FLEXIBILITY AND ANALYTIC SMOOTHING IN AVERAGING THEORY

Santiago Barbieri, Jean-Pierre Marco, Jessica Elisa Massetti

▶ **To cite this version:**

Santiago Barbieri, Jean-Pierre Marco, Jessica Elisa Massetti. FLEXIBILITY AND ANALYTIC SMOOTHING IN AVERAGING THEORY. 2021. hal-03226888v1

HAL Id: hal-03226888

https://hal.science/hal-03226888v1

Preprint submitted on 16 May 2021 (v1), last revised 5 Nov 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

FLEXIBILITY AND ANALYTIC SMOOTHING IN AVERAGING THEORY

SANTIAGO BARBIERI, JEAN-PIERRE MARCO, AND JESSICA ELISA MASSETTI

Abstract. Using a new strategy, we extend the classical Nekhoroshev's estimates to the case of Hölder regular steep near-integrable hamiltonian systems, the stability times being polynomially long in the inverse of the size of the perturbation. We prove that the stability exponents can be taken to be $(\ell-1)/(2n\boldsymbol{\alpha}_1...\boldsymbol{\alpha}_{n-2})$ for the time of stability and $1/(2n\boldsymbol{\alpha}_1...\boldsymbol{\alpha}_{n-1})$ for the radius of stability, $\ell > n+1$ being the regularity and the $\boldsymbol{\alpha}_i$'s being the indices of steepness. Our strategy consists in deriving a perturbation theory which exploits a sharp analytic smoothing theorem to approximate any Hölder function by an analytic one. In addition, an appropriate choice of the free parameters in the problem enables us to have a first grasp on the relation connecting the time and radius of stability to the threshold that the size of the perturbation must satisfy in order for the theorem to apply. Particular attention is payed to a geometric presentation of the construction of the so-called *resonant blocks*, in order to shed a definitive light on the nature of the steepness condition. We also investigate the convex setting, using a similar approach.

Contents

1. Introduction and main results

1. The main goal of this work is to introduce a unified way for proving "long time stability" of the action variables for perturbations of completely integrable Hamiltonian systems which belong to a large class of function spaces. We will limit ourselves here to Hölder perturbations of analytic systems, but our method is flexible enough to be adapted to many other settings[1].

The effective stability theory for near-integrable hamiltonian systems was initiated by the pioneering work of J.E. Littlewood [24] and reached a first main achievement in the seventies with the work of N.N. Nekhoroshev [31]; it was then developed by many authors. The usual setting is that of Hamiltonian systems of the form

$$H(I,\theta) = h(I) + f(I,\theta), \tag{1.1}$$

where $(I,\theta) \in \mathbb{R}^n \times \mathbb{T}^n$ are the action-angle variables. In Nekhoroshev's work the Hamiltonian H is analytic and h satisfies a steepness condition (see the definition below). The theory has been then developed in various

[1]Assuming that the unperturbed system is analytic is just a matter of simplification.

settings: H can be assumed to be Gevrey (which includes the analytic case) or C^k with $k \geq 2$ and integer, while h can be assumed to be convex or quasi-convex.

The norm of f, relative to the function space at hand, is denoted by ε. For systems as (1.1), the previous results assert that the action variables are confined in a ball of radius $\mathbf{R}(\varepsilon)$ centered at the initial action during a time $\mathbf{T}(\varepsilon)$, provided that ε is smaller that some threshold $\mathbf{E}$. We say that $\mathbf{R}(\varepsilon)$ is the *confinement radius*, $\mathbf{T}(\varepsilon)$ is the stability time and $\mathbf{E}$ is the *applicability threshold*. The remarkable fact is that – h being given – the results depend only on the norm of f and not on its particular form. In this setting, most papers on long time stability focus on the problem of finding the largest possible stability time $\mathbf{T}(\varepsilon)$ allowed by the geometric and regularity constraints, together with a relevant radius $\mathbf{R}(\varepsilon)$, while a precise derivation of the threshold $\mathbf{E}$ is seldom investigated.

However, several works have been dedicated to obtain physically relevant thresholds for problems of Celestial Mechanics (see e.g. [15], [21], [33], [14], [4]). In these problems the specific form of the perturbation plays a crucial role. In addition to our main goal, the method introduced in the present paper enables us to investigate in an abstract setting the various relations between $\mathbf{R}(\varepsilon)$, $\mathbf{T}(\varepsilon)$ and $\mathbf{E}$, which we see as *parameters* of the stability theory. We expect this approach to be a first step to take advantage of the specific form of the perturbation.

2. The classical results. Let us briefly describe the classical abstract results. In the 70's Nekhoroshev proved his seminal theorem [31], which asserts that for a steep real-analytic function h and for any real-analytic perturbation f with analytic extension to a complex domain $\mathcal{D}$, all solutions are stable at least over exponentially long time intervals. Namely, there exist positive *exponents* a, b and a positive threshold $\mathbf{E}$, depending only on h, such that if $|f|_{\mathcal{D}} \leq \mathbf{E}$, then any initial condition (I_0, θ_0) gives rise to a solution $\big(I(t), \theta(t)\big)$ which is defined at least for $|t| \leq \exp\big(c(1/\varepsilon)^a\big)$ and satisfies $|I(t) - I_0| \leq C\varepsilon^b$ in that range. Here $|f|_{\mathcal{D}}$ is the C^0 sup-norm on the domain $\mathcal{D}$ and c, C are positive constants which also depend only on h. With our notation, for such systems:

$$\mathbf{T}(\varepsilon) = \exp(c(1/\varepsilon)^a), \qquad \mathbf{R}(\varepsilon) = C\varepsilon^b, \tag{1.2}$$

while the expression of the threshold $\mathbf{E}$ is quite involved, see [31]. Since the constants c and C are less significant than the exponents we will get rid of them in our subsequent description.

Nekhoroshev's proof is based on the construction of a partition (a "patchwork") of the phase space into zones of approximate resonances of different multiplicities, over which one can construct adapted normal forms. The global stability result necessitates a very delicate control of the size and disposition of the elements of the patchwork in order to produce a "dynamical confinement" preventing the orbits from fast motions along distances larger than the confinement radius (see below for a discussion).

In the convex case, as noticed in [20] and [5], a shrewd use of energy conservation leads to a much simpler and "physical" way to confine the orbits. This gave rise to two distinct series of works, originating in the articles of Lochak [26] - where the simultaneous approximation method were introduced - and Pöschel [39] - where the construction of Nekhoroshev's patchwork was made much easier - both relying on the convexity or quasi-convexity of the integrable Hamiltonian.

The simplicity of these methods made it possible to prove that the Nekhoroshev Theorem in the analytic case holds with

$$\mathbf{T}(\varepsilon) = \exp(c(1/\varepsilon)^{1/2n}), \qquad \mathbf{R}(\varepsilon) = C\varepsilon^{1/2n}, \tag{1.3}$$

if h is assumed to be quasi-convex (see [26, 29, 39]). Moreover, besides the global result, one can state local results for neighborhoods of resonant surfaces. For $m \in \{1, \dots, n-1\}$, consider a sublattice $\Lambda \in \mathbb{Z}^n_K := \{k \in \mathbb{Z}^n : |k|_1 \leq K\}$ of rank m and the resonant subset $\mathcal{M}_\Lambda := \{I \in \mathbb{R}^n \mid \nabla h(I) \in \Lambda^\perp\}$. Then, for all trajectories starting at a distance of order $\varepsilon^{1/2}$ of $\mathcal{M}_\Lambda$, one gets larger stability exponents, namely $a = b = 1/(2(n-m))$. Moreover, in the resonant block D_Λ (which is obtained by eliminating from $\mathcal{M}_\Lambda$ all the intersections with other resonant subsets $\mathcal{M}'_\Lambda$, with rank $\Lambda' = m+1$, see section 5) one can even take $a = 1/(2(n-m))$, $b = 1/2$.

As alluded to above, long time stability does not require *a priori* the analyticity of the Hamiltonian at hand. For general Gevrey quasi-convex systems[2], the fast decay of the Fourier coefficients also yields exponentially long stability times. Namely, for β-Gevrey systems (where β is the Gevrey exponent) it is proved in [30] that

$$\mathbf{T}(\varepsilon) = \exp\left(c/\varepsilon^{1/(2n\beta)}\right), \qquad \mathbf{R}(\varepsilon) = C\varepsilon^{1/(2n\beta)}.$$

The proof is based on a direct construction of normal forms for Gevrey systems. This study was initiated by M. Herman for proving the optimality of the stability exponents by constructing explicit examples taking advantage of the flexibility of the Gevrey category, see below.

Soon after, finitely differentiable systems have been investigated in [7] using a *direct* implementation of Lochak's scheme in this setting, which yields the estimates

$$\mathbf{T}(\varepsilon) = c/\varepsilon^{(\ell-2)/(2n)} \qquad \mathbf{R}(\varepsilon) = C\varepsilon^{1/(2n)}$$

for quasi-convex C^ℓ systems with $\ell \geq 2$ and integer. On the other hand, the stability of C^ℓ systems, with ℓ an integer such that $\ell \geq \ell^* n + 1$ for some $\ell^* \in \mathbb{N}^*$, satisfying a property known as Diophantine-Morse condition[3], was investigated in [8], where the values

$$\mathbf{T}(\varepsilon) = c/\varepsilon^{\ell^*/[3(4(n+1))^n]} \qquad \mathbf{R}(\varepsilon) = C\varepsilon^{1/(4(n+1))^n}$$

were found.

The case $\ell = +\infty$ has been studied in [2], where the authors find that, in the case $h(I) = I^2/2$ and for fixed $b \in (0, 1/2)$, for any $M > 0$ there exists $C_M > 0$ such that

$$\mathbf{T}(\varepsilon) = \frac{C_M}{\varepsilon^M} \qquad \mathbf{R}(\varepsilon) = C_M\varepsilon^b$$

and $\mathbf{E}$ decreases with M. The result is achieved by implementing an innovative *global* normal form in Pöschel's framework.

Finally, we also refer to the recent work [11] and references therein for much more information about stability in various functional classes.

3. Purpose of the work. The objective of this paper is twofold. Our first goal is to make a systematic use of sharp analytic smoothing methods to derive normal forms in a very simple way - whatever the regularity of the Hamiltonians at hand - from the usual anaytic ones. This way we get maximal flexibility to adapt the different long-time stability proofs to a large class of function spaces. We will investigate here only the case of Hölder differentiable Hamiltonians, but our method extends to any steep functions belonging to any regularity class which admits an analytic smoothing. More precisely, the proposed strategy (see Section 4.3) allows us to prove, in a very simple way, the first Nekoroshev-type result of stability for Hölder steep Hamiltonians with presumed sharp exponents[4]. Since in this case one cannot expect to get more than polynomial stability times relative to the size ε of the perturbation [7], in the course of the proof we need to adjust in a rather unusual way the size of the various parameters (ultraviolet cutoff, analyticity widths) as a function of the size ε of the perturbation.

Our second goal is to take advantage of this freedom in the choice of the previous parameters to analyze the mutual dependence of the quantities $\mathbf{E}, \mathbf{T}(\varepsilon), \mathbf{R}(\varepsilon)$, which amounts to fixing one of them and finding the best possible values for the other two. Thereby, we get more flexibility also in the applicability of our results, choosing for instance the best possible threshold at the cost of reducing the time of stability, if necessary. This is indeed a challenging problem in the abovementioned practical applications, which we investigate here from a purely abstract point of view.

---

[2]See [30] for the definition.

[3]The Diophantine-Morse property is a special case of the Diophantine-steep condition introduced in [37] which, in turn, is a prevalent condition on integrable systems that ensures long time stability once these are perturbed. All steep functions are Diophantine-steep.

[4]Sharpness has the same meaning as in [23], i.e. it is the best values of the exponents for $\mathbf{T}(\varepsilon)$ and $\mathbf{R}(\varepsilon)$ that one can obtain with these techniques.

4. Main results. Let us fix the main definitions and assumptions. In the following, given $\nu \in \{1, \dots, \infty\}$, we denote by $|\cdot|_\nu$ the corresponding ℓ^ν-norm in $\mathbb{R}^n$ or $\mathbb{C}^n$. We use the same symbol for the norm in L^ν function spaces when there is no risk of confusion. We denote by $B_\nu(I_0, R)$ the open ball centered at I_0 of radius R for the norm $|\cdot|_\nu$ in $\mathbb{R}^n$.

Consider a Hamiltonian of the form (1.1), where

$$h \in C^\omega(B_\infty(0,R)_{\rho_0}), \qquad f \in C^\ell(B_\infty(0,R) \times \mathbb{T}^n), \tag{1.4}$$

where $B_\infty(0,R)_{\rho_0}$ is the complex extension of analyticity width $\rho_0 \geq 1$ of $B_\infty(0,R)$, and $\ell \in (1, +\infty)$ (meaning that f is Hölder differentiable when ℓ is not an integer, see section 3 for a brief overview on this class of functions). The small parameter is

$$\varepsilon := |f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)}, \tag{1.5}$$

(see (3.2)). We denote by $\omega = \nabla h : \mathbb{R}^n \to \mathbb{R}^n$ the frequency map attached to h.

We will assume that the Hessian of h is uniformly bounded from above:

$$M := \sup_{I \in B_\infty(0,R)_{\rho_0}} \left\| D^2 h(I) \right\|_{op} < \infty, \tag{1.6}$$

where $\| \ \|_{op}$ stands for the operator norm induced by the Hermitian norm on $\mathbb{C}^n$.

We will also assume that the Hamiltonian h is either steep or uniformly strictly convex, according to the following definitions.

Definition 1.1 (Steepness)**.** Fix $\delta > 0$. A C^1 function $h : B_\infty(0, R+\delta) \to \mathbb{R}$ is *steep* with *steepness indices* $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{n-1} \geq 1$ and *steepness coefficients* $C_1, \dots, C_{n-1}, \delta$ if:

(1) $\inf_{B_\infty(0,R)} |\omega(I)|_2 > 0$;

(2) for any $I \in B_\infty(0,R)$ and any m-dimensional subspace Γ orthogonal to $\omega(I)$, with $1 \leq m < n$:

$$\max_{0 \leq \eta \leq \xi} \min_{u \in \Gamma, |u|_2 = \eta} |\pi_\Gamma \omega(I+u)|_2 > C_m \xi^{\boldsymbol{\alpha}_m}, \quad \forall \xi \in (0, \delta], \tag{1.7}$$

where π_Γ stands for the orthogonal projection on Γ.

Definition 1.2 (Uniform strict convexity)**.** We say that h is *uniformly strictly convex* on $B_\infty(0,R)$ when there is a $\mu > 0$ such that the spectrum of $D^2h(I)$ is contained in $[\mu, +\infty)$ for all $I \in B_\infty(0,R)$.

Note that a uniformly strictly convex function is steep with steepness indices equal to 1.
The following main Theorems 1.1 -1.2-1.3- will be proved by adapting to our framework the two usual methods designed to prove exponentially long stability times: the "patchwork method" initiated by Nekhoroshev [31], lately improved by Pöschel [39] and Guzzo-Chierchia-Benettin [23], and the "periodic orbits method" introduced by Lochak [25] and then developed by Niederman [13].
We state our three results separately.

The first one extends the usual statements in the steep case to Hölder perturbations of analytic integrable Hamiltonians. The case where h is Hölder is only slightly more complicated and would yield the same estimates.

Theorem 1.1 (Stability estimates in the steep case)**.** *Consider a near-integrable Hamiltonian system* (1.1) *satisfying* (1.4) *and assume* $\ell \geq n+1$. *Suppose, moreover, that* h *is steep in* $B_\infty(0,R)$ *with steepness indices* $\boldsymbol{\alpha} := (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{n-1})$ *and set:*

$$a = \frac{1}{2n\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_{n-2}} \quad , \qquad b = \frac{1}{2n\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_{n-1}} \ .$$

Then, there exist positive constants $\mathbf{E} = \mathbf{E}(n, \ell, \boldsymbol{\alpha}), \mathtt{C}''_{\mathbf{I}} := \mathtt{C}''_{\mathbf{I}}(n, \ell, \boldsymbol{\alpha})$, $\mathtt{C}''_{\mathbf{T}} := \mathtt{C}''_{\mathbf{T}}(n, \ell, \boldsymbol{\alpha})$ *such that, for* $\varepsilon \leq \mathbf{E}$ *the radius and time of confinement relative to any initial condition in the set* $B_\infty(0, R/2)$ *satisfy:*

$$\mathbf{R}(\varepsilon) \leq \mathtt{C}''_{\mathbf{I}} \varepsilon^b \tag{1.8}$$

$$\mathbf{T}(\varepsilon) \leq \mathtt{C}''_{\mathbf{T}} \frac{\varepsilon^{-a(\ell-1)}}{|(1+a\ell)\ln\varepsilon|^{\ell-1}}\,. \tag{1.9}$$

The form of the other two results is more unusual since tuning parameters are introduced in order to analyze the mutual dependence of the main quantities $\mathbf{E}$, $\mathbf{T}(\varepsilon)$ and $\mathbf{R}(\varepsilon)$.

Theorem 1.2 (Tuned stability estimates in the convex case with the patchwork method)**.**
Consider a near-integrable Hamiltonian system (1.1) *satisfying* (1.4) *and assume* $\ell \geq n+1$*. Suppose, moreover, that* h *is uniformly strictly convex in* $B_\infty(0,R)$ *and fix a tuning parameter* $a \in \big(0, 1/2n\big)$*. Then, there exist positive constants* $\mathtt{C}_{\mathbf{E}} := \mathtt{C}_{\mathbf{E}}(n,\ell)$, $\mathtt{C}_{\mathbf{T}} := \mathtt{C}_{\mathbf{T}}(n,\ell)$, $\mathtt{C}_{\mathbf{I}} := \mathtt{C}_{\mathbf{I}}(n,\ell)$, $\mathtt{C}_\omega$ *such that, relative to any initial action* I *in the set* $B_\infty(0,R/2)$ *such that* $|\omega(I)|_2 \leq \mathtt{C}_\omega$*, the applicability threshold and the time and radius of confinement satisfy:*

$$\begin{aligned}
&|\ln \mathbf{E}|^{2n}\ \mathbf{E}^{1-2an} \leq \frac{\mathtt{C}_{\mathbf{E}}}{(1+a\ell)^{2n}} \\
&\mathbf{T}(\varepsilon) \leq \frac{\mathtt{C}_{\mathbf{T}}}{[(1+a\ell)\ |\ln\varepsilon|]^{2n}\ \varepsilon^{1+a(\ell-1)-2an}} \qquad \forall \varepsilon \leq \mathbf{E}, \\
&\mathbf{R}(\varepsilon) \leq \mathtt{C}_{\mathbf{I}} \frac{\varepsilon^a}{(1+a\ell)\ |\ln\varepsilon|} \qquad \forall \varepsilon \leq \mathbf{E}.
\end{aligned} \tag{1.10}$$

When $|\omega(I)|_2 \leq \mathtt{C}_\omega$*, one has the estimate*

$$\mathbf{R}(\varepsilon) \leq \mathtt{C}_{\mathbf{I}} \tag{1.11}$$

over an infinite time, for the same threshold $\mathbf{E}$*.*

Theorem 1.3 (Tuned stability estimates in the convex case with the periodic averaging method)**.**
Consider the near-integrable Hamiltonian system (1.1) *which satisfies* (1.4) *and assume* $\ell \geq n+1$*. Suppose, moreover, that* h *is uniformly strictly convex in* $B_\infty(0,R)$*, that for some constant* $\mathtt{C}'_\omega > 0$ *the norm of the frequency satisfies* $\inf_{I \in B_\infty(0,R)} |\omega(I)|_2 \geq \mathtt{C}'_\omega$ *and set two tuning parameters* $a, c \in (0, 1/2n)$*, with* $a \leq c$*. Then there exist positive constants* $\mathtt{C}'_{\mathbf{E}} := \mathtt{C}'_{\mathbf{E}}(n,\ell)$, $\mathtt{C}'_{\mathbf{T}} := \mathtt{C}'_{\mathbf{T}}(n,\ell)$, $\mathtt{C}'_{\mathbf{I}} := \mathtt{C}'_{\mathbf{I}}(n,\ell)$ *such that, relative to any initial action* I *in the set* $B_\infty(0,R/2)$ *such that* $|\omega(I)|_2 \leq \mathtt{C}'_\omega$*, the applicability threshold and the time and radius of confinement satisfy:*

$$\begin{aligned}
&|\ln \mathbf{E}|^{2n} \mathbf{E}^{1-2cn} \leq \frac{\mathtt{C}'_{\mathbf{E}}}{(1+a\ell)^{2n}} \\
&\mathbf{T}(\varepsilon) \leq \frac{\mathtt{C}'_{\mathbf{T}}}{(1+a\ell)^{2n}|\ln\varepsilon|^{2n}\ \varepsilon^{1+a(\ell-1)-2cn}} \qquad \forall \varepsilon \leq \mathbf{E}, \\
&\mathbf{R}(\varepsilon) \leq \mathtt{C}'_{\mathbf{I}} \frac{\varepsilon^c}{(1+a\ell)\ |\ln\varepsilon|} \qquad \forall \varepsilon \leq \mathbf{E}.
\end{aligned} \tag{1.12}$$

Remark 1.1. As is well known (see [7]), when ℓ is an integer, the exponents for the time and for the radius of confinement in the convex case can be taken equal to $\frac{\ell-2}{2n}$ and $\frac{1}{2n}$ respectively. By slighting modifying the dependence on ε of the parameters in the proofs of Theorem 1.2 and Theorem 1.3 (see expressions in (5.6) and (6.5)), we could reach the same exponents of stability. However, due to technical issues discussed later in Remark 5.1, this would result in losing the mutual dependence between $\mathbf{T}, \mathbf{R}$ and $\mathbf{E}$ and we will not consider this case in Theorem 1.2 and Theorem 1.3. The exponents of [7] can be recovered - and slightly improved for the stability time - by applying Theorem 1.1, taking into account the fact that a convex function is steep with steepness indices all equal to one.

5. Prospects. We believe that the framework of analytic smoothing could be used in order to give a rigorous setting to deal with the long time stability of concrete physical systems – this issue will be investigated in another work. For example, it is common when studying problems in Celestial Mechanics to consider suitable

truncations of the perturbation, obtained by selecting only a certain number of physically relevant harmonics. This way, one usually obtains realistic thresholds on the size of the perturbation, at the expense of a lack of control on the dynamics generated by the neglected remainder (see e.g. [15], [21] for some examples, and [4] for a comparison on thresholds in two examples with a complete and with a truncated perturbation respectively). Our main remark is that our analytic smoothing amounts to producing a precisely controlled truncation of the perturbation[5], which allows us to expect a better understanding on the shift in dynamics between truncated and non-truncated systems.

In a more abstract setting, the sharpness of the exponents in Theorem 1.1 should be proved in the same way as in the case of convex system. The first attempt to tackle this problem led to work in the Gevrey category instead of the analytic one and construct examples with unstable orbits, which experience a drift in action of the same order as the confinement radius within a time of the same order as the stability time, see [30]. It has then be realized that the initial conjecture ($a \sim 1/2n$, see [17] and Lochak [26]) was in fact incorrect: as proved in [12] using a purely topological argument together with the previous remark on the *local* exponents near simple resonances, one can choose $a = 1/(2(n-1))$ as a global stability exponent for $\mathbf{T}(\varepsilon)$ in quasi-convex analytic systems. This result was improved soon after with $a \sim 1/(2(n-2))$ (see [43]). The construction of unstable system proving the optimality of these latter exponents was achieved in [30], [27], [43]. A remarkable fact is that the unstable mechanism introduced by Arnold in the 60's, with its subsequent improvements, is exactly what is needed to produce the unstable examples in the quasi-convex case.

As for the steep case, a careful construction of the geography of resonances leads with strong evidence to the conjecture that the exponents $a = 1/(2n\boldsymbol{\alpha}_1...\boldsymbol{\alpha}_{n-2})$ and $b = 1/(2n\boldsymbol{\alpha}_1...\boldsymbol{\alpha}_{n-1})$ are sharp (see ref. [23]). The question of constructing explicit examples with unstable orbits proving this sharpness is still open nowadays and is maybe the last challenging problem in the general long time stability theory, relying more on approximation on non-steep examples exhibiting superconductivity channels than on the use of Arnold diffusion ideas.

Continuing along these lines, a natural question raised by our work is that of the optimality of the *simultaneous choice* of $\mathbf{E}, \mathbf{T}(\varepsilon), \mathbf{R}(\varepsilon)$. This problem necessitates a much better understanding of the instability mechanisms at work in perturbations of integrable systems, such as for instance the splitting of separatrices of partially hyperbolic tori, which admit upper bounds directly related to the stability times and the construction of unstable orbits (see [28]).

2. General setting and classical methods: a geometric framework

In this section we give a short overview of the classical methods on which the present work strongly relies.

1. Resonances, resonant normal forms and the steepness condition. Consider a Hamiltonian system of the form (1.1) defined on $O \times \mathbb{T}^n$, where O is an open subset of $\mathbb{R}^n$. The main feature underlying Hamiltonian perturbation theory is that one can modify the form of the perturbation f by composing H with properly chosen local Hamiltonian diffeomorphisms, in order to remove a large number of "nonessential harmonics". The result of this process - a local normal form - strongly depends on the location of the domain of the normalizing diffeomorphism w.r.t the resonances of the unperturbed part h, and enables one to discriminate between fast drift and extremely slow drift directions in the action space, according to this location.

Let us first make this idea more precise. Given an integer lattice $\Lambda \subset \mathbb{R}^n$ of dimension $m \in \{1, \dots, n-1\}$ – a *resonance lattice* – one associates with Λ the resonance vector subspace $\Lambda^\perp \subset \mathbb{R}^n$ in the frequency space $\mathbb{R}^n$, together with the corresponding resonance subset in the action space

$$\mathcal{M}_\Lambda := \omega^{-1}(\Lambda^\perp) = \{I \in O \mid \omega(I) \in \Lambda^\perp\},$$

[5]Which is not simply a truncation on its Fourier series, but is optimal in some sense.

(where $\omega = \nabla h$ is the frequency map). The dimension m of Λ is said to be the multiplicity of the resonance $\mathcal{M}_\Lambda$. Of course, given a resonance module $\Lambda' \supset \Lambda$ with $\dim \Lambda' > \dim \Lambda$, the resonance $\mathcal{M}_{\Lambda'}$ is contained in $\mathcal{M}_\Lambda$, so that a resonance subset contains in general infinitely many resonances of higher multiplicity. The complement $\mathcal{M}_0 \subset O$ of the union of all resonance subsets is the *non-resonant subset*. In general, a resonance subset $\mathcal{M}_\Lambda$ has no particular structure, however, one can think of $\mathcal{M}_\Lambda$ as a submanifold of $\mathbb{R}^n$ of the same dimension as $\Lambda^\perp$ (with perharps singular loci).

As a rule, when ε is small enough[6], for a small enough ε-depending neighborhood W_Λ of the parts of the resonance subset $\mathcal{M}_\Lambda$ *located far enough from resonances of higher multiplicity*[7], one can iteratively construct a symplectic diffeomorphism Ψ_Λ, whose image contains $W_\Lambda \times \mathbb{T}^n$, such that the pull-back $H_\Lambda = H \circ \Psi_\Lambda$ takes the following form

$$H_\Lambda = h + N_\Lambda + R_\Lambda. \tag{2.1}$$

Here R_Λ is a remainder whose C^2 norm is (very) small[8] with respect to ε and the *resonant part* N_Λ contains only harmonics belonging to Λ, that is:

$$N_\Lambda(I,\theta) = \sum_{k\in\Lambda,\, |k|_1 \le K(\varepsilon)} a_k(I)\, e^{ik\cdot\theta},$$

where $K(\varepsilon)$ is an ultraviolet cutoff which has to be properly chosen[9]. Both terms N_Λ and R_Λ of course depend on ε. A subset W_Λ for which such a normal form is proved to exist will be called a *resonant neighborhood* associated with Λ, with multiplicity $\dim \Lambda$.

The iterative process to construct the normalizing diffeomorphism involves the control of small denominators which appear during the resolution of the so-called homological equation, and which depend on the location of the normalization domain with respect to the resonances (see for instance [39] and Section 5). This can be seen as a drawback of the method which could be greatly simplified by an idea due to Lochak (see below), however the general method presented here give precise dynamical informations which would not be reachable otherwise.

The Hamilton equations generated by (2.1) yield the following form for the evolution of the action variables:

$$\begin{aligned} I(t) - I(0) &= \int_0^t \partial_\theta N_\Lambda\big(I(s),\theta(s)\big) + \partial_\theta R_\Lambda\big(I(s),\theta(s)\big)\, ds \\ &= \sum_{k\in\Lambda,\, |k|_1 \le K(\varepsilon)} k \cdot \Big(\int_0^t i\, a_k(I(s))\, e^{ik\cdot\theta(s)}\, ds \Big) + \mathcal{R}(t). \end{aligned} \tag{2.2}$$

The variation of I is therefore the sum of the main part

$$\mathcal{D}(t) := \sum_{k\in\Lambda,\, |k|_1 \le K(\varepsilon)} k \cdot \mathcal{N}^{(k)}(t), \qquad \mathcal{N}^{(k)}(t) = \int_0^t i\, a_k(I(s))\, e^{ik\cdot\theta(s)}\, ds, \tag{2.3}$$

and the very small remainder term $\mathcal{R}(t)$.

To simplify the presentation in the following, we will *forget about the angles* and consider only the action part of the solutions of our system (which is legitimized by the fact that the angles play no role in the various estimates).

The whole theory relies firstly on the obvious fact that the main drift term $\mathcal{D}(t)$ in (2.3) belongs to the vector space $\operatorname{Vect}\Lambda$ spanned by Λ (which is often called *"plane" of fast drift*), and secondly on the smallness of the

[6] As it is habit, ε represents the size of the norm of the perturbation

[7] In fact, only a finite ε-depending subset (related to the cutoff $K(\varepsilon)$ introduced below) of these resonances has to be taken into account.

[8] The smallness depends on the regularity of the system.

[9] This choice is indeed a main issue in the theory.

remainder term $\mathcal{R}$. A solution starting from some initial condition $I(0) \in W_\Lambda$ will therefore remain very close to the fast drift space

$$I(0) + \operatorname{Vect}\Lambda$$

during a very long time –governed by the smallness of $\mathcal{R}$– *as long as it is contained inside the neighborhood* W_Λ. This makes it necessary to understand first the intersections of the fast drift planes $I + \operatorname{Vect}\Lambda$ and the neighborhoods W_Λ to which they are attached.

As an extreme example, let us consider the Hamiltonian

$$h(I) = \frac{1}{2}(I_1^2 - I_2^2)$$

on $\mathbb{A}^2$, with (invertible) frequency map $\omega(I_1, I_2) = (I_1, -I_2)$. We focus on the resonance module $\Lambda = \mathbb{Z}(1,-1)$, so that $\Lambda^\perp = \mathbb{R}(1,1)$ and $\operatorname{Vect}\Lambda = \mathcal{M}_\Lambda$. Hence, given an initial action $I(0) \in \mathcal{M}_\Lambda$, the entire fast drift affine subspace $I(0) + \operatorname{Vect}\Lambda$ coincides with $\mathcal{M}_\Lambda$, so that nothing prevents the fast drift to take place during the whole motion provided the perturbation is well-chosen: the resonance $\mathcal{M}_\Lambda$ is called a *superconductivity channel*. No long time stability result can be expected in this case: indeed, when $f(I,\theta) = \sin(\theta_1 - \theta_2)$, the initial condition $I = 0$, $\theta = 0$ yields the fast evolution $(I_1(t), I_2(t)) = (-\varepsilon t, \varepsilon t)$ for the action variables [10].

In constrast with the previous example, for the Hamiltonian

$$H(I,\theta) = \frac{1}{2}|I|_2^2 + \varepsilon f(\theta)$$

on $\mathbb{A}^n$, for any $\Lambda \subset \mathbb{Z}^n_K$, the the resonant set $\mathcal{M}_\Lambda$ coincides with $\Lambda^\perp$, so that the affine planes of fast drift are always orthogonal to $\mathcal{M}_\Lambda$. In this case a fast drift - if it happens - makes the orbits move away from the resonance in a very short time.

These extreme examples illustrate the role of the Nekhoroshev condition: steepness is an intermediate quantitative property, which prevents from the existence of the superconductivity channels by ensuring a certain amount of transversality between the fast drift planes and the corresponding resonances in action. Starting from an action $I = I(0)$ located at some resonance $\mathcal{M}_\Lambda$, so that its associated frequency $\omega(I)$ is orthogonal to $\Gamma := \operatorname{Vect}\Lambda$, the condition

$$\max_{0\le\eta\le\xi} \min_{u\in\Gamma, |u|_2=\eta} |\pi_\Gamma \omega(I+u)|_2 > C_m \xi^{\boldsymbol{\alpha}_m}, \quad \forall \xi \in (0,\delta], \tag{2.4}$$

(where π_Γ stands for the orthogonal projection on Γ) imposes that a drift of length ξ starting from I and occuring along the fast drift plane $I + \Gamma$ makes the projection $\pi_\Gamma(\omega)$ change by an amount of $C_m \xi^{\boldsymbol{\alpha}_m}$ during the way.

This admits an easy geometric interpretation (see Figure 1). Assume $\dim\Lambda = m$ and consider the vector space Γ spanned by Λ, together with its orthogonal space $\Lambda^\perp$ - of dimension $n - m$. Then one can define a family of tubular neighborhoods of $\Lambda^\perp$ of width $\delta > 0$ by

$$\mathbf{T}_\delta(\Lambda^\perp) = \{\varpi \in \mathbb{R}^n \mid \pi_\Gamma(\varpi) < \delta\}, \quad \delta > 0\,. \tag{2.5}$$

Each such neighborhood gives rise to a neighborhood of the resonance $\mathcal{M}_\Lambda$ in action, namely:

$$\mathbf{W}_\delta(\mathcal{M}_\Lambda) = \omega^{-1}\big(\mathbf{T}_\delta(\Lambda^\perp)\big). \tag{2.6}$$

Therefore, condition (2.4) just says that *any* orbit starting from I and drifting to a distance ξ from I along the plane of fast drift Γ must exit the neighborhood $\mathbf{W}_\delta(\mathcal{M}_\Lambda)$ with $\delta = C_m \xi^{\boldsymbol{\alpha}_m}$.

Note finally that given disjoint subsets $\mathbf{T}$, $\mathbf{T}'$ of tubular neighborhoods of the form (2.5), the associated neighborhoods $\omega^{-1}(\mathbf{T})$ and $\omega^{-1}(\mathbf{T}')$ are disjoint too, whatever the geometric assumptions on the frequency map ω.

2. Nekhoroshev's hierarchy. This section is inspired by Nekhoroshev's ideas as presented in the very nice paper [23]. We also refer to [22] for further details and to [37] for a different approach. Nekhoroshev's strategy

[10] Here a proper choice of the initial angles is needed.

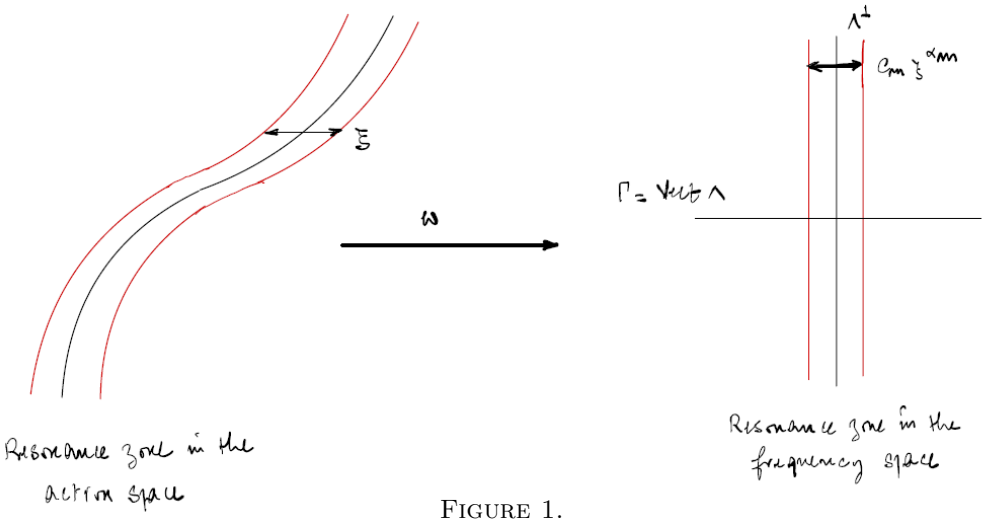

FIGURE 1.

to prove long-time stability results for perturbations of steep Hamiltonians is based on the previous description of resonance neighborhoods, and relies on the following key observation.

Given ε small enough, there exist $T(\varepsilon)$, $R(\varepsilon)$ and a covering of the action space O by resonant "blocks" $(D_{m,p})_{0\leq p\leq p_m}$, for $0\leq m\leq n-1$ which satisfy the following properties:

(1) $T(\varepsilon)\to+\infty$ and $R(\varepsilon)\to 0$ when $\varepsilon\to 0$;

(2) each block $D_{m,p}$ is contained in a resonance neighborhood of multiplicity m and admits an enlargement $\widehat{D}_{m,p}\supset\mathcal{B}_{m,p}$ contained in the same resonance neighborhood;

(3) any solution starting from an initial condition in $D_{m,p}$ either stays inside $\widehat{D}_{m,p}$ for $0\leq t\leq T(\varepsilon)$ or admits a first exit time t_1 such that $I(t_1)$ belongs to a block $D_{m',p'}$ with $m'<m$;

(4) for any initial condition $I(0)$ inside a block $D_{m,p}$ and for any interval $\mathcal{I}$ such that $I(t)\in\widehat{D}_{m,p}$ for all $t\in\mathcal{I}$, then

$$|I(t)-I(0)|_2<R(\varepsilon),\quad\forall t\in\mathcal{I}.$$

We say that m is the multiplicity of the block $D_{m,P}$. Taking the previous observation for granted, the stability of the action variable over a timescale $T(\varepsilon)$ is easy to prove by finite induction. Given an initial condition $I(0)$ located in some block D_{m_0,p_0}, either $I(t)\in\widehat{D}_{m_0,p_0}$ for $0\leq t\leq T(\varepsilon)$, or there is a t_1 such that $I(t)\in\widehat{D}_{m_0,p_0}$ for $0\leq t<t_1$ and $I(t_1)$ belongs to a block D_{m_1,p_1} with $m_1<m_0$. Consequently, there is a finite sequence $(m_0,p_0),\dots,(m_j,p_j)$ such that $m_0>m_1>\cdots>m_j$ (with maybe $m_j=0$) and a finite sequence of times $t_0=0<t_1<\cdots<t_p=T(\varepsilon)$ such that for $0\leq i<j$:

$$I(t)\in\widehat{D}_{(m_i,p_i)},\quad\forall t\in[t_i,t_{i+1}].$$

In words, any orbits crosses a finite number of enlarged blocks during the interval $[0,T(\varepsilon)]$ and get trapped inside the last one. To conclude, one just has to use (4), which proves that the distance between $I(0)$ and $I(t)$ is at most $nR(\varepsilon)$ for $t\in[0,T(\varepsilon)]$.

One should be aware that the covering by the blocks is *not* a partition of O: two distinct blocks may have a nonempty intersection. However, one can *choose* the blocks visited by the orbits according to a hierarchical order, in such a way that their multiplicity decreases as t increases [11]. We say that a covering of O by blocks satisfying the previous properties is a Nekhoroshev patchwork.

3. Construction of Nekhoroshev patchworks. Let us now describe how the blocks are constructed so as to possess their covering and confinement properties[12].

[11] This raises the question of the existence of local finite time Lyapunov functions on the phase space, a still unclear issue.

[12] A source of inspiration for nowadays governments.

Given $\varepsilon > 0$, we first fix an ultraviolet cutoff $K(\varepsilon)$ and consider only the set $\mathbf{M}_\varepsilon$ of resonance modules which are spanned by vectors of length smaller than $K(\varepsilon)$. Given a resonance module $\Lambda \in \mathbf{M}_\varepsilon$ of multiplicity m, we start with the resonant zone of "width" δ_Λ

$$Z_\Lambda := W_{\delta_\Lambda}(\mathcal{M}_\Lambda) = \omega^{-1}\big\{\varpi \in \mathbb{R}^n \mid |\pi_\Gamma(\varpi)|_2 < \delta_\Lambda\big\},$$

where δ_Λ has to be properly chosen as a function of ε and the various geometric invariants of the module (see section 7). We then define the (ε-dependent) resonant zone $\mathcal{Z}_m$ of multiplicity m as

$$Z_m = \bigcup_{\Lambda\in\mathbf{M}_\varepsilon,\, \dim\Lambda=m} Z_\Lambda.$$

Given $\Lambda \in \mathbf{M}_\varepsilon$, $\dim\Lambda = m$, the block attached to Λ is obtained by removing from $\mathcal{Z}_\Lambda$ its intersection with the complete resonant zone of multiplicity $m+1$:

$$D_\Lambda = Z_\Lambda \setminus Z_{m+1}.$$

The blocks $D_{m,p}$ are the connected components of $\mathcal{Z}_m$. With no great loss of generality, one can think of (the closure of) a block as a submanifold with boundary and corners –even if it is not necessary.

The following figure shows the construction of the blocks in the case $n = 3$ (and in a transverse section). The resonance zone of multiplicity 2 if the disjoint union of the blue blocks, the resonance zone of multiplicity 1 is the union on the strips with red boundaries, while the 0-multiplicity zone is the complement of the 1-multiplicity zone.

In any case, the blocks satisfy two main properties.

– The closures of two different blocks can intersect only when their multiplicities are distinct.

This comes from a very careful choice of the widths of the various resonance zones (see [23] and Section 7), which in fact ensures a more stringent (and crucial) property: the enlargement of a block contained in some D_Λ cannot intersect *any other block* contained in the zone D_Λ, neither *any other neighborhood* $\mathcal{M}_{\Lambda'}$ *with* $\dim\Lambda' = \dim\Lambda$ (see below for precisions on the construction of the enlargement).

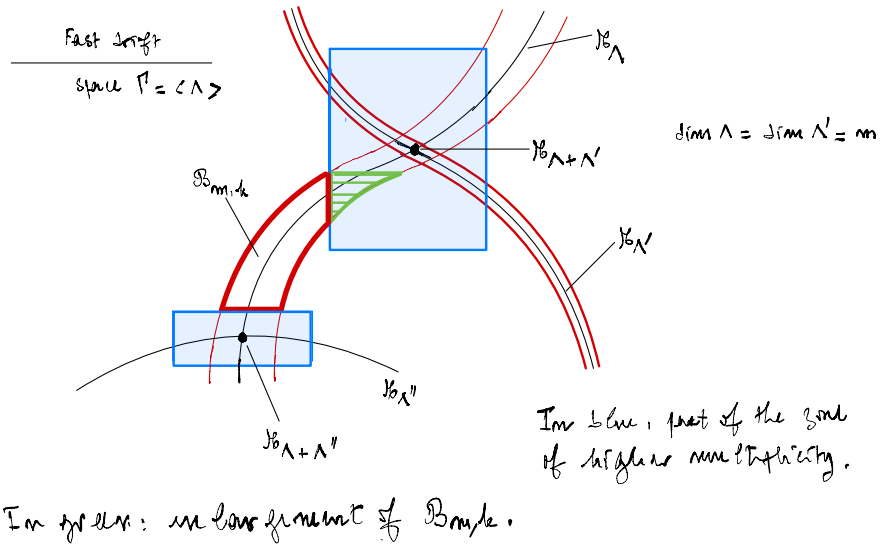

FIGURE 2.

We state the second property in the spirit of Conley's isolating blocks theory.

– The frontier $\partial D_{m,p}$ of $D_{m,p}$ is the union of two subsets

$$\partial D_{m,p} = \partial^+ D_{m,p} + \partial^- D_{m,p}$$

where $\partial^+ D_{m,p}$ (resp. $\partial^- D_{m,p}$) is contained in blocks $D_{m',p'}$ with $m' > m$ (resp. $m' < m$).

This raises new questions which could be the starting point of a better understanding of the relations between diffusion along invariant subsets and long-time stability theory. Indeed, given a block $\mathcal{B}_{m,k}$, a description of the (generic) features of the Hamiltonian vector field X_{H_ε} at the frontier $\partial\mathcal{B}_{m,p}$ has never been done. In particular, nothing is known on the locus where X_{H_ε} "enters the block" and the locus where X_{H_ε} "exits the block". These two subsets are crucial for the understanding of the homology of the invariant sets contained into the blocks, following Conley's theory, and could provide one with a new tool for constructing diffusing orbits in the steep setting.

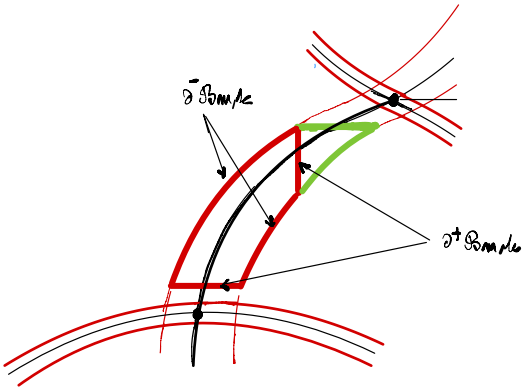

FIGURE 3.

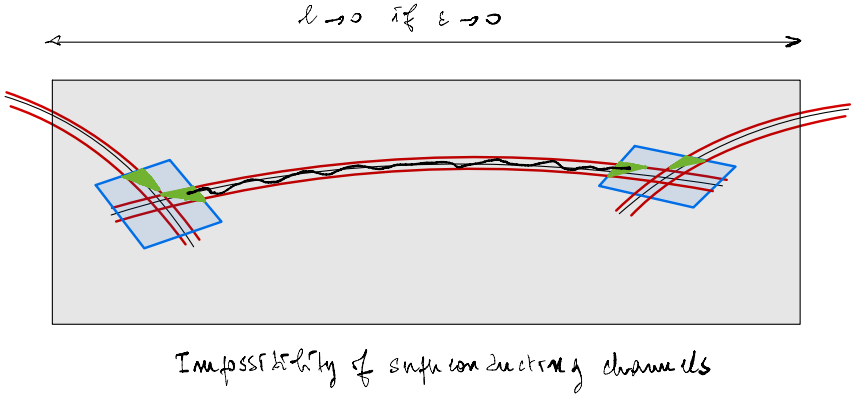

FIGURE 4.

Going back to the construction of a Nekhoroshev patchwork, we have to make precise the process conducting to the enlargement of a block and its stability property. Here we will again make a crucial use of the fact that an orbit starting from an initial condition $I := I(0)$ located in $\mathcal{B}_{m,k}$ will remain extremely close to the fast drift space $I + \operatorname{Vect}\Lambda$ for $0 \leq t \leq T(\varepsilon)$, as long as it stays inside the resonant neighborhood $\mathcal{M}_\Lambda$ and far enough to the higher multiplicity resonance zones. Hence, to enlarge the block $\mathcal{B}_{m,k}$, we just have to add to it the collection of all the parts of the disks centered at points $I \in \mathcal{B}_{m,k}$ which are contained in the intersection of

the fast drift spaces $I + \operatorname{Vect}\Lambda$ with the resonance neighborhood $\mathcal{M}_\Lambda$ (the resulting added subset is the green part in the previous two figures). We have in fact to add a very small neighborhood of these union of disks, in order to prevent the solutions to exit the extended block under the influence of the remainder part $\mathcal{R}$ of the dynamics during the time $T(\varepsilon)$, but this would not change our description significantly. Finally, one has to make sure that the extension will not intersect any other block of the same neighborhood $\mathcal{B}_\Lambda$ or any other resonance neighborhood, which can be done by a careful tuning of the width of the zone (see Section 7).

This concludes our description of Nekhoroshev's method.

4. Geometric confinement in the convex case. As proved by Nekhoroshev in [32], steeepness is a generic property in the C^∞ category (see also [36] and [42], [3] for an explanation of the algebraic properties that are fulfilled by steep functions). However, lots of physical examples of Hamiltonian systems are perturbations of convex or quasi-convex integrable Hamiltonians. This makes this class of perturbed systems particularly relevant from the point of view of long-time stability theory, even if it is far from being the generic situation. In addition, as noted first by Gallavotti [20], the convexity of the integrable part confers the perturbed system much more efficient confinement properties in the neighborhood of a block, which yields a simpler scheme than the original Nekhoroshev's one to prove long-time stability. Indeed, the convexity (or quasi convexity) property of the integrable part allows one to confine the orbits by using only the energy conservation.

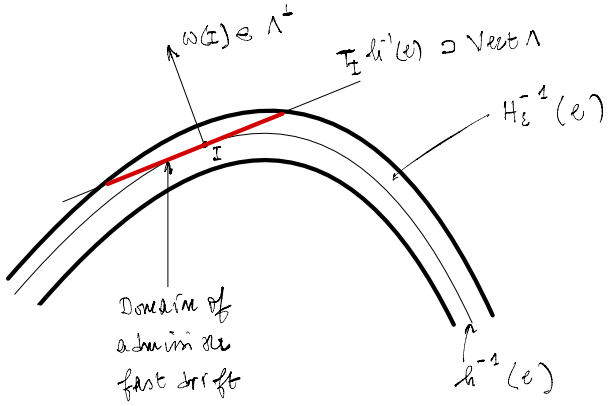

FIGURE 5.

In the same way as for our description of the Nekhoroshev mechanism, we find it useful to neglect the slow drift due to the remainder term, in a first approximation (see [39] for a complete description). Given an initial action I located on some unperturbed energy level $h^{-1}(e)$, with frequency $\omega(I)$ orthogonal to some resonant module Λ, the main remark is that the fast drift space $\operatorname{Vect}\Lambda$ is contained in $\omega(I)^\perp$, which is nothing but the tangent space to $h^{-1}(e)$ at I. By energy preservation, the motion of the actions starting at $I(0)$ along the fast drift space is constrained to stay inside the intersection of this space with the projection of the initial energy level of H_ε on the action space, which is small when ε is small (see Figure 5, which also shows the kind of tangency that $\operatorname{Vect}\Lambda$ and $h^{-1}(e)$ can have since h is convex). As a consequence, an orbit starting from a block contained inside a neighborhood of the form (2.6) is constrained to stay in a very small neighborhood of the block, which is itself a block of the same multiplicity, with slightly larger width. This way, one no longer needs to invoke the Nekhoroshev hierarchical scheme since one and the same block is enough to confine an orbit.

Let us give a more quantitative argument for the "enlarged block" inside which the orbit will stay. The variation of the integrable part of the Hamiltonian starting from an initial action $I(0)$ and ending at $I(T)$ reads

$$\Delta h = \omega_0 \cdot \Delta I + \int_0^1 (1-s)\big(D^2h(I(s))\Delta I \cdot \Delta I\big)\, ds$$

where $\omega_0 = \omega\big(I(0)\big)$, $\Delta I = I(T) - I(0)$ and $I(s) = I(0) + s\Delta I$, so that if h is μ-uniformly strictly convex

$$\frac{\mu}{2}|\Delta I|_2^2 \le |\Delta h| + |\omega_0 \cdot \Delta I|.$$

Assuming that the projection $\pi_\Gamma(\omega)$ is bounded by δ in the block yields

$$|\omega_0 \cdot \Delta I| \leq |\pi_\Gamma(\omega_0)|_2 |\pi_\Gamma(\Delta I)|_2 \leq \delta\, |\pi_\Gamma(\Delta I)|_2,$$

so that, forgetting about the slow drift term and replacing $\pi_\Gamma(\Delta I)$ by ΔI:

$$|\omega_0 \cdot \Delta I| \leq \delta\, |\Delta I|_2 \leq \frac{1}{2}\Big(\frac{2\delta^2}{\mu} + \frac{\mu}{2}|\Delta I|_2^2\Big).$$

By the first and last inequalities:

$$|\Delta I|_2^2 \leq \frac{4}{\mu}\Big(|\Delta h| + \frac{\delta^2}{\mu}\Big) \leq C\varepsilon + \frac{4\delta^2}{\mu^2}.$$

The length of the fast drift is therefore *uniformly* bounded, which easily yields the conclusion that an orbit starting inside a block will stay in a slightly larger one as long as the small drift due to the remainder term in 2.2 can be bounded from above by a controlled quantity, which in turn yields the long-time stability.

5. Lochak's method. As alluded to above, an alternative proof of the long-time stability of the actions which completely avoids the small divisors problems appeared at the beginning of the 90's with Lochak's work [25], where a new method based on *single frequency* averaging was also introduced. The main idea is to "stick to the dynamics" from the very beginning and consider only the neighborhood of the periodic tori of h (that is, those Lagrangian tori which are foliated by periodic orbits) to derive normal forms. To determine those neighborhoods, one is therefore led to analyze the approximation of frequency vectors by periodic ones (that is, vectors $\omega \in \mathbb{R}^n$ such that there exists $T > 0$ with $T\omega \in \mathbb{Z}^n$), while in the previous scheme the main constraint consists in controlling the small denominators $k \cdot \omega$ – this is the classical opposition between simultaneous approximation and classical Diophantine approximation. Besides its simplicity, a major conceptual advantage of simultaneous approximation is that is has an obvious dynamical interpretation: the closeness of a frequency vector to a periodic one has an immediate translation in terms of the recurrence properties of the associated linear flow, while the dynamical interpretation of the small denominators is much more intricate[13]. However, the patchwork methods yield more information about the local motions and geometric features of the invariant subsets (splitting of separatrices for instance) since they carefully analyze the dynamical data close to resonances of any multiplicity, while the periodic orbits method considers only resonances of "maximal" multiplicity.

So Lochak's method is based on a simpler patchwork than Nekhoroshev's one, – in this case all the blocks are neighborhoods of periodic tori – over which normalizing symplectic diffeomorphisms are proved to exist. In the same way as for the construction mentioned above, in any of such blocks the associated normal form for the initial Hamiltonian is the sum of h, a resonant part N and a "small remainder" Z. Using this strategy, stability theorems can be obtained both in the convex and in the generic steep or Diophantine-steep [14] cases (see e.g. [29], [8], [13]) but sharp estimates of stability under the two latter constraints seem to be reached only by Nekhoroshev's original construction[15] (see [23] and [35], [38]). On the other hand, the Lochak method turns out to be more adapted when trying to extend Nekhoroshev results to other contexts (for example to elliptic points [19], [34], [9], [10] or to infinite-dimensional systems [1], [40]) since it involves much easier computations.

3. Functional setting

For $n \geq 1$, we denote the standard n-dimensional torus by $\mathbb{T}^n = \mathbb{R}^n/2\pi\mathbb{Z}^n$ and the standard $2n$-dimensional annulus by $\mathbb{A}^n = \mathbb{R}^n \times \mathbb{T}^n$.

1. Hölder differentiable functions. Given an integer $q \geq 0$ and an open subset D of $\mathbb{R}^n$, we denote by $C^q(D)$ the set of q-times continuously differentiable maps $f : D \to \mathbb{R}$ ($C^0(D)$ being the set of continuous

[13] see [6, 18] for an interpretation in terms of ergodization rates.

[14] See note 3.

[15] The authors conjecture that, in order to obtain sharp estimates of stability in the steep case with the Lochak method, one should use refined theorems on diophantine approximations.

functions on D). We identify $C^q(\mathbb{T}^n)$ with the subset of $C^q(\mathbb{R}^n)$ formed by the functions that are $2\pi\mathbb{Z}^n$-periodic and $C^q(D\times\mathbb{T}^n)$ with the subset of $C^q(D\times\mathbb{R}^n)$ formed by the functions which are $2\pi\mathbb{Z}^n$-periodic with respect to their last n variables.

We use the conventional notation for partial derivatives: given $f\in C^q(D)$ and $\alpha\in\mathbb{N}^n$, we set for $x\in D$:

$$\partial^\alpha f(x)=\frac{\partial^{|\alpha|}f}{\partial x_1^{\alpha_1}\dots\partial x_n^{\alpha_n}}(x),$$

with $|\alpha|=\alpha_1+\cdots+\alpha_n$.

We denote by $C^q_b(D)$ the set of $f\in C^q(D)$ such that

$$\|f\|_{C^q(D)}:=\sup_{|\alpha|\le q}\sup_{x\in D}|\partial^\alpha f(x)|<+\infty, \tag{3.1}$$

so that $\left(C^q_b(D),\|\cdot\|_{C^q(D)}\right)$ is a Banach space with multiplicative norm[16]. It is understood that, for a function defined on a compex domain D, the $\|\cdot\|_{C^0(D)}$ is the usual sup-norm.

If $\ell>0$ is a non-integer real number, we write $q:=\lfloor\ell\rfloor$ for its integer part and $\mu=\ell-q\in(0,1)$ for its fractional part. Given a non-negative integer q and $\mu\in(0,1)$, we denote by $C^{q,\mu}_b(D)$ the space formed by those functions $f\in C^q(D)$ such that

$$|f|_{C^{q,\mu}(D)}:=\|f\|_{C^q(D)}+\sup_{\alpha\in\mathbb{N}^n:|\alpha|=q}\sup_{\substack{x,y\in D:\\ 0<|x-y|<1}}\frac{|\partial^\alpha f(x)-\partial^\alpha f(y)|}{|x-y|^\mu}<+\infty. \tag{3.2}$$

It is well-known that $\left(C^{q,\mu}_b(D),|\cdot|_{C^{q,\mu}(D)}\right)$ is also a Banach space with multiplicative norm. Functions belonging to such spaces are called Hölder-differentiable functions.

Given a non-integer real number $\ell>0$, together with its integer part $q:=\lfloor\ell\rfloor$ and its fractional part $\mu=\ell-q\in(0,1)$, we also write $C^\ell_b(D)$ instead of $C^{q,\mu}_b(D)$ and $|\cdot|_{C^\ell(D)}$ instead of $|\cdot|_{C^{q,\mu}(D)}$. Clearly $C^\ell_b(D)\subset C^{\ell'}_b(D)$ when $\ell\ge\ell'$ and if $f\in C^\ell_b(D)$

$$|f|_{C^{\ell'}(D)}\le|f|_{C^\ell(D)}. \tag{3.3}$$

2. Domains and their complex extensions.

Let us define the complex n-dimensional torus $\mathbb{T}^n_{\mathbb{C}}$ and the complex $2n$-dimensional annulus $\mathbb{A}^n_{\mathbb{C}}$ as

$$\mathbb{T}^n_{\mathbb{C}}=\mathbb{C}^n/2\pi\mathbb{Z}^n\qquad\text{and}\qquad\mathbb{A}^n_{\mathbb{C}}=\mathbb{C}^n\times\mathbb{T}^n_{\mathbb{C}}. \tag{3.4}$$

We use angle coordinates θ on $\mathbb{T}^n_{\mathbb{C}}$ (with the usual abuse $\theta\in\mathbb{C}^n$ when there is no ambiguity) and action-angle coordinates (I,θ) on $\mathbb{A}^n_{\mathbb{C}}$. We see $\mathbb{T}^n_{\mathbb{C}}$ as a real n-dimensional vector bundle over $\mathbb{T}^n$. Consequently, we write

$$|\theta|:=\max_j\left(|\operatorname{Im}\theta_j|\right),\qquad |I|:=\max_j|I_j|\,,\qquad |(I,\theta)|=\max\left(|I|,|\theta|\right) \tag{3.5}$$

For integer vectors $k\in\mathbb{Z}^n$, we use the "dual" ℓ^1-norm, which we write $|k|$ only when there is no risk of confusion.

We need to introduce specific domains in $\mathbb{A}^n_{\mathbb{C}}$. First, given $r>0$, for a domain $D\subset\mathbb{R}^n$, we set

$$D_r:=\left\{z\in\mathbb{C}^n:\exists z^*\in D:|z-z^*|_2<r\right\}. \tag{3.6}$$

As for the torus, given $s>0$, we introduce the global complex neighborhood

$$\mathbb{T}^n_s:=\left\{\theta\in\mathbb{T}^n_{\mathbb{C}}\,:\,|\theta|<s\right\}. \tag{3.7}$$

We will essentially deal with complex domains of the form

$$\mathcal{D}_{r,s}\;:=D_r\times\mathbb{T}^n_s\subset\mathbb{A}^n_{\mathbb{C}}. \tag{3.8}$$

We finally write $D^{\mathbb{R}}_r$ and $\mathcal{D}^{\mathbb{R}}_{r,s}$ for the projections of D_r and $\mathcal{D}_{r,s}$ on $\mathbb{R}^n$ and $\mathbb{A}^n$ respectively.

[16]That is, satisfying an inequality of the form $|fg|\le C|f||g|$ for a suitable constant C.

3. Analytic functions and norms. If g is a bounded holomorphic function defined on $\mathbb{T}^n_s, D_r$ or $\mathcal{D}_{r,s}$ we denote the corresponding classical sup-norms by

$$|g|_s = \sup_{\theta\in\mathbb{T}^n_s} |g(\theta)|, \quad |g|_r = \sup_{I\in D_r} |g(I)|, \quad |g|_{r,s} = \sup_{(I,\theta)\in\mathcal{D}_{r,s}} |g(I,\theta)|. \tag{3.9}$$

Fix a bounded holomorphic function $g : \mathcal{D}_{r,s+2\sigma} \to \mathbb{C}$, where $\sigma > 0$, and let $g(I,\theta) = \sum_{k\in\mathbb{Z}^n} \hat{g}_k(I) e^{\mathrm{i}\, k\cdot\theta}$ be its Fourier expansion, where $k\cdot\theta = k_1\theta_1 + \cdots + k_n\theta_n$. We then introduce the *weighted Fourier norm*

$$||g||_{r,s} := \sup_{I\in D_r} \sum_{k\in\mathbb{Z}^n} |\hat{g}_k(I)|\, e^{|k|s}, \tag{3.10}$$

which is finite and satisfies

$$|g|_{r,s} \le ||g||_{r,s} \le \coth^{\mathrm{n}}\sigma\, |g|_{r,s+\sigma}. \tag{3.11}$$

We denote by $\mathcal{A}_{r,s}$ the space of holomorphic functions on $\mathcal{D}_{r,s}$ with finite Fourier norm. Endowed with this norm, $\mathcal{A}_{r,s}$ is a Banach algebra.

Finally, the norm of a vector valued function will be the maximum of the norms of its components.

4. Analytic smoothing

We state in this section the key ingredient of the present work. We first recall the analytic smoothing method as developed by Jackson-Moser-Zehnder for Hölder functions on $\mathbb{R}^n$: given a Hölder function $f \in C^\ell(\mathbb{R}^n)$ and a positive number $s \le 1$, this yields an analytic function on the complex neighborhood $\mathbb{R}^n_s$ whose restriction to $\mathbb{R}^n$ is close to f in the C^k topology, for $1 \le k \le \ell$.

We then adapt their method to our specific setting of functions defined on $\mathbb{A}^n$ (see Section 4.2) and, in addition, we derive the new estimate (4.22) for the weighted Fourier norm of the smoothed function.

4.1. Analytic smoothing in $\mathbb{R}^n$.

We recall here the result by Jackson, Moser and Zehnder, following the presentation by [16] and [41].

Proposition 4.1 (Jackson-Moser-Zehnder)**.** *Fix an integer $n \ge 1$ and a real number $\ell > 0$. Fix $f \in C^\ell_b(\mathbb{R}^n)$. Then there is a constant $\mathtt{C}_\mathtt{J} = \mathtt{C}_\mathtt{J}(\ell, n)$ such that for every $0 < s \le 1$ there exists a function $\mathtt{f}_s$, analytic on $\mathbb{R}^n_s$, which satisfies*

$$\left| \partial^\alpha \mathtt{f}_s(x) - \sum_{\beta\in\mathbb{N}^n : |\beta| \le \lfloor \ell \rfloor - |\alpha|} \partial^{\alpha+\beta} f(\mathrm{Re}\, x) \frac{(\mathrm{Im}\, x)^\beta}{\beta!} \right| \le \mathtt{C}_\mathtt{J} s^{\ell-|\alpha|} |f|_{C^\ell(\mathbb{R}^n)}, \quad \forall x \in \mathbb{R}^n_s, \tag{4.1}$$

for all multi-integer $\alpha \in \mathbb{N}^n$ such that $|\alpha| \le \lfloor \ell \rfloor$. More precisely, given any even C^∞ function Φ with compact support in $\mathbb{R}^n$ and setting

$$K(\xi) := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \Phi(x) e^{\mathrm{i} x\cdot\xi} dx, \quad \xi \in \mathbb{R}^n_s, \tag{4.2}$$

the function

$$\mathtt{f}_s(x) := \int_{\mathbb{R}^n} K\left(\frac{x}{s} - \xi\right) f(s\xi)\, d\xi\ , \tag{4.3}$$

satisfies the previous requirements (where the constant $\mathtt{C}_\mathtt{J}(\ell, n)$ depends on the choice of Φ).

Observe that $\mathtt{f}_s$ takes real values when its argument is in $\mathbb{R}^n$.

4.2. **Analytic smoothing in $\mathbb{A}^n$.** In the following, the Hölder regularity ℓ is assumed to satisfy $\lfloor \ell \rfloor \geq n+1$ as in the hypothesis of Theorems 1.1-1.2-1.3.
We now specialize the previous result to our setting and give a more detailed description of the method in the case of functions on $\mathbb{A}^n$. In that case, the analytic smoothing is a truncation of the Fourier series of the initial Hölder function with suitably modified Fourier coefficients (the so-called Jackson polynomials). Our main concern here is to derive an estimate on the weighted Fourier norm of an s-smoothed C^ℓ function over a complex strip of width s.

To make the whole presentation more explicit *and take the anisotropy of the weighted Fourier norm into account*, we first separately consider functions defined on $\mathbb{R}^n$ and $\mathbb{T}^n$. This then yields an easy statement for functions on $\mathbb{A}^n$.

• *The non-periodic case.* Fix an even function $\Phi : \mathbb{R}^n \to [0,1]$, of class C^∞, with support in the ball $\overline{B}_2(0,1)$ and let $K : \mathbb{C}^n \to \mathbb{C}$ be its Fourier-Laplace transform:

$$K(y) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \Phi(\eta) e^{-i\eta\cdot y} d\eta. \tag{4.4}$$

Since Φ is compactly supported, then K is an entire function . Moreover its restriction to $\mathbb{R}^n$ is in the Schwartz class $\mathcal{S}(\mathbb{R}^n)$ since Φ is, and this is also the case for the translates $y \mapsto K(y-z)$ for $y \in \mathbb{R}^n$ and fixed $z \in \mathbb{C}^n$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^ℓ function with $\lfloor \ell \rfloor \geq n+1$, with compact support contained in the ball $\overline{B}_\infty(0, R_0)$ for some $R_0 > 0$. Given $s \in]0,1]$, set for $x \in \mathbb{R}^n$:

$$\mathbf{f}_s(x) = \frac{1}{s^n} \int_{\mathbb{R}^n} K\Big(\frac{x-y}{s}\Big) f(y) dy = \int_{\mathbb{R}^n} K\Big(\frac{x}{s} - y\Big) f(sy) dy = \int_{\mathbb{R}^n} K(y) f(x - sy) dy. \tag{4.5}$$

By Fourier reciprocity:

$$\mathbf{f}_s(x) = \int_{\mathbb{R}^n} \Phi(\eta) \widehat{f(x-sy)}(\eta) d\eta,$$

with:

$$\widehat{f(x-sy)}(\eta) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} f(x-sy) e^{-iy\cdot\eta} dy = \frac{1}{(2\pi)^n s^n} \int_{\mathbb{R}^n} f(u) e^{-i(x-u)\cdot\eta/s} du = \frac{e^{-ix\cdot\eta/s}}{s^n} \widehat{f}\Big(\frac{-\eta}{s}\Big).$$

Therefore, since Φ is even:

$$\mathbf{f}_s(x) = \frac{1}{s^n} \int_{\mathbb{R}^n} \Phi(\eta) \widehat{f}\Big(\frac{-\eta}{s}\Big) e^{-ix\cdot\eta/s} d\eta = \int_{\mathbb{R}^n} \Phi(s\eta) \widehat{f}(-\eta) e^{-ix\cdot\eta} d\eta = \int_{\mathbb{R}^n} \Phi(s\eta) \widehat{f}(\eta) e^{ix\cdot\eta} d\eta. \tag{4.6}$$

Hence $\mathbf{f}_s$ is the inverse Fourier-Laplace transform of the "truncation"

$$\eta \mapsto \Phi(s\eta) \widehat{f}(\eta).$$

The first term of (4.5) shows that $\mathbf{f}_s$ extends to $\mathbb{C}^n$ and is an entire function. To get our final estimate we go back to the second term in (4.5), which yields

$$|(\mathbf{f})_s(z)| \leq \|f\|_{C^0(\mathbb{R}^n)} \int_{\mathbb{R}^n} \Big| K\Big(\frac{z}{s} - y\Big)\Big| dy, \qquad z \in \mathbb{C}^n. \tag{4.7}$$

By the Schwartz estimate of Lemma A.1, there exists a constant C_n such that

$$\Big| K\Big(\frac{z}{s} - y\Big)\Big| \leq C_n \frac{e^{\mathrm{Im}(z/s - y)}}{(1 + |z/s - y|_2)^{n+1}},$$

so that, for $y \in \mathbb{R}^n, z \in \mathbb{C}^n$ and $|\operatorname{Im} z|_2 \leq s$:

$$\Big| K\Big(\frac{z}{s} - y\Big)\Big| \leq C_n \frac{e}{(1 + |\operatorname{Re}(z/s - y)|_2)^{n+1}}.$$

Hence:

$$|(\mathbf{f})_s(z)| \leq \|f\|_{C^0(\mathbb{R}^n)} C_n e \int_{\mathbb{R}^n} \frac{dy}{(1 + |y|_2)^{n+1}}. \tag{4.8}$$

since z/s is fixed and can be eliminated by a simple translation. We finally get the following estimate:

$$\|\mathbf{f}_s\|_s = \sup_{z\in\mathbb{C}^n:|\mathrm{Im}\, z|_2\le s} \le C_1(n)\,\|f\|_{C^0(\mathbb{R}^n)}\,, \tag{4.9}$$

with

$$C_1(n) := C_n e \int_{\mathbb{R}^n} \frac{dy}{(1+|y|_2)^{n+1}} < \infty.$$

• *The periodic case.* Fix now an even function $\Psi : \mathbb{R}^n \to [0,1]$, of class C^∞, with support in the ball $\overline{B}_1(0,1)$ and define the associate kernel K as in (4.4).

Fix a $2\pi\mathbb{Z}^n$-periodic function $f \in C^\ell(\mathbb{R}^n)$ with $\ell \ge n+1$. Then the Fourier expansion

$$f(\theta) = \sum_{k\in\mathbb{Z}^n} \widehat{f}_k e^{ik\cdot\theta}, \qquad \widehat{f}_k = \frac{1}{(2\pi)^n}\int_{\mathbb{T}^n} f(\varphi)e^{-ik\cdot\varphi}d\varphi,$$

converges normally since, by Lemma A.2 in Appendix A, for $k \in \mathbb{Z}^n \setminus \{0\}$, there exists a universal constant $\mathtt{C}_\mathtt{F}(n,\ell)$ satisfying

$$\left|\widehat{f}_k\right| \le \mathtt{C}_\mathtt{F}(n,\ell)\frac{||f||_{C^{\lfloor\ell\rfloor}}}{|k|_\infty^{\lfloor\ell\rfloor}} \tag{4.10}$$

and $\lfloor\ell\rfloor \ge n+1$ by hypothesis. For $s \in]0,1]$, the function

$$\mathbf{f}_s(\theta) = \frac{1}{s^n}\int_{\mathbb{R}^n} K\Big(\frac{\theta-\varphi}{s}\Big) f(\varphi)d\varphi$$

is well-defined and, by the Fubini interversion theorem:

$$\mathbf{f}_s(\theta) = \sum_{k\in\mathbb{Z}^n} \widehat{f}_k \int_{\mathbb{R}^n} K(\varphi)e^{ik\cdot(\theta-s\varphi)}d\varphi = \sum_{k\in\mathbb{Z}^n} \widehat{f}_k e^{ik\cdot\theta}\int_{\mathbb{R}^n} K(\varphi)e^{-isk\cdot\varphi}d\varphi.$$

Hence, since K is the inverse Fourier transform of Ψ, by the Fourier inversion theorem:

$$\mathbf{f}_s(\theta) = \sum_{k\in\mathbb{Z}^n} \widehat{f}_k \Psi(sk)\, e^{ik\cdot\theta}, \quad \theta\in\mathbb{R}^n. \tag{4.11}$$

As in the non-periodic case, this makes apparent that $\mathbf{f}_s$ is a continuous truncation of the Fourier expansion of f with a Ψ-dependent modification of its Fourier coefficients (the so-called Jackson polynomial):

$$\widehat{(\mathbf{f}_s)}_k = \Psi(sk)\widehat{f}_k\ . \tag{4.12}$$

Consequently, the Fourier norm

$$\|\mathbf{f}_s\|_s = \sum_{k\in\mathbb{Z}^n} \left|\widehat{(\mathbf{f}_s)}_k\right| e^{s|k|_1}$$

depends only on the harmonics such that $|k|_1 \le 1/s$ and satisfies

$$\|\mathbf{f}_s\|_s \le \sum_{|k|_1\le 1/s} \left|\widehat{(\mathbf{f}_s)}_k\right| e^{s|k|_1} \le e \sum_{|k|_1\le 1/s} \left|\widehat{(\mathbf{f}_s)}_k\right| \le e \sum_{k\in\mathbb{Z}^n} \left|\widehat{f}_k\right|.$$

Hence, by (4.10):

$$\|\mathbf{f}_s\|_s \le C_2(\ell)|f|_{C^{\lfloor\ell\rfloor}} \tag{4.13}$$

with

$$C_2(\ell) := e\left(1 + \mathtt{C}_\mathtt{F}(n,\ell)\sum_{k\in\mathbb{Z}^n\setminus\{0\}} \frac{1}{|k|_\infty^{[\ell]}}\right) \tag{4.14}$$

• *Functions on* $\mathbb{A}^n$. We finally gather together the previous two cases. Let $\Phi\otimes\Psi : \mathbb{R}^n\times\mathbb{R}^n \to [0,1]$ be defined by

$$\Phi\otimes\Psi(x,\theta) = \Phi(x)\Psi(\theta),$$

and define the kernel

$$K(y,\varphi)=\int_{\mathbb{R}^{2n}}\Phi\otimes\Psi(x,\theta)\,e^{-i(x,\theta)\cdot(y,\varphi)}\,dxd\theta=K_\Phi(y)K_\Psi(\varphi)=K_\Phi\otimes K_\Psi(y,\varphi)$$

where K_Φ and K_Ψ are defined as above.

Fix a function $f:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{C}$, $2\pi\mathbb{Z}^n$-periodic with respect to its last n variables, with support in $\overline{B}_2(0,R_0)\times\mathbb{R}^n$ for some $R_0>0$, belonging to $C^\ell(\mathbb{R}^{2n})$ with $\lfloor\ell\rfloor\geq n+1$. For $s\in]0,1]$ and $(x,\theta)\in\mathbb{R}^n\times\mathbb{R}^n$, set

$$\begin{aligned}\mathbf{f}_s(x,\theta)&=\int_{\mathbb{R}^{2n}}K(y,\varphi)f(x-sy,\theta-s\varphi)dyd\varphi\\&=\int_{\mathbb{R}^{2n}}K(y,\varphi)\sum_{k\in\mathbb{Z}^n}\widehat{f}_k(x-sy)e^{ik\cdot(\theta-s\varphi)}dyd\varphi\end{aligned}$$

with

$$\widehat{f}_k(u)=\frac{1}{(2\pi)^n}\int_{\mathbb{T}^n}f(u,v)e^{-ik\cdot v}dv. \tag{4.15}$$

Note that f_k is C^ℓ, with support in $\overline{B}_2(0,R_0)$, so that the previous study applies to f_k.

By Fubini interversion

$$\begin{aligned}\mathbf{f}_s(x,\theta)&=\sum_{k\in\mathbb{Z}^n}\int_{\mathbb{R}^{2n}}K(y,\varphi)\widehat{f}_k(x-sy)e^{ik\cdot(\theta-s\varphi)}dyd\varphi\\&=\sum_{k\in\mathbb{Z}^n}\Big(\int_{\mathbb{R}^n}K_\Phi(y)\widehat{f}_k(x-sy)dy\Big)\Big(\int_{\mathbb{R}^n}K_\Psi(\varphi)e^{ik\cdot(\theta-s\varphi)}d\varphi\Big)\\&=\sum_{k\in\mathbb{Z}^n}(\widehat{\mathbf{f}}_k)_s(x)\Psi(sk)e^{ik\cdot\theta}\end{aligned} \tag{4.16}$$

where $(\widehat{\mathbf{f}}_k)_s$ stands for the analytic smoothing of the Fourier coefficient $\widehat{f}_k$. This proves that the Fourier coefficient $(\widehat{\mathbf{f}}_s)_k(x)$ relative to the periodic variable θ reads

$$(\widehat{\mathbf{f}}_s)_k(x)=\Psi(sk)(\widehat{\mathbf{f}}_k)_s(x),\quad k\in\mathbb{Z}^n. \tag{4.17}$$

Expressions (4.16) and (4.17) make clear that the whole smoothing procedure of a function depending both on action and angle variables consists in constructing a Jackson trigonometric polynomial by smoothing the Fourier coefficients and by suitably truncating the Fourier series.

Using the definition of Ψ, $(\widehat{\mathbf{f}}_s)_k=0$ when $|k|_1>1/s$ and, by (4.17) and (4.9): $\mathbf{f_k}$

$$|(\widehat{\mathbf{f}}_s)_k(z)|\leq|(\widehat{\mathbf{f}}_k)_s(z)|\leq C_1(n)\left\|\widehat{f}_k\right\|_{C^0(\mathbb{R}^n)}\leq C_1(n)\mathtt{C}_\mathtt{F}(n,\ell)\frac{|f|_{C^{\lfloor\ell\rfloor}(\mathbb{R}^n)}}{|k|_\infty^{\lfloor\ell\rfloor}},\quad k\neq0,\ |k|_1\leq1/s, \tag{4.18}$$

and

$$|(\widehat{\mathbf{f}}_s)_0(z)|\leq C_1(n)\left\|\widehat{f}_0\right\|_{C^0(\mathbb{R}^n)}\leq C_1(n)\left\|f\right\|_{C^0(\mathbb{R}^n)}. \tag{4.19}$$

As for the weighted Fourier norm of $\mathbf{f}_s$, we finally get:

$$\begin{aligned}||\mathbf{f}_s||_{s,s}&=\sup_{|\mathrm{Im}\,z|_2\leq s}\sum_{k\in\mathbb{Z}^n}\left|(\widehat{\mathbf{f}}_s)_k(z)\right|e^{s|k|_1}\\&\leq C_1(n)\left\|f\right\|_{C^0(\mathbb{R}^n)}+\sum_{\substack{k\in\mathbb{Z}^n\setminus\{0\}:\\|k|_1\leq1/s}}eC_1(n)\mathtt{C}_\mathtt{F}(n,\ell)\frac{|f|_{C^{\lfloor\ell\rfloor}(\mathbb{R}^n)}}{|k|_\infty^{\lfloor\ell\rfloor}}\leq\mathtt{C}_B(n,\ell)|f|_{C^\ell(\mathbb{R}^n)},\end{aligned}$$

where

$$\mathtt{C}_B(n,\ell):=C_1(n)\left(1+e\mathtt{C}_\mathtt{F}(n,\ell)\sum_{k\in\mathbb{Z}^n}\frac{1}{|k|_\infty^{\lfloor\ell\rfloor}}\right)<+\infty. \tag{4.20}$$

4.3. **The main result with an application to normal forms.**

4.3.1. *Main result.* Gathering together the elements of the previous section, we get the following result.

Lemma 4.1 (Analytic smoothing)**.** *Fix an integer $n \geq 1$, $R > 0$ and $s \in]0,1]$. Let f be a C^ℓ function on $B_\infty(0,2R)\times\mathbb{T}^n$. Then, for any integer $p \leq \lfloor \ell \rfloor$ there exist two constants $\mathtt{C}_A(R,\ell,n), \mathtt{C}_B(R,\ell,n)$ and an analytic function $\mathtt{f}_s$ on the set $\mathbb{A}^n_s$ satisfying*

$$\|f-\mathtt{f}_s\|_{C^p(B_\infty(0,R)\times\mathbb{T}^n)} \leq \ \mathtt{C}_A(R,\ell,n)\ s^{\ell-p}|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)} \tag{4.21}$$

and

$$\|\mathtt{f}_s\|_s \leq \mathtt{C}_B(R,\ell,n)|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)}. \tag{4.22}$$

Moreover, $\mathtt{f}_s$ is a trigonometric polynomial in the angular variables.

Proof. Fix a function $\chi \in C^\infty(\mathbb{R}^n)$, with values in $[0,1]$, equal to 1 on the ball $B_\infty(0,R)$ and with support in $B_\infty(0,2R)$. Then the product $\overline{f} := \chi f$ is C^∞ on $\mathbb{A}^n$, has compact support in $B_\infty(0,2R)\times\mathbb{T}^n$ and coincides with f on $B_\infty(0,R)\times\mathbb{T}^n$. Moreover

$$|\overline{f}|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)} \leq \mathtt{C}_B|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)}$$

where $\mathtt{C}_B = C|\chi|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)}$ and C is a universal constant. By the Jackson-Moser-Zehnder theorem applied to $\overline{f}$, there is an analytic function $\bar{\mathtt{f}}_s$ on $\mathbb{A}^n_s$ satisfying

$$\left|\partial^\alpha \bar{\mathtt{f}}_s(I,\theta) - \sum_{\substack{\beta\in\mathbb{N}^{2n}:\\ |\beta|\leq\lfloor\ell\rfloor-|\alpha|}} \partial^{\alpha+\beta}\bar{f}(\mathrm{Re}(I,\theta))\frac{(\mathrm{Im}(I,\theta))^\beta}{\beta!}\right| \leq \mathtt{C}_{\mathtt{J}} s^{\ell-|\alpha|}|\bar{f}|_{C^\ell(\mathbb{A}^n)}, \tag{4.23}$$

so that for any $p \leq \lfloor \ell \rfloor$:

$$\left\|\bar{f}-\bar{\mathtt{f}}_s\right\|_{C^p(\mathbb{A}^n)} \leq \ \mathtt{C}_{\mathtt{J}} s^{\ell-p}|\bar{f}|_{C^\ell(\mathbb{A}^n)}. \tag{4.24}$$

As a consequence, taking the form of χ into account, one gets

$$\|f-\mathtt{f}_s\ \|_{C^p(B_\infty(0,R)\times\mathbb{T}^n)} \leq \mathtt{C}_B\mathtt{C}_{\mathtt{J}} s^{\ell-p}|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)}. \tag{4.25}$$

Setting $\mathtt{C}_A := \mathtt{C}_B\mathtt{C}_{\mathtt{J}}$ and, since the analyticity width ρ of the integrable part h is greater than s, the bound (4.21) follows. The proof of (4.22) is an immediate consequence of the previous paragraphs. □

4.3.2. *An easy way to derive normal forms for Hölder functions from analytic ones.* Let

$$H(I,\theta) := h(I) + f(I,\theta) \tag{4.26}$$

be C^ℓ on $B_\infty(0,2R)\times\mathbb{T}^n$. Given $s \in]0,1]$, let $\mathbf{H}_s$ be the s-smoothed analytic function given by Lemma 4.1 applied to the function H. By classical constructions (alluded to in the introduction and which will be recalled in the following), there exist (close to identity) symplectic analytic local diffeomorphisms Φ defined on domains $D \subset \mathbb{A}^n$ which bring $\mathbf{H}_s = \mathbf{h}_s + \mathbf{f}_s$ to the normal forms $\mathbf{H}_s \circ \Phi : D \to \mathbb{R}$:

$$\mathbf{H}_s \circ \Phi = \mathbf{h}_s + g + f^* \tag{4.27}$$

where $\mathbf{h}_s$ is nothing else than the smoothed initial integrable Hamiltonian, g is a resonant part which controls the fast drift in certain directions and f^* is a very small remainder – all these functions being analytic on D. The keypoint in our subsequent constructions is the following very simple equality

$$H \circ \Phi = \mathbf{H}_s \circ \Phi + (H - \mathbf{H}_s)\circ\Phi = \mathbf{h}_s + g + \big[f^* + (H-\mathbf{H}_s)\circ\Phi\big]. \tag{4.28}$$

This is a normal form for H, obtained by composition of H with an *analytic* diffeomorphism, in which the first two terms are analytic on D and only the last one is C^ℓ. So $H \circ \Phi$ has the same structure and dynamical interpretation as $\mathbf{H}_s \circ \Phi$, *provided that the C^ℓ size of the additional remainder $(H-\mathbf{H}_s)\circ\Phi$ is of the same order as the size of the initial remainder f^**. This issue strongly depends on the analytic smoothing method in use,

we will show in the sequel that the Jackson-Moser-Zehnder method is relevant for our purposes. Our study will be even easier since we assume from the beginning that the integrable part h is analytic.

It turns out that the same smoothing method - and the same simple way to get a normal form from an analytical one - are also relevant in many other functional classes, the main ones being the Gevrey classes already used in [30], but also other ultradifferentiable ones. This will be developed in a further work.

5. Stability estimates in the convex case with the patchwork method

We prove here the tuned estimates of stability of Theorem 1.2. In order to obtain a global result, we shall proceed by steps and analyze stability both in the completely non-resonant and in the resonant blocks respectively.

To our purpose, we will make use of Pöschel's resonant normal form and resonant patchwork for analytic Hamiltonians (see for convenience Appendix B) and we will suitably choose the dipendence of the involved quantities on the perturbative parameter ε. This allows for a flexible framework in which the role of thresholds of applicability, time of stability and radius of confinement appears clearly. In the following section, we start by introducing and deduce some necessary conditions that the tuning parameters must satisfy.

5.1. Initializing the tuning parameters.

We start by see that, by monotonicity of the Fourier norm w.r.t. the action variables and (4.22) we immediately get

$$||\mathtt{f}_s||_{r,s} \leq \mathtt{C}_B(R,\ell,n)|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)} =: \mathtt{C}_B(R,\ell,n)\varepsilon =: \epsilon, \tag{5.1}$$

which will be the "analytic perturbative" parameter of Poschel's normal form.

Let $\mathtt{H}_s := h(I) + \mathtt{f}_s$ be the analytic hamiltonian defined on $\mathcal{B}_{s,s}$ which was introduced in Lemma 4.1 and let Λ be sublattice of $\mathbb{Z}^n_K$. By Pöschel's Lemma B.1 applied in this framework, with $\varrho' \rightsquigarrow r, \varrho \rightsquigarrow s, \sigma \rightsquigarrow s$, given a domain D_Λ which is (α, K)-nonresonant modulo Λ, with [17] $D_{\Lambda,r} \times \mathbb{T}^n_s \subset \mathcal{B}^2_{r,s}$, if for some constant $\xi > 1$ the conditions

$$\epsilon \leq \frac{1}{256}\frac{\alpha_\Lambda r}{\xi K} \ , \quad r \leq \min\left(\frac{\alpha_\Lambda}{2MK}, s\right) \ , \quad Ks \geq 6 \tag{5.2}$$

are satisfied, then $\mathtt{H}_s$ is transformed into the following resonant normal form

$$\mathtt{H}_s \circ \Psi_\Lambda = h(I) + \mathtt{g} + \mathtt{f}^*_s \ , \quad \{h, \mathtt{g}\} = 0, \quad \Psi_\Lambda : \mathcal{D}_{\Lambda,r/2,s/6} \longrightarrow \mathcal{D}_{\Lambda,r,s} \tag{5.3}$$

where Ψ_Λ is a symplectic diffeomorphism. In particular the resonant and non-resonant part satisfy, respectively,

$$||\mathtt{g} - \mathtt{g_0}||_{r/2,s/6} \leq \frac{\epsilon}{4\xi} \ , \quad ||\mathtt{f}^*_s||_{r/2,s/6} \leq e^{-\frac{Ks}{6}}\epsilon \tag{5.4}$$

where $\mathtt{g_0} := P_\Lambda P_K \mathtt{f}_s$.
Moreover, by Pöschel's Covering Lemma B.2, each resonant block D_Λ is (α_Λ, K)-nonresonant modulo Λ, with

$$\alpha_\Lambda = pM\gamma_0 A^{d-n}\frac{K^{d-n+1}}{|\Lambda|} \ , \tag{5.5}$$

where p, γ_0 are free parameters, A is a constant satisfying some appropriate lower bound and $|\Lambda|$ denotes the fundamental volume of the resonant lattice.

Let us now introduce the ε-dependence of the analyticity widths and of the ultraviolet cut-off K; we shall carefully analyze the constraints and bounds given by Poschel's normal form in terms of such choices.
Given positive parameters a, b, c and $\mathtt{r}_*$ we set

$$r := \mathtt{r}_*\frac{\varepsilon^c|\ln(\varepsilon^{6+b})|^{d-n}}{|\Lambda|}, \quad s := \varepsilon^a, \quad K := \varepsilon^{-a}|\ln(\varepsilon^{6+b})| \, . \tag{5.6}$$

[17] recall notations (3.6)

Remark 5.1. In order to have a polynomial remainder, the quantity Ks must be of logarithmic order, which explains the presence of $|\ln(\varepsilon^{6+b})|$ in the definition of K. On the other hand s must be of order epsilon to have a good control on the analytically smoothed remainder (see expression (4.21)). Of course one could define

$$r := \mathtt{r}_* \frac{\varepsilon^c}{|\Lambda|}, \quad s := \varepsilon^a |\ln(\varepsilon^{6+b})|, \quad K := \varepsilon^{-a}. \tag{5.7}$$

instead. This would result in achieving the typical exponents $1/2n, (\ell-1)/2n$ for the radius and time of stability, at the cost of losing the mutual dependence of time, threshold and radius through a.

- **Condition over r and Ks.** Since by (5.2) r must satisfy $r \leq \frac{\alpha_\Lambda}{2MK}$, by (5.5) and the definition of K in (5.6) we have

$$r \leq \frac{pA^{d-n}}{2}\frac{\gamma_0 K^{d-n}}{|\Lambda|} = \frac{pA^{d-n}}{2}\frac{\gamma_0(\varepsilon^{-a}|\ln(\varepsilon^{6+b})|)^{d-n}}{|\Lambda|} = \frac{p\gamma_0 A^{d-n}}{2|\Lambda|}\varepsilon^{a(n-d)}|\ln(\varepsilon^{6+b})|^{d-n}, \tag{5.8}$$

which together with the definition of r gives

$$\mathtt{r}_*\varepsilon^c \leq \frac{p\gamma_0 A^{d-n}}{2}\varepsilon^{a(n-d)}, \tag{5.9}$$

which is automatically satisfied if we set

$$c \geq a(n-d), \quad \mathtt{r}_* \leq \min\left\{\frac{pA^{d-n}\gamma_0}{2}, 1, \frac{R}{2}\right\}, \tag{5.10}$$

where the condition $\mathtt{r}_* \leq R/2$ will be used in the sequel. Moreover, since $|\Lambda| \geq 1$, if

$$\varepsilon \leq e^{-1}, \tag{5.11}$$

by the choices in (5.6) together with (5.10), we have automatically that $r \leq s$, $s \leq 1$ and $Ks \geq 6$.

- **Condition over ε.** By (5.6) and the bound in (5.2)

$$\epsilon = \mathtt{C}_B\varepsilon \leq \frac{1}{256}\frac{\alpha_\Lambda r}{\xi K}$$

which together with the form of r and α_Λ gives

$$|\Lambda|^2\varepsilon^{1-a(n-d)-c}|\ln(\varepsilon^{6+b})|^{2(n-d)} \leq \mathtt{C}_S, \tag{5.12}$$

where we set

$$\mathtt{C}_S = \mathtt{C}_S(d,n,R,\ell,p,\xi) := \frac{pMA^{d-n}\gamma_0\mathtt{r}_*}{256\ \mathtt{C}_B\xi} \tag{5.13}$$

and we write $\mathtt{C}_B = \mathtt{C}_B(R,\ell,n)$ not to burden notations.

The exponent of ε in (5.12) must be positive if we want such expression to make sense, so we must have

$$1 - a(n-d) - c > 0 \tag{5.14}$$

which, together with the first constraint in (5.10), yields

$$0 \leq a < \frac{1}{2(n-d)}. \tag{5.15}$$

- **Diffeomorphisms and remainders.** Finally, the size of the transformation Ψ_Λ is given by Lemma B.1 and reads

$$|\Pi_I\Psi_\Lambda - I|_2 \leq 4\frac{K\epsilon}{\alpha_\Lambda} \overset{(5.2)}{\leq} \frac{r}{64\xi} \overset{r}{=} \frac{\mathtt{r}_*}{64\xi}\frac{\varepsilon^c|\ln(\varepsilon^{6+b})|^{d-n}}{|\Lambda|}, \tag{5.16}$$

while (5.4) yields

$$||\mathtt{f}_s^*||_{r/2,s/6} \leq \exp\left(-\frac{\varepsilon^a\varepsilon^{-a}|\ln(\varepsilon^{6+b})|}{6}\right)\epsilon \overset{(5.1)}{=} \mathtt{C}_B(R,\ell,n)\exp\left(\frac{\ln(\varepsilon^{6+b})}{6}\right)\varepsilon = \mathtt{C}_B(R,\ell,n)\varepsilon^{2+b/6}, \tag{5.17}$$

which is in fact polynomial in ε and

$$||\mathtt{g} - \mathtt{g}_0||_{r/2,s/6} \leq \frac{\epsilon}{4\xi} = \mathtt{C}_B \frac{\varepsilon}{4\xi} \tag{5.18}$$

In conclusion, for the applicability of Pöschel's result, a, b, c and ε must meet the following conditions

$$\begin{gathered} 0 \leq a < \frac{1}{2(n-d)}, \qquad a(n-d) \leq c < 1 - a(n-d), \qquad b > 0 \\ |\Lambda|^2 \varepsilon^{1-a(n-d)-c} |\ln(\varepsilon^{6+b})|^{2(n-d)} \leq \mathtt{C}_S \qquad \varepsilon \leq e^{-1} \qquad \mathtt{r}_* \leq \min\left\{ \frac{pA^{d-n}\gamma_0}{2}, 1, \frac{R}{2} \right\} . \end{gathered} \tag{5.19}$$

Remark 5.2. We observe that, if $d = n$ and $\Lambda = \mathbb{Z}^n_K$, condition (5.19) yields $a \geq 0$ and $c \in [0, 1)$.

With the setup of this section, we are ready to state local results of stability, both in the resonant blocks and in the completely non-resonant one.

5.2. **Stability in the completely non-resonant domain.** In this framework we consider $\Lambda = \{0\}$ and the block $D_0 := D_\Lambda \cap B_\infty(0, \frac{R}{2})$ corresponding to completely non-resonant actions for the unperturbed hamiltonian.

Theorem 5.1 (Non-resonant Stability Estimates)**.** *Let $a \in \left(0, \dfrac{1}{2n}\right)$, $c \in [an, 1 - an)$ and $b = 6a\ell$. If the following thresholds hold*

$$\varepsilon^{1-an-c} |\ln \varepsilon^{6+b}|^{2n} \leq \mathtt{C}_S(0, n, R, \ell) , \quad \varepsilon \leq e^{-1} \tag{5.20}$$

where $\mathtt{C}_S$ is the constant defined in (5.13)*, then, there exist two constants $\mathtt{C}'_1 = \mathtt{C}'_1(n, \ell), \mathtt{C}'_2 = \mathtt{C}'_2(n)$ such that, for any time t satisfying*

$$|t| \leq \frac{\mathtt{C}'_1(n, \ell, h)}{\{[1 + a(\ell + n)] \ |\ln \varepsilon|\}^n \ \varepsilon^{1+a(\ell-1)-c}} , \tag{5.21}$$

any initial condition $I(0) \in D_0$ drifts as

$$|I(t) - I(0)|_2 \leq \mathtt{C}'_2(n, h) \frac{\varepsilon^c}{\{[1 + a(\ell + n)] \ |\ln \varepsilon|\}^n} . \tag{5.22}$$

Proof. We consider the setup of the previous section with $|\Lambda| = 1$, $d = 0$ and

$$\frac{b}{6} = a\ell , \tag{5.23}$$

where such choice for parameter b in the ultra-violet cut-off K (remember formula (5.6)) shall be justified in the sequel. By the smallness conditions in (5.20) (which correspond to (5.11) and (5.12) for $d = 0$) the smoothed hamiltonian $\mathtt{H}_s := h + \mathtt{f}_s$ can be put into normal form (5.3), which in this completely non-resonant case reads

$$\mathtt{H}_s \circ \Psi_0 = h + \mathtt{f}^*_s , \quad \Psi_0 := \Psi_\Lambda . \tag{5.24}$$

For the Hölder Hamiltonian $H = h + f + \mathtt{H}_s - \mathtt{H}_s$, we therefore get

$$H \circ \Psi_0 = \mathtt{H}_s \circ \Psi_0 + (H - \mathtt{H}_s) \circ \Psi_0 = h + \mathtt{f}^*_s + (f - \mathtt{f}_s) \circ \Psi_0 . \tag{5.25}$$

Let $D^{\mathbb{R}}_{0,r}$ be the real extension of width r around D_0, which is contained in $B_\infty(0, R)$ since $r \leq R/2$. The Faà Di Bruno formula, together with (4.21) and (5.6), insures that there exist constants $\mathtt{C}_A = \mathtt{C}_A(n, \ell), \mathtt{C}_\mathcal{F} = \mathtt{C}_\mathcal{F}(n, \ell)$ such that

$$\|(f - \mathtt{f}_s) \circ \Psi_0\|_{C^1(D^{\mathbb{R}}_{0,r/2} \times \mathbb{T}^n)} \leq \mathtt{C}_\mathcal{F} \|f - \mathtt{f}_s\|_{C^1(D^{\mathbb{R}}_{0,r} \times \mathbb{T}^n)} \leq \mathtt{C}_\mathcal{F} \mathtt{C}_A s^{\ell-1} |f|_{C^\ell(B_\infty(0,R) \times \mathbb{T}^n)} = \mathtt{C}_\mathcal{F} \mathtt{C}_A \varepsilon^{1+a(\ell-1)} . \tag{5.26}$$

Now, any initial condition $(I(0),\theta(0)) \in D_0 \times \mathbb{T}^n$ is mapped by Ψ_0 in $(\mathtt{I}(0),\vartheta(0)) \in D^{\mathbb{R}}_{0,\frac{r}{64\xi}} \times \mathbb{T}^n$ by (5.16). For any time t such that the normalized flow $\Phi^t_{H\circ\Psi_0} : (\mathtt{I}(0),\vartheta(0)) \longmapsto (\mathtt{I}(t),\vartheta(t))$ starting at $D^{\mathbb{R}}_{0,\frac{r}{64\xi}} \times \mathbb{T}^n$ does not exit from $D^{\mathbb{R}}_{0,r/2} \times \mathbb{T}^n$, the evolution of the normalized variables reads

$$
\begin{aligned}
|\mathtt{I}(t)-\mathtt{I}(0)|_\infty &\le \int_0^t \sup_{(\mathtt{I},\vartheta)\in D^{\mathbb{R}}_{0,\frac{r}{64\xi}}\times\mathbb{T}^n} \left(\left|(\partial_\vartheta \mathtt{f}^*_s)\circ\Phi^t_{H\circ\Psi_0}\right| + \left|\{\partial_\vartheta[(f-\mathtt{f}_s)\circ\Psi_0]\}\circ\Phi^t_{H\circ\Psi_0}\right| \right) dt \\
&\le \int_0^t \left(\sup_{(\mathtt{I},\vartheta)\in D^{\mathbb{R}}_{0,r/2}\times\mathbb{T}^n} |\partial_\vartheta \mathtt{f}^*_s| + \sup_{(\mathtt{I},\vartheta)\in D^{\mathbb{R}}_{0,r/2}\times\mathbb{T}^n} |\partial_\vartheta[(f-\mathtt{f}_s)\circ\Psi_0]| \right) dt \qquad (5.27)\\
&\le |t| \left[6\frac{||\mathtt{f}^*_s||_{r/2,s/6}}{es} + \|(f-\mathtt{f}_s)\circ\Psi_0\|_{C^1(D^{\mathbb{R}}_{0,r/2}\times\mathbb{T}^n)} \right] \\
&\le |t| \left[\frac{6\mathtt{C}_B}{e}\varepsilon^{2-a+b/6} + \mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{1+a(\ell-1)} \right] ,
\end{aligned}
$$

where the factor e at the denominator of the first term at the third row is an improvement of the standard Cauchy estimates which is proven in Lemma B.3 of ref. [39], whereas in the last inequality we used (5.17) and (5.26). By the appropriate choice (5.23) of[18] the parameter b (recall the definitions in (5.6)) and from inequality (5.27) we finally get

$$
|\mathtt{I}(t)-\mathtt{I}(0)|_2 \le |t| 2\mathtt{C}'_3\varepsilon^{1+a(\ell-1)} , \qquad (5.28)
$$

where $\mathtt{C}'_3 = \mathtt{C}'_3(n,\ell) = \sqrt{n} \times \max\left\{ \frac{6\mathtt{C}_B}{e}, \mathtt{C}_A\mathtt{C}_{\mathcal{F}} \right\}$.

Hence, for any time t satisfying

$$
|t| \le \frac{31r}{128\xi\mathtt{C}'_3}\frac{1}{\varepsilon^{1+a(\ell-1)}} \equiv \frac{31\mathtt{r}_*}{128\xi\mathtt{C}'_3}\frac{1}{\left|\ln\varepsilon^{6(1+a\ell)}\right|^n\varepsilon^{1+a(\ell-1)-c}} \qquad (5.29)
$$

the variation of the normalized actions is bounded by

$$
|\mathtt{I}(t)-\mathtt{I}(0)|_2 \le \frac{31}{64\xi}r . \qquad (5.30)
$$

Then, since $\mathtt{I}(0) \in D^{\mathbb{R}}_{0,\frac{r}{64\xi}}$ and $\xi > 1$ by hypothesis, one has $\mathtt{I}(t) \in D^{\mathbb{R}}_{0,r/2}$, so that $\mathtt{I}(t)$ is still in the domain of Ψ_0 and, using also (5.16), the maximal variation of the original action $I(0) \in D_0$ reads

$$
|I(t)-I(0)|_2 \le |I(t)-\mathtt{I}(t)|_2 + |\mathtt{I}(t)-\mathtt{I}(0)|_2 + |\mathtt{I}(0)-I(0)|_2 = \frac{r}{64\xi} + \frac{31}{64\xi}r + \frac{r}{64\xi} \le \frac{33}{64\xi}r \qquad (5.31)
$$

From the expression of r and setting

$$
\mathtt{C}'_1 := \frac{31\mathtt{r}_*}{6^n 128\xi\mathtt{C}'_3} \qquad \mathtt{C}'_2 := \frac{33\mathtt{r}_*}{6^n 64\xi},
$$

we get the claimed bound.

□

5.3. **Stability in resonant domains.** Let us now consider a domain D_Λ corresponding to a non-trivial resonant maximal lattice Λ of dimension $d \in \{1,...,n-1\}$. The setup is the one of Pöschel's Normal form B.1 and Covering Lemma B.2 with

$$
\mathtt{b} \le \frac{\mathtt{r}_* A^{n-d}}{64\gamma_0\mu\xi} , \quad \xi > \frac{(1+\sqrt{2})M}{4\mu} . \qquad (5.32)
$$

[18] One might wonder why we did not choose $1+b/6 = a\ell$, since with such choice the two summands in the r.h.s of (5.27) would have been of the same order in ε. Actually, since b must be positive, $1+b/6 = a\ell$ implies the spurious condition $a > 1/\ell$, which we do not want.

Let $\hat{\omega} := \sup_{B_\infty(0,R)} |\omega(I)|_\infty$ and $\mathtt{C}_0 = \mathtt{C}_0(d,n,\ell,\xi)$

$$\mathtt{C}_0 = \min\left\{\mathtt{C}_S(d,n,\ell,\xi), \frac{4\xi\hat{\omega}\mathtt{r}_*}{\mathtt{C}_B(4\xi+1)}, \frac{n\hat{\omega}\mathtt{C}_A\mathtt{C}_\mathcal{F}\xi\mathtt{r}_*}{6\mathtt{C}_B^2(4\xi+1)}, \frac{\mathtt{C}_\mathcal{F}\mathtt{C}_A}{6\mathtt{C}_B}\right\}.$$

Theorem 5.2 (Resonant Stability Estimates)**.** *Assume that $D_\Lambda \cap B_\infty(0,\frac{R}{2}) \neq \varnothing$. For any $a \in \left(0, \frac{1}{2(n-d)}\right)$ and for any couple of parameters b, c satisfying $b/6 = a(\ell+n)$ and $c = a(n-d)$, if*

$$|\Lambda|^2 \varepsilon^{1-a(n-d)-c} |\ln(\varepsilon^{6+b})|^{2(n-d)} \leq \mathtt{C}_0 \ , \quad \varepsilon \leq e^{-1} \tag{5.33}$$

then there exists two positive explicit constants $\mathtt{C}_1 = \mathtt{C}_1(n,d,\ell)$ and $\mathtt{C}_2 = \mathtt{C}_2(n,d)$ such that any initial datum $I(0) \in D_\Lambda \cap B_\infty(0,\frac{R}{2})$ varies as

$$|I(t) - I(0)|_2 \leq \mathtt{C}_2 \frac{\varepsilon^c}{|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}} \tag{5.34}$$

for any time t satisfying

$$|t| \leq t^* = \frac{\mathtt{C}_1}{|\Lambda|^2 |\ln \varepsilon^{6+b}|^{2(n-d)}\, \varepsilon^{1+a(\ell-1)-2c}} \ . \tag{5.35}$$

Proof. By (5.11) and (5.12), if

$$|\Lambda|^2 \varepsilon^{1-a(n-d)-c} |\ln(\varepsilon^{6+b})|^{2(n-d)} \leq \mathtt{C}_S \ , \quad \varepsilon \leq e^{-1} \tag{5.36}$$

the smoothed hamiltonian $\mathtt{H}_s$ is brought into resonant normal form (5.3) and estimates (5.4) hold. As in the completely non-resonant case, we write

$$H \circ \Psi_\Lambda = \mathtt{H}_s \circ \Psi_\Lambda + (H - \mathtt{H}_s) \circ \Psi_\Lambda = h + \mathtt{g} + \mathtt{f}_s^* + (f - \mathtt{f}_s) \circ \Psi_\Lambda \tag{5.37}$$

and we have

$$\|(f-\mathtt{f}_s)\circ\Psi\|_{C^1(D^\mathbb{R}_{\Lambda,r/2}\times\mathbb{T}^n)} \leq \mathtt{C}_\mathcal{F} \|f - \mathtt{f}_s\|_{C^1(D^\mathbb{R}_{\Lambda,r}\times\mathbb{T}^n)} \leq \mathtt{C}_\mathcal{F}\mathtt{C}_A s^{\ell-1} |f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)} = \mathtt{C}_\mathcal{F}\mathtt{C}_A \varepsilon^{1+a(\ell-1)} \,, \tag{5.38}$$

for some positive constants $\mathtt{C}_A, \mathtt{C}_\mathcal{F}$. By (5.16), the image $\mathtt{I}(0)$ of any initial condition $I(0) \in D_\Lambda$ by Ψ_Λ does not exit from $D^\mathbb{R}_{\Lambda,\frac{r}{64\xi}}$. Let t_e be the time of escape of the normalized flow $\Phi^t_{H\circ\Psi_\Lambda} : (\mathtt{I}(0), \vartheta(0)) \longmapsto (\mathtt{I}(t)\vartheta(t))$ from $D^\mathbb{R}_{\Lambda,r/4} \times \mathbb{T}^n$. For any time $t \leq t_e$ the Taylor formula and the uniform strict convexity of h (recall definition 1.2) yield

$$|h(\mathtt{I}(t)) - h(\mathtt{I}(0))| + |\omega(\mathtt{I}(0)) \cdot (\mathtt{I}(t) - \mathtt{I}(0))| \geq \frac{\mu}{2\sqrt{n}} |\mathtt{I}(t) - \mathtt{I}(0)|_2{}^2 \ . \tag{5.39}$$

Set now $\mathtt{R} := \mathtt{f}_s^* + (f - \mathtt{f}_s) \circ \Psi_\Lambda$, by (5.37) and energy conservation we have

$$\begin{aligned}
|h(\mathtt{I}(t)) - h(\mathtt{I}(0))| \leq & |\mathtt{g}(\mathtt{I}(t),\theta(t)) - \mathtt{g}(\mathtt{I}(0),\theta(0))| + |\mathtt{R}(\mathtt{I}(t),\theta(t)) - \mathtt{R}(\mathtt{I}(0),\theta(0))| \\
\leq & \int_0^t \left| \frac{d}{d\tau} \mathtt{g}(\mathtt{I}(\tau),\theta(\tau)) \right| d\tau + \int_0^t \left| \frac{d}{d\tau} \mathtt{R}(\mathtt{I}(\tau),\theta(\tau)) \right| d\tau \\
= & \int_0^t \left| \{H\circ\Psi_\Lambda, \mathtt{g}\} \circ \Phi^\tau_{H\circ\Psi_\Lambda} \right| d\tau + \int_0^t \left| \{H\circ\Psi_\Lambda, \mathtt{R}\} \circ \Phi^\tau_{H\circ\Psi_\Lambda} \right| d\tau \\
\leq & 2\int_0^t \left| \{\mathtt{R}, \mathtt{g}\} \circ \Phi^\tau_{H\circ\Psi_\Lambda} \right| d\tau + \int_0^t \left| \{h, \mathtt{R}\} \circ \Phi^\tau_{H\circ\Psi_\Lambda} \right| d\tau \\
\leq & 2 \|\{\mathtt{R},\mathtt{g}\}\|_{C^0(D^\mathbb{R}_{\Lambda,r/4}\times\mathbb{T}^n)} |t| + \|\{h,\mathtt{R}\}\|_{C^0(D^\mathbb{R}_{\Lambda,r/4}\times\mathbb{T}^n)} |t| \ .
\end{aligned} \tag{5.40}$$

The standard Cauchy estimates, together with expressions (5.1), (5.17) and (5.38), yield the following estimates on the first summand at the r.h.s.:

$$\begin{aligned}\|\{\mathtt{R},\mathtt{g}\}\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)} \leq& \|\{\mathtt{f}^*_s,\mathtt{g}\}\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)} + \|\{(f-\mathtt{f}_s)\circ\Psi_\Lambda,\mathtt{g}\}\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)}\\ \leq& \frac{||\mathtt{f}^*_s||_{r/2,s/6}\ ||\mathtt{g}||_{r/2,s/6}}{r/4\times s/6} + 2n\mathtt{C}_{\mathcal{F}}\left\|f-\mathtt{f}_s\right\|_{C^1(D^{\mathbb{R}}_{\Lambda,r}\times\mathbb{T}^n)}\frac{||\mathtt{g}||_{r/2,s/6}}{r/4}\\ \leq& \left(\frac{24}{rs}\mathtt{C}_B\varepsilon^{2+b/6} + 8n\frac{\mathtt{C}_{\mathcal{F}}\mathtt{C}_A}{r}\varepsilon^{1+a(\ell-1)}\right)\left(||\mathtt{g}-\mathbf{g_0}||_{r/2,s/6}+||\mathbf{g_0}||_{r/2,s/6}\right)\\ \leq& \left(\frac{24}{rs}\mathtt{C}_B\varepsilon^{2+b/6} + 8n\frac{\mathtt{C}_{\mathcal{F}}\mathtt{C}_A}{r}\varepsilon^{1+a(\ell-1)}\right)\left(\frac{\mathtt{C}_B}{4\xi}\varepsilon+\mathtt{C}_B\varepsilon\right)\\ =& \mathtt{C}_B\frac{4\xi+1}{4\xi}\left(\frac{24}{rs}\mathtt{C}_B\varepsilon^{3+b/6} + 8n\frac{\mathtt{C}_{\mathcal{F}}\mathtt{C}_A}{r}\varepsilon^{2+a(\ell-1)}\right)\ .\end{aligned} \tag{5.41}$$

Substititung the dependence (5.6) of r and s on ε in the above expression finally yields

$$\|\{\mathtt{R},\mathtt{g}\}\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)} \leq \mathtt{C}_B\frac{4\xi+1}{4\xi\mathtt{r}_*}|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}\left(24\mathtt{C}_B\varepsilon^{3-a+b/6-c} + 8n\mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{2+a(\ell-1)-c}\right)\ . \tag{5.42}$$

On the other hand, setting $\hat{\omega} := \sup_{B_\infty(0,R/2)_r}|\omega(I)|_\infty$, the second Poisson bracket in (5.40) can be estimated as

$$\begin{aligned}\|\{h,\mathtt{R}\}\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)} \leq& n\hat{\omega}\left(\left\|\frac{\partial\mathtt{f}^*_s}{\partial\vartheta}\right\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)} + \left\|\frac{\partial(f-\mathtt{f}_s)\circ\Psi}{\partial\vartheta}\right\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)}\right)\\ \leq& n\hat{\omega}\left(\frac{6\,||\mathtt{f}^*_s|\,|_{r/2,s/6}}{es} + \mathtt{C}_{\mathcal{F}}\left\|f-\mathtt{f}_s\right\|_{C^1(D^{\mathbb{R}}_{\Lambda,r}\times\mathbb{T}^n)}\right)\\ \leq& n\hat{\omega}\left(\frac{6\mathtt{C}_B}{e}\varepsilon^{2-a+b/6} + \mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{1+a(\ell-1)}\right)\ .\end{aligned} \tag{5.43}$$

By the hypothesis on a,b,c and (5.33), the summands at the r.h.s. of (5.42)-(5.43) are in fact bounded by $n\hat{\omega}\mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{1+a(\ell-1)}$ and, setting

$$\mathtt{C}_3(n,\ell) = 8n\hat{\omega}\mathtt{C}_{\mathcal{F}}\mathtt{C}_A\ ,$$

we obtain

$$|h(\mathtt{I}(t))-h(\mathtt{I}(0))| \leq \frac{\mathtt{C}_3}{2}|t|\varepsilon^{1+a(\ell-1)}\ . \tag{5.44}$$

Concerning the second term of (5.39), if I^* denotes the resonant action closest to $\mathtt{I}(0)$ we have

$$|\omega(\mathtt{I}(0))\cdot(\mathtt{I}(t)-\mathtt{I}(0))| \leq |\omega(I^*)\cdot(\mathtt{I}(t)-\mathtt{I}(0))| + |(\omega(\mathtt{I}(0))-\omega(I^*))\cdot(\mathtt{I}(t)-\mathtt{I}(0))|\ . \tag{5.45}$$

By Pöschel's Covering Lemma B.2 and estimate (5.16) on the size of the normal form transformation we also have

$$|\mathtt{I}(0)-I^*|_2 \leq |\mathtt{I}(0)-I(0)|_2 + |I(0)-I^*|_2 \leq \frac{r}{64\xi}+\delta_\Lambda\ ,$$

so that, by Taylor's formula, taking the dependences on ε into account and the bound (1.6) together with the definition of δ_Λ, the second summand at the r.h.s of (5.45) reads

(5.46)

$$\begin{aligned}|(\omega(\mathtt{I}(0))-\omega(I^*))\cdot(\mathtt{I}(t)-\mathtt{I}(0))| \leq& M\left(\frac{\mathtt{r}_*\varepsilon^c}{64\xi|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}} + \mathtt{b}\mu\gamma_0 A^{d-n}\frac{\varepsilon^{a(n-d)}}{|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}}\right)|\mathtt{I}(t)-\mathtt{I}(0)|_2\\ =& M\left(\frac{\mathtt{r}_*}{64\xi}+\mathtt{b}\mu\gamma_0 A^{d-n}\right)\frac{\varepsilon^{a(n-d)}}{|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}}|\mathtt{I}(t)-\mathtt{I}(0)|_2\ .\end{aligned}$$

where the last equality comes from the choice $c=a(n-d)$.
By definition of Hamiltonian vector field and the fundamental theorem of Calculus one has

$$|\omega(I^*)\cdot(\mathtt{I}(t)-\mathtt{I}(0))| \leq \|\{h,\mathtt{R}\}\|_{C^0(D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n)}\,|t| \tag{5.47}$$

and the same estimate as in (5.44) applies.

Plugging (5.44) and (5.46) into (5.39) we finally obtain

$$\frac{\mu}{2\sqrt{n}}|\mathtt{I}(t)-\mathtt{I}(0)|_2{}^2 - M\left(\frac{\mathtt{r}_*}{64\xi}+\mathtt{b}\mu\gamma_0 A^{d-n}\right)\frac{\varepsilon^{a(n-d)}}{|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}}|\mathtt{I}(t)-\mathtt{I}(0)|_2 - \mathtt{C}_3\,\varepsilon^{1+a(\ell-1)}|t| \le 0\ .$$

which gives

$$\begin{aligned}|\mathtt{I}(t)-\mathtt{I}(0)|_2 \le & \mathtt{c}(\xi,\mathtt{b})\frac{\varepsilon^{a(n-d)}}{|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}} \\ & + \left\{\left[\mathtt{c}(\xi,\mathtt{b})\frac{\varepsilon^{a(n-d)}}{|\Lambda||\ln(\varepsilon^{6+b})|^{n-d}}\right]^2 + \frac{3\mathtt{C}_3\sqrt{n}}{\mu}\,\varepsilon^{1+a(\ell-1)}|t|\right\}^{1/2}\ .\end{aligned} \tag{5.48}$$

where we have set

$$\mathtt{c}(\xi,\mathtt{b}) := \frac{M}{\mu}\left(\frac{\mathtt{r}_*}{64\xi}+\mathtt{b}\mu\gamma_0 A^{d-n}\right). \tag{5.49}$$

Hence, over a time

$$|t| \le \frac{\mathtt{c}(\xi,\mathtt{b})^2\mu}{3\mathtt{C}_3\sqrt{n}}\frac{\varepsilon^{2a(n-d)}}{|\Lambda|^2|\ln(\varepsilon^{6+b})|^{2(n-d)}\,\varepsilon^{1+a(\ell-1)}} =: t^* \tag{5.50}$$

the variation of the action variables is bounded by [19]

$$|\mathtt{I}(t)-\mathtt{I}(0)|_2 \le (1+\sqrt{2})\mathtt{c}(\xi,\mathtt{b})\frac{\varepsilon^{a(n-d)}}{|\Lambda||\ln\varepsilon^{6+b}|^{(n-d)}} = (1+\sqrt{2})\mathtt{c}(\xi,\mathtt{b})\frac{r}{\mathtt{r}_*} \tag{5.51}$$

where we have exploited the definition of r in (5.8) together with the choice $c = a(n-d)$.

Note that *a posteriori* any time satisfying (5.50) is in fact less than the time of escape from the domain $D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n$, as we supposed at the beginning of this reasonment.
In fact, since the parameter $\mathtt{b}$, representing the tuning on the width of the resonant blocks in Covering Lemma B.2, and ξ satisfy the constraint in (5.32), estimate (5.51) finally reads

$$|\mathtt{I}(t)-\mathtt{I}(0)|_2 \le \frac{(1+\sqrt{2})M}{32\mu\xi}r \le \frac{r}{8}\ , \tag{5.52}$$

so that the bootstrap argument is self-consistent and the normalized actions starting from $D^{\mathbb{R}}_{\Lambda,\frac{r}{64\xi}}\times\mathbb{T}^n$ stay inside the domain $D^{\mathbb{R}}_{\Lambda,r/4}\times\mathbb{T}^n$ up to a time $t^* \le t_e$. The variation of the initial action coordinates is obtained by summing to (5.52) the size (5.16) of the normal form. By the second relation in (5.9), moreover, the initial actions do not get out of the ball $B_\infty(0,R)$ where the initial hamiltonian is defined. The claim is therefore proven with

$$\mathtt{C}_1 := \frac{\mathtt{c}^2(\xi,b)\mu}{3\mathtt{C}_3\sqrt{n}}\quad,\qquad \mathtt{C}_2 := \frac{1}{32\xi}\left(\frac{(1+\sqrt{2})M}{\mu}+1\right)\ . \tag{5.53}$$

□

5.4. **Stability in the completely resonant domain.** As it was already remarked by Pöschel (see Theorems 1 and 1* of ref. [39]), the behaviour in the resonant domain corresponding to $\Lambda = \mathbb{Z}^n_K$ is rather different from that in the other blocks, in the sense that, as we shall see, one can only insure a drift of order one in the action variables, still over an infinite time. Confinement of the actions (i.e. a drift of order ε^a for some $a > 0$) can be obtained only at the expense of worse exponents of stability in the remaining blocks.

[19] Since $x+\sqrt{x^2+y|t|} \le (1+\sqrt{2})x$ for any $|t| \le x^2/y$ with $x, y > 0$.

Theorem 5.3. *Assume that* $D_{\mathbb{Z}^n_K} \cap B_\infty(0, \frac{R}{2}) \neq \varnothing$. *If*

$$\varepsilon \leq \frac{(M\mathfrak{b}\gamma_0)^2}{4\mu} \tag{5.54}$$

then, for any initial condition $I(0) \in D_{\mathbb{Z}^n_K} \cap B_\infty(0, \frac{R}{2})$, *one has*

$$|I(t) - I(0)|_2 \leq (1+\sqrt{2})M\mathfrak{b}\gamma_0 \tag{5.55}$$

over an infinite time.

Proof. Let $I(0)$ belong to $D_{\mathbb{Z}^n_K}$. Exploiting the convexity of the integrable part we have, as in (5.39),

$$|h(I(t)) - h(I(0))| + |\omega(I(0)) \cdot (I(t) - I(0))| \geq \frac{\mu}{2\sqrt{n}} |I(t) - I(0)|_2{}^2 \ . \tag{5.56}$$

By the conservation of energy we have $|h(I(t)) - h(I(0))| \leq 2\varepsilon$ and, by making use of the same arguments as in the proof of Lemma 5.2 and Lemma B.2 with $d = n$, we get

$$|I(t) - I(0)|_2 \leq M\mathfrak{b}\gamma_0 + \left\{ (M\mathfrak{b}\gamma_0)^2 + \ 4\mu\varepsilon \right\}^{1/2} \ . \tag{5.57}$$

Observe that such expression does not depend on time since in the completely resonant domain corresponding to $\Lambda = \mathbb{Z}^n_K$ one has $\omega(I^*) = 0$[20]. By threshold (5.54) we finally obtain

$$|I(t) - I(0)|_2 \leq (1+\sqrt{2})M\mathfrak{b}\gamma_0 \tag{5.58}$$

over any time t. □

Remark 5.3. One could impose $\gamma_0 \sim \varepsilon^\delta$ in order to confine the action variables even in the completely resonant case, but this would worsen the estimates of stability (thresholds, times and radii) in the other blocks.

Remark 5.4. Taking Lemma B.2 into account, for any initial condition $I(0)$ in $D_{\mathbb{Z}^n}$ the frequency is bounded as

$$|\omega(I(0))|_2 = |\omega(I(0)) - \omega(I^*)|_2 \leq M|I(0) - I^*|_2 \leq M\mathfrak{b}\mu\gamma_0 \ . \tag{5.59}$$

5.5. **Proof of Theorem 1.2.** Fixing b, c as in the hypotheses of Lemma 5.2, we see that Theorem 1.2 is a consequence of Theorems 5.1, 5.2 and 5.3. By (5.20) and (5.33) we see that the worst thresholds in ε are obtained in the non-resonant case, whereas comparing (5.22) and (5.34) the largest variation in the action variables is obtained in resonant blocks corresponding to resonance lattices of maximal order $d = n-1$ and fundamental volume $|\Lambda| \sim 1$. Finally, by closely looking at expression (5.21) and (5.35), one easily sees that the worst time of stability is obtained for resonances of maximal volume $|\Lambda| \sim K^d = \varepsilon^{ad}|\{[1 + a(\ell + n)]|\ln\varepsilon|\}^d$. Finally, Theorem 1.2 is proven once we set

$$\mathtt{C}_{\mathbf{E}} = \min\left\{ \min_{1\leq d\leq n-1} \mathtt{C}_0(d, n, \ell, p, \xi), e^{-\frac{1}{6}}, \frac{(M\mathfrak{b}\gamma_0)^2}{4\mu} \right\} \ , \quad \mathtt{C}_{\mathbf{T}} = \min\left\{\mathtt{C}_1, \mathtt{C}_1'\right\} \ , \quad \mathtt{C}_{\mathbf{I}} = \max\left\{ \mathtt{C}_2, \mathtt{C}_2', (1+\sqrt{2})M\mathfrak{b}\gamma_0 \right\} .$$

6. Stability estimates in the convex case with the periodic averaging method

The goal of this section is to prove Theorem 1.3. While in the previous section the stability around each possible resonance was carefully analyzed, here one only focuses on resonances corresponding to lattices Λ of dimension $n-1$, corresponding to periodic actions. In this way, one insures stability in the vicinity of these resonances with the help of perturbation theory and then obtains a global result of stabiliy with the help of Dirichlet's Theorem on simultaneous approximations. We refer to appendix B for the statement of a Normal Form Lemma in the vicinity of periodic actions due to Lochak-Neishtadt-Niederman [29] and for a statement of Dirichlet's Theorem. A more complete description of this strategy of proof can be found in [25] and [26].

We proceed as in the previous case: we tune the parameters, we get local extimates of stability and then we analyze stability in the whole phase space.

[20]Indeed, at the exact resonance I^* there exist n relations of the kind $k_i \cdot \omega(I^*) = 0$, where $i \in \{1, ..., n\}$ and the k_i's are independent vectors of $\mathbb{Z}^n_K$. This implies immediately that $\omega(I^*) = 0$.

6.1. **Initializing the tuning parameters.** By the definition of $\bar{\mathtt{f}}_s$ in the proof of Lemma 4.1, we have for $(I,\theta)\in B_\infty(0,R)\times\mathbb{T}^n$

(6.1)
$$|\mathtt{f}_s(I,\theta)|=|\bar{\mathtt{f}}_s(I,\theta)|\leq\int_{\mathbb{R}^n}\left|K\left(\frac{I}{s}-\xi,\frac{\theta}{s}-\eta\right)\bar{f}(s\xi,s\eta)\right|d\xi d\eta\leq\mathtt{C}_L|\bar{f}|_{C^0(B_\infty(0,R)\times\mathbb{T}^n)}\leq\mathtt{C}_L|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)},$$

since, as it is shown in Lemma A.1, there exists a uniform constant $\mathtt{C}_L=\mathtt{C}_L(n,\ell)$, such that

$$\int_{\mathbb{R}^n}|K(x,y)|dxdy\leq\mathtt{C}_L$$

hence

$$|\mathtt{f}_s(I,\theta)|\leq\mathtt{C}_L\varepsilon\,. \tag{6.2}$$

Let now $I_0\in B_\infty(0,\frac{R}{2})$ be an action correspondig to a T-periodic frequency $\omega:=\nabla h(I_0)$. Let $r\leq s$. We remark that $B^{\mathbb{R}}_{2,r}(I_0,r)=B_2(I_0,2r)$ and, following notations in (3.8), we set $\mathcal{B}_{r,s}(I_0,r):=B_{2,r}(I_0,r)\times\mathbb{T}^n_s$. Without loss of generality and following the notations in (3.9), we can assume that $|h|_r=1$, since this amounts to rescaling time. Then, by Theorem B.3 with $\varrho\rightsquigarrow r$, $\sigma\rightsquigarrow s$, and $E\rightsquigarrow 1$ if the following bounds are satisfied

$$mrT\leq 3\times10^{-3}\frac{s}{M}\;,\quad\frac{\varepsilon T}{r}\leq\frac{s}{72\xi\;\mathtt{C}_4(n,\ell)}\;,\quad\varepsilon\leq\frac{M}{20\;\mathtt{C}_4(n,\ell)}r^2 \tag{6.3}$$

there exists a symplectic map $\Psi:\mathcal{B}_{r/6,s/6}(I_0,r)\longrightarrow\mathcal{B}_{r,s}(I_0,r)$ of size

$$|\Pi_I\Psi-I|_2\leq\frac{r}{6\xi}\;,\quad|\Pi_\theta\Psi-\theta|_2\leq\frac{s}{6}$$

taking $\mathtt{H}_s$ into

$$\mathtt{H}_s\circ\Psi=h+\mathtt{g}+\mathtt{f}^*_s\;,$$

where

$$|\mathtt{f}^*_s|_{r/6,s/6}\leq\mathtt{C}_4(n,\ell)\left(324\frac{MTr}{s}\right)\;2^{-m}\varepsilon\;,\quad|\mathtt{g}|_{r/6,s/6}\leq\mathtt{C}_4(n,\ell)\left(1+324\frac{MTr}{s}\right)\varepsilon\;. \tag{6.4}$$

As we did in the patchwork method, we now set the dependence of the different parameters on the size of the perturbation. For $\varepsilon>0$ and $a,b,\mathtt{r}_*>0$, we set

$$s:=\varepsilon^a\;,\quad m:=\frac{|\ln\varepsilon^{1+b}|}{\ln 2}\;,\quad r:=\mathtt{r}_*\frac{\varepsilon^c}{T|\ln\varepsilon^{1+b}|}\;. \tag{6.5}$$

- Condition on the thresholds in (6.3) The choices in (6.5), once injected in the thresholds (6.3), immediately give out the following implications:

$$r\leq s\;,\;\;\Longrightarrow a\leq c\;, \tag{6.6}$$

(6.7)
$$mrT\leq3\times10^{-3}\frac{s}{M}\Longrightarrow(1+b)\frac{|\ln\varepsilon|}{\ln 2}\mathtt{r}_*\frac{\varepsilon^c}{(1+b)T|\ln\varepsilon|}T\leq3\times10^{-3}\frac{\varepsilon^a}{M}\;\Longrightarrow\;\mathtt{r}_*\leq3\times10^{-3}\times\frac{\ln 2}{M}\;,\;\;a<c\;,\;\;\varepsilon\leq1\;,$$

$$\frac{\varepsilon T}{r}\leq\frac{s}{72\xi\;\mathtt{C}_4(n,\ell)}\;\Longrightarrow\;\frac{72\xi\;\mathtt{C}_4(n,\ell)}{\mathtt{r}_*}(1+b)T^2|\ln\varepsilon|\;\varepsilon^{1-c-a}\leq1\;,\quad c+a<1\,, \tag{6.8}$$

$$\varepsilon\leq\frac{M}{20\;\mathtt{C}_4(n,\ell)}r^2\Longrightarrow\frac{20\;\mathtt{C}_4(n,\ell)}{M\mathtt{r}_*^2}(1+b)^2T^2|\ln\varepsilon|^2\;\varepsilon^{1-2c}\leq1\;,\quad c<\frac{1}{2}\;. \tag{6.9}$$

We set $\mathtt{r}_*=3\times10^{-3}\times\dfrac{\ln 2}{M}$ and

$$\mathtt{C}_N(n,\ell):=\frac{9\times10^{-6}\times\ln^2 2}{20M\mathtt{C}_4(n,\ell)} \tag{6.10}$$

so that from (6.6)-(6.7)-(6.8)-(6.9) we obtain the following conditions

$$(1+b)^2T^2|\ln\varepsilon|^2\varepsilon^{1-2c} \leq \frac{\mathtt{C}_N(n,\ell)}{\xi} \qquad 0<a\leq c<\frac{1}{2} \qquad c+a<1\ . \tag{6.11}$$

- **Diffeomorphisms and remainders.** If (6.11) is satisfied, Normal Form Lemma (B.3) can be applied and one has

$$(h+f)\circ\Psi = (h+\mathtt{f}_s)\circ\Psi + (f-\mathtt{f}_s)\circ\Psi = h+\mathtt{g}+\mathtt{f}_s^* + (f-\mathtt{f}_s)\circ\Psi := h+\mathtt{g}+\mathtt{R}\ . \tag{6.12}$$

By the first threshold in (6.3), expression (6.4) and the definition of m in (6.5), the estimates on the resonant and non-resonant remainders read

$$\|\mathtt{g}\|_{C^0(\mathcal{B}_{r/6,s/6}(I_0,r))} \leq 2\mathtt{C}_4(n,\ell)\varepsilon\ , \quad \|\mathtt{f}_s^*\|_{C^0(\mathcal{B}_{r/6,s/6}(I_0,r))} \leq \mathtt{C}_4(n,\ell)\varepsilon^{2+b}\ . \tag{6.13}$$

In order not to burden notations, since the analyticity width of domain does not change, we set

$$\|\mathtt{g}\| := \|\mathtt{g}\|_{C^0(\mathcal{B}_{r/6,s/6}(I_0,r))} \qquad \|\mathtt{f}_s^*\| := \|\mathtt{f}_s^*\|_{C^0(\mathcal{B}_{r/6,s/6}(I_0,r))}\ .$$

On the other hand, the Hölder remainder is estimated in the usual way as in (5.38)

$$\|(f-\mathtt{f}_s)\circ\Psi\|_{C^1\left(\mathcal{B}^{\mathbb{R}}_{r/6,s/6}(I_0,r)\right)} \leq \mathtt{C}_{\mathcal{F}}\|f-\mathtt{f}_s\|_{C^1(B_2(I_0,2r)\times\mathbb{T}^n)} \leq \mathtt{C}_{\mathcal{F}}\mathtt{C}_A s^{\ell-1}|f|_{C^\ell(B_\infty(0,R)\times\mathbb{T}^n)} \leq \mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{1+a(\ell-1)}\ . \tag{6.14}$$

6.2. **Stability in the neighborhood of a periodic torus.** With the settings of the previous section, we are now able to prove the following result of stability around an action corresponding to a periodic frequency, i.e. around a periodic torus for the unperturbed system. In the sequel we denote

$$\mathtt{C}_0'' := \min\left\{\mathtt{C}_N, n\hat{\omega}\frac{\mathtt{C}_{\mathcal{F}}\mathtt{C}_A}{\mathtt{C}_4}, \frac{\mathtt{C}_{\mathcal{F}}\mathtt{C}_A}{\mathtt{C}_4}, \hat{\omega}\right\}\ . \tag{6.15}$$

Theorem 6.1 (Stability in the neighborhood of a periodic torus)**.** *Let I_0 be an action corresponding to a T-periodic torus for the unperturbed hamiltonian h. Suppose that for three positive numbers a, c, ρ the following relations are fulfilled*

$$(1+a\ell)^2T^2|\ln\varepsilon|^2\varepsilon^{1-2c} \leq 6\left(\frac{\mu}{M(1+\sqrt{2})}-\rho\right)\mathtt{C}_0''\ , \quad 0<a\leq c<\frac{1}{2}\ , \quad \rho<\frac{\mu}{M(1+\sqrt{2})}\ . \tag{6.16}$$

Then for any initial condition $I(0)$ satisfying

$$|I(0)-I_0|_2 \leq \rho\frac{\varepsilon^c}{(1+a\ell)\ T|\ln\varepsilon|} \tag{6.17}$$

there exist explicit constants $\mathtt{C}_1''(n,\ell)$ and $\mathtt{C}_2''(n,\ell)$ such that one has

$$|I(t)-I(0)|_2 \leq \mathtt{C}_2''(n,\ell)\frac{\varepsilon^c}{(1+a\ell)\ T|\ln\varepsilon|} \tag{6.18}$$

over a time

$$|t| \leq \frac{\mathtt{C}_1''(n,\ell)}{[(1+a\ell)\ T]^2|\ln\varepsilon|^2\ \varepsilon^{1+a(\ell-1)-2c}}\ . \tag{6.19}$$

Proof. The first threshold in (6.16) is equivalent to (6.11) with

$$b=a\ell\ , \quad \xi=\frac{1}{6}\left(\frac{\mu}{M(1+\sqrt{2})}-\rho\right)^{-1}\ , \tag{6.20}$$

so that the normal form in Lemma B.3 can be applied for this choice of parameters. Let now $I(0)\in B_2(I_0,\rho r)$. By B.3 $\Psi(I(0)) =: \mathtt{I}(0) \in B_2(I_0,\rho r+\frac{r}{6\xi}) \subset B_2(I_0,r)$, where the inclusion comes from the choice of ξ in (6.20) and from the fact that $\rho<1$ since $\mu\leq M$ by construction. Since h is convex, for any time t inferior to the (possibly infinite) time of escape t_e of the normalized actions starting at $B_2\left(I_0,\rho r+\frac{r}{6\xi}\right)$ from the domain $B_2\left(I_0,r\right)$, we have

$$|h(\mathtt{I}(t))-h(\mathtt{I}(0))| + |\nabla h(\mathtt{I}(0))(\mathtt{I}(t)-\mathtt{I}(0)| \geq \frac{\mu}{2}|\mathtt{I}(t)-\mathtt{I}(0)|_2^{\ 2}\ . \tag{6.21}$$

The first term in the previous expression is estimated in the standard way thanks to the conservation of energy

$$|h(\mathtt{I}(t)) - h(\mathtt{I}(0))| \leq 2||\{\mathtt{R},\mathtt{g}\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)}|t| + ||\{h,\mathtt{R}\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)}|t| \ . \tag{6.22}$$

We have

$$||\{\mathtt{R},\mathtt{g}\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} \leq ||\{\mathtt{f}_s^*,\mathtt{g}\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} + ||\{(f-\mathtt{f}_s)\circ\Psi,\mathtt{g}\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} \ , \tag{6.23}$$

so that, exploiting (6.5), (6.13) and (6.14), we have

$$\begin{aligned} ||\{\mathtt{f}_s^*,\mathtt{g}\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} &\leq \frac{\|\mathtt{f}_s^*\|\,\|\mathtt{g}\|}{r/6\times s/6} \\ &\leq 12\frac{M\mathtt{C}_4^2(n,\ell)}{10^{-3}\times\ln 2}(1+b)T|\ln\varepsilon|\varepsilon^{3+b-a-c} \end{aligned} \tag{6.24}$$

and

$$\begin{aligned} \|\{(f-\mathtt{f}_s)\circ\Psi,\mathtt{g}\}\|_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} \leq& 2n\frac{\|\mathtt{g}\|}{r/6}\times\mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{1+a(\ell-1)} \\ \leq& 8n\frac{M\mathtt{C}_4(n,\ell)\mathtt{C}_{\mathcal{F}}\mathtt{C}_A}{10^{-3}\times\ln 2}(1+b)T|\ln\varepsilon|\varepsilon^{2+a(\ell-1)-c} \ . \end{aligned} \tag{6.25}$$

As for the estimate of the second Poisson bracket in (6.22), we take (6.13) into account, and write

$$||\{h,\mathtt{f}_s^*\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} \leq n\hat{\omega}\|\partial_\theta\mathtt{f}_s^*\|_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} \leq 6n\hat{\omega}\frac{\|\mathtt{f}_s^*\|}{s} \leq 6n\hat{\omega}\mathtt{C}_4(n,\ell)\varepsilon^{2+b-a} \tag{6.26}$$

and, by (6.14), also

$$||\{h,(f-\mathtt{f}_s)\circ\Psi\}||_{C^0(B_2(I_0,r)\times\mathbb{T}^n)} \leq n\hat{\omega}\mathtt{C}_{\mathcal{F}}\mathtt{C}_A\varepsilon^{1+a(\ell-1)} \ . \tag{6.27}$$

Since $b = a\ell$ and $c < 1/2$, by comparing (6.24), (6.25), (6.26) and (6.27) and by the thresholds in (6.16) and the definition of $\mathtt{C}_0''$, we have that the larger of such terms is (6.27). Therefore, if we set

$$\mathtt{C}_3'' := 8n\hat{\omega}\mathtt{C}_{\mathcal{F}}\mathtt{C}_A \tag{6.28}$$

we finally get

$$|h(\mathtt{I}(t)) - h(\mathtt{I}(0))| \leq \frac{\mathtt{C}_3''}{2}\ \varepsilon^{1+a(\ell-1)}|t| \ . \tag{6.29}$$

The linear term in (6.21) is also estimated in the usual manner:

$$|\nabla h(\mathtt{I}(0))\cdot(\mathtt{I}(t)-\mathtt{I}(0))| \leq |\omega\cdot(\mathtt{I}(t)-\mathtt{I}(0))| + |(\nabla h(\mathtt{I}(0))-\omega)\cdot(\mathtt{I}(t)-\mathtt{I}(0))| \tag{6.30}$$

and the two terms on the r.h.s. of the inequality are bounded in as in Section 5.3, namely, as in (5.47), we have

$$|\omega\cdot(\mathtt{I}(t)-\mathtt{I}(0))| \leq \frac{\mathtt{C}_3''}{2}\varepsilon^{1+a(\ell-1)}|t| \ , \tag{6.31}$$

and, as in (5.46),

$$\begin{aligned} |(\nabla h(\mathtt{I}(0))-\omega)\cdot(\mathtt{I}(t)-\mathtt{I}(0))| &\leq M|\mathtt{I}(0)-I_0|_2|\mathtt{I}(t)-\mathtt{I}(0)|_2 \\ &\leq \left(\rho+\frac{1}{6\xi}\right)Mr|\mathtt{I}(t)-\mathtt{I}(0)|_2 \\ &= \left(\rho+\frac{1}{6\xi}\right)M\mathtt{r}_*\frac{\varepsilon^c}{T|\ln(\varepsilon^b)|}|\mathtt{I}(t)-\mathtt{I}(0)|_2 \ . \end{aligned} \tag{6.32}$$

Plugging all of the above estimates into (6.21) yields

$$\frac{\mu}{2}|\mathtt{I}(t)-\mathtt{I}(0)|_2{}^2 - \left(\rho+\frac{1}{6\xi}\right)M\mathtt{r}_*\frac{\varepsilon^c}{T|\ln(\varepsilon^b)|}|\mathtt{I}(t)-\mathtt{I}(0)|_2 - \mathtt{C}_3''\varepsilon^{1+a(\ell-1)}|t| \leq 0 \ , \tag{6.33}$$

whose solution is

$$|\mathtt{I}(t)-\mathtt{I}(0)|_2 \leq \mathtt{C}_G(\rho,\xi)\mathtt{r}_*\frac{\varepsilon^c}{T|\ln(\varepsilon^b)|} \pm \left\{\left[\mathtt{C}_G(\rho,\xi)\mathtt{r}_*\frac{\varepsilon^c}{T|\ln(\varepsilon^b)|}\right]^2 + \frac{2\mathtt{C}_3''(n,\ell)}{\mu}\varepsilon^{1+a(\ell-1)}|t|\right\}^{1/2} \tag{6.34}$$

where we have set

$$\mathtt{C}_G(\rho,\xi) := \left(\rho + \frac{1}{6\xi}\right)\frac{M}{\mu} \ .$$

If

$$|t| \leq \frac{\mu \mathtt{C}_G^2 \mathtt{r}_*^2}{2\mathtt{C}_3''(n,\ell)} \frac{1}{T^2 |\ln \varepsilon^{1+b}|^2 \ \varepsilon^{1+a(\ell-1)-2c}} \ , \tag{6.35}$$

then the actions are bounded by

$$|\mathtt{I}(t) - \mathtt{I}(0)|_2 \leq (1+\sqrt{2})\mathtt{C}_G \mathtt{r}_* \frac{\varepsilon^c}{T|\ln \varepsilon^{1+b}|} = (1+\sqrt{2})\mathtt{C}_G \ r \ . \tag{6.36}$$

By the choices of ρ and ξ in (6.15) one has $\mathtt{C}_G \leq 1/(1+\sqrt{2})$, so that, by a bootstrap argument identical to the one of the patchwork method, the normalized actions stay in the domain $B_2(I_0, r)$. The total variation of the action variables is obtained by coming back in the initial coordinates, so that we finally find

$$|I(t) - I(0)|_2 \leq \left[(1+\sqrt{2})\left(\rho + \frac{1}{6\xi}\right)\frac{M}{\mu} + \frac{1}{6\xi}\right] \mathtt{r}_* \frac{\varepsilon^c}{T|\ln \varepsilon^{1+a\ell}|} \ . \tag{6.37}$$

The claim is therefore proven with

$$\mathtt{C}_1'' = \frac{\mu \mathtt{C}_G^2 \mathtt{r}_*^2}{2\mathtt{C}_3''(n,\ell)} \ , \quad \mathtt{C}_2'' = \left[(1+\sqrt{2})\left(\rho + \frac{1}{6\xi}\right)\frac{M}{\mu} + \frac{1}{6\xi}\right] \mathtt{r}_* \ .$$

□

6.3. **Proof of Theorem 1.3.** We are now able to prove Theorem 1.3 with the help of Corollary B.1 of Dirichlet's Theorem (B.1) on simultaneous approximations, whose notations shall henceforth be adopted.

Q represents the "velocity" at which rational frequencies of unperturbed periodic tori are approached. Take two real numbers $Q_0 > \mu^2$ and $q > 0$, set

$$Q := \left(\frac{Q_0 |\ln \varepsilon^{1+b}| \varepsilon^{-q}}{\mu^2}\right)^{n-1} > 1 \tag{6.38}$$

so that, for any action $I(0) \in B_\infty(0, \frac{R}{2})$ corresponding to a frequency $\Omega := \partial_I h(I(0))$ there exists a T-periodic frequency ω fulfilling

$$|\Omega - \omega|_2 \leq \sqrt{n-1} \frac{\mu^2 \varepsilon^q}{Q_0 |\ln \varepsilon^{1+b}| T} \ .$$

Hence, since the frequency map is invertible, there exists an action I_0 such that

$$\mu |I(0) - I_0|_2 \leq \sqrt{n-1} \frac{\mu^2 \varepsilon^q}{Q_0 |\ln \varepsilon^{1+b}| T} \ . \tag{6.39}$$

We apply the local stability Lemma 6.1 with $\rho = \dfrac{\mu}{2M(1+\sqrt{2})}$ and we impose that $I(0)$ is contained in the "influence zone" of the periodic torus corresponding to action I_0 that is $B_2(I_0, \rho r)$, hence by (6.5)-(6.39),

$$\sqrt{n-1} \frac{\mu \varepsilon^q}{Q_0 |\ln \varepsilon^{1+b}| T} \leq \rho r \implies \frac{\sqrt{n-1}\ \mu}{Q_0 |\ln \varepsilon^{1+b}| T} \varepsilon^q \leq \frac{\mu}{2M(1+\sqrt{2})} \times 3 \times 10^{-3} \times \frac{\ln 2}{M} \frac{\varepsilon^c}{T|\ln \varepsilon^{1+b}|} \ . \tag{6.40}$$

Setting

$$Q_0 = 2000 \times \sqrt{n-1} \frac{M^2(1+\sqrt{2})}{3 \ln 2} \ , \quad q := c \tag{6.41}$$

such condition yields, trivially, $\varepsilon \leq 1$. Now, Lemma (6.1) applies everywhere in the phase space provided that condition (6.16) is satisfied uniformly for all $T \geq 1$. In order to insure this, we observe that, by Corollary B.1, we have $T = u/|\Omega|_\infty$ and $1 \leq u < Q$, since by hypothesis we have $\underline{\omega} := \inf_{I' \in B_\infty(0,R)} |\partial_I h(I')| > 0$ we can impose the stronger condition

$$(1+a\ell)^2 Q^2 |\ln \varepsilon|^2 \varepsilon^{1-2c} \leq 6\underline{\omega} \ \left(\frac{\mu}{M(1+\sqrt{2})} - \rho\right) \mathtt{C}_0'' \tag{6.42}$$

which is equivalent to

$$(1+a\ell)^{2n}|\ln\varepsilon|^{2n}\varepsilon^{1-2nc} \leq \frac{3\mu\underline{\omega}}{M(1+\sqrt{2})}\mathtt{C}_0''\left(\frac{\mu^2}{Q_0}\right)^{n-1} \quad , \quad 0<a\leq c<\frac{1}{2n} \ . \tag{6.43}$$

By expression (6.37) we see that the worst radius of confinement is obtained for $T=1$, that is

$$|I(t)-I(0)| \leq \mathtt{C}_2''(n,\ell)\frac{\varepsilon^c}{(1+a\ell)\ |\ln\varepsilon|} \ , \quad 0<c<\frac{1}{2n} \tag{6.44}$$

and by (6.35) we see that the worst time of confinement is obtained when $T=q/|\Omega|_\infty$ is maximal, namely when $T=Q/\underline{\omega}$, so that by taking expression (6.38) into account we get

$$|t| \leq \frac{\mathtt{C}_1''(n,\ell)\underline{\omega}}{[(1+a\ell)\ Q]^2|\ln\varepsilon|^2\ \varepsilon^{1+a(\ell-1)-2c}} = \left(\frac{\mu^2}{Q_0}\right)^{n-1}\underline{\omega}\mathtt{C}_1''(n,\ell)\frac{1}{(1+a\ell)^{2n}|\ln\varepsilon|^{2n}\ \varepsilon^{1+a(\ell-1)-2cn}} \ . \tag{6.45}$$

This proves Theorem 1.3, once we set

$$\mathtt{C}'_{\mathbf{E}} = \frac{3\mu\underline{\omega}}{M(1+\sqrt{2})}\mathtt{C}_0''\left(\frac{\mu^2}{Q_0}\right)^{n-1} \qquad \mathtt{C}'_{\mathbf{T}} = \left(\frac{\mu^2}{Q_0}\right)^{n-1}\underline{\omega}\mathtt{C}_1''(n,\ell) \qquad \mathtt{C}'_{\mathbf{I}} = \mathtt{C}_2''(n).$$

7. Stability estimates in the steep case

The aim of this section is to prove Theorem 1.1.

7.1. **Construction of the resonant patchwork.** The first step in order to obtain stability estimates in the steep case consists in building an appropriate resonant covering of the phase space for the integrable hamiltonian h, exactly as it was done for the convex case with the patchwork method. In the sequel, we perform this task by following closely ref. [23]. The construction is more complicated than in the convex case and the dependence of the involved quantities on the ultraviolet cut-off K is substantially different. We refer to [23] for a heuristic explanation of the choice of such dependences. All the notations of the previous sections are assumed hereafter but, for the sake of simplicity, when possible we shall avoid the explicit computation of constants, which will simply be indicated with a dot " $\cdot$ " in the sequel.

Now we set some parameters, depending on the steepness indices $\boldsymbol{\alpha}_1,..,\boldsymbol{\alpha}_{n-1}$ of h, that will be useful throughout this section

$$p_j := \begin{cases} \Pi_{i=j}^{n-2}\boldsymbol{\alpha}_i & , \quad \text{if } j\in\{1,...,n-2\} \\ 1 & , \quad \text{if } j\in\{n-1,n\} \end{cases} \quad ; \quad q_j := np_j - j \ , \ \ j\in\{1,...,n\}; \quad c_j := q_j - q_{j+1} \ , \ \ j\in\{1,...,n-1\} \tag{7.1}$$

and recall that

$$a := \frac{1}{2n\boldsymbol{\alpha}_1...\boldsymbol{\alpha}_{n-2}} = \frac{1}{2np_1} \quad , \qquad b := \frac{1}{2n\boldsymbol{\alpha}_1...\boldsymbol{\alpha}_{n-1}} = \frac{a}{\boldsymbol{\alpha}_{n-1}} \quad , \qquad \mathtt{R}(\varepsilon) = \cdot\,\varepsilon^b \ . \tag{7.2}$$

With this setting, we fix an action $I_0\in B_\infty(0,R/2)$ and we consider its neighborhood $B_2(I_0,\mathtt{R}(\varepsilon))$.

Since h is steep in $B_\infty(0,R)$, the norm of the frequency $\omega := \partial_I h(I)$ at any point of this set admits a uniform lower positive bound, that is $\inf_{I\in B_\infty(0,R)}||\omega(I)|| > \cdot 1$. Hence, when studying the geography of resonances for h, for sufficiently small ε and without any loss of generality we can just consider maximal lattices $\Lambda\subset\mathbb{Z}^n_K$ of dimension $j\in\{0,...,n-1\}$, with $K\geq 1$ the ultraviolet cut-off. For a given Λ, we define its associated *resonant zone* as

$$Z_\Lambda := \{I\in B_2(I_0,\mathtt{R}(\varepsilon)):\ \forall k\in\Lambda \text{ one has} \quad |k\cdot\omega(I)|<\delta_\Lambda\} \ , \quad \delta_\Lambda := \cdot\frac{1}{|\Lambda|K^{q_j}} \ . \tag{7.3}$$

The *resonant block* D_Λ is defined as that part of the resonant zone Z_Λ which does not contain any other resonances other than the one associated to Λ, namely

$$D_\Lambda := Z_\Lambda \setminus \bigcup_{\Lambda':\,\dim\Lambda'=j+1} Z_{\Lambda'} \ . \tag{7.4}$$

In particular, this implies that for the completely non-resonant block associated to $\Lambda = \{0\}$ and for any block Λ corresponding to a maximal resonance of dimension $j = n-1$ one has, respectively

$$D_0 := B(I_0, \mathtt{R}(\varepsilon)) \setminus \bigcup_{\Lambda' :\ \dim \Lambda' = 1} Z_{\Lambda'} \quad \text{and} \quad D_\Lambda = Z_\Lambda \ . \tag{7.5}$$

For any $j \in \{0, ..., n-1\}$ we set

$$D_j := \bigcup_{\Lambda :\ \dim \Lambda = j} D_\Lambda \quad , \qquad Z_j := \bigcup_{\Lambda :\ \dim \Lambda = j} Z_\Lambda \tag{7.6}$$

and it is easy to see from (7.4) that

$$D_j = Z_j \setminus Z_{j+1} \tag{7.7}$$

so that one has the decomposition

$$B(I_0, \mathtt{R}(\varepsilon)) = \bigcup_{i=0}^{n-1} D_i \quad , \qquad B(I_0, \mathtt{R}(\varepsilon)) = \left(\bigcup_{i=0}^{j-1} D_i \right) \cup Z_j \qquad \forall j = 1, ..., n-1 \ . \tag{7.8}$$

Since, as we have explained in the introduction, a large drift over a short time of any action variable I belonging to the resonant block D_Λ associated to a maximal lattice $\Lambda \neq \{0\}$ is only possible along the plane of fast drift $I + \langle \Lambda \rangle$ spanned by the vectors belonging to Λ, we are naturally taken to consider the intersection of a neighborhood of $I + \langle \Lambda \rangle$ with Z_Λ. Indeed, as we have heuristically explained in the introduction, the fast motion of the orbit starting at I along $I + \langle \Lambda \rangle$ can take the actions out of the block D_Λ. So, we are interested in understanding what happens when the actions leave D_Λ but keep staying in Z_Λ. In this spirit, we fix

$$\rho(\varepsilon) := \frac{\mathtt{R}(\varepsilon)}{2n} \tag{7.9}$$

and, for any $0 < \eta \leq \rho(\varepsilon)$ and for any action $I \in D_\Lambda$ with $\Lambda \neq \{0\}$, we define the *disc* associated to I as

$$\mathcal{D}^\rho_{\Lambda,\eta}(I) := \left(\left(\bigcup_{I' \in I + \langle \Lambda \rangle} B_2(I', \eta) \right) \cap Z_\Lambda \cap B\big(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon)\big) \right)_I \tag{7.10}$$

where the subscript I denotes the connected component of the set containing the action I. Since we are going to study the fate of all orbits starting at a fixed block D_Λ, with $\dim \Lambda = j \in \{1, ..., n-1\}$, that exit such block in a short time along the plane of fast drift, we are also led to define the *extended resonant block*

$$D^\rho_{\Lambda, r_\Lambda} := \left(\bigcup_{I \in D_\Lambda \cap B(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon))} \mathcal{D}^\rho_{\Lambda, r_\Lambda}(I) \right) \ \subset \ Z_\Lambda \cap B\big(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon)\big) \quad , \qquad r_\Lambda := \cdot \frac{1}{|\Lambda| K^{q_j}} \ . \tag{7.11}$$

In the same way, the *extended non-resonant block* is defined as

$$D^\rho_0 := D_0 \cap B(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon)) \ . \tag{7.12}$$

7.2. **The resonant blocks.** In this paragraph we shall give a rigorous framework to some of the heuristics of the introduction. As we have explained there, Nekhoroshev proved in [32] that, if h is steep, when any action $I \in D_\Lambda$, with $\Lambda \neq \{0\}$, moves along the plane of fast drift, it must exit the resonant zone Z_Λ. Indeed, if h is steep with steepness indices $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_{n-1}$ one can prove that the diameter of the intersection of a neighborhood of the fast drift plane with the resonant zone is small in the sense given by the following

Lemma 7.1. *For any $\Lambda \neq 0$, $\dim \Lambda = j \in \{1, ..., n-1\}$, for any $I \in D_\Lambda \cap B(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon))$ and for any $I' \in \mathcal{D}^\rho_{\Lambda, r_\Lambda}(I)$ one has*

$$|I - I'|_2 \leq r_j \quad , \qquad \textit{where} \quad r_j := \cdot \ \frac{1}{K^{q_j/\alpha_j}} \ . \tag{7.13}$$

By this lemma, we see that a smaller value of ε, i.e. a higher value of K since the ultraviolet cut-off is always a decreasing function of ε, leads to a closer maximal distance between any action I belonging to a resonant block and any action belonging to its disc. For a proof of this result we refer to Lemma 2.1 of ref. [23].

Clearly, in order to perform normal forms in the (extended) resonant blocks, we also need an estimate of the small divisors in these sets, namely we have

Lemma 7.2. *For any maximal lattice $\Lambda \in \mathbb{Z}^n_K$ of dimension $j \in \{0, ..., n-1\}$, for any $k \in \mathbb{Z}^n_K \backslash \Lambda$ and for any $I \in D^\rho_{\Lambda, r_\Lambda}$ one has*

$$|\langle k, \omega(I)\rangle| \geq \cdot \frac{1}{|\Lambda| K^{q_j - c_j}} \ , \tag{7.14}$$

whereas for any action I in the completely non-resonant block D_0 and for any $k \in \mathbb{Z}^n_K$ one has

$$|\langle k, \omega(I)\rangle| \geq \cdot \frac{1}{K^{q_1}} \ . \tag{7.15}$$

We refer again to [23, Lemma 2.2] for a proof of this result.

Finally, a key ingredient in order to insure stability in the steep case is the fact that, when possibly exiting a resonant zone along the plane of fast drift, the actions must enter another resonant zone associated to a lattice of lower dimension. This is the content of

Lemma 7.3. *Let Λ, Λ' two maximal lattices of $\mathbb{Z}^n_K$ having the same dimension $j \in \{1, ..., n-1\}$. Then one has*

$$\text{closure}\left(D^\rho_{\Lambda, r_\Lambda}\right) \cap Z_{\Lambda'} = \varnothing \ . \tag{7.16}$$

Once again, the proof of this Lemma can be found in [23] (Lemma 2.3).

With the ingredients of this paragraph, we are able to prove stability in the steep case.

7.3. **Proof of Theorem 1.1.** Theorem 1.1, whose proof is at the end of the section, is a consequence of the following Theorem 7.1 and Lemma 7.4. We start by giving the standard estimates of stability in the completely non-resonant extended block D^ρ_0. As we have already seen in the proof of Theorem 5.1, such estimates do not require any geometric assumption on the integrable part h.

Theorem 7.1 (Non-resonant Stability Estimates)**.** *For any sufficiently small ε, there exist implicit constants such that for any time t satisfying*

$$|t| \leq T_0 := \cdot \frac{1}{\sqrt{\varepsilon}\ |(1 + a\ell) \ln \varepsilon|^{\ell - 1}\ \varepsilon^{a(\ell-1)}} \tag{7.17}$$

any initial condition $I(0) \in D^\rho_0$ drifts at most as

$$|I(t) - I(0)|_2 \leq \cdot\ \varepsilon^{1/2} \ , \tag{7.18}$$

where a was defined in (7.2)*.*

Proof. The proof is essentially the same as the one of Theorem 5.1. One only needs to consider a different dependence on K of the small divisors (see (7.15)) and implement a slightly different dependence on ε of the parameters, namely

$$K := \cdot\ \varepsilon^{-a} \ , \quad s := \cdot \varepsilon^a |\ln \varepsilon^{6(1+a\ell)}| \ , \quad r := \cdot \frac{1}{K^{1+q_1}} = \cdot\ \varepsilon^{1/2} \ . \tag{7.19}$$

Then, for sufficiently small ε, it is a matter of standard computations analogous to those in the proof of Theorem 5.1 to show that Pöschel's normal form applies in D^ρ_0 and that the Fundamental Theorem of calculus yields the result. □

As for the dynamics in the resonant blocks, we have the following

Lemma 7.4. *Consider a maximal lattice $\Lambda \subset \mathbb{Z}^n_K$ of dimension $j \in \{1, ..., n-1\}$. For any sufficiently small ε and for any initial condition $(I(0), \theta(0)) \in \left(D_\Lambda \cap B\big(I_0, \mathtt{R}(\varepsilon) - (j+1)\rho(\varepsilon)\big)\right) \times \mathbb{T}^n$, set*

$$T_j := \cdot \frac{1}{|\ln \varepsilon^{6(1+\delta\ell)}|^{\ell-1}\, \varepsilon^{\delta(\ell-1+np_{j+1})}} \quad , \qquad \delta := \frac{1}{n(p_j + p_{j+1})} \tag{7.20}$$

and consider the time of escape of the flow generated by H from the extended resonant block

$$\tau_e := \inf \left\{ t \in \mathbb{R} : \Phi^t_H \left(D_\Lambda \cap B\big(I_0, \mathtt{R}(\varepsilon) - (j+1)\rho(\varepsilon)\big) \times \mathbb{T}^n \right) \not\subset D^\rho_{\Lambda, r_\Lambda} \times \mathbb{T}^n \right\} \ , \tag{7.21}$$

Then

- *If $|\tau_e| \geq T_j$ one has*

$$|I(t) - I(0)|_2 < \rho(\varepsilon) \tag{7.22}$$

  over a time $|t| \leq T_j$;
- *If $|\tau_e| < T_j$ there exists $i \in \{0, ..., j-1\}$ such that*

$$I(\tau_e) \in D_i \cap \left(B\big(I_0, \mathtt{R}(\varepsilon) - j\rho(\varepsilon)\big)\right) \ .$$

Proof. We start by considering the case $|\tau_e| \geq T_j$. By exploiting the same arguments of paragraph 5.1, it is easy to see that, for sufficiently small ε and by suitably adjusting the implicit constants, Pöschel's Normal Form (see Lemma B.1) applies in $D^\rho_{\Lambda, r_\Lambda}$ with parameters

$$K := \cdot\, \varepsilon^{-\delta} \ , \quad s := \cdot\, \varepsilon^{\delta} |\ln \varepsilon^{6(1+\delta\ell)}| \ , \quad r_\Lambda := \cdot \frac{1}{|\Lambda| K^{q_j}} \tag{7.23}$$

and with a small divisor estimate given by formula (7.14) in Lemma 7.2, namely

$$\gamma_\Lambda = \cdot \frac{1}{|\Lambda| K^{q_j - c_j}} \ . \tag{7.24}$$

Therefore, by taking into account the notations for the analytic smoothing introduced in the previous sections, there exists a symplectic transformation $\Psi_\Lambda : (D^\rho_{\Lambda, r_\Lambda})_{r_\Lambda/2} \times \mathbb{T}^n_{s/6} \longrightarrow (D^\rho_{\Lambda, r_\Lambda})_{r_\Lambda} \times \mathbb{T}^n_s$, $(\mathtt{I}, \vartheta) \longmapsto (I, \theta)$ taking H into resonant normal form

$$H \circ \Psi_\Lambda = \mathtt{H}_s \circ \Psi_\Lambda + (H - \mathtt{H}_s) \circ \Psi_\Lambda = h + \mathtt{g} + \mathtt{f}^*_s + (f - \mathtt{f}_s) \circ \Psi_\Lambda \tag{7.25}$$

with $\{h, \mathtt{g}\} = 0$, $||\mathtt{f}^*_s||_{r/2, s/6} \leq \cdot\, e^{-Ks/6}\, \varepsilon$ and $|I - \mathtt{I}|_2 \leq \cdot\, r_\Lambda$, where the implicit constant in the estimate can be taken suitably small without loss of generality (see Lemma B.1 and the rôle that the constant ξ plays in it; for a smaller threshold the normalized actions stay closer to the original ones).

Now, for any time t such that $|t| \leq |\tau_e|$, the dynamics on the subspace orthogonal to $\langle \Lambda \rangle$ can be controlled in the usual way by exploiting the smallness of the non-resonant remainder $\mathtt{f}^*_s$ as well as that of $(f - \mathtt{f}_s) \circ \Psi_\Lambda$. Namely, for any initial position in the actions $I(0) \in D^\rho_{\Lambda, r_\Lambda}$ one has that the normalized coordinate satisfies $\mathtt{I}(0) \in \big(D^\rho_{\Lambda, r_\Lambda}\big)_{\cdot r_\Lambda}$ and

$$\begin{aligned} \big|\Pi_{\langle\Lambda\rangle^\perp}\big(\mathtt{I}(t) - \mathtt{I}(0)\big)\big|_2 &\leq \cdot \left(\frac{e^{-Ks/6}}{s} + s^{\ell-1} \right) \varepsilon |t| \leq \left(\frac{\varepsilon^{2+\delta(\ell-1)}}{|\ln \varepsilon^{6(1+\delta\ell)}|} + |\ln \varepsilon^{6(1+\delta\ell)}|^{\ell-1} \varepsilon^{1+\delta(\ell-1)} \right) |t| \\ &\leq \cdot \left(|\ln \varepsilon^{6(1+\delta\ell)}|^{\ell-1} \varepsilon^{1+\delta(\ell-1)} \right) |t| \ . \end{aligned} \tag{7.26}$$

Then, if for a suitably small implicit constant we set

$$T_\Lambda := \cdot \frac{r_\Lambda}{|\ln \varepsilon^{6(1+\delta\ell)}|^{\ell-1}\, \varepsilon^{1+\delta(\ell-1)}} \tag{7.27}$$

with the choice of r_Λ in (7.23), by considering that $|\Lambda| \leq K^j$ and thanks to the choice of δ, it is straightforward to show that $T_j \leq T_\Lambda$ and, for any time t satisfying $|t| \leq T_j \leq T_\Lambda$ estimate (7.26) implies

$$\big|\Pi_{\langle\Lambda\rangle^\perp}\big(\mathtt{I}(t) - \mathtt{I}(0)\big)\big|_2 \leq \frac{r_\Lambda}{4} \ . \tag{7.28}$$

Since for any $|t| \leq |\tau_e|$ the dynamics of the action variables can be decomposed as

$$\begin{aligned} I(t) - I(0) =& I(t) - \mathtt{I}(t) + \mathtt{I}(t) - \mathtt{I}(0) + \mathtt{I}(0) - I(0) \\ =& I(t) - \mathtt{I}(t) + \Pi_{\langle\Lambda\rangle^\perp}\big(\mathtt{I}(t) - \mathtt{I}(0)\big) + \Pi_{\langle\Lambda\rangle}\big(\mathtt{I}(t) - \mathtt{I}(0)\big) + \mathtt{I}(0) - I(0) \ , \end{aligned} \tag{7.29}$$

inequality (7.28), together with the estimate on the size of the normal form, implies that, for any time t fulfilling $|t| \leq T_j$ the motion in the direction perpendicular to the fast drift plane is bounded by

$$\begin{aligned} |I(t) - I(0) - \Pi_{\langle\Lambda\rangle}\big(\mathtt{I}(t) - \mathtt{I}(0)\big)|_2 \leq& |I(t) - \mathtt{I}(t)|_2 + |\Pi_{\langle\Lambda\rangle^\perp}\big(\mathtt{I}(t) - \mathtt{I}(0)\big)|_2 + |\mathtt{I}(0) - I(0)|_2 \\ \leq& \cdot \ r_\Lambda + \frac{r_\Lambda}{4} + \cdot r_\Lambda \leq \frac{3}{4} r_\Lambda \ , \end{aligned} \tag{7.30}$$

where we have used the fact that, as we said above, the implicit constant for the size of the change of variables can be taken suitably small without any loss of generality. Since $I(t) \in D^\rho_{\Lambda, r_\Lambda}$ for all $|t| \leq |\tau_e|$, $I(0)$ is connected to $I(t)$ in such set and, by (7.30) and the definition of disc in (7.10), we also have $I(t) \in \mathcal{D}^\rho_{\Lambda, \frac{3}{4} r_\Lambda}(I(0)) \subset \mathcal{D}^\rho_{\Lambda, r_\Lambda}(I(0))$ and this, together with Lemma 7.1, yields

$$|I(t) - I(0)|_2 \leq r_j \quad , \qquad \text{where} \qquad r_j := \cdot \ \frac{1}{K^{q_j/\alpha_j}} \ . \tag{7.31}$$

As it is shown in [23] (formula (38)), a careful choice of the implicit constants leads to

$$\max_{j \in \{1, \dots, n-1\}} r_j < \rho(\varepsilon) \ ,$$

which concludes the proof of the first claim of this Lemma.

We now consider the second claim. In this case, for any time t such that $|t| < |\tau_e| \leq T_j \leq T_\Lambda$ we can repeat the same arguments above and find $I(t) \in \mathcal{D}^\rho_{\Lambda, \frac{3}{4} r_\Lambda}(I(0))$. Then, by construction, the escape time satisfies

$$I(\tau_e) \in \text{closure}\big(\mathcal{D}^\rho_{\Lambda, \frac{3}{4} r_\Lambda}(I(0))\big). \tag{7.32}$$

Again, by Lemma 7.1, this implies $|I(\tau_e) - I(0)|_2 < \rho(\varepsilon)$, so that, since $I(0) \in B_2\big(I(0), \mathtt{R}(\varepsilon) - (j+1)\rho(\varepsilon)\big)$ one has

$$I(\tau_e) \in B_2\big(I(0), \mathtt{R}(\varepsilon) - j\rho(\varepsilon)\big) \, . \tag{7.33}$$

Now, we shall prove that $I(\tau_e) \notin Z_\Lambda$. By definition we have $I(\tau_e) \notin D^\rho_{\Lambda, r_\Lambda}$ and, thanks to (7.11), this means that there does not exist any action $I^* \in D_\Lambda \cap B\big(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon)\big)$ such that $I(\tau_e)$ belongs to its disc $\mathcal{D}^\rho_{\Lambda, r_\Lambda}(I^*)$. Hence, by (7.10), $I(\tau_e)$ must satisfy at least one of the three following conditions:

(1) $\nexists I^* \in D_\Lambda \cap B_2\big(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon)\big) : I(\tau_e) \in \bigcup_{I' \in I^* + \langle\Lambda\rangle} B_2(I', r_\Lambda)$;
(2) $I(\tau_e) \notin Z_\Lambda$;
(3) $I(\tau_e) \notin B_2\big(I_0, \mathtt{R}(\varepsilon) - \rho(\varepsilon)\big)$.

By taking (7.32) and (7.33) into account, we see that the first and the third possibility cannot occur. Therefore, there must exist a maximal lattice $\Lambda' \neq \Lambda$ and a resonant zone $Z_{\Lambda'}$ such that $I(\tau_e) \in Z_{\Lambda'}$. Moreover, Lemma 7.3, insures that $\dim \Lambda' \neq \dim \Lambda$ so that, by taking the second decomposition in (7.8) as well as (7.33) into account, the second claim is proven. □

Remark 7.1. The decompositions in (7.8) are a covering of $B(I_0, \mathtt{R}(\varepsilon))$ but they are not a partition since, in general, $D_i \cap D_j \neq \varnothing$ for $j > i+1$. Hence, nothing prevents $I(\tau_e)$ from belonging to a resonant block of strictly higher multiplicity than the starting one. If this happens, however, thanks to the construction in (7.8), one is insured that $I(\tau_e)$ will also belong to another block associated to a lower order resonance. One therefore chooses the block in which to study the evolution of the actions once they leave the resonant zone they started at. This is at the core of the *resonant trap argument* we discuss in the sequel.

Proof of Theorem 1.1. Theorem 1.1 follows from Theorem 7.1 and Lemma 7.4. Indeed, for any initial condition in the action variables $I_0 \in B_\infty(0, R/2)$, we consider the ball $B_2(I_0, \mathtt{R}(\varepsilon))$ and the following dichotomy holds:

(1) either I_0 belongs to the completely non-resonant domain D^ρ_0, in which case the proof ends here thanks to Theorem 7.1;

(2) or for some $j \in \{1, ..., n-1\}$ and some maximal $\Lambda \subset \mathbb{Z}^n_K$ of rank j, $I_0 \in D_\Lambda \cap B\big(I_0, \mathtt{R}(\varepsilon) - (j+1)\rho(\varepsilon)\big)$.

In the second case, Lemma 7.4 applies and one has another dichotomy:

(1) either one has $|I(t) - I(0)|_2 \leq \cdot\, \rho(\varepsilon) := \cdot\, \varepsilon^b$ over a time T_j; in such case the Theorem is proven since, with the choice of a in (7.2), one has

$$\mathbf{T}(\varepsilon) := \cdot\, \frac{1}{|\ln \varepsilon^{6(1+a\ell)}|^{\ell-1} \varepsilon^{a(\ell-1)}} \leq T_j \tag{7.34}$$

if ε is sufficiently small ;

(2) or the actions enter a resonant block $D_i \cap \Big(B\big(I_0, \mathtt{R}(\varepsilon) - j\rho(\varepsilon)\big)\Big)$ corresponding to a resonant lattice of dimension $i < j$ after having travelled a distance $\rho(\varepsilon)$ over a time inferior to T_j. In such block, the above arguments can be repeated so that, after having possibly visited at most $n-1$ blocks, overall the actions can travel at most a distance $(n-1)\rho(\varepsilon)$ before entering the completely non-resonant block, in which they are trapped for a time T_0 given by Lemma 7.1 and they travel for another length $\rho(\varepsilon)$. Thanks to (7.9), by construction one has $|I(t) - I(0)| \leq n\rho(\varepsilon) = \mathtt{R}(\varepsilon) = \cdot\, \varepsilon^b$.

This is called the *resonant trap argument* and concludes the proof of Theorem 1.1.

□

Appendix A. Smoothing estimates

Lemma A.1. *The derivatives of K satisfy*

$$\forall p \in \mathbb{N}, \quad \exists C_p \,:\, \big|\partial^\beta K(x)\big| \leq C_p \frac{e^{|\operatorname{Im} x|}}{(1+|x|_2)^p},\ \forall\, |\beta| \leq p.$$

For the proof see [16, Lemma 9].

Lemma A.2. *Let $f \in C^\ell_b(\mathbb{A}^n)$, with $\ell \geq 1$, and let $\sum_{k\in\mathbb{Z}^n} \hat f_k(I) e^{ik\cdot\theta}$ be its Fourier series. Then, for any fixed $k \in \mathbb{Z}^n\backslash\{0\}$, there exists a uniform constant $\mathtt{C_F}(n,\ell)$ satisfying*

$$\Big\|\hat f_k\Big\|_{C^0(\mathbb{R}^n)} \leq \mathtt{C_F}(n,\ell) \frac{\|f\|_{C^q(\mathbb{A}^n)}}{|k|^q}\ , \tag{A.1}$$

where $q := \lfloor \ell \rfloor$.

Proof. Fix a multi-index $j = (j_1, ..., j_n) \in \mathbb{N}^n$ such that $|j|_1 \leq q := \lfloor \ell \rfloor$, one obviously has

$$\partial^j_\theta f(I,\theta) = \sum_{k\in\mathbb{Z}^n} (i)^{|j|} k_1^{j_1} ... k_n^{j_n} \hat f_k(I) e^{ik\cdot\theta}\,. \tag{A.2}$$

From

$$\partial^j_\theta f(I,\theta) := \sum_{k\in\mathbb{Z}^n} (\widehat{\partial^j f})_k(I) e^{ik\cdot\theta}\ , \tag{A.3}$$

and by the unicity of Fourier's coefficients one also has

$$\hat f_k(I) := \frac{(\widehat{\partial^j f})_k(I)}{k_1^{j_1} ... k_n^{j_n}}\ . \tag{A.4}$$

As in expression (A.4) the multi-index $j \in \mathbb{Z}^n$ is arbitrary, for each value of $k \in \mathbb{Z}^n\backslash\{0\}$ we can choose j so that

$$\hat f_k(I) = \frac{(\widehat{\partial^j f})_k(I)}{(\max_{i=1,...,n}\{k_i\})^{|j|}}\ . \tag{A.5}$$

Moreover, for any $k \in \mathbb{Z}^n\backslash\{0\}$ one has the trivial inequality

$$\max_{i=1,...,n} \{|k_i|\} \geq \frac{|k|}{n}\ .$$

This, together with (A.5) and the choice $|j| = q$ yields

$$|\hat{f}_k(I)| = n^\ell \frac{|(\widehat{\partial^j f})_k(I)|}{|k|^q} = n^\ell \frac{1/(2\pi)^n |\int_0^{2\pi} \partial^j f(I,\theta) e^{\mathrm{i}k\cdot\theta} d\theta|}{|k|^q} \leq n^\ell \frac{|\partial^j f(I,\theta)|}{|k|^q} \ , \tag{A.6}$$

which, once the supremum over the actions is taken, implies the result. □

Appendix B. Analytic and arithmetic tools for the convex case

B.1. **Tools of the patchwork method.** Given a function F in $\mathcal{D}_{r,s}$, the notations $\mathcal{P}_\Lambda$ and $\mathcal{P}_K$ stand for the projections

$$\mathcal{P}_\Lambda F(I,\theta) := \sum_{k\in\mathbb{Z}^n : k\in\Lambda} F_k(I) e^{ik\cdot\theta} \ , \quad \mathcal{P}_K F(I,\theta) := \sum_{k\in\mathbb{Z}^n : |k|_1 \leq K} F_k(I) e^{ik\cdot\theta}$$

Accordingly with our notations, we state here the results of Pöschel [39].

Lemma B.1 (Poschel's normal form)**.** *Let $\varrho, \sigma > 0$ and $\mathbf{H}(I,\theta) = \mathbf{h}(I) + \mathbf{f}(I,\theta)$ be analytic on*

$$\mathcal{D}_{\Lambda,\varrho,\sigma} := \{(I,\theta) \in \mathbb{C}^n : |I - D_\Lambda|_2 < \varrho \ , \quad \theta \in \mathbb{T}^n_\sigma\}$$

where D_Λ is (α, K)-nonresonant modulo Λ with respect to the integrable hamiltonian $\mathbf{h}$. Also, let M denote the hermitian norm of the hessian of $\mathbf{h}$ over $\mathcal{D}_{\Lambda,\varrho,\sigma}$.

If, for some $\varrho' > 0$, one is insured

$$||\mathbf{f}||_{\varrho,\sigma} \leq \epsilon \leq \frac{1}{256\xi} \frac{\alpha\varrho'}{K}, \qquad \varrho' \leq \left(\varrho, \frac{\alpha}{2\xi MK}\right) \tag{B.1}$$

for some $\xi > 1$ and

$$K\sigma \geq 6, \tag{B.2}$$

then there exists a real-analytic, symplectic transformation $\Psi : \mathcal{D}_{\Lambda,\varrho'/2,\sigma/6} \longrightarrow \mathcal{D}_{\Lambda,\varrho,\sigma}$ taking $\mathbf{H}$ into resonant normal form, that is

$$\mathbf{H} \circ \Psi = \mathbf{h} + \mathbf{g} + \mathbf{f}^* \ , \quad \{\mathbf{h}, \mathbf{g}\} = 0 \ . \tag{B.3}$$

Moreover, denoting by $\mathbf{g_0} := P_\Lambda P_K \mathbf{f}$ the resonant part of $\mathbf{f}$, we have the estimates

$$||\mathbf{g} - \mathbf{g_0}||_{\varrho'/2,\sigma/6} \leq 64 \frac{K}{\alpha\varrho'} \epsilon^2 \ , \quad ||\mathbf{f}^*||_{\varrho'/2,\sigma/6} \leq e^{-K\sigma/6}\epsilon. \tag{B.4}$$

Furthermore, Ψ is close to the identity, in the sense that, for any $I \in \mathcal{D}_{\Lambda,\varrho'/2,\sigma/6}$, one has

$$|\Pi_I \Psi - I|_2 \leq 4 \frac{K}{\alpha} \epsilon \ , \tag{B.5}$$

where Π_I denotes the projection on the action variables.

Lemma B.2 (Pöschel's Covering Lemma)**.** *Given positive parameters $p, \mathtt{b}, \gamma_0$ and $K \geq 1$, fix a real constant $A \geq \frac{pM}{\mathtt{b}\mu} + \sqrt{2}$, with M and μ the convexity constants defined in* (1.6) *and definision 1.2. Moreover, denote $\mathbb{Z}^n_K := \{k \in \mathbb{Z}^n : |k| \leq K\}$ and, for each maximal lattice $\Lambda \in \mathbb{Z}^n_K$ with $\dim \Lambda = d \in \{0, \dots, n\}$, set*

$$r_\Lambda := \gamma_0 \frac{A^{d-n} K^{d-n}}{|\Lambda|} \ , \quad \delta_\Lambda := \mathtt{b}\mu r_\Lambda \ , \quad \alpha_\Lambda := pMKr_\Lambda \tag{B.6}$$

where $|\Lambda|$ denotes the fundamental volume of the maximal sub-lattice Λ (which is set to be equal to one for the trivial lattice).

Then, there exists a covering of $B_\infty(0,R)$ by resonance blocks D_Λ such that each block D_Λ is (α_Λ, K) non-resonant modulo Λ and, in case Λ is non-trivial, D_Λ is also δ_Λ-close in euclidean norm to exact Λ-resonances in frequency space.

B.2. **Tools of the periodic averaging method.** Let now I_0 be the action corresponding to a T-periodic frequency vector $\omega \in \mathbb{R}^n \setminus \{0\}$, that is

$$\omega := \frac{\partial \mathbf{h}}{\partial I}(I_0) \quad , \qquad T > 0 : T\omega \in \mathbb{Z}^n \, ,$$

and for $\rho, \sigma > 0$ let

$$\mathcal{B}_{\varrho,\sigma}(I_0, \varrho) := \{(I, \theta) \in \mathbb{C}^n : |I - B_2(I_0, \varrho)|_2 < \varrho \ , \ \ \theta \in \mathbb{T}^n_\sigma\} \ ,$$

and

$$M := \sup_{B_{2,\rho}(I_0,\rho)} \left\| \partial^2 h(I) \right\|_{op} , \qquad E := ||\mathbf{h}||_{C^0(B_{2,\varrho}(I_0,\varrho))} \, . \tag{B.7}$$

Lemma B.3 (Lochak's normal form [29])**.** *Let $\mathbf{H}(I, \theta) = \mathbf{h}(I) + \mathbf{f}(I, \theta)$ be analytic in the complex neighborhood*[21] *$\mathcal{B}_{\varrho,\sigma}(I_0, \varrho)$. If for some $\xi > 1$ and $m \in \mathbb{N}$*

$$||\mathbf{f}||_{C^0(\mathcal{B}_{\varrho,\sigma}(I_0,\varrho))} \leq \epsilon E \leq \min\left\{ \frac{1}{72\ \xi} \frac{\sigma \varrho}{T} \ , \ \ \frac{M}{20} \varrho^2 \right\} \quad , \qquad \frac{m \varrho T}{\sigma} \leq \frac{3 \times 10^{-3}}{M} \tag{B.8}$$

then there exists a real-analytic, symplectic transformation

$$\Psi : \mathcal{B}_{\varrho/6,\sigma/6}(I_0, \varrho) \longrightarrow \mathcal{B}_{\varrho,\sigma}(I_0, \varrho)$$

taking $\mathbf{H}$ *into resonant normal form, that is*

$$\mathbf{H} \circ \Psi = \mathbf{h} + \mathbf{g} + \mathbf{f}^* \ , \ \ \{\mathbf{h}, \mathbf{g}\} = 0 \ . \tag{B.9}$$

Moreover, we have the estimates

$$||\mathbf{g}||_{C^0(\mathcal{B}_{\varrho,\sigma}(I_0,\varrho))} \leq \left(1 + 324 \frac{MT\varrho}{\sigma}\right) \epsilon E \ , \ \ ||\mathbf{f}^*||_{C^0(\mathcal{B}_{\varrho,\sigma}(I_0,\varrho))} \leq 324 \frac{MT\varrho}{\sigma} \ 2^{-m} \epsilon E. \tag{B.10}$$

Furthermore, Ψ is close to the identity, in the sense that, for any $(I, \theta) \in U_{\varrho/6,\sigma/6}(I_0, \varrho)$, one has

$$|\Pi_I \Psi - I|_2 \leq \frac{\varrho}{6\xi} \ , \ \ |\Pi_\theta \Psi - \theta|_2 \leq \frac{\sigma}{6} \ , \tag{B.11}$$

where Π_I and Π_θ denote the projectors on the action and angle variables respectively.

Theorem B.1 (Dirichlet)**.** *For any real number $Q > 1$ and any vector $\alpha \in \mathbb{R}^n$ there exist an integer $1 \leq u < Q$ and $\zeta \in \mathbb{Z}^n$ such that*

$$\left| \alpha - \frac{\zeta}{u} \right|_2 \leq \frac{\sqrt{n}}{u Q^{\frac{1}{n}}} \ . \tag{B.12}$$

Corollary B.1. *Take $Q > 1$ and $\Omega \in \mathbb{R}^n$, with $|\Omega|_\infty = w > 0$. Then there exist an integer $1 \leq u < Q$ and a vector $\omega \in \mathbb{Q}^n$ of period $T = u/w$ such that*

$$|\Omega - \omega|_2 \leq \frac{\sqrt{n-1}}{T Q^{\frac{1}{n-1}}} \ . \tag{B.13}$$

A proof of such Corollary can be found in [29], p.30.

[21] Notice that in such framework the radius around the action I_0 is chosen to be equal to the analyticity width. Such choice is arbitrary but helpful in simplifying expressions.

References

[1] D. Bambusi. Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations. Math. Z., 230:345–387, 1999.

[2] D. Bambusi and B. Langella. A c^∞ nekhoroshev theorem. Mathematics in Engineering, 3(2):1–17, 2020.

[3] S. Barbieri. On the algebraic properties of exponentially stable integrable hamiltonian systems. Annales de la Faculté des Sciences de Toulouse, Accepted.

[4] S. Barbieri and L. Niederman. Sharp Nekhoroshev estimates for the three-body problem around periodic orbits. Journal of Differential Equations, 268(7):3749–3780, 2020.

[5] G. Benettin and G. Gallavotti. Stability of motions near resonances in quasi-integrable Hamiltonian systems. Journal of statistical physics, (44):293–338, 1986.

[6] M. Berti, L. Biasco, and P. Bolle. Drift in phase space: a new variational mechanism with optimal diffusion time. J. Math. Pures Appl. (9), 82(6):613–664, 2003.

[7] A. Bounemoura. Nekhoroshev estimates for finitely differentiable, quasi-convex hamiltonian systems. Journal of Differential Equations, 249(11):2905–2920, 2010.

[8] A. Bounemoura. Effective Stability for Gevrey and Finitely Differentiable Prevalent Hamiltonians. Communications in Mathematical Physics, 307:157–183, 2011.

[9] A. Bounemoura, B. Fayad, and L. Niederman. Nekhoroshev estimates for steep real-analytic elliptic equilibrium points. Nonlinearity, (33):1–33, 2020.

[10] A. Bounemoura, B. Fayad, and L. Niederman. Super-exponential stability for generic real-analytic elliptic equilibrium points. Advances in Mathematics, 366:1–30, 2020.

[11] A. Bounemoura and J. Féjoz. Hamiltonian perturbation theory for ultra-differentiable functions. Memoirs of the American Mathematical Society, 2018.

[12] A. Bounemoura and J.-P. Marco. Improved exponential stability for near-integrable quasi-convex Hamiltonians. Nonlinearity, 24(1):97–112, 2011.

[13] A. Bounemoura and L. Niederman. Generic Nekhoroshev theory without small divisors. Annales de l'Institut Fourier, (62):277–324, 2012.

[14] T. Castan. Stability in the plane, planetary three-body problem. PhD Thesis, Université Paris VI, 2017.

[15] A. Celletti and L. Chierchia. KAM stability and Celestial Mechanics. Memoirs of the American Mathematical Society, 187(878), 2007.

[16] L. Chierchia. Kam lectures. Pubbl. Cent. Ric. Mat. Ennio De Giorgi, pages 1–55, 2003.

[17] B. Chirikov. A universal instability of many-dimensional oscillator systems. Phys. Rep., 52(5):264–379, 1979.

[18] H. S. Dumas. Ergodization rates for linear flow on the torus. J. Dynam. Differential Equations, 3(4):593–610, 1991.

[19] F. Fassò, M. Guzzo, and G. Benettin. Nekhoroshev-Stability of Elliptic Equilibria ofHamiltonian Systems. Communications in Mathematical Physics, 197:347–360, 1998.

[20] G. Gallavotti. Stability near resonances in classical mechanics. Helv. Phys. Acta, 59(2):278–291, 1986.

[21] A. Giorgilli, U. Locatelli, and M. Sansottera. Kolmogorov and Nekhoroshev theory for the problem of three bodies. Cel. Mech and Dyn. Astr., 104:159–175, 2009.

[22] M. Guzzo. The Nekhoroshev theorem and long term stabilities in the Solar System. Serbian Astronomical Journal, 190:1–10, 2015.

[23] M. Guzzo, L. Chierchia, and G. Benettin. The Steep Nekhoroshev's Theorem. Commun. Math. Phys., 342:569–601, 2016.

[24] J. E. Littlewood. The Lagrange Configuration in Celestial Mechanics. Proc. London Math. Soc. (3), 9(4):525–543, 1959.

[25] P. Lochak. Canonical perturbation theory via simultaneous approximation. Russian Mathematical Surveys, 47(6):57–133, 1992.

[26] P. Lochak. Hamiltonian perturbation theory: periodic orbits, resonances and intermittency. Nonlinearity, (6):885–904, 1993.

[27] P. Lochak and J.-P. Marco. Diffusion times and stability exponents for nearly integrable analytic systems. Cent. Eur. J. Math., 3(3):342–397, 2005.

[28] P. Lochak, J.-P. Marco, and D. Sauzin. On the splitting of invariant manifolds in multidimensional near-integrable Hamiltonian systems. (163):viii+145, 2003.

[29] P. Lochak, A. Neishtadt, and L. Niederman. Stability of nearly integrable hamiltonian systems over exponentially long times. Proceedings of the 1991 Euler institute conference on dynamical systems. Birkhäuser Basel, 1994.

[30] J-P. Marco and D. Sauzin. Stability and instability for gevrey quasi-convex near-integrable hamiltonian systems. Publ. Math. Inst. Hautes Études Sci., 96:199–275, 2003.

[31] N. N. Nekhoroshev. An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. Russian Mathematical Surveys, 32(6):1–65, 1977.

[32] N. N. Nekhorošev. Stable lower estimates for smooth mappings and for the gradients of smooth functions. Mat. Sb. (N.S.), 90(132):432–478, 480, 1973.

[33] L. Niederman. Stability over exponentially long times in the planetary problem. Nonlinearity, 9:1703–1751, 1996.

[34] L. Niederman. Nonlinear stability around an elliptic equilibrium point in a Hamiltonian system. Nonlinearity, 11(6):1465–1479, 1998.
[35] L. Niederman. Exponential stability for small perturbations of steep integrable Hamiltonian systems. Ergodic Theory and Dynamical Systems, 24:593–608, 2004.
[36] L. Niederman. Hamiltonian stability and subanalytic geometry. Annales de l'Institut Fourier, 56(3):795–813, 2006.
[37] L. Niederman. Prevalence of exponential stability among nearly-integrable hamiltonian systems. Ergodic Theory and Dynamical Systems, 27:905–928, 2007.
[38] L. Niederman. Corrigendum on the article: exponential stability for small perturbations of steep integrable Hamiltonian systems. Ergodic Theory and Dynamical Systems, 3:1055–1056, 2014.
[39] J. Pöschel. Nekhoroshev estimates for quasi-convex hamiltonian systems. Math Z., (213):187–216, 1993.
[40] J. Pöschel. On Nekhoroshev estimates for a nonlinear Schrödinger equation and a theorem by Bambusi. Nonlinearity, 12(6):1587–1600, 1999.
[41] D. A. Salamon. The Kolmogorov-Arnold-Moser theorem. Math. Phys. Electron. J., 10:Paper 3, 37, 2004.
[42] G. Schirinzi and M. Guzzo. On the formulation of new explicit conditions for steepness from a former result of N.N. Nekhoroshev. Journal of Mathematical Physics, 54(7):1–23, 2013.
[43] J. Zhang and K. Zhang. Improved stability for analytic quasi-convex nearly integrable systems and optimal speed of Arnold diffusion. Nonlinearity, 30(7):2918–2929, 2017.

Université Paris-Saclay and Università degli Studi Roma Tre
Email address: santiago.barbieri@universite-paris-saclay.fr

Sorbonne Université
Email address: jean-pierre.marco@imj-prg.fr

Università degli Studi Roma Tre
Email address: jmassetti@mat.uniroma3.it