



**HAL**  
open science

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

Abed Bounemoura, Vadim Kaloshin

► **To cite this version:**

Abed Bounemoura, Vadim Kaloshin. Generic diffusion for a class of non-convex Hamiltonians with two degrees of freedom. 2013. hal-00811640v1

**HAL Id: hal-00811640**

**<https://hal.science/hal-00811640v1>**

Preprint submitted on 10 Apr 2013 (v1), last revised 8 Nov 2013 (v3)

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

Abed Bounemoura <sup>\*</sup> and Vadim Kaloshin <sup>†</sup>

April 10, 2013

## 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  $\varepsilon$  is non-zero, in which case the system defined by  $H$  can be considered as an  $\varepsilon$ -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  $\varepsilon$  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  $\varepsilon$  for all  $\varepsilon$  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.

---

<sup>\*</sup>Institut des Hautes Études Scientifiques (abedbou@gmail.com)

<sup>†</sup>University of Maryland at College Park (vadim.kaloshin@gmail.com)

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  $\varepsilon$  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  $\varepsilon$  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  $\varepsilon$  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. 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

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

for some positive constant  $c_1, c_2, a$  and  $b$ , provided  $\varepsilon$  is small enough and the system 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  $\varepsilon$ . 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  $\varepsilon$ , 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 Kolmogorov, Arnold and Nekhoroshev restrict the possibility of diffusion, both in space and in time, at least provided the corresponding non-degeneracy assumptions

are met. 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], [DdILS06], [GR07], [Ber08], [DH09], [CY09], and [GR09].

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. Similar results were independently obtained by Cheng ([Che12]) and announced by Marco ([Mar12a], [Mar12b]). 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. 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. 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.

Yet for some non-convex non-steep integrable Hamiltonians, the construction of examples 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  $\varepsilon$ .

The example above is rather specific, as  $f$  does not depend on the action variables, but 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, §1.2, 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  $\varepsilon$ . The conditions defining this class of integrable Hamiltonians  $h$  is the existence of a line with rational slope such that  $h$  is constant along this line (which means that the gradient of the restriction of  $h$  to this line vanishes), but its gradient  $\nabla h$  is not identically zero along this line (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  $\varepsilon$ ) 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  $h$  to be  $C^4$  and  $f$  to be  $C^3$ . As for the proof, it is in fact rather simple. The only ingredient is a normal form, in the spirit of [BKZ11], which is valid on a domain in the action space whose size is independent of  $\varepsilon$ , even though, unlike [BKZ11] where the normal form is used for an other purpose (namely to construct a normally hyperbolic invariant cylinder which is then used to locate an unstable orbit), we need a slightly stronger statement in order to derive the existence of an unstable orbit directly from the normal form.

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  $\varepsilon$  for all  $\varepsilon$  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 result

Let us now state precisely the main result of the paper. Given  $R > 0$ , let  $B_R$  be the closed ball of  $\mathbb{R}^2$  of radius  $R$  with respect to the supremum norm  $|\cdot|$ , 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  $h : B_R \rightarrow \mathbb{R}$  of class  $C^4$ , which satisfy the following two conditions:

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

(A.2) There exists a point  $I^* \in L \cap B_{R^*}$ , for some  $0 \leq R^* < R$ , such that  $\nabla h(I^*) \neq 0$ .

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  $h|_L$  vanishes identically on  $L \cap B_R$ , hence the function  $h|_L$  has a set of critical points which contains  $L \cap B_R$  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 in the interior of  $L \cap B_R$ . 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 [Bou12b], functions which do satisfy (A.1) are functions which are not rationally steep.

Given a small parameter  $0 < \varepsilon < 1$ , our perturbation  $\varepsilon f$  will be a “generic” function  $\varepsilon f : \mathbb{T}^2 \times B_R \rightarrow \mathbb{R}$  which is “small” for the  $C^3$  topology. For an integer  $r \geq 2$ , let  $C^r(\mathbb{T}^2 \times B_R)$  the space of  $C^r$  function  $f : \mathbb{T}^2 \times B_R \rightarrow \mathbb{R}$ , which is Banach space with respect to the norm

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

where we have used the standard multi-index notation. We extend the definition of the  $C^r$ -norm for vector-valued functions  $F = (f_1, \dots, f_m) : \mathbb{T}^2 \times B_R \rightarrow \mathbb{R}^m$ , 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  $\varepsilon f$  will be such that  $f$  belongs to an open and dense subset  $\mathcal{F}_k^*$  of  $C_1^3(\mathbb{T}^2 \times B_R)$ , depending on the vector  $k$  defined in (A.1) and the point  $I_*$  defined in (A.2). For a given function  $f \in C_1^3(\mathbb{T}^2 \times B_R)$ , we define  $f_k^* \in C_1^3(\mathbb{T}^2)$  by

$$f_k^*(\theta) = \int_0^1 f(\theta + tk, I_*) dt,$$

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

$$\mathcal{F}_k^* = \{f \in C_1^3(\mathbb{T}^2 \times B_R) \mid \exists \theta^* \in \mathbb{T}^2, \partial_\theta f_k^*(\theta^*) \neq 0\}.$$

In words,  $\mathcal{F}_k^*$  is the subset of functions  $f \in C_1^3(\mathbb{T}^2 \times B_R)$  such that  $f_k^*$  is a non-constant function: this is obviously an open and dense subset of  $C_1^3(\mathbb{T}^2 \times B_R)$ . Note that  $f_k^*$  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  $\mathbb{T}$ .

We can finally state our 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  $f \in \mathcal{F}_k^*$ . Then there exists a positive constant  $C$ , depending only on  $h$ , and positive constants  $\varepsilon_0$  and  $\delta$  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*

$$|I(\tau) - I(0)| \geq C\delta^2, \quad \tau = \delta\varepsilon^{-1}.$$

It is a statement of diffusion for the action variables, in the sense that they have a variation which is bounded from below independently of  $\varepsilon$ , for all  $\varepsilon$  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

$$C\delta^2 \leq |I(\tau) - I_0| \leq \delta.$$

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 (C.1), the norm of  $\nabla h(I^*)$  and  $R^*$  that appeared in (C.2), while the dependence on  $f$  is only through the absolute value of  $\partial_\theta f_k^*(\theta^*)$ . We refer to Theorem 2.1 in §2.1 for a more concrete and precise statement.

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  $L$  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  $\omega$  is resonant, that is  $l \cdot \omega = 0$  for some  $l \in \mathbb{Z}^2 \setminus \{0\}$ , and  $\omega$  is non-zero. On the other hand, if  $\omega$  is non-resonant, it follows from [Bou12b] that the statement of Theorem 1.1 cannot be true for any sufficiently small perturbation, since for all sufficiently small perturbation, one has stability for an interval of time which is strictly larger than  $[-\tau, \tau]$  with  $\tau$  as above. In particular, if  $\omega$  is Diophantine, one has stability for an interval of time which is exponentially long with respect to  $\varepsilon^{-1}$ , up to an exponent depending only on the Diophantine exponent of  $\omega$ .

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  $h(I) = \alpha_1^2 I_1^2 - \alpha_2^2 I_2^2$ , (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  $\alpha_2/\alpha_1$  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  $\alpha_2/\alpha_1$  is a Diophantine number, the quadratic Hamiltonian falls into the class of Diophantine step 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  $\tau$  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  $\tau$ . This is indeed the case for linear or quadratic integrable Hamiltonians as we described above, and the general case is conjectured in [Bou12b].

## 2 Proof of Theorem 1.1

In §2.1, we will perform some preliminary transformations to reduce Theorem 1.1 to an equivalent but more concrete statement, which is Theorem 2.1. Theorem 2.1 will be proved in §2.3, based on a normal form result which is stated and proved in §2.2.

## 2.1 Preliminary reductions

Let us first give more concrete formulations of the conditions (A.1) and (A.2).

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  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  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  ${}^t M^{-1}$  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  $\Phi_M(\theta, I) = (M\theta, {}^t M^{-1}I)$  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  $\Phi_T$  and  $\Phi_M$  do change the domain  $B_R$  in the space of actions, but to simplify the notations, we will assume that the latter is fixed.

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

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

The condition (A.1) is that  $h$  is constant on  $L = \{(I_1, I_2) \in \mathbb{R}^2 \mid I_2 = 0\}$ , which is obviously equivalent to  $\partial_{I_1} h(I_1, 0) = \omega_1(I_1, 0) = 0$  for all  $I_1$  such that  $|I_1| \leq R$ . The condition (A.2) is that there exists  $I_1^*$  and  $0 \leq R^* < R$  with  $|I_1^*| \leq R^*$  such that  $\omega_2(I^*) = \omega_2(I_1^*, 0) = \omega^* \neq 0$ . Changing  $H$  to  $-H$  if necessary and reversing the time accordingly, we may assume that  $\omega^* > 0$ .

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

(B.1) For all  $I_1$  such that  $|I_1| \leq R$ , we have  $\omega_1(I_1, 0) = 0$ .

(B.2) There exists  $I_1^*$  and  $0 \leq R^* < R$  with  $|I_1^*| \leq R^*$  such that  $\omega_2(I^*) = \omega_2(I_1^*, 0) = \omega^* > 0$ .

Then the definition of  $\mathcal{F}_{e_2}^*$  also simplifies: one easily check that for  $f \in C_1^3(\mathbb{T}^2 \times B_R)$ , we have  $f_{e_2}^* \in C_1^3(\mathbb{T})$  where

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

so that  $f \in \mathcal{F}_{e_2}^*$  if and only if there exists  $\theta_1^* \in \mathbb{T}$  for which  $\partial_{\theta_1} f_{e_2}^*(\theta_1^*) \neq 0$ . For simplicity, we write  $f_{e_2}^* = f^*$  and  $\mathcal{F}_{e_2}^* = \mathcal{F}^*$ , and for  $f \in \mathcal{F}^*$ , we denote by  $\lambda$  a lower bound on the absolute value of  $\partial_{\theta_1} f^*(\theta_1^*)$ .

From the previous discussion, it follows that Theorem 1.1 is implied by the following statement.

**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  $f \in \mathcal{F}^*$ . Then there exists a positive constant  $C$ , depending only on  $R$ ,  $R^*$  and  $\omega^*$ , and a positive constant  $\varepsilon_0$  depending also on  $\lambda$ , such that for any  $0 < \varepsilon \leq \varepsilon_0$ , if we set  $\delta = \lambda(4C)^{-1}$ , the Hamiltonian system defined by  $H$  has a solution  $(\theta(t), I(t))$  such that*

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



## 2.2 A normal form

The main ingredient of the proof of Theorem 2.1 will be a normal form on a domain in the space of action whose size, in the direction given by the first action variables  $I_1$ , is independent of  $\varepsilon$ . Let us define

$$\rho = \min \{(R - R^*)/2, \omega^*/4\}$$

and the domain  $D_\rho^*$  by

$$D_\rho^* = \{(I_1, I_2) \in \mathbb{R}^2 \mid |I_1 - I_1^*| \leq \rho, I_2 = 0\}.$$

Let us furthermore define  $\kappa = 4/\omega^*$  and consider the  $\kappa\varepsilon$ -neighbourhood  $D_\rho^*(\kappa\varepsilon)$  of the domain  $D_\rho^*$ , defined by

$$D_\rho^*(\kappa\varepsilon) = \{I \in \mathbb{R}^2 \mid d(I, D_\rho^*) \leq \kappa\varepsilon\} = \{(I_1, I_2) \in \mathbb{R}^2 \mid |I_1 - I_1^*| \leq \rho + \kappa\varepsilon, |I_2| \leq \kappa\varepsilon\}$$

where the distance is the one induced by the supremum norm on  $\mathbb{R}^2$ . Eventually, we define  $\mathcal{D}_\rho^*(\kappa\varepsilon) = \mathbb{T}^2 \times D_\rho^*(\kappa\varepsilon)$ .

In the statement and proof of the proposition below, to avoid cumbersome notations, when convenient we will use a dot  $\cdot$  in replacement of any constant depending only on  $R$ ,  $R^*$  and  $\omega^*$ , that is for any two quantities  $u$  and  $v$ , an expression  $u \prec v$  means that there exists a constant  $c$  depending only on  $R$ ,  $R^*$  and  $\omega^*$  such that  $u \leq cv$ .

**Proposition 2.2.** *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  $f \in \mathcal{F}^*$ . Assume that  $\kappa\varepsilon \leq \rho$ . Then there exists a symplectic embedding  $\Phi : \mathcal{D}_\rho^*(\kappa\varepsilon/2) \rightarrow \mathcal{D}_\rho^*(\kappa\varepsilon)$  of class  $C^2$  such that*

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

and, if  $\Phi = (\Phi_\theta, \Phi_I)$ , we have the following estimates

$$|\Phi_I - \text{Id}|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \leq \kappa\varepsilon/2, \quad |\partial_\theta f'|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \prec \varepsilon, \quad |\partial_I f'|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \prec 1.$$

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

*Proof.* First of all, note that since  $\kappa\varepsilon \leq \rho$ , by definition of  $\rho$  the domain  $D_\rho^*(\kappa\varepsilon)$  is included in  $B_R$ . For a function  $\chi : \mathcal{D}_\rho^*(\kappa\varepsilon) \rightarrow \mathbb{R}$  of class  $C^3$  to be chosen below, the transformation  $\Phi$  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  $\mathcal{D}_\rho^*(\kappa\varepsilon/2)$  for  $|t| \leq 1$ , let  $\Phi = X_{\varepsilon\chi}^1$ . Using the relation

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

for an arbitrary function  $K$ , 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

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

where  $\bar{f}$  is the function defined in the statement. It would be natural to choose  $\chi$  to solve the equation  $\{h, \chi\} + g = 0$  where  $g = f - \bar{f}$ , which can be written as  $\{\chi, h\} = g$ , but to avoid the use of Fourier expansions and hence regularity issues, we will only solve this equation approximatively.

Let us define the projection  $\Pi : D_\rho^*(\kappa\varepsilon) \rightarrow D_\rho^*(\kappa\varepsilon) \cap \{I_2 = 0\} = D_{\rho+\kappa\varepsilon}^*$ , and for  $I \in D_\rho^*(\kappa\varepsilon)$ , we write  $\Pi(I) = \tilde{I}$ . For any  $I = (I_1, I_2) \in D_\rho^*(\kappa\varepsilon)$ , we obviously have

$$|\tilde{I} - I| \leq \kappa\varepsilon \quad (2)$$

and  $\omega(\tilde{I}) = (\omega_1(I_1, 0), \omega_2(I_1, 0)) = (0, \omega_2(I_1, 0))$  by (B.1). Moreover, for any  $I = (I_1, I_2) \in D_\rho^*(\kappa\varepsilon)$ ,  $|I_1 - I_1^*| \leq \rho + \kappa\varepsilon \leq 2\rho$  and as  $|h|_{C^1_1(B_R)} \leq 1$ , we have

$$|\omega_2(I_1, 0) - \omega_2(I_1^*, 0)| \leq 2\rho \leq \omega^*/2$$

by definition of  $\rho$ . Since  $\omega_2(I_1^*, 0) = \omega^* > 0$  by (B.2), it follows that  $\omega_2(I_1, 0) \geq \omega^*/2$  for any  $I = (I_1, I_2) \in D_\rho^*(\kappa\varepsilon)$ . Now observe that the equation  $\{\chi, h\} = g$  can be written again as

$$\omega(I) \cdot \partial_\theta \chi(\theta, I) = g(\theta, I), \quad (\theta, I) \in \mathcal{D}_\rho^*(\kappa\varepsilon). \quad (3)$$

Instead of solving this equation, we will solve the equation

$$\omega(\tilde{I}) \cdot \partial_\theta \chi(\theta, I) = g(\theta, I), \quad (\theta, I) \in \mathcal{D}_\rho^*(\kappa\varepsilon) \quad (4)$$

which can be written again as

$$\omega_2(I_1, 0) \partial_{\theta_2} \chi(\theta, I) = g(\theta, I), \quad (\theta, I) \in \mathcal{D}_\rho^*(\kappa\varepsilon) \quad (5)$$

since  $\omega(\tilde{I}) = (0, \omega_2(I_1, 0))$ . We claim that the equation (5) is solved by

$$\chi(\theta, I) = \frac{1}{\omega_2(I_1, 0)} \int_0^1 g(\theta + te_2, I) t dt. \quad (6)$$

where  $e_2 = (0, 1)$ . First recall that  $g = f - \bar{f}$  and therefore

$$\int_0^1 g(\theta + te_2, I) dt = \int_0^1 f(\theta + te_2, I) dt - \int_0^1 \bar{f}(\theta + te_2, I) dt = \bar{f}(\theta_1, I) - \bar{f}(\theta_1, I) = 0.$$

Then we compute

$$\begin{aligned}
\omega_2(I_1, 0) \partial_{\theta_2} \chi(\theta, I) &= \partial_{\theta_2} \left( \int_0^1 g(\theta + te_2, I) t dt \right) = \int_0^1 \partial_{\theta_2} g(\theta + te_2, I) t dt \\
&= \int_0^1 \left( \frac{d}{dt} g(\theta + te_2, I) \right) t dt = g(\theta + te_2, I) t \Big|_0^1 - \int_0^1 g(\theta + te_2, I) dt \\
&= g(\theta + e_2, I) = g(\theta, I)
\end{aligned}$$

where we have used the chain rule and an integration in the second line.

Now  $h \in C_1^4(B_R)$ ,  $f \in C_1^3(\mathbb{T}^2 \times B_R)$  and  $\omega_2(I_1, 0) \geq \omega^*/2$  for any  $I = (I_1, I_2) \in D_\rho^*(\kappa\varepsilon)$ : it follows from (6) and Leibniz formula (inequality (18) of Appendix A) that  $\chi$  is of class  $C^3$  and

$$|\chi|_{C^3(\mathcal{D}_\rho^*(\kappa\varepsilon))} \ll 1 \quad (7)$$

whereas

$$|\partial_\theta^j \chi|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon), \mathbb{R}^2)} \leq 2/\omega^* = \kappa/2 \quad (8)$$

for any  $j \in \mathbb{N}^2$ ,  $|j| \leq 2$ . So in particular, from (7) and (8), the function  $\varepsilon\chi \in C^3(\mathbb{T}^2 \times B_R)$  satisfies

$$|\varepsilon\chi|_{C^2(\mathcal{D}(\kappa\varepsilon))} \ll \varepsilon, \quad |\partial_\theta \varepsilon\chi|_{C^0(\mathcal{D}(\kappa\varepsilon), \mathbb{R}^2)} \leq \kappa\varepsilon/2,$$

hence we can apply Lemma A.1 of Appendix A: for all  $|t| \leq 1$ ,  $X_{\varepsilon\chi}^t : \mathcal{D}(\kappa\varepsilon/2) \rightarrow \mathcal{D}(\kappa\varepsilon)$  is a well-defined symplectic embedding of class  $C^2$ , and if we write  $X_{\varepsilon\chi}^t = (\Phi_\theta^t, \Phi_I^t)$ , we have the estimates

$$|\Phi_I^t - \text{Id}|_{C^0(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^2)} \leq \kappa\varepsilon/2, \quad |X_{\varepsilon\chi}^t - \text{Id}|_{C^1(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^4)} \ll \varepsilon, \quad |X_{\varepsilon\chi}^t|_{C^1(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^4)} \ll 1. \quad (9)$$

In particular, the first estimate gives

$$|\Phi_I - \text{Id}|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \leq \kappa\varepsilon/2.$$

Now let us define  $R_1$  by  $R_1(\theta, I) = (\omega(I) - \omega(\tilde{I})) \cdot \partial_\theta \chi(\theta, I)$  for  $(\theta, I) \in \mathcal{D}_\rho^*(\kappa\varepsilon)$  and

$$f' = R_1 + R_2, \quad R_2 = \varepsilon \int_0^1 \{(1-t)\{h, \chi\} + f, \chi\} \circ X_{\varepsilon\chi}^t$$

It follows from the equalities (1), (3) and (4) that

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

so it remains only to estimate the partial derivatives of  $f'$ . The estimates

$$|\partial_\theta R_1|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \ll \varepsilon, \quad |\partial_I R_1|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \ll 1$$

follow easily from (2), (7) and (8). Then we have

$$\begin{aligned} |R_2|_{C^1(\mathcal{D}_\rho^*(\kappa\varepsilon/2))} &\ll \varepsilon |\{(1-t)\{h, \chi\} + f, \chi\}|_{C^1(\mathcal{D}_\rho^*(\kappa\varepsilon))} |X_{\varepsilon\chi}^t|_{C^1(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^4)} \\ &\ll \varepsilon |\{\{h, \chi\} + f, \chi\}|_{C^1(\mathcal{D}_\rho^*(\kappa\varepsilon))} \\ &\ll \varepsilon |\{h, \chi\} + f|_{C^2(\mathcal{D}_\rho^*(\kappa\varepsilon))} |\chi|_{C^2(\mathcal{D}_\rho^*(\kappa\varepsilon))} \\ &\ll \varepsilon \left( |\{h, \chi\}|_{C^2(\mathcal{D}_\rho^*(\kappa\varepsilon))} + |f|_{C^2(\mathcal{D}_\rho^*(\kappa\varepsilon))} \right) \\ &\ll \varepsilon \left( |h|_{C^3(\mathcal{D}_\rho^*(\kappa\varepsilon))} |\chi|_{C^3(\mathcal{D}_\rho^*(\kappa\varepsilon))} + |f|_{C^2(\mathcal{D}_\rho^*(\kappa\varepsilon))} \right) \\ &\ll \varepsilon \end{aligned}$$

where we have used the last part of (9), the fact that  $h \in C_1^4(B_R)$  and  $f \in C_1^3(\mathbb{T}^2 \times B_R)$ , the estimate (7) and the inequality (19) of Appendix A several times. This implies that

$$|\partial_\theta R_2|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \ll \varepsilon, \quad |\partial_I R_2|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \ll \varepsilon$$

and since  $f' = R_1 + R_2$ , we eventually obtain

$$|\partial_\theta f'|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \ll \varepsilon, \quad |\partial_I f'|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \ll 1$$

which concludes the proof.  $\square$

### 2.3 Proof of Theorem 2.1

The proof of Theorem 2.1 is now a consequence of our normal form Proposition 2.2. Since the latter is defined on a domain which is independent of  $\varepsilon$  in the  $I_1$ -direction, 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  $\Phi$  is 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  $f \in \mathcal{F}^*$ , so we can apply Proposition 2.2: assuming  $\kappa\varepsilon \leq \rho$ , there exist positive constants  $C_1$  and  $C_2$  depending only on  $R$ ,  $R^*$  and  $\omega^*$  and a symplectic embedding  $\Phi : \mathcal{D}_\rho^*(\kappa\varepsilon/2) \rightarrow \mathcal{D}_\rho^*(\kappa\varepsilon)$  of class  $C^2$  such that

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

Moreover, if  $\Phi = (\Phi_\theta, \Phi_I)$ , we have the following estimates

$$|\Phi_I - \text{Id}|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \leq \kappa\varepsilon/2, \quad (10)$$

and

$$|\partial_\theta f'|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \leq C_1\varepsilon, \quad |\partial_I f'|_{C^0(\mathcal{D}_\rho^*(\kappa\varepsilon/2), \mathbb{R}^2)} \leq C_2 \quad (11)$$

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

Let us consider the Hamiltonian  $\tilde{H} = H \circ \Phi$  defined on  $\mathcal{D}_\rho^*(\kappa\varepsilon/2)$ , and we shall write  $\Phi(\tilde{\theta}, \tilde{I}) = (\theta, I)$ . Let  $I^* = (I_1^*, 0)$  be given by (B.2), and  $\theta_1^*$  such that

$$|\partial_{\tilde{\theta}_1} f^*(\theta_1^*)| = |\partial_{\tilde{\theta}_1} \bar{f}(\theta_1^*, I^*)| \geq \lambda. \quad (12)$$

Note that necessarily  $\lambda \leq 1$  since  $f \in C_1^3(\mathbb{T}^2 \times B_R)$ . We 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  $\tilde{\theta}_2(0) \in \mathbb{T}$  arbitrary: we have the equations

$$\begin{cases} \frac{d}{dt} \tilde{I}_1(t) = -\varepsilon \partial_{\tilde{\theta}_1} \bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) - \varepsilon \partial_{\tilde{\theta}_1} f'(\tilde{\theta}(t), \tilde{I}(t)), \\ \frac{d}{dt} \tilde{I}_2(t) = -\varepsilon \partial_{\tilde{\theta}_2} f'(\tilde{\theta}(t), \tilde{I}(t)), \\ \frac{d}{dt} \tilde{\theta}_1(t) = \omega_1(I(t)) + \varepsilon \partial_{\tilde{I}_1} \bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) + \varepsilon \partial_{\tilde{I}_1} f'(\tilde{\theta}(t), \tilde{I}(t)), \\ \frac{d}{dt} \tilde{\theta}_2(t) = \omega_2(I(t)) + \varepsilon \partial_{\tilde{I}_2} \bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) + \varepsilon \partial_{\tilde{I}_2} f'(\tilde{\theta}(t), \tilde{I}(t)), \end{cases} \quad (13)$$

since  $\bar{f}$  is independent of the second angular variables. For a positive constant  $\delta$  to be chosen later in terms of  $\lambda$ , we let  $\tau = \delta\varepsilon^{-1}$ . From the second equation of (13) and the first estimate of (11), we get

$$|\tilde{I}_2(t) - \tilde{I}_2(0)| = |\tilde{I}_2(t)| \leq C_1\varepsilon\delta, \quad |t| \leq \tau,$$

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  $C_1\delta \leq \kappa/2$ , that is  $\delta \leq \kappa(2C_1)^{-1}$ . 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

$$|\omega_1(\tilde{I}(t))| = |\omega_1(\tilde{I}_1(t), \tilde{I}_2(t))| = |\omega_1(\tilde{I}_1(t), \tilde{I}_2(t)) - \omega_1(\tilde{I}_1(t), \tilde{I}_2(0))| \leq |\tilde{I}_2(t) - \tilde{I}_2(0)| \leq C_1\varepsilon\delta.$$

Therefore, from the third equation of (13), the second estimate of (11) and the fact that  $\bar{f} \in C_1^3(\mathbb{T} \times B_R)$ , we have

$$\left| \frac{d}{dt} \tilde{\theta}_1(t) \right| \leq C_1 \varepsilon \delta + \varepsilon + C_2 \varepsilon = (C_1 \delta + 1 + C_2) \varepsilon \leq C \varepsilon, \quad |t| \leq \tau$$

with  $C = C_1 + 1 + C_2$ , provided  $\delta \leq 1$ . This implies that

$$|\tilde{\theta}_1(t) - \theta_1^*| = |\tilde{\theta}_1(t) - \tilde{\theta}_1(0)| \leq C \delta, \quad |t| \leq \tau. \quad (14)$$

Moreover, recall that  $|\tilde{I}_2(t) - \tilde{I}_2(0)| \leq C_1 \varepsilon \delta \leq C \delta$  for  $|t| \leq \tau$  by the definition of  $C$  and since  $\varepsilon \leq 1$ , and from the first equation of (13), the first estimate of (11) and the fact that  $\bar{f} \in C_1^3(\mathbb{T} \times B_R)$ , we also have

$$|\tilde{I}_1(t) - I_1^*| = |\tilde{I}_1(t) - \tilde{I}_1(0)| \leq \delta + C_1 \varepsilon \delta, \quad |t| \leq \tau,$$

which makes sense if  $\delta \leq \kappa(2C_1)^{-1}$  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

$$|\tilde{I}(t) - I^*| = |\tilde{I}(t) - \tilde{I}(0)| \leq C \delta, \quad |t| \leq \tau. \quad (15)$$

Using the fact that  $\bar{f} \in C_1^3(\mathbb{T} \times B_R)$ , from (14) and (15) we obtain

$$|\partial_{\tilde{\theta}_1} \bar{f}(\tilde{\theta}_1(t), \tilde{I}(t)) - \partial_{\tilde{\theta}_1} \bar{f}(\theta_1^*, I^*)| \leq C \delta, \quad |t| \leq \tau. \quad (16)$$

We eventually choose  $\delta = \lambda(4C)^{-1}$ , and hence  $\tau = \lambda(4C)^{-1} \varepsilon^{-1}$ . We have to make sure that  $\delta \leq 1$  and  $\delta \leq \kappa(2C_1)^{-1}$ . The first requirement is obviously satisfied since  $C \geq 1$  and  $\lambda \leq 1$ . For the second, note that  $\omega^* \leq 1$  since  $h \in C_1^4(B_R)$ , so  $\kappa \geq 1$  hence  $\lambda \leq 1 \leq 2\kappa$  and this implies that  $\delta \leq \kappa(2C_1)^{-1}$  as  $C \geq C_1$ . Now from (12), (16) and the definition of  $\delta$ , we have for all  $|t| \leq \tau$ ,

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

Moreover, if we assume that  $\varepsilon \leq (4C_1)^{-1} \lambda$ , then from the first estimate of (11), we have

$$|\partial_{\tilde{\theta}_1} f'(\tilde{\theta}(t), \tilde{I}(t))| \leq C_1 \varepsilon \leq \lambda/4, \quad |t| \leq \tau,$$

and this gives, as before,

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

Now from the first equation of (13), we obtain

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

which eventually gives

$$|\tilde{I}_1(\tau) - \tilde{I}_1(0)| \geq \tau \varepsilon \lambda / 2 = \lambda^2 (8C)^{-1}.$$

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 from the estimate (10), we have

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

as long as  $\tilde{I}(t) \in D_\rho^*(\kappa\varepsilon/2)$ , 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  $\varepsilon \leq \lambda^2(16C\kappa)^{-1}$ , this gives

$$|I_1(\tau) - I_1(0)| \geq |\tilde{I}_1(\tau) - I_1(t)| - |\tilde{I}_1(\tau) - I_1(\tau)| - |\tilde{I}_1(0) - I_1(0)| \geq \lambda^2(8C)^{-1} - \kappa\varepsilon \geq \lambda^2(16C)^{-1}.$$

Summing up, if we define

$$\varepsilon_0 = \min\{\rho\kappa^{-1}, \lambda(4C_1)^{-1}, \lambda^2(16C\kappa)^{-1}\}$$

then for  $\varepsilon \leq \varepsilon_0$ , if  $\delta = \lambda(4C)^{-1}$  and  $\tau = \delta\varepsilon^{-1}$ , the Hamiltonian  $H$  has a solution  $(\theta(t), I(t))$  for which

$$|I_1(\tau) - I_1(0)| \geq \lambda^2(16C)^{-1} = C\delta^2.$$

This was the statement to prove.  $\square$

## A Technical estimates

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

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

$$|\partial^j f|_{C^{r-|j|}(\mathcal{D}(\kappa\varepsilon))} \leq |f|_{C^r(\mathcal{D}(\kappa\varepsilon))}. \quad (17)$$

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

$$|X_f|_{C^{r-1}(\mathcal{D}(\kappa\varepsilon), \mathbb{R}^4)} \leq |f|_{C^r(\mathcal{D}(\kappa\varepsilon))}.$$

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

$$|fg|_{C^r(\mathcal{D}(\kappa\varepsilon))} \leq c(r)|f|_{C^r(\mathcal{D}(\kappa\varepsilon))}|g|_{C^r(\mathcal{D}(\kappa\varepsilon))}. \quad (18)$$

for some constant depending only on  $r$ . By (17) and (18), the Poisson Bracket  $\{f, g\}$  belongs to  $C^{r-1}(\mathcal{D}(\kappa\varepsilon))$  and

$$|\{f, g\}|_{C^{r-1}(\mathcal{D}(\kappa\varepsilon))} \leq c(r)|f|_{C^r(\mathcal{D}(\kappa\varepsilon))}|g|_{C^r(\mathcal{D}(\kappa\varepsilon))}. \quad (19)$$

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

We shall also need the following lemma, which follows easily from Faa di Bruno's formula (see for instance [AR67]) and classical results on the existence and regularity of solutions of differential equations.

**Lemma A.1.** *Let  $f \in C^3(\mathcal{D}(\kappa\varepsilon))$ , and assume that*

$$|f|_{C^2(\mathcal{D}(\varepsilon))} \leq C\varepsilon, \quad |\partial_\theta f|_{C^0(\mathcal{D}(\varepsilon), \mathbb{R}^2)} \leq \kappa\varepsilon/2,$$

*for some positive constant  $C$ . Then, for all  $|t| \leq 1$ ,  $X_f^t : \mathcal{D}(\kappa\varepsilon/2) \rightarrow \mathcal{D}(\kappa\varepsilon)$  is a well-defined symplectic embedding of class  $C^2$ , and if we write  $X_f^t = (\Phi_\theta^t, \Phi_I^t)$ , we have the estimates*

$$|\Phi_I^t - \text{Id}|_{C^0(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^2)} \leq \kappa\varepsilon/2, \quad |X_f^t - \text{Id}|_{C^1(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^4)} \leq C_1\varepsilon, \quad |X_f^t|_{C^1(\mathcal{D}(\kappa\varepsilon/2), \mathbb{R}^4)} \leq C_2$$

*for some constant  $C_1, C_2$  depending only on  $C, \rho$  and  $\kappa$ .*

Note that the constants  $C_1$  and  $C_2$  depend only on  $C$  and on the diameter of  $D(\kappa\varepsilon)$ , but since  $\varepsilon < 1$ , the latter is bounded by  $2(\rho + \kappa)$  where  $\rho$  is the diameter of  $D$ .

The proof of the above lemma is a simple adaptation of Lemma 3.15 in [DH09], see also Lemma A.1 in [Bou12a].

## 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.
- [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.
- [Bou12a] A. Bounemoura, *Normal forms, stability and splitting of invariant manifolds II. Finitely differentiable Hamiltonians*, preprint.
- [Bou12b] ———, *Optimal stability and instability for near-linear hamiltonians*, Annales Henri Poincaré **13** (2012), no. 4, 857–868.

- [Che12] C.-Q. Cheng, *Arnold diffusion in nearly integrable hamiltonian systems*, Preprint (2012), arXiv:1207.4016.
- [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.
- [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.
- [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*, Matu. 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.