HAL open science

Generic Nekhoroshev theory without small divisors

Abed Bounemoura, Laurent Niederman

To cite this version:

Abed Bounemoura, Laurent Niederman. Generic Nekhoroshev theory without small divisors. Annales de l'Institut Fourier, 2012, 62 (1), pp.277-324. hal-00441999v2

HAL Id: hal-00441999

https://hal.science/hal-00441999v2

Submitted on 7 Nov 2010

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.

Generic Nekhoroshev theory without small divisors.

Abed Bounemoura * and Laurent Niederman †

November 7, 2010

Dedicated to the memory of N.N. Nekhoroshev (1946-2008)

Abstract

In this article, we present a new approach of Nekhoroshev's theory for a generic unperturbed Hamiltonian which completely avoids small divisors problems. The proof is an extension of a method introduced by P. Lochak, it combines averaging along periodic orbits with simultaneous Diophantine approximation and uses geometric arguments designed by the second author to handle generic integrable Hamiltonians. This method allows to deal with generic non-analytic Hamiltonians and to obtain new results of generic stability around linearly stable tori.

1 Introduction

1. In this article, we are concerned with the stability properties of near-integrable analytic Hamiltonian systems. According to a classical theorem of Liouville-Arnold (see [AKN06]), such systems are locally governed by a Hamiltonian of the form

$$\begin{cases} H(\theta, I) = h(I) + f(\theta, I) \\ |f| < \varepsilon \ll 1 \end{cases}$$

where $(\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n$ are action-angle coordinates for h and f is a small perturbation in some suitable topology. For the integrable system, that is when $f = 0$, the action variables of solutions are trivially constant for all times, but when $f \neq 0$ they are no longer constant of motions and we are interested in studying their evolution for long intervals of time.

*Abed.Bounemoura@math.u-psud.fr, Laboratoire Mathématiques d'Orsay & Institut Mathématiques de Jussieu

†Laurent.Niederman@math.u-psud.fr, Laboratoire Mathématiques d'Orsay & IMCCE, Observatoire de Paris

2. But first it is important to understand the integrable case. When $H = h$ depends only on the action variables, as the latter are constant for all times, the phase space is trivially foliated into invariant tori $\mathcal{T}_{I_0} = \mathbb{T}^n \times \{I_0\}$, for $I_0 \in \mathbb{R}^n$, and on each torus $\mathcal{T}_{I_0}$ the flow is quasi-periodic with frequency vector $\omega_0 = \nabla h(I_0) \in \mathbb{R}^n$. The dynamics of such a flow is completely understood and depends on the frequency vector ω_0, more precisely on its resonant module

$$\mathcal{M}(\omega_0) = \{k \in \mathbb{Z}^n \mid k.\omega_0 = 0\},$$

where the dot denotes the Euclidean scalar product. If $\mathcal{M}(\omega_0)$ is trivial, then the dynamics is minimal and uniquely ergodic. Otherwise, we have a relation of the form $k.\omega_0 = 0$ for some $k \in \mathbb{Z}^n \setminus \{0\}$, which is usually called a resonance, and denoting by m the rank of $\mathcal{M}(\omega_0)$, the torus $\mathcal{T}_0$ splits into a continuous m-parameter family of invariant sub-tori of dimension $n - m$, on which the dynamics is minimal and uniquely ergodic. These are called resonant tori, and in case of maximal resonances (*i.e.* $m = n - 1$ if h does not have critical points), the tori are foliated into periodic orbits. Under some non-degeneracy assumption on h, both resonant and non-resonant tori form a dense subset of the phase space.

3. Returning to the perturbed system, since Poincaré we know that resonant tori do not survive (actually he proved that for a periodic tori, generically only a finite number of periodic orbits persist). But it was a remarkable idea of Kolmogorov ([Kol54]) to focus on non-resonant tori to prove that a set of large measure of invariant tori survives under some regularity and non-degeneracy assumptions. This has now become a rich and vast subject called KAM theory (see [Pös01], [dlL01] or [Bos86] for some nice introductions on this theory). Such tori persist in a $\sqrt{\varepsilon}$-neighbourhood of the unperturbed ones and therefore for a set of large measure of initial conditions, the variation of the actions is of order $\sqrt{\varepsilon}$ for *all time*. But on the other hand, this set of KAM tori is typically a Cantor family (hence with no interior) and the theory gives no information on the complement, except when $n = 2$ where these two-dimensional invariant tori disconnect the three-dimensional energy level leaving *all solutions* stable for *all time*. However for $n \geq 3$, it is still possible to find solutions for which the variation of the action components is of order one. A proof of this fact was outlined by Arnold in his famous paper ([Arn64]) where he proposed a mechanism to produce examples of near-integrable Hamiltonian systems where such a drift occurs no matter how small the perturbation is. This phenomenon is usually referred to Arnold diffusion.

4. Hence for $n \geq 3$, results of stability for near-integrable Hamiltonian systems which are valid for an open set of initial conditions can only be proved over finite times. This picture was completed by Nekhoroshev in the seventies (see [Nek77],[Nek79] and [Nie09] for a recent overview of the

theory) who proved the following: if the system is analytic and the unperturbed Hamiltonian h satisfies some quantitative transversality condition called *steepness*, then there exist positive constants a, b, ε_0, c_1, c_2 and c_3 depending only on h, such that every solution $(\theta(t), I(t))$ of the perturbed system starting at time $t = 0$ satisfies

$$|I(t) - I(0)| \leq c_1 \varepsilon^b, \quad |t| \leq c_2 \exp\left(c_3 \varepsilon^{-a}\right), \tag{1}$$

provided that the size of the perturbation ε is smaller than the threshold ε_0. The constants a and b are called the stability exponents. If property (1) is satisfied, we shall say that the integrable Hamiltonian h is exponentially stable. Hence, KAM and Nekhoroshev's theory yield different type of stability results, but they both ultimately rely on the same tool which is the construction of normal forms, and we shall described it below.

5. The basic idea is to look at a "more integrable" Hamiltonian which yields a good approximation of the perturbed system. By the averaging principle (see [AKN06]), this simpler Hamiltonian is given by the time average of the system along the unperturbed flow, that is

$$[H] = h + [f],$$

where

$$[f] = \lim_{t \to \infty} \left(\frac{1}{t} \int_0^t f \circ \Phi_s^h ds \right),$$

and Φ_s^h is the Hamiltonian flow of the integrable part h. Actually, this average depends on the dynamics of the unperturbed Hamiltonian and hence on resonant modules associated to frequencies. So given a sub-module $\mathcal{M} \subseteq \mathbb{Z}^n$, we define its resonant manifold by

$$S_{\mathcal{M}} = \{ I \in \mathbb{R}^n \mid k.\nabla h(I) = 0 \text{ for } k \in \mathcal{M} \}.$$

Due to the ergodic properties of the linear flow with vector $\nabla h(I)$ over the torus $\mathbb{T}^n$, the time average over $S_{\mathcal{M}}$ equals the space average along a torus of dimension $n - m$ if m is the multiplicity of the resonance (*i.e.* the rank of $\mathcal{M}$), hence $n - m$ angles have been removed in this case. From a physical point of view, the guiding principle is that rapidly oscillating terms discarded in averaging cause only small oscillations which are superimposed to the solutions of the averaged system. In order to prove this claim, one should check that any solution of the perturbed system remains close to the solution of the averaged system with the same initial condition. Especially, this will be the case if one finds a canonical transformation ε-close to identity which conjugates the perturbed Hamiltonian to its average. Hence we are reduced to a problem of *normal form* where one tries to conjugate the system to a simpler one, that is we look for a convenient system of coordinates.

However, constructing such a good system of coordinates is not an easy task. The linearised equation of conjugation reads

$$\{\chi, h\} = f - [f],$$

if χ is the function generating the conjugation. This is usually called a *homological equation* and to solve it we need to invert the linear operator $L_h = \{., h\}$ acting on a suitable space of functions. Here our operator is invertible, but its inverse is generally unbounded: this is the *small divisors* phenomenon. To see this, just note that once an action $I \in S_{\mathcal{M}}$ is fixed (and hence a frequency $\omega = \nabla h(I)$ satisfying $k.\omega \neq 0$ for $k \notin \mathcal{M}$), the homological equation is a just a first-order, linear with constant coefficients partial differential equation on $\mathbb{T}^n$, namely

$$\omega.\nabla\chi = f - [f].$$

Such equations are known to be well-suited for Fourier analysis, in our case the operator L_h is easily diagonalized in a Fourier basis and we find that the eigenvalues are proportional to the scalar products $k.\omega$, for $k \in \mathbb{Z}^n$. More precisely, expanding χ and f as

$$\chi(\theta) = \sum_{k\in\mathbb{Z}^n} \hat{\chi}_k e^{i2\pi k.\theta}, \quad f(\theta) = \sum_{k\in\mathbb{Z}^n} \hat{f}_k e^{i2\pi k.\theta},$$

then

$$[f] = \sum_{k\in\mathcal{M}} \hat{f}_k e^{i2\pi k.\theta},$$

and so formally

$$\hat{\chi}_k = \begin{cases} (i2\pi k.\omega)^{-1}\, \hat{f}_k, \ k \notin \mathcal{M}, \\ 0, \ k \in \mathcal{M}. \end{cases} \tag{2}$$

The scalar products $k.\omega$ appearing in the denominators of (2) are not zero by assumption, but they can be arbitrarily small and this is inevitable for large integers k (see the estimate (3) below). This can cause the divergence of the Fourier series of χ and hence the unboundedness of the inverse of L_h. Classical small divisors techniques are concerned with obtaining lower bounds for these scalar products to ensure the convergence of the series and this leads necessarily to complicated estimates. Furthermore, to obtain a result applying to all solutions, a partition of the phase space into resonant manifolds associated to different modules, usually called the *geometry of resonances*, has to be achieved and this is a delicate task. All these techniques are very important, in particular to study Arnold diffusion and related problems, however we will show that they are not necessary to prove Nekhoroshev's estimates.

6. Indeed, all these problems are completely bypassed if we only average along periodic orbits of the unperturbed flow. We first recall the following definition.

Definition 1.1. *A vector $\omega \in \mathbb{R}^n$ is said to be periodic if there exists a real number $t > 0$ such that $t\omega \in \mathbb{Z}^n$. In this case, the number*

$$T = \inf\{t > 0 \mid t\omega \in \mathbb{Z}^n\}$$

is called the period of ω.

A basic example is given by a vector with rational components, the period of which is just the least common multiple of the denominators of its components. Geometrically, if ω is T-periodic, an invariant torus with a linear flow with vector ω is filled with T-periodic orbits. In this case, the average along such a periodic solution is given by

$$[f] = \lim_{t\to\infty}\left(\frac{1}{t}\int_0^t f\circ\Phi_s^l ds\right) = \frac{1}{T}\int_0^T f\circ\Phi_s^l ds,$$

where l denotes the linear Hamiltonian with frequency ω, that is $l(I) = \omega.I$. Then the homological equation $\{\chi, l\} = f - [f]$ is easily solved without using Fourier expansions and is given by an explicit integral formula

$$\chi = \frac{1}{T}\int_0^T (f - [f])\circ\Phi_s^l s ds.$$

So in this case, there is no small divisors. To understand more concretely the previous sentence, consider a vector $\omega \in \mathbb{R}^n$ and multi-integers k that do not resonate with ω (that is $k \notin \mathbb{Z}^n \cap \omega^\perp$). Then in general we don't have a lower bound on the divisors $k.\omega$ that appears in (2), and by a theorem of Dirichlet one has the upper bound

$$\min_{0<|k|\leq K} |k.\omega| \leq \frac{|\omega|}{K^{n-1}}. \tag{3}$$

In that context, small divisors techniques use Diophantine vectors for which $|k.\omega| \geq \gamma|k|_1^{-\tau}$, with $\gamma > 0$, $\tau \geq n-1$ and where $|\,.\,|_1$ stands for the ℓ^1-norm, but nevertheless the lower bound deteriorates as $|k|_1$ increases, causing extra difficulties (which are usually handled by the so-called ultra-violet cut-off). However if the vector ω is T-periodic, one simply has $|k.\omega| \geq T^{-1}$ and the lower bound is uniform in $|k|_1$.

7. Lochak ([Loc92], see also [LN92] and [LNN94] for refinements) has shown that averaging along the periodic orbits of the integrable Hamiltonian is enough to obtain Nekhoroshev's estimates of stability when the unperturbed Hamiltonian is strictly *convex* (or strictly quasi-convex, that is the Hamiltonian is strictly convex when restricted to its energy sub-levels). Indeed, using convexity, Lochak obtains open sets around periodic orbits over which exponential stability holds. Then, Dirichlet's theorem about *simultaneous Diophantine approximation* ensures easily that these open sets

recover the whole action space and yields the global result, avoiding the difficult geometry of resonances. Put it differently, in the convex case one only needs dynamical informations near resonances of maximal multiplicities, which are completely characterized by periodic orbits.

The goal of this paper is to extend Lochak's approach for a generic set of integrable Hamiltonians. To do so, we will have to analyze the dynamics in a neighbourhood of suitable resonances of any multiplicities by using only successive averagings along periodic orbits together with Dirichlet's theorem, and this will lead to exponential estimates of stability for perturbation of a generic integrable Hamiltonian, as stated below.

Theorem 1.2. *Consider an arbitrary real analytic integrable Hamiltonian h defined on a neighbourhood of a closed ball in $\mathbb{R}^n$. Then for almost any $\xi \in \mathbb{R}^n$, the integrable Hamiltonian $h_\xi(x) = h(I) - \xi.I$ is exponentially stable with the exponents $a = b = 3^{-1}(2n)^{-3n}$.*

This will be a direct consequence of Theorems 2.2 and 2.4, see below in section 2.1. This result is not new, see [Nie07], but the novelty here is our method of proof, which avoids completely the fundamental problem of small divisors and hence all the associated technicalities (non-resonant domains, Fourier series, Fourier norm, ultra-violet cut-off and so on). The analytic part of our proof of Nekhoroshev's estimates is therefore reduced to its bare minimum, it is nothing but a classical one-phase averaging, while our geometric part is based on a clever use of Dirichlet's theorem along each solution. Applications of our method to other problems will be discussed below, in section 2.2.

To conclude this introduction, we point out that the method of averaging along periodic orbits has also been used successfully to re-prove recently some KAM theorems without small divisors (see [KLDM06] and [KLDM07]), even though their techniques are much more complicated.

2 Statement of results

2.1 Set-up and results

1. Let $B = B_R$ be the open ball centered at the origin of $\mathbb{R}^n$ of radius R with respect to the supremum norm, the domain $\mathcal{D} = \mathbb{T}^n \times B$ will be our phase space. To avoid trivial situations, we assume $n \geq 2$. Our Hamiltonian function H is real-analytic and bounded on $\mathcal{D}$ and it admits a holomorphic extension to some complex neighbourhood of $\mathcal{D}$ of the form

$$\mathcal{D}_{r,s} = \{(\theta, I) \in (\mathbb{C}^n/\mathbb{Z}^n) \times \mathbb{C}^n \mid |\mathcal{I}(\theta)| < s,\ d(I, B) < r\},$$

with two fixed numbers $r > 0$, $s > 0$, and where $\mathcal{I}(\theta)$ is the imaginary part of θ, $|.|$ the supremum norm on $\mathbb{C}^n$ and d the associated distance

on $\mathbb{C}^n$. Equivalently, one can start with a Hamiltonian H, defined and holomorphic on $\mathcal{D}_{r,s}$ and which preserves reality, that is H is real-valued for real arguments. Without loss of generality, we may assume that $r < 1$ and $s < 1$. The space of such analytic functions on $\mathcal{D}_{r,s}$, equipped with the supremum norm $|\,.\,|_{r,s}$, is obviously a Banach algebra with respect to the multiplication of functions, and we shall denote it by $\mathcal{A}_{r,s}$.

Our Hamiltonian $H \in \mathcal{A}_{r,s}$ is assumed to be close to integrable, that is of the form

$$\begin{cases} H(\theta, I) = h(I) + f(\theta, I) \\ |f|_{r,s} < \varepsilon \ll 1, \end{cases} \tag{$*$}$$

where h is the integrable part and f a small perturbation. Moreover, the derivatives up to order 3 of h are assumed to be bounded by some constant $M > 1$, that is

$$|\partial^k h(I)| \leq M, \quad 1 \leq |k|_1 \leq 3, \quad I \in B,$$

where $|k|_1 = |k_1| + \cdots + |k_n|$.

2. In order to obtain results of exponential stability, we do need to impose some non-degeneracy condition on the unperturbed Hamiltonian. Let $G(n,k)$ be the set of all vector subspaces of $\mathbb{R}^n$ of dimension k. We equip $\mathbb{R}^n$ with the Euclidean scalar product, $\|\,.\,\|$ stands for the Euclidean norm, and given an integer $L \in \mathbb{N}^*$, we define $G^L(n,k)$ as the subset of $G(n,k)$ consisting of those subspaces whose orthogonal complement can be spanned by vectors $k \in \mathbb{Z}^n$ with $|k|_1 \leq L$.

Definition 2.1. *A function $h \in C^2(B)$ is said to be SDM if there exist $\gamma > 0$ and $\tau \geq 0$ such that for any $L \in \mathbb{N}^*$, any $k \in \{1, \ldots, n\}$ and any $\Lambda \in G^L(n,k)$, there exists $(e_1, \ldots, e_k)$ (resp. $(f_1, \ldots, f_{n-k})$), an orthonormal basis of Λ (resp. of $\Lambda^\perp$), such that the function h_Λ defined on B by*

$$h_\Lambda(\alpha, \beta) = h\left(\alpha_1 e_1 + \cdots + \alpha_k e_k + \beta_1 f_1 + \cdots + \beta_{n-k} f_{n-k}\right),$$

satisfies the following: for any $(\alpha, \beta) \in B$,

$$\|\partial_\alpha h_\Lambda(\alpha, \beta)\| \leq \gamma L^{-\tau} \Longrightarrow \|\partial_{\alpha\alpha} h_\Lambda(\alpha, \beta).\eta\| > \gamma L^{-\tau} \|\eta\|$$

for any $\eta \in \mathbb{R}^n \setminus \{0\}$.

In other words, for any $(\alpha, \beta) \in B$, we have the following alternative: either $\|\partial_\alpha h_\Lambda(\alpha, \beta)\| > \gamma L^{-\tau}$ or $\|\partial_{\alpha\alpha} h_\Lambda(\alpha, \beta).\eta\| > \gamma L^{-\tau} \|\eta\|$ for any $\eta \in \mathbb{R}^n \setminus \{0\}$. This technical definition, which is a slight variation of a notion introduced in [Nie07], is basically a quantitative transversality condition which is stated in adapted coordinates. It is inspired on the one hand by the steepness condition introduced by Nekhoroshev ([Nek77]) where one has to look at the projection of the gradient map ∇h onto affine subspaces,

and on the other hand by the quantitative Morse-Sard theory of Yomdin ([Yom83], [YC04]) where critical or "nearly-critical" points of h have to be quantitatively non degenerate. The abbreviation SDM stands for "Simultaneous Diophantine Morse" functions, and we refer to Appendix B for more explanations on this condition and some justifications on the latter terminology.

3. The set of SDM functions on B with respect to $\gamma > 0$ and $\tau \geq 0$ will be denoted by $SDM^\tau_\gamma(B)$, and we will also use the notations

$$SDM^\tau(B) = \bigcup_{\gamma>0} SDM^\tau_\gamma(B), \quad SDM(B) = \bigcup_{\tau\geq 0} SDM^\tau(B).$$

The following result states that SDM functions are generic among sufficiently smooth functions.

Theorem 2.2. *Let $\tau > 2(n^2+1)$ and $h \in C^{2n+2}(B)$. Then for Lebesgue almost all $\xi \in \mathbb{R}^n$, the function $h_\xi(I) = h(I) - \xi.I$ belongs to $SDM^\tau(B)$.*

More precisely, there is a good notion of "full measure" in an infinite dimensional vector space, which is called prevalence (see [OY05] and [HK10] for nice surveys), and the previous theorem immediately gives the following result.

Corollary 2.3. *For $\tau > 2(n^2+1)$, $SDM^\tau(B)$ is prevalent in $C^{2n+2}(B)$.*

4. Now we can state the main result of the paper.

Theorem 2.4. *Let H as in $(*)$ and assume that the integrable part h belongs to $SDM^\tau_\gamma(B)$ with $\tau \geq 2$ and $\gamma \leq 1$. Then there exist positive constants a and b depending only on n and τ, and ε_0 depending only on h, such that if $\varepsilon \leq \varepsilon_0$, for every initial action $I(0) \in B_{R/2}$ the following estimates*

$$|I(t) - I(0)| < (n+1)^2\varepsilon^b, \quad |t| < \exp(\varepsilon^{-a}),$$

hold true.

More precisely, we can choose the exponents

$$a = b = 3^{-1}(2(n+1)\tau)^{-n},$$

and ε_0 depending on the whole set of parameters n, R, r, s, M, γ and τ, but no efforts was made to improved the stability exponents since the optimality of the constants involved is not our goal. Actually, this optimality is not relevant for generic integrable Hamiltonians.

Let us add that the only property used on the integrable part h to derive these estimates is a specific steepness property, therefore the proof is also valid, and in fact simpler, assuming the original steepness condition

of Nekhoroshev (see Appendix B). However, note that this is precisely this "weaker" genericity assumption that allows new results of stability near linearly stable invariant tori (see [Bou09]).

We emphasized again that this is not the result itself, but the method of proof which is new and leads to many improvements as we explain below.

2.2 Comments and prospects

To conclude this section we mention other problems for which our method should apply, mainly the study of elliptic fixed points, Nekhoroshev's estimates in lower regularity and finally estimates in large or infinite dimensional Hamiltonian systems. In all these topics, the method of periodic averagings have already proved to be very useful.

5. First our analytic arguments are very intrinsic and this is important in the study of the stability of elliptic fixed points in Hamiltonian systems. Actually, in this case the transformation in action-angle variables (via the symplectic polar coordinates) admits singularities which do not allow to derive directly stability results from Nekhoroshev's theory. In the convex case, this problem has been overcomed independently by Fassò, Guzzo and Benettin([FGB98]) and by Niederman ([Nie98]). Both use Cartesian coordinates, the first study uses the classical approach and adapted Fourier expansions while the second one relies on periodic averagings and simultaneous Diophantine approximation. The latter proof was clarified by Pöschel ([Pös99b]). With our approach, we can remove the convexity hypothesis to have exponential stability around an elliptic fixed point under a generic assumption on the non-linear part. Furthermore, assuming a Diophantine condition on the normal frequency it is well-known since Morbidelli and Giorgilli ([MG95]) that one can even obtain super-exponential stability by combining a sufficiently large number of Birkhoff normalizations with Nekhoroshev's estimates. Here, with our method generic results of super-exponential stability around elliptic fixed points are also available, and similarly around invariant Diophantine Lagrangian tori and even isotropic reducible linearly stable tori. All this results are contained in [Bou09].

6. Furthermore, one should mention that periodic averagings are well-suited for non-analytic Hamiltonians and our formalism should also carry on in this context. The advantage of periodic averagings is clear already at the linear level when solving the homological equation: if the system is of finite differentiability, then for a Diophantine frequency vector the solution of the homological equation is subjected to a disastrous loss of derivatives (larger than the number of degrees of freedom) and one has to use rather cumbersome Fourier expansions, while for a periodic frequency vector, this loss of derivatives is minimal and one can use a more elegant integral formula. Hence in finite differentiability, for a convex or generic unperturbed

Hamiltonian system, we can expect a proof of stability estimates (with of course a polynomial bound on the time of stability) which is both simple (no small divisors) and direct (no need to use the result in the analytic case and smoothing techniques, which is the usual approach in KAM theory in finite differentiability). Note that the analyticity of the studied system is only needed for the construction of normal forms up to an exponentially small remainder, but our steepness condition is generic for Hamiltonians of *finite* but sufficiently high regularity. Concerning Gevrey regularity, Marco and Sauzin ([MS02]) have already proved exponential estimates of stability in the convex case and for the C^k regularity, polynomial estimates of stability are indeed available (see [Bou10]). Both results use only periodic averagings, so with our method they should also hold for a generic integrable Hamiltonian. It can also be noticed that the analytical properties of the expansions arising in periodic averagings are accurately known ([Nei84],[RS96]).

7. Finally, results of stability for large Hamiltonian systems as a model for statistical mechanics have been obtained by Bambusi and Giorgilli ([BG93]) and Bourgain ([Bou04]), and for non-linear evolution PDE seen as an infinite dimensional Hamiltonian system mostly by Bambusi ([Bam99], [BN02]) and then clarified by Pöschel ([Pös99a]). All these works use Lochak's approach in the convex case. We believe that our method should allow to remove the convexity assumption in those results to obtain more general statements.

8. The paper is organized as follows. In the next section, we state our normal form and explain the main ideas, and then we give the proof of Theorem 2.4. The complete proof of the normal form is deferred to Appendix A, and in Appendix B we collect the basic properties of SDM functions that we shall need and we prove Theorem 2.2 and Corollary 2.3.

9. In the text, we shall adopt the following notation taken from [Pös99b]: we will write $u \lessdot v$ if there exists a constant $C \geq 1$ such that $u < Cv$, where C depends only on n, R, r, s, M, but not on τ and on the small parameters ε and γ. Similarly, we will use the notations $u \cdot\!< v$, $u =\!\cdot\, v$ and $u \cdot\!= v$.

We shall use the following norms for vectors $v \in \mathbb{R}^n$ or $v \in \mathbb{C}^n$: $|\,.\,|$ will be the supremum norm, $|\,.\,|_1$ the ℓ^1-norm and $\|\,.\,\|$ the Euclidean (or Hermitian) norm.

3 Proof of Theorem 2.4

In this section, we consider the Hamiltonian $(*)$, that is

$$\begin{cases} H(\theta, I) = h(I) + f(\theta, I) \\ |f|_{r,s} < \varepsilon \end{cases}$$

with $H \in \mathcal{A}_{r,s}$. As usual, the proof of exponential stability estimates splits into an analytic part and a geometric part.

The analytic part is contained in section 3.1. It consists in the construction normal forms on a neighbourhood of specific resonances, that is suitable coordinates which display the relevant part of the perturbation on such a neighbourhood. Basically, we will reduce the perturbation to a so-called resonant term which is dynamically significant, and a general term which will only cause exponentially small deviations.

The geometric part is expanded in section 3.2, and it is mainly based on the properties of the underlying integrable system. The strategy will be first to defined a class of solutions, which we call *restrained*, and for which it is obvious from our normal forms that they are stable for an exponentially long time. Using this intermediate result, we will then show that all solutions are in fact exponentially stable, and our main tools to do this will be an adapted steepness property satisfied by our integrable system, as well as a basic theorem of Dirichlet on simultaneous Diophantine approximation.

3.1 Analytical part

1. Let us begin by describing the neighbourhoods of resonances we will consider. Given a sequence of linearly independent periodic vectors $(\omega_1, \dots, \omega_n)$, with periods $(T_1, \dots, T_n)$, we define in the complex phase space, for $j \in \{1, \dots, n\}$, the domains

$$\mathcal{D}_{r_j,s_j}(\omega_j) = \{(\theta, I) \in \mathcal{D}_{r_j,s_j} \mid |\nabla h(I) - \omega_j| \lessdot r_j\},$$

with two sequences $(r_1, \dots, r_n)$ and $(s_1, \dots, s_n)$.

Remark 3.1. *It is important to note that there is an implicit constant in the previous definition represented by the dot, and we will not make it explicit in order to avoid cumbersome and meaningless expressions. We just mention that it depends only on n, M and $j \in \{1, \dots, n\}$ and for subsequent arguments it has to be chosen sufficiently large.*

Informally, one has to view the domain $\mathcal{D}_{r_j,s_j}(\omega_j)$ as a neighbourhood, in frequency space, of a periodic torus with a linear flow of frequency ω_j. Such domains will therefore be called *nearly-periodic* tori. We will also use the real part of those domains, which are $\mathbb{T}^n \times \mathcal{B}_{r_j}(\omega_j)$ where

$$\mathcal{B}_{r_j}(\omega_j) = \{I \in B_{r_j} \mid |\nabla h(I) - \omega_j| \lessdot r_j\},$$

with $B_{r_j} = \{I \in \mathbb{R}^n \mid d(I, B) < r_j\}$.

2. Given an analytic function f defined on $\mathcal{D}_{r_j,s_j}(\omega_j)$, we simply denote its supremum norm by

$$|f|_{r_j,s_j} = |f|_{\mathcal{D}_{r_j,s_j}(\omega_j)}.$$

For vector-valued functions, this definition is extended component-wise, that is

$$|\partial_\theta f|_{r_j,s_j} = \max_{1 \le i \le n} |\partial_{\theta_i} f|_{r_j,s_j}, \ |\partial_I f|_{r_j,s_j} = \max_{1 \le i \le n} |\partial_{I_i} f|_{r_j,s_j}.$$

We will write l_j for the linear integrable Hamiltonian with frequency ω_j, that is $l_j(I) = \omega_j.I$ for $j \in \{1, \dots, n\}$. For any function f, we will denote $[f]_j$ its average along the periodic flow generated by l_j, that is

$$[f]_j = \frac{1}{T_j} \int_0^{T_j} f \circ \Phi_s^{l_j} ds.$$

3. Our interest here is to obtain normal forms on nearly-periodic tori up to an exponentially small remainder with respect to some parameter $m \in \mathbb{N}$, that we will choose later of order ε^{-1} (during the proof of Theorem 2.4). To this end, we will need the following conditions (A_j), for $j \in \{1, \dots, n\}$, where (A_1) is

$$\begin{cases} mT_1\varepsilon \cdot\!< r_1^2,\ mT_1 r_1 \cdot\!< s_1,\ 0 < r_1 <\!\cdot s_1, \\ \mathcal{B}_{r_1}(\omega_1) \neq \emptyset,\ r_1 \cdot\!< r,\ s_1 \cdot\!< s, \end{cases} \tag{A_1}$$

and for $j \in \{2, \dots, n\}$, (A_j) is

$$\begin{cases} mT_j\varepsilon \cdot\!< r_1 r_j,\ mT_j r_j \cdot\!< s_j,\ 0 < r_j <\!\cdot s_j, \\ \mathcal{B}_{r_j}(\omega_j) \neq \emptyset,\ \mathcal{D}_{r_j,s_j}(\omega_j) \subseteq \mathcal{D}_{2r_{j-1}/3, 2s_{j-1}/3}(\omega_{j-1}). \end{cases} \tag{A_j}$$

Let us explain briefly our assumptions.

First, the condition on the inclusion of nearly-periodic tori is really crucial. Indeed, since ω_1 is periodic, the nearly-periodic torus $\mathcal{D}_{r_1,s_1}(\omega_1)$ describes a neighbourhood of a resonance of multiplicity $n-1$. Now for $j \in \{2, \dots, n\}$, since $(\omega_1, \dots, \omega_j)$ are periodic and independent, the inclusion assumption, together with the non triviality assumption, imply that the nearly-periodic torus $\mathcal{D}_{r_j,s_j}(\omega_j)$ also describes a neighbourhood of a resonance, but of multiplicity $n-j$. Note that such a condition will put an important restriction on our choice of the sequence $(\omega_1, \dots, \omega_n)$ as they will have to be sufficiently close to each other to ensure these inclusions.

Then, the condition on our parameter $m \in \mathbb{N}$,

$$mT_j r_j \cdot\!< s_j, \quad j \in \{1, \dots, n\},$$

is also important as it will later determine m in terms of ε and hence the precise size of the exponentially small term.

Finally, the other conditions are only technical (and will be easily arranged in the sequel), as they only give smallness conditions on ε.

4. Our normal form is described in the next proposition.

Proposition 3.2. *Consider $H = h + f$ as in $(*)$ and let $j \in \{1, \dots, n\}$. If (A_i) is satisfied for any $i \in \{1, \dots, j\}$, then there exists an analytic symplectic transformation*

$$\Psi_j : \mathcal{D}_{2r_j/3, 2s_j/3}(\omega_j) \to \mathcal{D}_{r_1,s_1}(\omega_1)$$

such that

$$H \circ \Psi_j = h + g_j + f_j,$$

with $\{g_j, l_i\} = 0$ *for* $i \in \{1, \dots, j\}$ *and the estimates*

$$|\partial_\theta g_j|_{2r_j/3, 2s_j/3} \lessdot \varepsilon, \quad |\partial_\theta f_j|_{2r_j/3, 2s_j/3} \lessdot e^{-m}\varepsilon.$$

Moreover, we have $\Psi_j = \Phi_1 \circ \dots \circ \Phi_j$ *with*

$$\Phi_i : \mathcal{D}_{2r_i/3, 2s_i/3}(\omega_i) \to \mathcal{D}_{r_i, s_i}(\omega_i)$$

such that $|\Phi_i - \mathrm{Id}|_{2r_i/3, 2s_i/3} \lessdot r_i$, *for* $i \in \{1, \dots, j\}$.

The proof of Proposition 3.2 goes by induction, it is not difficult but quite long, and so it is deferred to Appendix A. Here we will try to give a sketch in the case $j = 2$, explaining the main ideas without any technicalities.

The first step is to prove the case $j = 1$, that is to find a transformation Ψ_1 such that $H \circ \Psi_1 = h + g_1 + f_1$ with $\{g_1, l_1\} = 0$ and f_1 exponentially small with m. This is very classical. First observe that we can write our original Hamiltonian as $H = h + g^0 + f^0$, where $g^0 = 0$ trivially satisfies $\{g^0, l_1\} = 0$ and $f^0 = f$ is order ε. Now it is easy to produce a transformation φ^0 such that $H \circ \varphi^0 = h + g^1 + f^1$, with $\{g^1, l_1\} = 0$, but thanks to our assumption (A_1) the remainder f^1 can be made smaller, of order $e^{-1}\varepsilon$: this is an averaging process, $g_1 = [f^0]_1$ and the remainder is estimated by Cauchy inequality. Now we only have to iterate this process m times, and writing $\Psi_1 = \Phi_1 = \varphi^0 \circ \dots \varphi^{m-1}$, $g_1 = g^m$ and $f_1 = f^m$, we end up with $H \circ \Psi_1 = h + g_1 + f_1$ with the required properties.

For the second step, we use the first one and consider $H \circ \Psi_1 = h + g_1 + f_1$ which, by our assumption on the inclusion of domains (this is part of (A_2)), is also defined on $\mathcal{D}_{r_2, s_2}(\omega_2)$. We can forget for a moment about f_1 which is already exponentially small and consider g_1 as the new perturbation. Now as in the first step, we can construct a transformation Φ_2 such that $(h + g_1) \circ \Phi_2 = h + g_2 + f^2$ with $\{g_2, l_2\} = 0$ and f^2 is exponentially small: we start with $h + g_1 = h + g_1^0 + f_1^0$, where $g_1^0 = 0$, $f_1^0 = g_1$ and we find φ^1 such that $(h + g_1) \circ \varphi^1 = h + g_1^1 + f_1^1$ where $g_1^1 = [f_1^0]_2$. After m iterations we finally have $g_2 = g_1^m$ and $f^2 = f_1^m$. Assuming we still have $\{g_2, l_1\} = 0$, the conclusion follows: let $\Psi_2 = \Psi_1 \circ \Phi_2 = \Phi_1 \circ \Phi_2$ and $f_2 = f^2 + f_1 \circ \Phi_2$, then $H \circ \Psi_2 = h + g_2 + f_2$ has the desired properties.

So it remains to explain why $\{g_2, l_1\} = 0$. The key observation is the following: if g_1 satisfies $\{g_1, l_1\} = 0$, then

$$[g_1]_2 = \frac{1}{T_2} \int_0^{T_2} g_1 \circ \Phi_s^{l_2} ds$$

and

$$\chi = \frac{1}{T_2} \int_0^{T_2} (g_1 - [g_1]_2) \circ \Phi_s^{l_2} s ds$$

also satisfy $\{[g_1]_2, l_1\} = 0$ and $\{\chi, l_1\} = 0$. In Appendix A, this will be done by direct computations, but this is in fact a more general phenomenon in normal form theory and it is not restricted to the situation we consider here. Indeed, since $\{l_1, l_2\} = 0$, the linear operators $L_{l_1} = \{., l_1\}$ and $L_{l_2} = \{., l_2\}$ commutes, so that the kernel of L_{l_1} is invariant by L_{l_2}, and as L_{l_2} is semi-simple, it is also invariant under the projection onto the kernel of L_{l_2} which is given by the map $[.]_2$. This explains why $\{[g_1]_2, l_1\} = 0$. Now $g_1 - [g_1]_2$ is in the kernel of L_{l_1}, and its unique pre-image by L_{l_2} is given by χ, hence $\{\chi, l_1\} = 0$.

Remark 3.3. *Note that this property was actually used by Bambusi ([Bam99], Lemma 8.4).*

5. Let us now examine the dynamical consequences of our normal form. As usual, it will be used to control the directions, if any, in which the action variables in these new coordinates can actually drift, and we shall come back to our original coordinates at the beginning of section 3.2.

Under the assumptions of Proposition 3.2, consider the Hamiltonian

$$H_j = H \circ \Psi_j = h + g_j + f_j$$

on the domain $\mathcal{D}_{2r_j/3, 2s_j/3}(\omega_j)$. Let $\mathcal{M}_j$ be the $\mathbb{Z}$-module

$$\mathcal{M}_j = \{k \in \mathbb{Z}^n \mid k.\omega_i = 0,\ i \in \{1, \dots, j\}\},$$

whose rank is $n - j$, and $\Lambda_j = \mathcal{M}_j \otimes \mathbb{R}$ the vector space spanned by $\mathcal{M}_j$.

The following lemma is completely obvious using the definition of the Poisson bracket.

Lemma 3.4. *The equality $\{g_j, l_i\} = 0$, for all $i \in \{1, \dots, j\}$, is equivalent to $\partial_\theta g_j \in \Lambda_j$.*

Now consider a solution $(\theta^j(t), I^j(t))$ of H_j with an initial action $I^j(t_j) \in \mathcal{B}_{2r_j/3}(\omega_j)$ for some $t_j \in \mathbb{R}$, and define the time of escape of this solution as the smallest time $\tilde{t}_j \in]t_j, +\infty]$ for which $I^j(\tilde{t}_j) \notin \mathcal{B}_{2r_j/3}(\omega_j)$. The only information we shall use from our normal form is contained in the next proposition.

Proposition 3.5. *Let Π_j be the projection onto the linear subspace Λ_j, then with the previous notations, we have*

$$|I^j(t) - I^j(t_j) - \Pi_j(I^j(t) - I^j(t_j))| < \varepsilon, \quad t \in [t_j, e^m[\cap [t_j, \tilde{t}_j[.$$

In particular,

$$|I^n(t) - I^n(t_n)| < \varepsilon, \quad t \in [t_n, e^m[.$$

Proof. Let $\Pi_j^\perp$ be the projection onto the orthogonal complement of Λ_j, so that $\Pi_j + \Pi_j^\perp$ is the identity and therefore

$$|I^j(t) - I^j(t_j) - \Pi_j(I^j(t) - I^j(t_j))| = |\Pi_j^\perp(I^j(t) - I^j(t_j))|.$$

Now, as long as $t < \tilde{t}_j$, the equations of motion for $H_j = h + g_j + f_j$ and the mean value theorem give

$$|I^j(t) - I^j(t_j)| \leq |t - t_j||\partial_\theta(g_j + f_j)|_{2r_j/3, 2s_j/3}.$$

But $\{g_j, l_i\} = 0$ for $i \in \{1, \ldots, j\}$, so by Lemma 3.4 we have $\partial_\theta g_j \in \Lambda_j$, hence if we first project the equations onto the orthogonal complement of Λ_j we have

$$|\Pi_j^\perp(I^j(t) - I^j(t_j))| \leq |t - t_j||\partial_\theta f_j|_{2r_j/3, 2s_j/3}.$$

Now since $|t - t_j| < e^m$ and $|\partial_\theta f_j|_{2r_j/3, 2s_j/3} \lessdot e^{-m}\varepsilon$, the previous estimate gives

$$|\Pi_j^\perp(I^j(t) - I^j(t_j))| \lessdot \varepsilon,$$

and therefore

$$|I^j(t) - I^j(t_j) - \Pi_j(I^j(t) - I^j(t_j))| \lessdot \varepsilon$$

for $t \in [t_j, e^m[\cap [t_j, \tilde{t}_j[$.

Finally, note that Π_n is identically zero, so that the mean value theorem immediately gives $\tilde{t}_n \geq e^m$ and the estimate

$$|I^n(t) - I^n(t_n)| \lessdot \varepsilon, \quad t \in [t_n, e^m[,$$

follows easily. This concludes the proof. □

The interpretation of the above proposition is the following: if λ_j is the affine subspace passing through $I^j(t_j)$ with direction space Λ_j, then as long as $I^j(t)$ remains in the domain $\mathcal{B}_{r_j}(\omega_j)$, it is ε-close to λ_j for an exponentially long time with respect to m. This means that for that interval of time, there is almost no variation of the action components in the direction transversal to λ_j, so that any potential drift has to occur along that space.

3.2 Geometric part.

In this section we will give the proof of Theorem 2.4 using the method introduced by Niederman in [Nie04] and [Nie07]. Without loss of generality, we will consider only solutions $(\theta(t), I(t))$ starting at time $t_0 = 0$ and evolving in positive time $t > 0$. We will first show that some specific solutions are exponentially stable, but to define them we shall need some extra notations.

6. Consider a sequence of linearly independent periodic vectors $(\omega_1, \ldots, \omega_n)$, with periods $(T_1, \ldots, T_n)$, and two decreasing sequences of real numbers

$(r_1, \dots, r_n)$ and $(s_1, \dots, s_n)$ satisfying conditions (A_j), for $j \in \{1, \dots, n\}$. Recall that from Proposition 3.2 we have a transformation

$$\Psi_j : \mathcal{D}_{2r_j/3, 2s_j/3}(\omega_j) \to \mathcal{D}_{r_1, s_1}(\omega_1), \quad j \in \{1, \dots, n\},$$

such that $\Psi_j = \Phi_1 \circ \cdots \circ \Phi_j$, where

$$\Phi_i : \mathcal{D}_{2r_i/3, 2s_i/3}(\omega_i) \to \mathcal{D}_{r_i, s_i}(\omega_i), \quad i \in \{1, \dots, j\},$$

satisfies the estimate $|\Phi_i - \mathrm{Id}|_{2r_i/3, 2s_i/3} \cdot\!< r_i$.

By construction, our transformations preserve reality so that

$$\Phi_i : \mathbb{T}^n \times \mathcal{B}_{2r_i/3}(\omega_i) \to \mathbb{T}^n \times \mathcal{B}_{r_i}(\omega_i), \quad i \in \{1, \dots, j\},$$

with $|\Phi_i - \mathrm{Id}|_{2r_i/3} \cdot\!< r_i$. In particular, arranging the implicit constant in the previous estimate ensure that the image of $\mathcal{B}_{2r_i/3}(\omega_i)$ under Φ_i contains the smaller domain $\mathcal{B}_{r_i/3}(\omega_i)$. From now on, we shall simply write

$$\mathcal{B}_i = \mathcal{B}_{r_i/3}(\omega_i), \quad i \in \{1, \dots, n\},$$

and for completeness $\mathcal{B}_0 = B$.

Given a solution $(\theta(t), I(t)) \in B$ starting at time $t_0 = 0$, we can define inductively the "averaged" solution $(\theta^i(t), I^i(t))$ for $i \in \{1, \dots, n\}$ by

$$\Phi_i(\theta^i(t), I^i(t)) = (\theta^{i-1}(t), I^{i-1}(t))$$

as long as $I^{i-1}(t) \in \mathcal{B}_i$, with $(\theta^0(t), I^0(t)) = (\theta(t), I(t))$. Moreover, using our estimate on Φ_i we have

$$|I^i(t) - I^{i-1}(t)| \cdot\!< r_i, \quad i \in \{1, \dots, n\}, \tag{4}$$

during that time interval.

7. We can finally make our definition.

Definition 3.6. *Given $r_0 > 0$ and $m \in \mathbb{N}$, a solution $(\theta(t), I(t))$ of the Hamiltonian $(*)$, starting at time $t_0 = 0$, is said to be restrained (by r_0, up to time e^m) if we can find sequences of:*

(1) radii $(r_1, \dots, r_n)$, with $0 < r_n < \cdots < r_1 < r_0$;

(2) widths $(s_1, \dots, s_n)$, with $0 < s_n < \cdots < s_1$;

(3) independent periodic vectors $(\omega_1, \dots, \omega_n)$, with periods $(T_1, \dots, T_n)$;

(4) times $(t_1, \dots, t_n)$, with $0 = t_0 \le t_1 \le \cdots \le t_n \le t_{n+1} = e^m$,

satisfying, for $j \in \{0, \dots, n-1\}$, conditions (A_{j+1}) and the following conditions (B_j) defined by

$$\begin{cases} |I^j(t) - I^j(t_j)| < r_j, & t \in [t_j, t_{j+1}], \\ |\nabla h(I^j(t_{j+1})) - \omega_{j+1}| < r_{j+1}. \end{cases} \tag{B_j}$$

Before explaining this definition, we need to make several remarks. First, for $j \in \{0, \dots, n-2\}$ we will see that the first condition of (B_{j+1}) is well defined by the second condition of (B_j). Furthermore, for $j \in \{0, \dots, n-1\}$ the last condition in (B_j) implies in particular that the set $\mathcal{B}_{j+1}(\omega_{j+1})$ is non-empty so we may remove this assumption from (A_{j+1}). Finally, we can choose the same sequence of widths $(s_1, \dots, s_n)$ for all solutions, therefore we may already fix $s_i \cdot= s$ with a suitable constant and this simplifies some conditions (for instance, the condition $mT_j r_j \cdot< s_j$ appearing in (A_j) will be replaced by $mT_j r_j \cdot< 1$).

We have chosen the word "restrained" because for such a solution the actions $I(t)$ (or some properly normalized actions $I^j(t)$) are forced to pass close to a resonance at the time $t = t_j$, the multiplicity of which decreases as j increases, and moreover the variation of these (normalized) actions is controlled on each time interval $[t_j, t_{j+1}]$. Hence after the time t_n, the actions are in a domain free of resonances and they are easily confined in view of the last part of Proposition 3.5. This is reminiscent of the original mechanism of Nekhoroshev, but the fact that we consider each solution individually will greatly simplify this geometric part.

8. Let us see how the actions of a restrained solution are easily confined for an exponentially long time with respect to m. We shall write

$$\rho_j = r_1 + \cdots + r_j,$$

for $j \in \{1, \dots, n\}$.

Proposition 3.7. *Consider a restrained solution $(\theta(t), I(t))$, with an initial action $I(0) \in B_{R/2}$. If*

(i) $\varepsilon \cdot< r_n$;

(ii) $r_0 \cdot< R$,

then the estimates

$$|I(t) - I(0)| < (n+1)^2 r_0, \quad 0 \le t < e^m,$$

hold true.

Proof. First observe that for each $j \in \{1, \dots, n-1\}$, for $t \in [t_j, t_{j+1}]$ we have

$$|I(t) - I(t_j)| \le |I(t) - I^j(t)| + |I^j(t) - I^j(t_j)| + |I^j(t_j) - I(t_j)|,$$

so the first part of (B_j) and (4) yields

$$|I(t) - I(t_j)| < 2\rho_j + r_j \tag{5}$$

while for $t \in [0, t_1]$, the first part of (B_0) reads

$$|I(t) - I(0)| < r_0. \tag{6}$$

Now let $t \in [0, e^m]$, then $t \in [t_j, t_{j+1}]$ for some $j \in \{0, \dots, n\}$ (recall that $t_{n+1} = e^m$), and we will distinguish three cases.

First assume that $t \in [0, t_1]$, in this case the conclusion follows by (6) since $(n+1)^2 \geq 1$. Now assume $t \in [t_j, t_{j+1}]$ for some $j \in \{1, \dots, n-1\}$, then we can write

$$|I(t) - I(0)| \leq |I(t) - I(t_j)| + \sum_{i=0}^{j-1} |I(t_{i+1}) - I(t_i)|,$$

and by (5) and (6)

$$|I(t) - I(0)| < \sum_{i=1}^{j} (2\rho_i + r_i) + r_0 < (n+1)^2 r_0,$$

since $r_i < r_0$ for $i \in \{1, \dots, j\}$. Finally, assume that $t \in [t_n, t_{n+1}]$, then we can apply the second inequality of Proposition 3.5 and (i) to estimate

$$|I^n(t) - I^n(t_n)| < \varepsilon < r_n,$$

and so

$$|I(t) - I(t_n)| < 2\rho_n + r_n$$

which gives

$$|I(t) - I(0)| < \sum_{i=1}^{n} (2\rho_i + r_i) + r_0 < (n+1)^2 r_0.$$

To conclude, just note $I(0) \in B_{R/2}$ and (ii) ensures that $I(t)$ remains in B_R for $t < e^m$. □

9. Restrained solutions are exponentially stable, and now we will show that this is in fact true for all solutions. However, to use our steepness arguments this will be done quite indirectly, and so it is useful to introduce the following definition.

Definition 3.8. *Given $r_0 > 0$ and $m \in \mathbb{N}$, a solution $(\theta(t), I(t))$ of the Hamiltonian $(*)$, starting at time $t_0 = 0$, is said to be drifting (to r_0, before time e^m) if there exists a time t_* satisfying*

$$|I(t_*) - I(0)| = (n+1)^2 r_0, \quad 0 < t_* < e^m.$$

Of course, this definition makes sense only if $(n+1)^2 r_0 < R/2$. In view of Proposition 3.7, drifting solutions cannot be restrained. However, we will prove below that if such a drifting solution exists, it has to be restrained under some assumptions on r_0, m and ε, which will eventually prove that all solutions are in fact exponentially stable.

More precisely, assuming the existence of a drifting solution, we will construct a sequence of radii $(r_1, \dots, r_n)$, an increasing sequence of times $(t_1, \dots, t_n)$ and a sequence of linearly independent vectors $(\omega_1, \dots, \omega_n)$, with periods $(T_1, \dots, T_n)$ satisfying, for $j \in \{0, \dots, n-1\}$, assumptions (A_{j+1}) and (B_j). All sequences will be built inductively, and we first describe the tools that we shall need.

10. For $j \in \{1, \dots, n\}$, recall that Λ_j is the vector space spanned by

$$\mathcal{M}_j = \{k \in \mathbb{Z}^n \mid k.\omega_i = 0,\ i \in \{1, \dots, j\}\},$$

and that Π_j (resp. $\Pi_j^\perp$) is the projection onto Λ_j (resp. $\Lambda_j^\perp$). Let us define the integer

$$L_j = \sup_{i \in \{1,\dots,j\}} \{|T_i \omega_i|\} \in \mathbb{N}^*, \quad j \in \{1, \dots, n\}.$$

For completeness, we set $\Lambda_0 = \mathbb{R}^n$, $L_0 = 1$ and in this case Π_0 is nothing but the identity. To construct the sequence of times, we will rely on the fact that our integrable part h belongs to $SDM_\gamma^\tau(B)$, so that it satisfies the following steepness property (see Appendix B).

Lemma 3.9. *For $j \in \{0, \dots, n-1\}$, let λ_j be any affine subspace with direction Λ_j, and take $r < 1$. Then for any continuous curve $\Gamma : [0,1] \to \lambda_j \cap B$ with length*

$$|\Gamma(0) - \Gamma(1)| = r \cdot < \gamma L_j^{-\tau},$$

there exists a time $t_* \in [0,1]$ such that

$$\begin{cases} |\Gamma(t) - \Gamma(0)| < r, & t \in [0, t_*], \\ |\Pi_j(\nabla h(\Gamma(t_*)))| \cdot > r^2. \end{cases}$$

Proof. For any $j \in \{1, \dots, n-1\}$, the orthogonal complement of Λ_j is spanned by $\omega_1, \dots, \omega_j$, hence by the integer vectors $T_1\omega_1, \dots, T_j\omega_j$, so that Λ_j belongs to $G^{L_j}(n, n-j)$ with the integer L_j defined above. Therefore one can apply the Proposition B.2 in Appendix B to get the required properties (note that here we are using the supremum norm instead of the Euclidean norm, so the implicit constants are different).

For $j = 0$, $\Gamma : [0,1] \to B = B \cap \mathbb{R}^n$, but since the orthogonal complement of $\mathbb{R}^n$ is trivial one can take $L_0 = 1$. □

11. To construct the sequence of periodic vectors, we shall use the following lemma, which is a straightforward application of Dirichlet's theorem on simultaneous Diophantine approximation (see [Cas57]).

Lemma 3.10. *Given any vector $v \in \mathbb{R}^n$ and any real number $Q > 0$, there exists a T-periodic vector ω satisfying*

$$|v - \omega| \leq T^{-1} Q^{-\frac{1}{n-1}}, \quad |v|^{-1} \leq T \leq Q|v|^{-1}.$$

Proof. Fix any real number $Q > 0$. We can write the vector v, up to reordering its components, as $v = |v|(\pm 1, x)$ with $x \in \mathbb{R}^{n-1}$, and it will be enough to approximate x by a periodic vector. By a theorem of Dirichlet, we can find an integer q, with $1 \leq q < Q$, such that

$$|qx - p| \leq Q^{-\frac{1}{n-1}},$$

for some $p \in \mathbb{Z}^{n-1}$. The vector $q^{-1}p$ is trivially q-periodic, hence the vector $\omega = |v|(\pm 1, q^{-1}p)$ is T-periodic, with $T = |v|^{-1}q$, therefore

$$|v|^{-1} \leq T \leq Q|v|^{-1},$$

and we have the estimate

$$|v - \omega| \leq T^{-1}|qx - p| \leq T^{-1} Q^{-\frac{1}{n-1}}.$$

□

12. Now we can finally prove that drifting solutions are in fact restrained under some assumptions. This will be done inductively, and for technical reasons we separate the first step (Proposition 3.11) from the general inductive step (Proposition 3.12).

Proposition 3.11. *Let $(\theta(t), I(t))$ be a drifting solution. If $r_0 \cdot\!< \gamma$, then there exist a time t_1, a T_1-periodic vector ω_1 and $r_1 =\!\cdot T_1^{-1}\varepsilon^{a_1}$ for some constant a_1, satisfying (B_0). Moreover, we have the estimates*

$$1 \cdot\!< T_1 \cdot\!< \varepsilon^{-a_1(n-1)} r_0^{-2}, \quad 1 \leq L_1 \cdot\!< \varepsilon^{-a_1(n-1)} r_0^{-2}. \tag{7}$$

Proof. We need to construct t_1, ω_1 and r_1 satisfying

(a) $|I(t) - I(0)| < r_0, \quad t \in [0, t_1]$;

(b) $|\nabla h(I(t_1)) - \omega_1| < r_1$,

and the estimate (7). Consider the curve

$$\Gamma_1 : t \in [0, t_*] \longmapsto I(t) \in B \subseteq \mathbb{R}^n.$$

Since we have a drifting solution, we can select $t_0^* \in [0, t_*]$ such that

$$|\Gamma_1(t_0^*) - \Gamma_1(0)| = r_0.$$

Now using the fact that $h \in SDM_\gamma^\tau(B)$ and $r_0 \cdot\!< \gamma$ (recall that $L_0 = 1$), we can apply Lemma 3.9 (the case $j = 0$) to the curve Γ_1 restricted to $[0, t_0^*]$ to find a time $t_1 \in [0, t_0^*]$ for which

$$\begin{cases} |I(t) - I(0)| < r_0, & t \in [0, t_1], \\ |\nabla h(I(t_1))| \cdot\!> r_0^2. \end{cases} \tag{8}$$

The first inequality of (8) gives (a).

Now choose $Q_1 = \varepsilon^{-a_1(n-1)}$, for some constant a_1 yet to be chosen, and apply Lemma 3.10 to approximate $\nabla h(I(t_1))$ by a T_1-periodic vector ω_1, that is

$$|\nabla h(I(t_1)) - \omega_1| \leq T_1^{-1} Q_1^{-\frac{1}{n-1}} = T_1^{-1} \varepsilon^{a_1}. \tag{9}$$

Moreover, since

$$r_0^2 \cdot\!< |\nabla h(I(t_1))| \cdot\!< 1,$$

the period T_1 satisfies the following estimate

$$1 \cdot\!< T_1 \cdot\!< \varepsilon^{-a_1(n-1)} r_0^{-2}. \tag{10}$$

Now choose $r_1 =\!\cdot T_1^{-1} \varepsilon^{a_1}$ so that (9) gives (b). Finally, as $L_1 = |T_1 \omega_1|$ and

$$|\omega_1| \leq |\nabla h(I(t_1))| + |\nabla h(I(t_1)) - \omega_1| \cdot\!< 1$$

we obtain

$$1 \leq L_1 \cdot\!< \varepsilon^{-a_1(n-1)} r_0^{-2} \tag{11}$$

where the lower bound follows from the fact that $T_1 \omega_1$ is a non-zero integer vector. The estimates (10) and (11) give (7). □

Proposition 3.12. *Let $(\theta(t), I(t))$ be a drifting solution, $j \in \{1, \ldots, n-1\}$ and assume that there exist sequences $(t_1, \ldots, t_j)$, $(\omega_1, \ldots, \omega_j)$ linearly independent and $(r_1, \ldots, r_j)$, satisfying assumptions (A_i) and (B_{i-1}), for $i \in \{1, \ldots, j\}$. Assume also that*

(i) $r_j \cdot\!< \min\{r, s\}$;

(ii) $m T_j \varepsilon \cdot\!< r_1 r_j$;

(iii) $m T_j r_j \cdot\!< 1$;

(iv) $\left(T_j r_j L_j^{-1}\right)^\tau \cdot\!< \gamma L_j^{-\tau}$;

(v) $\left(T_j r_j L_j^{-1}\right)^\tau \cdot\!< r_j$;

(vi) $\varepsilon \cdot\!< \left(T_j r_j L_j^{-1}\right)^{2\tau}$;

(vii) $r_1 \cdot\!< r_0^2$.

Then there exist a time t_{j+1}, a T_{j+1}-periodic vector ω_{j+1} and $r_{j+1} =\cdot T_{j+1}^{-1}\varepsilon^{a_{j+1}}$ for some constant a_{j+1}, satisfying (A_{j+1}) and (B_j). Moreover, we have the estimates

$$1 <\cdot T_{j+1} <\cdot \varepsilon^{-a_{j+1}(n-1)} r_0^{-2}, \quad 1 \leq L_{j+1} <\cdot \max_{i\in\{1,\dots,j+1\}} \{\varepsilon^{-a_i(n-1)}\} r_0^{-2}, \quad (12)$$

and if

$(viii)$ $r_{j+1} \cdot< \left(T_j r_j L_j^{-1}\right)^{2\tau}$,

then ω_{j+1} is linearly independent of $(\omega_1, \dots, \omega_j)$.

Proof. First note that for $j = 1$, we do not require that t_1, ω_1 and r_1 satisfy (A_1) since this is implied by the conditions $(i), (ii)$ and (iii), and for $j > 1$, the same conditions reduce assumption (A_{j+1}) to the inclusion of real domains $\mathcal{B}_{r_{j+1}}(\omega_{j+1}) \subseteq \mathcal{B}_{2r_j/3}(\omega_j)$ (recall that by condition (B_{j-1}) these domains are non-empty, and that we have already fixed $s_j \cdot= s$).

Therefore, we need to construct t_{j+1}, ω_{j+1} and r_{j+1} satisfying

(a) $|I^j(t) - I^j(t_j)| < r_j, \quad t \in [t_j, t_{j+1}]$;

(b) $|\nabla h(I^j(t_{j+1})) - \omega_{j+1}| < r_{j+1}$;

(c) ω_{j+1} is independent of $(\omega_1, \dots, \omega_j)$;

(d) $\mathcal{B}_{r_{j+1}}(\omega_{j+1}) \subseteq \mathcal{B}_{2r_j/3}(\omega_j)$,

and the estimates (12).

Let $\tilde{t}_j$ be the maximal time of existence within $\mathcal{B}_j$ of the solution $I^j(t)$ starting at $I^j(t_j)$. Since (A_j) is satisfied, we can apply Proposition 3.5 and for $t \in [t_j, \tilde{t}_j] \cap [t_j, e^m]$, we have

$$|I^j(t) - I^j(t_j) - \Pi_j(I^j(t) - I^j(t_j))| <\cdot \varepsilon. \quad (13)$$

Now consider the curve

$$\Gamma_{j+1} : t \in [t_j, \tilde{t}_j] \cap [t_j, e^m] \longmapsto I^j(t_j) + \Pi_j(I^j(t) - I^j(t_j)) \in \lambda_j \cap B,$$

where λ_j is the affine subspace $I^j(t_j) + \Lambda_j$.

Claim: there exists a time $t_j^* \in [t_j, \tilde{t}_j] \cap [t_j, e^m]$ such that

$$|\Gamma_{j+1}(t_j^*) - \Gamma_{j+1}(t_j)| = |\Pi_j(I^j(t_j^*) - I^j(t_j))| = \left(T_j r_j L_j^{-1}\right)^{\tau}.$$

Let us prove the claim. We have to distinguish two cases.

First case: $\tilde{t}_j \leq e^m$. We have

$$|\nabla h(I^j(t_j)) - \omega_j| \leq |\nabla h(I^j(t_j)) - \nabla h(I^{j-1}(t_j))| + |\nabla h(I^{j-1}(t_j)) - \omega_j|,$$

and therefore

$$|\nabla h(I^j(t_j)) - \omega_j| <\cdot\, r_j,$$

while by definition,

$$|\nabla h(I^j(\tilde{t}_j)) - \omega_j| =\cdot\, r_j$$

with a sufficiently *larger* implicit constant (see Remark 3.1). Hence

$$|\nabla h(I^j(\tilde{t}_j)) - \nabla h(I^j(t_j))| >\cdot\, r_j,$$

and this implies

$$|I^j(\tilde{t}_j) - I^j(t_j)| >\cdot\, r_j. \tag{14}$$

But conditions (v) and (vi) give in particular

$$\varepsilon \cdot< r_j,$$

so that (13) and (14) yields

$$|\Pi_j(I^j(\tilde{t}_j) - I^j(t_j))| >\cdot\, r_j.$$

Now using (v) again, this gives

$$|\Pi_j(I^j(\tilde{t}_j) - I^j(t_j))| >\cdot\, \left(T_j r_j L_j^{-1}\right)^\tau,$$

and so we can certainly find a time $t_j^* \in [t_j, \tilde{t}_j]$ such that

$$|\Pi_j(I^j(t_j^*) - I^j(t_j))| = \left(T_j r_j L_j^{-1}\right)^\tau.$$

Second case: $\tilde{t}_j > e^m$. We will first prove that $t_* \in [t_j, e^m]$. Indeed, otherwise t_* belongs to $[t_k, t_{k+1}]$ for some $k \in \{0, \ldots, j-1\}$ and we can write

$$|I(t_*) - I(0)| \leq |I(t_*) - I(t_k)| + \sum_{i=0}^{k-1} |I(t_{i+1}) - I(t_i)|. \tag{15}$$

Each term of the right-hand side of (15) is easily estimated: using (B_i) for $i \in \{0, \ldots, k-1\}$ we have

$$|I^i(t_{i+1}) - I^i(t_i)| < r_i, \quad |I^k(t_*) - I^k(t_k)| < r_k,$$

which implies, by the triangle inequality and the estimate (4)

$$|I(t_{i+1}) - I(t_i)| < 2\rho_i + r_i, \quad |I(t_*) - I(t_k)| < 2\rho_k + r_k.$$

Moreover,

$$|I(t_1) - I(0)| < r_0,$$

hence we find

$$|I(t_*) - I(0)| < \sum_{i=1}^{k}(2\rho_i + r_i) + r_0 < (n+1)^2 r_0,$$

which of course contradicts the definition of our drifting time t_*.

Now to prove the claim, we argue by contradiction and suppose that

$$|\Pi_j(I^j(t) - I^j(t_j))| < \left(T_j r_j L_j^{-1}\right)^{\tau}, \quad t \in [t_j, e^m].$$

Since $t_* \in [t_j, e^m]$, we can use the previous inequality together with the estimate (13) and both conditions (v) and (vi) to first obtain

$$|I^j(t_*) - I^j(t_j)| < r_j,$$

and then with the triangle inequality

$$|I(t_*) - I(t_j)| < 2\rho_j + r_j.$$

Now, as the argument above, writing

$$|I(t_*) - I(0)| \leq |I(t_*) - I(t_j)| + \sum_{i=0}^{j-1} |I(t_{i+1}) - I(t_i)|$$

we find the same contradiction on the time t_*, which completes the proof of the claim.

Now consider the restriction of the curve Γ_{j+1} on the interval $[t_j, t_j^*]$. Using our claim together with conditions (iv) and (v), we can apply Lemma 3.9 to find a time $t_{j+1} \in [t_j, t_j^*]$ such that

$$\begin{cases} |\Pi_j(I^j(t) - I^j(t_j))| < \left(T_j r_j L_j^{-1}\right)^{\tau}, \quad t \in [t_j, t_{j+1}], \\ |\Pi_j(\nabla h(\Gamma_{j+1}(t_{j+1})))| \cdot> \left(T_j r_j L_j^{-1}\right)^{2\tau}. \end{cases} \tag{16}$$

The first inequality of (16), together with (13) and conditions (v) and (vi) give

$$|I^j(t) - I^j(t_j)| < r_j$$

for $t \in [t_j, t_{j+1}]$, hence (a) is verified. Now as in the first step, choose $Q_{j+1} = \varepsilon^{-a_{j+1}(n-1)}$ for some constant a_{j+1} to be chosen later, and apply Lemma 3.10 to approximate $\nabla h(I^j(t_{j+1}))$ by a T_{j+1}-periodic vector ω_{j+1}, that is

$$|\nabla h(I^j(t_{j+1})) - \omega_{j+1}| \leq T_{j+1}^{-1} Q_{j+1}^{-\frac{1}{n-1}} = T_{j+1}^{-1} \varepsilon^{a_{j+1}}. \tag{17}$$

Let $r_{j+1} =\cdot T_{j+1}^{-1}\varepsilon^{a_{j+1}}$ so that (b) is verified by (17). To estimate the period T_{j+1} and the number L_j, we need a lower bound for $|\nabla h(I^j(t_{j+1}))|$ and we will use the fact that we have such a lower bound for $|\nabla h(I(t_1)|$ (see the second inequality of (8)). First note that one has easily

$$|I^j(t_{j+1}) - I(t_1)| <\cdot r_1,$$

since $r_i < r_1$ for $i \in \{1, \dots, j\}$, and therefore

$$|\nabla h(I^j(t_{j+1})) - \nabla h(I(t_1))| <\cdot r_1,$$

so choosing properly the constant in the condition (vii) we can ensure that

$$|\nabla h(I^j(t_{j+1})) - \nabla h(I(t_1))| \cdot< r_0^2$$

and hence

$$|\nabla h(I^j(t_{j+1}))| \geq |\nabla h(I(t_1)| - |\nabla h(I^j(t_{j+1})) - \nabla h(I(t_1))| \cdot> r_0^2. \tag{18}$$

By Lemma 3.10, this gives the estimate

$$1 <\cdot T_{j+1} <\cdot \varepsilon^{-a_{j+1}(n-1)} r_0^{-2}. \tag{19}$$

Now as $|\omega_j| <\cdot 1$ this easily implies that

$$1 \leq L_{j+1} <\cdot \max_{i\in\{1,\dots,j+1\}} \{\varepsilon^{-a_i(n-1)}\} r_0^{-2}. \tag{20}$$

The estimates (19) and (20) give (12).

Next having built r_{j+1}, we need to check that ω_{j+1} is independent of $(\omega_1, \dots, \omega_j)$. First, by using the mean value theorem, the estimate (13) and our condition (vi), we have

$$|\nabla h(I^j(t_{j+1})) - \nabla h(\Gamma_{j+1}(t_{j+1}))| \cdot< \left(T_j r_j L_j^{-1}\right)^{2\tau},$$

and together with the second estimate of (16), this gives

$$|\Pi_j(\nabla h(I^j(t_{j+1})))| \cdot> \left(T_j r_j L_j^{-1}\right)^{2\tau}. \tag{21}$$

Furthermore, using (17)

$$|\Pi_j(\nabla h(I^j(t_{j+1})) - \omega_{j+1})| \leq |\nabla h(I^j(t_{j+1})) - \omega_{j+1}| \cdot< r_{j+1}$$

hence with $(viii)$, we get

$$|\Pi_j(\nabla h(I^j(t_{j+1})) - \omega_{j+1})| \cdot< \left(T_j r_j L_j^{-1}\right)^{2\tau}. \tag{22}$$

Now by the estimates (21) and (22)

$$|\Pi_j(\omega_{j+1})| \geq |\Pi_j(\nabla h(I^j(t_{j+1})))| - |\Pi_j(\nabla h(I^j(t_{j+1})) - \omega_{j+1})| \cdot\!> \left(T_j r_j L_j^{-1}\right)^{2\tau}$$

and so $\Pi_j(\omega_{j+1})$ is non zero, which means that ω_{j+1} is not a linear combination of $\{\omega_1, \ldots, \omega_j\}$. This proves (c).

Finally we can write

$$\begin{aligned} |\omega_{j+1} - \omega_j| &\leq |\omega_{j+1} - \nabla h(I^j(t_{j+1}))| + |\nabla h(I^j(t_{j+1})) - \nabla h(I^j(t_j))| \\ &+ |\nabla h(I^j(t_j)) - \nabla h(I^{j-1}(t_j))| + |\nabla h(I^{j-1}(t_j)) - \omega_j|, \end{aligned}$$

and hence

$$|\omega_{j+1} - \omega_j| \cdot\!< (r_j + r_{j+1}) \cdot\!< r_j.$$

So given any $I \in \mathcal{B}_{r_{j+1}}(\omega_{j+1})$, we have

$$|\nabla h(I) - \omega_j| \leq |\nabla h(I) - \omega_{j+1}| + |\omega_{j+1} - \omega_j| \cdot\!< r_j,$$

so that $I \in \mathcal{B}_{2r_j/3}(\omega_j)$, which gives (d). This ends the proof. □

13. Now we can eventually complete the proof of the main Theorem 2.4.

Proof of Theorem 2.4. As a consequence of Propositions 3.7, 3.11 and 3.12, we know that

$$|I(t) - I(0)| < (n+1)^2 r_0, \quad 0 \leq t < e^m$$

provided that the parameters r_0, m and ε satisfy the following eleven conditions:

(i) $r_{j+1} \cdot\!< \left(T_j r_j L_j^{-1}\right)^{2\tau}$, $j \in \{1, \ldots, n-1\}$;

(ii) $\varepsilon \cdot\!< \left(T_j r_j L_j^{-1}\right)^{2\tau}$, $j \in \{1, \ldots, n-1\}$;

(iii) $\left(T_j r_j L_j^{-1}\right)^{\tau} \cdot\!< r_j$, for $j \in \{1, \ldots, n-1\}$;

(iv) $m T_j r_j \cdot\!< 1$, for $j \in \{1, \ldots, n\}$;

(v) $r_1 \cdot\!< r_0^2$;

(vi) $m T_j \varepsilon \cdot\!< r_1 r_j$, for $j \in \{1, \ldots, n\}$;

(vii) $\varepsilon \cdot\!< r_n$;

$(viii)$ $\left(T_j r_j L_j^{-1}\right)^{\tau} \cdot\!< \gamma L_j^{-\tau}$, for $j \in \{1, \ldots, n-1\}$;

(ix) $r_0 \cdot\!< \gamma$;

(x) $r_0 \cdot< R$;

(xi) $r_j \cdot< \min\{r, s\}$, for $j \in \{1, \dots, n\}$,

where $r_j =\cdot T_j^{-1}\varepsilon^{a_j}$, with a_j to be defined for $j \in \{1, \dots, n\}$, and

$$1 <\cdot T_j <\cdot \varepsilon^{-a_j(n-1)} r_0^{-2}, \quad 1 \le L_j <\cdot \max_{i\in\{1,\dots,j\}} \{\varepsilon^{-a_i(n-1)}\} r_0^{-2}. \tag{23}$$

So let us choose $m \cdot= \varepsilon^{-a}$ and $r_0 = \varepsilon^b$, for two constants a and b also to be determined.

Using the estimates (23) on the periods T_j, $j \in \{1, \dots, n\}$ and the numbers L_j, $j \in \{1, \dots, n-1\}$, as well as the form of r_0, r_j for $j \in \{1, \dots, n\}$ and m, one can see that conditions (i) to (xi) are implied by the following conditions:

(i') $a_{j+1} - 2n\tau \left(\max_{i\in\{1,\dots,j\}}\{a_i\}\right) - 4\tau b > 0$, $j \in \{1, \dots, n-1\}$;

(ii') $1 - 2n\tau a_j - 4\tau b > 0$, $j \in \{1, \dots, n-1\}$;

(iii') $(\tau - 1)a_j - 2b > 0$, for $j \in \{1, \dots, n-1\}$;

(iv') $a_j > a$, for $j \in \{1, \dots, n\}$;

(v') $a_1 - 2b > 0$;

(vi') $1 - a - (2n-1)a_j - na_1 - 6b > 0$, for $j \in \{1, \dots, n\}$;

(vii') $1 - na_n - 2b > 0$;

$(viii')$ $\varepsilon < \gamma^{(\tau a_j)^{-1}}$, for $j \in \{1, \dots, n-1\}$;

(ix') $\varepsilon < \gamma^{b^{-1}}$;

(x') $\varepsilon < R^{b^{-1}}$;

(xi') $\varepsilon < (\min\{r, s\})^{(na_j+2b)^{-1}}$, for $j \in \{1, \dots, n\}$.

So we need to choose constants a_j, $j \in \{1, \dots, n\}$, a and b such that the previous conditions are satisfied. First note that by (i'), the sequence a_j, for $j \in \{1, \dots, n\}$, has to be increasing, hence

$$\max_{i\in\{1,\dots,j\}} \{a_i\} = a_j, \quad j \in \{1, \dots, n\}.$$

Then using (v'), we observe that (i') is satisfied if $a_{j+1} = 2\tau(n+1)a_j$ for $j \in \{1, \dots, n-1\}$, that is

$$a_j = (2\tau(n+1))^{j-1} a_1.$$

Now for (ii') to be satisfied, one can choose

$$a_1 = (2\tau(n+1))^{-n},$$

so a_j, for $j \in \{2, \dots, n\}$, is determined by

$$a_j = (2\tau(n+1))^{-n-1+j}.$$

Then, since $\tau \geq 2$, we may choose

$$b = 3^{-1}a_1 = 3^{-1}(2\tau(n+1))^{-n}$$

and (iii') easily holds. Finally, we may also choose

$$a = b = 3^{-1}(2\tau(n+1))^{-n}$$

so that (iv') is satisfied. With those values, it is easy to check that (v'), (vi') and (vii') holds, recalling that $\tau \geq 2$ and $n \geq 2$. To conclude, just note that $(viii')$, (ix'), (x') and (xi') are satisfied if $\varepsilon \leq \varepsilon_0$ with a sufficiently small ε_0 depending on n, R, r, s, M, γ and τ. This ends the proof. □

A Proof of the normal form

In this first appendix we will give the proof of the normal form 3.2. We will closely follow the method of [Pös99b] and deduce our result from an equivalent version in terms of vector fields (Proposition A.4 below).

A.1 Preliminary estimates

Before giving the proof, we will need some general estimates based on the classical Cauchy inequality.

1. First consider the case of a function f analytic on some domain $\mathcal{D}_{r,s}$, and recall that

$$|\partial_\theta f|_{r,s} = \max_{1\leq i\leq n} |\partial_{\theta_i} f|_{r,s}, \ |\partial_I f|_{r,s} = \max_{1\leq i\leq n} |\partial_{I_i} f|_{r,s}.$$

We take r', s' such that $0 < r' < r$ and $0 < s' < s$. The first estimate is classical, but we repeat the proof for convenience.

Lemma A.1. *Under the previous assumptions, we have*

$$|\partial_I f|_{r-r',s} < \frac{1}{r'}|f|_{r,s}, \quad |\partial_\theta f|_{r,s-s'} < \frac{1}{s'}|f|_{r,s}.$$

Proof. For $x = (\theta, I) \in \mathcal{D}_{r-r',s}$ and any unit vector $v \in \mathbb{C}^n$, consider the function

$$F_{x,v} : t \in \mathbb{C} \longmapsto f(\theta, I + tv) \in \mathbb{C}.$$

This function is well-defined and holomorphic on the disc $|t| < r'$, so the classical Cauchy estimate gives

$$|F'_{x,v}(0)| < \frac{1}{r'}|f|_{r,s},$$

from which the inequality for $\partial_I f$ follows easily by optimizing with respect to x and v. The estimate for $\partial_\theta f$ is completely similar. □

2. Now let $j \in \{1, \ldots, n\}$, and let f and g be analytic functions defined on the domain

$$\mathcal{D}_{r_j,s_j}(\omega_j) = \{(\theta, I) \in \mathcal{D}_{r_j,s_j} \mid |\nabla h(I) - \omega_j| \lessdot r_j\},$$

where ω_j is a periodic vector. We can define a vector field norm on $\mathcal{D}_{r_j,s_j}(\omega_j)$ by

$$|X_f|_{r_j,s_j} = \max\left(|\partial_I f|_{r_j,s_j}, |\partial_\theta f|_{r_j,s_j}\right).$$

However, it will more convenient to use the following "weighted" norm

$$||X_f||_{r_j,s_j} = \max\left(|\partial_I f|_{r_j,s_j}, s_1 r_1^{-1}|\partial_\theta f|_{r_j,s_j}\right),$$

since the components $|\partial_I f|_{r_j,s_j}$ and $|\partial_\theta f|_{r_j,s_j}$ may have very different sizes when estimated from the size of f by a Cauchy estimate (this idea is used in [DG96]).

Remark A.2. *Note that under assumption* (A), $s_1 r_1^{-1} > 1$, *so we have the inequality* $|X_f|_{r_j,s_j} \leq ||X_f||_{r_j,s_j}$ *and the equality holds if* f *is integrable. Moreover, note that each norm* $||\,.\,||_{r_j,s_j}$ *is normalized with* $s_1 r_1^{-1}$ *(and not with* $s_j r_j^{-1}$*): by our inclusions of domains, this implies in particular that* $||\cdot||_{r_{j+1},s_{j+1}} \leq ||\cdot||_{2r_j/3,2s_j/3}$.

It is well-known how to use the Cauchy inequality to estimate the size of the Poisson bracket $\{f, g\}$ in terms of f and g. Similarly, our second estimate is concerned with the size of the vector field $[X_f, X_g]$ in terms of X_f and X_g. We take r', s' such that $0 < r' < r_j$ and $0 < s' < s_j$.

Lemma A.3. *Under the previous assumptions, we have*

$$||[X_f, X_g]||_{r_j-r',s_j-s'} < \frac{1}{r'}||X_f||_{r_j,s_j}||X_g||_{r_j,s_j},$$

and moreover, if g *is integrable, then*

$$||[X_f, X_g]||_{r_j-r',s_j-s'} < \frac{1}{s'}||X_f||_{r_j,s_j}||X_g||_{r_j}.$$

Proof. First recall that

$$[X_f, X_g] = \frac{d}{dt}(\Phi_t^g)^* X_f \Big|_{t=0}.$$

Now fix $x \in \mathcal{D}_{r_j - r', s_j - s'}$, and let us define the vector-valued function

$$F_x : t \in \mathbb{C} \longmapsto (\Phi_t^g)^* X_f(x) \in \mathbb{C}^{2n}.$$

Clearly, the map Φ_t^g is analytic, and it sends $\mathcal{D}_{r_j - r', s_j - s'}(\omega_j)$ into $\mathcal{D}_{r_j, s_j}(\omega_j)$ for complex values of t satisfying

$$|t| < r' ||X_g||_{r_j, s_j}^{-1},$$

hence the function F_x is well-defined and analytic on the disc $|t| < r' ||X_g||_{r_j, s_j}^{-1}$. So applying the classical Cauchy estimate to each component of F_x and optimizing with respect to $x \in \mathcal{D}_{r_j - r', s_j - s'}$ we obtain the desired inequality

$$||[X_f, X_g]||_{r_j - r', s_j - s'} < \frac{1}{r'} ||X_g||_{r_j, s_j} ||X_f||_{r_j, s_j}.$$

In case g is integrable, the map Φ_t^g leaves invariant the action components, so the same reasoning can be applied on the larger disc

$$|t| < s' ||X_g||_{r_j, s_j}^{-1},$$

giving the improved estimate

$$||[X_f, X_g]||_{r_j - r', s_j - s'} < \frac{1}{s'} ||X_f||_{r_j, s_j} ||X_g||_{r_j}.$$

□

A.2 Proof of Proposition 3.2

Now we can pass to the proof of Proposition 3.2. Given $\tilde{\varepsilon} > 0$ which will be the size of our perturbating vector field X_f, let us introduce a slightly modified set of conditions $(\tilde{A}_j)$, for $j \in \{1, \dots, n\}$, where $(\tilde{A}_1)$ is

$$\begin{cases} mT_1 \tilde{\varepsilon} \cdot\!< r_1, \ mT_1 r_1 \cdot\!< s_1, \ 0 < r_1 <\!\cdot s_1, \\ \mathcal{B}_{r_1}(\omega_1) \neq \emptyset, \end{cases} \tag{$\tilde{A}_1$}$$

and for $j \in \{2, \dots, n\}$, $(\tilde{A}_j)$ is

$$\begin{cases} mT_j \tilde{\varepsilon} \cdot\!< r_j, \ mT_j r_j \cdot\!< s_j, \ 0 < r_j <\!\cdot s_j, \\ \mathcal{B}_{r_j}(\omega_j) \neq \emptyset, \ \mathcal{D}_{r_j, s_j}(\omega_j) \subseteq \mathcal{D}_{2r_{j-1}/3, 2s_{j-1}/3}(\omega_{j-1}). \end{cases} \tag{$\tilde{A}_j$}$$

These modifications take into account the fact that we will use the weighted norms $||\,.\,||_{r_j, s_j}$, for $j \in \{1, \dots, n\}$.

3. The normal form lemma in terms of vector fields is the following.

Proposition A.4. *Consider $H = h + f$ on the domain $\mathcal{D}_{r_1,s_1}(\omega_1)$, with $||X_f||_{r_1,s_1} < \tilde{\varepsilon}$, and let $j \in \{1,\dots,n\}$. If $(\tilde{A}_i)$ is satisfied for any $i \in \{1,\dots,j\}$, then there exists an analytic symplectic transformation*

$$\Psi_j : \mathcal{D}_{2r_j/3,2s_j/3}(\omega_j) \to \mathcal{D}_{r_1,s_1}(\omega_1)$$

such that

$$H \circ \Psi_j = h + g_j + f_j,$$

with $\{g_j, l_i\} = 0$ for $i \in \{1,\dots,j\}$, and the estimates

$$||X_{g_j}||_{2r_j/3,2s_j/3} \lessdot \tilde{\varepsilon}, \quad ||X_{f_j}||_{2r_j/3,2s_j/3} \lessdot e^{-m}\tilde{\varepsilon}.$$

Moreover, we have $\Psi_j = \Phi_1 \circ \cdots \circ \Phi_j$ with

$$\Phi_i : \mathcal{D}_{2r_i/3,2s_i/3}(\omega_i) \to \mathcal{D}_{r_i,s_i}(\omega_i)$$

such that $|\Phi_i - \mathrm{Id}|_{2r_i/3,2s_i/3} \cdot\!< r_i$.

Let us see how this implies our Proposition 3.2.

Proof of Proposition 3.2. We know that $|f|_{r,s} < \varepsilon$, so we can apply Lemma A.1 with $r' = r_1$ and $s' = s_1$ to obtain

$$|\partial_I f|_{r-r_1,s} < r_1^{-1}|f|_{r,s}, \quad |\partial_\theta f|_{r,s-s_1} < s_1^{-1}|f|_{r,s},$$

and hence

$$||X_f||_{r-r_1,s-s_1} < r_1^{-1}\varepsilon.$$

Now since $r_1 \cdot\!< r$ and $s_1 \cdot\!< s$ (this is part of assumption (A_1)), we have the inclusion $\mathcal{D}_{r_1,s_1}(\omega_1) \subseteq \mathcal{D}_{r-r_1,s-s_1}$ and hence

$$||X_f||_{r_1,s_1} \lessdot r_1^{-1}\varepsilon.$$

Set $\tilde{\varepsilon} = r_1^{-1}\varepsilon$, then for any $i \in \{1,\dots,j\}$, (A_i) implies $(\tilde{A}_i)$ so that the Proposition A.4 can be applied: there exists an analytic symplectic transformation

$$\Psi_j : \mathcal{D}_{2r_j/3,2s_j/3}(\omega_j) \to \mathcal{D}_{r_1,s_1}(\omega_1)$$

such that

$$H \circ \Psi_j = h + g_j + f_j,$$

with $\{g_j, l_i\} = 0$ for $i \in \{1,\dots,j\}$, and the estimates

$$||X_{g_j}||_{2r_j/3,2s_j/3} \lessdot \varepsilon r_1^{-1}, \quad ||X_{f_j}||_{2r_j/3,2s_j/3} \lessdot e^{-m} r_1^{-1}\varepsilon.$$

Recalling the definition of our norm $||\,.\,||_{r_j,s_j}$, this readily implies

$$|\partial_\theta g_j|_{2r_j/3,2s_j/3} \lessdot \varepsilon s_1^{-1} \lessdot \varepsilon, \quad |\partial_\theta f_j|_{2r_j/3,2s_j/3} \lessdot e^{-m}\varepsilon s_1^{-1} \lessdot e^{-m}\varepsilon.$$

Moreover, we have $\Psi_j = \Phi_1 \circ \cdots \circ \Phi_j$ with

$$\Phi_i : \mathcal{D}_{2r_i/3,2s_i/3}(\omega_i) \to \mathcal{D}_{r_i,s_i}(\omega_i)$$

such that $|\Phi_i - \mathrm{Id}|_{2r_i/3,2s_i/3} \cdot\!< r_i$. □

4. Hence it remains to prove Proposition A.4. This will be done by induction on $j \in \{1, \dots, n\}$, and for that we shall need two iterative lemmas. The first iterative lemma is needed for the first step, that is to prove the statement for $j = 1$, and it can be seen as an averaging process with respect to one fast angle.

Lemma A.5 (First iterative lemma)**.** *Consider $H = h+g+f$ on the domain $\mathcal{D}_{r_1,s_1}(\omega_1)$, with h integrable, $\{g, l_1\} = 0$, and assume that*

$$||X_g||_{r_1,s_1} < \tilde{\varepsilon}, \quad ||X_f||_{r_1,s_1} < \tilde{\varepsilon}.$$

If we have

$$T_1\tilde{\varepsilon} < r' < s'$$

with two real numbers r', s' satisfying $0 < r' < r_1$ and $0 < s' < s_1$, then there exists an analytic symplectic transformation

$$\varphi_1 : \mathcal{D}_{r_1-r',s_1-s'}(\omega_1) \to \mathcal{D}_{r_1,s_1}(\omega_1)$$

such that $|\varphi_1 - \mathrm{Id}|_{r_1-r',s_1-s'} < T_1\tilde{\varepsilon}$ and

$$H \circ \varphi_1 = h + g_+ + f_+,$$

with $\{g_+, l_1\} = 0$ and the estimates

$$||X_{g_+}||_{r_1,s_1} < \tilde{\varepsilon}, \quad ||X_{g_+} - X_g||_{r_1,s_1} < \tilde{\varepsilon}, \quad ||X_{f_+}||_{r_1-r',s_1-s'} < \left(\frac{r_1}{s'} + \frac{\tilde{\varepsilon}}{r'}\right) T_1\tilde{\varepsilon}.$$

Proof. We have $H = h + g + f$, with h integrable, g satisfying $\{g, l_1\}$ and f a general term. Let us write

$$[f]_1 = \frac{1}{T_1}\int_0^{T_1} f \circ \Phi_t^{l_1} dt,$$

the average of f along the Hamiltonian flow of l_1.

Our transformation $\varphi_1 = \Phi_1^{\chi}$ will be the time-one map of the Hamiltonian flow generated by some auxiliary function χ which satisfies

$$\{\chi, l_1\} = f - [f]_1.$$

The latter equation is easily solved by

$$\chi = \frac{1}{T_1}\int_0^{T_1} (f - [f]_1) \circ \Phi_t^{l_1} t dt, \tag{24}$$

and by Taylor formula, our transformed Hamiltonian writes

$$H \circ \varphi_1 = h + g_+ + f_+,$$

with

$$g_+ = g + [f]_1, \quad f_+ = \int_0^1 \{h - l_1 + g + f_t, \chi\} \circ \Phi_t^\chi dt,$$

and $f_t = tf + (1-t)[f]_1$. By construction, g_+ still satisfies $\{g_+, l_1\} = 0$, and

$$X_{g_+} - X_g = X_{[f]_1} = \frac{1}{T_1} \int_0^{T_1} (\Phi_t^{l_1})^* X_f dt.$$

Our hypothesis $||X_f||_{r_1,s_1} < \tilde{\varepsilon}$ immediately gives $||X_{g_+} - X_g||_{r_1,s_1} < \tilde{\varepsilon}$ and also $||X_{g_+}||_{r_1,s_1} \lessdot \tilde{\varepsilon}$. Similarly using (24) we have the expression

$$X_\chi = \frac{1}{T_1} \int_0^{T_1} (\Phi_t^{l_1})^* X_{f-[f]_1} t dt,$$

and hence $||X_\chi||_{r_1,s_1} < T_1\tilde{\varepsilon}$. By the hypothesis $T_1\tilde{\varepsilon} < r' < s'$ our transformation φ_1 maps $\mathcal{D}_{r_1-r',s_1-s'}(\omega_1)$ into $\mathcal{D}_{r_1,s_1}(\omega_1)$ and

$$|\varphi_1 - \mathrm{Id}|_{r_1-r',s_1-s'} < T_1\tilde{\varepsilon}.$$

Therefore it remains to estimate the vector field

$$X_{f_+} = \int_0^1 (\Phi_t^\chi)^* [X_{h-l_1} + X_g + X_{f_t}, X_\chi] dt,$$

and for that it is enough to estimate the brackets $[X_{f_t}, X_\chi]$, $[X_g, X_\chi]$ and $[X_{h-l_1}, X_\chi]$. Using Lemma (A.3), we find

$$||[X_{f_t}, X_\chi]||_{r_1-r',s_1-s'} < \frac{1}{r'} ||[X_{f_t}||_{r_1,s_1} ||X_\chi||_{r_1,s_1} \lessdot \frac{\tilde{\varepsilon}}{r'} T_1\tilde{\varepsilon}$$

and

$$||[X_g, X_\chi]||_{r_1-r',s_1-s'} < \frac{1}{r'} ||[X_g||_{r_1,s_1} ||X_\chi||_{r_1,s_1} \lessdot \frac{\tilde{\varepsilon}}{r'} T_1\tilde{\varepsilon}.$$

For the last bracket, note that $h - l_1$ is integrable so that we can use the improved estimate in Lemma (A.3). By definition of the domain $\mathcal{D}_{r_1,s_1}(\omega_1)$, we have $||X_{h-l_1}||_{r_1} \lessdot r_1$ and hence

$$||[X_{h-l_1}, X_\chi]||_{r_1-r',s_1-s'} < \frac{1}{s'} ||X_{h-l_1}||_{r_1} ||X_\chi||_{r_1,s_1} \lessdot \frac{r_1}{s'} T_1\tilde{\varepsilon}.$$

Putting the last three estimates together we arrive at

$$||X_{f_+}||_{r_1-r',s_1-s'} \lessdot \left(\frac{r_1}{s'} + \frac{\tilde{\varepsilon}}{r'} \right) T_1\tilde{\varepsilon}.$$

□

Our second iterative lemma is needed for the inductive step, that is to go from j to $j+1$. This is just a simple extension of the previous one. Let $j \in \{1, \ldots, n-1\}$.

Lemma A.6 (Second iterative lemma)**.** *Consider $H = h + g + f$ on the domain $\mathcal{D}_{r_{j+1},s_{j+1}}(\omega_{j+1})$, with h integrable, $\{g, l_i\} = 0$ for $i \in \{1, \dots, j+1\}$, $\{f, l_{i'}\} = 0$ for $i' \in \{1, \dots, j\}$, and assume that*

$$||X_g||_{r_{j+1},s_{j+1}} <\cdot \tilde{\varepsilon}, \quad ||X_f||_{r_{j+1},s_{j+1}} <\cdot \tilde{\varepsilon}.$$

If we have

$$T_{j+1}\tilde{\varepsilon} \cdot< r' \cdot< s'$$

with two real numbers r', s' satisfying $0 < r' < r_{j+1}$ and $0 < s' < s_{j+1}$, then there exists an analytic symplectic transformation

$$\varphi_{j+1} : \mathcal{D}_{r_{j+1}-r',s_{j+1}-s'}(\omega_{j+1}) \to \mathcal{D}_{r_{j+1},s_{j+1}}(\omega_{j+1})$$

such that $|\varphi_{j+1} - \mathrm{Id}|_{r_{j+1}-r',s_{j+1}-s'} <\cdot T_{j+1}\tilde{\varepsilon}$ and

$$H \circ \varphi_{j+1} = h + g_+ + f_+,$$

with $\{g_+, l_i\} = 0$ for $i \in \{1, \dots, j+1\}$, $\{f_+, l_{i'}\} = 0$ for $i' \in \{1, \dots, j\}$, and the estimates

$$||X_{g_+}||_{r_{j+1},s_{j+1}} <\cdot \tilde{\varepsilon}, \quad ||X_{g_+} - X_g||_{r_{j+1},s_{j+1}} <\cdot \tilde{\varepsilon},$$

$$||X_{f_+}||_{r_{j+1}-r',s_{j+1}-s'} <\cdot \left(\frac{r_{j+1}}{s'} + \frac{\tilde{\varepsilon}}{r'}\right) T_{j+1}\tilde{\varepsilon}.$$

Proof. Our Hamiltonian is $H = h+g+f$, h is integrable and we have $\{g, l_i\} = 0$ for $i \in \{1, \dots, j+1\}$ and $\{f, l_{i'}\} = 0$ for $i' \in \{1, \dots, j\}$. Once again, our transformation $\varphi_{j+1} = \Phi_1^\chi$ will be the time-one map of the Hamiltonian flow generated by some auxiliary function χ.

We choose

$$\chi = \frac{1}{T_{j+1}} \int_0^{T_{j+1}} (f - [f]_{j+1}) \circ \Phi_t^{l_{j+1}} t dt, \tag{25}$$

where $[.]_{j+1}$ is the averaging along the Hamiltonian flow of l_{j+1}. Introducing the notation $f_t = tf + (1-t)[f]_{j+1}$, like in Lemma A.5 we have

$$H \circ \varphi_{j+1} = h + g_+ + f_+$$

with

$$g_+ = g + [f]_{j+1}, \quad f_+ = \int_0^1 \{h - l_{j+1} + g + f_t, \chi\} \circ \Phi_t^\chi dt.$$

We need to verify that we still have $\{g_+, l_i\} = 0$ for $i \in \{1, \dots, j+1\}$ and $\{f_+, l_{i'}\} = 0$ for $i' \in \{1, \dots, j\}$. By definition, $\{[f]_{j+1}, l_{j+1}\} = 0$, and for

$i' \in \{1, \dots, j\}$, we compute

$$\begin{aligned}
\{[f]_{j+1}, l_{i'}\} &= \frac{1}{T_{j+1}} \int_0^{T_{j+1}} \{f \circ \Phi_t^{l_{j+1}}, l_{i'}\} dt \\
&= \frac{1}{T_{j+1}} \int_0^{T_{j+1}} \{f \circ \Phi_t^{l_{j+1}}, l_{i'} \circ \Phi_t^{l_{j+1}}\} dt \\
&= \frac{1}{T_{j+1}} \int_0^{T_{j+1}} \{f, l_{i'}\} \circ \Phi_t^{l_{j+1}} dt \\
&= 0.
\end{aligned}$$

This proves that $\{g_+, l_i\} = \{g + [f]_{j+1}, l_i\} = 0$ for $i \in \{1, \dots, j+1\}$. Now a completely similar calculation shows that for $i' \in \{1, \dots, j\}$, $\{\chi, l_{i'}\} = 0$, hence $l_{i'} \circ \Phi_t^{\chi} = l_{i'}$ and therefore

$$\{f_+, l_{i'}\} = \int_0^1 \{\{h - l_{j+1} + g + f_t, \chi\}, l_{i'}\} \circ \Phi_t^{\chi} dt.$$

The double bracket in the expression above is zero, as a consequence of Jacobi identity and the fact that $\{h - l_{j+1} + g + f_t, l_{i'}\} = \{\chi, l_{i'}\} = 0$, hence $\{f_+, l_{i'}\} = 0$ for $i' \in \{1, \dots, j\}$.

To conclude, using our hypothesis $T_{j+1}\tilde{\varepsilon} \cdot\!< r' \cdot\!< s'$, as in Lemma A.5 we can show that our transformation φ_{j+1} maps $\mathcal{D}_{r_{j+1}-r', s_{j+1}-s'}(\omega_{j+1})$ into $\mathcal{D}_{r_{j+1}, s_{j+1}}(\omega_{j+1})$ with $|\varphi_{j+1} - \mathrm{Id}|_{r_{j+1}-r', s_{j+1}-s'} \cdot\!< T_{j+1}\tilde{\varepsilon}$ and the estimates

$$||X_{g_+}||_{r_{j+1}, s_{j+1}} \cdot\!< \tilde{\varepsilon}, \quad ||X_{g_+} - X_g||_{r_{j+1}, s_{j+1}} \cdot\!< \tilde{\varepsilon},$$

$$||X_{f_+}||_{r_{j+1}-r', s_{j+1}-s'} \cdot\!< \left(\frac{r_{j+1}}{s'} + \frac{\tilde{\varepsilon}}{r'} \right) T_{j+1}\tilde{\varepsilon},$$

are obtained in a completely analogous way. □

5. We can eventually complete the proof of our normal form A.4.

Proof of Proposition A.4. The proof is by induction on $j \in \{1, \dots, n\}$.

First step. Here we assume $(\tilde{A}_1)$ and we will apply m times our first iterative Lemma A.5, starting with the Hamiltonian

$$H^0 = H = h + g^0 + f^0$$

where $g^0 = 0$ and $f^0 = f$ and choosing uniformly at each step

$$r' = (3m)^{-1} r_1, \quad s' = (3m)^{-1} s_1.$$

Since $m \geq 1$, we have $0 < r' < r_1$, $0 < s' < s_1$ and using $(\tilde{A}_1)$, we have

$$T_1 \tilde{\varepsilon} < r' < s',$$

so that the lemma can indeed be applied at each step. For $i \in \{0, \dots, m-1\}$, the Hamiltonian $H^i = h + g^i + f^i$ at step i is transformed into

$$H^{i+1} = H^i \circ \varphi_1^i = h + g^{i+1} + f^{i+1}.$$

For each $i \in \{0, \dots, m\}$, we obviously have $\{g^i, l_1\} = 0$ and we claim that the estimates

$$||X_{g^i}||_{r_1^i, s_1^i} \cdot\!< \tilde{\varepsilon}, \quad ||X_{f^i}||_{r_1^i, s_1^i} < \tilde{\varepsilon}_i, \tag{26}$$

hold true, where we have set $\tilde{\varepsilon}_i = e^{-i}\tilde{\varepsilon}$, $r_1^i = r_1 - ir'$ and $s_1^i = s_1 - is'$. Assuming this claim, given $i \in \{0, \dots, m-1\}$, we have

$$\varphi_1^i : \mathcal{D}_{r_1^{i+1}, s_1^{i+1}}(\omega_1) \longrightarrow \mathcal{D}_{r_1^i, s_1^i}(\omega_1),$$

so that $\Psi_1 = \varphi_1^0 \circ \cdots \circ \varphi_1^{m-1}$ is well defined from $\mathcal{D}_{2r_1/3, 2s_1/3}(\omega_1)$ to $\mathcal{D}_{r_1, s_1}(\omega_1)$. Setting $g_1 = g^m$ and $f_1 = f^m$, we finally obtain

$$H \circ \Psi_1 = h + g_1 + f_1$$

with the desired properties, that is $\{g_1, l_1\} = 0$ and the estimates

$$||X_{g_1}||_{2r_1/3, 2s_1/3} < \tilde{\varepsilon}, \quad ||X_{f_1}||_{2r_1/3, 2s_1/3} < e^{-m}\tilde{\varepsilon}.$$

Note that since $||X_{f^i}||_{r_1^i, s_1^i} < \tilde{\varepsilon}_i$ for $i \in \{0, \dots, m-1\}$, we obtain

$$|\varphi_1^i - \mathrm{Id}|_{r_1^{i+1}, s_1^{i+1}} < T_1 \tilde{\varepsilon}_i,$$

which gives

$$|\Psi_1 - \mathrm{Id}|_{2r_1/3, 2s_1/3} \leq \sum_{k=0}^{m-1} T_1 \tilde{\varepsilon}_k \cdot\!< T_1 \tilde{\varepsilon}.$$

But recall that $mT_1\tilde{\varepsilon} \cdot\!< r_1$ and hence we can arrange

$$|\Psi_1 - \mathrm{Id}|_{2r_1/3, 2s_1/3} \cdot\!< r_1.$$

Therefore to conclude the proof we need to establish the estimates (26), and we may proceed by induction. For $i = 0$, $g^0 = 0$ and $f^0 = f$ so there is nothing to prove. Now assume that the estimates (26) are satisfied for each $k \leq i$, where $i \in \{0, \dots, m-1\}$. For $k \in \{0, \dots, i\}$, since $||X_{f^k}||_{r_1^k, s_1^k} < \tilde{\varepsilon}_k$ we get that

$$||X_{g^{k+1}} - X_{g^k}||_{r_1^{k+1}, s_1^{k+1}} < \tilde{\varepsilon}_k,$$

and therefore

$$||X_{g^{i+1}}||_{r_1^{i+1}, s_1^{i+1}} \leq \sum_{k=0}^{i} \tilde{\varepsilon}_k \cdot\!< \tilde{\varepsilon},$$

so this gives the desired estimate for $X_{g^{i+1}}$. For $X_{f^{i+1}}$, note that

$$||X_{f^{i+1}}||_{r_1^{i+1},s_1^{i+1}} \lessdot T\left(\frac{r_1}{s'}+\frac{\tilde{\varepsilon}}{r'}\right)||X_{f^i}||_{r_1^i,s_1^i},$$

but

$$T_1\left(\frac{r_1}{s'}+\frac{\tilde{\varepsilon}}{r'}\right) =\cdot \left(\frac{mT_1r_1}{s_1}+\frac{mT_1\tilde{\varepsilon}}{r_1}\right),$$

so choosing properly the implicit constants in $(\tilde{A}_1)$ we can ensure that

$$T_1\left(\frac{r_1}{s'}+\frac{\tilde{\varepsilon}}{r'}\right) \cdot\!< \frac{1}{e}$$

which implies the estimate for $X_{f^{i+1}}$ and concludes this first step.

Inductive step. Now assume that the statement holds true for some $j \in \{1,\dots,n-1\}$, and we have to show that it remains true for $j+1$. By assumptions, there exists an analytic symplectic transformation

$$\Psi_j : \mathcal{D}_{2r_j/3,2s_j/3}(\omega_j) \to \mathcal{D}_{r_1,s_1}(\omega_1)$$

such that

$$H \circ \Psi_j = h + g_j + f_j,$$

with $\{g_j, l_i\} = 0$ for $i \in \{1,\dots,j\}$ and the estimates

$$||X_{g_j}||_{2r_j/3,2s_j/3} \lessdot \tilde{\varepsilon}, \quad ||X_{f_j}||_{2r_j/3,2s_j/3} \lessdot e^{-m}\tilde{\varepsilon}.$$

Also, $\Psi_j = \Phi_1 \circ \cdots \circ \Phi_j$ with

$$\Phi_i : \mathcal{D}_{2r_i/3,2s_i/3}(\omega_i) \to \mathcal{D}_{r_i,s_i}(\omega_i)$$

such that $|\Phi_i - \mathrm{Id}|_{2r_i/3,2s_i/3} \cdot\!< r_i$ for $i \in \{1,\dots,n\}$. Furthermore, $(\tilde{A}_{j+1})$ holds. Now consider the Hamiltonian $h+g_j$, it is defined on $\mathcal{D}_{2r_j/3,2s_j/3}(\omega_j)$, hence by $(\tilde{A}_{j+1})$, it is also defined on the domain $\mathcal{D}_{r_{j+1},s_{j+1}}(\omega_{j+1})$ and it satisfies $\{g_j, l_i\} = 0$ for $i \in \{1,\dots,j\}$. Moreover, we have the estimate

$$||X_{g_j}||_{r_{j+1},s_{j+1}} \leq ||X_{g_j}||_{2r_j/3,2s_j/3} \lessdot \tilde{\varepsilon}.$$

As in the first step, starting this time with the Hamiltonian

$$h + g_j = h + g_j^0 + f_j^0,$$

with $g_j^0 = 0$ and $f_j^0 = g_j$, we can apply m times our second iterative Lemma A.6 to have the following: there exists an analytic symplectic transformation

$$\Phi_{j+1} : \mathcal{D}_{2r_{j+1}/3,2s_{j+1}/3}(\omega_{j+1}) \to \mathcal{D}_{r_{j+1},s_{j+1}}(\omega_{j+1})$$

of the form $\Phi_{j+1} = \varphi^0_{j+1} \circ \cdots \circ \varphi^{m-1}_{j+1}$ such that $|\Phi_{j+1} - \mathrm{Id}|_{2r_{j+1}/3, 2s_{j+1}/3} \cdot\!< r_{j+1}$ and

$$(h + g_j) \circ \Phi_{j+1} = h + g_j^m + f_j^m,$$

with $\{g_j^m, l_i\} = 0$ for $i \in \{1, \ldots, j+1\}$, and the estimates

$$||X_{g_j^m}||_{2r_{j+1}/3, 2s_{j+1}/3} \cdot\!< \tilde{\varepsilon}, \quad ||X_{f_j^m}||_{2r_{j+1}/3, 2s_{j+1}/3} \cdot\!< e^{-m}\tilde{\varepsilon}.$$

Now we set

$$\Psi_{j+1} = \Psi_j \circ \Phi_{j+1} : \mathcal{D}_{2r_{j+1}/3, 2s_{j+1}/3}(\omega_{j+1}) \to \mathcal{D}_{r_1, s_1}(\omega_1),$$

which is well-defined by $(\tilde{A}_{j+1})$, to have

$$\begin{aligned} H \circ \Psi_{j+1} &= (H \circ \Psi_j) \circ \Phi_{j+1} \\ &= (h + g_j + f_j) \circ \Phi_{j+1} \\ &= (h + g_j) \circ \Phi_{j+1} + f_j \circ \Phi_{j+1} \\ &= h + g_j^m + f_j^m + f_j \circ \Phi_{j+1} \\ &= h + g_{j+1} + f_{j+1} \end{aligned}$$

with $g_{j+1} = g_j^m$ and $f_{j+1} = f_j^m + f_j \circ \Phi_{j+1}$. The conclusions follow: $\{g_{j+1}, l_i\} = 0$ for $i \in \{1, \ldots, j+1\}$, we have the estimate

$$||X_{g_{j+1}}||_{2r_{j+1}/3, 2s_{j+1}/3} \cdot\!< \tilde{\varepsilon},$$

and since

$$||X_{f_j \circ \Phi_{j+1}}||_{2r_{j+1}/3, 2s_{j+1}/3} \leq ||X_{f_j}||_{r_{j+1}, s_{j+1}} \leq ||X_{f_j}||_{2r_j/3, 2s_j/3}$$

we also have

$$\begin{aligned} ||X_{f_{j+1}}||_{2r_{j+1}/3, 2s_{j+1}/3} &\leq ||X_{f_j^m}||_{2r_{j+1}/3, 2s_{j+1}/3} + ||X_{f_j}||_{2r_j/3, 2s_j/3} \\ &\cdot\!< e^{-m}\tilde{\varepsilon}. \end{aligned}$$

The proof is therefore complete. □

B SDM functions

In this appendix, we will study our class of SDM functions. We will first show in B.1 that they satisfy an adapted steepness property, which we used in the proof of our exponential estimates, and then in B.2 we will prove that they are generic. These results are similar to [Nie07].

B.1 Steepness.

1. We denote by $GA_B(n,k)$ the set of all affine subspaces of $\mathbb{R}^n$ of dimension k intersecting the ball B, and by $GA_B^L(n,k)$ those subspaces with direction in $G^L(n,k)$ (the latter is the space of linear subspaces of $\mathbb{R}^n$ of dimension k whose orthogonal complement is spanned by integer vectors of length less than or equal to L). Let us recall the classical steepness condition, originally introduced by N.N. Nekhoroshev ([Nek77]).

Definition B.1. *A function $h \in C^2(B)$ is said to be steep if it has no critical points and if for any $k \in \{1,\dots,n-1\}$, there exist an index $p_k > 0$ and coefficients $C_k > 0$, $\delta_k > 0$ such that for any affine subspace $\lambda_k \in GA_B(n,k)$ and any continuous curve $\Gamma : [0,1] \to \lambda_k \cap B$ with*

$$\|\Gamma(0) - \Gamma(1)\| = r < \delta_k,$$

there exists $t_* \in [0,1]$ such that:

$$\begin{cases} \|\Gamma(t) - \Gamma(0)\| < r, \quad t \in [0,t_*], \\ \|\Pi_{\Lambda_k}(\nabla h(\Gamma(t_*)))\| > C_k r^{p_k} \end{cases}$$

where Π_{Λ_k} is the projection onto Λ_k, the direction of λ_k.

The function is said to be symmetrically steep (or shortly S-steep) if the above property is also satisfied for $k = n$, with an index $p_n > 0$ and coefficients $C_n > 0$, $\delta_n > 0$.

Let us remark that S-steep functions are allowed to have critical points. Those definitions are rather obscure, but in fact it can be given a simpler and more geometric interpretation, as was shown by Ilyashenko ([Ily86]) and Niederman ([Nie06]). Important examples of steep functions are given by the class of strictly convex (or quasi-convex) functions, with all the steepness indices equal to one.

2. A typical example of non-steep function, which is due to Nekhoroshev, is $h(I_1,I_2) = I_1^2 - I_2^2$, and it is not exponentially stable: for the perturbation $h_\varepsilon(I_1,I_2) = I_1^2 - I_2^2 + \varepsilon\sin(I_1+I_2)$, any solution with $I_1(0) = I_2(0)$ has a fast drift, that is a drift of order one on a time scale of order ε^{-1} (this is obviously the fastest drift possible). But adding a third order term in the previous example (for example I_2^3) we recover steepness, and this is in fact a general phenomenon. Indeed, non-steep functions has infinite codimension among smooth functions, or more precisely, if $J_r(n)$ is the space of r-jets of C^∞ functions on an open set of $\mathbb{R}^n$, then Nekhoroshev proved in [Nek79] that the set of r-jets of non-steep functions is an algebraic subset of $J_r(n)$ which codimension goes to infinity has r goes to infinity. In this sense, steep functions are "generic". However, for $n \geq 3$, a quadratic Hamiltonian is

steep only if it is sign definite, which is a strong assumption, and more generally a polynomial is generically steep only if its degree is sufficiently high (of order n^2 if n is the number of degrees of freedom). Hence polynomials of lower degree are generically non-steep (see [LM88]). This is clearly a shortcoming, and we will see at the end of the next section the advantage of our genericity condition.

3. Steepness (or S-steepness) is a sufficient condition to ensure exponential stability, but this is not necessary, as was first noticed by Morbidelli and Guzzo (see [MG96]). They considered the Hamiltonian $h(I_1, I_2) = I_1^2 - \alpha I_2^2$, which is non-steep for any value of $\alpha > 0$, and noticed that a "fast drift" is not possible if $\sqrt{\alpha}$ is "strongly" irrational. Therefore a Diophantine condition on $\sqrt{\alpha}$ should ensure exponential stability.

Such considerations were then generalized by Niederman who introduced the class of "Diophantine Morse" functions and who proved that they are exponentially stable ([Nie07]). The only difference between these functions and the "Simultaneous Diophantine Morse" functions we use in this paper is that Diophantine Morse functions consider subspaces in $G_L(n,k)$, which are generated by integer vectors of length bounded by L, while here we are looking at subspaces in $G^L(n,k)$ where the latter condition is imposed on the orthogonal complement. This reflects the difference between the method of proof: in ([Nie07]) the analytic part was based on classical small divisors techniques (that is linear Diophantine approximation) and therefore required an adapted geometric assumption, while here we simply rely on the most basic theorem of simultaneous Diophantine approximation (and this explains the name Simultaneous Diophantine Morse functions).

4. In both cases, the use of such a class of functions has two advantages. The first one is that these functions are generic in a much more clearer sense than steep functions, and this will be explained in the next section. The second advantage is that they are in some sense more general than the usual steep functions, since we only have to consider curves in some specific affine subspaces. This is explained in the proposition below.

Proposition B.2. *Let $h \in SDM_\gamma^\tau(B)$, assume that $|h|_{C^3(B)} < M$ and take $r < 1$. Then for any affine subspace $\lambda \in GA_B^L(n,k)$ and any continuous curve $\Gamma : [0,1] \to \lambda \cap B$ with*

$$\|\Gamma(0) - \Gamma(1)\| = r < (2M)^{-1}\gamma L^{-\tau},$$

there exists $t_* \in [0,1]$ such that:

$$\begin{cases} \|\Gamma(t) - \Gamma(0)\| \leq r, & t \in [0, t_*], \\ \|\Pi_\Lambda(\nabla h(\Gamma(t_*)))\| > \frac{1}{2}r^2 \end{cases}$$

where Π_Λ is the projection onto Λ, the direction of λ.

Proof. It is enough to check that these properties are satisfied for a vector space $\Lambda \in G^L(n,k)$, since any affine subspace $\lambda \in GA^L_B(n,k)$ is of the form $\lambda = v + \Lambda$ with $\Lambda \in G^L(n,k)$ for some vector v. So consider a continuous curve $\Gamma : [0,1] \to \Lambda \cap B$ with length $r < 1$ satisfying

$$\|\Gamma(0) - \Gamma(1)\| = r < (2M)^{-1}\gamma L^{-\tau}.$$

We will denote by $(\alpha(t), \beta)$ the coordinates of $\Gamma(t)$ for $t \in [0,1]$ in a basis adapted to the orthogonal decomposition $\Lambda \oplus \Lambda^\perp$. Therefore

$$\|\Pi_\Lambda(\nabla h(\Gamma(t)))\| = \|\partial_\alpha h_\Lambda(\alpha(t), \beta)\|$$

for all $t \in [0,1]$. We will distinguish distinguish two cases.

For the first one, we suppose that

$$\|\partial_\alpha h_\Lambda(\alpha(0), \beta)\| > 2^{-1}r^2,$$

so the conclusion trivially holds for $t_* = 0$.

For the second one, we have

$$\|\partial_\alpha h_\Lambda(\alpha(0), \beta)\| \leq 2^{-1}r^2, \tag{27}$$

but since $r^2 < r < \gamma L^{-\tau}$, this gives

$$\|\partial_\alpha h_\Lambda(\alpha(0), \beta)\| \leq \gamma L^{-\tau}.$$

Now $h \in SDM^\tau_\gamma(B)$, so we can apply the definition at the point $(\alpha(0), \beta)$, and for any $\eta \in \mathbb{R}^k \setminus \{0\}$ we obtain

$$\|\partial_{\alpha\alpha} h_\Lambda(\alpha(0), \beta).\eta\| > \gamma L^{-\tau}\|\eta\|. \tag{28}$$

Take any $\tilde{\alpha}$ such that $\|\tilde{\alpha} - \alpha(0)\| < (2M)^{-1}\gamma L^{-\tau}$. We can apply Taylor formula with integral remainder to obtain

$$\partial_\alpha h_\Lambda(\tilde{\alpha}, \beta) - \partial_\alpha h_\Lambda(\alpha(0), \beta) = \int_0^1 \partial_{\alpha\alpha} h_\Lambda(\alpha(0) + t(\tilde{\alpha} - \alpha(0)), \beta).(\tilde{\alpha} - \alpha(0))dt.$$

Now since M bounds the third derivative of h, we have

$$\|\partial_{\alpha\alpha} h_\Lambda(\alpha(0) + t(\tilde{\alpha} - \alpha(0)), \beta) - \partial_{\alpha\alpha} h_\Lambda(\alpha(0), \beta)\| \leq Mt\|\tilde{\alpha} - \alpha(0)\| \leq 2^{-1}\gamma L^{-\tau}t,$$

and this yields

$$\begin{aligned} \|\partial_\alpha h_\Lambda(\tilde{\alpha}, \beta) - \partial_\alpha h_\Lambda(\alpha(0), \beta)\| \quad \geq \quad & \|\partial_{\alpha\alpha} h_\Lambda(\alpha(0), \beta).(\tilde{\alpha} - \alpha(0))\| \\ & -2^{-1}\gamma L^{-\tau}\int_0^1 t\|\tilde{\alpha} - \alpha(0)\|dt, \end{aligned}$$

which in turns, using (28) with $\eta = \tilde{\alpha} - \alpha(0)$, gives

$$\begin{aligned} \|\partial_\alpha h_\Lambda(\tilde{\alpha}, \beta) - \partial_\alpha h_\Lambda(\alpha(0), \beta)\| &\geq \left(\gamma L^{-\tau} - 2^{-1}\gamma L^{-\tau} \int_0^1 t dt\right) \|\tilde{\alpha} - \alpha(0)\| \\ &\geq 2^{-1}\gamma L^{-\tau} \|\tilde{\alpha} - \alpha(0)\|. \end{aligned} \tag{29}$$

Now we define

$$t_* = \inf_{t \in [0,1]} \{\|\Gamma(t) - \Gamma(0)\| = r\},$$

so trivially we have

$$\|\Gamma(t) - \Gamma(0)\| \leq r, \quad t \in [0, t_*].$$

Furthermore, we have

$$\|\partial_\alpha h_\Lambda(\alpha(t_*), \beta)\| \geq \|\partial_\alpha h_\Lambda(\alpha(t_*), \beta) - \partial_\alpha h_\Lambda(\alpha(0), \beta)\| - \|\partial_\alpha h_\Lambda(\alpha(0), \beta)\|,$$

and so using (27), (29) and recalling that $\|\alpha(t_*) - \alpha(0)\| = r$ and $\gamma L^{-\tau} > 2r$ we obtain

$$\begin{aligned} \|\partial_\alpha h_\Lambda(\alpha(t_*), \beta)\| &\geq 2^{-1}\gamma L^{-\tau} r - 2^{-1} r^2 \\ &> r^2 - 2^{-1} r^2 \\ &= 2^{-1} r^2, \end{aligned}$$

and this is the desired estimate. □

B.2 Prevalence

5. Here we will prove our results of genericity concerning SDM functions, that is Theorem 2.2 and Corollary 2.3. Our main tool is the following lemma, which is proved in [Nie07] and relies on the quantitative Morse-Sard theory developed by Yomdin (see [YC04] and [Yom83]). Let us denote by λ_k the k-dimensional Lebesgue measure.

Lemma B.3. *Let $g \in C^{2n+1}(B, \mathbb{R}^k)$. Then for any $\kappa \in]0, 1[$ there exist a subset $\mathcal{C}_\kappa \subseteq \mathbb{R}^k$ with*

$$\lambda_k(\mathcal{C}_\kappa) \leq c_k \sqrt{\kappa},$$

where c_k only depends on k, such that for any $\zeta \notin \mathcal{C}_\kappa$, the function g^ζ defined by $g^\zeta(x) = g(x) - \zeta$ satisfies the following: for any $x \in B$,

$$\|g^\zeta(x)\| \leq \kappa \Longrightarrow \|dg^\zeta(x).\nu\| > \kappa \|\nu\|,$$

for any $\nu \in \mathbb{R}^n \setminus \{0\}$.

In the above statement, the set $\mathcal{C}_\kappa$ is a "nearly-critical set" for the function g.

6. Let us prove Theorem 2.2.

Proof of Theorem 2.2. Recall that we are given a function $h \in C^{2n+2}(B)$. The proof is divided in two steps: first, we will describe the set of parameters $\xi \in \mathbb{R}^n$ for which the function h_ξ, defined by $h_\xi(I) = h(I) - \xi.I$, is not in $SDM^\tau(B)$, and then, in a second step, we will show that this set has zero Lebesgue measure, for $\tau > 2(n^2+1)$. In the sequel, given $k \in \{1, \dots, n\}$, we denote by λ_k the Lebesgue measure of $\mathbb{R}^k$.

First step. Given an element $\Lambda \in G^L(n,k)$, let Π_Λ the projection onto this subspace and consider the associate function h_Λ (recall that h_Λ is just the function h written in coordinates adapted to the orthogonal decomposition $\Lambda \oplus \Lambda^\perp$). Let us define the function

$$g = \partial_\alpha h_\Lambda,$$

which belongs $C^{2n+1}(B, \mathbb{R}^k)$, and apply to this function Lemma B.3 with the value $\kappa = \gamma L^{-\tau}$. We find a "nearly-critical" set $\mathcal{C}_\kappa = \mathcal{C}_{\gamma,\tau,L} \subseteq \mathbb{R}^k$ with the measure estimate

$$\lambda_k(\mathcal{C}_{\gamma,\tau,L}) \leq c_k \gamma^{\frac{1}{2}} L^{-\frac{\tau}{2}}, \tag{30}$$

such that for any $\zeta \notin \mathcal{C}_{\gamma,\tau,L}$ and any $(\alpha,\beta) \in B$,

$$\|g^\zeta(\alpha,\beta)\| \leq \kappa \Longrightarrow \|dg^\zeta(\alpha,\beta).\nu\| > \kappa\|\nu\|, \tag{31}$$

for any $\nu \in \mathbb{R}^n \setminus \{0\}$.

Now take any $\zeta \notin \mathcal{C}_{\gamma,\tau,L}$, any $\xi \in \Pi_\Lambda^{-1}(\zeta)$ and consider the modified function h_ξ as well as its version $h_{\xi,\Lambda}$. Since

$$\partial_\alpha h_{\xi,\Lambda} = \partial_\alpha h_\Lambda - \zeta = g - \zeta = g^\zeta,$$

and $\partial_{\alpha,\alpha} h_{\xi,\Lambda} = \partial_{\alpha,\alpha} h_\Lambda$ is just some restriction of dg, the estimate (31) gives for any $(\alpha,\beta) \in B$,

$$\|\partial_\alpha h_{\xi,\Lambda}(\alpha,\beta)\| \leq \gamma L^{-\tau} \Longrightarrow \|\partial_{\alpha,\alpha} h_{\xi,\Lambda}(\alpha,\beta).\eta\| > \gamma L^{-\tau}\|\eta\| \tag{32}$$

for any $\eta \in \mathbb{R}^n \setminus \{0\}$. So let $\mathcal{C}_{\gamma,\tau,L,\Lambda} = \Pi_\Lambda^{-1}(\mathcal{C}_{\gamma,\tau,L})$, and define

$$\mathcal{C}_{\gamma,\tau} = \bigcup_{L \in \mathbb{N}^*} \bigcup_{k \in \{1,\dots,n\}} \bigcup_{\Lambda \in G^L(n,k)} \mathcal{C}_{\gamma,\tau,L,\Lambda}.$$

As a consequence of the estimate (32), the function $h_\xi \in SDM^\tau_\gamma(B)$ provided that $\xi \notin \mathcal{C}_{\gamma,\tau}$, hence $h_\xi \in SDM^\tau(B)$ provided that $\xi \notin \mathcal{C}_\tau$, where

$$\mathcal{C}_\tau = \bigcap_{\gamma > 0} \mathcal{C}_{\gamma,\tau}.$$

Second step. It remains to prove that $\mathcal{C}_\tau$ has zero Lebesgue measure under our assumption that $\tau > 2(n^2+1)$. For an integer $m \in \mathbb{N}^*$, we define

$\mathcal{C}^m_{\gamma,\tau,L,\Lambda}$ (resp. $\mathcal{C}^m_{\gamma,\tau}$ and $\mathcal{C}^m_\tau$) as the intersection of $\mathcal{C}_{\gamma,\tau,L,\Lambda}$ (resp. $\mathcal{C}_{\gamma,\tau}$ and $\mathcal{C}_\tau$) with the ball of $\mathbb{R}^n$ of radius m centered at the origin. As a consequence of (30) and Fubini-Tonelli theorem, one has

$$\lambda_n(\mathcal{C}^m_{\gamma,\tau,L,\Lambda}) \leq V_{n,m} c_k \gamma^{\frac{1}{2}} L^{-\frac{\tau}{2}}$$

where $V_{n,m} = m^n \pi^{n/2} \Gamma(n/2+1)^{-1}$ is the volume of the ball of $\mathbb{R}^n$ of radius m centered at the origin. Therefore

$$\lambda_n \left(\bigcup_{\Lambda \in G^L(n,k)} \mathcal{C}^m_{\gamma,\tau,L,\Lambda} \right) \leq |G^L(n,k)| V_{n,m} c_k L^{-\frac{\tau}{2}} \gamma^{\frac{1}{2}},$$

with $|G^L(n,k)|$ the cardinal of $G^L(n,k)$. But obviously $|G^L(n,k)| \leq L^{n^2}$ and hence

$$\lambda_n \left(\bigcup_{\Lambda \in G^L(n,k)} \mathcal{C}^m_{\gamma,\tau,L,\Lambda} \right) \leq V_{n,m} c_k L^{n^2 - \frac{\tau}{2}} \gamma^{\frac{1}{2}}.$$

Now

$$\lambda_n \left(\bigcup_{k \in \{1,\ldots,n\}} \bigcup_{\Lambda \in G^L(n,k)} \mathcal{C}^m_{\gamma,\tau,L,\Lambda} \right) \leq V_{n,m} \left(\sum_{k=1}^n c_k \right) L^{n^2 - \frac{\tau}{2}} \gamma^{\frac{1}{2}},$$

and so

$$\lambda_n(\mathcal{C}^m_{\gamma,\tau}) \leq V_{n,m} \left(\sum_{k=1}^n c_k \right) \left(\sum_{L=1}^{+\infty} L^{n^2 - \frac{\tau}{2}} \right) \gamma^{\frac{1}{2}}$$

where the sum in the right-hand side of the last estimate is finite since we are assuming $\tau > 2(n^2+1)$. This shows that

$$\lambda_n(\mathcal{C}^m_\tau) = \inf_{\gamma > 0} \lambda_n(\mathcal{C}^m_{\gamma,\tau}) = 0,$$

and as $\mathcal{C}_\tau = \bigcup_{m \geq 1} \mathcal{C}^m_\tau$ we finally obtain

$$\lambda_n(\mathcal{C}_\tau) = 0,$$

and this concludes the proof. □

7. As we mentioned in the introduction, there is a notion of genericity in infinite dimensional vector spaces called prevalence, first introduced in a different setting by Christensen ([Chr73]) and rediscovered by Hunt, Sauer and Yorke ([HSY92], see also [OY05] and [HK10]).

Definition B.4. *Let E be a completely metrizable topological vector space. A Borel subset $S \subseteq E$ is said to be shy if there exists a Borel measure μ on E, with $0 < \mu(C) < \infty$ for some compact set $C \subseteq E$, and such that $\mu(x + S) = 0$ for all $x \in E$.*

An arbitrary set is called shy if it is contained in a shy Borel subset, and finally the complement of a shy set is called prevalent.

The following "genericity" properties are easy to check ([OY05], [HK10]): a prevalent set is dense, a set containing a prevalent set is also prevalent, and prevalent sets are stable under translation and countable intersection.

Furthermore, we have an easy but useful criterion for a set to be prevalent.

Proposition B.5 ([HSY92])**.** *Let P be a subset of E. Suppose there exists a finite-dimensional subspace F of E such that $x + P$ has full λ_F-measure for all $x \in E$. Then P is prevalent.*

8. Now we can prove our Corollary 2.3.

Proof of Corollary 2.3. Let $E = C^{2n+2}(B)$, $P = SDM^\tau(B)$ for $\tau > 2(n^2 + 1)$ and F the space of linear forms of $\mathbb{R}^n$ restricted to B. Then F is a linear subspace of $C^{2n+2}(B)$ of dimension n, and the conclusion follows immediately from Theorem 2.2 and the above Proposition B.5. □

9. To conclude, let us compare our generic condition with the usual steepness property. First, our condition is prevalent in the space $C^k(B)$, with $k \geq 2n+2$, and this is not true for steep functions. But more importantly, as prevalence is nothing but "full Lebesgue measure" in finite dimension, given any non zero integers m and n, Lebesgue almost all polynomial Hamiltonian h_m of degree m with n degrees of freedom is SDM, but not steep unless m is of order n^2. This remark turns out to be very useful when studying the stability of invariant tori under generic conditions (see [Bou09]).

Acknowledgements.

The authors wish to thank Jacques Féjoz and Jean-Pierre Marco for useful discussions, and A.B. also thanks the ASD team at Observatoire de Paris for its hospitality, especially Alain Albouy and Alain Chenciner for their kind support. Both authors thank the CRM of Barcelona where this work was initiated during the semester "Stability and Instability in Mechanical Systems". Finally, the authors wish to thank the referee for his careful reading.

References

[AKN06] V.I. Arnold, V.V. Kozlov, and A.I. Neishtadt, *Mathematical aspects of classical and celestial mechanics*, [Dynamical Systems III], Transl. from the Russian original by E. Khukhro, Third edition, Encyclopedia of Mathematical Sciences, 3 ed., Springer-Verlag, Berlin, 2006.

[Arn64] V.I. Arnold, *Instability of dynamical systems with several degrees of freedom*, Sov. Math. Doklady **5** (1964), 581–585.

[Bam99] D. Bambusi, *Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations*, Math. Z. **230** (1999), no. 2, 345–387.

[BG93] D. Bambusi and A. Giorgilli, *Exponential stability of states close to resonance in infinite-dimensional hamiltonian systems*, J. Statist. Phys. **71** (1993), no. 3-4, 569–606.

[BN02] D. Bambusi and Nekhoroshev N.N., *Long time stability in perturbations of completely resonant PDE's*, Acta Appl. Math. **70** (2002), no. 3, 1–22.

[Bos86] J.-B. Bost, *Tores invariants des systemes dynamiques Hamiltoniens*, Séminaire Bourbaki **133-134** (1986), 113–157.

[Bou04] J. Bourgain, *Remarks on stability and diffusion in high-dimensional Hamiltonian systems and partial differential equations*, Erg. Th. Dyn. Sys. **24** (2004), no. 5, 1331–1357.

[Bou09] A. Bounemoura, *Generic super-exponential stability of invariant tori*, Erg. Th. Dyn. Sys. (2009), to appear.

[Bou10] ———, *Nekhoroshev estimates for finitely differentiable quasi-convex hamiltonians*, Journal of Differential Equations **249** (2010), no. 11, 2905–2920.

[Cas57] J.W.S. Cassels, *An introduction to Diophantine approximations*, Cambridge Tracts in Mathematics and Mathematical Physics, Cambridge University Press, 1957.

[Chr73] J.P.R. Christensen, *On sets of Haar measure zero in abelian Polish groups*, Isr. J. Math. **13** (1973), 255–260.

[DG96] A. Delshams and P. Gutiérrez, *Effective stability and KAM theory*, J. Differ. Equations **128** (1996), no. 2, 415–490.

[dlL01] R. de la Llave, *A tutorial on KAM theory*, Katok, Anatole (ed.) et al., Smooth ergodic theory and its applications (Seattle, WA, 1999). Providence, RI: Amer. Math. Soc. (AMS). Proc. Symp. Pure Math. 69, 175-292, 2001.

[FGB98] F. Fassò, M. Guzzo, and G. Benettin, *Nekhoroshev-stability of elliptic equilibria of Hamiltonian systems*, Comm. Math. Phys. **197** (1998), no. 2, 347–360.

[HK10] B. Hunt and V. Kaloshin, *Prevalence*, H. Broer, F. Takens and B. Hasselblatt (eds.), Handbook of Dynamical Systems Volume 3. North Holland Title, Elsevier, 2010.

[HSY92] B.R. Hunt, T. Sauer, and J.A. Yorke, *Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces*, Bull. of the Amer. Math. Soc. **27** (1992), 217–238.

[Ily86] I.S. Ilyashenko, *A steepness test for analytic functions*, Russian Math. Surveys **41** (1986), 229–230.

[KLDM06] K. Khanin, J. Lopes Dias, and J. Marklof, *Renormalization of multidimensional Hamiltonian flows*, Nonlinearity **19** (2006), no. 12, 2727–2753.

[KLDM07] ———, *Multidimensional continued fractions, dynamical renormalization and KAM theory*, Comm. Math. Phys. **270** (2007), no. 1, 197–231.

[Kol54] A.N. Kolmogorov, *On the preservation of conditionally periodic motions for a small change in Hamilton's function*, Dokl. Akad. Nauk. SSSR **98** (1954), 527–530.

[LM88] P. Lochak and C. Meunier, *Multiphase averaging for classical systems. With applications to adiabatic theorems. Transl. from the French by H. S. Dumas*, Applied Mathematical Sciences, 72, New York etc, Springer-Verlag. xi, 360 pp. , 1988.

[LN92] P. Lochak and A.I. Neishtadt, *Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian*, Chaos **2** (1992), no. 4, 495–499.

[LNN94] P. Lochak, A.I. Neistadt, and L. Niederman, *Stability of nearly integrable convex Hamiltonian systems over exponentially long times*, Kuksin, S. (ed.) et al., Seminar on dynamical systems. Basel: Birkhäuser. Prog. Nonlinear Differ. Equ. Appl. 12, 15-34 , 1994.

[Loc92] P. Lochak, *Canonical perturbation theory via simultaneous approximation*, Russ. Math. Surv. **47** (1992), no. 6, 57–133.

[MG95] A. Morbidelli and A. Giorgilli, *Superexponential stability of KAM tori*, J. Stat. Phys. **78** (1995), 1607–1617.

[MG96] A. Morbidelli and M. Guzzo, *The nekhoroshev theorem and the asteroid belt dynamical system*, Celestial Mech. Dynam. Astronom. **65** (1996), no. 1-2, 107–136.

[MS02] J.-P. Marco and D. Sauzin, *Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems*, Publ. Math. Inst. Hautes Études Sci. **96** (2002), 199–275.

[Nei84] A.I. Neishtadt, *The separation of motions in systems with rapidly rotating phase*, J. Appl. Math. Mech. **48** (1984), no. 2, 133–139.

[Nek77] N.N. Nekhoroshev, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems*, Russian Math. Surveys **32** (1977), no. 6, 1–65.

[Nek79] ———, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems II*, Trudy Sem. Petrovs **5** (1979), 5–50.

[Nie98] L. Niederman, *Nonlinear stability around an elliptic equilibrium point in a Hamiltonian system*, Nonlinearity **11** (1998), no. 6, 1465–1479.

[Nie04] ———, *Exponential stability for small perturbations of steep integrable Hamiltonian systems*, Erg. Th. Dyn. Sys. **24** (2004), no. 2, 593–608.

[Nie06] ———, *Hamiltonian stability and subanalytic geometry*, Ann. Inst. Fourier **56** (2006), no. 3, 795–813.

[Nie07] ———, *Prevalence of exponential stability among nearly integrable Hamiltonian systems*, Erg. Th. Dyn. Sys. **27** (2007), no. 3, 905–928.

[Nie09] ———, *Nekhoroshev theory*, Springer Encyclopedia of Complexity and Systems Science, 2009.

[OY05] W. Ott and J.A. Yorke, *Prevalence*, Bull. of the Amer. Math. Soc. **42** (2005), no. 3, 263–290.

[Pös99a] J. Pöschel, *On Nekhoroshev estimates for a nonlinear Schrödinger equation and a theorem by Bambusi*, Nonlinearity **12** (1999), no. 6, 1587–1600.

[Pös99b] _____, *On Nekhoroshev's estimate at an elliptic equilibrium*, Internat. Math. Res. Notices **4** (1999), 203–215.

[Pös01] _____, *A lecture on the classical KAM theory*, Katok, Anatole (ed.) et al., Smooth ergodic theory and its applications (Seattle, WA, 1999). Providence, RI: Amer. Math. Soc. (AMS). Proc. Symp. Pure Math. 69, 707-732, 2001.

[RS96] J.-P. Ramis and R. Schäfke, *Gevrey separation of fast and slow variables*, Nonlinearity **9** (1996), no. 2, 353–384.

[YC04] Y. Yomdin and G. Comte, *Tame geometry with application in smooth analysis*, Lecture Notes in Mathematics, Springer Verlag, Berlin, 2004.

[Yom83] Y. Yomdin, *The geometry of critical and near-critical values of differentiable mappings*, Math. Ann. **264** (1983), 495–515.