Generic fast diffusion for a class of non-convex Hamiltonians with two degrees of freedom

Abed Bounemoura [*] and Vadim Kaloshin [†]

November 11, 2013

To Yulij Ilyashenko on his 70th birthday with deep respect and admiration

Abstract

In this paper, we study small perturbations of a class of non-convex integrable Hamiltonians with two degrees of freedom, and we prove a result of diffusion for an open and dense set of perturbations, with an optimal time of diffusion which grows linearly with respect to the inverse of the size of the perturbation.

1 Introduction and statement of the result

1.1 Introduction

In this paper, we consider small perturbations of integrable Hamiltonian systems which are defined by a Hamiltonian function of the form

$$H(\theta, I) = h(I) + \varepsilon f(\theta, I), \quad (\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n, \quad 0 \leq \varepsilon < 1,$$

where $n \geq 2$ is an integer and $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$. When $\varepsilon = 0$, $H = h$ is integrable in the sense that the action variables $I(t)$ of all solutions $(\theta(t), I(t))$ of the system associated to h are first integrals, $I(t) = I(0)$ for all times $t \in \mathbb{R}$. The sets $I = I_0$, for $I_0 \in \mathbb{R}^n$, are thus invariant tori of dimension n in the phase space $\mathbb{T}^n \times \mathbb{R}^n$, which moreover carry quasi-periodic motions with frequency $\omega(I_0) = \nabla h(I_0)$, that is $\theta(t) = \theta(0) + t\omega(I_0)$ modulo $\mathbb{Z}^n$. From now on we will assume that the small parameter ε is non-zero, in which case the system defined by H can be considered as an ε-perturbation of the integrable system defined by h.

In the sixties, Arnold conjectured that for a generic h, the following phenomenon should occur: "for any points I' and I'' on the connected level hypersurface of h in the action space there exist orbits connecting an arbitrary small neighbourhood of the torus $I = I'$ with an arbitrary small neighbourhood of the torus $I = I''$, provided that ε is sufficiently small and that f is generic" (see [Arn94]). This is a strong form of instability. A weaker form of this conjecture would be to ask for the existence of orbits for which the variation of the actions is of order one, that is bounded from below independently of ε for all ε sufficiently small. To support his conjecture, Arnold gave an example in [Arn64] where this weaker form of instability is satisfied, with $n = 2$, h convex and f a specific time-periodic perturbation (so this is equivalent to $n = 3$, h quasi-convex and f a specific time-independent perturbation). The phenomenon highlighted in [Arn64] is now known as Arnold diffusion.

---

[*]CNRS-Université Paris Dauphine & Observatoire de Paris (abedbou@gmail.com)

[†]University of Maryland at College Park (vadim.kaloshin@gmail.com)

1.1.1 KAM stability

Obstructions to Arnold diffusion, and to any form of instability in general, are widely known following the works of Kolmogorov and Arnold on the one hand, and the work of Nekhoroshev on the other hand. In [Kol54], Kolmogorov proved that for a non-degenerate h and for all f, the system defined by H still has many invariant tori, provided it is analytic and ε is small enough. What he showed is that among the set of unperturbed invariant tori, there is a subset of positive measure (the complement of which has a measure going to zero when ε goes to zero) who survives any sufficiently small perturbation, the tori being only slightly deformed. The non-degeneracy assumption on h is that at all points, the determinant of its Hessian matrix $\nabla^2 h(I)$ is non-zero. Then, under a different non-degeneracy assumption on h, namely that the determinant of the square matrix

$$\begin{pmatrix} \nabla^2 h(I) & {}^t\nabla h(I) \\ \nabla h(I) & 0 \end{pmatrix}$$

is non-zero at all points, Arnold proved in [Arn63a], [Arn63b] a similar statement but with a set of tori inside a fixed level hypersurface. In particular, for $n = 2$, a level hypersurface is 3-dimensional and the complement of the set of invariant 2-dimensional tori is disconnected, and each connected component is bounded with a diameter going to zero as ε goes to zero. As a consequence, it can be proved more precisely that for $n = 2$ and if h is non-degenerate in the sense of Arnold, along all solutions we have

$$|I(t) - I(0)| \leq c\sqrt{\varepsilon}, \quad t \in \mathbb{R},$$

for some positive constant c. Therefore we have stability for all solutions and for all time. Now for any $n \geq 2$, and if h is either Kolmogorov or Arnold non-degenerate, we have perpetual stability only for most solutions, those lying on invariant tori, and Arnold's example shows that this cannot be true for all solutions. The consequence of these results is that Arnold diffusion cannot exist for $n = 2$ if h is Arnold non-degenerate, and for $n \geq 2$ and h Kolmogorov or Arnold non-degenerate, the unstable solution, if it exists, must live in a set of relatively small measure.

1.1.2 Nekhoroshev stability

In an other direction, in the seventies Nekhoroshev proved ([Nek77], [Nek79]) that for any $n \geq 2$, for a non-degenerate h and for all f, along all solutions we have

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

for some positive constant c_1, c_2, a and b, provided ε is small enough and the system is analytic. So solutions which do not lie on invariant tori are stable not for all time, but during an interval of time which is exponentially long with respect to some power of the inverse of ε. The consequence on Arnold diffusion is that the time of diffusion, that is the time it takes for the action variables to drift independently of ε, is exponentially large. The integrable systems non-degenerate in the sense of Nekhoroshev, which are called *steep*, were originally quite complicated to define, but an equivalent definition was found in [Ily86] and [Nie06]: h is steep if and only if its restriction to any affine subspace has only isolated critical points. Such functions can be proved to be generic in a rather strong sense ([Nek73]), and the simplest

(and also steepest) functions are the convex or quasi-convex ones (convex or quasi-convex functions are those for which the stability exponent a in Nekhoroshev estimates is the best). Note that convex (respectively quasi-convex) functions are Kolmogorov (respectively Arnold) non-degenerate.

So the results of Kolmorogov, Arnold and Nekhoroshev restrict the possibility of diffusion, both in space and in time, at least provided the corresponding non-degeneracy assumptions are met.

1.1.3 Arnold's example and "a priori" unstable systems

Following the original insight of Arnold in [Arn64], much study have been devoted to perturbations of a special class of Hamiltonian systems, which are called "a priori" unstable, where these restrictions are much less stringent. We won't try to give a precise definition of "a priori" unstable systems, but these systems are integrable in the larger sense of symplectic geometry (they have n first integrals in involution and independent almost everywhere) but display hyperbolic features (typically they have a normally hyperbolic invariant manifold), and by opposition, the systems we are considering are called "a priori" stable. These simpler "a priori" unstable systems are now well-understood, and many results confirm that instability occurs for a generic perturbation, see for instance [Tre04], [CY04], [DdlLS06], [GR07], [Ber08], [DH09], [CY09], and [GR09].

1.1.4 "A priori" stable systems

The situation for "a priori" stable systems is much more complicated. In [Mat04] (see also [Mat12] for a recent corrected version), Mather announced a proof of Arnold conjecture in a special case, that is a strong form of Arnold diffusion for a generic time-dependent perturbation of a convex integrable Hamiltonian with $n = 2$ (and also for a generic time-independent perturbation of a quasi-convex integrable Hamiltonian with $n = 3$) based on his variational techniques. Mather never gave a complete proof of the announced results, but his work and unpublished preprints played a fundamental role in the subsequent developments. First, Bernard, Kaloshin and Zhang in [BKZ11] proved a weaker form of Arnold conjecture, still with the convexity requirement but for an arbitrary number of degrees of freedom. Then, Kaloshin and Zhang in [KZ12] proved the strong form of Arnold conjecture, for $n = 2$, h convex and f time-periodic. A different approach to this problem was proposed by Cheng ([Che13]) and announced by Marco ([Mar12a], [Mar12b]).

1.1.5 Role of normally hyperbolic invariant cylinders

The central and common point in all these works, which was not present in the work of Mather, is the use of normally hyperbolic invariant manifolds as a "skeleton" for the unstable orbits. Construction of such cylinders under generic hypothesis relies on resonant normal forms and the theory of normally hyperbolic manifolds, and they were already discussed in [KZZ10] and [Ber10].

On the other hand, most of these works do rely strongly on Mather's variational techniques, once the normally hyperbolic invariant manifolds have been constructed. It has to be noted that these variational techniques, and to a lesser extent, the existence of normally hyperbolic cylinders, use in an essential way the convexity assumption, so that none of these works apply to non-convex integrable Hamiltonians.

1.1.6 Non-convex systems

It can be said that a typical non-degenerate integrable system (in the sense of Kolmogorov, Arnold or Nekhoroshev) is non-convex nor quasi-convex, but for these systems, essentially nothing is known: for the simplest integrable Hamiltonians h which are non-convex nor quasi-convex but steep and non-degenerate (in the sense of Kolmogorov for $n \geq 2$ or Arnold for $n \geq 3$), it is not even known how to construct a single f such that $H = h + \varepsilon f$ has unstable orbits. This is a bit paradoxical from the point of view of Nekhoroshev estimates, as the time of diffusion for perturbations of steep non-convex integrable Hamiltonians should be smaller and hence diffusion should be easier to observe.

Another evidence of the difficulty connected to the lack of convexity is the problem of existence of periodic orbits (not to mention, of course, the problem of extending Aubry-Mather theory). Recall that Bernstein and Katok proved, in [BK87], that given a periodic tori (filled with periodic orbits of common period) of a convex integrable Hamiltonian system with n degrees of freedom, at least n periodic orbits persist after a small perturbation, no matter how large is the period (that is, the threshold of the perturbation is independent of the period). Moreover, these periodic orbits are uniformly (with respect to the period) close to the unperturbed ones. For non-convex non-degenerate integrable Hamiltonian system, this was partly generalized in [Che92]: one can still get the existence of at least n periodic orbits but without the uniform estimates (one can find in [Che92] an example due to Herman which shows that these uniform estimates are indeed not possible without convexity).

1.1.7 Examples of non-convex non-steep Hamiltonians

Yet for some non-convex non-steep integrable Hamiltonians, the construction of examples of instability is much easier and has been known for a long time. A prototype of such an integrable Hamiltonian with two degrees of freedom, which can be found in [Nek77] (but a completely analogous example, in a slightly different setting, was already considered in [Mos60]), is given by $h(I_1, I_2) = \frac{1}{2}(I_1^2 - I_2^2)$: letting $f(\theta_1, \theta_2) = (2\pi)^{-1} \sin(2\pi(\theta_1 - \theta_2))$, the system $H = h + \varepsilon f$ admits the unstable solution $I(t) = (-\varepsilon t, \varepsilon t)$, $\theta(t) = -\frac{1}{2}(\varepsilon t^2, \varepsilon t^2)$. This Hamiltonian h is obviously non-convex, but it is also non-steep since the restriction of h to the lines $\{I_1 \pm I_2 = 0\}$ is constant so this restriction has only critical points, which are thus non-isolated. Also it is degenerate in the sense of Arnold, so diffusion can and do already occur for $n = 2$, even though it is non degenerate in the sense of Kolmogorov so that it admits many invariant tori (circles). Moreover, the time of diffusion in this example is the smallest possible, as it is linear with respect to the inverse of ε. Let us point out that for integrable systems with two degrees of freedom, the Arnold non-degeneracy condition is in fact equivalent to quasi-convexity which is also equivalent to steepness.

It is obvious that the above example can be generalized to the case where h is a quadratic form that has an isotropic vector with rational components (such a quadratic form is thus indefinite) for any number of degrees of freedom n, but in fact more is true. On the one hand, it was proved by Nekhoroshev in [Nek79] that one can consider an even more general class of integrable Hamiltonians. Indeed, suppose there exist an affine subspace L of $\mathbb{R}^n$, whose associated vector space is spanned by vectors with rational components, and a curve $\sigma : [0, 1] \to L$ such that the gradient of the restriction of h to L vanishes identically along σ. Then one can construct an arbitrarily small perturbation εf such that the system $H = h + \varepsilon f$ has an orbit $(\theta(t), I(t))$ for which $I(0) = \sigma(0)$ and $I(\tau) = \sigma(1)$, where τ proportional to ε^{-1}.

On the other hand, in the case where h is a quadratic form with a rational isotropic vector, Herman had constructed examples of perturbation for which one can find a dense G_δ set of initial conditions leading to orbits whose action components are unbounded (note that such a quadratic form can be non-degenerate, so at the same time most of the orbits lie on invariant tori). We refer to [Her92] for such examples and many other interesting examples related to non-convex integrable Hamiltonians.

1.1.8 Preliminary description of the class of Hamiltonians studied in the paper

The examples above are rather specific. For instance, for the prototype of non-convex integrable Hamiltonian $h(I_1, I_2) = \frac{1}{2}(I_1^2 - I_2^2)$, the perturbation f does not depend on the action variables and more importantly, it depends only on a specific combination of the angular variables. The purpose of this paper is to investigate the question whether such a phenomenon remains true for a generic perturbation. We will show in Theorem 1.1 that we have diffusion for a class of non-convex non-steep Hamiltonians h with two degrees of freedom, which includes the example $h(I_1, I_2) = \frac{1}{2}(I_1^2 - I_2^2)$ as a particular case, and for an open and dense set of perturbations, with a time of diffusion which is linear with respect to the inverse of ε. Under stronger assumptions, we will also prove a stronger statement of diffusion in Theorem 1.2. The conditions defining this class of integrable Hamiltonians h is the existence of a segment with rational slope contained in an energy level of h such that the gradient ∇h does not vanish along this segment (these are the assumptions $(A.1)$ and $(A.2)$ in §1.2). For integrable Hamiltonians which are compatible with "fast" diffusion (that is, with a time of diffusion which is linear with respect to the inverse of ε) for some perturbation, we expect these conditions to be quite sharp. The set of admissible perturbations is very easily described: we only required that some "averaged" perturbation is a non-constant function. Moreover, we only require h and f to be of finite regularity.

1.1.9 Heuristic description of the proof

Let us now explain the idea of the proof on the specific example $h(I_1, I_2) = I_1^2 - I_2^2$. Up to a linear symplectic change of coordinates, we can equivalently consider $h(I_1, I_2) = I_1 I_2$. If the perturbation f depends only on θ_1 and is non-constant, then we are essentially back to the example described in §1.1.7. Indeed, the equations of motion in this case are

$$\dot{I}_1(t) = -\varepsilon f'(\theta_1(t)), \quad \dot{I}_2(t) = 0, \quad \dot{\theta}_1(t) = I_2(t), \quad \dot{\theta}_2(t) = I_1(t)$$

so that if we choose an initial condition $(\theta(0), I(0))$ with $f'(\theta_1(0)) \neq 0$ (which is possible since f is non constant) and $I_2(0) = 0$, then $I_2(t) = 0$ and $\theta_1(t) = \theta_1(0)$ for all time $t \in \mathbb{R}$ and so the I_1 variable drift with a speed of order ε:

$$I_1(t) = I_1(0) - t\varepsilon f'(\theta_1(0)).$$

One should observe that this drift happens because the solution is locked in the resonance $\{I_2 = 0\}$ for all time.

Now for a general perturbation f depending on (θ, I), on a suitable domain one can average the θ_2 variables: more precisely the Hamiltonian can be conjugated to a Hamiltonian of the form $h(I) + \varepsilon \bar{f}(\theta_1, I) + \varepsilon^2 f'(\theta, I)$, where $\bar{f}$ is obtained from f by averaging over θ_2. This is close to the example we described before, except that $\bar{f}$ depends also on the action variables

and there is a remainder of order ε^2, so we cannot find a solution which stays locked in the resonance $\{I_2 = 0\}$ for all time. However, it is still possible to find a solution which stays ε-close to the resonance $\{I_2 = 0\}$ for a time of order ε^{-1} and this is sufficient to prove that the I_1 variable can drift independently of ε, provided of course that the function $\theta_1 \mapsto \bar{f}(\theta_1, I)$ is non-constant for a fixed value of I (this is satisfied for an open and dense set of perturbations).

The main difficulty is that usually, the domain on which one can conjugated the Hamiltonian to the special form we described above (which is called a resonant normal form with a remainder) is ε-dependent and in the space of action, its diameter goes to zero as ε goes to zero. We will actually prove (in Proposition 2.3) that in our situation, we can construct such a conjugacy on a domain in the action space which contains a segment, in the I_1 direction, whose length is independent of ε, so that the existence of a drifting orbit for the normal form will actually yield the existence of a drifting orbit for the original Hamiltonian (this is Theorem 1.1). Assuming moreover that f is action-independent and constructing a more accurate normal form with a remainder of order ε^3 instead of ε^2 (as in Proposition 2.4), we will also show that the size of the drift can be as large as we want (this is Theorem 1.2).

Let us point out that the main ingredient of the proof, which is the normal form, is similar in spirit to the normal form constructed in [BKZ11]. However, in [BKZ11] the normal form is used for another purpose (namely to construct a normally hyperbolic invariant cylinder which is then used to locate an unstable orbit) so they only need the remainder to be of order $\delta\varepsilon$, for some $\delta > 0$ independent of ε. In our case, we need a stronger statement (a remainder of order at least ε^2) in order to derive the existence of an unstable orbit directly from the normal form.

1.1.10 Prospects

To conclude, let us note that the statement of Theorem 1.1 gives a diffusion in a weak sense, that is the action variables drift independently of ε for all ε sufficiently small, but we cannot find an orbit which connects arbitrary neighbourhoods in the space of action. Also, for the moment, it is restricted to two degrees of freedom, which is the minimal number of degrees of freedom for which instability can occur for Arnold degenerate integrable systems. The normal form we used is in fact valid for any number of degrees of freedom, but in general it appears too weak to derive the result directly from it, and therefore we expect that additional restrictions on the set of admissible perturbations has to be imposed for more degrees of freedom. We plan to come back to these issues in a subsequent work.

1.2 Main results

1.2.1 Geometric assumptions

Given $R > 0$, let B_R be the closed ball of $\mathbb{R}^2$ of radius R with respect to the supremum norm $|\,.\,|$, that is $B_R = \{(I_1, I_2) \in \mathbb{R}^2 \mid |I_1| \leq R, \ |I_2| \leq R\}$. Our integrable Hamiltonian h will be a function $h : B_R \to \mathbb{R}$ of class C^4, which satisfy the following two conditions:

$(A.1)$ There exist a vector $k = (k_1, k_2) \in \mathbb{Z}^2 \setminus \{0\}$, a constant $a \in \mathbb{R}$ and a closed segment $S \subseteq L \cap B_R$, where $L = \{(I_1, I_2) \in \mathbb{R}^2 \mid k_1 I_1 + k_2 I_2 + a = 0\}$, such that the restriction of h to S is constant.

$(A.2)$ There exists a closed segment $S^* \subseteq S$, such that for all $I \in S^*$, $\nabla h(I) \neq 0$.

Such a segment S is sometimes called a channel of superconductivity. Note that the condition $(A.1)$ obviously rules out convex functions, but it also rules out steep functions. Indeed, $(A.1)$ is equivalent to the assertion that the gradient of $h_{|L}$, the restriction of h to L, vanishes identically on S, therefore the function $h_{|L}$ has a set of critical points which contains S and hence is non-isolated. As for the condition $(A.2)$, it is a non-degeneracy assumption, as we want to avoid that the gradient of h vanishes identically on S: note that $(A.2)$ is satisfied if there exists $I^* \in S$ such that $\nabla h(I^*) \neq 0$. The condition $(A.1)$ is crucial, whereas $(A.2)$ is somehow just technical, as we believe it can be removed in general. Following the terminology of [Bou12], functions which do satisfy $(A.1)$ are functions which are not rationally steep.

1.2.2 Regularity assumptions

Given a small parameter $0 < \varepsilon < 1$, our perturbation εf will be a "generic" function $\varepsilon f :$ $\mathbb{T}^2 \times B_R \to \mathbb{R}$ which is "small" for the C^r topology, for r sufficiently large. For an integer $r \geq 2$, let $C^r(\mathbb{T}^2 \times B_R)$ the space of C^r function $f : \mathbb{T}^2 \times B_R \to \mathbb{R}$, which is Banach space with respect to the norm

$$|f|_{C^r(\mathbb{T}^2 \times B_R)} = \sup_{j \in \mathbb{N}^4, \, |j| \leq r} \left(\sup_{(\theta, I) \in \mathbb{T}^n \times B_R} |\partial^j f(\theta, I)| \right)$$

where we have used the standard multi-index notation. We extend the definition of the C^r-norm for vector-valued functions $F = (f_1, \ldots, f_m) : \mathbb{T}^2 \times B_R \to \mathbb{R}^m$, for an arbitrary integer $m \geq 1$, by setting

$$|F|_{C^r(\mathbb{T}^2 \times B_R, \mathbb{R}^m)} = \sup_{1 \leq i \leq m} |f_i|_{C^r(\mathbb{T}^2 \times B_R)}.$$

Let us denote by $C_1^r(\mathbb{T}^2 \times B_R)$ the unit ball of $C^r(\mathbb{T}^2 \times B_R)$ with respect to this norm, that is

$$C_1^r(\mathbb{T}^2 \times B_R) = \{f \in C^r(\mathbb{T}^2 \times B_R) \mid |f|_{C^r(\mathbb{T}^2 \times B_R)} \leq 1\}.$$

Our perturbation εf will be such that f belongs to an open and dense subset $\mathcal{F}_k^r$ of $C_1^r(\mathbb{T}^2 \times B_R)$, depending on the vector k defined in $(A.1)$. For a given function $f \in C_1^r(\mathbb{T}^2 \times B_R)$ and a given I^* in the interior of S^*, we define $\bar{f}_k^* \in C_1^r(\mathbb{T}^2)$ by

$$\bar{f}_k^*(\theta) = \int_0^1 f(\theta + tk, I^*)dt,$$

then $\mathcal{F}_k^r$ is defined by

$$\mathcal{F}_k^r = \{f \in C_1^r(\mathbb{T}^2 \times B_R) \mid \exists I^* \in \text{int}(S^*), \, \exists \theta^* \in \mathbb{T}^2, \, \partial_\theta \bar{f}_k^*(\theta^*) \neq 0\}.$$

In words, $\mathcal{F}_k^r$ is the subset of functions $f \in C_1^r(\mathbb{T}^2 \times B_R)$ such that, for some I^* in the interior of S^*, the associated function $\bar{f}_k^*$ is non-constant: this is obviously an open and dense subset of $C_1^r(\mathbb{T}^2 \times B_R)$. Note that $\bar{f}_k^*$ is a function on $\mathbb{T}^2$, but by definition it is constant on the orbits of the linear flow of frequency k, hence it can be considered as being defined on the space of orbits (the leaf space) of this flow, which is diffeomorphic to $\mathbb{T}$.

1.2.3 Statements

We can finally state our first main result.

Theorem 1.1. *Let $H = h + \varepsilon f$ be defined on $\mathbb{T}^2 \times B_R$, with $h \in C_1^4(B_R)$ satisfying (A.1) and (A.2) and $f \in \mathcal{F}_k^7$. Then there exist positive constants $C \geq 1$ and c, depending only on h, and positive constants $\varepsilon_0 \leq 1$ and $\delta \leq 1$ depending also on f, such that for any $0 < \varepsilon \leq \varepsilon_0$, the Hamiltonian system defined by H has a solution $(\theta(t), I(t))$ such that for $\tau = \delta\varepsilon^{-1}$,*

$$|I(0) - I^*| \leq c\varepsilon, \quad |I(\tau) - I(0)| \geq C\delta^2.$$

Moreover, for all $t \in [0, \tau]$, we have $d(I(t), S^*) \leq c\varepsilon$ where d is the distance induced by the supremum norm.

It is a statement of diffusion for the action variables, in the sense that they have a variation along $S^* \subseteq S$ which is bounded from below independently of ε, for all ε small enough. It has to be noted that the time of diffusion $\tau = \delta\varepsilon^{-1}$ is essentially optimal in the sense that for all $f \in C^2(\mathbb{T}^2 \times B_R) \cap C_1^1(\mathbb{T}^2 \times B_R)$, for all $\varepsilon > 0$ and for all $0 < \delta \leq 1$, we have

$$|I(\tau) - I(0)| \leq \delta$$

for all solutions of $H = h + \varepsilon f$. In particular, for the solution given by Theorem 1.1, one has the inequalities

$$C\delta^2 \leq |I(\tau) - I(0)| \leq \delta.$$

A stronger statement of diffusion can be reached, assuming that the perturbation is independent of the action variables and that it is slightly more regular. Indeed, let us define

$$\mathcal{G}_k^r = \{f \in C_1^r(\mathbb{T}^2) \mid \exists \theta^* \in \mathbb{T}^2, \ \partial_\theta \bar{f}_k(\theta^*) \neq 0\}$$

where

$$\bar{f}_k(\theta) = \int_0^1 f(\theta + tk)dt.$$

Then we can state our second main result.

Theorem 1.2. *Let $H = h + \varepsilon f$ be defined on $\mathbb{T}^2 \times B_R$, with $h \in C_1^{10}(B_R)$ satisfying (A.1) and (A.2) and $f \in \mathcal{G}_k^{19}$. Then, given any two points $I' \in S^*$ and $I'' \in S^*$, there exists a positive constant c, depending only on h, and positive constants $\varepsilon_0 \leq 1$ and $\delta \leq 1$ depending also on f and on the distance between I' and I'', such that for any $0 < \varepsilon \leq \varepsilon_0$, the Hamiltonian system defined by H has a solution $(\theta(t), I(t))$ such that for $\tau \leq \delta\varepsilon^{-1}$,*

$$|I(0) - I'| \leq c\varepsilon, \quad |I(\tau) - I''| \leq c\varepsilon.$$

Moreover, for all $t \in [0, \tau]$, we have $d(I(t), S^*) \leq c\varepsilon$.

The conclusion of Theorem 1.2 is indeed stronger than the conclusion of Theorem 1.1, as we not only have a variation along $S^* \subseteq S$ which is independent of ε, but we can also connect ε-neighborhoods of any two points in S^*. It should be noted that both theorems give a new obstruction to extend Nekhoroshev estimates in the non-steep case, even if one is willing to consider only a generic perturbation.

Concerning the dependence of the constants involved, the dependence on h is only through R, the vector k and the constant a that appeared in (A.1), the length of the segment S^*

that appeared in $(A.2)$ and a lower bound on the norm of $\nabla h(I)$ for $I \in S^*$ (which is positive by $(A.2)$), while the dependence on f is through the absolute value of $\partial_\theta \bar{f}_k^*(\theta^*)$ (respectively $\partial_\theta \bar{f}_k(\theta^*)$) for Theorem 1.1 (respectively for Theorem 1.2) and the distance of I^* to the boundary of S^* for Theorem 1.1. We refer to Theorem 2.1 and Theorem 2.2 in §2.1 for more concrete and precise statements.

1.2.4 Comments on the regularity assumptions

The first statement (Theorem 1.1) requires the integrable part to be C^4 and the perturbation to be C^7, while the second statement (Theorem 1.2) requires the integrable part to be C^{10} and the perturbation to be C^{19}. Theses regularity assumptions are far from being optimal, and no efforts were made to improve them.

For instance, using analytic approximations as in [BKZ11] or polynomial approximations, it is certainly possible to lower these regularities. Moreover, concerning the first statement, one can use a different method which would give the same result assuming only that the integrable part and the perturbation are of class C^3 (unfortunately, this method cannot be applied directly to prove the second statement).

1.2.5 Comments on the geometric assumptions

Let us now discuss some particular cases of functions h satisfying $(A.1)$ and $(A.2)$, and therefore for which one has diffusion for a generic perturbation. As we will explain later, we can always assume without loss of generality that $a = 0$ in $(A.1)$, and upon adding an irrelevant additive constant, we can assume that the restriction of h to S is identically zero.

For a linear Hamiltonian $h(I) = \omega \cdot I$, it follows that $(A.1)$ and $(A.2)$ are satisfied if and only if ω is resonant, that is $l \cdot \omega = 0$ for some $l \in \mathbb{Z}^2 \setminus \{0\}$, and ω is non-zero. On the other hand, if ω is non-resonant, it follows from [Bou12] that the statement of Theorem 1.1 cannot be true since for all sufficiently small perturbation, one has stability for an interval of time which is strictly larger than $[-\tau, \tau]$ with τ as above. In particular, if ω is Diophantine, one has stability for an interval of time which is exponentially long with respect to ε^{-1}, up to an exponent depending only on the Diophantine exponent of ω.

Now for a quadratic Hamiltonian $h(I) = AI \cdot I$ where A is a 2 by 2 symmetric matrix, $(A.1)$ and $(A.2)$ are satisfied if and only if there exists a vector $l \in \mathbb{Z}^2 \setminus \{0\}$ such that $Al \cdot l = 0$ and $Al \neq 0$. Assuming that A is diagonal, its eigenvalues have to be of different sign, and writing $h(I) = \alpha_1^2 I_2 - \alpha_2^2 I_2^2$, $(A.1)$ and $(A.2)$ are satisfied if and only if $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ and $\alpha_2/\alpha_1 \in \mathbb{Q}$. The example described in the introduction corresponds to $\alpha_1 = \alpha_2 = 1$. On the other hand, one knows that if α_2/α_1 is irrational, the statement of Theorem 2.1 cannot be true for any sufficiently small perturbation for the same reason as above: for instance, if α_2/α_1 is a Diophantine number, the quadratic Hamiltonian falls into the class of Diophantine steep functions introduced in [Nie07] and it follows from results in [Nie07] or [BN12] that such Hamiltonians are stable for an exponentially long interval of time.

Note that in these two special cases, the condition $(A.2)$, which amounts to $\omega \neq 0$ in the first case and $Al \neq 0$ in the second case, can be easily removed.

We already explained that the time of diffusion τ is in some sense optimal, regardless of the integrable Hamiltonian h. Now we believe that if we fix the time of diffusion, the condition $(A.1)$ on the integrable Hamiltonian h is also in some sense optimal, as if h does not satisfy this assumption, one can have diffusion but with a time strictly greater than τ.

This is indeed the case for linear or quadratic integrable Hamiltonians as we described above, and the general case is conjectured in [Bou12].

2 Proof of Theorem 1.1 and Theorem 1.2

In §2.1, we will perform some preliminary transformations to reduce Theorem 1.1 and Theorem 1.2 to equivalent but more concrete statements, which are Theorem 2.1 and Theorem 2.2. Theorem 2.1 and Theorem 2.2 will be proved in §2.3, based on normal form results which are stated and proved in §2.2.

2.1 Preliminary reductions

2.1.1 Preliminary transformations

First we may assume that the line L in $(A.1)$ passes through the origin, that is $L = \{(I_1, I_2) \in \mathbb{R}^2 \mid k_1 I_1 + k_2 I_2 = 0\}$: indeed, we can always find a translation of the action variables $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that T sends $\{(I_1, I_2) \in \mathbb{R}^2 \mid k_1 I_1 + k_2 I_2 = 0\}$ to $\{(I_1, I_2) \in \mathbb{R}^2 \mid k_1 I_1 + k_2 I_2 + a = 0\}$, and since the map $\Phi_T(\theta, I) = (\theta, TI)$ is symplectic, the statement holds true for H if and only if it holds true for $H \circ \Phi_T$, up to constants depending on a.

Then we can suppose that the components of the vector $k = (k_1, k_2) \in \mathbb{Z}^2 \setminus \{0\}$ are relatively prime, since changing k by k/p, where p is the greatest common divisor of k_1 and k_2, does not change the definition of L. Hence we may assume that in fact $k = e_2 = (0, 1)$, that is $L = \{(I_1, I_2) \in \mathbb{R}^2 \mid I_2 = 0\}$: indeed we can always find a matrix $M \in GL_2(\mathbb{Z})$ such that its second row is k, hence $Me_2 = k$ and $^tM^{-1}$ sends $\{(I_1, I_2) \in \mathbb{R}^2 \mid I_2 = 0\}$ to $\{(I_1, I_2) \in \mathbb{R}^2 \mid k_1 I_1 + k_2 I_2 = 0\}$. The map $\Phi_M(\theta, I) = (M\theta, {}^tM^{-1}I)$ is well-defined since $M\mathbb{T}^2 = \mathbb{T}^2$, and it is symplectic, so the statement holds true for H if and only if it holds true for $H \circ \Phi_M$, up to constants depending on k.

Note that the symplectic transformations Φ_T and Φ_M do change the domain B_R in the space of actions, but to simplify the notations, we will assume that the latter is fixed.

2.1.2 Simplified assumptions

Now for all $I = (I_1, I_2) \in B_R$, let us write

$$\nabla h(I) = \omega(I) = (\omega_1(I), \omega_2(I)) = (\omega_1(I_1, I_2), \omega_2(I_1, I_2)) \in \mathbb{R}^2.$$

Since $L = \{(I_1, I_2) \in \mathbb{R}^2 \mid I_2 = 0\}$, S, which by definition is a closed segment contained in $L \cap B_R$, is of the form $S = \{(I_1, 0) \in \mathbb{R}^2 \mid I_1 \in S_1\}$ where S_1 is a closed segment of $\mathbb{R}$ contained in $[-R, R]$. Similarly, $S^* = \{(I_1, 0) \in \mathbb{R}^2 \mid I_1 \in S_1^*\}$ where S_1^* is contained in the interior of S_1. The condition $(A.1)$ is then obviously equivalent to $\partial_{I_1} h(I_1, 0) = \omega_1(I_1, 0) = 0$ for all $I_1 \in S_1$, while the condition $(A.2)$ is that $\omega_2(I_1, 0) \neq 0$ for all $I_1 \in S_1^*$. Changing H to $-H$ if necessary and reversing the time accordingly, we may assume that $\omega_2(I_1, 0) \geq \varpi > 0$ for all $I_1 \in S_1^*$.

We can eventually formulate simplified conditions, that we call $(B.1)$ and $(B.2)$:

$(B.1)$ There exists a closed segment $S_1 \subseteq [-R, R]$ such that for all $I_1 \in S_1$, we have $\omega_1(I_1, 0) = 0$.

(B.2) There exist a closed segment $S_1^* \subseteq S_1$ and $\varpi > 0$ such that for all $I_1 \in S_1^*$, we have $\omega_2(I_1, 0) \geq \varpi$.

Then the definition of $\mathcal{F}_{e_2}^r$ also simplifies: one easily check that for $f \in C_1^r(\mathbb{T}^2 \times B_R)$, we have, for some $I^* = (I_1^*, 0)$ in the interior of S^*, $\bar{f}_{e_2}^* \in C_1^r(\mathbb{T})$ where

$$\bar{f}_{e_2}^*(\theta_1) = \int_{\mathbb{T}} f(\theta_1, \theta_2, I^*) d\theta_2$$

so that $f \in \mathcal{F}_{e_2}^r$ if and only if there exist $I^* = (I_1^*, 0)$ in the interior of S^* and $\theta_1^* \in \mathbb{T}$ for which $\partial_{\theta_1} \bar{f}_{e_2}^*(\theta_1^*) \neq 0$. The definition of $\mathcal{G}_{e_2}^r$ is analogous. For simplicity, we write $\bar{f}_{e_2}^* = \bar{f}^*$ and $\mathcal{F}_{e_2}^r = \mathcal{F}^r$, $\bar{f}_{e_2} = \bar{f}$ and $\mathcal{G}_{e_2}^r = \mathcal{G}^r$, and for $f \in \mathcal{F}^r$ (respectively $f \in \mathcal{G}^r$), we denote by λ a lower bound on the absolute value of $\partial_{\theta_1} \bar{f}^*(\theta_1^*)$ (respectively a lower bound on the absolute value of $\partial_{\theta_1} \bar{f}(\theta_1^*)$). For Theorem 1.1, we denote by δ^* the distance of I_1^* to the boundary of S_1^*, and for Theorem 1.2, we denote by ρ the distance between $I_1' \in S_1^*$ and $I_1'' \in S_1^*$, where $I' = (I_1', 0)$ and $I'' = (I_1'', 0)$.

2.1.3 Simplified statements

From the previous discussion, it follows that Theorem 1.1 and Theorem 1.2 are implied by the following statements.

Theorem 2.1. *Let $H = h + \varepsilon f$ be defined on $\mathbb{T}^2 \times B_R$, with $h \in C_1^4(B_R)$ satisfying (B.1) and (B.2) and $f \in \mathcal{F}^7$. Then there exist positive constants $C \geq 1$ and c depending only on R, the length of S_1^* and ϖ, and a positive constant $\varepsilon_0 \leq 1$ depending also on λ, such that for any $0 < \varepsilon \leq \varepsilon_0$, if we set $\delta = \min\{\lambda(4C)^{-1}, \delta^*\}$, the Hamiltonian system defined by H has a solution $(\theta(t), I(t))$ such that for $\tau = \delta\varepsilon^{-1}$,*

$$|I_1(0) - I_1^*| \leq c\varepsilon, \quad |I_1(\tau) - I_1(0)| \geq C\delta^2.$$

Moreover, for all $t \in [0, \tau]$, we have $|I_2(t)| \leq c\varepsilon$ and $d(I_1(t), S_1^*) \leq c\varepsilon$.

Theorem 2.2. *Let $H = h + \varepsilon f$ be defined on $\mathbb{T}^2 \times B_R$, with $h \in C_1^{10}(B_R)$ satisfying (B.1) and (B.2) and $f \in \mathcal{F}^{19}$. Then there exists a positive constant c, depending only on R, the length of S_1^*, and ϖ, and a positive constant $\varepsilon_0 \leq 1$ depending also on λ and ρ, such that for any $0 < \varepsilon \leq \varepsilon_0$, if we set $\delta = 2\rho\lambda^{-1}$, the Hamiltonian system defined by H has a solution $(\theta(t), I(t))$ such that for $\tau \leq \delta\varepsilon^{-1}$,*

$$|I_1(0) - I_1'| \leq c\varepsilon, \quad |I_1(\tau) - I_1''| \leq c\varepsilon.$$

Moreover, for all $t \in [0, \tau]$, we have $|I_2(t)| \leq c\varepsilon$ and $d(I_1(t), S_1^*) \leq c\varepsilon$.

2.2 Normal forms

2.2.1 Domain of the normal forms

The main ingredient of the proofs of Theorem 2.1 and Theorem 2.2 will be normal forms on a domain which, in the space of action, is centred around S^*. In particular, in the direction given by the first action variables I_1, it contains the segment S_1^* whose length is independent

of ε. For a constant $\kappa > 0$ to be determined later (in the proof of Proposition 2.3), let us consider the $\kappa\varepsilon$-neighbourhood of S^* in $\mathbb{R}^2$

$$S^*(\kappa\varepsilon) = \{I \in \mathbb{R}^2 \mid d(I, S^*) \le \kappa\varepsilon\} = \{(I_1, I_2) \in \mathbb{R}^2 \mid d(I_1, S_1^*) \le \kappa\varepsilon, \ |I_2| \le \kappa\varepsilon\}.$$

We also define the domain $\mathcal{D}^*(\kappa\varepsilon) = \mathbb{T}^2 \times S^*(\kappa\varepsilon)$. Let r be an integer (we will choose $r = 7$ for Theorem 2.1 and $r = 19$ for Theorem 2.2, but for convenience we consider it as a free parameter for the moment).

In the sequel, to avoid cumbersome notations, when convenient we will use a dot $\cdot$ in replacement of any constant depending only on r, R, the length of S_1^* and ϖ, that is for any two quantities u and v, an expression $u \lessdot v$ means that there exists a constant c depending only on r, R, the length of S_1^* and ϖ such that $u \le cv$. Similarly, we will use the notation $u = \cdot v$. For instance, we will have $\kappa = \cdot 1$.

To simplify the exposition further, for any integer $j \le r$ we will simply denote by $|\,.\,|_j$ the C^j norm of a function or a vector-valued function, without referring to its domain of definition nor to the domain where it takes values.

2.2.2 One-step normal form

Proposition 2.3. *Let $H = h + \varepsilon f$ be defined on $\mathbb{T}^2 \times B_R$, $l \ge 1$ an integer, and assume that $h \in C_1^{l+3}(B_R)$ satisfies (B.1) and (B.2) and $f \in C_1^r(\mathbb{T}^2 \times B_R)$, for $r \ge 2(l+1) + 3$. Assume that $\varepsilon \lessdot 1$. Then there exists a symplectic embedding*

$$\Phi : \mathcal{D}^*(\kappa\varepsilon/2) \to \mathcal{D}^*(\kappa\varepsilon)$$

of class C^{l+1} such that

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon^2 f', \quad \bar{f}(\theta_1, I) = \int_{\mathbb{T}} f(\theta_1, \theta_2, I)d\theta_2$$

and we have the following estimates

$$|\Phi - \mathrm{Id}|_0 \le \kappa\varepsilon/2, \quad |f'|_l \lessdot 1.$$

The proof of this proposition uses some elementary estimates which are recalled in the Appendix A.

Proof of Proposition 2.3. First of all, since $\varepsilon \lessdot 1$, we can assume that $S^*(\kappa\varepsilon)$ is included in B_R. For a function $\chi : \mathcal{D}^*(\kappa\varepsilon) \to \mathbb{R}$ of class C^{l+2} to be chosen below, the transformation Φ in the statement will be obtained as the time-one map of the Hamiltonian flow generated by $\varepsilon\chi$. Let $X_{\varepsilon\chi}$ be the Hamiltonian vector field generated by $\varepsilon\chi$, and $X_{\varepsilon\chi}^t$ the time-t map. Assuming that $X_{\varepsilon\chi}^t$ is well-defined on $\mathcal{D}^*(\kappa\varepsilon/2)$ for $|t| \le 1$, let $\Phi = X_{\varepsilon\chi}^1$. Using the relation

$$\frac{d}{dt}\left(G \circ X_{\varepsilon\chi}^t\right) = \varepsilon\{G, \chi\} \circ X_{\varepsilon\chi}^t$$

for an arbitrary function G, and writing

$$H \circ \Phi = h \circ \Phi + \varepsilon f \circ \Phi$$

we can apply Taylor's formula to the right-hand side of the above equality, at order two for the first term and at order one for the second term, to get

$$
\begin{aligned}
H \circ \Phi & = h + \varepsilon\{h, \chi\} + \varepsilon^2 \int_0^1 (1-t)\{\{h, \chi\}, \chi\} \circ X_{\varepsilon\chi}^t dt + \varepsilon f + \varepsilon^2 \int_0^1 \{f, \chi\} \circ X_{\varepsilon\chi}^t dt \\
& = h + \varepsilon(\{h, \chi\} + f) + \varepsilon^2 \int_0^1 \{(1-t)\{h, \chi\} + f, \chi\} \circ X_{\varepsilon\chi}^t dt \\
& = h + \varepsilon\bar{f} + \varepsilon(\{h, \chi\} + f - \bar{f}) + \varepsilon^2 \int_0^1 \{(1-t)\{h, \chi\} + f, \chi\} \circ X_{\varepsilon\chi}^t dt \quad (1)
\end{aligned}
$$

where $\bar{f}$ is the function defined in the statement. It would be natural to choose χ to solve the equation $\{h, \chi\} + g = 0$ where $g = f - \bar{f}$, which can be written again as $\{\chi, h\} = g$, but we will only solve this equation approximatively.

We expand f in Fourier series with respect to the variables θ:

$$
f(\theta, I) = \sum_{k \in \mathbb{Z}^2} f_k(I) e^{2\pi i k \cdot \theta} = \sum_{(k_1, k_2) \in \mathbb{Z}^2} f_{(k_1, k_2)}(I) e^{2\pi i (k_1 \theta_1 + k_2 \theta_2)}
$$

and for a parameter $K \geq 1$ to be chosen below, we write

$$
f(\theta, I) = f_K(\theta, I) + f^K(\theta, I) = \sum_{k \in \mathbb{Z}^2, \, |k| \leq K} f_k(I) e^{2\pi i k \cdot \theta} + \sum_{k \in \mathbb{Z}^2, \, |k| > K} f_k(I) e^{2\pi i k \cdot \theta}.
$$

Instead of solving the equation $\{\chi, h\} = g$ with $g = f - \bar{f}$, we will actually choose χ to solve the equation $\{\chi, h\} = g_K$ where $g_K = f_K - \bar{f}_K$. Observe that

$$
\bar{f}(\theta_1, I) = \sum_{k_1 \in \mathbb{Z}} f_{(k_1, 0)}(I) e^{2\pi i k_1 \theta_1}
$$

and hence

$$
\bar{f}_K(\theta_1, I) = \sum_{k_1 \in \mathbb{Z}, \, |k_1| \leq K} f_{(k_1, 0)}(I) e^{2\pi i k_1 \theta_1}.
$$

It is then easy to check that the solution of $\{\chi, h\} = g_K$, which can be written again as

$$
\omega(I) \cdot \partial_\theta \chi(\theta, I) = g_K(\theta, I), \quad (\theta, I) \in \mathcal{D}^*(\kappa\varepsilon) \quad (2)
$$

is given by $\chi(\theta, I) = \sum_{k \in \mathbb{Z}^2, \, |k| \leq K} \chi_k(I) e^{2\pi i k \cdot \theta}$ where

$$
\chi_k(I) = \begin{cases} (2\pi i k \cdot \omega(I))^{-1} f_k(I), & |k| \leq K, \, k_2 \neq 0 \\ 0, & |k| \leq K, \, k_2 = 0. \end{cases} \quad (3)
$$

Since h is C^{l+3}, χ is C^{l+2} provided we can prove that $k \cdot \omega(I)$ is non-zero, for $|k| \leq K$ such that $k_2 \neq 0$ and $I \in S^*(\kappa\varepsilon)$. By definition, given $I \in S^*(\kappa\varepsilon)$ we can find $\tilde{I} \in S^*$ such that $|I - \tilde{I}| \leq \kappa\varepsilon$ and therefore $|\omega(I) - \omega(\tilde{I})| \leq \kappa\varepsilon$ since $h \in C_1^{l+3}(B_R)$. For any $k \in \mathbb{Z}^2$ such that $k_2 \neq 0$, by (B.1) we have $\omega_1(\tilde{I}) = \omega_1(\tilde{I}_1, 0) = 0$ (since $\tilde{I} = (\tilde{I}_1, 0) \in S^* \subseteq S$). Moreover by (B.2) we have $\omega_2(\tilde{I}_1, 0) \geq \varpi$ (since $\tilde{I}_1 \in S_1^*$), hence

$$
|k \cdot \omega(\tilde{I})| = |k \cdot \omega(\tilde{I}_1, 0)| = |k_1 \omega_1(\tilde{I}_1, 0) + k_2 \omega_2(\tilde{I}_1, 0)| = |k_2| \omega_2(\tilde{I}_1, 0) \geq \varpi.
$$

It follows that for $k \in \mathbb{Z}^2$ such that $k_2 \neq 0$ and $|k| \leq K$, and for $I \in S^*(\kappa\varepsilon)$, we have

$$|k \cdot \omega(I)| \geq |k \cdot \omega(\tilde{I})| - |k||\omega(I) - \omega(\tilde{I})| \geq \varpi - K\kappa\varepsilon \geq \varpi/2 \tag{4}$$

provided we define $K = \varpi(2\kappa\varepsilon)^{-1} =\cdot \varepsilon^{-1}$. Now $f \in C_1^r(\mathbb{T}^2 \times B_R)$, so an integration by parts gives that $|f_k|_j \lessdot |k|^{j-r}|f|_r \lessdot |k|^{j-r}$ for any $j \leq r$, and therefore using (3), (4) and Leibniz formula (inequality (19) of Appendix A), we obtain $|\chi_k|_{l+1} \lessdot |k|^{l+1-r}$. Since $r \geq 2(l+1) + 3$, we can therefore bound the C^{l+1} norm of χ independently of K as

$$|\chi|_{l+1} \lessdot \sum_{k \in \mathbb{Z}^2, \, |k| \leq K} |\chi_k|_{l+1}|k|^{l+1} \lessdot \sum_{k \in \mathbb{Z}^2, \, |k| \leq K} |k|^{l+1-r}|k|^{l+1} \lessdot \sum_{k \in \mathbb{Z}^2} |k|^{-3} \lessdot 1$$

and so

$$|\varepsilon\chi|_{l+1} \leq \gamma\varepsilon, \quad \gamma =\cdot 1. \tag{5}$$

It is easy to see that γ is independent of κ, so we now choose $\kappa = 2\gamma =\cdot 1$, and as $\varepsilon \lessdot 1$, we can apply Lemma A.1 of Appendix A: for all $|t| \leq 1$, $X_{\varepsilon\chi}^t : \mathcal{D}^*(\kappa\varepsilon/2) \to \mathcal{D}^*(\kappa\varepsilon)$ is a well-defined symplectic embedding of class C^{l+1} with the estimates

$$|X_{\varepsilon\chi}^t - \mathrm{Id}|_0 \leq \kappa\varepsilon/2, \quad |X_{\varepsilon\chi}^t|_l \lessdot 1. \tag{6}$$

In particular, the first estimate of (6) gives

$$|\Phi - \mathrm{Id}|_0 \lessdot \varepsilon.$$

Now from the equalities (1) and (2) we can write

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon f^K - \varepsilon\bar{f}^K + \varepsilon^2 \int_0^1 \{(t-1)g_K + f, \chi\} \circ X_{\varepsilon\chi}^t dt$$

so that, if we set

$$f' = \varepsilon^{-1}f^K - \varepsilon^{-1}\bar{f}^K + \int_0^1 \{(t-1)g_K + f, \chi\} \circ X_{\varepsilon\chi}^t dt = \varepsilon^{-1}f^K - \varepsilon^{-1}\bar{f}^K + R,$$

then

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon^2 f'.$$

It remains to estimate f'. Using the fact that $|f_k|_j \lessdot |k|^{j-r}|f|_r \lessdot |k|^{j-r}$ for any $j \leq r$, one easily obtain

$$\varepsilon^{-1}|f^K|_l \lessdot \varepsilon^{-1}K^{l+2-r} \lessdot 1 \tag{7}$$

by definition of K. Similarly

$$\varepsilon^{-1}|\bar{f}^K|_l \lessdot 1. \tag{8}$$

Then we have

$$\begin{aligned}
|R|_l &\lessdot |\{(t-1)g_K + f, \chi\}|_l |X_{\varepsilon\chi}^t|_l^l \\
&\lessdot |\{(t-1)g_K + f, \chi\}|_l \\
&\lessdot |(t-1)g_K + f|_{l+1}|\chi|_{l+1} \\
&\lessdot 1
\end{aligned} \tag{9}$$

where we have used Faa di Bruno formula (inequality (21) of Appendix A), the last part of (6), the inequality (20) of Appendix A and the fact that $|g_K|_{l+1} \lessdot 1$, $|f|_{l+1} \lessdot 1$ and $|\chi|_{l+1} \lessdot 1$. The estimates (7), (8) and (9) implies that

$$|f'|_l \lessdot 1$$

which concludes the proof. $\qquad\square$

2.2.3 Two-steps normal form

Proposition 2.4. *Let $H = h + \varepsilon f$ be defined on $\mathbb{T}^2 \times B_R$, $l \geq 1$ and $l' \geq 1$ integers, and assume that $h \in C_1^{l+3}(B_R)$ satisfies $(B.1)$ and $(B.2)$ and $f \in C_1^r(\mathbb{T}^2 \times B_R)$, for $r \geq 2(l+1)+3$ and $l \geq 2(l'+1)+3$. Assume that $\varepsilon < 1$. Then there exists a symplectic embedding*

$$\Phi : \mathcal{D}^*(\kappa\varepsilon/4) \to \mathcal{D}^*(\kappa\varepsilon)$$

of class $C^{l'+1}$ such that

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon^2\bar{f}' + \varepsilon^3 f'', \quad \bar{f}(\theta_1, I) = \int_{\mathbb{T}} f(\theta_1, \theta_2, I)d\theta_2, \quad \bar{f}'(\theta_1, I) = \int_{\mathbb{T}} f'(\theta_1, \theta_2, I)d\theta_2$$

and we have the following estimates

$$|\Phi - \mathrm{Id}|_0 \leq 3\kappa\varepsilon/4, \quad |f'|_l < 1, \quad |f''|_{l'} < 1.$$

The proof of Proposition 2.4 consists essentially of applying twice Proposition 2.3; in particular the estimates are analogous so they will not be repeated below.

Proof of Proposition 2.4. First of all, the assumptions allow us to apply Proposition 2.3: there exists a symplectic embedding

$$\Phi_1 : \mathcal{D}^*(\kappa\varepsilon/2) \to \mathcal{D}^*(\kappa\varepsilon)$$

of class C^{l+1} such that

$$H_1 = H \circ \Phi_1 = h + \varepsilon\bar{f} + \varepsilon^2 f', \quad \bar{f}(\theta_1, I) = \int_{\mathbb{T}} f(\theta_1, \theta_2, I)d\theta_2$$

and we have the following estimates

$$|\Phi_1 - \mathrm{Id}|_0 \leq \kappa\varepsilon/2, \quad |f'|_l < 1.$$

Now consider the Hamiltonian $H_1 = h + \varepsilon\bar{f} + \varepsilon^2 f'$ defined on $\mathcal{D}^*(\kappa\varepsilon/2)$ and of class C^{l+1}. The transformation Φ in the statement will be obtained as a composition $\Phi = \Phi_1 \circ \Phi_2$, where Φ_2 will be the time-one map of the Hamiltonian flow generated by $\varepsilon^2\chi$, for some function $\chi : \mathcal{D}^*(\kappa\varepsilon/2) \to \mathbb{R}$ of class $C^{l'+2}$ to be determined. As before, we write

$$H_1 \circ \Phi_2 = h \circ \Phi_2 + (\varepsilon\bar{f} + \varepsilon^2 f') \circ \Phi_2$$

and by a Taylor expansion we have

$$\begin{aligned}
H_1 \circ \Phi_2 &= h + \varepsilon^2\{h, \chi\} + \varepsilon^4 \int_0^1 (1-t)\{\{h, \chi\}, \chi\} \circ X_{\varepsilon\chi}^t dt \\
&\quad + \varepsilon\bar{f} + \varepsilon^2 f' + \varepsilon^3 \int_0^1 \{\bar{f} + \varepsilon f', \chi\} \circ X_{\varepsilon\chi}^t dt \\
&= h + \varepsilon\bar{f} + \varepsilon^2\bar{f}' + \varepsilon^2(\{h, \chi\} + f' - \bar{f}') + \varepsilon^3 R.
\end{aligned}$$

As before also, we will choose χ to solve the equation $\{\chi, h\} = g'_K$ where $g_K = f'_K - \bar{f}'_K$ and $K = \cdot\varepsilon^{-1}$. Note that H_1 is of class C^{l+1}, but we have a bound only on the C^l norm of f', and as $l \geq 2(l'+1)+3$, it can be proved that χ is $C^{l'+2}$ with

$$|\varepsilon^2\chi|_{l'+1} < \varepsilon^2$$

15

and hence

$$|\Phi_2 - \mathrm{Id}|_0 < \varepsilon^2 \leq \kappa\varepsilon/4$$

since $\varepsilon < 1$. Therefore Φ_2 is well-defined on $\mathcal{D}^*(\kappa\varepsilon/4)$ and of class $C^{l'+1}$, and so $\Phi : \mathcal{D}^*(\kappa\varepsilon/4) \to \mathcal{D}^*(\kappa\varepsilon)$ is of class $C^{l'+1}$ and satisfies

$$|\Phi - \mathrm{Id}|_0 \leq |\Phi_1 - \mathrm{Id}|_0 + |\Phi_2 - \mathrm{Id}|_0 \leq 3\kappa\varepsilon/4.$$

If we set

$$f'' = \varepsilon^{-1} f'^K - \varepsilon^{-1} \bar{f}'^K + R,$$

then it can be proved that $|f''|_{l'} < 1$ and we have

$$H_1 \circ \Phi_2 = h + \varepsilon\bar{f} + \varepsilon^2 \bar{f}' + \varepsilon^3 f''$$

hence

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon^2 \bar{f}' + \varepsilon^3 f''.$$

This concludes the proof. $\qquad\square$

2.3 Proof of Theorem 2.1 and Theorem 2.2

2.3.1 Proof of Theorem 2.1

The proof of Theorem 2.1 is now a consequence of the normal form Proposition 2.3. Since the latter is defined on a domain which contains the segment S^*, whose length is independent of ε, it will be possible to prove the statement of Theorem 2.1 for the normal form $H \circ \Phi$ by analyzing directly the equation of motions, and using the fact that Φ is ε-close to the identity, we will prove that the statement remains true for H.

Proof of Theorem 2.1. Recall that we are considering $H = h + \varepsilon f$ defined on $\mathbb{T}^2 \times B_R$, with $h \in C_1^4(B_R)$ satisfying $(B.1)$ and $(B.2)$ and $f \in \mathcal{F}^7$, so we can apply Proposition 2.3 with $l = 1$: there exist positive constants ε_0' and C' depending only on R, the length of S_1^* and ϖ such that if $\varepsilon \leq \varepsilon_0'$, there exists a symplectic embedding

$$\Phi : \mathcal{D}^*(\kappa\varepsilon/2) \to \mathcal{D}^*(\kappa\varepsilon)$$

of class C^2 such that

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon^2 f', \quad \bar{f}(\theta_1, I) = \int_{\mathbb{T}} f(\theta_1, \theta_2, I) d\theta_2$$

and we have the following estimates

$$|\Phi - \mathrm{Id}|_0 \leq \kappa\varepsilon/2, \quad |f'|_1 \leq C' \tag{10}$$

where $S^*(\kappa\varepsilon) \subseteq B_R$ and $\mathcal{D}^*(\kappa\varepsilon) \subseteq \mathbb{T}^2 \times B_R$ have been defined in §2.2.

Let us consider the Hamiltonian $\tilde{H} = H \circ \Phi$ defined on $\mathcal{D}^*(\kappa\varepsilon/2)$, and we shall write $\Phi(\tilde{\theta}, \tilde{I}) = (\theta, I)$. Since $f \in \mathcal{F}^7$, there exist $I^* = (I_1^*, 0)$ in the interior of S^* and $\theta_1^* \in \mathbb{T}$ such that

$$|\partial_{\tilde{\theta}_1} \bar{f}^*(\theta_1^*)| = |\partial_{\tilde{\theta}_1} \bar{f}(\theta_1^*, I^*)| \geq \lambda. \tag{11}$$

Note that necessarily $\lambda \leq 1$ since $\bar{f} \in C_1^7(\mathbb{T}^2 \times B_R)$, and recall that δ^* is the distance of I^* to the boundary of S^*. Since $\tilde{H}$ is C^2, we can consider a solution $(\tilde{\theta}(t), \tilde{I}(t))$ of the system defined by $\tilde{H}$ with an initial condition $(\tilde{\theta}(0), \tilde{I}(0))$ such that $\tilde{I}(0) = I^*$, $\tilde{\theta}_1(0) = \theta_1^*$ and $\tilde{\theta}_2(0) \in \mathbb{T}$ is arbitrary: we have the equations

$$\begin{cases} \frac{d}{dt}\tilde{I}_1(t) = -\partial_{\tilde{\theta}_1}\tilde{H}(\tilde{\theta}(t), \tilde{I}(t)) = -\varepsilon\partial_{\tilde{\theta}_1}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) - \varepsilon^2\partial_{\tilde{\theta}_1}f'(\tilde{\theta}(t), \tilde{I}(t)), \\ \frac{d}{dt}\tilde{I}_2(t) = -\partial_{\tilde{\theta}_2}\tilde{H}(\tilde{\theta}(t), \tilde{I}(t)) = -\varepsilon^2\partial_{\tilde{\theta}_2}f'(\tilde{\theta}(t), \tilde{I}(t)), \\ \frac{d}{dt}\tilde{\theta}_1(t) = \partial_{\tilde{I}_1}\tilde{H}(\tilde{\theta}(t), \tilde{I}(t)) = \omega_1(I(t)) + \varepsilon\partial_{\tilde{I}_1}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) + \varepsilon^2\partial_{\tilde{I}_1}f'(\tilde{\theta}(t), \tilde{I}(t)), \\ \frac{d}{dt}\tilde{\theta}_2(t) = \partial_{\tilde{I}_2}\tilde{H}(\tilde{\theta}(t), \tilde{I}(t)) = \omega_2(I(t)) + \varepsilon\partial_{\tilde{I}_2}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) + \varepsilon^2\partial_{\tilde{I}_2}f'(\tilde{\theta}(t), \tilde{I}(t)), \end{cases} \quad (12)$$

since $\bar{f}$ is independent of the second angular variables. For a positive constant δ to be chosen later in terms of λ and δ^*, we let $\tau = \delta\varepsilon^{-1}$. From the second equation of (12) and the first estimate of (10), we get

$$|\tilde{I}_2(t) - \tilde{I}_2(0)| = |\tilde{I}_2(t)| \leq C'\varepsilon\delta, \quad |t| \leq \tau,$$

which makes sense provided that $|\tilde{I}_2(t) - \tilde{I}_2(0)| \leq \kappa\varepsilon/2$ for $|t| \leq \tau$, and this is satisfied if $C'\delta \leq \kappa/2$, that is $\delta \leq \kappa(2C')^{-1}$. Now for $|t| \leq \tau$, recalling that $\omega_1(\tilde{I}_1(t), \tilde{I}_2(0)) = 0$ and $h \in C_1^4(B_R)$, we have

$$|\omega_1(\tilde{I}(t))| = |\omega_1(\tilde{I}_1(t), \tilde{I}_2(t))| = |\omega_1(\tilde{I}_1(t), \tilde{I}_2(t)) - \omega_1(\tilde{I}_1(t), \tilde{I}_2(0))| \leq |\tilde{I}_2(t) - \tilde{I}_2(0)| \leq C'\varepsilon\delta.$$

Therefore, from the third equation of (12), the second estimate of (10) and the fact that $\bar{f} \in C_1^7(\mathbb{T} \times B_R)$, we have

$$\left|\frac{d}{dt}\tilde{\theta}_1(t)\right| \leq C'\varepsilon\delta + \varepsilon + C'\varepsilon^2 = (C'\delta + 1 + C'\varepsilon)\varepsilon \leq C\varepsilon, \quad |t| \leq \tau$$

with $C = 1 + 2C'$, provided $\delta \leq 1$ and since $\varepsilon \leq 1$. This implies that

$$|\tilde{\theta}_1(t) - \theta_1^*| = |\tilde{\theta}_1(t) - \tilde{\theta}_1(0)| \leq C\delta, \quad |t| \leq \tau. \quad (13)$$

Moreover, recall that $|\tilde{I}_2(t) - \tilde{I}_2(0)| \leq C'\varepsilon\delta \leq C\delta$ for $|t| \leq \tau$ by the definition of C and since $\varepsilon \leq 1$, and from the first equation of (12), the first estimate of (10) and the fact that $\bar{f} \in C_1^7(\mathbb{T} \times B_R)$, we also have

$$|\tilde{I}_1(t) - I_1^*| = |\tilde{I}_1(t) - \tilde{I}_1(0)| \leq \delta + C'\varepsilon\delta, \quad |t| \leq \tau,$$

which makes sense if $\delta \leq \delta^*$ as this implies that $|\tilde{I}_1(t) - I_1^*| \leq \delta^* + \kappa\varepsilon/2$. In particular

$$|\tilde{I}_1(t) - I_1^*| = |\tilde{I}_1(t) - \tilde{I}_1(0)| \leq C\delta, \quad |t| \leq \tau,$$

by the definition of C and since $\varepsilon \leq 1$ and therefore

$$|\tilde{I}(t) - I^*| = |\tilde{I}(t) - \tilde{I}(0)| \leq C\delta, \quad |t| \leq \tau. \quad (14)$$

Using the fact that $\bar{f} \in C_1^7(\mathbb{T} \times B_R)$, from (13) and (14) we obtain

$$|\partial_{\tilde{\theta}_1}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) - \partial_{\tilde{\theta}_1}\bar{f}(\theta_1^*, I^*)| \leq C\delta, \quad |t| \leq \tau. \quad (15)$$

We eventually choose $\delta = \min\{\lambda(4C)^{-1}, \delta^*\}$, and hence $\tau = \delta\varepsilon^{-1} \leq \lambda(4C)^{-1}\varepsilon^{-1}$. We have to make sure that $\delta \leq 1$ and $\delta \leq \min\{\delta^*, \kappa(2C')^{-1}\}$. The first requirement is obviously satisfied since $C \geq 1$ and $\lambda \leq 1$, and hence $\delta \leq \lambda(4C)^{-1} \leq 1$. For the second one, which reduces to $\delta \leq \kappa(2C')^{-1}$, note that $\varpi \leq 1$ since $h \in C_1^4(B_R)$, so $\kappa \geq 1$ hence $\lambda \leq 1 \leq 2\kappa$ and this implies that $\delta \leq \kappa(2C')^{-1}$ as $C \geq C'$. Now from (11), (15) and the definition of δ, we have for all $|t| \leq \tau$,

$$|\varepsilon\partial_{\tilde{\theta}_1}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t))| \geq |\varepsilon\partial_{\tilde{\theta}_1}\bar{f}(\theta_1^*, I^*)| - |\varepsilon\partial_{\tilde{\theta}_1}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) - \varepsilon\partial_{\tilde{\theta}_1}\bar{f}(\theta_1^*, I^*)| \geq \varepsilon\lambda - C\varepsilon\delta \geq 3\varepsilon\lambda/4.$$

Moreover, if we assume that $\varepsilon \leq (4C')^{-1}\lambda$, then from the second estimate of (10), we have

$$|\varepsilon^2\partial_{\tilde{\theta}_1}f'(\tilde{\theta}(t), \tilde{I}(t))| \leq C'\varepsilon^2 \leq \varepsilon\lambda/4, \quad |t| \leq \tau,$$

and this gives, as before,

$$|\varepsilon\partial_{\tilde{\theta}_1}\bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) + \varepsilon^2\partial_{\tilde{\theta}_1}f'(\tilde{\theta}(t), \tilde{I}(t))| \geq 3\varepsilon\lambda/4 - \varepsilon\lambda/4 = \varepsilon\lambda/2, \quad |t| \leq \tau.$$

Now from the first equation of (12), we obtain

$$\left|\frac{d}{dt}\tilde{I}_1(t)\right| \geq \varepsilon\lambda/2, \quad |t| \leq \tau,$$

which eventually gives

$$|\tilde{I}_1(\tau) - \tilde{I}_1(0)| \geq \tau\varepsilon\lambda/2 \geq 2C\delta^2.$$

Coming back to the original Hamiltonian, $\Phi(\tilde{\theta}(t), \tilde{I}(t)) = (\theta(t), I(t))$ is a solution of the Hamiltonian H, and from the first estimate of (10), we have

$$|\tilde{I}_1(t) - I_1(t)| \leq \kappa\varepsilon/2, \quad |\tilde{I}_2(t) - I_2(t)| \leq \kappa\varepsilon/2$$

as long as $\tilde{I}(t) \in S^*(\kappa\varepsilon/2)$, so in particular

$$|\tilde{I}_1(0) - I_1(0)| \leq \kappa\varepsilon/2, \quad |\tilde{I}_1(\tau) - I_1(\tau)| \leq \kappa\varepsilon/2.$$

Assuming that $\varepsilon \leq C\delta^2/\kappa$, this gives

$$|I_1(\tau) - I_1(0)| \geq |\tilde{I}_1(\tau) - \tilde{I}_1(0)| - |\tilde{I}_1(\tau) - I_1(\tau)| - |\tilde{I}_1(0) - I_1(0)| \geq 2C\delta^2 - \kappa\varepsilon \geq C\delta^2.$$

Summing up, if we define
$$\varepsilon_0 = \min\{\varepsilon_0', \lambda(4C')^{-1}, C\delta^2\kappa^{-1}\}$$
and $c = \kappa$, then for $\varepsilon \leq \varepsilon_0$, $\delta = \min\{\delta^*, \lambda(4C)^{-1}\}$ and $\tau = \delta\varepsilon^{-1}$, the Hamiltonian H has a solution $(\theta(t), I(t))$ for which

$$|I_1(0) - I_1^*| \leq c\varepsilon, \quad |I_1(\tau) - I_1(0)| \geq \lambda^2(16C)^{-1} = C\delta^2.$$

Moreover, for all $t \in [0, \tau]$,

$$|I_2(t)| \leq c\varepsilon, \quad d(I_1(t), S_1^*) \leq c\varepsilon.$$

This was the statement to prove. $\qquad\qquad\square$

2.3.2 Proof of Theorem 2.2

The proof of Theorem 2.2 is based on the normal form Proposition 2.4, and is similar to the proof of Theorem 1.1 so we will not repeat several details. The only difference is that the perturbation f (and hence its average $\bar{f}$) is action independent so that the action dependence in the normal form is of order ε^2, and moreover the dependence on θ_2 in the normal form is of order ε^3. These facts will be used to control the solution on a longer time τ, and this will eventually give a larger drift of the action variable I_1.

Proof of Theorem 2.2. Recall that we are considering $H = h + \varepsilon f$ defined on $\mathbb{T}^2 \times B_R$, with $h \in C_1^{10}(B_R)$ satisfying $(B.1)$ and $(B.2)$ and $f \in \mathcal{G}^{19}$, so we can apply Proposition 2.3 with $l = 7$ and $l' = 1$: there exist positive constants ε_0'', C' and C'' depending only on R, the length of S_1^* and ϖ such that if $\varepsilon \leq \varepsilon_0'$, there exists a symplectic embedding

$$\Phi : \mathcal{D}^*(\kappa\varepsilon/4) \to \mathcal{D}^*(\kappa\varepsilon)$$

of class C^2 such that

$$H \circ \Phi = h + \varepsilon\bar{f} + \varepsilon^2\bar{f}' + \varepsilon^3 f'', \quad \bar{f}(\theta_1, I) = \int_{\mathbb{T}} f(\theta_1, \theta_2, I)d\theta_2, \quad \bar{f}'(\theta_1, I) = \int_{\mathbb{T}} f'(\theta_1, \theta_2, I)d\theta_2$$

and we have the following estimates

$$|\Phi - \mathrm{Id}|_0 \leq 3\kappa\varepsilon/4, \quad |f'|_7 \leq C', \quad |f''|_1 \leq C''.$$

Let $I' \in S^*$ and $I'' \in S^*$, without loss of generality we may assume that $I_1' \leq I_1''$ and recall that $\rho = |I' - I''| = I_1'' - I_1'$. As before, we consider a solution $(\tilde{\theta}(t), \tilde{I}(t))$ of the system defined by $\tilde{H} = H \circ \Phi$ with an initial condition $(\tilde{\theta}(0), \tilde{I}(0))$ such that $\tilde{I}(0) = I'$, $\tilde{\theta}_1(0) = \theta_1^*$ and $\tilde{\theta}_2(0) \in \mathbb{T}$ is arbitrary, and for a positive constant δ to be chosen later in terms of λ and ρ, we let $\tau' = \delta\varepsilon^{-1}$. Moreover, we define

$$\tilde{\tau} = \inf\{t \geq 0, \; | \; |\tilde{I}_1(t) - I_1'| \geq \rho\} \in \mathbb{R}^+ \cup \{+\infty\}$$

and we let $\tau = \min\{\tau', \tilde{\tau}\} \in \mathbb{R}^+$. Reversing time if necessary, we may assume that $I_1' \leq \tilde{I}_1(\tau)$. Note that in the normal form, only f'' depends on θ_2, so from the equations of motion of $\tilde{H}$ and the estimate on f'',

$$|\tilde{I}_2(t) - \tilde{I}_2(0)| = |\tilde{I}_2(t)| \leq C'\varepsilon^2\delta, \quad |t| \leq \tau,$$

which makes sense provided that $|\tilde{I}_2(t) - \tilde{I}_2(0)| \leq \kappa\varepsilon/4$ for $|t| \leq \tau$, and this is satisfied if $C'\delta\varepsilon \leq \kappa/4$, that is $\varepsilon \leq \kappa(4\delta C')^{-1}$. This implies, as before, that

$$|\omega_1(\tilde{I}(t))| \leq |\tilde{I}_2(t) - \tilde{I}_2(0)| \leq C'\varepsilon^2\delta$$

and hence, using the fact that $\bar{f}$ is action-independent,

$$\left|\frac{d}{dt}\tilde{\theta}_1(t)\right| \leq C''\varepsilon^2\delta + C'\varepsilon^2 + C''\varepsilon^3 = (C''\delta + C' + C''\varepsilon^2)\varepsilon^2 \leq C\varepsilon^2, \quad |t| \leq \tau$$

with $C = C''\delta + C' + C''$, since $\varepsilon \leq 1$. Hence

$$|\tilde{\theta}_1(t) - \theta_1^*| = |\tilde{\theta}_1(t) - \tilde{\theta}_1(0)| \leq C\varepsilon\delta, \quad |t| \leq \tau. \tag{16}$$

19

Since $\bar{f} \in C_1^{19}(\mathbb{T} \times B_R)$, we obtain from the last estimate that
$$|\partial_{\tilde{\theta}_1} \bar{f}(\tilde{\theta}_1(t)) - \partial_{\tilde{\theta}_1} \bar{f}(\theta_1^*)| \leq C\varepsilon\delta, \quad |t| \leq \tau. \tag{17}$$
Now recall that since $f \in \mathcal{G}^{19}$, we have
$$|\varepsilon\partial_{\tilde{\theta}_1} \bar{f}(\theta_1^*)| \geq \varepsilon\lambda$$
and assuming that $\varepsilon \leq \lambda(4C\delta)^{-1}$, we have $C\varepsilon\delta \leq \lambda/4$ so that the last estimate together with (17) implies that
$$|\varepsilon\partial_{\tilde{\theta}_1} \bar{f}(\tilde{\theta}_1(t))| \geq 3\varepsilon\lambda/4, \quad 0 \leq t \leq \tau.$$
Moreover, if we assume that $\varepsilon \leq \lambda(8C')^{-1}$ and $\varepsilon \leq \lambda^{1/2}(8C'')^{-1/2}$, then
$$|\varepsilon^2 \partial_{\tilde{\theta}_1} \bar{f}'(\tilde{\theta}(t), \tilde{I}(t))| \leq C'\varepsilon^2 \leq \varepsilon\lambda/8, \quad |t| \leq \tau,$$
and
$$|\varepsilon^3 \partial_{\tilde{\theta}_1} f''(\tilde{\theta}(t), \tilde{I}(t))| \leq C''\varepsilon^3 \leq \varepsilon\lambda/8, \quad |t| \leq \tau$$
which gives
$$|\varepsilon\partial_{\tilde{\theta}_1} \bar{f}(\tilde{\theta}_1(t)) + \varepsilon^2 \partial_{\tilde{\theta}_1} \bar{f}'(\tilde{\theta}(t), \tilde{I}(t)) + \varepsilon^3 \partial_{\tilde{\theta}_1} f''(\tilde{\theta}(t), \tilde{I}(t))| \geq 3\varepsilon\lambda/4 - \varepsilon\lambda/8 - \varepsilon\lambda/8 = \varepsilon\lambda/2, \quad |t| \leq \tau.$$
From the equations of motion of $\tilde{H}$, this implies
$$\left| \frac{d}{dt}\tilde{I}_1(t) \right| \geq \varepsilon\lambda/2, \quad |t| \leq \tau,$$
and hence
$$|\tilde{I}_1(\tau) - \tilde{I}_1(0)| = \tilde{I}_1(\tau) - \tilde{I}_1(0) \geq \tau\varepsilon\lambda/2.$$
We have imposed no restriction on the choice of δ, so we eventually choose $\delta = 2\rho\lambda^{-1}$. If $\tau = \tilde{\tau}$, then by definition of $\tilde{\tau}$ we have $\tilde{I}_1(\tau) - I' = \rho$, so $\tilde{I}(\tau) = I''$ and we have $\tau \leq \tau' = \delta\varepsilon^{-1}$. If $\tau = \tau' = \delta\varepsilon^{-1}$, then from the last estimate and the choice of δ we obtain
$$\tilde{I}_1(\tau) - \tilde{I}_1(0) \geq \delta\lambda/2 = \rho$$
which implies that $\tilde{I}_1(\tau) - \tilde{I}_1(0) = \rho$ and $\tau = \tilde{\tau}$, and hence $\tilde{I}(\tau) = I''$. Therefore, in any cases, $\tilde{\tau}$ is finite, and we have $\tilde{I}(\tau) = I''$ for $\tau \leq \delta\varepsilon^{-1}$. Coming back to the original Hamiltonian, $\Phi(\tilde{\theta}(t), \tilde{I}(t)) = (\theta(t), I(t))$ is a solution of the Hamiltonian H, and using the estimate on Φ, we have
$$|\tilde{I}_1(t) - I_1(t)| \leq 3\kappa\varepsilon/4, \quad |\tilde{I}_2(t) - I_2(t)| \leq 3\kappa\varepsilon/4$$
as long as $\tilde{I}(t) \in S^*(\kappa\varepsilon/2)$, so in particular
$$|\tilde{I}_1(0) - I_1(0)| \leq 3\kappa\varepsilon/4, \quad |\tilde{I}_1(\tau) - I_1(\tau)| \leq 3\kappa\varepsilon/4.$$
Summing up, if we define
$$\varepsilon_0 = \min\{\varepsilon_0'', \kappa(4C'\delta)^{-1}, \lambda(4C\delta)^{-1}, \lambda(8C')^{-1}, \lambda^{1/2}(8C'')^{-1/2}\}$$
and $c = \kappa$, then for $\varepsilon \leq \varepsilon_0$, $\delta = 2\rho\lambda^{-1}$ and $\tau \leq \delta\varepsilon^{-1}$, the Hamiltonian H has a solution $(\theta(t), I(t))$ for which
$$|I_1(0) - I_1'| \leq c\varepsilon, \quad |I_1(\tau) - I_1''| \leq c\varepsilon.$$
Moreover, for all $t \in [0, \tau]$,
$$|I_2(t)| \leq c\varepsilon, \quad d(I_1(t), S_1^*) \leq c\varepsilon.$$
This was the statement to prove. □

A   Technical estimates

Let S^* be a bounded domain in $\mathbb{R}^2$, and for $0 < \varepsilon < 1$ and a positive constant κ, consider the domains $S^*(\kappa\varepsilon) = \{I \in \mathbb{R}^2 \mid d(I, S^*) \leq \kappa\varepsilon\}$ and $\mathcal{D}(\kappa\varepsilon) = \mathbb{T}^2 \times S^*(\kappa\varepsilon)$.

Let us begin by recalling some elementary estimates. First if $f \in C^r(\mathcal{D}(\kappa\varepsilon))$ for $r \geq 2$, then for $j \in \mathbb{N}^4$, $|j| \leq r$, $\partial^l f \in C^{r-|j|}(\mathcal{D}(\kappa\varepsilon))$ and obviously

$$|\partial^l f|_{r-|j|} \leq |f|_r. \tag{18}$$

In particular, this implies that if $f \in C^r(\mathcal{D}(\kappa\varepsilon))$, then its Hamiltonian vector field X_f is of class C^{r-1} and

$$|X_f|_{r-1} \leq |f|_r.$$

Then, given two functions $f, g \in C^r(\mathcal{D}(\kappa\varepsilon))$, the product fg belongs to $C^r(\mathcal{D}(\kappa\varepsilon))$ and by the Leibniz formula

$$|fg|_r \leq c(r)|f|_r|g|_r. \tag{19}$$

for some constant depending only on r. By (18) and (19), the Poisson Bracket $\{f, g\}$ belongs to $C^{r-1}(\mathcal{D}(\kappa\varepsilon))$ and

$$|\{f, g\}|_{r-1} \leq c(r)|f|_r|g|_r. \tag{20}$$

for another constant $c(r)$ depending only on r.

Given any two vector-valued functions F and G of class C^r, defined on appropriate domains in $\mathbb{R}^m$ with values in $\mathbb{R}^m$, such that the composition $F \circ G$ makes sense, from Faa di Bruno's formula (see for instance [AR67]) it is easy to see that $F \circ G$ is of class C^r and

$$|F \circ G|_r \leq c(m, r)|F|_r|G|_r^r. \tag{21}$$

for a constant $c(m, r)$ depending only on m and r. Faa di Bruno's formula and classical results on the existence and regularity of solutions of differential equations can be used to prove the following lemma.

Lemma A.1. *Let $\chi \in C^{r+2}(\mathcal{D}(\kappa\varepsilon))$, and assume that*

$$|\chi|_{r+1} \leq \kappa\varepsilon/2, \quad \varepsilon \leq c$$

for some positive constant c. Then, for all $|t| \leq 1$, $X_\chi^t : \mathcal{D}(\kappa\varepsilon/2) \to \mathcal{D}(\kappa\varepsilon)$ is a well-defined symplectic embedding of class C^{r+1}, and we have the estimates

$$|X_\chi^t - \mathrm{Id}|_0 \leq \kappa\varepsilon/2, \quad |X_\chi^t|_r \leq C$$

for some constant C depending only on c, κ and the diameter of S^*.

Note that the constants C depend only on c, κ and on the diameter of $S^*(\kappa\varepsilon)$, but since $\varepsilon < 1$, the latter is bounded by $d + 2\kappa$ where d is the diameter of S^*.

The proof of the above lemma is a simple adaptation of Lemma 3.15 in [DH09], see also Lemma $A.1$ in [Bou13].

References

[AR67]   R. Abraham and J. Robbin, *Transversal mappings and flows*, Benjamin, New-York, 1967.

[Arn63a]   V.I. Arnol'd, *Proof of a theorem of A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations*, Russ. Math. Surv. **18** (1963), no. 5, 9–36.

[Arn63b]   _____, *Small denominators and problems of stability of motion in classical and celestial mechanics*, Russ. Math. Surv. **18** (1963), no. 6, 85–191.

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

[Arn94]   _____, *Mathematical problems in classical physics*, Trends and perspectives in applied mathematics, Appl. Math. Sci., vol. 100, Springer, New York, 1994, pp. 1–20.

[Ber08]   P. Bernard, *The dynamics of pseudographs in convex Hamiltonian systems*, Journal of the American Math. Society **21** (2008), no. 3, 615–669.

[Ber10]   _____, *Large normally hyperbolic cylinders in a priori stable Hamiltonian systems*, Ann. Henri Poincaré **11** (2010), no. 5, 929–942.

[BK87]   D. Bernstein and A. Katok, *Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with convex Hamiltonians*, Invent. math. **88** (1987), 225–241.

[BKZ11]   P. Bernard, V. Kaloshin, and K. Zhang, *Arnold diffusion in arbitrary degrees of freedom and crumpled 3-dimensional normally hyperbolic invariant cylinders*, Preprint (2011), arXiv:1112.2773.

[BN12]   A. Bounemoura and L. Niederman, *Generic Nekhoroshev theory without small divisors*, Ann. Inst. Fourier **62** (2012), no. 1, 277–324.

[Bou12]   A. Bounemoura, *Optimal stability and instability for near-linear hamiltonians*, Annales Henri Poincaré **13** (2012), no. 4, 857–868.

[Bou13]   _____, *Normal forms, stability and splitting of invariant manifolds II. Finitely differentiable Hamiltonians*, Regul. Chaotic Dyn. **18** (2013), no. 3, 261–276.

[Che92]   W. Chen, *Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with nondegenerate Hessian*, Twist mappings and their applications (R. McGehee and K.R. Meyer, eds.), vol. IMA Math. Appl. 44, Springer, 1992, pp. 87–94.

[Che13]   C.-Q. Cheng, *Arnold diffusion in nearly integrable hamiltonian systems*, Preprint (2013), arXiv:1207.4016v2.

[CY04]   C.-Q. Cheng and J. Yan, *Existence of diffusion orbits in a priori unstable Hamiltonian systems*, J. Differ. Geom. **67** (2004), no. 3, 457–517.

[CY09] _____, *Arnold diffusion in Hamiltonian systems: a priori unstable case*, J. Differ. Geom. **82** (2009), no. 2, 229–277.

[DdlLS06] A. Delshams, R. de la Llave, and T. M. Seara, *A geometric mechanism for diffusion in Hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model*, Mem. Amer. Math. Soc. **179** (2006), no. 844.

[DH09] A. Delshams and G. Huguet, *Geography of resonances and Arnold diffusion in a priori unstable Hamiltonian systems*, Nonlinearity **22** (2009), no. 8, 1997–2077.

[GR07] M. Gidea and C. Robinson, *Shadowing orbits for transition chains of invariant tori alternating with Birkhoff zones of instability*, Nonlinearity **20** (2007), no. 5, 1115–1143.

[GR09] _____, *Obstruction argument for transition chains of tori interspersed with gaps*, Discrete Contin. Dyn. Syst., Ser. S **2** (2009), no. 2, 393–416.

[Her92] M.R. Herman, *Dynamics connected with indefinite normal torsion*, Twist mappings and their applications (R. McGehee and K.R. Meyer, eds.), vol. IMA Math. Appl. 44, Springer, 1992, pp. 153–182.

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

[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.

[KZ12] V. Kaloshin and K. Zhang, *A strong form of Arnold diffusion for two and a half degrees of freedom*, preprint (2012), arXiv:1212.1150.

[KZZ10] V. Kaloshin, K. Zhang, and Y. Zheng, *Almost dense orbit on energy surface*, XVIth International Congress on Mathematical Physics, World Sci. Publ., Hackensack, NJ, 2010, pp. 314–322.

[Mar12a] J.-P. Marco, *Generic properties of classical systems on the torus $\mathbb{T}^2$*, unpublished preprint.

[Mar12b] _____, *Nets of hyperbolic annuli in generic nearly integrable systems on $\mathbb{A}^3$*, unpublished preprint.

[Mat04] J.N. Mather, *Arnold diffusion I : Announcement of results*, J. of Math. Sciences **124** (2004), no. 5, 5275–5289.

[Mat12] _____, *Arnold diffusion by variational methods*, Essays in mathematics and its applications, Springer, Heidelberg, 2012, pp. 271–285.

[Mos60] J. Moser, *On the elimination of the irrationality condition and Birkhoff's concept of complete stability*, Bol. Soc. Mat. Mexicana (2) **5** (1960), 167–175.

[Nek73] N.N. Nekhoroshev, *Stable lower estimates for smooth mappings and the gradients of smooth functions*, Math. USSR Sbornik **19** (1973), no. 3, 425–467.

[Nek77] ———, *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.

[Nie06] L. Niederman, *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.

[Tre04] D. Treschev, *Evolution of slow variables in a priori unstable Hamiltonian systems*, Nonlinearity **17** (2004), no. 5, 1803–1841.