Properly–degenerate KAM theory (following V.I. Arnold)*

Luigi Chierchia
Dipartimento di Matematica
Università "Roma Tre"
Largo S. L. Murialdo 1, I-00146 Roma (Italy)
luigi@mat.uniroma3.it

Gabriella Pinzari
Dipartimento di Matematica ed Applicazioni "R. Caccioppoli"
Università di Napoli "Federico II"
Monte Sant'Angelo – Via Cinthia I-80126 Napoli (Italy)
pinzari@mat.uniroma3.it

October 7, 2009

This paper is dedicated to the memory of Professor Nikolaĭ Nekhoroshev

Abstract

Arnold's "Fundamental Theorem" on properly–degenerate systems [3, Chapter IV] is revisited and improved with particular attention to the relation between the perturbative parameters and to the measure of the Kolmogorov set. Relations with the planetary many–body problem are shortly discussed.

Keywords: KAM theory, Kolmogorov set, many–body problem, small divisors, invariant tori, degeneracies.

Contents

1 Introduction and Result **2**

2 Tools: Averaging, Birkhoff normal form and two–scale KAM **7**

3 Proof of Theorem 1.4 **11**

4 Proof of Theorem 1.3 **18**

A Averaging theory (Proposition 2.1) **19**

B Two–scale KAM theory (proof of Proposition 2.3) **20**

C Measure estimates (proof of Lemma 3.1) **28**

References **30**

* **Acknowledgments.** The authors are grateful to Jacques Féjoz for many helpful discussions.
This paper is based on the PhD thesis [12].
Partially supported by the Italian MIUR project "Metodi variazionali e equazioni differenziali nonlineari" and by European Research Council under F P 7 project "New connections between dynamical systems and Hamiltonian PDEs with small divisors phenomena".

1 Introduction and Result

1.1 A problem that one often encounters in applications of KAM theory is related to the presence of degeneracies.
An important example (which actually motivated the birth of KAM theory) is the problem of finding a positive measure set in phase space corresponding to quasi–periodic motions in the planetary $(1+n)$–body problem (i.e., $1+n$ point masses interacting only under a gravitational potential modeling a system formed by a star and n planets). In this case the integrable limit (i.e., the n uncoupled two–body systems formed by the star and one planet) does not depend upon a full set of action–variables ("proper degeneracy") and therefore typical non–degeneracy conditions (such as Kolmogorov's non–degeneracy or Arnold's iso–energetical non–degeneracy) are strongly violated.
To deal with properly–degenerate systems V.I. Arnold developed in [3] a new KAM technique, which is summarized in what he called the "*Fundamental Theorem*" [3, Chapter IV]. Arnold then applied the *Fundamental Theorem* to the planar, planetary, nearly–circular three–body problem ($n = 2$) proving for the first time relatively bounded motions for a positive set of initial data.
A full proof of this result in the general spatial many–body problem turned out to be more difficult than expected. After an extension to the spatial three–body case [14], a first complete proof was published only in 2004 [9], where a different (smooth) KAM technique (due to M.R. Herman) was used; for a real–analytic proof, see [6].
In this paper we revisit and extend Arnold's Fundamental theorem so as to weaken its hypotheses and to improve the measure estimates on the *Kolmogorov set* (i.e., the union of maximal invariant quasi-periodic tori).

1.2 In properly–degenerate KAM theory it is not enough to make non–degeneracy assumptions on the unperturbed limit (as in standard KAM theory).
To describe a typical setting, let us consider a Hamiltonian function of the form

$$H(I,\varphi,p,q;\mu) := H_0(I) + \mu P(I,\varphi,p,q;\mu) \ , \tag{1.1}$$

where[1] $(I,\varphi) \in V \times \mathbb{T}^{n_1} \subset \mathbb{R}^{n_1} \times \mathbb{T}^{n_1}$ and $(p,q) \in B \subset \mathbb{R}^{2n_2}$ are standard symplectic variables; here V is an open, connected set in $\mathbb{R}^{n_1}$ and B is a $(2n_2)$–ball around the origin; $2n$, where $n := n_1 + n_2$ is the dimension of the phase space

$$\mathcal{P} := V \times \mathbb{T}^{n_1} \times B \ , \tag{1.2}$$

which is endowed with the standard symplectic two–form

$$dI \wedge d\varphi + dp \wedge dq = \sum_{j=1}^{n_1} dI_j \wedge d\varphi_j + \sum_{j=1}^{n_2} dp_j \wedge dq_j \ .$$

The Hamiltonian H is assumed to be real–analytic.
When the perturbative parameter μ is set to be zero (in the planetary case μ measures the ratio between the masses of the planets and that of the star) the system is integrable but depends only on $n_1 < n$ action–variables. A typical further assumption is that the *averaged* (or *secular*) perturbation,

$$P_{\rm av}(p,q;I,\mu) := \int_{\mathbb{T}^{n_1}} P(I,\varphi,p,q;\mu)\ \frac{d\varphi}{(2\pi)^{n_1}} \ , \tag{1.3}$$

has an elliptic equilibrium in the origin with respect to the variables (p,q). Under suitable assumptions on the first and/or second order *Birkhoff invariants* (see [10] for general information) one can guarantee the existence of maximal KAM tori near the "secular tori"

$$\{I\} \times \mathbb{T}^{n_1} \times T_\eta^{n_2} \ , \tag{1.4}$$

[1] $\mathbb{T}^n$ denotes the standard n–dimensional flat torus $\mathbb{R}^n/(2\pi\mathbb{Z}^n)$.

where $\eta = (\eta_1, ..., \eta_{n_2})$, $T^{n_2}_\eta$ denotes a n_2–dimensional torus given by the product of n_2 circles of radii $\eta_j > 0$ and $\epsilon := \max \eta_j$ is small[2].
More precisely, Arnold makes the following assumptions[3]:

(A1) *$I \in V \to \partial_I H_0$ is a diffeomorphism;*

(A2) $P_{\rm av}(p, q; I) = P_0(I) + \sum_{i=1}^{n_2} \Omega_i(I) r_i + \frac{1}{2} \sum_{i,j=1}^{n_2} \beta_{ij}(I) r_i r_j + \sum_{i,j,k=1}^{n_2} \lambda_{ijk}(I) r_i r_j r_k + o_6$ *where* $r_i := \frac{p_i^2 + q_i^2}{2}$ *and* $o_6/|(p, q)|^6 \to 0$ *as* $(p, q) \to 0$;

(A3) *the matrix of the second order Birkhoff invariants is not singular, i.e.,* $|\det \beta(I)| \geq \text{const} > 0$ *for all* $I \in V$.

We can now state Arnold's Fundamental Theorem. Denote by $B_\epsilon = B^{2n_2}_\epsilon = \{y \in \mathbb{R}^{2n_2} : |y| < \epsilon\}$ the $2n_2$–ball of radius ϵ and let

$$\mathcal{P}_\epsilon := V \times \mathbb{T}^{n_1} \times B_\epsilon \qquad (1.5)$$

and recall the definitions of H and the phase space $\mathcal{P}$ in, respectively, (1.1) and (1.2).

Theorem 1.1 (Arnold's "Fundamental Theorem" [3, p. 143])
Let H be real–analytic on $\mathcal{P}$ and assume (A1)÷(A3). *Then, there exists $\epsilon_* > 0$ such that, for*

$$0 < \epsilon < \epsilon_* \ , \qquad 0 < \mu < \epsilon^8 \ , \qquad (1.6)$$

one can find a set $\mathcal{K} \subset \mathcal{P}_\epsilon \subset \mathcal{P}$ formed by the union of H–invariant n–dimensional tori close to the secular tori in (1.4), *on which the H–motion is analytically conjugated to linear Diophantine[4] quasi–periodic motions. The set $\mathcal{K}$ is of positive Liouville–Lebesgue measure and satisfies*

$$\text{meas}\, \mathcal{K} > (1 - \text{const}\, \epsilon^a)\, \text{meas}\, \mathcal{P}_\epsilon \ , \qquad \text{where } a := 1/(8(n+4)) \ . \qquad (1.7)$$

Remark 1.1 By Birkhoff's theory (compare Proposition 2.2 below), the expansion in (A2) for $P_{\rm av}$ may be achieved if one assumes that $(p, q) \to P_{\rm av}(p, q; I)$ has an elliptic equilibrium in $p = q = 0$ and the first order Birkhoff invariants Ω_i are non resonant up to order 6, *ie*[5],

$$\Big| \sum_{j=1}^{n_2} \Omega_j(I) k_j \Big| \geq \text{const} > 0 \quad \forall I \in V \ , \quad \forall \quad 0 < |k| \leq 6 \ , \quad k \in \mathbb{Z}^{n_2} \ . \qquad (1.8)$$

In this paper we relax condition (1.6) and replace assumption (A2) with either

(A2′) $(p, q) \to P_{\rm av}(p, q; I)$ *has an elliptic equilibrium in the origin $p = q = 0$ and the first order Birkhoff invariants are non resonant uo to order four, i.e. , they verify* (1.8) *with* 6 *replaced by* 4.

or

(A2″) $P_{\rm av}(p, q; I) = P_0(I) + \sum_{i=1}^{n_2} \Omega_i(I) r_i + \frac{1}{2} \sum_{i,j=1}^{n_2} \beta_{ij}(I) r_i r_j + o_4$ *with* $r_i := \frac{p_i^2 + q_i^2}{2}$ *and* $o_4/|(p, q)|^4 \to 0$ *as* $(p, q) \to 0$.

We shall prove the following two theorems.

---

[2] An interesting point is what is the relation between ϵ and μ, especially in view of physical applications (in the planetary case ϵ measures the eccentricities and relative inclinations of the star–planet motions): this matter will be further discussed in the following.

[3] From now on we drop the dependence on μ of the perturbation, assuming that such dependence is smooth enough, say C^1, and that the norms are uniform in μ.

[4] *I.e.*, the flow is conjugated to the Kronecker flow $\theta \in \mathbb{T}^n \to \theta + \omega\, t \in \mathbb{T}^n$, with $\omega = (\omega_1, \omega_2)$ satisfying (1.19) below.

[5] Here and below, for integer vectors $k \in \mathbb{Z}^m$, $|k| := |k|_1 = \sum_{j=1}^m |k_j|$. See also notation in Chapter 2.

Theorem 1.2 *Let H be real–analytic on $\mathcal{P}$ and assume* (A1), (A2′) and (A3)*. Then, there exist positive numbers ϵ_*, μ_*, C_* and b such that, for*

$$0 < \epsilon < \epsilon_* \ , \quad 0 < \mu < \mu_* \ , \quad \mu < \frac{1}{C_*(\log \epsilon^{-1})^{2b}} \ , \tag{1.9}$$

one can find a set $\mathcal{K} \subset \mathcal{P}$ formed by the union of H–invariant n–dimensional tori, on which the H–motion is analytically conjugated to linear Diophantine quasi–periodic motions. The set $\mathcal{K}$ is of positive Liouville–Lebesgue measure and satisfies

$$\operatorname{meas} \mathcal{P}_\epsilon > \ \operatorname{meas} \mathcal{K} > \Big(1 - C_*\Big(\sqrt{\mu}\ (\log \epsilon^{-1})^b + \sqrt{\epsilon}\Big)\Big) \operatorname{meas} \mathcal{P}_\epsilon \ . \tag{1.10}$$

Next theorem needs stronger hypotheses on μ but there are no conditions on the first order Birkhoff invariants.

Theorem 1.3 *Let H be real–analytic on $\mathcal{P}$ and assume* (A1), (A2″) *and* (A3)*. Then, there exist positive numbers ϵ_*, μ_*, C_* and b such that, for*

$$0 < \epsilon < \epsilon_* \ , \quad 0 < \mu < \mu_* \ , \quad \mu < \frac{\epsilon^6}{(\log \epsilon^{-1})^{2b}} \ , \tag{1.11}$$

one can find a set $\mathcal{K} \subset \mathcal{P}_\epsilon$ formed by the union of H–invariant n–dimensional tori, on which the H–motion is analytically conjugated to linear Diophantine quasi–periodic motions. The set $\mathcal{K}$ is of positive Liouville–Lebesgue measure and satisfies

$$\operatorname{meas} \mathcal{P}_\epsilon > \ \operatorname{meas} \mathcal{K} > \Big(1 - C_*\sqrt{\epsilon}\Big) \operatorname{meas} \mathcal{P}_\epsilon \ . \tag{1.12}$$

1.3 Let us make a few remarks.

(i) Under assumption (A2′), near $p = 0 = q$, the dynamics is approximated by the dynamics governed by the integrable "secular (averaged and truncated) Hamiltonian"

$$H_{\rm sec} := H_0(I) + \mu\Big(P_0(I) + \sum_{i=1}^{n_2} \Omega_i(I) \frac{p_i^2 + q_i^2}{2}\Big) \ . \tag{1.13}$$

The phase space $\mathcal{P}$ is foliated by $2n$ dimensional $H_{\rm sec}$–invariant tori as in (1.4) with $0 < \epsilon < \bar\epsilon$, where $\bar\epsilon$ denotes the radius of the ball B in (1.2). Indeed, in this case $T^{n_2}_\eta$ are simply given by $\{p_i^2 + q_i^2 = \eta_j \ , \ j \le n_2\}$ with $\eta_j \le \epsilon$. In the perturbed case the fate of the secular tori may be different according to the relation between ϵ and μ. In fact what happens is that, *if $\mu < \epsilon^\alpha$, with $\alpha > 1$ (in particular, if* (1.6) *or* (1.11) *holds), then $\mathcal{K} \subset \mathcal{P}_\epsilon$* as in Arnold's Theorem, but if $\mu > \epsilon^\alpha$ then, is general $\mathcal{K}$ is not contained in $\mathcal{P}_\epsilon$ and the persistent tori may be not so close to the secular tori $\{I\} \times \mathbb{T}^{n_1} \times \{p_i^2 + q_i^2 = \eta_j,\ j \le n_2\}$ but rather they are close to the translated tori $\{I\} \times \mathbb{T}^{n_1} \times \{(p_i - p_i^0)^2 + (q_i - q_i^0)^2 = \eta_j,\ j \le n_2\}$ where $(p_i^0, q_i^0) = \Big(p_i^0(I;\mu), q_i^0(I;\mu)\Big)$ are the coordinates of a "new equilibrium", which depend upon the full averaged system and which may be "logarithmically" distant from the origin (as far as $1/\log \epsilon^{-1}$). In any case, the set $\mathcal{K}$ fills almost completely a region diffeomorphic to and of equal measure of $\mathcal{P}_\epsilon$.

A precise geometrical description of the "Kolmogorov set" $\mathcal{K}$ is given in Step 6 of § 3.

(ii) As mentioned above, in the planetary problem, μ measures the mass ratio between the planets and the star, while ϵ is related to the eccentricities and inclinations of the (instantaneous) two–body systems planet–star. Condition (1.9) is much weaker than Arnold's condition (1.6) and allows, at least in principle, applications to a wider class of planetary systems.

Clearly, in order to apply properly–degenerate KAM theory to a concrete system such as the outer Solar system[6] one should also estimate ϵ_* and μ_* in (1.9), which would be quite a technical achievement[7].

(iii) Arnold declared [3, end of p. 142] that he made "no attempt to achieve elegance or precision in evaluating constants" adding that "the reader can easily strengthen the results". However, the authors are not aware of improvements on Arnold's results (in the "full torsion case", compare next item) and especially on the issue of giving possibly "sharp" estimates on the measure of the Kolmogorov set arising in properly–degenerate systems. At this respect it would be interesting to know whether estimate (1.10) could be improved or not.

(iv) Relaxing (1.8), *i.e.* , bringing to four the order of non–resonance to be checked, has an interesting application in the case of the $(1+n)$–body problem. In fact, Herman and Fejóz showed [9] that, in the spatial case ($n_1 = n$ and $n_2 = 2n$), the only linear relations satisfied by the first order Birkhoff invariants Ω_j are (up to rearranging indices):

$$\Omega_{2n} = 0 \ , \qquad \sum_{j=1}^{2n-1} \Omega_j(I) = 0 \ . \tag{1.14}$$

The first relation is due to rotation invariance of the system, while the second relation is usually called *Herman resonance*[8]. Now, since in the spatial case, Herman resonance is of order $2n-1$, one sees that for $n \geq 3$ it is not relevant for (A2′) (but it is for (1.8)).

Actually, at this respect, Theorem 1.3 might be even more useful since it involves no assumption on the Ω_j so that in possible application to the spatial $(1+n)$–body problem, Herman resonance plays no rôle.

(v) The properly–degenerate KAM theory developed in [9] (for the C^∞ case) and in [6] (for the analytic case), being based on weaker non–degeneracy assumptions, is different from Arnold's theory. Roughly speaking, while Arnold's approach is ultimately based on Kolmogorov's non–degeneracy condition ("full torsion in a two–scale setting"), the approach followed in [9, 6] (which might be called "weak properly–degenerate KAM theory") is based on the torsion of the frequency map, exploiting conditions studied by Arnold himself, Margulis, Pyartli, Parasyuk, Bakhtin and especially Rüßmann [15]; for a review, see [16]. Indeed, for Arnold's properly–degenerate theory one has to check that the matrix of the *second order* Birkhoff invariants is not singular (condition (A3) above), while for the weak properly–degeneracy theory it is enough to check a generic property involving only the *first order* Birkhoff invariants: Conditions (A2) and (A3) are replaced by the requirement that the re–scaled frequency map $I \in V \to \hat\omega(I) := (\partial H_0(I), \Omega(I))$ is non–planar, i.e., $\hat\omega(V)$ does not lie in any $(n-1)$–dimensional linear subspace of $\mathbb{R}^n$.

Incidentally, the presence of the resonances (1.14) makes difficult a direct application of weak properly–degenerate KAM theory to the spatial $(1+n)$–body problem in standard Poincaré variables[9].

Explicit measure estimates on the set of persistent tori in the context of weak properly–degenerate KAM theory are not readily available[10].

[6]In the outer Solar System (Sun, Jupiter, Saturn, Uranus and Neptune) μ is of order 10^{-3} and the largest eccentricity is of ~ 0.05 (Saturn).

[7]For partial results in this direction, see [5] and [11].

[8]Compare also [1].

[9]Application of the properly–degenerate KAM theory developed in the present paper using *Deprit variables* [7] will be matter of a future paper by the authors.

[10]In fact, although Pyartli's theorem on the measure of Diophantine points on a non–planar curve is quantitative (compare [9, Théorème 55]), explicit measure estimates of the Kolmogorov set in the N–body problem, following the strategy in [9], do not appear completely obvious.

(vi) Let us briefly (and informally) recall Arnold's scheme of proof. First, by classical averaging theory (see, *e.g.* , [2]) the Hamiltonian (1.1) is conjugated to a Hamiltonian $\tilde{H}$ satisfying, for any small[11] $\sigma > 0$,

$$\tilde{H} = H_0 + \mu P_{\rm av} + O(\mu^{2-\sigma}) \tag{1.15}$$

where $P_{\rm av}$ ia as in (A2). Denoting $P_{\rm av}^{[6]}$ the truncation in (p, q) at order 6 of $P_{\rm av}$, one sees that (1.15) can be rewritten as

$$\tilde{H} = H_0 + \mu P_{\rm av}^{[6]} + O(\mu\epsilon^7) + O(\mu^{2-\sigma}) \tag{1.16}$$

if $|(p, q)| < \epsilon$. In turn, (1.16) is of the form

$$\tilde{H} = H_0 + \mu P_{\rm av}^{[6]} + O(\mu\epsilon^7)$$

if (1.6) holds. At this point, a two–time scale KAM theorem can be applied.

The scheme of proof of Theorem 1.3 is similar, but we use more accurate estimates based on the averaging theory described in § 2.1 below and, especially, on the two–scale KAM theorem described in § 2.3 below; this last result, in particular, is not available in literature and we include its proof in Appendix B.

To relax significantly the relation between μ and ϵ, the above strategy has to be modified. The scheme to prove Theorem 1.2 is the following:

step 1: averaging over the "fast angles" φ's;

step 2: determination of the elliptic equilibrium for the "secular system";

step 3: symplectic diagonalization of the secular system;

step 4: Birkhoff normal form of the secular part;

step 5: global action–angle variables for the full system;

step 6: construction of the Kolmogorov set via an application of a two–scale KAM theorem and estimate of its measure.

1.4 Properly–degenerate systems present naturally *two different scales*: a scale of "order one" related to the unperturbed system (the typical velocity of the fast angles φ's) and a scale of order μ (typical size of the secular frequencies) related to the strength of the perturbation. Furthermore, a third scale appears naturally, namely, the distance from the elliptic equilibrium in the (p, q)–variables.

We now give a more technical and detailed statement, from which Theorem 1.2 follows at once.

Theorem 1.4 *Under the same notations of Theorem* 1.2 *and assumptions* (A1), (A2′) *and* (A3)*, let* $\tau > n_1$ *and*[12] $\tau_* > n := n_1 + n_2$*, with* n_1*,* n_2 *positive integers. Then, there exist* μ_**,* $\epsilon_* < 1$*,* γ_**,* $C_* > 1$ *such that, if* (1.9) *holds and if* $\bar{\gamma}$*,* γ_1*,* $\bar{\gamma}_2$ *are taken so as to satisfy* $\mu\bar{\gamma}_2 \le \gamma_1$ *and*

$$\begin{cases} \gamma_* \max\{\sqrt{\mu}(\log\epsilon^{-1})^{\tau+1},\ \sqrt[3]{\mu\epsilon}(\log\epsilon^{-1})^{\tau+1},\ \epsilon^2(\log\epsilon^{-1})^{\tau+1}\} < \bar{\gamma} < \gamma_* \\ \gamma_*\epsilon^{5/2} < \gamma_1 < \gamma_* \\ \gamma_*\epsilon^{5/2}(\log(\epsilon^5/\gamma_1{}^2)^{-1})^{\tau_*+1} < \bar{\gamma}_2 < \gamma_*\epsilon^2\ , \end{cases} \tag{1.17}$$

[11] The appearence of the exponents $(2 - \sigma)$ (rather than the more "natural" exponent 2) is due to the presence of small divisors.

[12] At contrast with classical KAM theory, where the Diophantine constant can be taken greater than $n - 1$, here one needs $\tau_* > n$ (in [3] it is taken $n + 1$): this is due to the asymmetry of the frequency–domain having n_1–dimensions of order one and $n - n_1 = n_2$ dimensions small with the perturbative parameters.

then, one can find a set $\mathcal{K} \subset \mathcal{P}$ formed by the union of H–invariant n–dimensional tori close to the secular tori in (1.4)*, on which the H–motion is analytically conjugated to linear Diophantine quasi–periodic motions. The set $\mathcal{K}$ is of positive measure and satisfies*

$$\operatorname{meas}\mathcal{P}_\epsilon > \operatorname{meas}\mathcal{K} > \Big[1 - C_*\Big(\bar\gamma + \gamma_1 + \frac{\bar\gamma_2}{\epsilon^2} + \epsilon^{n_2/2}\Big)\Big] \operatorname{meas}\mathcal{P}_\epsilon \ . \tag{1.18}$$

Furthermore, the flow on each H–invariant torus in $\mathcal{K}$ is analytically conjugated to a translation $\psi \in \mathbb{T}^n \to \psi + \omega t \in \mathbb{T}^n$ with Diophantine vector $\omega = (\omega_1, \omega_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ satisfying, for each $k = (k_1, k_2) \in \big(\mathbb{Z}^{n_1} \times \mathbb{Z}^{n_2}\big)\backslash\{0\}$,

$$|\omega_1 \cdot k_1 + \omega_2 \cdot k_2| \geq \begin{cases} \dfrac{\gamma_1}{|k|^{\tau_*}} & \text{if} \quad k_1 \neq 0 \ ; \\ \\ \mu\, \dfrac{\bar\gamma_2}{|k_2|^{\tau_*}} & \text{if} \quad k_1 = 0 \ , \quad k_2 \neq 0 \ . \end{cases} \tag{1.19}$$

To obtain Theorem 1.2 from Theorem 1.4 one can choose

$$\bar\gamma = \gamma_*^2 \max\Big\{\sqrt{\mu}(\log\epsilon^{-1})^{\tau+1},\ \epsilon^{2/3}(\log\epsilon^{-1})^{\tau+1}\Big\} \ , \qquad \gamma_1 = \bar\gamma_2 = \gamma_*^2\epsilon^{5/2} \ ; \tag{1.20}$$

then (1.10) follows easily[13], with $b = \tau + 1$.

The proof of Theorem 1.3, as already mentioned, is simpler and it will be shortly given in § 4.

2 Tools: Averaging, Birkhoff normal form and two–scale KAM

First of all we fix some notation, which will be used throughout the paper.

- in $\mathbb{R}^{n_1}$ we fix the 1–norm: $|I| := |I|_1 := \sum_{1\leq i\leq n_1} |I_i|$;
- in $\mathbb{T}^{n_1}$ we fix the "sup–metric": $|\varphi| := |\varphi|_\infty := \max_{1\leq i\leq n_1} |\varphi_i| \pmod{2\pi}$;
- in $\mathbb{R}^{n_2}$ we fix the sup norm: $|p| := |p|_\infty := \max_{1\leq i\leq n_2} |p_i|$, $|q| := |q|_\infty := \max_{1\leq i\leq n_2} |q_i|$;
- for matrices we use the "sup–norm": $|\beta| := |\beta|_\infty := \max_{i,j} |\beta_{ij}|$;
- if $A \subset \mathbb{R}^{n_i}$, or $A \subset \mathbb{T}^{n_1}$, and $r > 0$, we denote by $A_r := \bigcup_{x\in A}\Big\{z \in \mathbb{C}^{n_i} : \quad |z - x| < r\Big\}$ the complex r–neighbrhood of A (according to the prefixed norms/metrics above);
- if f is real–analytic on a complex domain of the form $U_v \times \mathbb{T}^m_s$ with $U \subset \mathbb{R}^d$, we denote by $\|f\|_{U_v\times\mathbb{T}^m_s}$, or, simply, $\|f\|_{v,s}$ its "sup–Fourier norm":

  $$\|f\|_{v,s} := \sum_{k\in\mathbb{Z}^m} \sup_{u\in U_v} |f_k(u)| e^{|k|s} \ , \quad |k| := \sum_{1\leq i\leq m} |k_i|$$

  where $f_k(u)$ denotes the k^{th} Fourier coefficient of $f = \sum_{k\in\mathbb{Z}^m} f_k(u) e^{ik\cdot\varphi}$;

[13] First, let us check that (1.17) holds. From (1.9) it follows that $\bar\gamma < \gamma_*$ (provided $C_* > \gamma_*$ and ϵ_* is small enough). The lower bound on $\bar\gamma$ is checked by considering the cases $\epsilon \leq \mu$ and $\mu < \epsilon$ separately. The bounds on γ_1 are obvious.The bounds on $\bar\gamma_2$ are true for γ_* big enough. Thus, (1.17) is checked. Finally, (1.20) and (1.18) imply easily (1.10).

- if f is as in the previous item, $K > 0$ and Λ is a sublattice of $\mathbb{Z}^m$, $T_K f$ and $\Pi_\Lambda f$ denote, respectively, the K–truncation and the Λ–projection of f:

$$T_K f := \sum_{|k|\le K} f_k(u) e^{ik\cdot\varphi} \ , \quad \Pi_\Lambda f := \sum_{k\in\Lambda} f_k(u) e^{ik\cdot\varphi} \ ;$$

- if $f : A \subset \mathbb{R}^d \to \mathbb{R}^n$ is a Lipschitz function and $\rho > 0$ a "weight", we denote its ρ–*Lipschitz norm* by

$$\|f\|^{\rm Lip}_{\rho,A} := \rho^{-1} \sup_A |f| + \mathcal{L}(f) \ , \qquad \mathcal{L}(f) := \sup_{I\ne I'\in A} \frac{|f(I)-f(I')|}{|I-I'|} \ . \tag{2.21}$$

- $\mathcal{D}_{\gamma_1,\gamma_2,\tau} \subset \mathbb{R}^{n_1+n_2}$ denotes the set of Diophantine (γ_1,γ_2,τ)–numbers, i.e., the set of vectors $\omega \in \mathbb{R}^{n_1+n_2}$ satisfying for any $k = (k_1,k_2) \in \mathbb{Z}^{n_1+n_2}\backslash\{0\}$, inequality (1.19) with $\tau_* = \tau$, $\bar\gamma_2 = \gamma_2$ and $\mu = 1$. When $\gamma_1 = \gamma_2 = \gamma$, we obtain the usual Diophantine set $\mathcal{D}_{\gamma,\tau}$.

2.1 *Averaging theory*

The first step of the proof of Theorem 1.4 (and hence of Theorem 1.2) is based upon averaging theory. We shall follow the presentation given in [4, Appendix A], which in turn is based upon [13].

Proposition 2.1 (Averaging theory) *Let $\bar K$, $\bar s$ and s be positive numbers such that $\bar K s \ge 6$ and let $\alpha_1 \ge \alpha_2 > 0$; let $A\times B\times B' \subset (\mathbb{R}^{\ell_1}\times\mathbb{R}^{\ell_2})\times\mathbb{R}^m\times\mathbb{R}^m$, and $v = (r, r_p, r_q)$ a triple of positive numbers. Let $H := h(I) + f(I,\varphi,p,q)$ be a real-analytic Hamiltonian on $W_{v,\bar s+s} := A_r \times B_{r_p} \times B'_{r_q} \times \mathbb{T}^{\ell_1+\ell_2}_{\bar s+s}$. Finally, let Λ be a (possibly trivial) sub–lattice of $\mathbb{Z}^{\ell_1+\ell_2}$ and let $\omega = (\omega_1,\omega_2)$ denote the gradient $(\partial_{I_1} h, \partial_{I_2} h) \in \mathbb{R}^{\ell_1+\ell_2}$. Let $k = (k_1,k_2) \in \mathbb{Z}^{\ell_1}\times\mathbb{Z}^{\ell_2}$ and assume that*

$$|\omega\cdot k| \ge \left\{\begin{array}{ll} \alpha_1 \ , & \text{if} \quad k_1 \ne 0 \\ \alpha_2 \ , & \text{if} \quad k_1 = 0 \end{array}\right. \qquad \forall\ I\in A_r \ , \ \forall\ k=(k_1,k_2)\notin\Lambda \ , \ |k|\le\bar K \tag{2.22}$$

$$E := \|f\|_{v,\bar s+s} < \frac{\alpha_2 d}{2^7 c_m \bar K s} \ , \ \text{where} \quad d = \min\{rs,\ r_p\, r_q\} \ , \quad c_m := \frac{e(1+em)}{2} \ . \tag{2.23}$$

Then, there exists a real-analytic, symplectic transformation

$$\Psi : \quad (I',\varphi',p',q') \in W_{v/2,\bar s+s/6} \to (I,\varphi,p,q) \in W_{v,\bar s+s} \tag{2.24}$$

such that

$$H_* := H\circ\Psi = h + g + f_* \ ,$$

with g in normal form and f_* small:

$$g = \sum_{k\in\Lambda} g_k(I',p',q')\, e^{ik\cdot\varphi'} \ , \qquad \|g - \Pi_\Lambda T_{\bar K} f\|_{v/2,\bar s+s/6} \le \frac{12}{11}\ \frac{2^7 c_m E^2}{\alpha_2 d} \le \frac{E}{4} \ ,$$

$$\|f_*\|_{v/2,\bar s+s/6} \le e^{-\bar K s/6}\ \frac{2^9 c_m E^2}{\alpha_2\, d} \le e^{-\bar K s/6}\, E \ . \tag{2.25}$$

Moreover, denoting by $z = z(I',\varphi',p',q')$, the projection of $\Psi(I',\varphi',p',q')$ onto the z–variables ($z = I_1$, I_2, φ, p or q) one has

$$\max\{\alpha_1 s|I_1 - I_1'|\,,\ \alpha_2 s|I_2 - {I_2}'|\,,\ \alpha_2 r\,|\varphi-\varphi'|\,,\ \alpha_2 r_q\,|p-p'|\,,\ \alpha_2 r_p\,|q-q'|\} \ \le 9E\,. \tag{2.26}$$

This Proposition is essentially Proposition A.1 of [4] with two slight improvements. The first improvement is trivial and concerns the introduction of the parameter $\bar{s}$ so as to separate the rôle of the analyticity loss in the angle–variables from the initial angle–domain. Such variation is important, for example, in applying the Averaging Theorem infinitely many times.

The second improvement is a bit more delicate and we use it in the proof of Proposition 2.3 below. It concerns the separation of two scales in the frequencies $\omega = \partial_I h$.

Proposition 2.1 holds also for $\ell_1 \neq 0$, $\ell_2 = 0$ (*i.e.* , there is only one action scale), in which case $\alpha_2 := \alpha_1 = \alpha$, and in the case $m = 0$ (*i.e.* , there are no (p, q)–variables), in which case one can take $d = rs$, $c_m = c_0 = e/2$.

In the following, Proposition 2.1 will be applied twice: in step 1 of § 3 (with $\ell_1 = n_1$, $\ell_2 = 0$, $m = n_2$) and in Appendix B with $m = 0$.
Proposition 2.1 is proved in Appendix A.

2.2 *Birkhoff normal form*

We now recall a fundamental result due to Birkhoff on normal forms. We follow [10].

Proposition 2.2 (Birkhoff normal form) *Let $\alpha > 0$, $s \geq 3$; let $\Omega = (\Omega_1, \cdots, \Omega_m) \in \mathbb{R}^m$ be non–resonant of order s, i.e. ,*

$$|\Omega \cdot k| \geq \alpha > 0 \ , \quad \forall \quad k \in \mathbb{Z}^m \quad \text{with} \quad 0 < |k| \leq s \tag{2.27}$$

and let $z = (p, q) \in B^{2m}_{\epsilon_0} = \{z : |z| < \epsilon_0\} \subset \mathbb{R}^{2m} \to \mathrm{H}(z)$ be a real–analytic function of the form

$$\mathrm{H}(z) = \sum_{i=1}^{m} \Omega_i \, r_i + O(|z|^3) \quad \text{where} \quad r_i := \frac{p_i^2 + q_i^2}{2} \ . \tag{2.28}$$

Then, there exists $0 < \tilde{\epsilon} \leq \epsilon_0$ and a real–analytic and symplectic[14] *transformation*

$$\phi : \quad \tilde{z} = (\tilde{p}, \tilde{q}) \in B^{2m}_{\tilde{\epsilon}} \to \tilde{z} + \hat{z}(\tilde{z}) \in B^{2m}_{\epsilon_0}$$

which puts H *into Birkhoff normal form up to order s, i.e.*[15],

$$\tilde{\mathrm{H}} := \mathrm{H} \circ \phi = \sum_{i=1}^{m} \Omega_i \, \tilde{r}_i + \sum_{j=2}^{[s/2]} Q_j(\tilde{r}) + O(|\tilde{z}|^{s+1}) \tag{2.29}$$

where, for $2 \leq j \leq [s/2]$, the Q_j's are homogeneous polynomials of degree j in $\tilde{r} = (\tilde{r}_1, \cdots, \tilde{r}_m)$ with $\tilde{r}_j := \dfrac{\tilde{p}_j^2 + \tilde{q}_j^2}{2}$. The polynomials Q_j do not depend on ϕ.

Following the proof of this classical result as presented in [10] one can easily achieve the following useful amplifications.

1. The construction of the transformation ϕ is iterative and can be described as follows. There exist positive numbers $\tilde{\epsilon} := \epsilon_{s-2} < \epsilon_{s-3} < \cdots < \epsilon_0$, and a symplectic transformations ϕ_i such that $\phi = \hat{\phi}_{s-2} := \phi_1 \circ \cdots \circ \phi_{s-2}$, that $H \circ \hat{\phi}_i$ is in Birkhoff normal form up to order $i + 2$ and
$$\phi_i : \quad \tilde{z} = (\tilde{p}, \tilde{q}) \in B^{2m}_{\epsilon_i} \to \tilde{z} + \hat{z}_{i-1}(\tilde{z}) \in B^{2m}_{\epsilon_{i-1}} \ , \qquad \forall \ 1 \leq i \leq s-2 \ , \quad \text{with}$$
$$\sup_{B^{2m}_{\epsilon_i}} |\hat{z}_{i-1}| \leq c_{i-1} \frac{m_{i-1}}{\alpha} (\epsilon_{i-1})^{i+1} \ ,$$

[14] With respect to the standard 2–form $dp \wedge dq = \sum_{1 \leq i \leq m} dp_i \wedge dq_i$.

[15] $[x]$ denotes the integer part of x.

where c_{i-1} depend only on the dimension m and m_{i-1} are defined as follows. For $i-1=0$, let $\mathcal{P}_0$ the homogeneous polynomial of degree 3 for which $H(z)-\sum_{j=1}^m \Omega_j r_j = \mathcal{P}_0 + O(|z|^4)$, while, for $i-1\ge 1$, let $\mathcal{P}_{i-1}$ the homogeneous polynomial of degree $i+2$ for which

$$H(z)\circ\hat\phi_{i-1} - \sum_{j=1}^m \Omega_j r_j - \sum_{j=2}^{[(i+1)/2]} Q_j(r) = \mathcal{P}_{i-1} + O(|z|^{i+3}) \ .$$

Write $\mathcal{P}_{i-1} = \sum_{|\alpha|+|\beta|=i+2} c_{\alpha,\beta}\prod_{j=1}^m (p_j+\mathrm{i}q_j)^{\alpha_j}(p_j-\mathrm{i}q_j)^{\beta_j}$, where i:=$\sqrt{-1}$. Then,

$$m_{i-1} := \max_{\alpha,\beta:\alpha\neq\beta} |c_{\alpha,\beta}| \ . \tag{2.30}$$

2. Proposition 2.2 can be easily extended to the case of a real–analytic function

$$\mathrm{H}(z;I) = \sum_{i=1}^m \Omega_i(I)\, r_i + O(|z|^3) \ ,$$

which also depends on suitable action variables I. More precisely, if A is an open subset of $\mathbb{R}^n$, ρ_0, s_0, ϵ_0 are positive numbers, (I,φ) and $z=(p,q)$, with $(I,\varphi,z)\in A_{\rho_0}\times\mathbb{T}^n_{\sigma_0}\times B^{2m}_{\epsilon_0}$, are conjugate couples of symplectic variables with respect to the standard 2–form $dI\wedge d\varphi + dp\wedge dq$ and $\Omega=(\Omega_1,\cdots,\Omega_m)$ is a suitable real–analytic function defined on A_{ρ_0} verifying (2.27) on A_{ρ_0}, then, one can prove that for suitable $0<\tilde\epsilon=\epsilon_{s-2}<\cdots<\epsilon_0$, $0<\tilde\sigma=\sigma_{s-2}<\cdots<\sigma_0$, $0<\tilde\rho=\rho_{s-2}<\cdots<\rho_0$, c_i, there exist $s-2$ real–analytic, symplectic transformations which we still denote ϕ_i,

$$(\tilde I,\tilde\varphi,\tilde z)\in A_{\rho_i}\times\mathbb{T}^n_{\sigma_i}\times B^{2m}_{\epsilon_i} \stackrel{\phi_i}{\to} (\tilde I,\tilde\varphi+\hat\varphi_{i-1}(\tilde z;\tilde I),\tilde z+\hat z_{i-1}(\tilde z;\tilde I))\in A_{\rho_{i-1}}\times\mathbb{T}^n_{\sigma_{i-1}}\times B^{2m}_{\epsilon_{i-1}}$$

such that (2.29) holds with $\phi=\hat\phi_{s-2}=\phi_1\circ\cdots\circ\phi_{s-2}$, $\Omega_i=\Omega_i(I)$ and suitable homogeneous polynomials $Q_j(\tilde r;I)$ of degree j in $\tilde r=(\tilde r_1,\cdots,\tilde r_m)$ whose coefficients are analytic functions on $A_{\tilde\rho}$. At each step, the functions $(\tilde z;\tilde I)\to\hat z_{i-1}(\tilde z;\tilde I)$, $(\tilde z;\tilde I)\to\hat\varphi_{i-1}(\tilde z;\tilde I)$ verify

$$\sup_{B^{2m}_{\epsilon_i}\times A_{\rho_i}} |\hat z_{i-1}| \le c_{i-1}\frac{m_{i-1}}{\alpha}(\epsilon_{i-1})^{i+1} \ , \qquad \sup_{B^{2m}_{\epsilon_i}\times A_{\rho_i}} |\hat\varphi_{i-1}| \le c_{i-1}\frac{m_{i-1}}{\alpha\rho_0}(\epsilon_{i-1})^{i+2} \tag{2.31}$$

where, if, for any fixed $I\in A_{\rho_{i-1}}$, $m_{i-1}(I)$ are defined as in (2.30) with $c_{\alpha,\beta}=c_{\alpha,\beta}(I)$, then, $m_{i-1}=\sup_{A_{\rho_{i-1}}} m_{i-1}(I)$.

2.3 *Two–scale KAM theory*

The invariant tori of Theorem 1.3 and 1.4 will be obtained as an application of a KAM Theorem, adapted to two different frequency scales, which is described in the following

Proposition 2.3 (Two–scale KAM Theorem) *Let n_1, $n_2\in\mathbb{N}$, $n:=n_1+n_2$, $\tau_*>n$, $\gamma_1\ge\gamma_2>0$, $0<4s\le\bar s<1$, $\rho>0$, $D\subset\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}$, $A:=D_\rho$, and let*

$$\mathrm{H}(J,\psi) = \mathrm{h}(J) + \mathrm{f}(J,\psi)$$

be real–analytic on $A\times\mathbb{T}^n_{\bar s+s}$. Assume that $\omega_0:=\partial\mathrm{h}$ is a diffeomorphism of A with non singular Hessian matrix $U:=\partial^2\mathrm{h}$ and let $\hat U$ denote the $n\times n_1$ submatrix of U, i.e. , the matrix with entries $\hat U_{ij}=U_{ij}$, for $n_1+1\le i\le n$, $1\le j\le n$. Let

$$M\ge\sup_A\|U\| \ , \quad \hat M\ge\sup_A\|\hat U\| \ , \quad \bar M\ge\sup_A\|U^{-1}\| \ , \quad E\ge\|\mathrm{f}\|_{\rho,\bar s+s} \ ;$$

define

$$\hat{c} := \max\left\{2^8 n,\ \frac{24^{\tau_*+1}}{6}\right\}$$
$$K := \frac{6}{s}\ \log_+\left(\frac{EM^2 L}{\gamma_1^2}\right)^{-1} \quad \text{where} \quad \log_+ a := \max\{1, \log a\}$$
$$\hat{\rho} := \min\left\{\frac{\gamma_1}{3MK^{\tau_*+1}}\ ,\ \frac{\gamma_2}{3\hat{M}K^{\tau_*+1}}\ ,\ \rho\right\}$$
$$L := \max\ \left\{\bar{M}\ , M^{-1}, \hat{M}^{-1}\right\}$$
$$\hat{E} := \frac{EL}{\hat{\rho}^2}\ ;$$

finally, let $\bar{M}_1$, $\bar{M}_2$ *upper bounds on the norms of the submatrices* $n_1 \times n$, $n_2 \times n$ *of* U^{-1} *of the first* n_1, *last* n_2 *rows*[16]. *Assume the perturbation* f *so small that the following "KAM condition" holds*

$$\hat{c}\hat{E} < 1\ . \tag{2.32}$$

Then, for any $\omega \in \Omega_* := \omega_0(D) \cap \mathcal{D}_{\gamma_1,\gamma_2,\tau_*}$, *one can find a unique real–analytic embedding*

$$\phi_\omega :\quad \vartheta \in \mathbb{T}^n \quad \to \quad (v(\vartheta;\omega), \vartheta + u(\vartheta;\omega)) \in \ \mathrm{Re}\,(D_r) \times \mathbb{T}^n \tag{2.33}$$

where $r := 20n\hat{E}\hat{\rho}$ *such that* $\mathrm{T}_\omega := \phi_\omega(\mathbb{T}^n)$ *is a real–analytic* n*–dimensional* H*–invariant torus, on which the* H*–flow is analytically conjugated to* $\vartheta \to \vartheta + \omega\, t$. *Furthermore, the map* $(\vartheta;\omega) \to \phi_\omega(\vartheta)$ *is Lipschitz and one–to–one and the invariant set* $\mathrm{K} := \bigcup_{\omega\in\Omega_*} \mathrm{T}_\omega$ *satisfies the following measure estimate*

$$\mathrm{meas}\Big(\,\mathrm{Re}\,(D_r) \times \mathbb{T}^n \setminus \mathrm{K}\Big) \le c_n\Big(\,\mathrm{meas}\,(D \setminus D_{\gamma_1,\gamma_2,\tau_*} \times \mathbb{T}^n) +\ \mathrm{meas}\,(\,\mathrm{Re}\,(D_r) \setminus D) \times \mathbb{T}^n\Big), \tag{2.34}$$

where $D_{\gamma_1,\gamma_2,\tau_*}$ *denotes the* ω_0*–preimage of* $\mathcal{D}_{\gamma_1,\gamma_2,\tau_*}$ *in* D *and* c_n *can be taken to be* $c_n = (1 + (1 + 2^8 n\hat{E})^{2n})^2$. *Finally, the following uniform estimates hold, on* $\mathbb{T}^n \times \Omega_*$

$$|v_1(\cdot;\omega) - I_1^0(\omega)| \le 10n\Big(\frac{\bar{M}_1}{\bar{M}} + \frac{\hat{M}}{M}\Big)\hat{E}\,\hat{\rho}\ ,\ |v_2(\cdot;\omega) - I_2^0(\omega)| \le 10n\Big(\frac{\bar{M}_2}{\bar{M}} + \frac{\hat{M}}{M}\Big)\hat{E}\,\hat{\rho}\ ,$$
$$|u(\cdot;\omega)| \le 2\,\hat{E}\,s \tag{2.35}$$

where v_i *denotes the projection of* $v \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ *over* $\mathbb{R}^{n_i}$ *and* $I^0(\omega) = (I_1^0(\omega), I_2^0(\omega)) \in D$ *is the* ω_0*–pre–image of* $\omega \in \Omega_*$.

This result is proved in Appendix B.

3 Proof of Theorem 1.4

In this section we prove Theorem 1.4 (and hence Theorem 1.2; compare the remark following the formulation of Theorem 1.4 in § 1, **1.4**).

In what follows, "C" denotes suitably positive constants greater than one independent of ϵ and μ, $\bar{\gamma}$, γ_1, $\bar{\gamma}_2$ but which may depend on n_1, n_2, H_0, s_0, *etc.* Without loss of generality, we may assume that H has an analytic extension to a domain $\mathcal{P}_{\rho_0,\epsilon_0,s_0} := V_{\rho_0} \times \mathbb{T}^{n_1}_{s_0} \times B_{\epsilon_0}$ with $s_0 < 1$ and with $\bar{\omega}_0 := \partial H_0$ a

[16] *I.e.,* $\bar{M}_i \ge \sup_{D_\rho} \|T_i\|$, $\quad i = 1,\ 2$, $\quad$ if $\quad U^{-1} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$.

diffeomorphism of V_{ρ_0}. We can also assume that the perturbation P has sup–fourier norm $\|P\|_{\rho_0,\epsilon_0,s_0} \equiv 1$ up to change the definition of μ.

Preliminary step. In view of (A2′) on p. 3, we can assume that the quadratic part of $P_{\rm av}(p,q;I)$ is in standard form $P_0(I) + \sum_{1\le i\le n_2} \Omega_i(I)\, r_i + o_2$, where $\Omega_i(I)$ are the first order Birkhoff invariants; compare [17]. Furthermore (again by (A2′)), since the $\Omega_i(I)$ are non resonant up to the order four, by Birkhoff theory (compare § 2.2 above), one can find a symplectic transformation, $O(|(p,q)|^2)$ close to the identity, which transforms the original Hamiltnian into[17] (1.1), with $P_{\rm av}$ as in the standard form in (A2″).

Step 1 *Averaging over the "fast angles" φ's*

Let $0 < \epsilon < e^{-1/5}$. The first step consists in removing, in H, the dependence on φ up to high orders (namely, up to $O(\mu\,\epsilon^5)$). To do this, we use Averaging theory (Proposition 2.1 above), with $\ell_1 = n_1$, $\ell_2 = 0$, $m = n_2$ $h = H_0$, $g \equiv 0$, $f = \mu P$, $B = B' = \{0\}$, $r_p = r_q = \epsilon_0$, $s = s_0$, $\bar s = 0$, $\Lambda = \{0\}$, $\bar K$ such that

$$e^{-\bar K s_0/6} := \epsilon^5 \qquad i.e.\ , \qquad \bar K = \frac{30}{s_0}\log\epsilon^{-1}\ , \tag{3.36}$$

$A = \bar D$, $r = \bar\rho$, where $\bar D$, $\bar\rho$ are defined as follows. Let $\tau > n_1$, $\bar M := \max_{i,j}\sup_{V_{\rho_0}} |\partial^2_{ij} H_0(I)|$, $\gamma_* \ge \max\{1, 2^5(30/s_0)^{\tau+1}\sqrt{c_{n_2}\bar M}\}$, $\bar\gamma \ge \gamma_*\sqrt{\mu}(\log\epsilon^{-1})^{\tau+1}$. Then, take

$$\bar D := \bar\omega_0^{-1}\Big(\mathcal{D}_{\bar\gamma,\tau}\Big) \cap V \qquad \text{and} \quad \bar\rho := \min\Big\{\frac{\bar\gamma}{2\bar M \bar K^{\tau+1}},\ \rho_0\Big\}\ , \tag{3.37}$$

where $\mathcal{D}_{\bar\gamma,\tau} \subset \mathbb{R}^{n_1}$ is defined just before § 2.1. From the Diophantine inequality it follows that $\bar\gamma \le |\bar\omega_0|_\infty$, so that

$$\gamma_*\sqrt{\mu}(\log\epsilon^{-1})^{\tau+1} \le \bar\gamma \le \|\bar\omega_0\|_{\rho_0}\ .$$

By the choice of $\bar D$, the following standard measure estimate holds

$$\text{meas}\Big(V\setminus\bar D\Big) \le C\bar\gamma\,\text{meas}\,(V) \tag{3.38}$$

where C depends on the C^1–norm of H_0. By the previous choices, when $I \in \bar D_{\bar\rho}$, the unperturbed frequency map $\bar\omega_0 = \partial H_0$ verifies (2.22), with $\alpha_1 = \alpha_2 = \bar\alpha := \frac{\bar\gamma}{2\bar K^\tau}$, in fact:

$$\inf_{\substack{J\in\bar D_{\bar\rho},\\ 0<|k|\le\bar K}} |\bar\omega_0(J)\cdot k| \ge \inf_{\substack{I\in\bar D,\\ 0<|k|\le\bar K}} |\bar\omega_0(I)\cdot k| - \sup_{\substack{J\in\bar D_{\bar\rho}, I\in\bar D,\\ 0<|k|\le\bar K}} |(\bar\omega_0(J)-\bar\omega_0(I))\cdot k| \ge \frac{\bar\gamma}{\bar K^\tau} - \bar\rho\bar K\bar M \ge \frac{\bar\gamma}{2\bar K^\tau}\ . \tag{3.39}$$

The smallness condition (2.23) is easily checked, provided $E = \mu$ is chosen small enough, because the choice of $\bar\gamma$ and γ_* implies

$$\mu\frac{2^7 c_{n_2}\bar K s_0}{\bar\alpha d} \le \max\Big\{\frac{1}{2},\ C\sqrt{\mu}\Big\} < 1\ .$$

Condition $\bar K s_0 \ge 6$ is trivially satisfied. Thus, by Proposition 2.1, we find a real–analytic symplectomorphism

$$\bar\phi:\ (\bar I,\bar\varphi,\bar p,\bar q) \in W_{\bar v,\bar s} \to (I,\varphi,p,q) \in W_{v_0,s_0} \quad \bar v := v_0/2 := (\bar\rho/2,\epsilon_0/2),\ \bar s := s_0/6$$

where $W_{v_0,s_0} := \bar D_{\rho_0}\times\mathbb{T}^{n_1}_{s_0}\times B_{\epsilon_0}$ ($v_0 = (\rho_0,\epsilon_0)$), and, by the choice of $\bar K$ in (3.36), H is transformed into[18]

$$\begin{aligned}\bar H(\bar I,\bar\varphi,\bar p,\bar q) &= H\circ\bar\phi(\bar I,\bar\varphi,\bar p,\bar q)\\ &= H_0(\bar I) + \mu\,\bar N(\bar I,\bar p,\bar q) + \mu e^{-\bar K s/6}\,\bar P(\bar I,\bar\varphi,\bar p,\bar q)\\ &= H_0(\bar I) + \mu\,\bar N(\bar I,\bar p,\bar q) + \mu\epsilon^5\,\bar P(\bar I,\bar\varphi,\bar p,\bar q)\ .\end{aligned} \tag{3.40}$$

[17] By abuse of notations, we use the same name for the variables, but, strictly speaking, they differ from the original variables by a quantity of $O(|(p,q)|^2)$ in (p,q), $O(|(p,q)|^3)$ in φ (the actions I are the same).

[18] For simplicity of notation, we do not write explicitely the dependence on $\mu,\epsilon,\bar\gamma$, that is, we write $H(\bar I,\bar\varphi,\bar p,\bar q)$, *etc.* , in place of $H(\bar I,\bar\varphi,\bar p,\bar q;\mu,\epsilon,\bar\gamma)$, *etc.*

By (2.25), $\|\bar{P}\|_{\bar{v},\bar{s}} \le C$ and

$$\sup_{\bar{D}_{\bar{\rho}/2}} |\bar{N} - P_{\rm av}| \le C \frac{\mu(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2} \ . \tag{3.41}$$

In view of (2.26), the transformation $\bar{\phi}$ verifies

$$|I - \bar{I}|,\ |p - \bar{p}|,\ |q - \bar{q}| \le C \frac{\mu(\log \epsilon^{-1})^{\tau}}{\bar{\gamma}} \ , \quad |\varphi - \bar{\varphi}| \le C \frac{\mu(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2} \ . \tag{3.42}$$

Remark 3.1 The right hand sides of (3.41) and (3.42) can be made small as we please, provided μ and ϵ are small and $\bar{\gamma}$ is chosen suitably. The precise choice will be discussed below.

Step 2 *Determination of the elliptic equilibrium for the "secular system"*
Since $P_{\rm av}$ has a 4–non resonant and non–degenerate elliptic equilibrium point at 0 and, in view of (3.41), $\bar{N} - P_{\rm av}$ is of order $\mu(\log \epsilon^{-1})^{2\tau+1}\bar{\gamma}^{-2}$, using the Implicit Function Theorem and standard Cauchy estimates[19], for small values of this parameter, for any fixed $\bar{I} \in \bar{D}_{\bar{\rho}/2}$, $\bar{N}$ also has a $\mu(\log \epsilon^{-1})^{2\tau+1}\bar{\gamma}^{-2}$–close–to–0 elliptic equilibrium point, which we call $(p^0(I), q^0(I))$. We can thus assume that $|(p^0(I), q^0(I))| < \epsilon_0/4$ for any I and consider a small neighborhood of radius $0 < \epsilon < \epsilon_0/4$ around $(p^0(I), q^0(I))$. We let

$$\tilde{\phi} : (\tilde{I}, \tilde{\varphi}, \tilde{p}, \tilde{q}) \in \ W_{\tilde{v},\tilde{s}} \to (\bar{I}, \bar{\varphi}, \bar{p}, \bar{q}) \in W_{\bar{v},\bar{s}} \quad \tilde{v} := (\bar{\rho}/4, \epsilon),\ \ \tilde{s} := s_0/12$$

be the transformation having as generating function

$$\tilde{s}(\tilde{I}, \tilde{p}, \bar{\varphi}, \bar{q}) = \tilde{I} \cdot \bar{\varphi} + \Big(\tilde{p} + p^0(\tilde{I})\Big) \cdot \Big(\bar{q} - q^0(\tilde{I})\Big) \ ,$$

which acts as the identity on the $\tilde{I}$–variables, while shifts the equilibrium point into the origin (and suitably lifts the angles $\tilde{\varphi}$) accordingly to

$$\bar{I} = \tilde{I} \ , \quad \bar{p} = p^0(\tilde{I}) + \tilde{p} \ , \quad \bar{q} = q^0(\tilde{I}) + \tilde{q} \ , \ \bar{\varphi} = \tilde{\varphi} - \partial_{\tilde{I}}\Big(\tilde{p} + p^0(\tilde{I})\Big) \cdot \Big(\bar{q} - q^0(\tilde{I})\Big) \ .$$

The transformation $\tilde{\phi}$ is close to the identity, since $\bar{I} = \tilde{I}$ and

$$|\bar{p} - \tilde{p}|,\ |\bar{q} - \tilde{q}| \le \frac{C}{\bar{\gamma}^2}\, \mu(\log \epsilon^{-1})^{2\tau+1} \ , \quad |\bar{\varphi} - \tilde{\varphi}| \le C \max\Big\{\frac{\epsilon^2(\log \epsilon^{-1})^{\tau+1}}{\bar{\gamma}}, \frac{\mu\epsilon(\log \epsilon^{-1})^{3\tau+2}}{\bar{\gamma}^3}\Big\} \ . \tag{3.43}$$

Let us check, for example, the bound on $|\bar{\varphi} - \tilde{\varphi}|$ (as the other ones are immediate): If

$$D(\mu, \epsilon, \bar{\gamma}) := \{(\tilde{I}, \tilde{p}, \bar{q}) :\ \tilde{I} \in \bar{D}_{\bar{\rho}(\epsilon)/4} \ , \quad (\tilde{p},\ \bar{q} - q^0(\tilde{I})) \in B_\epsilon\} \ ,$$

then, by Cauchy estimates,

$$\begin{aligned}
|\bar{\varphi}(\tilde{I}, \tilde{p}, \tilde{q}) - \tilde{\varphi}| &= \Big|\partial_{\tilde{I}}\Big(\big(\tilde{p} + p^0(\tilde{I})\big) \cdot \big(\bar{q} - q^0(\tilde{I})\big)\Big)\Big|_{\bar{q}=q^0+\tilde{q}} \\
&\le \sup_{D(\mu,\epsilon,\bar{\gamma})} \Big|\partial_{\tilde{I}}\Big(\big(\tilde{p} + p^0(\tilde{I})\big) \cdot \big(\bar{q} - q^0(\tilde{I})\big)\Big)\Big| \\
&\le \tilde{C} \frac{\sup_{D(\mu,\epsilon,\bar{\gamma})} \Big|\Big(\tilde{p} + p^0(\tilde{I})\Big) \cdot \Big(\bar{q} - q^0(\tilde{I})\Big)\Big|}{\bar{\rho}(\epsilon)/4} \\
&\le \tilde{C} \frac{\Big(\epsilon + \frac{\mu(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}\Big)\epsilon}{\frac{\bar{\gamma}}{(\log \epsilon^{-1})^{\tau+1}}} \\
&\le C \max\Big\{\frac{\epsilon^2(\log \epsilon^{-1})^{\tau+1}}{\bar{\gamma}} \ , \quad \frac{\mu\epsilon(\log \epsilon^{-1})^{3\tau+3}}{\bar{\gamma}^3}\Big\} \ .
\end{aligned}$$

[19] See, e.g., [4, Lemma A.1].

By construction, the transformation $\tilde{\phi}$ puts $\bar{H}$ into the form

$$\tilde{H} := \bar{H} \circ \tilde{\phi} = H_0(\tilde{I}) + \mu\tilde{N}(\tilde{I}, \tilde{p}, \tilde{q}) + \mu\epsilon^5\tilde{P}(\tilde{I}, \tilde{\varphi}, \tilde{p}, \tilde{q}), \text{ with } \tilde{N} := \bar{N} \circ \tilde{\phi}\ ,\ \tilde{P} := \bar{P} \circ \tilde{\phi}.$$

Observe that $\|\tilde{P}\|_{\tilde{v},\tilde{s}} \leq C$ and $\tilde{N}$ has a 4–non resonant and non–degenerate elliptic equilibrium point into the origin of the $(\tilde{p}, \tilde{q})$–coordinates.

Step 3 *Symplectic diagonalization of the secular system*

The standard "diagonal form" (2.28) can be achieved by a symplectic diagonalization as in [17]. In fact, by [17], one can find a symplectic map

$$\hat{\phi} : \quad (\hat{I}, \hat{\varphi}, \hat{p}, \hat{q}) \in W_{\hat{v},\hat{s}} \to (\tilde{I}, \tilde{\varphi}, \tilde{p}, \tilde{q}) \in W_{\tilde{v},\tilde{s}} \quad \hat{v} := (\bar{\rho}/8, \epsilon/2),\ \hat{s} := s_0/24$$

which acts as the identity on the $\hat{I}$–variables, is linear in the variables $(\hat{p}, \hat{q})$ and close to the identity in the sense

$$|\tilde{p} - \hat{p}|,\ |\tilde{q} - \hat{q}| \leq C\frac{\mu\epsilon(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}\ , \quad |\tilde{\varphi} - \hat{\varphi}| \leq C\frac{\mu\epsilon^2(\log \epsilon^{-1})^{3\tau+2}}{\bar{\gamma}^3}. \tag{3.44}$$

Such estimates are a consequence of the assumptions on $P_{\rm av}$ (compare the preliminary step above), the estimate $\tilde{N} = P_{\rm av} + O(\frac{\mu\epsilon^2(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2})$, for which $\tilde{N}$ is $O(\frac{\mu\epsilon^2(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2})$–close to be diagonal and Cauchy estimates[20] . Moreover, one has that[21] $\hat{N}(\hat{I}, \hat{p}, \hat{q}) := \tilde{N} \circ \hat{\phi}(\hat{I}, \hat{p}, \hat{q}) = \tilde{P}_0(\hat{I}) + \hat{\Omega}(\hat{I}) \cdot \hat{r} + \hat{R}$ where

$$|\hat{\Omega} - \Omega|, \quad |\hat{R}| \leq C\frac{\mu\epsilon(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2} \tag{3.45}$$

and $\hat{R}$ having a zero of order 3 for $(\hat{p}, \hat{q}) = 0$ and that $\hat{\phi}$ transforms $\tilde{H}$ into

$$\hat{H} := \tilde{H} \circ \hat{\phi} = H_0(\hat{I}) + \mu\hat{N}(\hat{I}, \hat{p}, \hat{q}) + \mu\epsilon^5\hat{P}(\hat{I}, \hat{\varphi}, \hat{p}, \hat{q}), \quad (\hat{P} := \tilde{P} \circ \hat{\phi})\ .$$

Step 4 *Birkhoff normal form of the secular part*

By Proposition 2.2 (and subsequent remark) there exists a Birkhoff transformation

$$\check{\phi} : \quad (\check{I}, \check{\varphi}, \check{p}, \check{q}) \in W_{\check{v},\check{s}} \to (\hat{I}, \hat{\varphi}, \hat{p}, \hat{q}) \in W_{\hat{v},\hat{s}} \qquad \check{v} := (\bar{\rho}/16, \epsilon/4),\ \check{s} := s_0/48$$

which acts as the identity on the $\check{I}$–variables, is close to the identity as (compare (2.31):

$$|\hat{p} - \check{p}|,\ |\hat{q} - \check{q}| \leq C\frac{\mu\epsilon^2(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}\ , \quad |\hat{\varphi} - \check{\varphi}| \leq C\frac{\mu\epsilon^3(\log \epsilon^{-1})^{3\tau+2}}{\bar{\gamma}^3}\ . \tag{3.46}$$

The previous estimates follow from the fact that, in (2.31), the coefficients $c_{\alpha,\beta}$ of the non–normal part of can be upper bounded by $m_1 := \frac{\mu(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}$, uniformly in I; α can be taken of order 1 in ϵ and μ; ρ_0 of $O(\frac{\bar{\gamma}}{(\log \epsilon^{-1})^{\tau+1}})$. Furthermore, $\check{\phi}$ puts $\hat{N}$ into Birkhoff normal form up to order 4, hence, transforms $\hat{H}$ into the form

$$\begin{aligned} \check{H} := \hat{H} \circ \check{\phi} \ &:= \ H_0(\check{I}) + \mu\left(\tilde{P}_0(\check{I}) + \hat{\Omega}(\check{I}) \cdot \check{r} + \frac{1}{2}\check{r}\check{\beta}(\check{I})\check{r} + O(|(\check{p}, \check{q})|^5)\right) + \mu\epsilon^5\hat{P} \circ \check{\phi} \\ &=: \ H_0(\check{I}) + \mu\check{N}(\check{I}, \check{r}) + \mu\epsilon^5\check{P}(\check{I}, \check{\varphi}, \check{p}, \check{q})\ , \end{aligned}$$

where $\check{N}(\check{I}, \check{r}) = \tilde{P}_0(\check{I}) + \hat{\Omega}(\check{I}) \cdot \check{r} + \frac{1}{2}\check{r}\check{\beta}(\check{I})\check{r}$, $\check{r}_i = \dfrac{\check{p}_i^2 + \check{q}_i^2}{2}$, $\|\check{P}\|_{\check{v},\check{s}} \leq C$.

[20] The generating function of this transformation is $\frac{\mu\epsilon^2(\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}$–close to the generating function $\hat{I} \cdot \tilde{\varphi} + \hat{p} \cdot \tilde{q}$ of the identity map. Taking the derivatives and using Cauchy estimates, with a loss of analiticity $\sim C\epsilon$ in $(\hat{p}, \hat{q})$, and $\sim \frac{\bar{\gamma}}{(\log(\epsilon^{-1}))^{\tau+1}}$ in the $\hat{\varphi}$, we find (3.44).

[21] $\hat{r}_i := (\hat{p}_i^2 + \hat{q}_i^2)/2$.

Step 5 *Global action–angle variables for the full system*
We finally introduce a set of action–angle variables using symplectic polar coordinates. Fix the real n_2–dimensional annulus[22]

$$\mathcal{A}(\epsilon) := \Big\{ J \in \mathbb{R}^{n_2} : \check{c}_1 \epsilon^{5/2} < J_i < \check{c}_2 \epsilon^2 \ , \quad 1 \le i \le n_2 \Big\} \tag{3.47}$$

where $\check{c}_1$ will be fixed later on so as to maximize the measure of preserved tori and ϵ small enough with respect to $1/\check{c}_1$, while $\check{c}_2$ is a constant depending only on the dimensions. Let

$$\mathcal{D} := \bar{D} \times \mathcal{A}(\epsilon) \ , \quad \check{\rho} := \min\{\check{c}_1 \epsilon^{5/2}/2, \ \bar{\rho}/16\} \ , \quad \check{s} := \hat{s} = s_0/48 \ , \tag{3.48}$$

where $\bar{D}$ is the set in (3.37). On $\mathcal{D}_{\check{\rho}} \times \mathbb{T}^n_{\check{s}}$, let $\check{\phi} : (J, \psi) = \Big((J_1, J_2), (\psi_1, \psi_2)\Big) \to (\check{I}, \check{\varphi}, \check{p}, \check{q})$ be defined by

$$J_1 = \check{I} \ , \quad \psi_1 = \check{\varphi} \ , \quad \check{p}_i = \sqrt{2 J_{2i}} \cos \psi_{2i} \ , \quad \check{q}_i = \sqrt{2 J_{2i}} \sin \psi_{2i} \quad 1 \le i \le n_2 \ . \tag{3.49}$$

For ϵ small enough, $(\check{p}, \check{q}) \in B_{\epsilon/4}$. The transformation $\check{\phi}$ puts $\check{H}$ into the form

$$\begin{aligned} \breve{H}(J,\psi) &:= \check{H} \circ \check{\phi} = H_0(J_1) + \mu \breve{N}(J) + \mu \epsilon^5 \breve{P}(J,\psi) \ , \quad \text{where} \\ \breve{N} &:= \check{N} \circ \check{\phi} = \tilde{P}_0(J_1) + \hat{\Omega}(J_1) \cdot J_2 + \frac{1}{2} J_2 \check{\beta}(J_1) J_2 \ , \quad \breve{P} := \hat{P} \circ \check{\phi} \end{aligned}$$

From the above construction there follows that the transformation

$$\phi := \bar{\phi} \circ \tilde{\phi} \circ \hat{\phi} \circ \check{\phi} \circ \breve{\phi} : (J, \psi) \to (I, \varphi, p, q) \tag{3.50}$$

is well defined[23] and verifies

$$\begin{cases} |I - J_1| \le C \frac{\mu (\log \epsilon^{-1})^\tau}{\bar{\gamma}} \\ |\varphi - \psi_1| \le C \max \Big\{ \frac{\mu (\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}, \ \frac{\epsilon^2 (\log \epsilon^{-1})^{\tau+1}}{\bar{\gamma}}, \ \frac{\mu \epsilon (\log \epsilon^{-1})^{3\tau+3}}{\bar{\gamma}^3} \Big\} \\ |p_i - p_i^0 - \sqrt{2 J_{2i}} \cos \psi_{2i}|, \ |q_i - q_i^0 - \sqrt{2 J_{2i}} \sin \psi_{2i}| \le C \max \Big\{ \frac{\mu (\log \epsilon^{-1})^\tau}{\bar{\gamma}}, \ \frac{\mu \epsilon (\log \epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2} \Big\} . \end{cases} \tag{3.51}$$

Step 6 *Construction of the Kolmogorov set and estimate of its measure*
Fix γ_1 and $\gamma_2 = \mu \bar{\gamma}_2$, with γ_1, $\bar{\gamma}_2$ satisfying $\mu \bar{\gamma}_2 \le \gamma_1$ and (1.17). We apply now the two–scale KAM Theorem (Proposition 2.3), with (compare Step 5 above)

$$\mathrm{H} = \breve{H} \ , \quad \mathrm{h} = H_0(J_1) + \mu \breve{N}(J) \ , \quad \mathrm{f} = \mu \epsilon^5 \breve{P}(J, \psi) \ , \quad D = \mathcal{D} \ , \quad \rho = \check{\rho}$$

and $s = \check{s}/5$, $\bar{s} = 4\check{s}/5$. It is easy to check that, for small values of[24] $\mu (\log \epsilon^{-1})^{2(\tau+1)} \bar{\gamma}^{-2}$, the frequency map $\omega_\mu := \partial (H_0(J_1) + \mu \breve{N}(J))$ is a diffeomorphism of $\mathcal{D}_{\check{\rho}}$, with non singular hessian matrix $\partial^2 (H_0(J_1) + \mu \breve{N}(J))$. Then, we see that (for a suitable constant C) we can take M, $\hat{M}$, $\cdots$, $\bar{M}_2$ in Proposition 2.3 as follows:

$$M = C \ , \quad \hat{M} = C \mu \ , \quad \bar{M} = C \mu^{-1} \ , \quad E = C \mu \epsilon^5 \ , \quad \bar{M}_1 = C \ , \quad \bar{M}_2 = C \mu^{-1} \ .$$

Then,

$$L \le C \mu^{-1} \ , \qquad K \le C \log{(\epsilon^5 / {\gamma_1}^2)^{-1}}$$

and (recall also (3.37))

$$\hat{\rho} \ge c \ \min \Big\{ \frac{\gamma_1}{(\log{(\epsilon^5/{\gamma_1}^2)^{-1}})^{\tau+1}} \ , \ \frac{\bar{\gamma}_2}{(\log{(\epsilon^5/{\gamma_1}^2)^{-1}})^{\tau+1}} \ , \frac{\bar{\gamma}}{(\log \epsilon^{-1})^{\bar{\tau}+1}} \ , \ \check{c}_1 \epsilon^{5/2} \ , \ \rho_0 \Big\} . \tag{3.52}$$

[22]This is needed to avoid the singularity introduced by the polar coordinates. Notice that $J_i \sim \epsilon^2$ compared to $(p,q) \sim \epsilon$.
[23]If $\bar{\gamma}$ is chosen as to satisfy the first inequality in (1.17), then, the right hand sides of (3.41), (3.42), (3.43), (3.44), (3.45) and (3.51) can all be bounded by $1/\gamma_*$. Choosing γ_* big enough, the quantities involved are small as we please.
[24]Such inequality is implied by $\gamma_* \sqrt{\mu} \log(\epsilon^{-1})^{\tau+1} < \bar{\gamma}$, which appears in (1.17).

Finally,

$$\hat{c}\hat{E} \le C \max\left\{\epsilon^5 \left(\log\left(\frac{{\gamma_1}^2}{\epsilon^5}\right)\right)^{2(\tau+1)} \max\{\frac{1}{\gamma_1^2}, \frac{1}{\bar{\gamma}_2^2}\}\ , \frac{\epsilon^5(\log\epsilon^{-1})^{2(\tau+1)}}{\bar{\gamma}^2}\ ,\ \frac{1}{\check{c}_1^2}\ ,\ \frac{\epsilon^5}{\rho_0}\right\}\ , \tag{3.53}$$

with a constant C not involving $\check{c}_1$. Then, from (1.17) it follows that

$$\hat{c}\hat{E} < C \max\left\{\frac{1}{\gamma_*}, \frac{1}{\check{c}_1^2}, \epsilon^5\right\} < 1 \tag{3.54}$$

provided γ_*, $\check{c}_1^2 > C$ and $\epsilon^5 < C^{-1}$. Finally, since the KAM condition $\hat{c}\hat{E} < 1$ is met, Proposition 2.3 holds in this case. In paricular, for any ω in the set $\Omega_* := \mathcal{D}_{\gamma_1,\mu\bar{\gamma}_2,\tau_*} \cap \omega_\mu(\mathcal{D})$, we find a real–analytic embedding

$$\phi_\omega : \quad \mathbb{T}^n \to \mathrm{T}_\omega := \phi_\omega(\mathbb{T}^n) \subset \mathrm{Re}\,(\mathcal{D}_r) \times \mathbb{T}^n$$

with $r \le \hat{C}\hat{\rho} \le C\bar{\gamma}_2$ such that, on T_ω, the $\breve{H}$–flow analytically conjugated to $\vartheta \to \vartheta + \omega\, t$. We set $\mathcal{T}_\omega := \phi(\mathrm{T}_\omega)$, where ϕ is the symplectic transformation defined in (3.50). Using (3.51) and (2.35), the parametric equations of $\mathcal{T}_\omega$ may be written as

$$\begin{aligned} & I = J_1^0 + \tilde{J}_1 \quad \text{with} \quad |\tilde{J}_1| \le C\left(\frac{\mu(\log\epsilon^{-1})^\tau}{\bar{\gamma}} + \mu\hat{\rho}\right) \\ & (p_i - p_i^0)^2 + (q_i - q_i^0)^2 = J_{2i}^0 + \tilde{J}_{2i} \quad \text{with} \quad \check{c}_1\epsilon^{5/2} < J_{2i}^0 < \check{c}_2\epsilon^2, \\ & \text{and} \quad |\tilde{J}_{2i}| \le C\Big(\frac{\mu^2(\log\epsilon^{-1})^{2\tau}}{\bar{\gamma}^2} + \frac{\mu^2\epsilon^2(\log\epsilon^{-1})^{4\tau+2}}{\bar{\gamma}^4} + \hat{\rho}\Big)\ , \end{aligned}$$

where (J_1^0, J_2^0) is the ω_μ pre–image of ω, $\hat{\rho}$ is much smaller than $\check{c}_1\epsilon^{5/2}$ (compare (3.52)); finally, by (3.51), $|(p_0, q_0)| \le C\frac{\mu(\log\epsilon^{-1})^{2\tau+1}}{\bar{\gamma}^2}$.

It remains to *estimate the measure of the Kolmogorov set*

$$\mathcal{K} := \phi(\mathrm{K}) = \bigcup_{\omega\in\Omega_*} \mathcal{T}_\omega\ , \quad \text{where} \quad \mathrm{K} := \bigcup_{\omega\in\Omega_*} \mathrm{T}_\omega.$$

namely, (1.18). Let $\mathcal{D}^*_{\gamma_1,\mu\bar{\gamma}_2,\tau_*} := \omega_\mu^{-1}(\mathcal{D}_{\gamma_1,\mu\bar{\gamma}_2,\tau_*}) \cap \mathcal{D}$, where $\mathcal{D}$ is the set in (3.48), with $\bar{D}$ as in (3.37) and $\mathcal{D}_{\gamma_1,\mu\bar{\gamma}_2,\tau_*}$ is defined just before § 2.1. Then, by (2.34) and because ϕ is volume preserving, we have

$$\begin{aligned} & \mathrm{meas}\,\mathcal{K} = \mathrm{meas}\,\mathrm{K} \\ & \ge \mathrm{meas}\,(\,\mathrm{Re}\,(\mathcal{D}_r) \times \mathbb{T}^n) - C\Big(\mathrm{meas}\,(\mathcal{D} \setminus \mathcal{D}^*_{\gamma_1,\mu\bar{\gamma}_2,\tau_*} \times \mathbb{T}^n) + \mathrm{meas}\,(\,\mathrm{Re}\,(\mathcal{D}_r) \setminus \mathcal{D} \times \mathbb{T}^n)\Big). \end{aligned} \tag{3.55}$$

Now, let $\mathcal{V} := V \times B^{n_2}_{\check{c}_2\epsilon^2}$, where $B^{n_2}_{\check{c}_2\epsilon^2}$ denotes the open set $\{|J_i| < \check{c}_2\epsilon^2\}$. Observe that $\mathcal{D} \subset \mathcal{V}$; define $\mathcal{P}_\epsilon := V \times \mathbb{T}^{n_1} \times \{p_i^2 + q_i^2 < \epsilon^2\}$ (compare (1.5)). Then, by the estimate (3.38) and the definition (3.47) of $\mathcal{A}(\epsilon)$,

$$\begin{aligned} \mathrm{meas}\,(\mathcal{D}_r \times \mathbb{T}^n) & \ge \mathrm{meas}\,(\mathcal{D} \times \mathbb{T}^n) \\ & = \mathrm{meas}\,\bar{D}\,\mathrm{meas}\,(\mathcal{A}(\epsilon))\,\mathrm{meas}\,(\mathbb{T}^n) \\ & \ge (1 - C\bar{\gamma} - C\epsilon^{n_2/2})\,\mathrm{meas}\,(\mathcal{V} \times \mathbb{T}^n) \\ & = (1 - C\bar{\gamma} - C\epsilon^{n_2/2})\,\mathrm{meas}\,(\mathcal{P}_{\sqrt{2\check{c}_2}\epsilon}) \end{aligned} \tag{3.56}$$

Similarly, denoting for short $B := B^{n_2}_{\check{c}_2\epsilon^2}$, one has that[25]

$$\mathrm{meas}\,(\,\mathrm{Re}\,(V_r) \setminus V) \le C\bar{\gamma}_2\,\mathrm{meas}\,V\ ,\ \ \mathrm{meas}\,(B_r \setminus B) \le \mathrm{meas}\,(B_{C\bar{\gamma}_2} \setminus B) \le C'\frac{\bar{\gamma}_2}{\epsilon^2}\,\mathrm{meas}\,B$$

$$\mathrm{meas}\,(\mathcal{V} \setminus \mathcal{D}) \le C(\bar{\gamma} + \epsilon^{n_2/2})\,\mathrm{meas}\,\mathcal{V}\ .$$

[25]Recall that $r < C\bar{\gamma}_2$.

Thus,

$$\begin{aligned}\text{meas Re}\,(\mathcal{D}_r)\setminus\mathcal{D}\times\mathbb{T}^n &\le \text{meas Re}\,(\mathcal{V}_r)\setminus\mathcal{D}\times\mathbb{T}^n \\ &\le \text{meas Re}\,(\mathcal{V}_r)\setminus\mathcal{V}\times\mathbb{T}^n + \text{meas}\,\mathcal{V}\setminus\mathcal{D}\times\mathbb{T}^n \\ &\le C(\frac{\bar\gamma_2}{\epsilon^2}+\bar\gamma+\epsilon^{n_2/2})\,\text{meas}\,(\mathcal{V}\times\mathbb{T}^n) \\ &= C(\frac{\bar\gamma_2}{\epsilon^2}+\bar\gamma+\epsilon^{n_2/2})\,\text{meas}\,(\mathcal{P}_{\sqrt{2\check c_2}\epsilon})\ .\end{aligned} \tag{3.57}$$

Finally, the frequency map $\omega_\mu := (\partial_{J_1}(H_0+\mu\check N), \mu\partial_{J_2}\check N)$ is a diffeomorphism of a $\check\rho$-neighborhood of $\bar D\times\overline{B}^{n_2}_{\check c_2\epsilon^2}$. Note that ω_μ as a function of J_1 is defined on $\bar D_{\check\rho}$ and as a function of J_2 is a polynomial; notice also that $\overline{B}^{n_2}_{\check c_2\epsilon^2}$ is just the full closed ball around the annulus $\mathcal{A}(\epsilon)$ (compare (3.47)). Then, the measure of the set $\mathcal{D}\setminus\mathcal{D}^*_{\gamma_1,\mu\bar\gamma_2,\tau_*}$ does not exceed the measure of the $(\gamma_1,\mu\bar\gamma_2)$–resonant set for ω_μ in the set $\bar D\times\overline{B}^{n_2}_{\check c_2\epsilon^2}$. Such set of resonant points may be estimated by the following technical Lemma, whose proof is deferred to Appendix C.

Lemma 3.1 *Let $n_1,\ n_2\in\mathbb{N}$, $\tau>n:=n_1+n_2$, $\gamma_1,\ \gamma_2>0$, $0<\hat r<1$, $\bar D$ be a compact set. Let*

$$\omega=(\omega_1,\omega_2):\quad \bar D\times\overline{B}^{n_2}_{\hat r}\to\Omega\subset\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}$$

be a function which can be extended to a diffeomorphism on an open neighborhood of $\bar D\times\overline{B}^{n_2}_{\hat r}$, with ω_2 of the form

$$\omega_2(I_1,I_2)=\omega_{02}(I_1)+\beta(I_1)I_2$$

where $I_1\to\beta(I_1)$ is a $(n_2\times n_2)$–matrix, non singular on $\bar D$. Let

$$R_1>\max_{\bar D\times\overline{B}^{n_2}_{\hat r}}|\omega_1|\ ,\quad a>\max_{\bar D}\|\beta\|\ ,\quad c(n,\tau):=\sum_{0\neq k\in\mathbb{Z}^n}\frac{1}{|k|^\tau}\ ,$$

and denote

$$\mathcal{R}_{\gamma_1,\gamma_2,\tau}:=\Big\{I=(I_1,I_2)\in\bar D\times\overline{B}^{n_2}_{\hat r}:\quad \omega(I)\notin\mathcal{D}_{\gamma_1,\gamma_2,\tau}\Big\}\ .$$

Then,

$$\text{meas}\Big(\mathcal{R}_{\gamma_1,\gamma_2,\tau}\Big)\le\Big(c_1\gamma_1+c_2\,\frac{\gamma_2}{\hat r}\Big)\ \text{meas}\Big(\bar D\times B^{n_2}_{\hat r}\Big)$$

where

$$\begin{cases} c_1:=\max\limits_{\bar D\times\overline{B}^{n_2}_{\hat r}}\|(\partial\omega)^{-1}\|^n\,\dfrac{R_1^{n_1-1}}{\text{meas}\,\bar D}\ a^{n_2}\ c(n,\tau)\,p \\ c_2:=\max\limits_{\bar D}\|\beta^{-1}\|^{n_2}\,a^{n_2-1}\ c(n_2,\tau)\end{cases} \tag{3.58}$$

for a suitable integer p depending on $\bar D$ and ω_1.

By Lemma 3.1 with γ_1 as in Step 6, $\bar D$ as in (3.37) and

$$\hat r=\check c_2\epsilon^2,\quad \omega=\omega_{\text{resc}}:=\Big(\partial_{J_1}(H_0+\mu\check N),\partial_{J_2}\check N\Big)\ ,\quad \gamma_2=\bar\gamma_2\ ,\quad a=R_1=C\ ,$$

we see that

$$\begin{aligned}\text{meas}\,(\mathcal{D}\setminus\mathcal{D}^*_{\gamma_1,\mu\bar\gamma_2,\tau_*}\times\mathbb{T}^n) &\le \text{meas}\,(\mathcal{R}_{\gamma_1,\bar\gamma_2,\tau_*}\times\mathbb{T}^n) \\ &\le \max\{c_1,\frac{c_2}{\check c_2}\}(\gamma_1+\frac{\bar\gamma_2}{\epsilon^2})\,\text{meas}\,(\bar D\times B_{\check c_2\epsilon^2}\times\mathbb{T}^n) \\ &\le \max\{c_1,\frac{c_2}{\check c_2}\}(\gamma_1+\frac{\bar\gamma_2}{\epsilon^2})\,\text{meas}\,\mathcal{P}_{\sqrt{2\check c_2}\epsilon}\end{aligned} \tag{3.59}$$

with c_i independent of ϵ and μ. Then, in view of (3.55)÷(3.59), (1.18) follows, with ϵ replaced by $\sqrt{2\check c_2}\epsilon$. The proof of Theorem 1.4 is finished. ∎

4 Proof of Theorem 1.3

Since most of the arguments are similar (but simpler) than the ones used in the proof of Theorem 1.2, we will skip most technical details.

We can write $P_{\rm av}(p,q;I) = N(I,r) + \tilde P_{\rm av}(p,q;I)$, where

$$N(I,r) := P_0(I) + \sum_{1\le i\le n_2} \Omega_i(I) r_i + \frac{1}{2} \sum_{1\le i,j\le n_2} \beta_{ij}(I) r_i r_j$$

and, for a suitable $C>0$,

$$\sup_{B^{2n_2}_\epsilon \times V_{\rho_0}} |\tilde P_{\rm av}| \le C\epsilon^5 , \qquad \forall\, 0<\epsilon<\epsilon_0 . \tag{4.1}$$

Step 1 Fix $\tau > n_1$, $0<\epsilon<\epsilon_0$, and

$$\mu < \frac{\epsilon^6}{(\log(\epsilon^{-1}))^{2\tau+1}} , \qquad \bar\gamma \ge \Big(\frac{30}{s_0}\Big)^{2\tau+1} \frac{\sqrt{\mu}(\log \epsilon^{-1})^{2\tau+1}}{\epsilon^{5/2}} . \tag{4.2}$$

In place of Proposition 2.1, we use Lemma A.1 below, where we take $r_p = r_q = \epsilon_0$, $\rho = \bar\rho$ with $\bar\rho$ as in (3.37), $\rho_p = \rho_q = \epsilon_0/4$, $\sigma = s_0/6$ and the remaining quantities as in Step 1 of the proof of Theorem 1.4, namely, $\ell_1 = n_1$, $\ell_2 = 0$, $m = n_2$ $h = H_0$, $g \equiv 0$, $f = \mu P$, $A = \bar D$, $r = \bar\rho$, as in (3.37) $B = B' = \{0\}$, $s = s_0$, $\alpha_1 = \alpha_2 = \bar\alpha = \frac{\bar\gamma}{2\bar K^\tau}$, where $\bar K$ as in (3.36) and $\Lambda = \{0\}$. With such choice, the check of the non resonance assumption (A.1) for $\bar\omega_0 = \partial H_0$ in $\bar D_{\bar\rho}$ is the same as in Step 1 of the proof of Theorem 1.4 and the smallness condition (A.2) is implied by (4.2). Then, there exists a symplectic transformation $\bar\phi$ such that $H\circ\bar\phi = H_0 + g_+ + f_+$ as in Lemma A.1. Since g_+ coincides with $\mu P_{\rm av}$, on the domain $W_{\tilde v,\tilde s}$ (recall the definition of $W_{v,s}$ just above the (3.40)), where $\tilde v = (\bar\rho/2, \epsilon_0/2)$ and $\tilde s = 2s_0/3$, we find

$$\begin{aligned} \bar H(\bar I,\bar\varphi,\bar p,\bar q) &:= H\circ\bar\phi(\bar I,\bar\varphi,\bar p,\bar q) \\ &= H_0(\bar I) + \mu P_{\rm av}(\bar p,\bar q;\bar I) + \tilde P(\bar I,\bar\varphi,\bar p,\bar q) \\ &= H_0(\bar I) + \mu N(\bar I,\bar r) + \mu\tilde P_{\rm av}(\bar p,\bar q;\bar I) + \tilde P(\bar I,\bar\varphi,\bar p,\bar q) . \end{aligned} \tag{4.3}$$

By (A.6) below, the transformation $\bar\phi$ satisfies the estimates (3.42). Furthermore, by (4.2), the choice of $\bar K$ in (3.36) and (A.5) below, the function $\tilde P$ in (4.3) satisfies

$$\|\tilde P\|_{\tilde v,\tilde s} \le \bar C\mu \max\{\frac{\mu\bar K^{2\tau+1}}{\bar\gamma^2}, \frac{\mu\bar K^\tau}{\bar\gamma}\ e^{-\bar K s_0/6}\} \le \bar C\mu\epsilon^5 . \tag{4.4}$$

By such estimate and (4.1), the perturbation $\bar P := \mu\tilde P_{\rm av} + \tilde P$, on the smaller domain $W_{\bar v,\bar s}$, where $\bar v = (\bar\rho/2, \epsilon/2)$, $\bar s = \tilde s$, is bounded by $C\mu\epsilon^5$.

Step 2 and conclusion At this point, we proceed as in Steps 5 and 6 of Theorem 1.4, with $W_{\bar v,\bar s}$, N and $\bar P$ replacing, respectively, $W_{\check v,\check s}$, $\check N$ and $\mu\epsilon^5\check P$. Now, choose γ_* big enough (so that the KAM condition (3.53), (3.54) is satisfied), and fix γ, $\bar\gamma_2$ satisfying $\mu\bar\gamma_2 \le \gamma_1$ and last two lines in (1.17). Then, we can find a set of invariant tori

$$\mathcal K_* \subset \bar D_r \times \mathbb T^{n_1} \times \{2\check c_1\epsilon^{5/2} < p_i^2 + q_i^2 < 2\check c_2\epsilon^2\}_r \subset (\mathcal P_{\sqrt{2\check c_2}\epsilon})_r$$

(with $r < C\bar\gamma_2$) satisfying the measure estimate

$$\text{meas}\,(\mathcal P_{\sqrt{2\check c_2}\epsilon} \setminus \mathcal K_*) \le\ \text{meas}\,((\mathcal P_{\sqrt{2\check c_2}\epsilon})_r \setminus \mathcal K_*) \le C(\bar\gamma + \gamma_1 + \frac{\bar\gamma_2}{\epsilon^2} + \epsilon^{n_2/2})\,\text{meas}\,\mathcal P_{\sqrt{2\check c_2}\epsilon}) .$$

Finally, taking, as in the proof of Theorem 1.2, γ_1, $\bar\gamma_2$ as in (1.20), and choosing as $\bar\gamma$ the value in the right hand side of (4.2), the theorem is proved with $\mathcal K := \mathcal K_* \cap \mathcal P_{\check c_2\epsilon}$ and $\sqrt{2\check c_2}\epsilon$ replacing ϵ. ∎

A Averaging theory (Proposition 2.1)

In this appendix we generalize Proposition A.1 in [4] to a two–frequency–scale, as needed in Appendix B below. Proposition A.1 in [4] is based on the application of an "iterative lemma". The following lemma is the (easy) generalization of the iterative lemma (Lemma A.5) in [4] suitable for our purpose.

Lemma A.1 *Let* $0 < \alpha_2 \le \alpha_1$, $\ell = \ell_1 + \ell_2$ *with* $\ell_i \in \mathbb{N}$ *and let* Λ *be a sublattice of* $\mathbb{Z}^\ell$. *Let* $g = \sum_{k\in\Lambda} g_k(u) e^{ik\cdot\varphi}$ *and* $H(u,\varphi) = h(I) + g(u,\varphi) + f(u,\varphi)$ *be real–analytic on* $W_{v,s} := A_r \times B_{r_p} \times B'_{r_q} \times \mathbb{T}^\ell_s$, *where* $A \times B \times B' \subset (\mathbb{R}^{\ell_1} \times \mathbb{R}^{\ell_2}) \times \mathbb{R}^m \times \mathbb{R}^m$ *and* $v = (r, r_p, r_q)$. *Let* $\rho < r/2$, $\rho_p < r_p/2$, $\rho_q < r_q/2$, $\sigma < s/2$, $\nu := (\rho, \rho_p, \rho_q)$. *Suppose that* $k = (k_1, k_2) \in \mathbb{Z}^{\ell_1} \times \mathbb{Z}^{\ell_2}$ *and that*

$$|\omega(I)\cdot k| \ge \begin{cases} \alpha_1 & \text{if} \quad k_1 \ne 0 \\ \alpha_2 & \text{if} \quad k_1 = 0 \end{cases} \qquad \forall \quad k \in \mathbb{Z}^{\ell_1} \times \mathbb{Z}^{\ell_2} \ , \quad k \notin \Lambda \ , \quad |k| \le \bar K \ , \quad I \in A_r \ . \tag{A.1}$$

Assume also that the following "smallness condition" holds:

$$\|f\|_{v,s} < \frac{\alpha_2 \delta}{c_m} \ , \quad \text{where} \quad \delta := \min\{\rho\sigma, \ \rho_p \rho_q\} \ , \quad c_m := e(1+em)/2 \ . \tag{A.2}$$

Then,there exists a real–analytic symplectic transformation

$$\phi : \quad (I', \varphi', p', q') \in W_{v-2\nu, s-2\sigma} \to (I, \varphi, p, q) \in W_{v,\sigma} \ , \tag{A.3}$$

such that

$$H_+ := H \circ \phi = h + g_+ + f_+ \ , \tag{A.4}$$

with $g_+ - g = \Pi_\Lambda T_{\bar K} f$ *and*

$$\|f_+\|_{v-2\nu, s-2\sigma} \le \left(1 - \frac{c_m}{\alpha_2\delta}\|f\|_{v,s}\right)^{-1} \left[\frac{c_m}{\alpha_2\delta}\|f\|^2_{v,s} + \|\{g, H_\phi\}\|_{v,s} + e^{-\bar K\sigma}\|f\|_{v,s}\right] \ . \tag{A.5}$$

Furthermore, the following uniform bounds hold:

$$\max\left\{\alpha_1\sigma|I_1 - I_1'|, \ \alpha_2\sigma|I_2 - I_2'|, \ \alpha_2\rho|\varphi - \varphi'|, \ \alpha_2\rho_q|p - p'|, \ \alpha_2\rho_p|q - q'|\right\} \le \|f\|_{v,s} \ . \tag{A.6}$$

Proof Assumptions (A.1), (A.2) allow to apply the iterative lemma [4, Lemma A.5], with $n = \ell$, $D = A$, $E = B$, $F = B'$, $K = \bar K$, $\alpha = \alpha_2$, so as to find an analytic transformation $\Phi := \phi$ verifying (A.3)÷(A.5) and the bounds on $|I_2 - I_2'|$, $|p - p'|$, $|q - q'|$ into (A.6). In order to prove the bound on $|I_1 - I_1'|$, we recall that such transformation is obtained[26] as the time–one map associated to the Hamiltonian flow of

$$H_\phi(u,\varphi) = \sum_{|k|\le\bar K, \ k\in\Lambda} \frac{f_k(u)}{ik\cdot\omega(I)} e^{ik\cdot\varphi} \ . \tag{A.7}$$

Then, one can split H_ϕ as

$$H_\phi(u,\varphi) = H_\phi^{(1)}(u,\varphi) + H_\phi^{(2)}(u,\varphi_2) \ , \quad \text{where} \quad \varphi = (\varphi_1, \varphi_2) \in \mathbb{T}^{\ell_1} \times \mathbb{T}^{\ell_2}$$

with

$$H_\phi^{(1)}(u,\varphi) := \sum_{|k|\le\bar K, \ k\in\Lambda, k_1\ne 0} \frac{f_k(u)}{ik\cdot\omega(I)} e^{ik\cdot\varphi} \ , \quad H_\phi^{(2)}(u,\varphi_2 :) = \sum_{|k_2|\le\bar K, \ (0,k_2)\in\Lambda} \frac{f_k(u)}{ik_2\cdot\omega_2(I)} e^{ik_2\cdot\varphi_2} \ .$$

Using (A.1), one finds that $H_\phi^{(1)}$, $H_\phi^{(2)}$ verify

$$\|H_\phi^{(1)}\|_{v,s} \le \frac{\|f\|_{v,s}}{\alpha_1} \ , \quad \|H_\phi^{(2)}\|_{v,s} \le \frac{\|f\|_{v,s}}{\alpha_2} \ . \tag{A.8}$$

[26] Compare [4, (A.14)].

Since $H_\phi^{(2)}$ is independent on φ_1, from the generating equations of ϕ, equation (A.8) and Cauchy estimates, the bounds for $|I_1 - I_1'|$ (A.6) follow.
Finally, when f does not depend on (p, q) one can simply take $m = 0$. ∎

We now may proceed to sketch the proof of the Averaging Theorem in § 2, i.e., Proposition 2.1.
By the same considerations of footnote 20 of [4], we can limit ourselves to the case $e^{-\bar{K}s/6} \leq 32c_m E/\alpha_2 d$. As in [4], we apply once Lemma A.1 with $\nu = \nu_0 := v/8$; $\sigma = \sigma_0 := s/6$, thus, constructing a transformation $\Phi_0 : \ \bar{W}_1 := W_{v_1,\bar{s}+s_1} \to W_{v,\bar{s}+s}$, with $v_1 = 3/4v$, $s_1 = 2/3s$ which transforms $H = h + f$ into $H_1 = h + g_0 + f_1$. Similarly to (A.19) of [4], it follows that $\|f_1\|_{\bar{W}_1} \leq \frac{E}{6}$. By (A.6), one can replace (A.20) of [4] with

$$\alpha_1 s|I_1^{(1)} - I_1| \ , \quad \alpha_2 s|I_2^{(1)} - I_2| \ , \quad \alpha_2 r_q|p^{(1)} - p| \ , \quad \alpha_2 r_p|q^{(1)} - q| \ , \quad \alpha_2 r|\varphi^{(1)} - \varphi| \leq 8E \ .$$

Next, one proceeds as in (A21)÷(A.26) of [4], with $\bar{W}_i := W_{v_i,\bar{s}+s_i}$ replacing W_i, α_2 replacing α, $E_i := \|f_i\|_{\bar{W}_i}$ replacing ϵ_i, in order to prove (2.25). Finally, (2.26) follows by the same telescopic argument as in[27] [4], except for taking into account, as done above, the double scale (A.6) of the α_i's. ∎

B Two–scale KAM theory (proof of Proposition 2.3)

The proof of Proposition 2.3 is based on the following iterative lemma. For the purpose of this proof, we replace τ_* with τ.

Lemma B.1 (Iterative Lemma) *Under the same assumptions and notations of Proposition 2.3, for any $1 \leq j \in \mathbb{N}$, there exists a domain $D_j \subset \mathbb{R}^n$, two positive numbers ρ_j, s_j and a real–analytic and symplectic transformation Φ_j on $W_j := (D_j)_{\rho_j} \times \mathbb{T}^n_{\bar{s}+s_j}$ which conjugates $\mathrm{H}_0 := \mathrm{H}$ to*

$$\mathrm{H}_j = \mathrm{H}_0 \circ \Phi_j = \mathrm{h}_j + \mathrm{f}_j$$

and such that the following holds. Letting, for $j = 0$, $D_0 := \omega_0^{-1}(\mathcal{D}_{\gamma_1,\gamma_2,\tau}) \cap D$, $s_0 := s$, $\rho_0 := \rho$, $M_0 := M$, $\hat{M}_0 := M$, $\bar{M}_0 := \bar{M}$, $L_0 := L$, $E_0 := E$, $K_0 := K$, $\hat{\rho}_0 := \hat{\rho}$, $\hat{E}_0 := \hat{E}$ and, for $j \geq 1$, $s_j := s_{j-1}/12$, $\bar{M}_j := 2\bar{M}_{j-1}$, $\hat{M}_j := 2\hat{M}_{j-1}$, $L_j := 2L_{j-1}$, $\rho_j := \hat{\rho}_{j-1}/16$,

$$E_j := \frac{E_{j-1}L_{j-1}M_{j-1}^2}{{\gamma_1}^2} \ , \quad K_j := \frac{6}{s_j}\log_+\left(\frac{E_jL_jM_j^2}{{\gamma_1}^2}\right)^{-1} \tag{B.1}$$

$$\hat{\rho}_j := \min\left\{\frac{\gamma_1}{3M_jK_j^{\tau+1}}, \ \frac{\gamma_2}{3\hat{M}_jK_j^{\tau+1}}, \ \rho_j\right\} \ , \quad \hat{E}_j = \frac{E_jL_j}{\hat{\rho}_j^2} \tag{B.2}$$

then

(i) $D_j \subseteq (D_{j-1})_{\hat{\rho}_{j-1}/16}$; the frequency map $\omega_j := \partial\mathrm{h}_j$ is a diffeomorphism of $(D_j)_{\rho_j}$ such that $\omega_j(D_j) = \omega_{j-1}(D_{j-1})$; the map

$$\hat{\imath}_j = (\hat{\imath}_{j1}, \hat{\imath}_{j2}) : \ I \in D_{j-1} \to \omega_j^{-1} \circ \omega_{j-1}(I) \in D_j$$

verifies

$$\sup_{D_{j-1}} |\hat{\imath}_{j1} - \mathrm{id}\,| \leq 5n\frac{\bar{M}_1}{\bar{M}}\hat{E}_{j-1}\hat{\rho}_{j-1} \leq 5n\hat{E}_{j-1}\hat{\rho}_{j-1} \ ,$$

$$\sup_{D_{j-1}} |\hat{\imath}_{j2} - \mathrm{id}\,| \leq 5n\frac{\bar{M}_2}{\bar{M}}\hat{E}_{j-1}\hat{\rho}_{j-1} \leq 5n\hat{E}_{j-1}\hat{\rho}_{j-1} \tag{B.3}$$

$$\mathcal{L}(\hat{\imath}_j - \mathrm{id}\,) \leq 2^6 n\hat{E}_{j-1} \tag{B.4}$$

(ii) the perturbation f_j has sup–Fourier norm on W_j

$$\|\mathrm{f}_j\|_{W_j} \leq E_j \tag{B.5}$$

[27]Compare [4]: last formula before Appendix B.

(iii) the real–analytic symplectomorphism Φ_j is obtained as $\Phi_j = \Psi_1 \circ \cdots \circ \Psi_j$, where

$$\Psi_k : (I_k, \varphi_k) \in W_k \to (I_{k-1}, \varphi_{k-1}) \in W_{k-1}$$

verifies

$$\begin{aligned}
\sup_{(D_k)_{\rho_k} \times \mathbb{T}^n_{\bar{s}+s_k}} |I_{k-1,1}(I_k, \varphi_k) - I_{k,1}| &\leq \frac{3}{4} \frac{\hat{M}}{M} \hat{E}_{k-1} \hat{\rho}_{k-1} \\
\sup_{(D_k)_{\rho_k} \times \mathbb{T}^n_{\bar{s}+s_k}} |I_{k-1,2}(I_k, \varphi_k) - I_{k,2}| &\leq \frac{3}{4} \hat{E}_{k-1} \hat{\rho}_{k-1} \\
\sup_{(D_k)_{\rho_k} \times \mathbb{T}^n_{\bar{s}+s_k}} |\varphi_{k-1}(I_k, \varphi_k) - \varphi_k| &\leq \frac{3}{4} \hat{E}_{k-1} s_{k-1}
\end{aligned} \tag{B.6}$$

and the rescaled dimensionless map $\check{\Psi}_k - \mathrm{id} := \mathbb{1}_{\hat{\rho},s} \Psi_k \circ \mathbb{1}^{-1}_{\hat{\rho},s} - \mathrm{id}$, has Lipschitz constant on $(\check{D}_k)_{\rho_k/\rho} \times \check{\mathbb{T}}^n_{(\bar{s}+s_k)/s}$

$$\mathcal{L}(\check{\Psi}_k - \mathrm{id}) \leq 4n \left(\frac{24^{\tau+1}}{6} \right)^{k-1} \hat{E}_{k-1} \tag{B.7}$$

where id *denotes the identity map, $\mathbb{1}_{\mathrm{d}}$ denotes the* $\mathrm{d} \times \mathrm{d}$ *identity matrix, $\mathbb{1}_{\rho,\sigma} := (\rho^{-1}\mathbb{1}_n, \sigma^{-1}\mathbb{1}_n)$, $\check{D}_k := \hat{\rho}^{-1} D_k$, $\check{\mathbb{T}} := \mathbb{R}/(2\pi/s)\mathbb{Z}$;*

(iv) for any $j \geq 1$, $\hat{E}_j < \hat{E}^2_{j-1}$.

Remark B.1 Lemma B.1 generalizes the *inductive theorem* of [3, p.144]. In [3], the quantities E, γ_1, γ_2 are estimated as $\sim \mu\epsilon^7$, ϵ^{2+a}, $\mu\epsilon^{2+a}$, respectively[28]. Indeed, our approach allows to have $E \sim \mu\epsilon^5$ (and hence, essentially, to replace assumption (A2) with (A2′)), taking for γ_1, γ_2 the smallest possible values compatibly with convergence, namely, $\gamma_1 \sim \epsilon^{5/2}$, $\gamma_2 \sim \mu\epsilon^{5/2}$ (compare (1.20) above). Such smaller choice of γ_1, γ_2 with respect to [3] is important in order to improve the density of the invariant set as in (1.18).

Proof The proof is based on Proposition 2.1 (with $m = 0$, $\ell_1 = n_1$, $\ell_2 = n_2$, $B = B' = \emptyset$). Notice that, by assumption and the choice of D_0, the following inequalities hold, for $j - 1 = 0$

$$\hat{c}\hat{E}_{j-1} < 1 \tag{B.8}$$

$$L_{j-1} \geq \max \left\{ \bar{M}_{j-1} \ , M^{-1}_{j-1}, \hat{M}^{-1}_{j-1} \right\} \tag{B.9}$$

$$\omega(D_{j-1}) \subset \mathcal{D}_{\gamma_1,\gamma_2,\tau} \ . \tag{B.10}$$

Let us assume (inductively) that, when $j - 1 \geq 0$ (B.8), (B.9) and (B.10) hold and that,for $j - 1 \geq 1$, the Lemma holds with $\Phi_{j-1} = \Psi_1 \circ \cdots \circ \Psi_{j-1}$.

In order to describe the j^{th} step, for simplicity, we write ρ, $\hat{\rho}$, s, M, $\hat{M}$, $\bar{M}$, L, K, E, $\hat{E}$, D, H, h, f, $\omega = (\omega_1, \omega_2)$ for ρ_{j-1}, $\hat{\rho}_{j-1}$, s_{j-1},*etc.* , $\cdots$ and ρ_+, $\hat{\rho}_+$, s_+, *etc.* , $\cdots$ for ρ_j, $\hat{\rho}_j$, s_j, *etc.* , $\cdots$ (the corresponding *initial* quantities will be called, as in the statement, ρ_0, $\hat{\rho}_0$, s_0, *etc.* , $\cdots$).

By (B.10) and the choice of $\hat{\rho}$ (equation (B.2)), when $0 < |k| \leq K$ and $I \in D_{\hat{\rho}}$ the following non resonance inequalities hold (which are checked as in (3.39) above)

$$|\omega_1(I) \cdot k_1 + \omega_2(I) \cdot k_2| \geq \begin{cases} \dfrac{2\gamma_1}{3K^\tau} =: \alpha_1 & \text{when} \quad k_1 \neq 0 \ ; \\[2ex] \dfrac{2\gamma_2}{3K^\tau} =: \alpha_2 & \text{when} \quad k_1 = 0 \ \& \ k_2 \neq 0 \ . \end{cases} \tag{B.11}$$

The inequality $Ks \geq 6$ is trivial by definition of K (see (B.1)) and also the smallness condition (2.23) is easily met, since in this case $d = \hat{\rho}s$ and hence

$$2^7 c_0 \frac{K}{\alpha_2 \hat{\rho}} \|f\|_{D_{\hat{\rho}} \times \mathbb{T}^n_{\bar{s}+s}} \leq 2^6 c_0 \frac{E}{\hat{M}\hat{\rho}^2} \leq \hat{c}E < 1$$

[28]With a as in (1.7) above.

having used $\alpha_2 \geq 2\hat{M}K\hat{\rho}$, $L \geq \hat{M}^{-1}$, $2^6 c_0 < \hat{c}$ and (B.8). Thus, by Proposition 2.1 (with $\Lambda = \{0\}$, $h = \mathrm{h}$, , $g \equiv 0$, $f = \mathrm{f}$, $W_{\hat{\rho},s} = D_{\hat{\rho}} \times \mathbb{T}^n_{\bar{s}+s}$), H may be conjugated to

$$\mathrm{H}_+ := \mathrm{H} \circ \Psi_+ = \mathrm{h}_+ + \mathrm{f}_+$$

where, by (2.25) and the choice of K,

$$\|f_+\|_{D_{\hat{\rho}/2} \times \mathbb{T}^n_{\bar{s}+s/6}} \leq e^{-Ks/6} E \leq \frac{ELM^2}{\gamma_1{}^2} E = E_+ \ . \tag{B.12}$$

The conjugation is realized by an analytic transformation

$$\Psi_+ : \quad (I_+, \varphi_+) \in D_{\hat{\rho}/2} \times \mathbb{T}^n_{\bar{s}+s/6} \to (I, \varphi) \in D_{\hat{\rho}} \times \mathbb{T}^n_{\bar{s}+s} \ .$$

Furthermore, in view of (2.26) (with α_1, α_2 as in (B.11)), of $\alpha_1 \geq 2LK\hat{\rho}M/\hat{M}$, of $\alpha_2 \geq 2LK\hat{\rho}$, of $Ks \geq 6$ and of $L \geq M^{-1}$, $\hat{M}^{-1}$, we have

$$\sup_{D_{\hat{\rho}/2} \times \mathbb{T}^n_{\bar{s}+s/6}} |I_1(I_+, \varphi_+) - I_{+,1}| \leq \frac{3}{4} \frac{\hat{M}}{M} \hat{E} \hat{\rho} \ . \tag{B.13}$$

Similarly,

$$\sup_{D_{\hat{\rho}/2} \times \mathbb{T}^n_{\bar{s}+s/6}} |I_2(I_+, \varphi_+) - I_{+,2}| \leq \frac{3}{4} \hat{E} \hat{\rho} \ , \qquad \sup_{D_{\hat{\rho}/2} \times \mathbb{T}^n_{\bar{s}+s/6}} |\varphi(I_+, \varphi_+) - \varphi_+| \leq \frac{3}{4} \hat{E} s \ . \tag{B.14}$$

Lemma B.2 *The new frequency map $\omega_+ := \partial \mathrm{h}_+$ is injective on $D_{\hat{\rho}/8}$ and maps $D_{\hat{\rho}/16}$ over $\omega(D)$. The map $\hat{\imath}_+ = (\hat{\imath}_{+1}, \hat{\imath}_{+2}) := \omega_+^{-1} \circ \omega|_D$ which assigns to a point $I_0 \in D$ the ω_+–preimage of $\omega(I_0)$ in $D_{\hat{\rho}/16}$ satisfies*

$$\begin{aligned} &\sup_D |\hat{\imath}_{+1} - \mathrm{id}\,| \leq 5n \frac{\bar{M}_1 E}{\hat{\rho}} \leq 5n \frac{\bar{M} E}{\hat{\rho}} \ , \quad \sup_D |\hat{\imath}_{+2} - \mathrm{id}\,| \leq 5n \frac{\bar{M}_2 E}{\hat{\rho}} \leq 5n \frac{\bar{M} E}{\hat{\rho}} \ , \\ &\mathcal{L}(\hat{\imath}_+ - \mathrm{id}\,) \leq 2^6 n \frac{\bar{M} E}{\hat{\rho}^2} \ . \end{aligned} \tag{B.15}$$

Finally, the Jacobian matrix $U_+ := \partial^2 \mathrm{h}_+$ is non singular on $D_{\hat{\rho}/8}$ and the following bounds hold

$$\begin{aligned} &M_+ := 2M \geq \sup_{I \in D_{\hat{\rho}/8}} \|U_+\| \ , \quad \hat{M}_+ := 2\hat{M} \geq \sup_{I \in D_{\hat{\rho}/8}} \|\hat{U}_+\| \ , \\ &\bar{M}_+ := 2\bar{M} \geq \sup_{I \in D_{\hat{\rho}/8}} \|U_+^{-1}\| \ , \quad \bar{M}_{i+} := 2\bar{M}_i \geq \sup_{I \in D_{\hat{\rho}/8}} \|T_{i+}\| \ , \quad i = 1, \ 2 \ . \end{aligned} \tag{B.16}$$

where $U_+^{-1} =: \begin{pmatrix} T_{+1} \\ T_{+2} \end{pmatrix}$.

Postponing for the moment the proof of this Lemma, we let $\rho_+ := \hat{\rho}/16$, $s_+ := s/12$, and $D_+ := \hat{\imath}_+(D)$. By Lemma B.2, D_+ is a subset of $D_{\hat{\rho}/16}$ and hence

$$(D_+)_{\rho_+} \subset D_{\hat{\rho}/8} \ . \tag{B.17}$$

At this point, (B.5) follows from (B.12) and (B.6) from (B.13), (B.14). We are now ready to prove that $\hat{E}_+ = \frac{E_+ L_+}{\rho_+^2} \leq \hat{E}^2$. Since

$$s_+ = \frac{s}{12} \quad \text{and} \quad x_+ := \left(\frac{E_+ L_+ M_+^2}{\gamma_1{}^2} \right)^{-1} = \frac{x^2}{8} \quad \text{where} \quad x := \left(\frac{ELM^2}{\gamma_1{}^2} \right)^{-1} \tag{B.18}$$

we have

$$K_+ = \frac{6}{s_+} \log x_+ = 12 \left(\frac{12}{s} \log x - \frac{3 \log 2}{s} \right) = 24K - \frac{36 \log 2}{s} < 24K \ .$$

Thus,

$$\hat{\rho}_+ = \min \left\{ \frac{\gamma_1}{3M_+ K_+^{\tau+1}} \ , \quad \frac{\gamma_2}{3\hat{M}_+ K_+^{\tau+1}} \ , \quad \rho_+ = \frac{\hat{\rho}}{16} \right\} \geq \frac{\hat{\rho}}{2(24)^{\tau+1}} \tag{B.19}$$

and

$$\hat{E}_+ = \frac{E_+ L_+}{\hat{\rho}_+^2} \le \frac{E^2 L M^2}{{\gamma_1}^2} \frac{2L}{\hat{\rho}^2} 4(24)^{2(\tau+1)} = 8(24)^{2(\tau+1)} \frac{E\,LM^2}{{\gamma_1}^2} \hat{E}$$

Now, using, in the last inequality, the bound

$$\frac{E\,LM^2}{{\gamma_1}^2} = \frac{1}{9K^{2(\tau+1)}} \frac{EL}{[\gamma_1/(3MK^{\tau+1})]^2} \le \frac{1}{9\cdot(6/s)^{2(\tau+1)}} \frac{EL}{\hat{\rho}^2} = \frac{1}{9}\left(\frac{s}{6}\right)^{2(\tau+1)} \hat{E}$$

(since $K \ge 6/s$) we find

$$\hat{E}_+ \le \frac{8}{9}(4s)^{\tau+1}\hat{E}^2 < \hat{E}^2 \tag{B.20}$$

(having used $s \le 1/4$).

The estimate in (B.7) is a consequence (B.13), (B.14), (B.17), (B.18), (B.19) and Cauchy estimates:

$$\begin{aligned} \mathcal{L}(\check{\Psi}_j - \mathrm{id}\,) &\le 2n \sup_{(D_j)\rho_j \times \mathbb{T}^n_{\bar{s}+s_j}} \|D(\check{\Psi} - \mathrm{id}\,)\|_\infty \\ &\le 2n \frac{\frac{3}{4}\hat{E}_{j-1}\max\{\hat{\rho}_{j-1}/\rho_0, s/s_0\}}{\min\{3\hat{\rho}_{j-1}/(8\hat{\rho}_0), s_{j-1}/(12s_0)\}} \\ &\le 2n \frac{3/4(1/12)^{j-1}}{3/8\left(\frac{1}{2(24)^{\tau+1}}\right)^{j-1}} \hat{E}_{j-1} = 4n\left(\frac{(24)^{\tau+1}}{6}\right)^{j-1} \hat{E}_{j-1}\ . \end{aligned}$$

Equations (B.3), (B.4) follow from (B.15), (B.16) (and the inequality $L \ge \bar{M}$) at once. Finally, by the bounds (B.16), we see that $L_+ = 2L$ is an upper bound for $\bar{M}_+$, M_+^{-1}, $\hat{M}_+^{-1}$; (B.20) easily implies $\hat{c}E_+ < 1$ and $\omega_+(D_+) = \omega(D) \subset \mathcal{D}_{\gamma_1,\gamma_2,\tau}$ by definition of D_+. Take then $\Phi_j := \Phi_{j-1} \circ \Psi_j$, where $\Psi_j = \Psi_+$, $\Phi_0 = \mathrm{id}$. Having also checked inequalities (B.8), (B.9), (B.10) after the $j^{\rm th}$ step, Lemma B.1 is proved. ∎

Proof of Lemma B.2 Since $\mathrm{h}_+ = \mathrm{h} + \mathrm{g}$ and, by (2.25),

$$\sup_{D_{\hat{\rho}/2}} |\mathrm{g}| \le \sup_{D_{\hat{\rho}/2}} |\mathrm{g} - \bar{\mathrm{f}}| + \sup_{D_{\hat{\rho}/2}} |\bar{\mathrm{f}}| \le \frac{5}{4}E\ , \tag{B.21}$$

(where $\bar{\mathrm{f}}$ denotes the average of f), by Cauchy inequality,

$$\sup_{D_{\hat{\rho}/4}} \|\partial^2\,\mathrm{g}\| = \sup_{D_{\hat{\rho}/4}} \|\partial^2\,\mathrm{g}\|_\infty \le \frac{\sup_{D_{3\hat{\rho}/8}} |\partial \mathrm{g}|_\infty}{\hat{\rho}/8} \le \frac{E}{(\hat{\rho}/8)^2} \le 2^7 \frac{E}{\hat{\rho}^2}\ ; \tag{B.22}$$

hence,

$$\begin{aligned} \sup_{D_{\hat{\rho}/4}} \|U_+\| &= \sup_{D_{\hat{\rho}/4}} \|\partial^2 \mathrm{h}_+\| = \sup_{D_{\hat{\rho}/4}} \|\partial^2 \mathrm{h} + \partial^2 \mathrm{g}\| \\ &\le \sup_{D_{\hat{\rho}/4}} \|\partial^2 \mathrm{h}\| + \sup_{D_{\hat{\rho}/4}} \|\partial^2 \mathrm{g}\| \le M + 2^7\frac{E}{\hat{\rho}^2} \le 2M\ . \end{aligned}$$

The proof of $\sup_{D_{\hat{\rho}/4}} \|\hat{U}_+\| \le 2\hat{M}$ is similar. Using

$$\sup_{D_{\hat{\rho}/4}} \|\partial^2\,\mathrm{g}(\partial^2 \mathrm{h})^{-1}\| \le \sup_{D_{\hat{\rho}/4}} \|\partial^2 \mathrm{g}\| \sup_{D_{\hat{\rho}/4}} \|U^{-1}\| \le 2^7 \frac{E\bar{M}}{\hat{\rho}^2} \le \frac{1}{2}\ , \tag{B.23}$$

and this implies $\|(\mathbb{1} + \partial^2\,\mathrm{g}(\partial^2 \mathrm{h})^{-1})^{-1}\| \le 2$, so,

$$\bar{M}_{+1} \le 2\,\bar{M}_1\ , \quad \bar{M}_{+2} \le 2\,\bar{M}_2\ , \quad \bar{M}_+ \le 2\,\bar{M}\ . \tag{B.24}$$

Injectivity of $\omega_+ = \omega + \partial \mathrm{g}$ on $D_{\hat\rho/8}$ follows from the non singularity of $\partial^2 \mathrm{h}_+$ over $D_{\hat\rho/4}$ and the observation that two points $I_+,\ I'_+ \in D_{\hat\rho/8}$ with the same image would be closer (in 1–norm) than $\hat\rho/8$:

$$\begin{aligned} |I_+ - I'_+| &= |I_+ - I'_+|_1 = |\omega^{-1}(\omega(I_+)) - \omega^{-1}(\omega(I'_+))|_1 \\ &\le \sup_{D_{\hat\rho/8}} \|\partial(\omega^{-1})\| \sup_{D_{\hat\rho/8}} |\omega(I'_+) - \omega(I_+)|_1 \\ &= \sup_{D_\rho} \|\partial(\omega^{-1})\| \sup_{D_{\hat\rho/8}} |\partial \mathrm{g}(I'_+) - \partial \mathrm{g}(I_+)|_1 \\ &\le \bar M \cdot n \frac{2\cdot 5/4E}{3/8\hat\rho} \le 2^3 n \frac{E\bar M}{\hat\rho} \le \frac{\hat\rho}{8} \ . \end{aligned}$$

From the inclusion $\omega_+(D_{\hat\rho/16}) \supset \omega_+(D)_{\hat\rho/(32\bar M)}$ (the latter set is defined with respect to 1–norm) and the uniform bound

$$\sup_D |\omega - \omega_+|_1 = \sup_D |\partial \mathrm{g}|_1 \le \frac{n5/4E}{\hat\rho/2} \le \hat\rho/(32\bar M) \ , \tag{B.25}$$

it follows that $\omega_+(D_{\hat\rho/16}) \supset \omega(D)$. The first bound in (B.15) follows from (B.25):

$$\begin{aligned} \sup_D |\hat\imath_{+1}(I) - I_1| &= \sup_D |\hat\imath_{+1}(I) - I_1|_1 \\ &= \sup_D |\omega_+^{-1} \circ \omega_1(I) - I_1|_1 \\ &= \sup_D |\omega_+^{-1} \circ \omega_1(I) - \omega_{+1}^{-1} \circ \omega_{+1}(I)|_1 \\ &\le \sup_D \bar M_{+1} |\omega_{+1}(I) - \omega_1(I)|_1 \\ &= 2\bar M_1 \sup_D |\partial \mathrm{g}|_1 \\ &\le 5n \frac{\bar M_1 E}{\hat\rho} \ . \end{aligned}$$

The bound for $\sup_D |\hat\imath_{+2}(I) - I_2|$ in (B.15) is similar. Finally, from

$$\sup_{D_{\hat\rho/8}} |\hat\imath_+(I) - I|_\infty \le 2\bar M \sup_{D_{\hat\rho/8}} |\partial \mathrm{g}|_\infty \le 2\bar M \frac{\sup_{D_{\hat\rho/2}} |\mathrm{g}|}{3\frac{\hat\rho}{8}} \le 2^3 \frac{\bar M E}{\hat\rho}$$

there follows

$$\mathcal{L}(\hat\imath_+ - \mathrm{id}\,) \le n \inf_{0<r<\frac{\hat\rho}{8}} \sup_{D_r} \|D(\hat\imath_+ - \mathrm{id}\,)\|_\infty \le n \inf_{0<r<\frac{\hat\rho}{8}} \frac{\sup_{D_{\frac{\hat\rho}{8}}} |\hat\imath_+(I) - I|_\infty}{\frac{\hat\rho}{8} - r} \le 2^6 n \frac{\bar M E}{\hat\rho^2}$$

($D(\hat\imath_+ - \mathrm{id}\,)$ denoting the matrix of the derivatives of $\hat\imath_+ - \mathrm{id}\,$). ∎

We are now ready for the **Proof of Proposition 2.3**

Step 1 *Construction of the "limit actions"*
Define, on $D_0 = \omega^{-1}(\mathcal{D}_{\gamma_1,\gamma_2,\tau}) \cap D$,

$$\check\imath_0 := \mathrm{id} \ , \quad \check\imath_j := \hat\imath_j \circ \hat\imath_{j-1} \circ \cdots \circ \hat\imath_1 \quad j \ge 1 \ . \tag{B.26}$$

The definition is well posed because (inductively) $\check\imath_{j-1}(D_0) \in \omega_{j-1}^{-1}(\mathcal{D}_{\gamma_1,\gamma_2,\tau})$. Usual telescopic arguments imply the uniform convergence of $\check\imath_j$ to $\check\imath = (\check\imath_1, \check\imath_2) := \lim_j \check\imath_j$ on D_0 and (compare (B.3)),

$$\sup_{D_0} |\check\imath_1 - \mathrm{id}\,| \le 10n \frac{\bar M_1}{\bar M} \hat\rho_0 \hat E_0 \ , \quad \sup_{D_0} |\check\imath_2 - \mathrm{id}\,| \le 10n \frac{\bar M_2}{\bar M} \hat\rho_0 \hat E_0 \tag{B.27}$$

and that

$$D_* := \check\imath(D_0) \subset (D_j)_{\rho_j} \quad \text{for all} \quad j \tag{B.28}$$

since

$$\sup_{D_0} |\breve{\imath}_i - \breve{\imath}| \le 5n\hat\rho_i \sum_{k=i}^{+\infty} \hat E_k = 5n\hat\rho_i \sum_{k=i}^{+\infty} \hat E_0^{2^k} \le 10n\hat\rho_i \hat E_0^{2^i} < \hat\rho_i \ . \tag{B.29}$$

In particular, (B.29) with $i = 0$ implies

$$D_* \subset (D_0)_{10n\hat\rho_0 E_0} \subset D_{10n\hat\rho_0 E_0} \ . \tag{B.30}$$

By (B.4), letting $\mu_j := 2^6 n\hat E_j$ one finds[29]

$$\mathcal{L}(\breve{\imath}_{j+1} - \mathrm{id}) \le \prod_{l=1}^{j}(1+\mu_l) - 1 \le \prod_{l=1}^{+\infty}(1+\mu_l) - 1 \ . \tag{B.31}$$

Here the infinite productory $\prod_{l=1}^{+\infty}(1+\mu_l)$ converges, being bounded by

$$\begin{aligned}\prod_{l=0}^{+\infty}(1+\mu_l) &= \exp\left[\sum_l \log(1+\mu_l)\right] \le \exp\left[\sum_l \mu_l\right]\\ &\le \exp\left[2^6 n\hat E \sum_l (\hat E)^l\right]\\ &\le \exp\left[2^7 n\hat E\right] \le 1 + 2^8 n\hat E\end{aligned} \tag{B.32}$$

(having used the elementary estimate $e^x \le 1+2x$ for $0 \le x \le 1$). It follows from (B.31), (B.32)

$$\mathcal{L}(\breve{\imath} - \mathrm{id}\,) \le \limsup_j \mathcal{L}(\breve{\imath}_{j+1} - \mathrm{id}) \le 2^8 n\hat E \ . \tag{B.33}$$

So, $\breve{\imath}$ is bi–Lipschitz, hence, injective on D_0, with lower and upper[30] Lipschitz constants

$$\mathcal{L}_-(\breve{\imath}) \ge (1 - 2^8 n\hat E) \ , \quad \mathcal{L}_+(\breve{\imath}) \le (1 + 2^8 n\hat E) \ .$$

Step 2 *Construction of* ϕ_ω

Put, on $W_j = (D_j)_{\rho_j} \times \mathbb{T}^n_{\bar s + s_j}$,

$$\Phi_j := \Psi_1 \circ \cdots \circ \Psi_j$$

and let $W_* := D_* \times \mathbb{T}^n$, so, by (B.28), $W_* \subset \cap_j W_j$. The sequence $\{\Phi_j\}$ converges uniformly on W_*, as it is easily seen using (B.6) and usual telescopic arguments. Let us call Φ this limit and define

$$\phi_\omega(\vartheta) = \Big((v_1(\vartheta;\omega), v_2(\vartheta;\omega)), \vartheta + u(\vartheta;\omega) \Big) := \Phi(\breve{\imath}(\omega_0^{-1}(\omega)), \vartheta)$$

[29] Write $i_{j+1} := \breve{\imath}_{j+1} - \mathrm{id}\ = (\hat\imath_{j+1} - \mathrm{id}\,) \circ (\,\mathrm{id}\, + i_j) + i_j$, so that

$$\begin{aligned}\mathcal{L}(i_{j+1}) &\le \mathcal{L}(\hat\imath_{j+1} - \mathrm{id}\,)\Big(1 + \mathcal{L}(i_j)\Big) + \mathcal{L}(i_j)\\ &= \mathcal{L}(i_j)\Big(\mathcal{L}(\hat\imath_{j+1} - \mathrm{id}\,) + 1\Big) + \mathcal{L}(\hat\imath_{j+1} - \mathrm{id}\,)\end{aligned}$$

Iterating the above formula, we find

$$\begin{aligned}\mathcal{L}(i_{j+1}) &\le \mathcal{L}(\hat\imath_{j+1} - \mathrm{id}\,) + (1 + \mathcal{L}(\hat\imath_{j+1} - \mathrm{id}\,))\mathcal{L}(\hat\imath_j - \mathrm{id}\,) + \cdots\\ &+ (1 + \mathcal{L}(\hat\imath_{j+1} - \mathrm{id}\,))\cdots(1 + \mathcal{L}(\hat\imath_2 - \mathrm{id}\,))\mathcal{L}(\hat\imath_1 - \mathrm{id}\,)\\ &= \prod_{k=1}^{j+1}(1 + \mathcal{L}(\hat\imath_k - \mathrm{id}\,)) - 1 \ .\end{aligned}$$

[30] We recall that, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is bi–Lipschitz, the lower, upper Lipschitz constant for f are defined as

$$\mathcal{L}_-(f) := \inf_{x \ne y} \frac{|f(x) - f(y)|}{|x-y|} \ , \quad \mathcal{L}_+(f) := \sup_{x \ne y} \frac{|f(x) - f(y)|}{|x-y|} \ .$$

for $\omega \in \mathcal{D}_{\gamma_1,\gamma_2,\tau} \cap \omega_0(D)$. Since (B.28) imply, on D_*[31],

$$|\Pi_{I_1}\Phi - \mathrm{id}\,|_1 = \lim_j \sup_{W_j} |\Pi_{I_1}\Phi - \mathrm{id}\,|_1 \le \lim_j \sum_{1\le k\le j} \sup_{W_k} |\Pi_{I_1}\Psi - \mathrm{id}\,|_1 \le \frac{3}{4}\frac{\hat M_0}{M_0}\hat\rho_0 \sum_k \hat E_k \le 2\frac{\hat M_0}{M_0}\hat E_0\hat\rho_0\ , \qquad \text{(B.34)}$$

namely,

$$\sup_{W_*} |\Pi_{I_1}\Phi - \mathrm{id}\,|_1 \le 2\frac{\hat M_0}{M_0}\hat E_0\hat\rho_0 \qquad \text{(B.35)}$$

and similarly,

$$\sup_{W_*} |\Pi_{I_2}\Phi - \mathrm{id}\,|_1 \le 2\hat E_0\hat\rho_0\ , \quad \sup_{W_*} |\Pi_{\varphi}\Phi - \mathrm{id}\,|_\infty \le 2\hat E_0 s_0\ , \qquad \text{(B.36)}$$

then, in view of (B.27), (B.35), (B.36), the definition of W_* and the triangular inequality, we have (2.35). Equations (B.30), (B.35), (B.36) also imply

$$\mathrm{T}_\omega = \phi_\omega(\mathbb{T}^n) \subset (D_*)_{2\hat E_0\hat\rho_0} \times \mathbb{T}^n \subset D_r \times \mathbb{T}^n \quad \text{where} \quad r = 20n\hat E_0\hat\rho_0\ .$$

Finally, with similar arguments as in Step 1, by (B.7), the rescaled map defined on $\check D_* \times \check{\mathbb{T}}^n_{\bar s/s_0}$, as $\check\Phi = \mathbb{1}_{\hat\rho_0,s_0}\Phi \circ \mathbb{1}_{\hat\rho_0,s_0} = \lim_j \check\Psi_1 \circ \cdots \circ \check\Psi_j$ has Lipschitz constant

$$\mathcal{L}(\check\Phi - \mathrm{id}\,) \le \lim_j \sup \mathcal{L}(\check\Psi_j - \mathrm{id}\,) \le \prod_0^{+\infty}(1+\varsigma_k) - 1 \le e^{\sum_k \varsigma_k} - 1 \le 2\sum_k \varsigma_k \le 2^4 n\hat E_0\ , \qquad \text{(B.37)}$$

where $\varsigma_k = 4n\left(\frac{24^{\tau+1}}{6}\right)^k \hat E_k$, provided $\dfrac{24^{\tau+1}}{6}\hat E < 1$. In particular, $\check\Phi$, hence, Φ, and, finally, the map $(\vartheta,\omega) \to \phi_\omega(\vartheta)$ are bi–Lipschitz, hence, injective.

Step 3 *For any $\omega \in \mathcal{D}_{\gamma_1,\gamma_2,\tau} \cap \omega_0(D)$, $\mathrm{T}_\omega := \phi_\omega(\mathbb{T}^n)$ is a Lagrangian H–invariant torus with frequency ω.*

We consider the action $I_* \in D_*$ as independent variable, so, if $\phi_t(I_0,\varphi_0)$ denotes the H–flow starting at (I_0,φ_0), then, we have to prove

$$\phi_t\Big(\Phi(I_*,\vartheta)\Big) = \Phi(I_*,\vartheta+\omega_*(I_*)\,t)\ , \quad \text{where} \quad \omega_*(I_*) := \omega_0(\imath^{-1}(I_*))\ .$$

We can write

$$\begin{aligned}
|\phi_t\Big(\Phi(I_*,\vartheta)\Big) - \Phi(I_*,\vartheta+\omega_*(I_*)\,t)| \quad &:= \quad |\phi_t\Big(\Phi(I_*,\vartheta)\Big) - \Phi(I_*,\vartheta+\omega_*(I_*)\,t)|_\infty \\
&\le \quad |\phi_t\Big(\Phi(I_*,\vartheta)\Big) - \phi_t\Big(\Phi_j(I_*,\vartheta)\Big)| \\
&+ \quad |\phi_t\Big(\Phi_j(I_*,\vartheta)\Big) - \Phi_j(I_*,\vartheta+\omega_*(I_*)t)| \\
&+ \quad |\Phi_j(I_*,\vartheta+\omega_*(I_*)t) - \Phi(I_*,\vartheta+\omega_*(I_*)t)|
\end{aligned}$$

Due to the uniform convergence of Φ_j to Φ on W_* and continuous dependence on the initial data, both $|\Phi_j(I_*,\vartheta+\omega_*(I_*)t) - \Phi(I_*,\vartheta+\omega_*(I_*)t)|$ and $|\phi_t\Big(\Phi(I_*,\vartheta)\Big) - \phi_t\Big(\Phi_j(I_*,\vartheta)\Big)|$ go to 0 as $j\to\infty$, so, to conclude, we have to prove that $|\phi_t\Big(\Phi_j(I_*,\vartheta)\Big) - \Phi_j(I_*,\vartheta+\omega_*(I_*)t)|$ also goes to 0. Due to the canonicity of Φ_j on W_j, we have

$$\phi_t\Big(\Phi_j(I_*,\vartheta)\Big) = \Phi_j\Big(\phi_t^j(I_*,\vartheta)\Big)$$

with $\phi_t^j(I_*,\vartheta)$ the $\mathrm{H}_j = \mathrm{H}\circ\Phi_j = \mathrm{h}_j + \mathrm{f}_j$–flow from (I_*,ϑ). Using (B.37), we are reduced to prove that

$$\phi_t^j(I_*,\vartheta) - (I_*,\vartheta+\omega_*(I_*)t) \to 0$$

[31] Π_z denotes the projection on the z–variables, where z is I_1, I_2 or φ.

as $j \to +\infty$. But this is an easy consequence of the generating equations of $\Big(I_j(t), \varphi_j(t)\Big) := \phi_t^j(I_*, \vartheta)$:

$$\begin{cases} I_j(t) = I_* - \int_0^t \partial_\varphi \mathrm{f}_j(I_j(\tau), \varphi_j(\tau)) d\tau \\ \varphi_j(t) = \vartheta + \int_0^t \partial_I (\mathrm{h}_j(I_j(\tau)) + \mathrm{f}_j(I_j(\tau), \varphi_j(\tau)) d\tau \end{cases}$$

making use of Cauchy estimates and the bounds of Lemma B.1.

Step 4 *Measure Estimates (proof of (2.34))*
In order to prove (2.34), we decompose

$$\mathrm{meas}\Big(\mathrm{Re}\,(D_r) \times \mathbb{T}^n \setminus \mathrm{K}\Big) \le \mathrm{meas}\Big((\mathrm{Re}\,(D_r) \setminus D) \times \mathbb{T}^n\Big) + \mathrm{meas}\Big(D \times \mathbb{T}^n \setminus \mathrm{K}\Big)\ . \tag{B.38}$$

The measure of the second set in (B.38) can be estimated as follows. Since the map $\check{\imath} - \mathrm{id}\ :\ D_0 \to D_*$ is Lipschitz on D_0 with Lipschitz constant verifying (B.33), by a Theorem of Kirszbraun[32], it can be extended to a Lipschitz function having the same Lipschitz constant on the domain $(D_0)_{\rho_1}$, where $\rho_1 = 10n\hat{E}\hat{\rho}/(1-2^8 n\hat{E})$. We denote $\check{\imath}_{\mathrm{e}} - \mathrm{id}$ such extension. Then, $\check{\imath}_{\mathrm{e}}$ is a bi–Lipschitz extension (hence, injective) of $\check{\imath}$ on $(D_0)_{\rho_1}$, with lower Lipschitz constant $\mathcal{L}_-(\check{\imath}_{\mathrm{e}}) \ge 1 - 2^8 n\hat{E}$. This implies that $\check{\imath}_{\mathrm{e}}$ sends a ball with radius ρ_1 centered at $I_0 \in D_0$ over a ball with radius $(1-2^8 n\hat{E})\rho_1 = 10n\hat{\rho}\hat{E} \ge \sup_{D_0} |\check{\imath} - \mathrm{id}\,|$ centered at $\check{\imath}(I_0)$, so as to conclude

$$\check{\imath}_{\mathrm{e}}\Big((D_0)_{\rho_1}\Big) \supset D_0\ .$$

Then, since $\check{\imath}_{\mathrm{e}}$ is injective[33],

$$\begin{aligned} \mathrm{meas}\Big(D_0 \setminus D_*\Big) &= \mathrm{meas}\Big(D_0 \setminus \check{\imath}(D_0)\Big) \\ &= \mathrm{meas}\Big(D_0 \setminus \check{\imath}_{\mathrm{e}}(D_0)\Big) \\ &\le \mathrm{meas}\Big(\check{\imath}_{\mathrm{e}}(\,\mathrm{Re}\,(D_0)_{\rho_1}) \setminus \check{\imath}_{\mathrm{e}}(D_0)\Big) \\ &\le \mathrm{meas}\Big(\check{\imath}_{\mathrm{e}}(\,\mathrm{Re}\,(D_0)_{\rho_1} \setminus D_0)\Big) \\ &\le \mathcal{L}(\check{\imath})^n\,\mathrm{meas}\Big(\mathrm{Re}\,(D_0)_{\rho_1} \setminus D_0\Big) \\ &\le \mathcal{L}(\check{\imath})^n\,\mathrm{meas}\Big(\mathrm{Re}\,(D_{\rho_1}) \setminus D_0\Big) \\ &\le \mathcal{L}(\check{\imath})^n\Big(\mathrm{meas}\Big(\mathrm{Re}\,(D_{\rho_1}) \setminus D\Big) + \mathrm{meas}\Big(D \setminus D_0\Big)\Big)\ . \end{aligned}$$

Then,

$$\begin{aligned} \mathrm{meas}\Big(D \setminus D_*\Big) &\le \mathrm{meas}\Big(D \setminus D_0\Big) + \mathrm{meas}\Big(D_0 \setminus D_*\Big) \\ &\le \Big(1 + \mathcal{L}(\check{\imath})^n\Big)\,\mathrm{meas}\Big(D \setminus D_0\Big) + \mathcal{L}(\check{\imath})^n\,\mathrm{meas}\Big(D_{\rho_1} \setminus D\Big) \\ &\le \sqrt{c_n}\Big(\mathrm{meas}\Big(D \setminus D_0\Big) + \mathrm{meas}\Big(D_{\rho_1} \setminus D\Big)\Big) \end{aligned} \tag{B.39}$$

[32]**Theorem** *Assume $A \subset \mathbb{R}^n$, and let $f : A \to \mathbb{R}^m$ be Lipschitz. Then, there exists a Lipschitz function $\bar{f} :\ \mathbb{R}^n \to \mathbb{R}^m$, such that $\bar{f} = f$ on A with $\mathcal{L}(\bar{f}) = \mathcal{L}(f)$.* For a proof, see [8].
[33]Recall also the following classical fact: *If $f : A \to \mathbb{R}^n$ is a Lipschitz function and $A \subset \mathbb{R}^n$ is Lebesgue–measurable, then* $\mathrm{meas}\,(f(A)) \le \mathcal{L}(f)^n\ \mathrm{meas}\,(A)$. For a proof, see [8].

where c_n is as in the statement of Proposition 2.3. In turn, using (B.34), (B.36), (B.37), we can repeat the above argument, with

$$\check{\Phi}\ ,\quad \check{\mathrm{K}}=\check{\Phi}(\check{D}_*\times\check{\mathbb{T}}^n)\ ,\quad \check{D}_*\times\check{\mathbb{T}}^n\ ,\quad 2n\ ,\quad \mathcal{L}(\check{\Phi})\ ,\quad \check{\rho}_2=2\hat{E}/(1-2^7n\hat{E})$$

replacing the previous

$$\check{\imath}\ ,\quad D_*\ ,\quad D_0\ ,\quad n\ ,\quad \mathcal{L}(\check{\imath})\ ,\quad \rho_1$$

respectively, and we find

$$\mathrm{meas}\left(\check{D}\times\check{\mathbb{T}}^n\setminus\check{\mathrm{K}}\right)\quad\le\quad\sqrt{c_n}\ \left(\mathrm{meas}\,(\check{D}\setminus\check{D}_*\times\check{\mathbb{T}}^n)+\ \mathrm{meas}\,(\,\mathrm{Re}\,(\check{D}_{\check{\rho}_2})\setminus\check{D}\times\check{\mathbb{T}}^n)\right)\ .$$

The result then follows by rescaling the variables, in view of (B.38), (B.39) and noticing that $\rho_1,\ \rho_2:=\hat{\rho}\check{\rho}_2\le r:=20n\hat{E}\hat{\rho}$. ∎

C Measure estimates (proof of Lemma 3.1)

Let

$$R_1>\max_{\bar{D}\times\overline{B}^{n_2}_{\hat{r}}}|\omega_1|\ ,\quad a>\max_{\bar{D}}\|\beta\|\ ,\quad R_2:=a\hat{r}\ ,\quad U(I_1):=B^{n_1}_{R_1}(0)\times B^{n_2}_{R_2}(\omega_{02}(I_1))\ ,\quad \mathcal{U}:=\bigcup_{I_1\in\bar{D}}U(I_1)\ . \tag{B.1}$$

By the choice of R_1 and since

$$|\omega_2(I)-\omega_{02}(I_1)|=|\beta(I_1)I_2|\le\|\beta(I_1)\||I_2|<a\hat{r}=R_2\qquad\text{for any}\quad I=(I_1,I_2)\in\bar{D}\times\overline{B}^{n_2}_{\hat{r}}\ ,$$

then, the set $\mathcal{U}$ in (B.1) is an open covering of the compact set $\Omega=\omega(\bar{D}\times\overline{B}^{n_2}_{\hat{r}})$. Hence, for a suitable positive integer p and $I_1^{(1)},\cdots,I_1^{(p)}\in\bar{D}$, we have that the sets $U_i:=U(I_1^{(i)})$ realize a covering of Ω, namely, $\bigcup_{1\le i\le p}U_i\supset\Omega$.

Now, we want to estimate the measure of the resonant set

$$\begin{aligned}&\mathcal{R}_{\gamma_1,\gamma_2,\tau}\\ &=\bigcup_{\substack{(k_1,k_2)\in\mathbb{Z}^{n_1}\times\mathbb{Z}^{n_2}\\ k_1\neq0}}\left\{I\in\bar{D}\times\overline{B}^{n_2}_{\hat{r}}\ :\ |\omega(I)\cdot k_1|\le\frac{\gamma_1}{|k_1|^\tau}\right\}\bigcup_{0\neq k_2\in\mathbb{Z}^{n_2}}\left\{I\in\bar{D}\times\overline{B}^{n_2}_{\hat{r}}\ :\ |\omega_2(I)\cdot k_2|\le\frac{\gamma_2}{|k_2|^\tau}\right\}.\end{aligned}\tag{B.2}$$

The measure of the first set in (B.2) is bounded by

$$\begin{aligned}&\mathrm{meas}\left(\bigcup_{k\in\mathbb{Z}^n,k_1\neq0}\left\{I\in\bar{D}\times\overline{B}^{n_2}_{\hat{r}}\ :\ |\omega(I)\cdot k|\le\frac{\gamma_1}{|k|^\tau}\right\}\right)\\
&\le\max_{\bar{D}\times\overline{B}^{n_2}_{\hat{r}}}\|(\partial\omega)^{-1}\|^n\,\mathrm{meas}\left(\bigcup_{k\in\mathbb{Z}^n,k_1\neq0}\left\{x\in\Omega:\ |x\cdot k|\le\frac{\gamma_1}{|k|^\tau}\right\}\right)\\
&\le\max_{\bar{D}\times\overline{B}^{n_2}_{\hat{r}}}\|(\partial\omega)^{-1}\|^n\,\mathrm{meas}\left(\bigcup_{k\in\mathbb{Z}^n,k_1\neq0}\ \bigcup_{i=1}^{p}\left\{x\in U_i:\ |x\cdot k|\le\frac{\gamma_1}{|k|^\tau}\right\}\right)\\
&\le\max_{\bar{D}\times\overline{B}^{n_2}_{\hat{r}}}\|(\partial\omega)^{-1}\|^n\sum_{k\in\mathbb{Z}^n,k_1\neq0}\ \sum_{1\le i\le p}\ \mathrm{meas}\left(\left\{x\in U_i:\ |x\cdot k|\le\frac{\gamma_1}{|k|^\tau}\right\}\right)\\
&=\max_{\bar{D}\times\overline{B}^{n_2}_{\hat{r}}}\|(\partial\omega)^{-1}\|^n\sum_{k\in\mathbb{Z}^n,k_1\neq0}\ \sum_{1\le i\le p}\int_{W_k^i}dx\end{aligned}\tag{B.3}$$

where

$$W_k^i:=\left\{x=(x_1,x_2)\in U_i:\ |x_1\cdot k_1+x_2\cdot k_2|\le\frac{\gamma_1}{|k|^\tau}\right\}\ .$$

Now, as $k_1 = (k_1^{(1)}, \cdots, k_1^{(n_1)}) \neq 0$, there exists $1 \leq j \leq n_1$ such that $|k_1^{(j)}| \neq 0$. Perform, then, the change of variables

$$\begin{cases} z_m = x_1^{(m)} \quad \text{for} \quad 1 \leq m \leq n_1 \ , \ m \neq j \ , \\ z_j = x_1 \cdot k_1 + x_2 \cdot k_2 \ , \\ z_m = x_2^{(m-n_1)} \quad \text{for} \quad n_1 + 1 \leq m \leq n = n_1 + n_2 \ , \end{cases}$$

where $x_i^{(j)}$ denothes the j^{th} component of x_i. Define the set $\tilde{W}_k^i$ as the set of points $z' = (z'_1, \cdots, z'_n)$ such that

$$\begin{cases} \left((z'_1, \cdots, z'_{j-1}, \frac{1}{k_j} \left[z'_j - \sum_{m \neq j} \bar{z}'_m k_1^{(m)} - \sum_{n_1+1 \leq m \leq n} z'_m k_2^{(m)} \right], z'_{j+1}, \cdots, z'_{n_1} \right) \in B_{R_1}^{n_1}(0) \\ \\ |z'_j| \leq \frac{\gamma_1}{|k|^\tau} \ , \ (z'_{n_1+1}, \cdots, z'_n) \in B_{R_2}^{n_2}(\omega_{02}(I_1^{(i)})) \ ; \end{cases}$$

define also the set $\mathcal{C}_k^i$ as the set of points $z' = (z'_1, \cdots, z'_n)$ such that

$$|z'_m| \leq R_1 \text{ for } 1 \leq m \neq j \leq n_1, \ |z'_j| \leq \frac{\gamma_1}{|k|^\tau} \ , \ |z'_m - \omega_{02}(I_1^{(i)})| \leq R_2 \text{ for } m = n_1 + 1, \cdots, n \ .$$

Then, $\mathcal{C}_k^i \supseteq \tilde{W}_k^i$ and, since $|k_j| \geq 1$,

$$\int_{W_k^i} dx = \frac{1}{|k_j|} \int_{\tilde{W}_k^i} dz \leq \frac{1}{|k_j|} \int_{\mathcal{C}_k^i} dz \leq \int_{\mathcal{C}_k^i} dz = R_1^{n_1-1} R_2^{n_2} \frac{\gamma_1}{|k|^\tau} \ .$$

Hence, inserting this expression into (B.3), we find

$$\begin{aligned} \text{meas} \left(\bigcup_{k \in \mathbb{Z}^n, k_1 \neq 0} \left\{ I : \ |\omega(I) \cdot k| \leq \frac{\gamma_1}{|k|^\tau} \right\} \right) &\leq \max_{\bar{D} \times \overline{B}_{\hat{r}}^{n_2}} \|(\partial\omega)^{-1}\|^n R_1^{n_1-1} R_2^{n_2} p \, \gamma_1 \sum_{k \in \mathbb{Z}^n, k_1 \neq 0} \frac{1}{|k|^\tau} \\ &\leq \max_{\bar{D} \times \overline{B}_{\hat{r}}^{n_2}} \|(\partial\omega)^{-1}\|^n R_1^{n_1-1} R_2^{n_2} c(n, \tau) \gamma_1 \\ &= \max_{\bar{D} \times \overline{B}_{\hat{r}}^{n_2}} \|(\partial\omega)^{-1}\|^n R_1^{n_1-1} a^{n_2} \hat{r}^{n_2} c(n, \tau) \gamma_1 \\ &= c_1 \gamma_1 \, \text{meas} \, (\bar{D} \times B_{\hat{r}}^{n_2}) \ , \end{aligned}$$

where c_1 is as in (3.58). We now estimate the measure of the second set in (B.2). By Fubini's Theorem,

$$\text{meas} \bigcup_{0 \neq k_2 \in \mathbb{Z}^{n_2}} \left\{ I : \ |\omega_2(I) \cdot k_2| \leq \frac{\gamma_2}{|k_2|^\tau} \right\} = \int_{\bar{D}} dI \int_{\bigcup_{k_2 \neq 0} \hat{W}_k(I)} dI_2 \tag{B.4}$$

where

$$\hat{W}_k(I_1) = \bigcup_{0 \neq k_2 \in \mathbb{Z}^{n_2}} \left\{ I_2 \in \overline{B}_{\hat{r}}^{n_2}(0) : \ |(\omega_{02}(I_1) + \beta(I_1) I_2) \cdot k_2| \leq \frac{\gamma_2}{|k_2|^\tau} \right\}$$

Perform, in the inner integral, the change of variable $x_2 = \omega_{02}(I_1) + \beta(I_1) I_2$ and let

$$\begin{aligned} \tilde{\mathcal{C}}_k(I_1) &:= \left\{ x_2 \in \mathbb{R}^{n_2} : \ \beta(I_1)^{-1}(x_2 - \omega_{02}(I_1)) \in \overline{B}_{\hat{r}}^{n_2}(0) \ , \quad |x_2 \cdot k_2| \leq \frac{\gamma_2}{|k_2|^\tau} \right\} \\ &\subseteq \left\{ x_2 \in \mathbb{R}^{n_2} : \ x_2 \in \overline{B}_{R_2}^{n_2}(\omega_{02}(I_1)) \ , \quad |x_2 \cdot k_2| \leq \frac{\gamma_2}{|k_2|^\tau} \right\} =: \hat{\mathcal{C}}_k(I_1) \ , \end{aligned}$$

since, if $x_2 \in \mathcal{C}_k(I_1)$, then, $x_2 \in \overline{B}_{\|\beta(I_1)\|\hat{r}}^{n_2} \subset B_{R_2}^{n_2}$. Then, proceeding as done for the first part of the proof (*i.e.* , with a suitable change of variable, for which $z'_j = x_2 \cdot k_2$ if $k_2^{(j)} \neq 0$; $z'_m = x_2^{(m)}$ for $m \neq j$), we find

$$\begin{aligned} \int_{\bigcup_{k_2 \neq 0} \hat{W}_k(I_1)} dI_2 &\leq \max_{\bar{D}} \|\beta^{-1}\|^{n_2} \int_{\bigcup_{k_2 \neq 0} \tilde{\mathcal{C}}_k(I_1)} dI_2 \\ &\leq \max_{\bar{D}} \|\beta^{-1}\|^{n_2} \int_{\bigcup_{k_2 \neq 0} \hat{\mathcal{C}}_k(I_1)} dx_2 \\ &\leq \max_{\bar{D}} \|\beta^{-1}\|^{n_2} R_2^{n_2-1} \gamma_2 \sum_{k_2 \neq 0} \frac{1}{|k|^\tau} \ . \end{aligned}$$

Hence, inserting this value into (B.4) and since

$$R_2 = a\hat{r} \quad \text{and} \quad \hat{r}^{n_2-1} = \frac{1}{\hat{r}}\text{meas}(B_{\hat{r}}^{n_2}) \ ,$$

we find

$$\begin{aligned}\text{meas}\left(\bigcup_{0\neq k_2\in\mathbb{Z}^{n_2}}\left\{I:\ |\omega_2(I)\cdot k_2|\leq\frac{\gamma_2}{|k_2|^\tau}\right\}\right)\leq & \quad \text{meas}\,(\bar{D})\max_{\bar{D}}\|\beta^{-1}\|^{n_2}R_2^{n_2-1}\gamma_2 c(n_2,\tau)\\ = & \quad c_2\frac{\gamma_2}{\hat{r}}\,\text{meas}\,(\bar{D}\times B_{\hat{r}}^{n_2})\end{aligned}$$

where c_2 is as in (3.58). ∎

References

[1] K. Abdullah and A. Albouy. On a strange resonance noticed by M. Herman. *Regul. Chaotic Dyn.*, 6(4):421–432, 2001.

[2] V. I. Arnol′d. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1989. Translated from the Russian by K. Vogtmann and A. Weinstein.

[3] V.I. Arnol'd. Small denominators and problems of stability of motion in classical and celestial mechanics. *Russian Math. Surveys*, 18(6):85–191, 1963.

[4] L. Biasco, L. Chierchia, and E. Valdinoci. Elliptic two-dimensional invariant tori for the planetary three-body problem. *Arch. Rational Mech. Anal.*, 170:91–135, 2003.

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

[6] L. Chierchia and F. Pusateri. Analytic Lagrangian tori for the planetary many-body problem. *Ergodic Theory Dynam. Systems*, 29(3):849–873, 2009.

[7] A. Deprit. Elimination of the nodes in problems of n bodies. *Celestial Mech.*, 30(2):181–195, 1983.

[8] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[9] J. Féjoz. Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman). *Ergodic Theory Dynam. Systems*, 24(5):1521–1582, 2004. Revised version (2007) at http://people.math.jussieu.fr/~fejoz/articles.html.

[10] H. Hofer, E. Zehnder. *Symplectic Invariants and Hamiltonian Dynamics.* Birkhäuser Verlag, Basel, 1994.

[11] U. Locatelli and A. Giorgilli. Invariant tori in the Sun-Jupiter-Saturn system. *Discrete Contin. Dyn. Syst. Ser. B*, 7(2):377–398 (electronic), 2007.

[12] G. Pinzari. *On the Kolmogorov set for many–body problems.* PhD thesis, Università degli Studi Roma Tre, April 2009. Available at www.mat.uniroma3.it/users/chierchia/TESI/PhD_Thesis_GPinzari.pdf.

[13] J. Pöschel. Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Math. Z.*, 213(2):187–216, 1993.

[14] P. Robutel. Stability of the planetary three-body problem. II. KAM theory and existence of quasiperiodic motions. *Celestial Mech. Dynam. Astronom.*, 62(3):219–261, 1995.

[15] H. Rüssmann. Nondegeneracy in the perturbation theory of integrable dynamical systems. In *Stochastics, algebra and analysis in classical and quantum dynamics (Marseille, 1988)*, volume 59 of *Math. Appl.*, pages 211–223. Kluwer Acad. Publ., Dordrecht, 1990.

[16] M.B. Sevryuk. The classical KAM theory at the dawn of the twenty-first century. *Mosc. Math. J.*, 3(3):1113–1144, 1201–1202, 2003. {Dedicated to Vladimir Igorevich Arnold on the occasion of his 65th birthday}.

[17] J. Williamson. On the algebraic problem concerning the normal forms of linear dynamical systems. *Amer. J. Math.*, 58(1):141–163, 1936.