space $\check{\mathcal{M}}^i \times \mathbb{T}^{n-1}$.

For every fixed $\hat{p} \equiv \hat{I} \in \hat{D}$, by Arnol'd–Liouville Theorem, 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

$$(p_1, q_1) \to \varphi_1^{(i)}(p_1, q_1; \hat{I}) = \varphi_1^{(i)}(p_1, \hat{I}, q_1).$$

Recall Definition 3.4 and define the normal domains

$$\mathcal{B}^i := \big\{I = (I_1, \hat{I}) \mid \hat{I} \in \hat{D},\ \ I_1^{(i)}(E_-^i(\hat{I}), \hat{I}) < I_1 < I_1^{(i)}(E_+^i(\hat{I}), \hat{I})\big\}. \tag{86}$$

Notice that, for i odd, $I_1^{(i)}(E_-^i(\hat{I}), \hat{I}) = 0$, which is the action of the elliptic point.

Then, recalling (73), one sees that the map

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

is surjective and invertible. Let us denote by

$$\check{\Phi}^i : (I, \varphi_1) \in \mathcal{B}^i \times \mathbb{T} \ \to\ (p, q_1) \in \check{\mathcal{M}}^i, \qquad (\hat{p} = \hat{I}), \tag{87}$$

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

$$\mathtt{H}_\flat \circ \check{\Phi}^i(I, \varphi_1) = \mathtt{E}^{(i)}(I), \qquad dp_1 \wedge dq_1|_{\hat{I}=\text{const}} = dI_1 \wedge d\varphi_1. \tag{88}$$

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

$$\Phi^i(I, \varphi) = \begin{cases} (\eta^i, \hat{I}, \psi^i, \hat{\varphi} + \chi^i), & \text{if } 0 < i < 2N, \\ (\eta^i, \hat{I}, \varphi_1 + \psi^i, \hat{\varphi} + \chi^i), & \text{if } i = 0, 2N, \end{cases} \tag{89}$$

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 (88), one has

$$(\mathtt{H}_\flat \circ \Phi^i)(I, \varphi) = (\mathtt{H}_\flat \circ \check{\Phi}^i)(I, \varphi_1) = \mathtt{E}^{(i)}(I), \qquad \forall\ 0 \leqslant i \leqslant 2N. \tag{90}$$

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

$$\lambda_{\rm max} = \lambda_{\rm max}(\hat{I}) := \big(E_+(\hat{I}) - E_-(\hat{I})\big)/\epsilon, \qquad \bar{\lambda}_{\rm max} := \big(\bar{E}_+ - \bar{E}_-\big)/\epsilon. \tag{91}$$

Notice that, by (45), Definitions 2.2, 3.1, and (44), one has

$$1/\mathfrak{g} \leqslant \beta/\epsilon \leqslant \bar{\lambda}_{\rm max} \leqslant 2. \tag{92}$$

Since $\mu \leqslant 1/\mathbf{c}^2$ and $\mathbf{c} \geqslant 2^8\mathfrak{g}^3$ (compare Theorem 3.7), by (71), one sees that

$$|\lambda_{\rm max} - \bar{\lambda}_{\rm max}| \leqslant 6\mathfrak{g}^3\mu, \qquad \lambda_{\rm max} \geqslant 1/2\mathfrak{g}. \tag{93}$$

Then, recalling the definition of $\mathtt{E}_\flat$ in (65), for $0 \leqslant \lambda \leqslant \lambda_{\rm max}$, we set:

$$\begin{aligned}
& a^i_\lambda(\hat{I}) := I^i_1(E^i_-(\hat{I}) + \lambda\epsilon, \hat{I}), \qquad 0 \leqslant i \leqslant 2N,\\
& b^i_\lambda(\hat{I}) := \begin{cases} I^i_1(E^i_+(\hat{I}) - \lambda\epsilon, \hat{I}), & 0 < i < 2N,\\ I^i_1\big(\mathtt{E}_\flat, \hat{I}\big), & i = 0, 2N, \end{cases}\\
& a^i(\hat{I}) := a^i_0(\hat{I}),\ b^i(\hat{I}) := b^i_0(\hat{I}), \quad 0 \leqslant i \leqslant 2N,\\
& \mathcal{B}^i(\lambda) := \{I = (I_1, \hat{I}) :\ \hat{I} \in \hat{D},\ a^i_\lambda(\hat{I}) < I_1 < b^i_\lambda(\hat{I})\}.
\end{aligned} \tag{94}$$

Remark 3.9 *(i) By the above definitions one has that*

$$a^{2j-1}(\hat{I}) := a^{2j-1}_0(\hat{I}) = I^{2j-1}_1(E^{2j-1}_-(\hat{I}), \hat{I}) \equiv 0, \tag{95}$$

reflecting the analyticity at the elliptic points; compare Remark 3.8-(ii) above.

(ii) By (86) and (94) one sees that $\mathcal{B}^i = \mathcal{B}^i(0) = \bigcup_{0<\lambda<\lambda_{\rm max}} \mathcal{B}^i(\lambda)$.

The holomorphic properties of the Arnol'd–Liouville symplectic maps are described in following theorem, proven in [16, Theorem 4.1]. Recall the definition of the constant $\mathbf{c}$ in Theorem 3.7.

Theorem 3.9 *Under the hypotheses of Theorem 3.7 there exists a constant $\hat{\mathbf{c}} = \hat{\mathbf{c}}(n, \mathfrak{g}) \geqslant 4\,\mathbf{c}^2$ depending only on n and $\mathfrak{g}$ such that, taking*

$$\mu \leqslant 1/\hat{\mathbf{c}}, \tag{96}$$

the symplectic transformation $\check{\Phi}^i$ *in* (87) *extends, for any* $0 \leqslant i \leqslant 2N$ *and* $0 < \lambda \leqslant 1/\hat{\mathbf{c}}$, *to a real analytic map*

$$\Phi^i : \big(\mathcal{B}^i(\lambda)\big)_{\rho_\lambda} \times \mathbb{T}^n_{\sigma_\lambda} \ \to\ D_{\mathtt{r}} \times \mathbb{T}^n_{\mathtt{s}/4}, \qquad \forall\, 0 < \lambda \leqslant 1/\hat{\mathbf{c}}, \tag{97}$$

where

$$\rho_\lambda := \frac{\sqrt{\epsilon}}{\hat{\mathbf{c}}}\,\lambda|\log\lambda|, \qquad \sigma_\lambda := \frac{1}{\hat{\mathbf{c}}|\log\lambda|}. \tag{98}$$

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*

$$\begin{aligned} &\big|\partial_{I_1}\mathtt{E}^i\big| \leqslant \hat{\mathbf{c}}\sqrt{\epsilon + |\mathtt{E}^i|}, \qquad \big|\partial^2_{I_1}\mathtt{E}^i\big| \leqslant \frac{\hat{\mathbf{c}}}{\hat{\lambda}},\\ &\big|\partial^2_{I_1\hat{I}}\mathtt{E}^i\big| \leqslant \hat{\mathbf{c}}\frac{\mu_o}{\hat{\lambda}}, \qquad \big|\partial^2_{\hat{I}}\mathtt{E}^i\big| \leqslant \hat{\mathbf{c}}\Big(\frac{\sqrt{\epsilon}}{\mathtt{r}}I_1^i + \frac{\mu_o}{\hat{\lambda}}\Big)\mu_o. \end{aligned} \tag{99}$$

Furthermore, defining

$$D^\flat := (-\mathtt{R} - \mathtt{r}/3, \mathtt{R} + \mathtt{r}/3) \times \hat{D}, \qquad \check{\mathcal{M}}^i(\lambda) := \check{\Phi}^i(\mathcal{B}^i(\lambda) \times \mathbb{T}), \tag{100}$$

one has

$$\mathrm{meas}\,\Big((D^\flat \times \mathbb{T}) \setminus \bigcup_{0\leqslant i\leqslant 2N} \check{\mathcal{M}}^i(\lambda)\Big) \leqslant \hat{\mathbf{c}}\,\sqrt{\epsilon}\,\mathrm{meas}(\hat{D})\ \lambda|\log\lambda|. \tag{101}$$

Remark 3.10 *Recall that* $\varepsilon < 1$ *(see* (1)*), and observe that, by* (49), (50), (20), (75), (32), (53) *and* (56), *it is*

$$1/\mathfrak{g} < \check{\mathtt{s}}/4, \qquad \frac{\varepsilon_k \chi_k}{\mathtt{r}}\,\mu < \mathtt{r}/6, \qquad \frac{4\varepsilon_k\chi_k}{\mathtt{r}^2}\,\mu < \frac{\check{\mathtt{s}}}{2^{20}\mathfrak{g}^3} < \check{\mathtt{s}}/2^{20}. \tag{102}$$

Since (as assumed in Theorem 3.9) $\lambda \leqslant 1/\hat{\mathbf{c}}$, *by* (75), *one sees that* σ_λ *in* (98) *satisfies*

$$\sigma_\lambda < \check{\mathtt{s}}/2^{20}. \tag{103}$$

4 Secondary nearly integrable structure at simple resonances

We are now ready to discuss the global nearly integrable structure of the Hamiltonians $\mathcal{H}_k(\mathtt{y}, \mathtt{x})$ in (37) near simple resonances.

As mentioned above (see beginning of § 3.2 and item (ii) in Remark 3.6), the problem here is that the symplectic transformations of Theorem 3.4, 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 problem, we shall exploit the particular form 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, (124) below.

Special sets of symplectic transformations

Besides the group $\mathfrak{G}_\dagger$ introduced in Definition 3.2 above, we shall introduce two new special classes of symplectic transformations, which will be used in the proof of Theorem 4.1. Recall the notation in (46).

Definition 4.1 *Fix a domain* $\hat{\mathrm{D}} \subseteq \mathbb{R}^{n-1}$*. Then,*

$\mathfrak{G}$ *denotes the set 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 over* $\hat{\mathrm{D}}$ *(i.e., D has the form* $\{(p_1, \hat{p}) : \alpha(\hat{p}) < p_1 < \beta(\hat{p}),\ \hat{p} \in \hat{\mathrm{D}}\}$*); the functions* η, ψ, χ *depend on* (p, q_1)*, are* 2π*-periodic in* q_1*; the* $(n+1)$*-dimensional the map*

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

is injective;

$\mathfrak{G}_0$ *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. 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 4.1 *(i) The Arnol'd–Liouville map* Φ^i *in the outer cases* (89) *(*$i = 0, 2N$*) belongs to* $\mathfrak{G}$ *(since* $\sup|\partial_{q_1}\psi| < 1$*), while* Φ^i *in the inner case* (89) *(*$0 < i < 2N$*) belongs to* $\mathfrak{G}_0$*. Notice also that* Φ_2 *in Theorem 3.4-(iii) is a close 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*

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

(iii) All maps in the group $\mathfrak{G}_\dagger$ *in Definition 3.2 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 *(but* $\pi_{\hat{p}}(\mathrm{D}) = \hat{\mathrm{D}}$ *is, instead, fixed). 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.*

(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,*

$$q \in \mathbb{T}^n \to Q = (q_1 + \psi, \hat{q} + \chi) \in \mathbb{T}^n$$

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*

$$\Psi_i \in \mathfrak{G}_\dagger \cup \mathfrak{G} \cup \mathfrak{G}_0 \quad \Longrightarrow \quad (\Psi_1 \circ \cdots \circ \Psi_m)\check{} = (\check{\Psi}_1 \circ \cdots \circ \check{\Psi}_m). \tag{104}$$

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

$$\Phi \in \mathfrak{G}_0 \quad \text{and} \quad \Psi \in \mathfrak{G}_\dagger \cup \mathfrak{G} \qquad \Longrightarrow \qquad \Psi \circ \Phi \in \mathfrak{G}_0. \tag{105}$$

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

For each $k \in \mathcal{G}^n_{\mathtt{K}_o}$, we may apply the theory of § 3.3 to the secular Hamiltonians described in Theorem 3.4 in standard form $\mathtt{H}_\flat = \mathtt{H}_k$; see (54), (55), and (56).

By (90), we get that, for every $k \in \mathcal{G}^n_{\mathtt{K}_o}$ and $0 \leqslant i \leqslant 2N_k$, the Arnol'd–Liouville map

$$\Phi^i \ : \ \mathcal{B}^i_k \times \mathbb{T}^n \xrightarrow{\text{onto}} \check{\mathcal{M}}^i_k \times \mathbb{T}^{n-1} \tag{106}$$

integrates $\mathtt{H}_k$, i.e., for every $0 \leqslant i \leqslant 2N_k$,

$$(\mathtt{H}_k \circ \Phi^i)(I, \varphi) = (\mathtt{H}_k \circ \check{\Phi}^i)(I, \varphi_1) = \mathtt{E}^{(i)}_k(I), \tag{107}$$

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 § 3.3 in the case $\mathtt{H}_\flat = \mathtt{H}_k$; compare, in particular, (72) and (73) for the definitions of $\mathcal{M}^i_k$ and $\check{\mathcal{M}}^i_k$, (86) for the definition of $\mathcal{B}^i_k$, (94) and (91) for the definition of $\mathcal{B}^i_k(\lambda)$, (100) for the definition of $\check{\mathcal{M}}^i_k(\lambda)$.

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

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

$$\jmath_\mathtt{g}(p) := (p_1 + \mathtt{g}(\hat{p}), \hat{p}). \tag{108}$$

Notice that, by the definition of $\Psi_\mathtt{g}$ in (47), one has

$$\check{\Psi}_\mathtt{g}(p, q_1) = (\jmath_\mathtt{g}(p), q_1). \tag{109}$$

Global action-angle variables at simple resonances

We are now ready to state and prove the first step of the proof of Theorem 2.1, which consists in showing how to construct symplectic action-angle maps, which put a generic nearly integrable mechanical system near simple resonances, *for all* $k \in \mathcal{G}^n_{\mathtt{K}_o}$, into uniform analytic nearly integrable form *with exponentially small perturbations*.

Let Assumption 3.1 and Definition 3.3 hold; let $\mathbf{c}_\mathbf{o}$ be as in Theorem 3.3, and let $\hat{\mathbf{c}}$ be as in Theorem 3.9 with $\mathfrak{g}$ as in (56). Let $\mathtt{g}_{\scriptscriptstyle 1}$ and $\mathtt{g}_{\scriptscriptstyle 3}$ be as in (ii) of Theorem 3.4, and define

$$B^i_k := \begin{cases} \mathcal{B}^i_k, & \text{if } 0 < i < 2N_k, \\ \jmath_{-\mathtt{g}_\star}(\mathcal{B}^i_k), & \text{if } i = 0, 2N_k. \end{cases} \qquad \mathtt{g}_\star := -(\mathtt{g}_{\scriptscriptstyle 1} + \mathtt{g}_{\scriptscriptstyle 3}). \tag{110}$$

Then, the following result holds (recall the notation (31), (22) and (35)).

Theorem 4.1 (Secondary nearly integrable structure at simple resonances) *There exists* $\mathbf{c}_\star = \mathbf{c}_\star(n, s, \beta, \delta) \geqslant \max\{c_2, \mathbf{c}_0, \hat{\mathbf{c}}\}$ *such that if* $\mathtt{K}_o \geqslant \mathbf{c}_\star$, *then for any* $k \in \mathcal{G}^n_{\mathtt{K}_o}$, $0 \leqslant i \leqslant 2N_k$, *there exist real analytic symplectomorphisms*

$$\phi^i_k : B^i_k \times \mathbb{T}^n \to \mathrm{Re}(\mathcal{R}^{1,k}_{r_k}) \times \mathbb{T}^n, \tag{111}$$

such that, if $\mathtt{E}^i_k = \mathtt{E}^i_k(I)$ *is the integrable Hamiltonian* $\mathtt{H}_k$ *of Theorem 3.4 in its Arnol'd–Liouville action variables,* $\widetilde{\mathtt{E}}^i_k := \mathtt{E}^i_k \circ j_{\mathtt{g}_\star}$, *and* $\hat{h}_k$ *is as in Theorem 3.4, then*

$$\begin{aligned} &\mathcal{H}^i_k := \mathtt{H} \circ \phi^i_k(I, \varphi) = h^i_k(I) + \varepsilon f^i_k(I, \varphi), \quad \textit{with:} \\ &h^i_k := \tfrac{|k|^2}{2}\mathtt{h}^i_k, \ \mathtt{h}^i_k := \begin{cases} \mathtt{E}^i_k + \hat{h}_k, & \textit{if } 0 < i < 2N_k, \\ \widetilde{\mathtt{E}}^i_k + \hat{h}_k, & \textit{if } i = 0, 2N_k. \end{cases} \end{aligned} \tag{112}$$

Furthermore, if $0 < \lambda \leqslant 1/\mathbf{c}_\star$ *and if we define*

$$\begin{aligned} &\rho_\star := \frac{\sqrt{\epsilon}}{\mathbf{c}_\star \mathtt{K}^n_o} \lambda |\log \lambda|, \qquad \sigma_\star := \frac{1}{\mathbf{c}_\star \mathtt{K}^n_o |\log \lambda|}, \\ &B^i_k(\lambda) := \begin{cases} \mathcal{B}^i_k(\lambda), & \textit{if } 0 < i < 2N_k, \\ j_{-\mathtt{g}_\star}\big(\mathcal{B}^i_k(\lambda)\big), & \textit{if } i = 0, 2N_k, \end{cases} \end{aligned} \tag{113}$$

then ϕ^i_k *admits a holomorphic extension*

$$\phi^i_k : (B^i_k(\lambda))_{\rho_\star} \times \mathbb{T}^n_{\sigma_\star} \to \mathcal{R}^{1,k}_{r_k} \times \mathbb{T}^n_{s_\star} \tag{114}$$

and the perturbation f^i_k *in* (112) *satisfies the exponential estimate*

$$\sup_{(B^i_k(\lambda))_{\rho_\star} \times \mathbb{T}^n_{\sigma_\star}} |f^i_k| \leqslant e^{-\mathtt{K}s/3}. \tag{115}$$

Remark 4.2 *(i) Recall* (49) *and the definition of* $\mathbf{c}$ *in Theorem 3.7. Now, notice that, since* $\mu = 1/\mathtt{K}^{5n}$, *and*

$$\mathtt{K} > \mathtt{K}_o \geqslant \mathbf{c}_\star > \mathbf{c},$$

condition (96), *which is stronger than condition* (75), *is implied by the assumption* $\mathtt{K}_o \geqslant \mathbf{c}_\star$. *Observe also that from the definitions of the constants in Theorem 4.1 and Theorem 3.7, and from* (93), *it follows that*

$$\mathbf{c}_\star \geqslant \mathbf{c} \geqslant 2^8 \mathfrak{g}^3 \geqslant 2^{14}, \qquad \lambda_{\max} \geqslant 2^{12}/\mathbf{c}_\star. \tag{116}$$

Finally, we remark that, since $\mathbf{c}_\star \geqslant \hat{\mathbf{c}}$, *one has (recall* (98)*)*

$$\rho_\star < \rho_\lambda, \qquad \sigma_\star < \sigma_\lambda. \tag{117}$$

(ii) In the proof of the theorem the maps ϕ^i_k *are explicitly given; compare* (122) *and* (128) *below.*

The following simple lemma will be one of the key points of the proof of Theorem 4.1. Recall Definition 4.1.

Lemma 4.2 *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* $\tau_{\mathrm{g}}\Phi$ *the map*

$$\tau_{\mathrm{g}}\Phi := \tau_{\mathrm{g}}\Phi(p, q) := \big(\eta_{\mathrm{g}} + \mathrm{g}, \hat{p}, q_1 + \psi_{\mathrm{g}}, \hat{q} + \chi_{\mathrm{g}} - \psi_{\mathrm{g}}\partial_{\hat{p}}\mathrm{g}\big), \tag{118}$$

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

$$u_{\mathrm{g}} := u \circ \check{\Psi}_{-g} : j_{\mathrm{g}}(D) \times \mathbb{T} \to \mathbb{R}^m. \tag{119}$$

Then, $\tau_{\mathrm{g}}\Phi$ *belongs to* $\mathfrak{G}$ *and satisfies*

$$\tau_{\mathrm{g}}\Phi : j_{\mathrm{g}}(\mathrm{D}) \times \mathbb{T}^n \overset{\text{onto}}{\longrightarrow} \big(\check{\Psi}_{\mathrm{g}} \circ \Phi(\mathrm{D} \times \mathbb{T}^n)\big) \times \mathbb{T}^{n-1}, \tag{120}$$

and

$$\big(\tau_{\mathrm{g}}\Phi\big)\check{} = \big(\eta_{\mathrm{g}} + \mathrm{g}, \hat{p}, q_1 + \psi_{\mathrm{g}}\big) = \check{\Psi}_{\mathrm{g}} \circ \check{\Phi} \circ \check{\Psi}_{-\mathrm{g}}. \tag{121}$$

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

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

is a well defined $\mathbb{T}^n$-map, so that (121) follows immediately by direct computation. Thus, $(\tau_{\mathrm{g}}\Phi)\check{}$ is injective being the composition of three injective maps, and, therefore, the whole map $\tau_{\mathrm{g}}\Phi$ is injective, and (120) follows. To check symplecticity, just note that, locally, on the universal cover $\mathbb{R}^{2n}$, $\tau_{\mathrm{g}}\Phi$ coincides (as it is immediate to check) with the composition $\Psi_{\mathrm{g}} \circ \Phi \circ \Psi_{-\mathrm{g}}$ of three symplectic maps. Hence $\tau_{\mathrm{g}}\Phi$ is symplectic and the claim follows. ∎

Proof of Theorem 4.1 We start by defining the maps ϕ_k^i. Consider, first, the inner case $0 < i < 2N_k$. Recall Definition 4.1. By Theorem 3.4-(iii), $\Phi_\star$ is the composition of maps in $\mathfrak{G}_{\dagger}$ and $\mathfrak{G}$ while $\Phi^i \in \mathfrak{G}_0$ (Remark 4.1-(i)). Hence, by (105), it follows that $\Phi_\star \circ \Phi^i \in \mathfrak{G}_0$ and we may define

$$\Phi_\star^i := \Phi_\star \circ \Phi^i, \quad \phi_k^i := \Psi^k \circ \Phi_\star^i : B_k^i \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n, \qquad 0 < i < 2N_k, \tag{122}$$

provided the composition is well defined; here Ψ^k is the transformation appearing in Theorem 3.3, and recall that, when $0 < i < 2N_k$, $B_k^i := \mathcal{B}_k^i$. To check that (122) is well posed, we observe that by (104), (97), (52), (53), (117), for $0 < \lambda \leqslant 1/\hat{\mathbf{c}}$, we get, for $0 < i < 2N_k$,

$$\check{\Phi}^i_\star = (\Phi_\star \circ \Phi^i)^\vee = \check{\Phi}_\star \circ \check{\Phi}^i : (\mathcal{B}^i_k(\lambda))_{\rho_\lambda} \times \mathbb{T}_{\sigma_\lambda} \to \mathcal{D}^k_{\tilde{r}_k} \times \mathbb{T}_{\check{\mathtt{s}}}. \tag{123}$$

Thus the composition is indeed well defined and (122) is well posed.

Let us turn the outer case $i = 0, 2N_k$. In this case $\Phi^i \in \mathfrak{G}$ (Remark 4.1-(i)). Recalling the definition in (118), (119), by Lemma 4.2, we may define

$$\Phi^i_{23} := \Phi_2 \circ \tau_{\mathtt{g}_3}\Phi^i, \qquad \text{and} \qquad \Phi^i_\star := \tau_{\mathtt{g}_1}\Phi^i_{23}. \tag{124}$$

Recalling that $\Phi_2 \in \mathfrak{G}$, by Lemma 4.2 and Remark 4.1-(iv), $\Phi^i_{23} \in \mathfrak{G}$ and, again by Lemma 4.2, $\Phi^i_\star \in \mathfrak{G}$, provided the compositions are well defined. To check that this is the case, as above, it is enough to control the complex domains of the first $(n+1)$ components. By (121) (used twice), (48), (57), and (104), one finds

$$\check{\Phi}^i_\star = \check{\Phi}_\star \circ \check{\Phi}^i \circ \check{\Psi}_{\mathtt{g}_\star}. \tag{125}$$

where $\mathtt{g}_\star = -(\mathtt{g}_1 + \mathtt{g}_3)$ is defined in (110). Then, by (144), we get,

$$j_{\mathtt{g}_\star}\big((B^i_k(\lambda))_{\rho'_\lambda}\big) \subseteq \big(j_{\mathtt{g}_\star}\big(B^i_k(\lambda)\big)\big)_{\rho_\lambda} \overset{(113)}{=} \big(\mathcal{B}^i_k(\lambda)\big)_{\rho_\lambda}, \qquad i = 0, 2N_k, \tag{126}$$

with $\rho'_\lambda := \rho_\lambda/(n+2)$. Observing that $\check{\Psi}_{\mathtt{g}_\star}(p, q_1) = \big(j_{\mathtt{g}_\star}(p), q_1\big)$, by (125), (126), (97), (52) and (53), we get, for $0 < \lambda \leqslant 1/\hat{\mathbf{c}}$,

$$\check{\Phi}^i_\star : (B^i_k(\lambda))_{\rho'_\lambda} \times \mathbb{T}_{\sigma_\lambda} \to \mathcal{D}^k_{\tilde{r}_k} \times \mathbb{T}_{\check{\mathtt{s}}}, \qquad i = 0, 2N_k. \tag{127}$$

Thus, the composition is well defined and (124) is well posed. So, we may define:

$$\phi^i_k := \Psi^k \circ \Phi^i_\star : B^i_k \times \mathbb{T}^n \to \mathbb{R}^n \times \mathbb{T}^n, \quad \Phi^i_\star \text{ as in (124)}, \qquad i = 0, 2N_k. \tag{128}$$

We can, now, prove (112). Recall the definition of $\bar{f}^k$ in Theorem 3.3 and define, for all $0 \leqslant i \leqslant 2N_k$,

$$f^i_k := f \circ \phi^i_k \overset{(37)}{=} \bar{f}^k \circ \Phi^i_\star. \tag{129}$$

Then, by definition of ϕ^i_k in (122) and (128), we have, for $0 \leqslant i \leqslant 2N_k$,

$$\mathcal{H}^i_k := \mathtt{H} \circ \phi^i_k(I, \varphi) := \mathtt{H} \circ \Psi^k \circ \Phi^i_\star \overset{(37,129)}{=} \overline{\mathtt{H}}_k \circ \Phi^i_\star + \varepsilon f^i_k. \tag{130}$$

Since $\overline{\mathtt{H}}_k$ in (38) depends only on the first $(n+1)$ variables, by (123) and (125), we find

$$\overline{\mathtt{H}}_k \circ \Phi^i_\star = \overline{\mathtt{H}}_k \circ \check{\Phi}^i_\star = \begin{cases} \overline{\mathtt{H}}_k \circ \check{\Phi}_\star \circ \check{\Phi}^i, & \text{if } 0 < i < 2N_k, \\ \overline{\mathtt{H}}_k \circ \check{\Phi}_\star \circ \check{\Phi}^i \circ \check{\Psi}_{\mathtt{g}_\star}, & \text{if } i = 0, 2N_k, \end{cases} \tag{131}$$

and, by (53) and (107),

$$\overline{\mathtt{H}}_k \circ \check{\Phi}_\star \circ \check{\Phi}^i = \tfrac{|k|^2}{2}(\mathtt{E}_k^{(i)} + \hat{h}_{_k}). \tag{132}$$

Thus, (112) follows from (130), (131), (132) and (109).

Next, we show that ϕ_k^i has, for $0 < \lambda \leqslant 1/\mathbf{c}_{\star}$, a holomorphic extension satisfying (114). To do this we have to consider the last $n-1$ components of $\Phi_\star^i$, namely$\pi_{\hat{\varphi}} \Phi_\star^i = \hat{\Phi}_\star^i$ (recall the notation introduced in (46)). By definition of $\Phi_\star^i$ in (122) and (124), recalling (89), (110), and Theorem 3.4-(iii), one finds

$$\hat{\Phi}_\star^i(I,\varphi) = \begin{cases} \hat{\varphi} + \chi_\star^i(I,\varphi_1), & \text{if } 0 < i < 2N_k,\\ \hat{\varphi} + \chi_\star^i(j_{\mathtt{g}_\star}(I),\varphi_1), & \text{if } i = 0, 2N_k, \end{cases} \tag{133}$$

with

$$\begin{cases} \chi_\star^i := \chi^i + \chi_{_2}^\flat + \psi^i\,\partial_{\hat{I}}\mathtt{g}_\star,\\ \chi_{_2}^\flat(I,\varphi_1) := \begin{cases} \chi_{_2}(\hat{I},\psi^i), & \text{if } 0 < i < 2N_k,\\ \chi_{_2}(\hat{I},\varphi_1+\psi^i), & \text{if } i = 0, 2N_k. \end{cases} \end{cases} \tag{134}$$

Now, we claim that, for all $0 \leqslant i \leqslant 2N_k$,

$$|\psi^i|_{\rho_\lambda,\sigma_\lambda} < \frac{3}{4}\check{\mathtt{s}}. \tag{135}$$

Indeed, if $0 < i < 2N_k$, (135) follows directly from (97) and (52); in the case $i = 0, 2N_k$, (135) follows again from (97) and (52) 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 (126), (134), (97), (52), (58), (102), (103), (135), (144), we find, for every $0 \leqslant i \leqslant 2N_k$, and for every $2 \leqslant \ell \leqslant n$,

$$\begin{aligned} |\operatorname{Im}\hat{\Phi}_{\star\ell}^i|_{\rho'_\lambda,\sigma_\lambda} &\leqslant |\operatorname{Im}(\varphi_\ell + \chi_{\star\ell}^i)|_{\rho_\lambda,\sigma_\lambda}\\ &\leqslant |\operatorname{Im}(\varphi_\ell + \chi_\ell^i)|_{\rho_\lambda,\sigma_\lambda} + |\chi_{_2}^\flat|_{\rho_\lambda,\sigma_\lambda} + |\psi^i|_{\rho_\lambda,\sigma_\lambda}\,|\partial_{\hat{I}}\mathtt{g}_\star|_{\rho_\lambda}\\ &\leqslant \frac{\check{\mathtt{s}}}{2} + \frac{\check{\mathtt{s}}}{2^{20}} + \frac{3}{4}(n+1)\check{\mathtt{s}} < 2n\check{\mathtt{s}}. \end{aligned} \tag{136}$$

Thus, by (123), (127) and (136), we get, for all $0 \leqslant i \leqslant 2N_k$,

$$\Phi_\star^i : (B_k^i(\lambda))_{\rho'_\lambda} \times \mathbb{T}^n_{\sigma_\lambda} \ \to\ \mathcal{D}^k_{\tilde{r}_k} \times \mathbb{T}^n_{2n\check{\mathtt{s}}}.$$

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

Lemma 4.3 *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

$$\zeta := \frac{1}{16n\, c_1\, \mathtt{c}_s\, \mathtt{K}_{\rm o}^n}. \tag{137}$$

Then, since $|k| \leqslant \mathtt{K}_{\rm o}$, by (53) and (49), we find

$$8\zeta(2n\check{\mathtt{s}}) < 16n\,\zeta\,\mathtt{K}_{\rm o}\max\{1,s\} \overset{(137)}{=} \frac{\max\{1,s\}}{c_1\ \mathtt{c}_s\,\mathtt{K}_{\rm o}^{n-1}} = \frac{s}{c_1\mathtt{K}_{\rm o}^{n-1}} \overset{(32)}{=} \tilde{s}_k.$$

Thus, by Lemma 4.3 (applied with $g = \hat{\Phi}^i_{\star\ell}$ for $2 \leqslant \ell \leqslant n$, ζ as in (137), and $\xi = 2n\check{\mathtt{s}}$), it follows, for all $0 \leqslant i \leqslant 2N_k$, that

$$\Phi^i_\star :\ (B^i_k(\lambda))_{\rho_\star} \times \mathbb{T}^n_{\sigma_\star} \ \to\ \mathcal{D}^k_{\tilde{r}_k} \times \mathbb{T}^n_{\tilde{s}_k},$$

with $\rho_\star$ and $\sigma_\star$ as in (113), provided

$$\mathbf{c}_\star := \max\{c_2, \mathbf{c_0},\ \hat{\mathbf{c}}\, c_1\ \mathtt{c}_s\, 16n\,(n+2)\}.$$

In conclusion, (114) follows by the definition of ϕ^i_k in (122), (128) and by (35).

Finally, estimate (115) follows at once from (129), (114) and (39). The proof of Theorem 4.1 is complete. ∎

Theorem 4.1 provides holomorphic action-angle variables in the phase space of $\mathtt{H}$ in a domain λ-away from separatrices. Next result gives a measure estimate of the portion of space space left out. Such measure estimate will play a crucial role in the proof of Theorem 2.1. Recall the definition of the sets $\widetilde{\mathcal{R}}^{1,k}$ given in (34).

Proposition 4.4 *For every $0 \leqslant \lambda < 1/\mathbf{c}_\star$, the following measure estimate holds:*

$$\operatorname{meas}\big((\mathcal{R}^{1,k}\times\mathbb{T}^n)\setminus \textstyle\bigcup_i \phi^i_k\big(B^i_k(\lambda)\times\mathbb{T}^n\big)\big) \leqslant \mathbf{c}_\star \operatorname{meas}\big(\widetilde{\mathcal{R}}^{1,k}\times\mathbb{T}^n\big)\,\lambda|\log\lambda|, \tag{138}$$

where the union runs from $i = 0$ up to $i = 2N_k$.

Proof Since $\check{\Phi}^i_\star$ depends only on the first $(n+1)$ variables, by (133), (122), (125), and the definitions of $B^i_k(\lambda)$ in (113) and $\check{\mathcal{M}}^i_k(\lambda)$ in (106), one has

$$\begin{aligned}\Phi^i_\star\big(B^i_k(\lambda)\times\mathbb{T}^n\big) &= \check{\Phi}^i_\star\big(B^i_k(\lambda)\times\mathbb{T}\big)\times\mathbb{T}^{n-1}\\ &= \big(\check{\Phi}_\star\circ\check{\Phi}^i(\mathcal{B}^i_k(\lambda)\times\mathbb{T})\big)\times\mathbb{T}^{n-1}\\ &\overset{(101)}{=} \big(\check{\Phi}_\star\circ\check{\mathcal{M}}^i_k(\lambda)\big)\times\mathbb{T}^{n-1}.\end{aligned} \tag{139}$$

Analogously, one has

$$\Phi_\star^{-1}(\mathcal{D}^k \times \mathbb{T}^n) = \check{\Phi}_\star^{-1}(\mathcal{D}^k \times \mathbb{T}) \times \mathbb{T}^{n-1}. \tag{140}$$

Since $\check{\Phi}_3^{-1}(p,q) = (p_1 - \mathtt{g}_3(\hat{p}), \hat{p}, q_1)$ and $\check{\Phi}_2^{-1}(p,q) = (p_1 - \eta_2(\hat{p}, q_1), \hat{p}, q_1)$, by (59), (58), and (the second estimate in) (102), it follows that

$$\check{\Phi}_3^{-1} \circ \check{\Phi}_2^{-1}(D \times \mathbb{T}) \subseteq \big((-\mathtt{R} - \mathtt{r}/3, \mathtt{R} + \mathtt{r}/3) \times \hat{D}\big) \times \mathbb{T} = (D^\flat \times \mathbb{T}), \tag{141}$$

where $D^\flat$ is defined in (100). Then, recalling Theorem 3.3, using the fact that $(\Psi^k)^{-1}$ and $\Phi_\star^{-1}$ are diffeomorphisms preserving Liouville measure, we find (the index i running from 0 to $2N_k$):

$$\begin{aligned}
&\operatorname{meas}(\mathcal{R}^{1,k} \times \mathbb{T}^n \setminus \textstyle\bigcup_i \phi_k^i(B_k^i(\lambda) \times \mathbb{T}^n))\\
&\overset{(122,128)}{=} \operatorname{meas}((\Psi^k)^{-1}(\mathcal{R}^{1,k} \times \mathbb{T}^n) \setminus \textstyle\bigcup_i \Phi_\star^i(B_k^i(\lambda) \times \mathbb{T}^n))\\
&\overset{(41)}{\leqslant} \operatorname{meas}((\mathcal{D}^k \times \mathbb{T}^n) \setminus \textstyle\bigcup_i \Phi_\star^i(B_k^i(\lambda) \times \mathbb{T}^n))\\
&= \operatorname{meas}(\Phi_\star^{-1}(\mathcal{D}^k \times \mathbb{T}^n) \setminus \textstyle\bigcup_i \Phi_\star^{-1}\Phi_\star^i(B_k^i(\lambda) \times \mathbb{T}^n))\\
&\overset{(139,140)}{=} (2\pi)^{n-1} \operatorname{meas}(\check{\Phi}_\star^{-1}(\mathcal{D}^k \times \mathbb{T}) \setminus \textstyle\bigcup_i \check{\mathcal{M}}_k^i(\lambda))\\
&\overset{(57,63)}{=} (2\pi)^{n-1} \operatorname{meas}(\check{\Phi}_3^{-1} \circ \check{\Phi}_2^{-1}(D \times \mathbb{T}) \setminus \textstyle\bigcup_i \check{\mathcal{M}}_k^i(\lambda))\\
&\overset{(141)}{\leqslant} (2\pi)^{n-1} \operatorname{meas}(D^\flat \times \mathbb{T} \setminus \textstyle\bigcup_i \check{\mathcal{M}}_k^i(\lambda))\\
&\overset{(101)}{\leqslant} (2\pi)^{n-1} \hat{\mathbf{c}} \sqrt{\epsilon} \operatorname{meas}(\hat{D})\ \lambda |\log \lambda|\\
&\overset{(51)}{<} (2\pi)^{n-1} \hat{\mathbf{c}}\ \mathtt{R} \operatorname{meas}(\hat{D})\ \lambda |\log \lambda|\\
&\overset{(62)}{=} \frac{\hat{\mathbf{c}}}{2\pi} \operatorname{meas}(\widetilde{\mathcal{R}}^{1,k} \times \mathbb{T}^n)\ \lambda |\log \lambda|,
\end{aligned}$$

which, since $\mathbf{c}_\star \geqslant \hat{\mathbf{c}}$, yields (138). ∎

Remark 4.3 *(i) The measure estimate (138) holds in view of the covering property (41), which takes care of the deformation near the boundaries. Note that $|\log \lambda|$ here has no direct connection with $|\log \varepsilon|$ appearing in Corollary 2.2; rather, it is related to the Lyapunov exponents of the hyperbolic equilibria of the secondary integrable systems at simple resonances.*

(ii) The KAM invariant tori for $\mathtt{H}$ will be found in the phase spaces $B_k^i(\lambda) \times \mathbb{T}^n$. Now, in order to obtain Corollary 2.2, one is allowed to disregard a neighborhood of

separatrices of measure at most $\sim \varepsilon$ (up to a logarithmic correction). In view of (138), since meas $\big(\widetilde{\mathcal{R}}^{1,k} \times \mathbb{T}^n\big) \sim \sqrt{\varepsilon}$*, we see that* λ *has to be taken, at most* $\sqrt{\varepsilon}$*, corresponding to a distance in energy from the separatrices of at least* $\sim \varepsilon\lambda \sim \varepsilon^{3/2}$*; for the actual choice of* λ*, see (237) below.*

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 necessary to control such a growth.

Proposition 4.5 *Assume that $\alpha < 1$. Then, there exists a constant $c_* = c_*(n) > 1$ such that*

$$\mathrm{diam}\, B^i_k \leqslant c_* |k|^{n-1}, \qquad \mathrm{meas}\, B^i_k \leqslant c_*. \tag{142}$$

Proof For the purpose of this proof, we denote by "c" suitable (possibly different) constants greater than one and depending only on n. First, we observe that, since $\gamma = 2(\nu + n)$, the hypothesis $\alpha = \varepsilon \mathtt{K}^\nu < 1$ is implied by the second condition in (11). Now, since $\alpha < 1$, by the definition of $\mathcal{B}^i_k$ in (86), by (45), and the definition of $\mathtt{R}$ in (49), we have, for every $0 \leqslant i \leqslant 2N_k$,

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

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

$$|\pi^\perp_k \hat{\mathrm{A}}^T \hat{I}| = \Big|\mathrm{A}^T \mathrm{U} \binom{0}{\hat{I}}\Big| \geqslant \frac{|\hat{I}|}{\|\mathrm{A}^{-1}\|\,\|\mathrm{U}^{-1}\|} \geqslant \frac{|\hat{I}|}{c|k|^{n-1}}.$$

Now, since, by (49),

$$\hat{D} \subseteq \{\hat{I} \in \mathbb{R}^{n-1} : \ |\pi^\perp_k \hat{\mathrm{A}}^T \hat{I}| < 1\},$$

it follows that $\mathrm{diam}\, \mathcal{B}^i_k \leqslant c|k|^{n-1}$, proving the first relation in (142) in the case $0 < i < 2N_k$.

In the case $i = 0, 2N_k$, we need to estimate the Lipschitz constant of $\mathtt{g}_\star$ in (110). The map $\mathtt{g}_1$ is linear and its gradient is given by $\hat{\mathrm{A}}k/|k|^2$, thus, by (33) one gets

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

By (59), the definitions in (49), and by Cauchy estimates (compare, e.g., [18]), one sees that

$$|\mathtt{g}_3|_{4\mathtt{r}} < \frac{2\varepsilon}{|k|^2} \frac{\mu}{\mathtt{r}} < \frac{c_2 \mathbf{c_o}}{2} \frac{\sqrt{\varepsilon}}{\mathtt{K}^{14n+5}}, \qquad |\partial_{\hat{I}} \mathtt{g}_3|_{3\mathtt{r}} \leqslant \frac{2\varepsilon\mu}{|k|^2 \mathtt{r}^2} \leqslant c\, \frac{\mathbf{c_o}}{\mathtt{K}^{14n+2}} < \frac{1}{4}, \tag{143}$$

by taking $\mathtt{K_o}$ big enough (recall that $\mathtt{K} \geqslant 6\mathtt{K_o}$). Hence, denoting by $\mathrm{Lip}_B(g)$ the absolute value of the Lipschitz constant of a function g over a domain B, one finds

$$|\partial_{\hat{I}} \mathtt{g}_\star|_{3\mathrm{r}} \leqslant n+1, \qquad \mathrm{Lip}_{D_{3\mathrm{r}}}(j_{\mathtt{g}_\star}) \leqslant n+2. \tag{144}$$

Choosing c_* suitably, the first relation in (142) follows also in this case. Let us check the second relation in (142). Since ϕ^i_k in (111) is symplectic, we have

$$\mathrm{meas}\, B^i_k = \frac{1}{(2\pi)^n}\,\mathrm{meas}(B^i_k \times \mathbb{T}^n) = \frac{1}{(2\pi)^n}\,\mathrm{meas}\,\big(\phi^i_k(B^i_k \times \mathbb{T}^n)\big)$$
$$\overset{(111)}{\leqslant} \mathrm{meas}\,\big(\,\mathrm{Re}(\mathcal{R}^{1,k}_{r_k})\big).$$

Now, since $\mathcal{R}^{1,k} \subseteq \mathrm{B}$ and $r_k \leqslant \alpha < 1$, choosing c_* suitably, also the second relation in (142) follows. ∎

5 Twist at simple resonances

In this section – which is the heart of the paper – we discuss the twist of the integrable (rescaled) secular Hamiltonians $\mathtt{h}^i_k$ in (112) near simple resonances and, in particular, in neighborhoods of secular separatrices, where the actions become singular.

In general, it has to be expected that *there are points where the twist of the secular Hamiltonians* $\mathtt{h}^i_k$ *vanishes*; compare Remark 5.1 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 3.7 – 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 5.4 below, from which the proof of the results described in § 2 will follow.

Remark 5.1 (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* (94), (76), *and* (79), *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). *By the chain rule,*

$$\partial_{I_1}^2 \mathtt{E}^i(I_1) = -\left.\frac{\partial_E^2 I_1^i}{(\partial_E I_1^i)^3}\right|_{\mathtt{E}^i(I_1)}, \tag{145}$$

and we see that $\partial_{I_1}^2 \mathtt{E}^i$ *must vanish at some points in the interval* (a^i, b^i) *defined in* (94).

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* (79), *the function* $E \to \partial_E^2 \bar{I}_1^i$ *tends to* $+\infty$ *when* $E \to \bar{E}_+^i$. *Thus, by* (145), $\partial_{I_1}^2 \bar{\mathtt{E}}^i(I_1)$ *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*

$$\delta := 3d_2 d_4 - 5d_3^2 > 0, \qquad d_j := (\partial_{q_1}^j \bar{\mathtt{G}})(\bar{\theta}_i), \tag{146}$$

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 (146) *is easily satisfied. For example, if* $\bar{\mathtt{G}}(\theta) = \cos\theta - \frac{1}{8}\cos(2\theta)$, *one finds that* $\delta = 3/2$, *so that, in this very simple case, 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 3.8.*

5.1 Twist Theorem near simple resonances (statement)

To state the Twist Theorem we need to introduce two parameters ($\xi > 0$, $\mathtt{m} \geqslant 1$) which measure the nondegeneracy of the energy as function of actions in the inner regions $0 < i < 2N_k$; compare Definition 5.3 below. This requires some preparation.

Nondegenerate functions and theirs sublevels

First, let us recall a standard quantitative definition of nondegenerate functions.

Definition 5.1 *Given $\xi > 0$, an open set $A \subseteq \mathbb{R}$, and $f \in C^m(A, \mathbb{R})$, we say that f is ξ-nondegenerate at order $m \geqslant 1$ on A (or, in short, (ξ, m)-nondegenerate), if*

$$\inf_{x \in A} \max_{1 \leqslant j \leqslant m} |f^{(j)}(x)| \geqslant \xi. \tag{147}$$

An important property of nondegenerate functions is that one can easily estimate the measure of their sublevels. In fact, denoting

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

one has:

Lemma 5.1 *Let f be a (ξ, m)-nondegenerate function on a bounded interval (a, b) and let $M := \|f\|_{C^{m+1}(a,b)}$. Then, there exist a constants $c_m > 1$, depending only on m, such that, for all $\eta > 0$, one has*

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

The proof of this lemma can be found, e.g., in [25, Lemma B.1]; compare, also, [41].

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

Consider a general Hamiltonian (43) in standard form. Recall Definition 3.4 and (94). For $0 \leqslant \lambda \leqslant \bar{\lambda}_{\max}$ (see (91)), and $0 \leqslant i \leqslant 2N_k$, define:

$$\bar{a}^i := a^i|_{\mu=0}, \quad \bar{b}^i := b^i|_{\mu=0}, \quad \bar{a}^i_\lambda := a^i_\lambda|_{\mu=0}, \quad \bar{b}^i_\lambda := b^i_\lambda|_{\mu=0}. \tag{148}$$

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 5.2 (Normalized second derivative of the energy function within separatrices) *Given $\mathtt{H}_\flat$ in standard form with reference potential $\bar{\mathtt{G}}$, we denote, for $0 < i < 2N_k$,*

$$\mathtt{F}^i_{\bar{\mathtt{G}}}(x) := (\partial^2_{I_1} \bar{\mathtt{E}}^i_{\bar{\mathtt{G}}})\big(\bar{a}^i_{\bar{\mathtt{G}}} + (\bar{b}^i_{\bar{\mathtt{G}}} - \bar{a}^i_{\bar{\mathtt{G}}})x\big), \quad \forall\, x \in (0, 1). \tag{149}$$

These functions satisfy a remarkable rescaling property:

Lemma 5.2 *If* $\mathtt{F}^i_{\bar{\mathtt{G}}}$ *is as in Definition 5.2, 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

$$\bar{I}^i_{1,\lambda\bar{\mathtt{G}}}(E) = \sqrt{\lambda}\bar{I}^i_{1,\bar{\mathtt{G}}}(E/\lambda), \qquad \bar{\mathtt{E}}^i_{\lambda\bar{\mathtt{G}}}(I_1) = \lambda\bar{\mathtt{E}}^i_{\bar{\mathtt{G}}}(I_1/\sqrt{\lambda}), \qquad \forall\lambda > 0. \tag{150}$$

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

$$\bar{I}^{2N_k}_{1,\lambda\bar{\mathtt{G}}}(E) \overset{(253)}{=} \frac{1}{2\pi}\int_0^{2\pi}\sqrt{E-\lambda\bar{\mathtt{G}}(x)}dx = \frac{\sqrt{\lambda}}{2\pi}\int_0^{2\pi}\sqrt{\frac{E}{\lambda}-\bar{\mathtt{G}}(x)}dx$$
$$= \sqrt{\lambda}\bar{I}^i_{1,\bar{\mathtt{G}}}(E/\lambda),$$

proving the first equality in (150), which, in turns, implies immediately the second equality. From (150), then , follows that

$$\bar{a}^i_{\lambda\bar{\mathtt{G}}} = \sqrt{\lambda}\bar{a}^i_{\bar{\mathtt{G}}}, \qquad \bar{b}^i_{\lambda\bar{\mathtt{G}}} = \sqrt{\lambda}\bar{b}^i_{\bar{\mathtt{G}}}. \tag{151}$$

The claim follows, now, at once from (150) and (151). ∎

Let us go back to the Hamiltonians in standard form $\mathtt{H}_k$ of Theorem 4.1, 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)-nondegenerate.

Lemma 5.3 *For every* $0 < i < 2N_k$*, the function* $\mathtt{F}^i_{\bar{\mathtt{G}}}$ *defined in* (149) *is* (ξ, m)*-nondegenerate for some* $\xi, m > 0$.

Proof We consider only the case i odd, the even case being similar. Deriving (145) we get, for $\mu = 0$,

$$\partial^3_{I_1}\bar{\mathtt{E}}^i(\bar{I}^i_1(E)) = -\frac{\partial^3_E\bar{I}^i_1(E)}{\big(\partial_E\bar{I}^i_1(E)\big)^4} + 3\frac{\big(\partial^2_E\bar{I}^i_1(E)\big)^2}{\big(\partial_E\bar{I}^i_1(E)\big)^5}. \tag{152}$$

By (76)÷(81) (which hold also for $\bar{I}^i_1$, corresponding to $\mu = 0$), we have that the dominant term in (152), for $z = (\bar{E}^i_+ - E)/\epsilon \to 0^+$, has the form $-1/(\mathtt{c}^3z^2\log^4 z)$, with $\mathtt{c} = \psi^i_+(0)|_{\mu=0}$. Then,

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

By (149), we obtain

$$\lim_{x\to1^-}|\partial_x\mathtt{F}^i_{\bar{\mathtt{G}}}(x)| = +\infty. \tag{153}$$

Moreover, $\partial_x \mathtt{F}^i_{\bar{\mathtt{G}}}(x)$ is analytic in a neighborhood of $x = 0$ (recall in particular, (80)). Assume now, by contradiction, that (147) does not hold, namely that there exists a sequence $x_m \in (0,1)$ such that

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

By (153), up to a subsequence, x_m converges to some $\bar{x} \in [0,1)$, such that $\partial_x^j \mathtt{F}^i_{\bar{\mathtt{G}}}(\bar{x}) = 0$ 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 (153). ∎

This lemma allows us to introduce *uniform* nondegeneracy parameters $\mathfrak{x} > 0$ and $\mathtt{m} \geqslant 1$ for the function $\mathtt{F}^i_{\bar{\mathtt{G}}}$ in (149) 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 5.2,

$$\mathtt{F}^i_{\bar{\mathtt{G}}} = \mathtt{F}^i_{\frac{2\varepsilon}{|k|^2}\pi_{\mathbb{Z}k} f} = \mathtt{F}^i_{\pi_{\mathbb{Z}k} f}, \tag{154}$$

and, by (17), every potential $\pi_{\mathbb{Z}k} f$ is β-Morse. By the above Lemma 5.3, every function in (154) is (ξ, m)-nondegenerate for some $\xi, m > 0$. We therefore can define uniform ε-independent nondegeneracy parameters $\mathfrak{x}$, $\mathtt{m}$ by setting:

Definition 5.3 *Let $\mathtt{F}^i_{\pi_{\mathbb{Z}k} f}$ be as in Definition 5.2 with rescaled reference potential $\bar{\mathtt{G}} = \pi_{\mathbb{Z}k} f$. We define $\mathfrak{x} > 0$ 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 $(\mathfrak{x}, \mathtt{m})$-nondegenerate (Definition 5.1).*

Twist Theorem

Let Assumptions 3.1 and Definitions 3.3 hold, let $\mathfrak{g}$ be as in (56), let $\mathfrak{x}, \mathtt{m}$ be as in Definition 5.3, let B^i_k be as in (110), let $\mathtt{h}^i_k$ be as in (112), and define

$$\mathfrak{d}_{\rm o} := |k|^{-2n}. \tag{155}$$

Then, the following result holds.

Theorem 5.4 (Twist theorem) *There exists $\mathfrak{c}_0 = \mathfrak{c}_0(n, \mathfrak{g}, \mathfrak{x}, \mathtt{m}) > 1$ such that, for $\mathtt{K}_{\rm o} \geqslant \mathfrak{c}_0$, $k \in \mathcal{G}^n_{\mathtt{K}_{\rm o}}$, $0 \leqslant i \leqslant 2N_k$, and $0 < \eta < \mathfrak{d}_{\rm o}/2^5$, one has:*

$$\mathrm{meas}\left(\left\{ I \in B^i_k : \ \left|\det \partial_I^2 \mathtt{h}^i_k(I)\right| \leqslant \eta \right\}\right) \leqslant \mathfrak{c}_0 (|k|^{2n}\eta)^{\mathtt{b}}\, \mathrm{meas}\, B^i_k, \tag{156}$$

with $\mathtt{b} := \min\{\frac{1}{9n^4}, \frac{1}{\mathtt{m}}\}$.

Theorem 5.4 will be proven in several steps, which we now summarize:

Step 1: Preliminaries

This is a preliminary step, where few simple computations are provided:

(a) Evaluation of the twist matrix in the inner case ($0 < i < 2N_k$).

(b) Analogous formulae for the outer case ($i = 0, 2N_k$); however, in view of the presence of the translation $j_{-\mathtt{g}_\star}$ in (108), 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 (110)).

(c) Uniform estimates on the sub-matrix $\partial^2_{\hat I}\hat h_{_k}$ of order $(n-1)$, depending only on the "trivial actions" $\hat I$.

Step 2: Coverings of the phase space into regions close to separatrices and far from separatrices

This step is needed 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*: Nondegeneracy 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 3.7, we will prove that a suitable regularization of the twist determinant is a nondegenerate function, allowing to control the measure of its sublevels. 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 nondegeneracy condition involving the parameters ξ 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

5.2 Proof of the Twist Theorem

Notation 2 *Throughout the proof the $(n-1)$ dimensional domain $\hat{D}$ (defined in (49)) will be kept fixed and often the variables $\hat{I}$ will not be indicated explicitly. Also, the label k will usually be omitted in the notation, as well as the suffix i (when this does not lead to confusion).*

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

Step 1: Preliminaries

(a) Let us give the analytic expression of the twist determinant inside separatrices, i.e., for $0 < i < 2N$. Recall that in this case, by (112) and (110), one has $\mathtt{h}^i = \mathtt{E}^i + \hat{h}_{{}_k}$ and $B^i = \mathcal{B}^i_k$. Observe also 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 $\det \mathrm{S} = \mathrm{s}_{11}\cdot \det\hat{\mathrm{S}} + \det \mathrm{S}_0$. Thus,

$$\det \partial^2_I \mathtt{h}^i = \det\big(\partial^2_I \mathtt{E}^i + \partial^2_I \hat{h}_{{}_k}\big) = \det\begin{pmatrix} \partial^2_{I_1}\mathtt{E}^i & \partial^T_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) \\ \partial_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) & \partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{\hat{I}}\hat{h}_{{}_k}\end{pmatrix} \tag{157}$$

$$= (\partial^2_{I_1}\mathtt{E}^i)\cdot\det(\partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{\hat{I}}\hat{h}_{{}_k}) + \det\begin{pmatrix} 0 & \partial^T_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) \\ \partial_{\hat{I}}(\partial_{I_1}\mathtt{E}^i) & \partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{\hat{I}}\hat{h}_{{}_k}\end{pmatrix}.$$

Remark 5.2 *At this point, we may give a motivation why, in view of the application in Corollary 2.2, the expected perturbative argument for the nondegeneracy of the determinant does not apply. When $\mu = 0$, the Hamiltonian $\mathtt{H}_\flat(p, q_1)$ in (43), does not depend on the dumb actions $\hat{\mathrm{p}}$ and, as a consequence, $\mathtt{E}^i$ depends only on I_1. Let us consider the low mode case $|k|_1 \leqslant \mathtt{N}$. We will prove in Lemma 5.6 below that $\det(\partial^2_{\hat{I}}\hat{h}_{{}_k})$ is nondegenerate. Then, recalling(74), we see that, for $\mu = 0$, the nondegeneracy of*

$$\det \partial^2_I \mathtt{h}^i|_{\mu=0} = \partial^2_{I_1}\bar{\mathtt{E}}^i \cdot \det(\partial^2_{\hat{I}}\hat{h}_{{}_k})$$

reduces to the one of the 1D-function $I_1 \to \partial^2_{I_1}\bar{\mathtt{E}}^i(I_1)$, whose nondegeneracy has been proved in Lemma 5.3 (recall also Definition 5.2). Let us now take $\mu > 0$. By (99), (44),

(49) *and the definition of* μ_o *in* (77) *we have, for a suitable constant* $\hat{\mathbf{c}}_*$, *that*

$$\left|\partial^2_{I_1\hat{I}}\mathtt{E}^i\right| \leqslant \hat{\mathbf{c}}_* \frac{\mu}{\hat{\lambda}}, \qquad \left|\partial^2_{\hat{I}}\mathtt{E}^i\right| \leqslant \hat{\mathbf{c}}_* \frac{\mu^2}{\hat{\lambda}} \qquad \textit{where} \quad \hat{\lambda} = \lambda|\log\lambda|^3, \tag{158}$$

where $\hat{\lambda}$ *has been defined in Theorem 3.9. Recall that, by* (49), $\mu = \mathtt{K}^{-5n}$. *In view of the application in Corollary 2.2, we take* $\mathtt{K} = \mathfrak{c}|\log\varepsilon|$ *(recall* (12)*) and, up to a logarithmic correction,* $\lambda \lesssim \sqrt{\varepsilon}$ *(recall Remark 4.3 (ii)). In conclusion, in* (158), *both* $\mu^2/\hat{\lambda}$ *and, a fortiori,* $\mu/\hat{\lambda}$ *behave as* $|\log\varepsilon|^{-c}\varepsilon^{-1/2}$ *for some* $c > 0$, *yielding a useless bound* $|\log\varepsilon|^{-c}\varepsilon^{-n/2}$, *coming from last determinant in* (157).

(b) We now consider the case outside the outer separatrices, i.e., $i = 0, 2N$. Recalling (112), (110) and (94), we see that the Hamiltonian $\mathtt{h}^i(I)$, in this case, is given by $\mathtt{h}^i(I) = \widetilde{\mathtt{E}}^i(I) + \hat{h}_{k}(\hat{I})$ for $I \in B^i = j_{-\mathtt{g}_\star}\left(\mathcal{B}^i_k\right)$. 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 $I = \mathrm{U}\mathrm{I} = j_{\mathtt{g}_1}(\mathrm{I})$, where U is defined in (60). Then, we observe that, defining

$$h^i(\mathrm{I}) := \mathtt{E}^i_*(\mathrm{I}) + \hat{h}_{k}(\hat{\mathrm{I}}), \qquad \mathtt{E}^i_* := \mathtt{E}^i_k \circ j_{-\mathtt{g}_3}, \tag{159}$$

one has that, for all $\mathrm{I} \in \mathrm{U}^{-1}B^i$,

$$j_{\mathtt{g}_\star} = j_{-\mathtt{g}_3} \circ \mathrm{U}^{-1}, \qquad \mathtt{h}^i(\mathrm{U}\mathrm{I}) = h^i(\mathrm{I}). \tag{160}$$

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

$$\det\left[\partial^2_{\mathrm{I}} h^i(\mathrm{I})\right] \overset{(160)}{=} \det\left[\partial^2_{\mathrm{I}}\left(\mathtt{h}^i(\mathrm{U}\mathrm{I})\right)\right] = \det\left[\mathrm{U}^T \partial^2_I \mathtt{h}^i(I)\ \mathrm{U}\right] = \det\left[\partial^2_I \mathtt{h}^i(I)\right].$$

Thus,

$$(\det \partial^2_I \mathtt{h}^i) \circ \mathrm{U} = \det \partial^2_{\mathrm{I}} h^i. \tag{161}$$

Recalling (159) we then obtain

$$\begin{aligned} \det \partial^2_{\mathrm{I}} h^i &= \det\left(\partial^2_{\mathrm{I}} \mathtt{E}^i_* + \partial^2_{\mathrm{I}} \hat{h}_{k}\right) \\ &= \det\begin{pmatrix} \partial^2_{\mathrm{I}_1}\mathtt{E}^i_* & \partial^T_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) \\ \partial_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) & \partial^2_{\hat{\mathrm{I}}}\mathtt{E}^i_* + \partial^2_{\hat{\mathrm{I}}}\hat{h}_{k} \end{pmatrix} \\ &= (\partial^2_{\mathrm{I}_1}\mathtt{E}^i_*) \cdot \det(\partial^2_{\hat{\mathrm{I}}}\mathtt{E}^i_* + \partial^2_{\hat{\mathrm{I}}}\hat{h}_{k}) \\ &\quad + \det\begin{pmatrix} 0 & \partial^T_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) \\ \partial_{\hat{\mathrm{I}}}(\partial_{\mathrm{I}_1}\mathtt{E}^i_*) & \partial^2_{\hat{\mathrm{I}}}\mathtt{E}^i_* + \partial^2_{\hat{\mathrm{I}}}\hat{h}_{k} \end{pmatrix}, \end{aligned} \tag{162}$$

and, by the chain rule,

$$
\begin{aligned}
(\partial^2_{\mathrm{I}_1}\mathtt{E}^i_*)\circ j_{\mathtt{g}_3} &= \partial^2_{I_1}\mathtt{E}^i,\\
(\partial^2_{\mathrm{I}_1\hat{\mathrm{I}}}\mathtt{E}^i_*)\circ j_{\mathtt{g}_3} &= \partial^2_{I_1\hat{I}}\mathtt{E}^i - \partial^2_{I_1}\mathtt{E}^i\partial_{\hat{I}}\mathtt{g}_3 =: \hat{\mathrm{v}},\\
(\partial^2_{\hat{\mathrm{I}}\hat{\mathrm{I}}}\mathtt{E}^i_*)\circ j_{\mathtt{g}_3} &= \hat{\mathrm{M}},
\end{aligned}
\tag{163}
$$

with

$$\hat{\mathrm{M}} := \partial^2_{\hat{I}}\mathtt{E}^i + \partial^2_{I_1}\mathtt{E}^i\partial^T_{\hat{I}}\mathtt{g}_3\partial_{\hat{I}}\mathtt{g}_3 - \partial_{I_1}\mathtt{E}^i\partial^2_{\hat{I}}\mathtt{g}_3 - \partial^T_{\hat{I}}\partial_{I_1}\mathtt{E}^i\partial_{\hat{I}}\mathtt{g}_3 - \partial^T_{\hat{I}}\mathtt{g}_3\partial_{\hat{I}}\partial_{I_1}\mathtt{E}^i.$$

Recalling that by (160) $j_{\mathtt{g}_\star} = j_{-\mathtt{g}_3}\circ \mathrm{U}^{-1}$, by (161), (162) and (163) we get

$$\delta^i_\sharp := (\det \partial^2_I \mathtt{h}^i)\circ j_{-\mathtt{g}_\star} = \partial^2_{I_1}\mathtt{E}^i\cdot\det(\hat{\mathrm{M}} + \partial^2_{\hat{I}}\hat{h}_k) + \det\begin{pmatrix} 0 & \hat{\mathrm{v}}^T \\ \hat{\mathrm{v}} & \hat{\mathrm{M}} + \partial^2_{\hat{I}}\hat{h}_k\end{pmatrix}. \tag{164}$$

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

$$\operatorname{meas} B^i_k = \operatorname{meas}\mathcal{B}^i_k, \qquad i = 0, 2N, \tag{165}$$

so that one obtains the following

Lemma 5.5 *Let $i = 0, 2N$ and $\delta^i_\sharp$ as in (164). Then,*

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

(c) Here we prove a uniform bound on the Hessian sub-matrix $\partial^2_{\hat{I}}\hat{h}_k$. Recall the definition of $\delta_{\rm o}$ in (155)

Lemma 5.6 *There exists $c_3 = c_3(n) > 1$ such that, if $\mathtt{K} \geqslant c_3$, then the following estimates on the sub-matrix $\partial^2_{\hat{I}}\hat{h}_k$ hold:*

$$\sup_{\hat{D}_{\mathrm{r}}}|\partial^2_{\hat{I}}\hat{h}_k| \leqslant 2n^5 + 1, \qquad \inf_{\hat{D}_{\mathrm{r}}\cap\mathbb{R}^{n-1}}\det\partial^2_{\hat{I}}\hat{h}_k \geqslant \delta_{\rm o}. \tag{166}$$

Proof By (19),

$$
\begin{aligned}
(\mathrm{A}^T\mathrm{U})I &= I_1 k + \hat{\mathrm{A}}^T\hat{I} - \frac{(\hat{\mathrm{A}}k)\cdot\hat{I}}{|k|^2}k = I_1 k + \hat{\mathrm{A}}^T\hat{I} - \frac{\hat{\mathrm{A}}^T\hat{I}\cdot k}{|k|^2}k\\
&= I_1 k + \pi^\perp_k\hat{\mathrm{A}}^T\hat{I}.
\end{aligned}
\tag{167}
$$

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

$$\begin{aligned}\partial_I^2\big(I_1^2+\hat{\mathcal{Q}}_k(\hat{I})\big)&=\partial_I^2\big(I_1^2+\frac{|\pi_k^\perp \hat{\mathrm{A}}^T\hat{I}|^2}{|k|^2}\big)\overset{(167)}{=}\frac{\partial_I^2|\mathrm{A}^T\mathrm{U}I|^2}{|k|^2}\\&=\frac{2(\mathrm{A}^T\mathrm{U})^T\mathrm{A}^T\mathrm{U}}{|k|^2},\end{aligned}\tag{168}$$

and

$$|\partial_{\hat{I}}^2\hat{\mathcal{Q}}_k|\leqslant 2|k|^{-2}|\mathrm{A}|^2|\mathrm{U}|^2\overset{(33),(64)}{\leqslant}2n^5.\tag{169}$$

Using that $|k|\leqslant \mathtt{K}/6$, by (20), (49), (53), and Cauchy estimates we get, for a suitable $c'=c'(n)$,

$$\sup_{\hat{D}_{\mathtt{r}}}|\partial_{\hat{I}}^2(\hat{h}_k-\hat{\mathcal{Q}}_k)|\leqslant\frac{c'}{\mathtt{K}^{14n+2}}.\tag{170}$$

By (169) and (170), taking $\mathtt{K}$ large enough (depending only on n), we get the first estimate in (166).

Let us prove the second estimate in (166). Observe that

$$\begin{aligned}2\det\partial_{\hat{I}}^2\hat{\mathcal{Q}}_k&=\det\partial_I^2\big(I_1^2+\hat{\mathcal{Q}}_k\big)\overset{(168)}{=}\frac{2^n}{|k|^{2n}}\det\big((\mathrm{A}^T\mathrm{U})^T\mathrm{A}^T\mathrm{U}\big)\\&\overset{(33,60)}{=}\frac{2^n}{|k|^{2n}}\geqslant\frac{4}{|k|^{2n}}=4\delta_{\mathtt{o}},\end{aligned}$$

and that

$$\begin{aligned}|(\partial_{\hat{I}}^2\hat{\mathcal{Q}}_k)^{-1}|\leqslant|(\partial_I^2(I_1^2+\hat{\mathcal{Q}}_k))^{-1}|&\overset{(168)}{\leqslant}\tfrac{|k|^2}{2}|\mathrm{A}^{-1}|^2|\mathrm{U}^{-1}|^2\\&\overset{(33),(64)}{\leqslant}\tfrac{1}{2}n^5(n-1)^{n-1}|k|^{2n}.\end{aligned}\tag{171}$$

Then, by (171), (170), using $|k|\leqslant\mathtt{K}/6$, we get, for a suitable constant $c''=c''(n)$,

$$|(\partial_{\hat{I}}^2\hat{\mathcal{Q}}_k)^{-1}|\cdot|\partial_{\hat{I}}^2(\hat{h}_k-\hat{\mathcal{Q}}_k)|\leqslant\frac{c''}{\mathtt{K}^{12n+2}}.\tag{172}$$

We now need an elementary result on perturbation of positive-definite matrices, whose proof is given in Appendix A.

Lemma 5.7 *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 $\det(P+Q)\geqslant(\det P)/2$.*

Observe that since $\partial_I^2\big(I_1^2+\hat{\mathcal{Q}}_k\big)$ is positive-definite (by (168)), so is $\partial_{\hat I}^2\hat{\mathcal{Q}}_k$. Therefore, since

$$\partial_{\hat I}^2\hat h_{_k} = \partial_{\hat I}^2\hat{\mathcal{Q}}_k + \partial_{\hat I}^2(\hat h_{_k} - \hat{\mathcal{Q}}_k),$$

in view of (172), taking $\mathtt{K} \geqslant c_{_3}$ for a suitable $c_{_3} = c_{_3}(n) > 1$, Lemma 5.7 implies also the second estimate in (166). The proof of Lemma 5.6 is complete. ∎

Step 2

Here we define suitable *coverings of the sets* $\mathcal{B}_k^i$ defined in (94), (49). Such coverings are made up of sets corresponding to zones close to separatrices and zones away from them.

Recall the definitions given in (94), (91) and (148); recall (116); finally, recall that for i odd $a_0^i \equiv 0$ (see (95)). Now, for any $\lambda_{\rm o} \in (0, 1/\mathbf{c}_\star)$, we are going to define suitable subsets of $\mathcal{B}_k^i$. In the following sets it is understood that the dumb action $\hat I$ varies in the set $\hat D$:

$$\begin{array}{ll}\begin{cases}\mathcal{B}^i_{_{\rm near}}(\lambda_{\rm o}) := \{I : b^i_{\lambda_{\rm o}}(\hat I) < I_1 < b^i(\hat I)\}, \\ \mathcal{B}^i_{_{\rm far}}(\lambda_{\rm o}) := \mathcal{I}^i \times \hat D, \quad \mathcal{I}^i := (0, \bar b^i_{\lambda_{\rm o}/2}),\end{cases} & \text{if } i \text{ is odd,} \\ \begin{cases}\mathcal{B}^i_{_{\rm near}}(\lambda_{\rm o}) := \{I : a^i(\hat I) < I_1 < a^i_{\lambda_{\rm o}}(\hat I)\}, \\ \mathcal{B}^i_{_{\rm far}}(\lambda_{\rm o}) := \mathcal{I}^i \times \hat D, \quad \mathcal{I}^i := (\bar a^i_{\lambda_{\rm o}/2}, b^i(\hat I)),\end{cases} & \text{if } i = 0, 2N_k,\end{array} \tag{173}$$

while, for i even and different from 0 and $2N_k$,

$$\begin{cases}\mathcal{B}^i_{_{\rm near}}(\lambda_{\rm o}) := \{I : a^i(\hat I) < I_1 < a^i_{\lambda_{\rm o}}(\hat I)\} \cup \{I : b^i_{\lambda_{\rm o}}(\hat I) < I_1 < b^i(\hat I)\}, \\ \mathcal{B}^i_{_{\rm far}}(\lambda_{\rm o}) := \mathcal{I}^i \times \hat D, \quad \mathcal{I}^i := (\bar a^i_{\lambda_{\rm o}/2}, \bar b^i_{\lambda_{\rm o}/2}).\end{cases}$$

Recall the definition of $\mathbf{c}_\star$ in Theorem 4.1, and of $\mathbf{c} \leqslant \mathbf{c}_\star$ in Theorem 3.7. Recall, also, (116). Then, one has:

Lemma 5.8 *Let* $0 \leqslant i \leqslant 2N$ *and assume that*

$$\lambda_{\rm o} < 1/\mathbf{c}_\star, \qquad \mu \leqslant \lambda_{\rm o}^2/2^8\mathbf{c}^4. \tag{174}$$

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

Proof We give a detailed proof in the case i odd, as there is no extra difficulty in adapting 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 (81)), the functions $\lambda \to \bar{b}_\lambda$ and $\lambda \to b_\lambda(\hat{I})$ are positive and strictly decreasing. We claim that

$$b_{\lambda_o}(\hat{I}) < \bar{b}_{\lambda_o/2} < b_{\lambda_o/4}(\hat{I}), \qquad \forall\, \hat{I} \in \hat{D}. \tag{175}$$

From such relations the claim follows. Indeed, the fact that $\mathcal{B}_{\rm far}$ is a subset of $\mathcal{B}$ follows from the second inequality in (175); the equality $\mathcal{B} = \mathcal{B}_{\rm near} \cup \mathcal{B}_{\rm far}$ follows from the first inequality in (175). Let us prove in detail the first inequality in (175) (the second one being analogous). Now, notice that, by (91), (94), and (95), one has $b_{\lambda_{\rm max}}(\hat{I}) = I_1(E_-(\hat{I}), \hat{I}) = a_0(\hat{I}) = 0$. Thus,

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

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 (91). Note that, by (71) and (116), $|\lambda_\sharp| \leqslant 3\mathfrak{g}^3 \mu \leqslant \mathbf{c}\mu$, and that by (91), (92) we have that $\lambda_o \leqslant 1/\mathbf{c}_\star \leqslant \min\{\lambda_{\rm max}/4, \bar{\lambda}_{\rm max}/8\}$. Then again by (92), (91), (93), and (174) we get that, for every $\hat{I} \in \hat{D}$,

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

We write

$$\begin{aligned}\frac{\bar{b}_{\lambda_o/2} - b_{\lambda_o}(\hat{I})}{\epsilon} &= \int_{\lambda_o/2}^{\lambda_o + \lambda_\sharp} \partial_E \bar{I}_1(\bar{E}_+ - \epsilon z)\, dz + \int_{\lambda_{\rm max} + \lambda_\sharp}^{\bar{\lambda}_{\rm max}} \partial_E \bar{I}_1(\bar{E}_+ - \epsilon z)\, dz \\ &\quad + \int_{\lambda_o + \lambda_\sharp}^{\lambda_{\rm max} + \lambda_\sharp} \big(\partial_E \bar{I}_1(\bar{E}_+ - \epsilon z) - \partial_E I_1(\bar{E}_+ - \epsilon z, \hat{I})\big)\, dz.\end{aligned}$$

Recall (83); note that (174) implies (82) with $\lambda = \lambda_o/8$, and observe 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 $\mathcal{E}_{\lambda_o/8}$. Then, by (81), (84) and (174) we get, for every $\hat{I} \in \hat{D}$,

$$\begin{aligned}\frac{\bar{b}_{\lambda_o/2} - b_{\lambda_o}(\hat{I})}{\epsilon} &\geqslant \frac{\lambda_o - 2|\lambda_\sharp|}{2\mathbf{c}\sqrt{\epsilon}} - \mathbf{c}^2 \frac{|\log \frac{\lambda_o}{8}|}{\sqrt{\epsilon}} \big(|\bar{\lambda}_{\rm max} - \lambda_{\rm max}| + |\lambda_\sharp|\big) - 8\mathbf{c}^2 \mu \frac{\lambda_{\rm max}}{\lambda_o \sqrt{\epsilon}} \\ &\overset{(93),(92)}{\geqslant} \frac{1}{2\mathbf{c}\sqrt{\epsilon}} \big(\lambda_o - 2\mathbf{c}\mu - 2^4 \mathbf{c}^4 \mu |\log \tfrac{\lambda_o}{8}| - 2^6 \mathbf{c}^3 \mu/\lambda_o\big) \geqslant \frac{\lambda_o}{4\mathbf{c}\sqrt{\epsilon}} > 0,\end{aligned}$$

completing the proof. ∎

Step 3*: Nondegeneracy in neighborhoods of separatrices

Here we show that (a suitable regularization of) the twist determinant $\det \partial_I^2 \mathtt{h}^i$ in (157) is a nondegenerate function in the sense of Definition 5.1 in *suitable neighborhoods of separatrices*. 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:

$$\delta_\mp(z,\hat I) := \det\big[\partial_I^2 \mathtt{E}\big(I_1(E^i_\mp(\hat I) \pm \epsilon z, \hat I), \hat I\big) + \partial_I^2 \hat h_k(\hat I)\big]. \tag{176}$$

The study of the twist determinant (176) will be based on the analytic properties described in Theorem 3.7. 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 *let, henceforth,* $\delta := \delta_+$.

The precise statement on the nondegeneracy of $z \to \delta(z,\hat I)$ (see Proposition 5.9 below) needs some preparation.

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

$$\zeta(z,\hat I) := z \cdot \big(\sqrt{\epsilon}\,\partial_E I_1(E_+(\hat I) - \epsilon z, \hat I)\big)^3, \tag{177}$$

and define *the regularized twist determinant* $\bar\delta$ by setting (recall (166)):

$$\bar\delta = \tilde\delta / \det \partial^2_{\hat I} \hat h_k, \quad \text{with} \quad \tilde\delta(z,\hat I) := \zeta(z,\hat I)^n \cdot \delta(z,\hat I). \tag{178}$$

The functions appearing in Theorem 3.7, as well as the functions in (177) and (178) belong to the following ring of functions $\mathcal F$.

Definition 5.4 *We denote by $\mathcal F$ the set of functions of the form*

$$f(z,\hat I) = z^h \sum_{j=0}^{\ell} u_j(z,\hat I) \log^j z, \tag{179}$$

where $h, \ell \in \mathbb Z$ with $\ell \geqslant 0$ and the u_j are real analytic functions on a (complex) neighborhood of $\{z=0\} \times \hat D \subset \mathbb C^n$; recall that the domain $\hat D$ is defined in (49), but, actually, here plays no role. 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 $f = u_1 f_1 + u_2 f_2$; for example, f in (179) can be written as

$$z^h\Big(\bigoplus_{j=0}^{\ell} \log^j z\Big).$$

We say that $f(z,\hat{I}) = \mathcal{O}_\varrho(h,\ell)$ *if* $f \in \mathcal{F}$ *as in* (179) *and*

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

Remark 5.3 *(i) The functions* $(z,\hat{I}) \to f(z,\hat{I}) = I_1^i\big(E_\mp^i(\hat{I}) \pm \epsilon z,\, \hat{I}\big)$ *in* (76) *of Theorem 3.7 belongs to* $\mathcal{F}$ *and, by* (77),

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

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

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

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

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

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 that there exist suitable polynomials $a_j(z)$ such that

$$\mathcal{L} = \sum_{j=1}^{\bar{\mathrm{m}}} a_j(z)\partial_z^j. \tag{180}$$

Actually,

$$\mathcal{L} = \sum_{j=\bar{n}+1}^{\bar{\mathrm{m}}} a_j z^{j-\bar{n}}\partial_z^j,$$

with a_j natural numbers.[13] Recall that the constants c_3 and $\mathbf{c}_\star$ have been introduced in, respetcively, Lemma 5.6 and Theorem 4.1; recall, also, (79). Then, the following proposition holds.

[13]For example, if $n = 2$ and $\bar{\mathrm{m}} = 7$, $\mathcal{L}$ is given by: $\mathcal{L} = z^6\partial_z^7 + 18z^5\partial_z^6 + 98z^4\partial_z^5 + 184z^3\partial_z^4 + 100z^2\partial_z^3 + 8z\partial_z^2$.

Proposition 5.9 *There exists* $\mathbf{c}_1 = \mathbf{c}_1(n, \mathfrak{g}) > c_3$ *such that if* $\mathtt{K} \geqslant \mathbf{c}_1$, *then the following holds.*

(i) One has

$$\mathcal{L}[\bar{\delta}] = \bar{n}!^{3\bar{n}+1}(3\bar{n})!\ \gamma^{3\bar{n}} + \mathcal{O}_\varrho(1, 3\bar{n}+1), \tag{181}$$

where

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

and

$$1/\mathbf{c} \leqslant \inf_{\hat{D}} |\gamma| \leqslant \sup_{\hat{D}} |\gamma| \leqslant \mathbf{c}. \tag{182}$$

(ii) There exist suitable positive constants

$$\xi_\sharp = \xi_\sharp(n, \mathfrak{g}) < 1, \qquad \lambda_{\rm o} = \lambda_{\rm o}(n, \mathfrak{g}) < 1/\mathbf{c}_\star,$$

such that, for $\hat{I} \in \hat{D}$, *the function* $z \to \bar{\delta}(z, \hat{I})$ *defined in* (178) *is* $\xi_\sharp$*-nondegenerate at order* $\bar{\mathrm{m}} = 3n^2 - 2n - 1$ *on the interval* $(0, \lambda_{\rm o})$.

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

Notation 3 *In the rest of this section, it is understood that in an expansion*

$$f = z^h \cdot \bigoplus_{j=0}^{\ell} \log^j z,$$

one has $|||f|||_\varrho \leqslant c$ *for a suitable constant* $c = c(n, \mathfrak{g})$; *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 5.10 *If* ζ *is as in* (177), $\tilde{I} = \tilde{I}(z, \hat{I}) := (b_z(\hat{I}), \hat{I})$ *and* $\mu_{\rm o}$ *is as in* (77), *one has, for* $2 \leqslant i, j \leqslant n$:

$$\begin{aligned} \zeta &= z\big(\gamma \log z + (1 \oplus z \log z)\big)^3 \\ &= \gamma^3 z \log^3 z + \mathcal{O}(1,2) + \mathcal{O}(2,3) = \mathcal{O}(1,3), \\ \zeta \cdot \partial^2_{I_1} \mathtt{E}|_{I=\tilde{I}} &= \gamma + z(1 \oplus \log z) = \gamma + \mathcal{O}(1,1), \\ \zeta \cdot \partial^2_{I_1 \hat{I}_i} \mathtt{E}|_{I=\tilde{I}} &= \mu_{\rm o}(1 \oplus z \log z \oplus z \log^2 z) = \mu_{\rm o}\, \mathcal{O}(0,2), \\ \zeta \cdot \partial^2_{\hat{I}_i \hat{I}_j} \mathtt{E}|_{I=\tilde{I}} &= \mu_{\rm o}(1 \oplus z \log z \oplus z \log^2 z \oplus z^2 \log^3 z) = \mu_{\rm o}\, \mathcal{O}(0,3). \end{aligned} \tag{183}$$

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

$$\partial_{I_1}\mathtt{E}^i = \frac{1}{\partial_E I_1}, \qquad \partial_{\hat I}\mathtt{E}^i = -\frac{\partial_{\hat I} I_1}{\partial_E I_1}, \qquad \partial^2_{I_1}\mathtt{E}^i = -\frac{\partial^2_E I_1}{(\partial_E I_1)^3},$$
$$\partial^2_{I_1\hat I}\mathtt{E}^i = \frac{\partial^2_E I_1 \partial_{\hat I} I_1}{(\partial_E I_1)^3} - \frac{\partial^2_{E\hat I} I_1}{(\partial_E I_1)^2}, \tag{184}$$
$$\partial^2_{\hat I}\mathtt{E}^i = -\frac{\partial^2_{\hat I} I_1}{\partial_E I_1} + \frac{\partial^T_{\hat I} I_1\ \partial_{\hat I}(\partial_E I_1) + \partial^T_{\hat I}(\partial_E I_1)\ \partial_{\hat I} I_1}{(\partial_E I_1)^2} - \frac{\partial^2_E I_1\ \partial^T_{\hat I} I_1\ \partial_{\hat I} I_1}{(\partial_E I_1)^3},$$

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 (184) and (76), we have

$$\begin{aligned} \sqrt{\epsilon}\partial_E I_1 &= \gamma\log z + (1\oplus z\log z) = 1\oplus\log z,\\ \partial_{\hat I_i} I_1 &= \mu_o(1\oplus z\log z),\\ \epsilon^{3/2}\partial^2_E I_1 &= -\gamma z^{-1} + (\log z\oplus z^{-1}) = \log z\oplus z^{-1},\\ \epsilon\partial^2_{E\hat I_i} I_1 &= \mu_o(1\oplus\log z),\\ \partial^2_{\hat I_i\hat I_j} I_1 &= \mu_o \mathtt{r}^{-1}(1\oplus z\log z) \overset{(44)}{=} \mu_o\epsilon^{-1/2}(1\oplus z\log z). \end{aligned} \tag{185}$$

Finally, by (185), (184), (76), and (77) we get

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

completing the proof. ∎

Lemma 5.11 *Let $\bar n := n-1$. Then, one has*

$$\bar\delta = \bar\delta(z,\hat I) = \gamma^{3\bar n} z^{\bar n}\log^{3\bar n} z + \mathcal{O}(\bar n+1, 3\bar n+1) + \mathcal{O}(0, 3\bar n-1). \tag{186}$$

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

$$|\delta(z,\hat I)| \geqslant \delta_o\ |\bar\delta(z,\hat I)|, \qquad \forall\, 0 < z \leqslant \lambda_o,\ \hat I \in \hat D. \tag{187}$$

Proof Recalling (157) we split $\tilde{\delta}$ in (178) in two terms. The first term is

$$\zeta^n(\partial^2_{I_1}\mathtt{E})\ \det(\partial^2_{\hat{I}}\hat{h}_k + \partial^2_{\hat{I}}\mathtt{E}) \overset{(183)}{=} \big(\gamma + \mathcal{O}(1,1)\big)\,\zeta^{\bar{n}}\det(\partial^2_{\hat{I}}\hat{h}_k + \partial^2_{\hat{I}}\mathtt{E}),$$

and, by (183), we have that

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

where in the last line we used (recall (77), (49), (166)):

$$\mu_{\rm o}\leqslant |k|^{-2n} = \delta_{\rm o}\leqslant \inf_{\hat{D}}\det\partial^2_{\hat{I}}\hat{h}_k. \tag{188}$$

The second term is,

$$\begin{aligned}
&\zeta^n\det\begin{pmatrix}0 & \partial^T_{\hat{I}}(\partial_{I_1}\mathtt{E})\\ \partial_{\hat{I}}(\partial_{I_1}\mathtt{E}) & \partial^2_{\hat{I}}\mathtt{E}+\partial^2_{\hat{I}}\hat{h}_k\end{pmatrix} = \mu_{\rm o}\det\begin{pmatrix}0 & w^T\\ w & \mathrm{N}\end{pmatrix}\\
&=\mu_{\rm o}^2(1\oplus z\log z\oplus z\log^2 z)^2(1\oplus z\log z\oplus z\log^2 z\oplus z\log^3 z)^{n-2}\\
&=\mu_{\rm o}^2\mathcal{O}(0,3n-4)+\mathcal{O}(n,3n-2),
\end{aligned}$$

where w is a $\bar{n}$-dimensional vector and N is an $(\bar{n}\times\bar{n})$ matrix satisfying by (183), for $2\leqslant i,j\leqslant n$,

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

Thus, the second term has the form $\mu_{\rm o}^2\big(\mathcal{O}(\bar{n}+1,3\bar{n}+1)+\mathcal{O}(0,3\bar{n}-1)\big)$. Summing up the two terms and using (188) we get (186).

By the first line in (183), we see that, taking $\mathbf{c}_1$ big enough, one has

$$|\zeta(z,\hat{I})|\leqslant 1,\qquad \forall\, 0<z\leqslant\lambda_{\rm o},\ \hat{I}\in\hat{D}.$$

Thus, by the definitions in (176), (178) and by (166), one obtains (187). ∎

Before giving the proof of Proposition 5.9, we need one more lemma. Define

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

Lemma 5.12 *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*

$$\mathcal{L}_{m,k}[z^m \log^k z + f_1 + f_2] = (m!)^{k+1} k! + f_3, \tag{189}$$

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

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

$$L\mathcal{O}_\varrho(0,\ell+1) = \mathcal{O}_{\frac{3}{4}\varrho}(1,\ell+1) + \mathcal{O}_{\frac{3}{4}\varrho}(0,\ell),$$

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

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. Indeed, the algebraic relations are just calculus, while the estimates follow easily by (iterated) use of Cauchy estimates.

Analogously, if $0 \leqslant \ell < k \leqslant \bar{\mathrm{m}}$ and $0 \leqslant q \leqslant \bar{\mathrm{m}}$, from the above relations, there follows that

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

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

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

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 lemma follows. ∎

Proof of Proposition 5.9 The estimates in (182) follow trivially from (79) and (77).

To check (181), observe that $\mathcal{L} = \mathcal{L}_{\bar{n},3\bar{n}}$, and use (186) in Lemma 5.11 and (189), 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 (181), and (182), we see that for $\lambda_o < 1/\mathbf{c}_\star$ small enough one has:

$$\frac{1}{\mathbf{c}^{3\bar{n}}} \leqslant \inf_{0<z\leqslant\lambda_o} \inf_{\hat{I}\in\hat{D}} \big|\mathcal{L}[\bar{\delta}]\big| \overset{(180)}{\leqslant} c'' \max_{1\leqslant j\leqslant \bar{\mathrm{m}}} |\partial_z^j \bar{\delta}|,$$

where $c'' = c''(n)$. Thus, for $\hat{I} \in \hat{D}$, $z \to \bar{\delta}(z, \hat{I})$ is $\xi_\sharp$-nondegenerate 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 5.13 *Let $k \in \mathcal{G}^n_{\mathtt{K}_o}$, $0 \leqslant i \leqslant 2N$, $\eta > 0$, and λ_o as in Proposition 5.9-(ii). Then, there exist a positive constant $\mathbf{c}_2 = \mathbf{c}_2(n, \mathfrak{g}) \geqslant \mathbf{c}$ such that, if $\mathtt{K} \geqslant \mathbf{c}_2$, then*

$$\mathrm{meas}\left\{I \in \mathcal{B}^i_{\mathrm{near}}(\lambda_o) : \ |\det \partial_I^2 \mathtt{h}^i(I)| \leqslant \eta\right\} \leqslant \mathbf{c}_2 \left(|k|^{2n}\eta\right)^{1/9n^4} \mathrm{meas}\, \mathcal{B}^i_k. \tag{190}$$

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

Notation 4 *Given two non negative functions f and g we say that $f \lessdot g$ if there exists a constant $c = c(n, \mathfrak{g}) \geqslant 1$, depending only on n and $\mathfrak{g}$, 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 $c = c(n, \mathfrak{g}) \geqslant 1$, such that $|f| \leqslant cg$.*

Recall that $\bar{\mathrm{m}} = 3n^2 - 2n - 1$ (compare Proposition 5.9).

Lemma 5.14 *There exists a constant $\mathbf{c}_3 = \mathbf{c}_3(n, \mathfrak{g}) > 1$ such that, for every $\hat{I} \in \hat{D}$ and $\eta > 0$, one has*

$$\mathrm{meas}\left\{z \in (0, \lambda_o] : |\bar{\delta}(z, \hat{I})| \leqslant \eta\right\} \leqslant \mathbf{c}_3 \eta^{\bar{\mathrm{a}}}, \qquad \bar{\mathrm{a}} := \tfrac{1}{\bar{\mathrm{m}}(\bar{\mathrm{m}}+3)}. \tag{191}$$

Proof If $z_0 \leqslant 2\eta^{\bar{\mathrm{a}}}$, estimate (191) is obvious. Consider the case $z_0 > 2\eta^{\bar{\mathrm{a}}}$. Let $\lambda_1 := \eta^{\bar{\mathrm{a}}} < z_0/2$. By (186), (166), and (182) we have that

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

where, as usual, as usual, $[\lambda_1, \lambda_o]_{\lambda_1/2}$ denotes the $\lambda_1/2$-complex-neighborhood of the real interval $[\lambda_1, \lambda_o]$. By Cauchy estimates

$$\sup_{\lambda_1\leqslant z\leqslant\lambda_o} \sup_{\hat{I}\in\hat{D}} \max_{1\leqslant j\leqslant\bar{\mathrm{m}}+1} |\partial_z^j \bar{\delta}(z, \hat{I})| \leqslant c_\flat/\lambda_1^{\bar{\mathrm{m}}+2} =: M, \tag{192}$$

for a suitable $c_\flat \geqslant 1$, depending only on $n, \mathfrak{g}$. Now, we can apply Lemma 5.1 with

$$f = \bar{\delta}, \quad m = \bar{\mathrm{m}}, \quad a = \lambda_1, \quad b = \lambda_o, \quad \xi = \xi_\sharp, \quad M \text{ as in (192)},$$

obtaining, for all $\hat{I} \in \hat{D}$,

$$\text{meas}\{z \in (\lambda_1, \lambda_o) : |\bar{\delta}(z, \hat{I})| \leqslant \eta\} < \eta^{\bar{\mathrm{a}}}. \tag{193}$$

Since the interval $(0, \lambda_1)$, has length $\lambda_1 = \eta^{\bar{\mathrm{a}}}$, from (193) we obtain the measure estimate (191). ∎

Now, recalling that $\delta_o = |k|^{-2n}$ (see (155)), we have:

Lemma 5.15 *There exists* $\mathbf{c}_4 = \mathbf{c}_4(n, \mathfrak{g}) \geqslant \max\{\mathbf{c}_1, \mathbf{c}_3\}$ *such that for* $k \in \mathcal{G}^n_{\mathtt{K}_o}$, *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 $\mathcal{Z}_\eta(\hat{I}) := \{z \in (0, \lambda_o] : \ |\delta(z, \hat{I})| \leqslant \eta\}$. By (187) and (191) we get, for all $\hat{I} \in \hat{D}$,

$$\mathrm{m}_\eta = \mathrm{m}_\eta(\hat{I}) := \text{meas}(\mathcal{Z}_\eta(\hat{I})) \leqslant \mathbf{c}_3 (\eta/\delta_o)^{\bar{\mathrm{a}}}. \tag{194}$$

Note that, since $\lambda_o \leqslant 1/2$ (see (174)), by definition

$$\mathrm{m}_\eta \leqslant \lambda_o \leqslant 1/2. \tag{195}$$

Recalling (94), 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, then, have that

$$\mathcal{I}_\eta(\hat{I}) = b_{\mathcal{Z}_\eta(\hat{I})}(\hat{I}) := \left\{I_1 = b_z(\hat{I}) : z \in \mathcal{Z}_\eta(\hat{I})\right\}, \tag{196}$$

since by definition of $\mathcal{Z}_\eta$, (176) and (173) $\delta(z, \hat{I}) = \det\left[\partial_{\hat{I}}^2 \hat{h}_k(\hat{I}) + \partial_{\hat{I}}^2 \mathtt{E}\big(b_z(\hat{I}), \hat{I}\big)\right]$. 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

$$\begin{aligned} \text{meas}(\mathcal{I}_\eta(\hat{I})) &= \int_{\mathcal{I}_\eta(\hat{I})} dI_1 \overset{(196)}{=} \int_{b_{\mathcal{Z}_\eta(\hat{I})}} dI_1 = \int_{\mathcal{Z}_\eta(\hat{I})} |\partial_z b_z(\hat{I})| dz \\ &\overset{(183)}{<} \sqrt{\epsilon} \int_{\mathcal{Z}_\eta(\hat{I})} |\log z| dz. \end{aligned}$$

Moreover, recalling (195),

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

Thus, since $\frac{1}{9n^4} < \bar{\mathrm{a}}$ (compare (191)), by (195), 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 (194), for every $\hat{I} \in \hat{D}$ and $\eta > 0$, $\mathrm{meas}(\mathcal{I}_\eta(\hat{I})) \lessdot \sqrt{\epsilon}\,(\eta/\delta_o)^{1/9n^4}$ and the claim follows from Fubini's Theorem. ∎

Proof of Proposition 5.13 By (94) we get

$$\begin{aligned} \mathrm{meas}\, \mathcal{B}_k^i &= \int_{\hat{D}} b(\hat{I}) d\hat{I} = \int_{\hat{D}} I_1\big(E_+(\hat{I}), \hat{I}\big) d\hat{I} \\ &= \int_{\hat{D}} d\hat{I} \int_{E_-(\hat{I})}^{E_+(\hat{I})} \partial_E I_1\big(E_+(\hat{I}), \hat{I}\big) dE \\ &\overset{(81)}{\geqslant} \frac{E_+(\hat{I}) - E_-(\hat{I})}{\mathbf{c}\sqrt{\epsilon}}\, \mathrm{meas}\, \hat{D} \overset{(91)}{\geqslant} \frac{\sqrt{\epsilon}}{2\mathfrak{g}\mathbf{c}}\, \mathrm{meas}\, \hat{D}. \end{aligned} \tag{197}$$

Lemma 5.15 and (197) imply at once (190), provided one takes $\mathbf{c}_2 \geqslant \mathbf{c}_4$ big enough. The proof of Proposition 5.13 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 (143), (49), (20), by Cauchy estimates, we have:

$$|\partial_{\hat{I}} \mathtt{g}_3|_{\hat{D}, 3\mathtt{r}} \lessdot 1/\mathtt{K}^{14n+2}, \qquad |\partial^2_{\hat{I}} \mathtt{g}_3|_{\hat{D}, 3\mathtt{r}} \lessdot 1/(\sqrt{\varepsilon}\, \mathtt{K}^{\frac{37}{2}n+2}). \tag{198}$$

Note that the term $\partial^2_{\hat{I}} \mathtt{g}_3$ in (198) has a "big" estimate, containing a $\sqrt{\varepsilon}$ at the denominator. However this does not cause any problem, since, by the first lines in (183), (184) and (185) one has

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

where, the regularizing term ζ is defined in Lemma 5.10, and the function in brackets belongs to $\mathcal{F}$ and has norm $|||\cdot|||_\varrho$ bounded by a constant depending only on n and $\mathfrak{g}$. At this point, mimicking the proof for the inner case, one gets easily (190) also in the outer case $i = 0, 2N$, provided $\mathbf{c}_2$ is taken big enough. The proof of of Proposition 5.13 is complete. ∎

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

Recall the definition of $\mathcal{B}_{\rm far} = \mathcal{B}^i_{\rm far}$ in (173).

Proposition 5.16 *Let $0 < i < 2N$. Then:*

(i) There exists a constant $\mathbf{c}_5 = \mathbf{c}_5(n, \mathfrak{g}) > 1$ such that, assuming $\mathtt{K} \geqslant \mathbf{c}_5$ and $\mathtt{N} \leqslant |k|_1 \leqslant \mathtt{K}_{\rm o}$, then one gets, on $\mathcal{B}^i_{\rm far}$, $|\det \partial^2_I \mathtt{h}| \geqslant \delta_{\rm o}/2^5$.

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

$$\text{meas}\left\{I \in \mathcal{B}^i_{\rm far} : \ |\det \partial^2_I \mathtt{h}(I)| \leqslant \eta\right\} \leqslant \hat{\mathfrak{c}}_0(|k|^{2n}\eta)^{\frac{1}{\mathtt{m}}}\ \text{meas}\,\mathcal{B}^i_k. \tag{199}$$

Remark 5.4 *Notice that by point (i), the set $\left\{I \in \mathcal{B}_{\rm far} : \ |\det \partial^2_I \mathtt{h}(I)| \leqslant \eta\right\}$, for $\eta < \delta_{\rm o}/2^5$, is empty. Therefore, in proving point (ii), one needs to consider only $|k|_1 < \mathtt{N}$.*

For definiteness, *in the proof of Proposition 5.16, 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.

Recall the definition of $\mathcal{I} = \mathcal{I}^i$ in (173), with $\lambda_{\rm o}$ as in Proposition 5.9-(ii), and notice that $\mathcal{I}$ depends only on $n, \mathfrak{g}$. Then, the following estimates hold.

Lemma 5.17 *There exists a constant $\mathbf{c}_6 = \mathbf{c}_6(n, \mathfrak{g}) > 1$ such that, defining $\mathtt{r}_* := \sqrt{\epsilon}/\mathbf{c}_6$, one has, for $I \in \mathcal{I}_{2\mathtt{r}_*} \times \hat{D}$,*

$$|\partial_{I_1}\mathtt{E}| \leqslant \mathbf{c}_6\sqrt{\epsilon}, \quad \left|\partial^2_{I_1}\mathtt{E}\right| \leqslant \mathbf{c}_6, \quad \left|\partial^2_{I_1\hat{I}}\mathtt{E}\right| \leqslant \mathbf{c}_6\mu_{\rm o}, \quad \left|\partial^2_{\hat{I}}\mathtt{E}\right| \leqslant \mathbf{c}_6\mu_{\rm o}, \tag{200}$$

and

$$|\partial_{I_1}\mathtt{E} - \partial_{I_1}\bar{\mathtt{E}}| \leqslant \mathbf{c}_6\sqrt{\epsilon}\mu. \tag{201}$$

Proof Recall the definitions in (94) and (148). Then:

$$\begin{aligned}
I_1(E_+(\hat{I}) - \epsilon\tfrac{\lambda_{\rm o}}{8}, \hat{I}) - \bar{b}_{\lambda_{\rm o}} &\geqslant \bar{I}_1(\bar{E}_+ - \epsilon\tfrac{\lambda_{\rm o}}{8}) - \bar{I}_1(\bar{E}_+ - \epsilon\frac{\lambda_{\rm o}}{2}) \\
&\quad - |I_1(E_+(\hat{I}) - \epsilon\tfrac{\lambda_{\rm o}}{8}, \hat{I}) - \bar{I}_1(\bar{E}_+ - \epsilon\tfrac{\lambda_{\rm o}}{8})| \\
&\overset{(81),(76)}{\geqslant} \frac{1}{\mathbf{c}\sqrt{\epsilon}}\big(\epsilon\tfrac{\lambda_{\rm o}}{2} - \epsilon\tfrac{\lambda_{\rm o}}{8}\big) - |\phi_+(\tfrac{\lambda_{\rm o}}{8}, \hat{I}) - \bar{\phi}_+(\tfrac{\lambda_{\rm o}}{8})| \\
&\quad - |\psi_+(\tfrac{\lambda_{\rm o}}{8}, \hat{I}) - \bar{\psi}_+(\tfrac{\lambda_{\rm o}}{8})| \cdot \tfrac{\lambda_{\rm o}}{8}|\log\tfrac{\lambda_{\rm o}}{8}| \\
&\overset{(78)}{\geqslant} \frac{3\sqrt{\epsilon}\lambda_{\rm o}}{8\mathbf{c}} - 2\mathbf{c}\sqrt{\epsilon}\mu \overset{(174)}{\geqslant} \frac{\sqrt{\epsilon}\lambda_{\rm o}}{4\mathbf{c}} > 0,
\end{aligned}$$

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 (98))

$$\mathcal{I}_{2\mathtt{r}_*} \times \hat{D} \subseteq \big(\mathcal{B}^i_k(\lambda_o/8)\big)_{\rho_{\lambda_o/8}}.$$

Thus, observing that (obviously) the first two estimates holds also for $\bar{\mathtt{E}} = \mathtt{E}|_{\mu=0}$, by (99), we get (200).

Next, observe that, by the definitions given in (173) and (148), one has $\bar{\mathtt{E}}(\mathcal{I}) = (\bar{E}_-, \bar{E}_+ - \epsilon\lambda_o/2)$. Then, recalling (83), by the first estimate in (200), for $\mathbf{c}_6$ big enough, we get, for any $\hat{I} \in \hat{D}$,

$$\bar{\mathtt{E}}(\mathcal{I}_{2\mathtt{r}_*}) \subseteq \mathcal{E}_{\lambda_o/8}, \qquad \mathtt{E}(\mathcal{I}_{2\mathtt{r}_*}, \hat{I}) \subseteq \mathcal{E}_{\lambda_o/8}. \tag{202}$$

Let us, now, prove (201). 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

$$\begin{aligned}
&\sup_{\mathcal{I}_{2\mathtt{r}_*} \times \hat{D}} |\partial_{I_1}\mathtt{E}(I) - \partial_{I_1}\bar{\mathtt{E}}(I_1)| \\
&\overset{(202)}{\leqslant} \sup_{\mathcal{E}_{\lambda_o/8} \times \hat{D}} \Big|\partial_E \bar{I}_1(E) - \partial_E I_1(E, \hat{I})\Big| \cdot \sup_{\mathcal{I}_{2\mathtt{r}_*} \times \hat{D}} \Big|\partial_{I_1}\mathtt{E}(I) \cdot \partial_{I_1}\bar{\mathtt{E}}(I_1)\Big| \\
&\overset{(84),(200)}{\lessdot} \sqrt{\epsilon}\mu,
\end{aligned}$$

completing the proof. ∎

By Cauchy estimates, from (201), there follows

$$\sup_{\mathcal{I}_{\mathtt{r}_*} \times \hat{D}} |\partial^2_{I_1}\mathtt{E} - \partial^2_{I_1}\bar{\mathtt{E}}| \lessdot \mu. \tag{203}$$

Then, by (157) and (200), on $\mathcal{I}_{\mathtt{r}_*} \times \hat{D}$, we get

$$\begin{aligned}
\det \partial^2_I \mathtt{h} &= (\partial^2_{I_1}\mathtt{E}) \cdot \det \partial^2_{\hat{I}} \hat{h}_k + O(\mu_o) \\
&= (\partial^2_{I_1}\bar{\mathtt{E}} + O(\mu)) \cdot \det \partial^2_{\hat{I}} \hat{h}_k + O(\mu_o).
\end{aligned} \tag{204}$$

Now, by (166), (77), one has that ${\delta_o}^{-1} \leqslant \mathtt{K}^{2n}$ and $\mu_o/\delta_o = O(\mathtt{K}^{-3n})$.

Finally, since, by (49), $\mu = 1/\mathtt{K}^{5n}$, from (204) one gets, at once, the following

Lemma 5.18 *Let* $\mathtt{r}_*$ *be as in Lemma 5.17,* $0 \leqslant i \leqslant 2N$*, and* $\mathcal{I} = \mathcal{I}^i$ *as in (173). Then, for all* $I \in \mathcal{I}_{\mathtt{r}_*} \times \hat{D}$*,*

$$\left|\det \partial_I^2 \mathtt{h}(I)\right| \geqslant \delta_o |g(I)|, \tag{205}$$

with

$$g(I) = \partial_{I_1}^2 \bar{\mathtt{E}}(I_1) + O(\mathtt{K}^{-3n}). \tag{206}$$

We can now proceed to give the

Proof of Proposition 5.16 (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 A. Hence, (85) in Proposition 3.8 holds, so that by (205) and (206), taking $\mathbf{c}_5$ large enough and $\mathtt{K} \geqslant \mathbf{c}_5$, the claimed estimate $|\det \partial_I^2 \mathtt{h}| \geqslant \delta_o / 2^5$ follows.

(ii) Recall (148). Since $\lambda \to \bar{b}_\lambda$ is decreasing, we get $\bar{b}_{\lambda_o/2} \leqslant \bar{b} = \bar{b}_0$. By 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 (94), by (76)÷(79) we have that $\bar{b} < \sqrt{\epsilon}$. Then, choosing $0 < \tilde{\mathtt{r}} \leqslant 1$ small enough, we have that

$$\bar{b}\tilde{\mathcal{I}}_{2\tilde{\mathtt{r}}} \subseteq \mathcal{I}_{\mathtt{r}_*}. \tag{207}$$

By (149) we get

$$\partial_{I_1}^2 \bar{\mathtt{E}}(I_1) = \mathtt{F}_{\bar{\mathtt{G}}}(I_1/\bar{b}). \tag{208}$$

By (206) and (207) we get

$$g(\bar{b}x, \hat{I}) = \mathtt{F}_{\bar{\mathtt{G}}}(x) + O(\mu) + O(\mu_o/\delta_o) \quad \text{uniformly in} \quad (x, \hat{I}) \in \tilde{\mathcal{I}}_{2\tilde{\mathtt{r}}} \times \hat{D}. \tag{209}$$

By (200), (207) and (208) we get

$$\sup_{\tilde{\mathcal{I}}_{2\tilde{\mathtt{r}}}} |\mathtt{F}_{\bar{\mathtt{G}}}| < 1. \tag{210}$$

Recall that $\mathcal{I} := (0, \bar{b}_{\lambda_o/2})$ (see, (173)), and define, for $\hat{I} \in \hat{D}$,

$$\begin{aligned} \mathcal{I}'_\eta(\hat{I}) &:= \left\{I_1 \in \mathcal{I} : \left|\det \partial_I^2 \mathtt{h}^i(I)\right| \leqslant \eta\right\}, \\ \tilde{\mathcal{I}}'_\eta(\hat{I}) &:= \{x \in \tilde{\mathcal{I}} : |g(\bar{b}x, \hat{I})| \leqslant \eta/\delta_o\}. \end{aligned} \tag{211}$$

By (211) and (205) we have that, for every $\hat{I} \in \hat{D}$,

$$\operatorname{meas} \mathcal{I}'_\eta(\hat{I}) \leqslant \bar{b} \operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I}) < \sqrt{\epsilon} \operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat{I}). \tag{212}$$

using $\bar{b} < \sqrt{\epsilon}$.

We, now, claim that

$$1/2 \leqslant \bar b_{\lambda_o/2}/\bar b. \tag{213}$$

To prove (213), recall (94), and observe that

$$\bar b = \bar I_1(\bar E_+) = \int_{\bar E_-}^{\bar E_+} \partial_E \bar I_1 \overset{(81)}{\geqslant} \frac{\bar E_+ - \bar E_-}{\mathbf{c}\sqrt{\epsilon}} \overset{(92)}{\geqslant} \frac{\sqrt{\epsilon}}{\mathbf{c}\mathfrak{g}}. \tag{214}$$

Observe that condition (82) reduces, here, to $\lambda \leqslant 1/\mathbf{c}$, since we are considering $\bar I_1$ (i.e., the case $\mu = 0$). Recalling (173) we have, by (84), that

$$\partial_E \bar I_1(\bar E_+ - \epsilon z) \leqslant \mathbf{c}^2 |\log z|/\sqrt{\epsilon}, \quad \forall\ 0 < z \leqslant \lambda_o/2.$$

Therefore

$$\begin{aligned} 1 - \frac{\bar b_{\lambda_o/2}}{\bar b} &= \frac{\bar I_1(\bar E_+) - \bar I_1(\bar E_+ - \epsilon\lambda_o/2)}{\bar b} = \frac{\epsilon}{\bar b}\int_0^{\lambda_o/2} \partial_E \bar I_1(\bar E_+ - \epsilon z)dz \\ &\overset{(214)}{\leqslant} \mathbf{c}^3\mathfrak{g}\int_0^{\lambda_o/2} |\log z| dz \leqslant \mathbf{c}^4 \lambda_o \log|\lambda_o| \overset{(174)}{\leqslant} \frac{1}{2}, \end{aligned}$$

proving (213).

Let us come back to the estimate of meas $\tilde{\mathcal{I}}'_\eta(\hat I)$. Recalling Definition 5.3, we have that $\mathtt{F}$ is ξ-nondegenerate at order $\mathtt{m}$. Recall the definition of δ_o in (155). By (209), (210), and Cauchy estimates, taking μ and μ_o/δ_o small enough (i.e., $\mathtt{K} \geqslant \hat{\mathfrak{c}}_0$ for a suitable $\hat{\mathfrak{c}}_0$ large enough) depending only on $\mathfrak{g}$, n, ξ and $\mathtt{m}$, we have that the function $x \mapsto g(\bar b x, \hat I)$ is $(\xi/2)$-nondegenerate at order $\mathtt{m}$. Now we want to apply Lemma 5.1 with η replaced by η/δ_o, and with:

$$f(x) = g(\bar b x, \hat I), \quad m = \mathtt{m}, \quad a = 0, \quad 1/2 \overset{(213)}{\leqslant} b = \bar b_{\lambda_o/2}/\bar b < 1, \quad \xi = \xi/2.$$

The constant M in Lemma 5.1, controlling the derivatives of f, using (209), (210), and Cauchy estimates, can be bounded as

$$1 \leqslant M \leqslant c_{n,\mathfrak{g}}/\tilde{\mathtt{r}}^{\mathtt{m}+1},$$

for a suitably large constant $c_{n,\mathfrak{g}}$, depending only on n and $\mathfrak{g}$ (recall that $\tilde{\mathtt{r}} = \tilde{\mathtt{r}}(n, \mathfrak{g})$ was chosen in (207) small enough). In conclusion, by Lemma 5.1, we get

$$\operatorname{meas} \tilde{\mathcal{I}}'_\eta(\hat I) \leqslant c_{\mathtt{m}}\Big(\frac{2c_{n,\mathfrak{g}}}{\xi\tilde{\mathtt{r}}^{\mathtt{m}+1}} + 1\Big)\Big(\frac{\eta}{\delta_o\xi}\Big)^{\frac{1}{\mathtt{m}}}.$$

Then, (199) follows by (166), (212), Fubini's theorem, and (197). The proof of Proposition 5.16 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 (164), and that $\delta_o = |k|^{-2n}$, (see (155)).

Proposition 5.19 *Let $i = 0, 2N$. Then, there exists a suitable constant $\mathbf{c}_7 = \mathbf{c}_7(n, \mathfrak{g}) > 1$, such that, if $\mathtt{K} \geqslant \mathbf{c}_7$, then, on $\mathcal{B}^i_{\rm far}$, $|\delta_\sharp| \geqslant \delta_o/2$.*

Proof Taking $\lambda = \lambda_o/2$ in (99), on $\mathcal{B}^i_{\rm far}$, we have that

$$|\hat{\mathrm{v}}|,\ |\hat{\mathrm{M}}|\ < 1/\mathtt{K}^{\frac{19}{2}n-1}, \tag{215}$$

where, $\hat{\mathrm{v}}$ and $\hat{\mathrm{M}}$ have been defined in (163). By (49) and (77), 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 3.4 and (49), 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 $\bar{\mathtt{E}}^i$ (defined in (68)) is strictly concave: This follows from the following lemma, which is a simple consequence of Jensen's inequality, and whose proof is given in Appendix A.

Lemma 5.20 *Let $i = 0, 2N$. Then, for every $E > \bar{E}_0 = \bar{E}_{2N}$, one has that $\partial^2_{I_1}\bar{\mathtt{E}}^i(\bar{I}^i_1(E)) \geqslant 2$.*

Now, since estimate (203) still holds in the present case $i = 2N$ we get by Lemma 5.20, $\partial^2_{I_1}\mathtt{E} \geqslant \frac{1}{2}\partial^2_{I_1}\bar{\mathtt{E}} \geqslant 1$, so that the claim follows by (164), (215) and (166). ∎

Step 7: Conclusion of the proof of the Twist Theorem

Let $\lambda_o = \lambda_o(n, \mathfrak{g})$ be as in Proposition 5.9-(ii), and let $\mathbf{c}_8 = \mathbf{c}_8(n, \mathfrak{g}) \geqslant 1$ be such that the second estimate in (174) holds for $\mathtt{K} = 1/\mu^{5n} \geqslant \mathbf{c}_8$ (recall (49)). Then, by Lemma 5.8,

$$\mathcal{B}^i_k = \mathcal{B}^i_{\rm near}(\lambda_o) \cup \mathcal{B}^i_{\rm far}(\lambda_o), \qquad \forall\ 0 \leqslant i \leqslant 2N. \tag{216}$$

Recall that $\mathbf{c}_2 \geqslant \mathbf{c}_4 \geqslant \max\{\mathbf{c}_1, \mathbf{c}_3\} \geqslant c_3$, and define:

$$\mathfrak{c}_0 := 2\max\{\mathbf{c}_2, \mathbf{c}_5, \mathbf{c}_6, \mathbf{c}_7, \mathbf{c}_8, \hat{\mathfrak{c}}_0\}. \tag{217}$$

Let us consider first the outer case $i = 0, 2N$. Recall the definition of $\mathtt{b}$ in (156). By Lemma 5.5, Proposition 5.19, Proposition 5.13, and by (217), we find

$$\begin{aligned}\mathrm{meas}\{\, I \in B^i_k : |\det \partial^2_I \mathtt{h}^i_k(I)| \leqslant \eta\} &= \mathrm{meas}\{\, I \in \mathcal{B}^i_k : |\delta^i_\sharp(I)| \leqslant \eta\} \\ &= \mathrm{meas}\{I \in \mathcal{B}^i_{\text{near}}(\lambda_\mathrm{o}) : |\det \partial^2_I \mathtt{h}^i(I)| \leqslant \eta\} \\ &\leqslant \mathbf{c}_2\, (|k|^{2n}\eta)^{1/9n^4}\ \mathrm{meas}\, \mathcal{B}^i_k \\ &\overset{(165)}{=} \mathbf{c}_2\, (|k|^{2n}\eta)^{1/9n^4}\ \mathrm{meas}\, B^i_k \\ &\overset{(217)}{<} \mathfrak{c}_0\, (|k|^{2n}\eta)^{\mathtt{b}}\ \mathrm{meas}\, B^i_k,\end{aligned}$$

proving Theorem 5.4 in the outer case $i = 0, 2N$.

In the inner case $0 < i < 2N$, $B^i_k = \mathcal{B}^i_k$ (compare (110)) and, since $\mathtt{K} \geqslant \mathfrak{c}_0$, (156) follows by (216), (190) in Proposition 5.13, and (199) in Proposition (5.16). The proof of Theorem 5.4 is complete.

6 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 2.

To construct such tori we shall use the following "KAM theorem".

Theorem 6.1 ([12]) *Fix $n \geqslant 2$ and let $\mathtt{D}$ be any non-empty, bounded, measurable set in $\mathbb{R}^n$. 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

$$\mathtt{M} := |\partial^2_\mathrm{p} \mathrm{h}|_\mathfrak{r}, \qquad |\mathrm{f}|_{\mathfrak{r},\mathfrak{s}}. \tag{218}$$

Assume that the frequency map $\mathrm{p} \in \mathtt{D} \to \omega = \partial_\mathrm{p} \mathrm{h}$ is a local diffeomorphism, namely, assume:

$$\mathrm{d} := \inf_{\mathtt{D}} |\det \partial^2_\mathrm{p} \mathrm{h}| > 0, \tag{219}$$

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

$$\epsilon_* := \frac{|\mathrm{f}|_{\mathfrak{r},\mathfrak{s}}}{\mathtt{M}\mathfrak{r}^2} \leqslant \frac{\mathrm{d}^8_*\ \mathfrak{s}^{4(\mathrm{n}+1)}}{\mathrm{C}_*}\ , \tag{220}$$

there exists a set $\mathcal{T} \subseteq (\mathtt{D}_{\mathrm{r}_*} \cap \mathbb{R}^{\mathrm{n}}) \times \mathbb{T}^{\mathrm{n}}$, which is a union of primary KAM tori, whose measure can be bounded as

$$\mathrm{meas}\left((\mathtt{D} \times \mathbb{T}^n) \setminus \mathcal{T}\right) \leqslant \mathrm{C}\sqrt{\epsilon_*}, \tag{221}$$

with:

$$\mathrm{C} := \left(\max\left\{\mathrm{d}_*^2 \mathfrak{r}\,, \mathrm{diam}\,\mathtt{D}\right\}\right)^{\mathrm{n}} \cdot \frac{\mathrm{C}_*}{\mathrm{d}_*^{n+5}\ \mathfrak{s}^{3(\mathrm{n}+1)}}.$$

This statement is an immediate corollary of the main result in [12]: In Theorem 1 of [12] take $\tau = n$ and substitute λ with its maximal value $2 \cdot n!\,\mathrm{d}_*^{-1}$ (compare (14) of [12]).

Remark 6.1 *(i) Note that in the formulation of Theorem 6.1 the action domain $\mathtt{D}$ is a completely arbitrary bounded measurable set and that the smallness quantitative condition (220) 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 [21].*

(ii) We point out that the smallness condition (220) can be rewritten as

$$|\mathrm{f}|_{\mathtt{D},\mathfrak{r},\mathfrak{s}} \leqslant \frac{\mathfrak{r}^2 \mathrm{d}^8\ \mathfrak{s}^{4n+4}}{\mathrm{C}_*\,\mathtt{M}^{8n-1}}. \tag{222}$$

(iii) Finally, observe that, since the absolute value of the eigenvalues of the symmetric matrix $\partial_{\mathrm{p}}^2 \mathrm{h}$ are bounded by $\mathtt{M}$, one has

$$\mathrm{d} \leqslant \sup_{\mathtt{D}} |\det \partial_{\mathrm{p}}^2 \mathrm{h}| \leqslant \mathtt{M}^n,$$

so that $\mathrm{d}_* \leqslant 1$. Therefore, estimate (221) implies

$$\mathrm{meas}\left((\mathtt{D} \times \mathbb{T}^n) \setminus \mathcal{T}\right) \leqslant \left(\max\left\{\mathfrak{r}\,, \mathrm{diam}\,\mathtt{D}\right\}\right)^n \cdot \frac{\mathrm{C}_*\,\mathtt{M}^{n^2+5n-1/2}}{\mathrm{d}^{n+5}\ \mathfrak{s}^{3n+3}\mathfrak{r}} \sqrt{|\mathrm{f}|_{\mathtt{D},\mathfrak{r},\mathfrak{s}}}. \tag{223}$$

6.1 KAM tori in the nonresonant region

Proposition 6.2 *Let the assumptions of Theorem 3.3 hold. There exists a constant $\mathtt{C}_{\mathrm{o}} = \mathtt{C}_{\mathrm{o}}(n,s) \geqslant \mathbf{c_o}$ such that, if $\mathtt{K}_{\mathrm{o}} \geqslant \mathtt{C}_{\mathrm{o}}$, then there exists a family of primary maximal KAM tori $\mathcal{T}^0$ invariant for the Hamiltonian $\mathtt{H}$ in (1), satisfying*

$$\mathrm{meas}\left((\mathcal{R}^0 \times \mathbb{T}^n) \setminus \mathcal{T}^0\right) \leqslant \mathtt{C}_{\mathrm{o}}\, e^{-\mathtt{K}_{\mathrm{o}} s/6}. \tag{224}$$

Remark 6.2 *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 3.3, which is introduced so that* (41) *holds; compare, also, Remark 3.2-(ii).*

Proof of Proposition 6.2 we apply the KAM Theorem 6.1 to the nearly integrable Hamiltonian $\mathtt{H}_{\rm o}$ in Theorem 3.3-(ii). More precisely, recall Theorem 3.3, definitions (20), (32), (21), and let

$$\mathrm{h}(\mathrm{p}) = \frac{|\mathrm{p}|^2}{2} + \varepsilon g^{\rm o}(\mathrm{p}), \quad \mathrm{f} = \varepsilon f^{\rm o}, \quad \mathtt{D} = \widetilde{\mathcal{R}}^0,$$

with

$$\mathfrak{r} = \frac{r_{\rm o}'}{2} = \frac{\sqrt{\varepsilon}\mathrm{K}^{\frac{9}{2}n+2}}{16\mathrm{K}_{\rm o}}, \qquad \mathfrak{s} = \min\{\tfrac{s}{2}, 1\}.$$

By (36) and Cauchy estimates we get

$$\mathtt{M} \leqslant 2, \qquad |\mathrm{f}|_{\mathfrak{r},\mathfrak{s}} \leqslant \varepsilon e^{-\mathrm{K}_{\rm o} s/3}, \qquad \mathrm{d} \geqslant 1/2.$$

Notice that the hypothesis $\mathtt{K} < \varepsilon^{-1/(9n+4)}$ implies that $\mathfrak{r} < 1$, so that

$$\max\left\{\mathrm{d}^2\mathtt{M}^{-2n}\mathfrak{r}\,, \mathrm{diam}\,\mathtt{D}\right\} = 2.$$

Now, if $\mathtt{K}_{\rm o}$ is taken large enough (larger than a constant despending on n and s), then the KAM smallness condition (222) is satisfied, and the KAM Theorem 6.1 yields the existence of a set $\widetilde{\mathcal{T}}^0$ of invariant tori for the Hamiltonian $\mathtt{H}_{\rm o}$ in Theorem 3.3-(ii), which, by (223), satisfies

$$\mathrm{meas}\left((\widetilde{\mathcal{R}}^0 \times \mathbb{T}^n)\setminus \widetilde{\mathcal{T}}^0\right) \leqslant \mathtt{C}_{\rm o}\, e^{-\mathrm{K}_{\rm o} s/6}, \tag{225}$$

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 (41) 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 (41) and the bound (225) imply (224). ∎

6.2 KAM tori around 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 4.1, for all $k \in \mathcal{G}^n_{\mathtt{K}_o}$ 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 (113), (110)), let us define:

$$B_k^i(\lambda, \eta) := \{I \in B_k^i(\lambda) : |\det \partial_I^2 \mathtt{h}_k^i(I)| > \eta\} \subset B_k^i. \tag{226}$$

Proposition 6.3 (KAM tori for $\mathcal{H}_k^i$)
Let the assumptions of Theorem 4.1 hold. There exist constants $\bar{\mathtt{C}}_1 = \bar{\mathtt{C}}_1(n, s, \beta) > 1$ and $\mathtt{C}_1 = \mathtt{C}_1(n, s, \beta, \delta) \geqslant \mathbf{c}_\star$, such that the following holds. Let $k \in \mathcal{G}^n_{\mathtt{K}_o}$, $0 \leqslant i \leqslant 2N_k$; $0 < \lambda \leqslant 1/\mathbf{c}_\star$ and $0 < \eta < 1/2$. Then, if

$$\mathtt{K} \geqslant \mathtt{C}_1 \log \frac{1}{\lambda \eta}, \tag{227}$$

there exists a set $\mathcal{T}_k^i$ of maximal KAM tori for the Hamiltonian $\mathcal{H}_k^i$ in (112) such that

$$\mathrm{meas}\left((B_k^i(\lambda, \eta) \times \mathbb{T}^n) \setminus \mathcal{T}_k^i\right) \leqslant \bar{\mathtt{C}}_1 e^{-\mathtt{K}s/7}. \tag{228}$$

Proof We apply the KAM Theorem 6.1 to the Hamiltonian $\mathcal{H}_k^i$ of Theorem 4.1 with (recall (49), (112) and (113)):

$$\begin{aligned} &\mathrm{h} = h_k^i = \tfrac{|k|^2}{2} \mathtt{h}_k^i, \qquad \mathrm{f} = \varepsilon f_k^i, \qquad \mathtt{D} = B_k^i(\lambda, \eta), \\ &\mathfrak{r} = \rho_\star = \frac{\sqrt{\epsilon}}{\mathbf{c}_\star \mathtt{K}_o^n} \lambda |\log \lambda|, \qquad \mathfrak{s} = \sigma_\star = \frac{1}{\mathbf{c}_\star \mathtt{K}_o^n |\log \lambda|}. \end{aligned} \tag{229}$$

Note that, by (116) and (56), $0 < \lambda \leqslant 1/\mathbf{c}_\star \leqslant 1/8c_2$, which implies easily $\mathfrak{r} \leqslant \mathtt{r}$ and $\mathfrak{s} \leqslant 1$. Also, since $\mathbf{c}_\star \geqslant \hat{\mathbf{c}}$ (see Theorem 4.1) and $\mathtt{K}_o^n \geqslant 2^n \geqslant n$, one has $\rho_\star \leqslant \rho_\lambda / n$.

In the following, we denote by $c(\cdot)$ possibly different constants depending only on the quantities inside the brackets.

We first have to estimate $\mathtt{M}$ in (218), namely, $\partial_I^2 \mathtt{h}_k^i$. By (99), (77) and (75) we get

$$\sup_{(\mathcal{B}_k^i(\lambda))_{\rho_\lambda}} \left|\partial_I^2 \mathtt{E}^i\right| \leqslant \frac{n\hat{\mathbf{c}}}{\lambda}. \tag{230}$$

In the case $0 < i < 2N_k$, by (113), we have $\mathcal{B}_k^i(\lambda) = B_k^i(\lambda)$. Therefore, recalling (112), 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 $j_{\mathtt{g}_\star}$ in (108), (110) we have that $\partial_I^2 j_{\mathtt{g}_\star}$ depends only on $\hat{I}$ and not on I_1. Moreover by (20), (33), (59), (49), and Cauchy estimates, we get

$$\sup_{\hat{I}\in\hat{D}_{3\mathtt{r}}} |\partial_I j_{\mathtt{g}_\star}| \leqslant c(n), \qquad \sup_{\hat{I}\in\hat{D}_{3\mathtt{r}}} |\partial_I^2 j_{\mathtt{g}_\star}| \leqslant \frac{c(n)|k|^2}{\sqrt{\varepsilon}\mathtt{K}^\nu}. \tag{231}$$

Recalling Definition 3.4, (99), and (49), we have that

$$\sup_{(\mathcal{B}_k^{2N_k}(\lambda))_{\rho_\lambda}} |\mathtt{E}^{2N_k}| \leqslant 4\mathtt{R}^2 = \frac{4\varepsilon\mathtt{K}^{2\nu}}{|k|^4}.$$

Recall that $\mathtt{K} > \mathtt{K}_{\rm o} > \mathtt{c}_s$ (see (20), (42)). Then, by (99) and (49), we get

$$\sup_{(\mathcal{B}_k^{2N_k}(\lambda))_{\rho_\lambda}} |\partial_{I_1}\mathtt{E}^{2N_k}| \leqslant \hat{\mathtt{c}}\sqrt{8\mathtt{c}_s\varepsilon + 4\varepsilon\mathtt{K}^{2\nu}|k|^{-4}} \leqslant 4\hat{\mathtt{c}}\sqrt{\varepsilon}\mathtt{K}^\nu|k|^{-2}.$$

Finally, recalling also (113), (126), (230), (231), 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 (218), (229), (112), (166) (and that $\mathfrak{r} \leqslant \mathtt{r}$), we finally get

$$\mathtt{M} \leqslant |k|^2\frac{c(n,s,\beta)}{\lambda}, \qquad \forall\ 0 \leqslant i \leqslant 2N_k. \tag{232}$$

Next, by (229) and (115),

$$|\mathrm{f}|_{\mathfrak{r},\mathfrak{s}} \leqslant \varepsilon\, e^{-\mathtt{K}s/3}. \tag{233}$$

By (219), (229) and (226), we get

$$\mathrm{d} \geqslant 2^{-n}|k|^{2n}\eta \qquad \text{and} \qquad \frac{\mathtt{M}^n}{\mathrm{d}} \leqslant \frac{c(n,s,\beta)}{\lambda^n\eta}\ . \tag{234}$$

By (49), (16) and using $\mathtt{K}_{\rm o} \leqslant 6\mathtt{K}$, we have:

$$\frac{\varepsilon}{\epsilon} \leqslant \frac{\mathtt{K}_{\rm o}^{n+2}}{8\mathtt{c}_s\delta}\ e^{\mathtt{K}_{\rm o}s} \leqslant \frac{\mathtt{K}^{n+2}}{6^{n+3}\mathtt{c}_s\delta}\ e^{\mathtt{K}s/6}. \tag{235}$$

It is now easy to check, by (232), (233), (234) and (235), that the KAM smallness condition (222) is satisfied, provided (227) holds with $\mathtt{C}_1$ large enough. By the KAM

Theorem 6.1 we get a set $\mathcal{T}_k^i$ of invariant tori for the Hamiltonian in (112), which, in view of (223) and by (232), (233), (234) and (235), satisfies (228) with a suitable constant $\bar{\mathtt{C}}_1 = \bar{\mathtt{C}}_1(n, s, \beta)$. In particular, note that, by (142) and (229), the maximum in (223) is estimated by $c(n)\mathtt{K}_{\rm o}^{n^2}$. ∎

Putting together these KAM statements and the Twist Theorem 5.4, the proof of the results stated in Section 2 follow easily.

6.3 Proof of the main theorem and its corollaries

By Lemma 2.5, since f belongs to $\mathbb{G}_s^n$, there exist $\delta, \beta > 0$ such that (16) and (17) hold with $\mathtt{N}$ as in (15). Let

$$\mathtt{K}_{\rm o} := \mathtt{K}/6,$$

with $\mathtt{K} \geqslant 12$ and let α be as in (20). Then, Assumptions 3.1 holds. Let $c_2 = c_2(n)$ and $\mathbf{c_o} = \mathbf{c_o}(n, s, \delta)$ be as in Theorem 3.4, and assume that

$$\mathtt{K} \geqslant 6 \max\{c_2, \mathbf{c_o}\}. \tag{236}$$

Notice that this relation implies that $\mathtt{K} > 12$. Then, Theorem 3.4 holds, and we may define the parameters $\xi > 0$ and $\mathtt{m} \geqslant 1$ as in Definition 5.3 with respect to standard Hamiltonians $\mathtt{H}_k$ (with $|k|_1 \leqslant \mathtt{K}_{\rm o}$) of Theorem 3.4-(ii).

We now let $\mathtt{b} < 1$ as in (156), $\mathtt{C}_1 = \mathtt{C}_1(n, s, \beta, \delta)$ be as in Proposition 6.3, and define

$$\eta := e^{-\frac{\mathtt{K}}{\mathtt{C}_1(1+\mathtt{b})}}, \qquad \lambda := \eta^{\mathtt{b}}. \tag{237}$$

Notice that, with such definitions, it is (compare (227)):

$$\mathtt{K} = \mathtt{C}_1 \log \frac{1}{\lambda \eta}. \tag{238}$$

With these premises, let us turn to the proof of the claims of Theorem 2.1.

Claim (ii) has already been proven in Lemma 3.1.

Now, we are ready to define the family of maximal KAM tori $\mathcal{T}$ for $\mathtt{H}$, appearing in item (iv) of Theorem 2.1.

Let $\mathtt{C}_{\rm o} = \mathtt{C}_{\rm o}(n, s)$ as in Proposition 6.2. There exists a constant

$$\hat{\mathfrak{c}} = \hat{\mathfrak{c}}(n, s, \beta, \delta, \mathtt{m}) \geqslant \max\{\mathtt{C}_{\rm o}, 2\mathtt{C}_1 \mathbf{c}_\star/\mathtt{b}\},$$

such that, if $\mathtt{K} \geqslant \hat{\mathfrak{c}}$, then

$$\mathtt{K}^{2n}\eta \overset{(237)}{=} \mathtt{K}^{2n} e^{-\frac{\mathtt{K}}{\mathtt{C}_1(1+\mathtt{b})}} \leqslant 1.$$

Assume that

$$\mathtt{K} \geqslant \hat{\mathtt{c}}. \tag{239}$$

Then, $\lambda = \eta^{\mathtt{b}}$ in (237) is smaller than $1/\mathbf{c}_\star$ and (recall (155))

$$\eta \leqslant \frac{\delta_{\mathrm{o}}}{2^5} < \frac{1}{2}. \tag{240}$$

Thus, in view of (238), by (239), the assumptions in Propositions 6.2 and 6.3 are satisfied. Thus, we can define the following families of tori:

$$\begin{aligned} &\mathcal{T}_i^{1,k} := \phi_k^i(\mathcal{T}_k^i), \quad \mathcal{T}^{1,k} := \bigcup_{0\leqslant i\leqslant 2N_k} \mathcal{T}_i^{1,k}, \quad \mathcal{T}^1 := \bigcup_{k\in\mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}} \mathcal{T}^{1,k}, \\ &\mathcal{T} := \mathcal{T}^0 \cup \mathcal{T}^1, \end{aligned} \tag{241}$$

where $\mathcal{T}_k^i$ is defined in Proposition 6.3, $\mathcal{T}^0$ in Proposition 6.2 and ϕ_k^i in (114).

Observe that $\mathcal{T}_k^i$ are invariant tori for $\mathcal{H}_k^i$ in (112), 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 $\mathtt{H}$.

Claim (i) follows, now, immediately by (23), if one defines:

$$\mathcal{A} := \big((\mathcal{R}^0 \cup \mathcal{R}^1) \times \mathbb{T}^n\big) \setminus \mathcal{T}. \tag{242}$$

It remains to prove (iii), namely, the exponential measure estimate on $\mathcal{A}$.

Observe that, by (242) and (241), one has:

$$\begin{aligned} \mathcal{A} &\subseteq \big((\mathcal{R}^0 \times \mathbb{T}^n)\setminus \mathcal{T}^0\big) \cup \big((\mathcal{R}^1 \times \mathbb{T}^n)\setminus \mathcal{T}^1\big) \\ &\subseteq \big((\mathcal{R}^0 \times \mathbb{T}^n)\setminus \mathcal{T}^0\big) \cup \bigcup_{k\in\mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}} (\mathcal{R}^{1,k} \times \mathbb{T}^n)\setminus \mathcal{T}^{1,k}. \end{aligned} \tag{243}$$

We now need the following elementary result, whose proof is given in Appendix A.

Lemma 6.4 *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 $\bar{\mathtt{c}} := \max\{4, \pi\sqrt{8/\beta}\}$.*

Obviously, the hypothesis of this lemma are met by our fixed potential in $\mathbb{G}^n_s$, and the following measure estimate holds.

Lemma 6.5 *Let λ be as in (237) and $\bar{\mathtt{c}}$ be as in Lemma 6.4. Then, for any k in $\mathcal{G}^n_{\mathtt{K}_{\mathrm{o}}}$, one has*

$$\begin{aligned} \mathrm{meas}\,\big((\mathcal{R}^{1,k} \times \mathbb{T}^n)\setminus \mathcal{T}^{1,k}\big) \leqslant &\mathbf{c}_\star\, \mathrm{meas}\,\big(\widetilde{\mathcal{R}}^{1,k} \times \mathbb{T}^n\big)\lambda|\log\lambda| \\ &+ \bar{\mathtt{c}} \max_{0\leqslant i\leqslant 2N_k} \mathrm{meas}\,\Big(\big(B_k^i(\lambda) \times \mathbb{T}^n\big)\setminus\mathcal{T}_k^i\Big). \end{aligned} \tag{244}$$

Proof Since ϕ_k^i in Theorem 4.1 is a diffeomorphism, one has (i runs from 0 to $2N_k$):

$$(\mathcal{R}^{1,k}\times\mathbb{T}^n)\setminus\mathcal{T}^{1,k} \overset{(241)}{=} (\mathcal{R}^{1,k}\times\mathbb{T}^n)\setminus\Big(\bigcup_i\phi_k^i(\mathcal{T}_k^i)\Big)$$
$$\subseteq\Big(\big(\mathcal{R}^{1,k}\times\mathbb{T}^n\big)\setminus\bigcup_i\phi_k^i\big(B_k^i(\lambda)\times\mathbb{T}^n\big)\Big)\cup\bigcup_i\phi_k^i\Big(\big(B_k^i(\lambda)\times\mathbb{T}^n\big)\setminus\mathcal{T}_k^i\Big),$$

then, passing to measures, using (138), the fact that ϕ_k^i is symplectic and Lemma 6.4, we get (244). ∎

We arrived at the end:

Proof of Theorem 2.1 Now, assume that, together with (236) and (239), it is also $\mathtt{K}\geqslant\mathfrak{c}_0$. Then, recalling (240), Theorem 5.4 holds. Thus, recalling (226), observing that

$$B_k^i(\lambda) \;=\; \big\{I\in B_k^i : |\det\partial_I^2\mathtt{h}_k^i(I)|\leqslant\eta\big\}\cup B_k^i(\lambda,\eta),$$

by (156) and (228) we get

$$\text{meas}\,\big((B_k^i(\lambda)\times\mathbb{T}^n)\setminus\mathcal{T}_k^i\big) \;\leqslant\; \mathfrak{c}_0(|k|^{2n}\eta)^{\mathtt{b}}\,\text{meas}\,B_k^i \;+\; \bar{\mathtt{C}}_1 e^{-\mathtt{K}s/7}. \tag{245}$$

Now, by (243), (224), (244), (245), (142), (237) and since $|k|_1\leqslant\mathtt{K}_{\rm o}=\mathtt{K}/6$, we get, for a suitable constant $\mathfrak{c}_1=\mathfrak{c}_1(n,s,\delta,\beta,\xi,\mathtt{m})$,[14]

$$\text{meas}(\mathcal{A})\leqslant\mathfrak{c}_1\,\mathtt{K}^{2n}e^{-\mathtt{K}/\mathfrak{c}_\star},\qquad \mathfrak{c}_\star:=\max\big\{36/s,2\mathtt{C}_1/\mathtt{b}\big\}. \tag{246}$$

Now, let

$$\mathfrak{c}=\mathfrak{c}(n,s,\delta,\beta,\xi,\mathtt{m})\geqslant 1+\mathfrak{c}_\star \tag{247}$$

be such that, if $\mathtt{K}\geqslant\mathfrak{c}$, then $\mathfrak{c}_1\,\mathtt{K}^{2n}e^{-\mathtt{K}/\mathfrak{c}_\star}\leqslant e^{-\mathtt{K}/(1+\mathfrak{c}_\star)}$. Then, if $\mathtt{K}\geqslant\mathfrak{c}$, claim (iii) follows, and the proof of Theorem 2.1 is complete. ∎

Remark 6.3 *Notice that $\mathcal{T}^0$ is a family of maximal primary tori for $\mathtt{H}$, and so are the families $\mathcal{T}_i^{1,k}$ for all $k\in\mathcal{G}^n_{\mathtt{K}_{\rm o}}$ and $i=0,2N_k$. On the other hand, $\mathcal{T}_i^{1,k}$ 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.*

[14]Remember the following facts: $\varepsilon\leqslant\mathtt{K}^{-\gamma}<1$ (compare (11)); meas $\big(\widetilde{\mathcal{R}}^{1,k}\times\mathbb{T}^n\big)\leqslant c(n)$, which holds since $\widetilde{\mathcal{R}}^{1,k}\subset\{y:|y|\leqslant 2\}$; $|\log\lambda|=\frac{\mathtt{b}}{\mathtt{C}_1(1+\mathtt{b})}\mathtt{K}$; meas $B_k^i\leqslant c_*$, which holds by (142); $\#\mathcal{G}^n_{\mathtt{K}_{\rm o}}<(2\mathtt{K}_{\rm o}+1)^n$.

Proof of Corollary 2.2 As already pointed out in § 2, the thesis follows trivially from Theorem 2.1 and the measure estimate (13), by taking

$$K := \mathfrak{c}|\log \varepsilon|, \quad \bar{\mathfrak{c}} := 1 + (2\pi)^n c_\star \mathfrak{c}^\gamma,$$

and ε_o so that $\varepsilon \mathtt{K}^\gamma < 1$ for $\varepsilon < \varepsilon_o$. ∎

Proof of Corollary 2.3 Let $n = 2$. We claim that $\mathcal{R}^2$ in (23) satisfies

$$\mathcal{R}^2 \subseteq \{y \in \mathbb{R}^2 : |y| < \varepsilon^{a/2}\}. \tag{248}$$

Fix $y \in \mathcal{R}^2$. Then $|y| < 1$ and there exists $k \in \mathcal{G}^n_{\mathtt{K}_o}$ 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

$$|\pi_k^\perp y \cdot \ell| \leqslant \frac{6\alpha \mathtt{K}}{|k|}. \tag{249}$$

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 (249), $|\pi_k^\perp y| \leqslant 6\alpha\mathtt{K}$. In conclusion

$$|y| = |\pi_k y + \pi_k^\perp y| < 7\alpha \mathtt{K} \overset{(20)}{=} 7\sqrt{\varepsilon}\mathtt{K}^{12}. \tag{250}$$

Now, let $\hat{a} := (1-a)/24$ and $\mathtt{K} := 1/(\sqrt[12]{7}\varepsilon^{\hat{a}})$. Then, (248) follows by (250).

Finally, let $\varepsilon_o < 1$ be so small that $\varepsilon\mathtt{K}^\gamma < 1$ is satisfied for any $\varepsilon < \varepsilon_o$. Then, by the estimate in Theorem 2.1-(iii), we get

$$\text{meas}\,\mathcal{A} \leqslant \text{meas}\left(\left(\{\varepsilon^{a/2} < |y| < 1\} \times \mathbb{T}^n\right) \setminus \mathcal{T}\right) \leqslant e^{-\frac{1}{7^{1/12}\mathfrak{c}\varepsilon^{\hat{a}}}} < e^{-\frac{1}{2\mathfrak{c}\varepsilon^{\hat{a}}}},$$

completing the proof of Corollary 2.3. ∎

A Proofs of elementary lemmata

Proof of Lemma 2.5

Assume $f \in \mathbb{G}^n_s$ for some $s > 0$. Then $f \in \mathbb{B}^n_s$. Fix a $0 < \delta_0 \leqslant 1$ smaller than

$$\liminf_{\substack{|k|_1 \to +\infty \\ k \in \mathcal{G}^n}} |f_k| e^{|k|_1 s} |k|_1^n > 0.$$

Then, there exists N_0 such that $|f_k| > \delta_0 |k|_1^{-n} e^{-|k|_1 s}$, for any $|k|_1 \geqslant N_0$, $k \in \mathcal{G}^n_{\mathtt{K}_o}$. 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 $k \in \mathcal{G}^n_{\mathtt{K}_o}$, (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 [17] (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*

$$\pi_{\mathbb{Z}k} f(\theta) = 2|f_k|\big(\cos(\theta + \theta_k) + F^k_\star(\theta)\big), \quad F^k_\star(\theta) := \frac{1}{2|f_k|} \sum_{|j| \geqslant 2} f_{jk} e^{\mathrm{i} j \theta}, \tag{251}$$

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

$$\frac{1}{2|f_k|} \sum_{|j|=1} f_{jk} e^{\mathrm{i} j \theta} = \mathrm{Re}\left(\frac{f_k}{|f_k|} e^{\mathrm{i}\theta}\right) = \mathrm{Re}\, e^{\mathrm{i}(\theta + \theta_k)} = \cos(\theta + \theta_k),$$

which is equivalent to (251). Now, since $\|f\|_s \leqslant 1$ (so that $|f_k| \leqslant e^{-|k|_1 s}$), one finds, for $|k|_1 \geqslant \mathtt{N}$,

$$\begin{aligned}
|F^k_\star|_1 &\overset{(251)}{\leqslant} \frac{1}{2|f_k|} \sum_{|j| \geqslant 2} |f_{jk}| e^{|j|} \overset{(16)}{\leqslant} \frac{|k|_1^n e^{|k|_1 s}}{2\delta} \sum_{|j| \geqslant 2} |f_{jk}| e^{|j|} \\
&\leqslant \frac{|k|_1^n e^{|k|_1 s}}{2\delta} \sum_{|j| \geqslant 2} e^{-|j|(|k|_1 s - 1)} \\
&\leqslant \frac{2e^2 |k|_1^n}{\delta}\, e^{-|k|_1 s} = \frac{2^{n+1} e^2}{s^n \delta} e^{-\frac{|k|_1 s}{2}} \quad \Big(\frac{|k|_1 s}{2}\Big)^n e^{-\frac{|k|_1 s}{2}} \\
&\leqslant \Big(\frac{2n}{es}\Big)^n \frac{2e^2}{\delta}\, e^{-\frac{|k|_1 s}{2}} \leqslant 2^{-40},
\end{aligned}$$

where last inequality follows since $|k|_1 \geqslant \mathtt{N}$ (see (15)). ∎

The second lemma is an elementary calculus result on functions C^2-close to a cosine. Denote by $\|F\|_{C^2}$ the C^2-norm $\max_{0\leqslant k\leqslant 2}\sup|F^{(k)}|$.

Lemma A.2 *Let $F \in C^2(\mathbb{T}, \mathbb{R})$. Let $\bar\theta$ and $\mathtt{c} < \frac{1}{2}$ be positive numbers such that $\|F - \cos(\theta + \bar\theta)\|_{C^2} \leqslant \mathtt{c}$. 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 $\theta = 0$ (observe 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 $F = g + \cos\theta$, 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. Now, 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. ∎

We can now conclude the proof of Lemma 2.5. Assume that (16) and (17) hold for some $\delta \in (0, 1]$ and $\beta > 0$. Obviously, (9) follows at once from (16). Moreover, if $k \in \mathcal{G}^n$ with $|k|_1 > \mathtt{N}$, by Lemma A.1 and Lemma A.2, since by Cauchy estimates, $\|F\|_{C^2} \leqslant 2|F|_1$, also Eq. (10) follows. ∎

Proof of Lemma 3.5 First note that, by (43), (44), and (45),

$$(1 - \mu)p_1^2 - (1 + \mu)2^{-16}\mathtt{r}^2 \leqslant \mathtt{H}_\flat(p, q_1) \leqslant (1 + \mu)(p_1^2 + 2^{-16}\mathtt{r}^2). \tag{252}$$

By the first inequality in (252), 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 $(\mathtt{R} + \mathtt{r}/2)^2$, in view of (66) and (45). This proves the second inclusion in (67). By the second inequality in (252), 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 (45)). This implies also the first inclusion in (67), completing the proof. ∎

Proof of Lemma 4.3 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

$$w(z) := \big(J_0 + \tfrac{z}{\zeta}(J - J_0), \operatorname{Re}\psi + \mathrm{i}\tfrac{z}{\zeta}\operatorname{Im}\psi\big),$$

with $\operatorname{Re}\psi := (\operatorname{Re}\psi_1, \dots, \operatorname{Re}\psi_n)$, and $\operatorname{Im}\psi := (\operatorname{Im}\psi_1, \dots, \operatorname{Im}\psi_n)$. Notice 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, $|\operatorname{Im} G| \leqslant \xi$, for all $|z| < 1$. 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 harmonic functions (compare, e.g., Theorem 2.10 in [28]), 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 $w(0) = (J_0, \operatorname{Re}\psi) \in D \times T^n$, and g is real analytic, we have that $G(0) = g(J_0, \operatorname{Re}\psi) \in \mathbb{R}$. Then, for any $0 < \zeta \leqslant 1/2$, by the the mean value theorem, we have that

$$|\operatorname{Im} g(J,\psi)| = |\operatorname{Im} G(\zeta)| = |\operatorname{Im} G(\zeta) - \operatorname{Im} G(0)| \leqslant |G(\zeta) - G(0)| \leqslant 8\zeta\xi.$$

The proof is finished. ∎

Proof of Lemma 5.7

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 5.7 follows, as $(1-\lambda)^d \geqslant 1 - d\lambda$. ∎

Proof of Lemma 5.20

First observe that the cases $i = 0$ and $i = 2N$ are identical since

$$\bar{I}^0_1(E) = \bar{I}^{2N}_1(E), \qquad \bar{\mathtt{E}}^0(I_1) = \bar{\mathtt{E}}^{2N}(I_1).$$

Let us then consider the case $i = 2N$. By definition,

$$\bar{I}_1^{2N}(E) = \frac{1}{2\pi}\int_0^{2\pi}\sqrt{E - \bar{\mathtt{G}}(x)}dx, \tag{253}$$

so that, by Jensen's inequality,

$$\begin{aligned}(2\partial_E \bar{I}_1^i(E))^3 &= \Big(\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{\sqrt{E-\bar{\mathtt{G}}(x)}}dx\Big)^3 \leqslant \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{(E-\bar{\mathtt{G}}(x))^{3/2}}dx\\ &= -4\partial_E^2\bar{I}_1^i(E).\end{aligned}$$

The claim, now, follows from (145). ∎

Proof of Lemma 6.4 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 a β-Morse function, 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 Lemma A.3 follows. ∎

We may conclude the proof of Lemma 6.4. 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. ∎

Acknowledgements

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.
Finally, the authors gratefully acknowledge the referees for their insightful comments and suggestions.

References

[1] V. I. Arnol'd. Proof of A. N. Kolmogorov's theorem on the conservation of conditionally periodic motions with a small variation in the Hamiltonian. *Russian Mathematical Surveys*, 18(5):9–36, 1963.

[2] V. I. Arnol'd. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspekhi Matematicheskikh Nauk*, 18(6):91–192, 1963. English transl.: Russian Mathematical Surveys 18 (1963), 85–191.

[3] V. I. Arnol'd. Instability of dynamical systems with many degrees of freedom. *Doklady Akademii Nauk SSSR*, 156:9–12, 1964.

[4] V. I. Arnol'd, V. V. Kozlov, and A. I. Neishtadt. *Dynamical Systems III*, volume 3 of *Encyclopaedia of Mathematical Sciences*. Springer, Berlin, 1988. Translated from the 1985 Russian edition by A. Iacob.

[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, Berlin, 3 edition, 2006. Translated from the 2002 Russian original by E. Khukhro.

[6] D. Bambusi, A. Fusè, and M. Sansottera. Exponential stability in the perturbed central motion. *Regular and Chaotic Dynamics*, 23:821–841, 2018.

[7] S. Barbieri. *Stability in Hamiltonian Systems: steepness and regularity in Nekhoroshev theory*. PhD thesis, Université Paris–Saclay, 2023.

[8] S. Barbieri, L. Biasco, L. Chierchia, and D. Zaccaria. Singular KAM theory for convex hamiltonian system. *Regular and Chaotic Dynamics*, 30(4):538–549, 2025.

[9] P. Bernard, V. Kaloshin, and K. Zhang. Arnol'd diffusion in arbitrary degrees of freedom and 3-dimensional normally hyperbolic invariant cylinders. *Acta Mathematica*, 217(1):1–79, 2016.

[10] M. Berti, L. Biasco, and P. Bolle. Drift in phase space: a new variational mechanism with optimal diffusion time. *Journal de Mathématiques Pures et Appliquées*, 82(6):613–664, 2003.

[11] L. Biasco and L. Chierchia. On the measure of Lagrangian invariant tori in nearly integrable mechanical systems. *Rendiconti Lincei - Matematica e Applicazioni*, 26:1–10, 2015.

[12] L. Biasco and L. Chierchia. Explicit estimates on the measure of primary KAM tori. *Annali di Matematica Pura ed Applicata*, 197:261–281, 2018.

[13] L. Biasco and L. Chierchia. On the measure of KAM tori in two degrees of freedom. *Discrete and Continuous Dynamical Systems*, 40(12):6635–6648, 2020.

[14] L. Biasco and L. Chierchia. On the topology of nearly-integrable Hamiltonians at simple resonances. *Nonlinearity*, 33:3526–3567, 2020.

[15] L. Biasco and L. Chierchia. Quasi-periodic motions in generic nearly integrable mechanical systems. *Atti Accademia Nazionale dei Lincei*, 33(3):575–580, 2022.

[16] L. Biasco and L. Chierchia. Complex Arnol'd–Liouville maps. *Regular and Chaotic Dynamics*, 28(4-5):395–424, 2023.

[17] L. Biasco and L. Chierchia. Global properties of generic real-analytic nearly integrable Hamiltonian systems. *Journal of Differential Equations*, 385:325–361, 2024.

[18] A. Celletti and L. Chierchia. A constructive theory of Lagrangian tori and computer–assisted applications. In *Dynamics Reported*, pages 60–129. Springer, Berlin, 1995.

[19] Q. Chen and R. de la Llave. Analytic genericity of diffusing orbits in a priori unstable Hamiltonian systems. *Nonlinearity*, 35(4):1986–2019, 2022.

[20] L. Chierchia and G. Gallavotti. Drift and diffusion in phase space. *Annales de l'Institut Henri Poincaré (Physique Théorique)*, 60:1–144, 1994. Erratum: ibid. 68 (1998), 135.

[21] L. Chierchia and C. E. Koudjinan. V. I. Arnol'd's global KAM theorem and geometric measure estimates. *Regular and Chaotic Dynamics*, 26(1):61–88, 2021.

[22] A. Delshams, R. de la Llave, and T. M. Seara. Instability of high dimensional Hamiltonian systems: multiple resonances do not impede diffusion. *Advances in Mathematics*, 294:689–755, 2016.

[23] A. Delshams and G. Huguet. Geography of resonances and Arnol'd diffusion in a priori unstable Hamiltonian systems. *Nonlinearity*, 22(8):1997–2077, 2009.

[24] H. S. Dumas. *The KAM Story*. World Scientific, 2014.

[25] H. Eliasson. Perturbation of linear quasi-periodic systems. In *Dynamical Systems and Small Divisors*, volume 1784 of *Lecture Notes in Mathematics*, pages 1–60. Springer, 2002.

[26] C. Falcolini and D. Zaccaria. On the analytic properties of the perturbing function in the PCR3 body problem. *Regular and Chaotic Dynamics*, 31:92–115, 2026.

[27] M. Gidea, R. de la Llave, and T. M. Seara. A general mechanism of diffusion in Hamiltonian systems: qualitative results. *Communications on Pure and Applied Mathematics*, 73(1):150–209, 2020.

[28] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*, volume 224 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin–New York, 1977.

[29] 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.

[30] A. N. Kolmogorov. On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian. *Doklady Akademii Nauk SSSR*, 98:527–530, 1954. English transl. in: Lecture Notes in Physics 93, Springer, 1979.

[31] V. F. Lazutkin. Concerning a theorem of Moser on invariant curves. *Voprosy Dinam. Teor. Rasprostran. Seism. Voln*, 14:109–120, 1974. (Russian).

[32] L. Lazzarini, J.-P. Marco, and D. Sauzin. *Measure and Capacity of Wandering Domains in Gevrey Near-Integrable Exact Symplectic Systems*, volume 257 of *Memoirs of the American Mathematical Society*. AMS, 2019.

[33] J. N. Mather. Arnol'd diffusion by variational methods. In *Essays in Mathematics and Its Applications*, pages 271–285. Springer, Heidelberg, 2012.

[34] A. G. Medvedev, A. I. Neishtadt, and D. V. Treschev. Lagrangian tori near resonances of near-integrable Hamiltonian systems. *Nonlinearity*, 28(7):2105–2130, 2015.

[35] J. Moser. *Lectures on Hamiltonian Systems*, volume 81 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence, RI, 1968.

[36] J. K. Moser. On invariant curves of area-preserving mappings of an annulus. *Nachrichten der Akademie der Wissenschaften Göttingen*, 1:1–20, 1962.

[37] A. I. Neishtadt. Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions. *Journal of Applied Mathematics and Mechanics*, 45(6):766–772, 1981.

[38] A. I. Neishtadt. *Problems of Perturbation Theory for Nonlinear Resonant Systems*. PhD thesis, Moscow State University, 1988. In Russian.

[39] J. Pöschel. Integrability of Hamiltonian systems on Cantor sets. *Communications on Pure and Applied Mathematics*, 35(5):653–696, 1982.

[40] J. Pöschel. Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Mathematische Zeitschrift*, 213:187–216, 1993.

[41] A. S. Pyartli. Diophantine approximations on submanifolds of Euclidean space. *Functional Analysis and Its Applications*, 3:303–306, 1969. Translated from Russian.

[42] A. A. Svanidze. Small perturbations of an integrable dynamical system with integral invariant. *Proceedings of the Steklov Institute of Mathematics*, 2:127–151, 1981.

[43] D. Treschev. Arnol'd diffusion far from strong resonances in multidimensional a priori unstable Hamiltonian systems. *Nonlinearity*, 25(9):2717–2757, 2012.

[44] Y. Yomdin and G. Comte. *Tame Geometry with Application in Smooth Analysis*, volume 1645 of *Lecture Notes in Mathematics*. Springer, Berlin, 2004.

[45] K. Zhang. Speed of Arnol'd diffusion for analytic Hamiltonian systems. *Inventiones Mathematicae*, 186(2):255–290, 2011.