

**Zeitschrift:** Commentarii Mathematici Helvetici  
**Herausgeber:** Schweizerische Mathematische Gesellschaft  
**Band:** 91 (2016)  
**Heft:** 4

**Artikel:** Nekhoroshev's estimates for quasi-periodic time-dependent perturbations  
**Autor:** Bounemoura, Abed  
**DOI:** <https://doi.org/10.5169/seals-630573>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 29.03.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

## Nekhoroshev's estimates for quasi-periodic time-dependent perturbations

Abed Bounemoura

**Abstract.** In this paper, we consider a Diophantine quasi-periodic time-dependent analytic perturbation of a convex integrable Hamiltonian system, and we prove a result of stability of the action variables for an exponentially long interval of time. This extends known results for periodic time-dependent perturbations, and partly solves a long standing conjecture of Chirikov and Lochak. We also obtain improved stability estimates close to resonances or far away from resonances, and a more general result without any Diophantine condition.

**Mathematics Subject Classification (2010).** 37J25, 37J40.

**Keywords.** Hamiltonian systems, perturbation theory, effective stability.

### 1. Introduction and results

**1.1. Introduction.** Let  $n \geq 1$  be an integer,  $D \subset \mathbb{R}^n$  an open bounded convex domain and  $\mathbb{T}^n := \mathbb{R}^n / (2\pi\mathbb{Z})^n$ . Consider a smooth Hamiltonian function  $H$  defined on the domain  $\mathbb{T}^n \times D$  of the form

$$H(\theta, I) = h(I) + \varepsilon f(\theta, I), \quad \varepsilon \geq 0, \quad (\theta, I) = (\theta_1, \dots, \theta_n, I_1, \dots, I_n) \in \mathbb{T}^n \times D, \quad (1.1)$$

and its associated Hamiltonian system

$$\begin{cases} \dot{\theta}_i(t) = \partial_{I_i} H(\theta(t), I(t)) = \partial_{I_i} h(I(t)) + \varepsilon \partial_{I_i} f(\theta(t), I(t)), \\ \dot{I}_i(t) = -\partial_{\theta_i} H(\theta(t), I(t)) = -\varepsilon \partial_{\theta_i} f(\theta(t), I(t)) \end{cases} \quad 1 \leq i \leq n.$$

For  $\varepsilon = 0$ , the system is stable in the sense that the action variables  $I(t)$  of all solutions are constant, and all solutions are quasi-periodic. Now for  $\varepsilon \neq 0$  but sufficiently small, a fundamental result of Nekhoroshev states that if the system is real-analytic and the integrable part  $h$  satisfies a steepness or S-steepness condition (these are generic conditions), then the action variables  $I(t)$  of all solutions are almost constant for an interval of time which is exponentially long with respect to the inverse of the

perturbation. More precisely, the following estimates hold true along all solutions:

$$\|I(t) - I(0)\| := \sqrt{\sum_{i=1}^n (I_i(t) - I_i(0))^2} \leq R_0 \varepsilon^b, \quad |t| \leq T_0 \exp\left(c \left(\frac{1}{\varepsilon}\right)^a\right) \quad (1.2)$$

for some positive constants  $R_0, T_0, c, a$  and  $b$ . We refer to [25] and [26] for the precise results and the definitions of steep and S-steep functions. The most important constants appearing in the estimates (1.2) are undoubtedly the constants  $a$  and  $b$ , which are called the stability exponents. The simplest class of steep (respectively S-steep) integrable Hamiltonians are quasi-convex (respectively convex) Hamiltonians, and these exponents depend then only on the number of degrees of freedom  $n$ . Nekhoroshev's original proof yielded the following dependence

$$a \sim b \sim \frac{1}{n^2}.$$

Such values for the exponents were however much worse than the values

$$a = b = \frac{1}{2n}$$

conjectured by Chirikov [12] on a basis of a heuristic argument and numerical simulations. This issue was later solved by Lochak [20]: more precisely, Lochak-Neishtadt [23] and independently Pöschel [27] proved that (1.2) holds true with the values

$$a = b = \frac{1}{2n}$$

in the quasi-convex case (in the convex case, these exponents are valid for solutions starting not too close to the minimum of  $h$ ; for these latter solutions the action variables are stable for all time). The proof in [27] also allows to increase the value of  $b$  at the expense of decreasing the value of  $a$ : given any  $0 \leq \mu \leq 1$ , one can obtain the exponents

$$a = \frac{\mu}{2n}, \quad b = \frac{\mu}{2n} + \frac{1-\mu}{2}$$

that recovers the latter result by setting  $\mu = 1$ . On the other hand, it is proved in [10] that one can slightly improve the value of  $a$  at the expense of greatly decreasing the value of  $b$ : for any  $0 \leq \delta \leq (2n)^{-1}$ , one can choose

$$a = \frac{1}{2(n-1)}(1-2\delta), \quad b = \delta,$$

and the classical result is recovered by letting  $\delta = (2n)^{-1}$ . Examples of Arnold diffusion [2,3,30] show that in any event  $a < (2(n-2))^{-1}$  so that in the convex case, the estimate (1.2) is, as far as the dependence on  $a$  is concerned, quite sharp.

In the general steep case, no improvement on the values of the stability exponents were made until a recent achievement by Guzzo, Chierchia and Benettin (see [17] for an announcement and [18] for the proof): denoting by  $\alpha_1, \dots, \alpha_{n-1}$  the steepness exponents (these are integers always larger than one), they obtained the values

$$a = \frac{1}{2n(\alpha_1 \cdots \alpha_{n-2})}, \quad b = \frac{a}{\alpha_{n-1}} = \frac{1}{2n(\alpha_1 \cdots \alpha_{n-1})}.$$

For a quasi-convex Hamiltonian, one simply has  $\alpha_1 = \cdots = \alpha_{n-1} = 1$  and thus the latter result extends the conjectured optimal values.

Now instead of an autonomous perturbation as in (1.1), one may consider a periodic time-dependent perturbation, that is one looks at the Hamiltonian

$$H(\theta, I) = h(I) + \varepsilon f(\theta, I, t), \quad \varepsilon \geq 0, \quad (\theta, I) \in \mathbb{T}^n \times D, \quad t \in \mathbb{T}. \quad (1.3)$$

Setting  $t = \varphi \in \mathbb{T}$  and introducing a variable  $J \in \mathbb{R}$  canonically conjugated to  $\varphi$ , it is equivalent to study the autonomous Hamiltonian

$$H(\theta, I, \varphi, J) = h(I) + J + \varepsilon f(\theta, I, \varphi), \quad \varepsilon \geq 0, \\ (\theta, I) \in \mathbb{T}^n \times B, \quad (\varphi, J) \in \mathbb{T} \times \mathbb{R}. \quad (1.4)$$

Indeed,  $(\theta(t), I(t), \varphi(t), J(t))$  is a solution of the system associated to (1.4) if, and only if,  $(\theta(t), I(t))$  is a solution of the system associated to (1.3). In his seminal work [25, 26], Nekhoroshev introduced a notion of  $P$ -steepness such that exponential stability holds true for a periodic time-dependent perturbation of a  $P$ -steep integrable Hamiltonian. As Nekhoroshev pointed out, the  $P$ -steepness of an integrable Hamiltonian  $I \mapsto h(I)$  turns out to be equivalent to the steepness of the extended integrable Hamiltonian  $(I, J) \mapsto h(I) + J$ , hence the stability of  $P$ -steep integrable Hamiltonians under a periodic time-dependent perturbation is a direct consequence of the stability of steep integrable Hamiltonians under an autonomous perturbation.

Once again, the simplest class of  $P$ -steep Hamiltonians are given by convex Hamiltonians: if  $I \mapsto h(I)$  is convex, then one easily check that  $(I, J) \mapsto h(I) + J$  is quasi-convex and thus steep. Thus the results of [23] and [27] apply to periodic time-dependent perturbation of convex integrable Hamiltonians, and the estimates (1.2) hold true for the Hamiltonian (1.3) with the exponents

$$a = b = \frac{1}{2(n+1)}.$$

Now a periodic time-dependent perturbation is nothing but a special case of a quasi-periodic time-dependent perturbation, and we may more generally consider a Hamiltonian of the form

$$H(\theta, I) = h(I) + \varepsilon f(\theta, I, t\alpha), \quad \varepsilon \geq 0, \\ (\theta, I) \in \mathbb{T}^n \times D, \quad t\alpha = t(\alpha_1, \dots, \alpha_m) \in \mathbb{T}^m \quad (1.5)$$



where  $\alpha \in \mathbb{R}^m$  is a vector which is assumed to be non-resonant, that is  $k \cdot \alpha \neq 0$  for any non-zero  $k \in \mathbb{Z}^m$  and where  $\cdot$  denotes the Euclidean scalar product. We will assume actually that  $\alpha$  satisfies a Diophantine condition: there exist  $\gamma > 0$  and  $\tau \geq m - 1$  such that

$$|k \cdot \alpha| \geq \gamma |k|^{-\tau}, \quad k = (k_1, \dots, k_m) \in \mathbb{Z}^m \setminus \{0\},$$

$$|k| := |k_1| + \dots + |k_m|. \quad (\text{Dio}_{\gamma, \tau})$$

As before, setting  $\varphi = t\alpha \in \mathbb{T}^m$  and introducing a vector  $J = (J_1, \dots, J_m) \in \mathbb{R}^m$  canonically conjugated to  $\varphi$ , the Hamiltonian (1.5) is equivalent to

$$H(\theta, I, \varphi, J) = h(I) + \alpha \cdot J + \varepsilon f(\theta, I, \varphi), \quad \varepsilon \geq 0,$$

$$(\theta, I) \in \mathbb{T}^n \times D, \quad (\varphi, J) \in \mathbb{T}^m \times \mathbb{R}^m. \quad (1.6)$$

The Hamiltonian (1.6), in the special case  $m = 1$ , reduces to Hamiltonian (1.4): indeed,  $\alpha \in \mathbb{R}$  and by a scaling one may assume that  $\alpha = 1$ , and moreover  $(\text{Dio}_{\gamma, \tau})$  is obviously satisfied for  $\gamma = |\alpha| = 1$  and  $\tau = m - 1 = 0$ . However, in the case  $m \geq 2$ , the extended integrable Hamiltonian does not satisfy any steepness condition, so it is not clear whether the estimates (1.2) hold true.

We can now state the most general form of a conjecture of Chirikov (see [12, 13] and [14]), stated in a more precise manner by Lochak [22].

**Conjecture 1.1.** *The estimates (1.2) hold true for the Hamiltonian (1.6), provided it is real-analytic and  $h$  convex, with the exponents*

$$a = b = \frac{1}{2(n+1+\tau)}.$$

Note that this conjecture is made plausible by the fact that in the periodic case, that is  $m = 1$  and  $\tau = 0$ , it is a theorem. Yet for  $m \geq 2$  and hence  $\tau \geq 1$ , it is an open question whether the estimates (1.2) hold true for the Hamiltonian (1.6), for some values of  $a$  and  $b$  whatsoever.

It is the purpose of this article to solve this problem: we will prove that the estimates (1.2) hold true for the Hamiltonian (1.6), provided it is real-analytic and  $h$  convex, with the exponents

$$a = \frac{1}{2(n+1)(1+\tau)}, \quad b = \frac{(n+1)\tau + 1}{2(n+1)(1+\tau)} = \frac{1}{2} - \frac{n}{2(n+1)(1+\tau)}. \quad (1.7)$$

We refer to Theorem 1.2 below for a more precise statement. Concerning the values of the exponents we obtain, let us just make two comments (a more detailed discussion is contained in Section 1.4). First, for the periodic case ( $m = 1$ ,  $\tau = 0$ ), we also recover the known values of the exponents, so our result can also be considered as a “correct” generalization of the periodic case. Then, in the non-periodic case ( $m \geq 2$ ,  $\tau \geq 1$ ), our exponent  $a$  is substantially worse than the one of Conjecture 1.1, but

at the same time our exponent  $b$  is always much better:  $b$  is always strictly bigger than  $1/4$ , and when  $m$  gets large, it is close to  $1/2$ . Moreover, in the non-periodic case,  $b$  is essentially independent of  $n$  which is an interesting feature. In any case, it is therefore still an open question whether the values of the exponents of Conjecture 1.1 can be reached or not (once, again, we refer to Section 1.4).

In fact, we do obtain more general results. First, the perturbation will be allowed to depend also on the  $J$  variables, that is we can replace  $f(\theta, I, \varphi)$  in (1.6) by  $f(\theta, I, \varphi, J)$ , provided  $f$  is bounded and real-analytic in  $J$ , when  $J$  varies in  $\mathbb{R}^m$ . Moreover, we will also be able to control the evolution of the  $J$  variables; in the periodic case  $\tau = 0$  we will obtain a stability result with the same exponents  $a$  and  $b$ , while in the quasi-periodic case  $\tau \geq 1$ , we will obtain the same exponent  $a$  but a worse confinement given by the exponent

$$b^* = b - na = \frac{1}{2} - \frac{2n}{2(n+1)(1+\tau)}$$

which is still strictly positive, and asymptotically close to  $1/2$  when  $m$  gets large. This will be the precise content of Theorem 1.2. To understand the interest of this seemingly mild extension, one can compare such a result with the preservation of invariant tori (that is, the KAM theory) for Hamiltonians as in (1.1), (1.4) or (1.6). In the autonomous case of (1.1), if the integrable Hamiltonian is convex then it is in particular Kolmogorov non-degenerate and the classical KAM theory applies (see [28] for a survey). In the periodic case (1.4), the integrable Hamiltonian is no longer Kolmogorov non-degenerate but as it is quasi-convex, it is Arnold (or iso-energetically) non-degenerate: it follows that tori are preserved at a fixed energy for (1.4) yielding invariant tori for (1.3). In this case the perturbation may also depend on  $J \in \mathbb{R}$  without affecting the result. Now in the more general quasi-periodic case (1.6), the integrable part is both Kolmogorov and Arnold degenerate. However, using the non-degeneracy with respect to the  $I$  variables and the fact that the perturbation is independent of  $J$ , it is not hard to prove, using classical KAM techniques, that many tori with prescribed Diophantine frequencies of the form  $(\omega, \alpha) \in \mathbb{R}^{n+m}$  are preserved (up to our knowledge, this was first observed by Galavotti in [15] in a restricted situation and later by Lochak in [21, 22] in a general situation). It is crucial here to have a perturbation which is independent of  $J$ : if not, the method simply breaks down and it is rather easy to construct counter-examples (as in [29]) to the preservation of (full dimensional) invariant tori. Therefore unlike the situation in KAM theory, our result is exactly the same when  $f$  is allowed to depend on the  $J$  variables.

Then, exactly as in [27], one can obtain a more general result by increasing the value of  $b$  (and therefore of  $b^*$ ) while decreasing the value of  $a$ . This will be the content of Theorem 1'.

Moreover, as in the autonomous or periodic case, we do obtain enhanced stability close to resonances. In our situation, the multiplicity of any resonance is at most  $d$ ,

where  $0 \leq d \leq n$  (with the convention that any frequency is resonant of multiplicity at least 0), and solutions who start sufficiently close to such a resonance are stable with the exponents

$$a(d) = \frac{1}{2((n+1)\tau + n + 1 - d)}, \quad b(d) = \frac{(n+1)\tau + 1}{2((n+1)\tau + n + 1 - d)}$$

and

$$b^*(d) = \frac{(n+1)\tau + 1 - n}{2((n+1)\tau + n + 1 - d)}.$$

In the case of a resonance of maximal multiplicity  $d = n$  these exponents read

$$a(n) = \frac{1}{2((n+1)\tau + 1)}, \quad b(n) = \frac{1}{2}.$$

This will be proved in Theorem 5.1. One recovers the improved stability exponents of the periodic case by setting  $\tau = 0$ , and our main result by setting  $d = 0$ . Using this improved stability at resonances, one can proceed as in [10] to slightly improve the value of the exponent  $a$  by greatly decreasing the value of the exponent  $b$  (and loosing control on the  $J$  variables for  $m \geq 2$ ). This will be the content of Theorem 5.2.

The complement of the neighborhoods of all resonances is the non-resonant domain (this domain contains, in particular, invariant tori, if any). Solutions starting in the non-resonant domain are stable with the exponents

$$a = \frac{1}{2(n+1)(\tau + 1)}, \quad b' = b'^* = \frac{1}{2}.$$

Moreover, the complement of this non-resonant domain is actually very small: its measure is of order

$$\varepsilon^b, \quad b = \frac{(n+1)\tau + 1}{2(n+1)(1 + \tau)}$$

and therefore goes to zero with  $\varepsilon$ . This will be stated as Theorem 6.1. This actually improves on the corresponding statement for  $\tau = 0$ , where the measure estimate of the complement is just of order one. This non-resonant domain comes from the proof of our main theorem, and ends up quite large as we need to exclude very small neighborhoods of resonances. A more natural definition of a non-resonant domain yield the following result: given any  $0 < \gamma' \leq \gamma$  and any  $\tau'$  such that  $\tau' > n + m - 1$  and  $\tau' \geq \tau$ , where  $\gamma$  and  $\tau$  are the constants appearing in  $(\text{Dio}_{\gamma, \tau})$ , there is a set whose complement has a measure of order  $\gamma'$ , such that on this set, the estimates (1.2) hold true with the exponents

$$a' = \frac{1}{2(\tau' + 1)}, \quad b' = b'^* = \frac{1}{2}.$$

This will be the content of Theorem 6.2. Hence on a smaller non-resonant subset, but which is still relatively large, we have a stronger stability result, with  $a'$  arbitrarily

close to the value conjectured and  $b'$  much better. As a matter of fact, when  $\tau > m - 1$  (for  $m \geq 2$ , the set of vectors  $\alpha$  for which  $\tau = m - 1$  has zero measure), one can choose  $\tau' = n + \tau$  and then  $a'$  coincides with the value conjectured.

Finally, as usual with results in Hamiltonian perturbation theory concerning long but finite time scale, the Diophantine condition ( $\text{Dio}_{\gamma, \tau}$ ) on the vector  $\alpha \in \mathbb{R}^m$  turns out to be unnecessary. For an arbitrary vector  $\alpha \in \mathbb{R}^m$  which is assumed to be non-resonant, that is

$$k \cdot \alpha \neq 0, \quad k \neq 0 \in \mathbb{Z}^m,$$

we will obtain in Theorem 7.3 a more general stability result which reduces to the main result in the case where  $\alpha$  is Diophantine.

**1.2. Main result.** Let us now state more precisely our main result. We consider a Hamiltonian of the form

$$H(\theta, \varphi, I, J) = h(I) + \alpha \cdot J + f(\theta, \varphi, I, J),$$

$$(\theta, I) \in \mathbb{T}^n \times D, \quad (\varphi, J) \in \mathbb{T}^m \times \mathbb{R}^m \quad (\text{H})$$

where  $h$  is the integrable part and  $f$  the perturbation. The Hamiltonian  $\bar{h}$ , defined on  $\bar{D} := D \times \mathbb{R}^m$  by

$$\bar{h}(I, J) := h(I) + \alpha \cdot J, \quad (I, J) \in \bar{D}$$

will be called the extended integrable part. The functions  $h$  and  $f$  are assumed to be real-analytic as follows. Given two parameters  $r_0 > 0$  and  $s_0 > 0$ , we define the complex domains

$$V_{r_0} D := \{I \in \mathbb{C}^n \mid \|I - D\| < r_0\},$$

$$V_{r_0} \bar{D} := \{(I, J) \in \mathbb{C}^{n+m} \mid \|(I, J) - \bar{D}\| < r_0\}$$

where

$$\|I - D\| := \inf_{I' \in D} \|I - I'\|, \quad \|(I, J) - \bar{D}\| := \inf_{(I', J') \in \bar{D}} \|(I, J) - (I', J')\|,$$

and

$$V_{s_0} \mathbb{T}^{n+m} := \left\{ (\theta, \varphi) \in \mathbb{C}^{n+m} / (2\pi\mathbb{Z})^{n+m} \mid \max_{1 \leq i \leq n} |\text{Im}(\theta_i)| < s_0, \right. \\ \left. \max_{1 \leq i \leq m} |\text{Im}(\varphi_i)| < s_0 \right\}.$$

Let us also define the associated real domains

$$U_{r_0} D := V_{r_0} D \cap \mathbb{R}^n,$$

$$U_{r_0} \bar{D} := V_{r_0} \bar{D} \cap \mathbb{R}^{n+m} = U_{r_0} D \times \mathbb{R}^m.$$

The function  $h$  is assumed to be real-analytic on  $V_{r_0}D$  so that  $\bar{h}$  is real-analytic on  $V_{r_0}\bar{D}$ , and its Hessian  $\nabla^2 h$  is assumed to be uniformly bounded on the complex domain  $V_{r_0}D$ , namely there exists  $M > 0$  such that

$$\sup_{I \in V_{r_0}D} \|\nabla^2 h(I)\| = \sup_{(I,J) \in V_{r_0}\bar{D}} \|\nabla^2 \bar{h}(I, J)\| \leq M \quad (M)$$

where the matrix norm is the one induced by the Euclidean norm. The gradient of  $h$  is also assumed to be uniformly bounded on the real domain  $U_{r_0}D$ , that is there exists  $\Omega > 0$  such that

$$\sup_{I \in U_{r_0}D} \|(\nabla h(I), \alpha)\| = \sup_{I \in U_{r_0}\bar{D}} \|\nabla \bar{h}(I)\| \leq \Omega. \quad (\Omega)$$

Moreover, the integrable Hamiltonian is assumed to be (strictly, uniformly) convex: there exists  $\kappa > 0$  such that for any  $v \in \mathbb{R}^n$ ,

$$\nabla^2 h(I)v \cdot v \geq \kappa \|v\|^2. \quad (\kappa)$$

Observe that  $\Omega \geq \|\alpha\|$  and  $\kappa \leq M$ .

Finally, the function  $f$  is real-analytic on  $V_{r_0}\bar{D} \times V_{s_0}\mathbb{T}^{n+m}$ , and moreover, given a small parameter  $\varepsilon \geq 0$ , it is assumed that

$$|f|_{r_0, s_0} \leq \varepsilon \quad (\varepsilon)$$

where the Fourier norm  $|f|_{r_0, s_0}$  of  $f$  is defined as follows: letting

$$f(\theta, \varphi, I, J) = \sum_{(k,l) \in \mathbb{Z}^{n+m}} f_{k,l}(I, J) e^{i(k,l) \cdot (\theta, \varphi)}$$

be the Fourier expansion of  $f$  with respect to  $(\theta, \varphi)$ , we define

$$|f|_{r_0, s_0} := \sup_{(I,J) \in V_{r_0}\bar{D}} \sum_{(k,l) \in \mathbb{Z}^{n+m}} |f_{k,l}(I, J)| e^{|(k,l)|s_0}, \quad |(k,l)| = |k| + |l|.$$

We can now state our main theorem.

**Theorem 1.2.** *Let  $H$  be as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , and let us define*

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

and

$$R_* = \frac{\kappa r_0 \gamma}{10M(n+1)^\tau}, \quad T_* = \frac{3s_0}{\Omega},$$

$$\varepsilon_0 = \frac{\kappa r_0^2 \gamma^2}{2^{10}(n+1)^{2\tau}} \left( \frac{\kappa}{10M} \right)^{2(n+1)}, \quad \varepsilon_* = \varepsilon_0 \left( \frac{r_0}{R_*} \right)^{\frac{1}{b}}.$$

If  $\kappa r_0 \leq \min\{80, 24\|\alpha\|\}$  and  $\varepsilon \leq \min\{\varepsilon_0, \varepsilon_*\}$ , for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D} \times \mathbb{T}^{n+m}$ , we have

$$\|I(t) - I_0\| \leq R(\varepsilon) := R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b, \quad |t| \leq T(\varepsilon) := T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_0}{\varepsilon} \right)^a \right).$$

Moreover, in the case where  $m = 1$  and hence  $\tau = 0$ , we have

$$|J(t) - J_0| \leq (\Omega/|\alpha| + 1) R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_0}{\varepsilon} \right)^a \right)$$

whereas in the case  $m \geq 2$  and hence  $\tau \geq 1$ , we have

$$|J(t) - J_0|_\infty := \sup_{1 \leq i \leq m} |J_i(t) - J_i(0)| \leq R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^{b^*}, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_0}{\varepsilon} \right)^a \right).$$

Let us just make one comment concerning the assumption  $(\varepsilon)$ , which requires the perturbation  $f(\theta, \varphi, I, J)$  to be uniformly bounded in  $J$ , with  $J$  belonging to the  $r_0$ -neighborhood of  $\mathbb{R}^m$  in  $\mathbb{C}^m$ . When  $H$  comes from a time-dependent quasi-periodic perturbation of a convex integrable Hamiltonian, the perturbation is in fact independent of  $J$  so the above requirement is void. As a matter of fact, since we are able to control the evolution of the  $J$  variables, one can obtain a stability result assuming only that the perturbation  $f(\theta, \varphi, I, J)$  is uniformly bounded in  $J$ , with  $J$  belonging to the  $r_0$ -neighborhood of some fixed bounded domain  $U \subset \mathbb{R}^m$  in  $\mathbb{C}^m$ . In the periodic case  $m = 1$  this is easily done and one obtains exactly the same result with this weakened assumption. The problem in the quasi-periodic case  $m \geq 2$  is that the control on the  $J$  variables we obtain (which is given by the exponent  $b^*$ ) is worse than the control on the  $I$  variables (given by the exponent  $b$ ) and as a result, under this weakened assumption one would obtain a worse stability result. The discrepancy between the exponents  $b$  and  $b^*$  will be discussed in more details in Section 1.4.

Next, exactly as in [27], replacing the radius of analyticity  $r_0$  by the smaller radius  $r_0(\varepsilon/\varepsilon_0)^{\frac{1-\mu}{2}}$  for  $0 \leq \mu \leq 1$ , we immediately obtain the following more general statement.

**Theorem 1'.** Let  $H$  be as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , and let us define

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

and

$$R_* = \frac{\kappa r_0 \gamma}{10M(n+1)^\tau}, \quad T_* = \frac{3s_0}{\Omega},$$

$$\varepsilon_0 = \frac{\kappa r_0^2 \gamma^2}{2^{10}(n+1)^{2\tau}} \left( \frac{\kappa}{10M} \right)^{2(n+1)}, \quad \varepsilon_* = \varepsilon_0 \left( \frac{r_0}{R_*} \right)^{\frac{1}{b}}.$$



If  $\kappa r_0 \leq \min\{80, 24\|\alpha\|\}$  and  $\varepsilon \leq \min\{\varepsilon_0, \varepsilon_*\}$ , for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D} \times \mathbb{T}^{n+m}$ , we have

$$\|I(t) - I_0\| \leq R_* \left(\frac{\varepsilon}{\varepsilon_0}\right)^{\mu b + \frac{1-\mu}{2}}, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^{\mu a}\right)$$

for any  $0 \leq \mu \leq 1$ . Moreover, in the case where  $m = 1$  and hence  $\tau = 0$ , we have

$$|J(t) - J_0| \leq (\Omega/|\alpha| + 1) R_* \left(\frac{\varepsilon}{\varepsilon_0}\right)^{\mu b + \frac{1-\mu}{2}}, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^{\mu a}\right)$$

whereas in the case  $m \geq 2$  and hence  $\tau \geq 1$ , we have

$$|J(t) - J_0|_\infty \leq R_* \left(\frac{\varepsilon}{\varepsilon_0}\right)^{\mu b^* + \frac{1-\mu}{2}}, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^{\mu a}\right).$$

Observe that Theorem 1.2 reduces to the case  $\mu = 1$  of Theorem 1'.

**1.3. Strategy of the proof.** There are two known methods to prove Nekhoroshev type estimates for small perturbations of integrable Hamiltonian systems. The first one is the Nekhoroshev–Pöschel’s method, introduced in the seminal work of Nekhoroshev [25, 26] and later improved by Pöschel [27] in the convex case (see also [18] for a further extension of the work of Nekhoroshev and Pöschel leading to an improved and conjecturally optimal value of the stability exponents in the steep case). The second method is the Lochak method, introduced by Lochak [20, 23] in the convex case (see also [11] for an extension to the steep case, though with worse values for the stability exponents). In the convex case, the latter method is undoubtedly the simplest and most elegant way to prove stability estimates.

The Lochak method crucially relies on the existence of periodic orbits for the integrable system, that is on the existence of periodic frequencies. Now in the case of a quasi-periodic perturbation, the space of frequencies is of the form  $(\omega, \alpha) \in \mathbb{R}^{n+m}$ , where  $\omega \in \mathbb{R}^n$  is free but  $\alpha \in \mathbb{R}^m$  fixed and non-resonant (in particular,  $\alpha$  is not periodic; if it were, one would be in fact looking at a time-dependent periodic perturbation). The issue is that this space does not contain periodic frequencies. As a matter of fact, it is not really necessary to have exact periodic frequencies, but only frequencies which can be approximated by periodic ones. Now any frequency of the form  $(\omega, \alpha)$  can be approximated by a periodic frequency, say  $(\omega', \alpha')$ . But then necessarily  $\alpha' \neq \alpha$ , and since our Hamiltonian is not convex in the  $J$  variables we were not able to prove stability close to such periodic frequencies. Therefore the method of Lochak does not seem to extend in a easy way to the case of a quasi-periodic perturbation.

Our strategy is therefore to try to extend the Nekhoroshev–Pöschel method, and we will succeed in doing so. This method consists of covering the space of frequencies

by resonant blocks, which are neighborhood of resonances (defined by a certain lattice of integer vectors) that are otherwise non-resonant (for integer vectors not in the lattice). Using convexity, one can then show that the solutions stay in the same resonant block for a long time (in the general steep case, they may leave their resonant block and the proof of the stability gets much harder). In the autonomous case where the frequency space is just  $\omega \in \mathbb{R}^n$ , resonances define linear subspaces which are orthogonal to arbitrary submodules of  $\mathbb{Z}^n$ . In the quasi-periodic case, our frequency space is still  $n$ -dimensional but resonances are associated to submodules of  $\mathbb{Z}^{n+m}$ . But not all submodules of  $\mathbb{Z}^{n+m}$  are associated to resonances: those that are not will be called non admissible. Now resonances associated to admissible submodules do not necessarily define linear subspaces but rather affine subspaces in the space of  $\omega \in \mathbb{R}^n$ . Therefore we are facing much more resonances than in the autonomous case, and the geometry of these resonances gets more involved. In particular, different admissible submodules might lead to different but parallel affine subspaces, and it is at this point that the assumption that  $\alpha$  is Diophantine (or in fact simply non-resonant) comes into play: it ensures that we can control the distance between these parallel affine subspaces. In particular, in the extreme case where these parallel affine subspaces are just points (that is, their associated vector space is trivial), they can get very close to each other, and this is precisely from this phenomenon that our values of the stability exponents come from. This improved geometry of resonances will lead to the fact that any frequency  $(\omega, \alpha)$  is close to some resonant frequency of the form  $(\omega', \alpha)$ . Using this, and the fact that our integrable Hamiltonian is convex in the  $I$  variables while linear in the  $J$  variables, we will be able to prove stability for the  $I$  variables. Once we know that the evolution of the  $I$  variables is bounded, using the fact that resonances are associated to admissible submodules, we will obtain stability for the  $J$  variables, with the same interval of time yet with a worse confinement.

**1.4. A discussion on the stability exponents.** First let us recall that the conjecture of Chirikov–Lochak predicts that

$$a = b = \frac{1}{2(n+1+\tau)} \quad (1.8)$$

while we proved

$$a = \frac{1}{2(n+1)(1+\tau)}, \quad b = \frac{(n+1)\tau+1}{2(n+1)(1+\tau)}.$$

For  $\tau = 0$ , the exponents are the same but not for  $\tau \geq 1$ . However there is certainly no contradiction here; our result yields a better confinement but on a worse time-scale. Chirikov initial conjecture is ultimately based on the *ansatz*

$$|(k, l) \cdot (\omega, \alpha)| \sim \frac{1}{K^{n+m-1}}, \quad |k| + |l| \leq K, \quad (1.9)$$



for an arbitrary fixed vector  $\alpha \in \mathbb{R}^m$ , leading to the exponents

$$a = b = \frac{1}{2(n+m)}. \quad (1.10)$$

We refer to [14, Equation (1.8)], for instance. When  $m = 0$  or  $m = 1$ , that is when the perturbation is autonomous or time-periodic, vectors  $\omega \in \mathbb{R}^n$  satisfying the above condition (1.9) do exist: they are called badly approximable, they are the “best” non-resonant vectors and they form a dense set but of zero Lebesgue measure. In those two cases  $m = 0$  and  $m = 1$ , the exponents (1.10) turned out to be correct, but proofs do not use existence (nor density) of those badly approximable vectors: on the contrary, in the work of Lochak it is the “worst” resonant vectors, namely periodic vectors, that play the major role. When  $m \geq 2$ , as observed by Lochak in [22], the ansatz (1.9) and hence the exponents (1.10) are clearly too optimistic: one should require  $\alpha$  to be Diophantine with some exponent  $\tau \geq m - 1$ , and replace (1.9) by

$$|(k, l) \cdot (\omega, \alpha)| \sim \frac{1}{K^{n+\tau}}, \quad |k| + |l| \leq K, \quad (1.11)$$

so that the exponents of (1.10) become those of (1.8). Vectors  $\omega \in \mathbb{R}^n$  satisfying (1.11) do exist, they form a dense set which is furthermore of full Lebesgue measure when  $\tau > m - 1$  (when  $\tau = m - 1$  they have zero Lebesgue measure). In our opinion, it is a very interesting open problem to derive first these exponents when  $m = 0$  or  $m = 1$  using badly approximable vectors, to see if and how one can try to use vectors satisfying (1.11) to possibly reach the values (1.8).

The fact that our exponent  $b$  is much better than the exponent conjectured, whereas our exponent  $a$  is much worse, lead to the following natural question: is it possible to improve the exponent  $a$  (possibly to the conjectured value) at the expense of deteriorating the exponent  $b$ ?<sup>1</sup> Of course, we cannot prove this is not possible (as this would mean that the conjecture is wrong, which of course we do not know) yet we will try to explain why this cannot be done if one strictly follows the confinement mechanism of Nekhoroshev (which is what we do here). Indeed, a crucial point in this mechanism is the so-called non-overlapping of resonances: resonant zones (which are small neighborhoods around exact resonances) of the same multiplicity should not be allowed to intersect, because if they do, this could create a path in action space along which orbits could have large variation. This property was fundamental in the original work of Nekhoroshev [25], but also in [27] (though in a rather disguised form) and then finally in [18]. As a matter of fact, in [25] this non-overlapping of resonances was forced by the definition of the resonant zones which ultimately lead to non-optimal exponents while in [18], using a different definition of resonant zones, which generalizes the one introduced in [27], this property was proved to be satisfied by a very careful analysis, leading to improved and conjecturally optimal exponents.

---

<sup>1</sup>We would like to thank an anonymous referee for pointing out this question.

The point here is that in the quasi-periodic case, there are much more resonances than in the periodic case, hence if one wants to separate resonant zones of the same multiplicity, then one needs to consider much thinner neighborhoods of resonances, and by doing so one obtains at the end a very small radius of confinement. To explain this in more detail, let us start by comparing the simplest yet non-trivial cases  $n = m = 1$  with  $n = 1, m = 2$ . In the first case, we consider the integrable part

$$h(I, J) = \frac{1}{2}I^2 + J, \quad (I, J) \in \mathbb{R}^2 \quad (1.12)$$

while in the second case, we consider the integrable part

$$h(I, J_1, J_2) = \frac{1}{2}I^2 + J_1 + \alpha J_2, \quad (I, J_1, J_2) \in \mathbb{R}^3, \quad \alpha \in \mathbb{R} \setminus \mathbb{Q}. \quad (1.13)$$

For simplicity, let us restrict the discussion to  $I \in [0, 1]$ . In the case (1.12), resonances of order at most  $K \geq 1$  correspond to rational values of  $I$  with denominators bounded by  $K$ , that is

$$\begin{cases} (q, p) \cdot \nabla h(I, J) = 0, \\ |(q, p)|_\infty \leq K, \end{cases} \iff \begin{cases} I = -\frac{p}{q}, \\ q = |(q, p)|_\infty \leq K. \end{cases}$$

Clearly, the distance between any two such rationals is bounded by

$$\left| \frac{p}{q} - \frac{p'}{q'} \right| \geq \frac{1}{qq'} \geq \frac{1}{qK}$$

and therefore to separate the resonant zones, one has to define them as

$$Z_{(q,p)} = \left\{ I \in \mathbb{R}, \mid \left| I - \frac{p}{q} \right| \leq r_{(p,q)} \right\}, \quad r_{(p,q)} \sim \frac{1}{qK}.$$

The analysis then requires to have  $\sqrt{\varepsilon} \lesssim r_{(q,p)}$ , and as  $q$  can be as large as  $K$ , the largest possible choice of  $K$  is  $K \sim \varepsilon^{-1/4}$  and this leads to the exponent of stability  $a = 1/4$ . Moreover, since  $q$  can be as small as one, one also has  $r_{(q,p)} \lesssim K^{-1} \sim \varepsilon^{1/4}$  and thus  $b = 1/4$ . Another way to see this (and this is the point of view taken in [27]) is to observe that at  $I = -p/q$ , for any  $k = (q', p') \in \mathbb{Z}^2$  with  $|k|_\infty \leq K$ , then either  $k \cdot \nabla h(-p/q, J) = 0$  (if  $k = (q', p')$  is proportional to  $(p, q)$ ) or else one has the “small divisor” estimate

$$|k \cdot \nabla h(-p/q, J)| = |(q', p') \cdot \nabla h(-p/q, J)| = \left| p' - \frac{q'p}{q} \right| = \left| \frac{p'q - q'p}{q} \right| \geq \frac{1}{q}.$$

It is then easy to see that this estimate, which holds true at  $I = -p/q$ , can then be extended to any  $I \in Z_{(q,p)}$  provided we choose  $r_{(q,p)}$  as above.

Let us now look at the case (1.13). Here, resonances are much more abundant, namely

$$\begin{cases} (q, p, l) \cdot \nabla h(I, J) = 0, \\ |(q, p, l)|_\infty \leq K, \end{cases} \iff \begin{cases} I = -\frac{p+l\alpha}{q}, \\ q \leq |(q, p, l)|_\infty \leq K. \end{cases}$$

To control the distance between any two such resonant values, assume that  $(1, \alpha)$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , then we have

$$\begin{aligned} \left| \frac{p+l\alpha}{q} - \frac{p'+l'\alpha}{q'} \right| &\geq \frac{|(q'p - qp') + (q'l - ql')\alpha|}{qq'} \\ &\geq \frac{1}{qq'} \frac{\gamma}{(|(q'p - qp')| + |(q'l - ql')|)^\tau} \end{aligned}$$

and using the fact that  $q' \leq K$ ,  $|(q'p - qp')| + |(q'l - ql')| \lesssim K^2$ , one obtains for instance

$$\left| \frac{p+l\alpha}{q} - \frac{p'+l'\alpha}{q'} \right| \gtrsim \frac{\gamma}{qK^{2\tau+1}} \quad (1.14)$$

and therefore resonant zones are separated if one define them as

$$Z_{(q,p,l)} = \left\{ I \in \mathbb{R}, \mid \left| I - \frac{p+l\alpha}{q} \right| \leq r_{(p,q,l)} \right\}, \quad r_{(p,q,l)} \sim \frac{\gamma}{qK^{2\tau+1}}.$$

Then since  $q$  can be as large as  $K$  and since we need  $\sqrt{\varepsilon} \lesssim r_{(p,q,l)}$ , the largest possible choice of  $K$  is  $K \sim \varepsilon^{-\frac{1}{2(2\tau+2)}} = \varepsilon^{-\frac{1}{4(\tau+1)}}$  and this gives the stability exponent  $a = 1/(4(\tau+1))$ , and moreover, as  $q$  can be as small as one, then  $r_{(p,q,l)} \lesssim K^{-(2\tau+1)} \sim \varepsilon^b$  with  $b = (2\tau+1)/(4(\tau+1))$ . As before, one can also see this (and this is the point of view we take in this work, following [27]) as the maximal width around exact resonances for which the small divisors estimate at  $I = -(p+l\alpha)/q$  can be extended to  $Z_{(p,q,l)}$ . One may argue that (1.14) can be estimated differently, for instance with a lower bound that depends on the norm of  $(q, p, l)$  and not just  $q$ . The first point we want to make is that this is not compatible with the approach we will take in this paper: in order to use convexity to control the evolution of the action variables  $I$ , we need to consider resonant zones as neighborhoods of affine subspaces in the  $I$  space (in the example here, these are neighborhoods of points in the  $I$  space) and not as neighborhoods of vectorial subspaces in the  $(I, J_1, J_2)$  space (which, in the example, would be neighborhoods of vectorial lines in the  $(I, J_1, J_2)$  space); indeed, in the second case exact resonances in frequency space would then correspond to vectors of the form  $(I, \beta) \in \mathbb{R}^3$  with  $\beta$  close to  $(1, \alpha)$  but not necessarily of this form, and our argument using convexity to bound the evolution in the  $I$  variables wouldn't work. The second point is that, independently of how one chooses to bound (1.14), in any event this quantity can be as small as  $K^{-2\tau+2} = K^{-2(\tau+1)}$  (it is easy to construct examples of integer

vectors  $(q, p, l)$  and  $(q', p', l')$  generating a maximal lattice in  $\mathbb{Z}^3$  for which this happens), so it seems hard to choose  $K$  larger.

Now in the general case  $n \geq 1$  and  $m \geq 1$ , the same discussion applies at maximal resonances (which are of multiplicity  $n$ ), and leads to the fact that the distance between resonant zones of the same multiplicity can be as small as  $K^{-(n+1)(\tau+1)}$  and therefore the largest choice of  $K$  is  $K \sim \varepsilon^{-\frac{1}{2(n+1)(\tau+1)}}$ , which gives the value of our exponents  $a$  and then  $b = a((n+1)\tau + 1)$ .

Let us also try to give another reason why it is not so unnatural to have such a small radius of confinement. For this we need to recall that the perturbation is allowed to depend on the  $J$  variables, and that we do control the evolution of these variables also, yet with a worse radius of confinement in the case  $m \geq 2$ : for the  $I$  variables the exponents are given by

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

whereas for the  $J$  variables, we obtain the same exponents if  $m = 1$ ,  $\tau = 0$  while in the case  $m \geq 2$ ,  $\tau \geq 1$ , we obtain

$$a = \frac{1}{2(n+1)(1+\tau)}, \quad b^* = \frac{(n+1)\tau + 1 - n}{2(n+1)(1+\tau)} = \frac{1}{2} - \frac{2n}{2(n+1)(1+\tau)}.$$

We do have  $0 < b^* < b$ , but  $b^*$  still gets close to  $1/2$  when  $m$  becomes large. Let us first briefly explain why we do not obtain the same exponents. Using the geometry of resonances and convexity with respect to  $I$ , one first obtain a control on the evolution of the  $I$  variables. As we already explained, in the geometry of resonances, given an admissible submodule  $\Lambda \subset \mathbb{Z}^{n+m}$  since we want to work in the  $I$  space, the width of its resonant zone is controlled by the covolume  $|\tilde{\Lambda}|$  of the projection  $\tilde{\Lambda}$  of  $\Lambda$  onto  $\mathbb{Z}^n$ , and not by the covolume  $|\Lambda|$  of  $\Lambda$  (in the case  $n = 1$  and  $m = 2$  we described in (1.13), the width of the resonant zone depends just on  $q$  and not on the norm of  $(q, p, l)$ ). Clearly one cannot control the  $J$  variables in the same way, that is using the geometry of resonances and convexity, since the integrable Hamiltonian is linear in the  $J$  variables. However, once we know that the  $I$  variables are stable, one can then use the normal form to control the  $J$  variables, using the fact that since  $\Lambda$  has rank at most  $n$ , any resonant normal form has at least  $m$  independent first integrals. But here it is no longer  $|\tilde{\Lambda}|$  but the covolume of the full lattice  $|\Lambda|$  that plays an important role, and since  $|\tilde{\Lambda}|$  and  $|\Lambda|$  might be of different size, we cannot reach  $b^* = b$ . We do believe that one could reach  $b^* = b$  even though we were not able to do so; since  $b^*$  still converge to  $1/2$  when  $m$  goes to infinity, we do believe that this is not so important for our purpose here. Indeed, the point we want to make is to compare our exponents for the “mixed non-linear linear” integrable Hamiltonian

$$\bar{h}(I, J) = h(I) + \alpha \cdot J$$

with the known exponents of the fully non-linear convex integrable Hamiltonian  $h(I)$  (under a periodic time-dependent perturbation, so we consider a quasi-convex integrable Hamiltonian with  $n + 1$  degrees of freedom) and with the known exponents of the fully linear integrable Hamiltonian  $\alpha \cdot J$ . In the first case, the known exponents are

$$a = \frac{1}{2(n+1)}, \quad b = \frac{1}{2(n+1)} \quad (1.15)$$

while in the second case, they are

$$a = \frac{1}{\tau+1}, \quad b = \frac{1}{2}. \quad (1.16)$$

When  $n$  is fixed, and  $\tau$  gets very large, it seems reasonable to expect that the mixed system should essentially behave like the linear system; and indeed, in this case our exponents  $a$  and  $b^* \sim b$  are asymptotically equivalent to those of (1.16) (it is precisely for this reason that an exponent  $b^*$  that converges to  $1/2$  when  $m$  converges to infinity is sufficient for our purpose<sup>2</sup>). The exponents  $a$  and  $b$  in (1.16) are known to be optimal (see [6] for easy examples), and therefore since our result do contain the linear case (one can even put formally  $n = 0$ ), it seems reasonable to have such large exponent  $b$  and  $b^*$ . But then in the same way, when  $\tau$  is fixed, and  $n$  gets very large, it seems reasonable to expect that the mixed system should essentially behave like the non-linear system; yet our exponents  $a$  and  $b$  are asymptotically equivalent to

$$a \sim \frac{1}{2(n+1)}, \quad b \sim \frac{1}{4}$$

so we obtain essentially the same exponent  $a$  but a much better exponent  $b$ . As a matter of fact, only the exponent  $a$  in (1.15) is known to be optimal; it is still an open question whether the exponents

$$a = \frac{1}{2(n+1)}, \quad b = \frac{1}{2} \quad (1.17)$$

can hold uniformly in phase space in the non-linear case; our result shows that for  $n$  large, if one adds linearity in just one degree of freedom then exponents close to (1.17) do hold uniformly in phase space.

Clearly this discussion does not give evidence that the exponents of the conjecture cannot be reached; it just shows that it could be hard to improve on the value of the exponent  $a$  by simply following the mechanism of Nekhoroshev. One could try to use another mechanism taking into account that the system is not fully non-linear (the mechanism of Nekhoroshev is, by nature, essentially non-linear); we believe that adding the assumption that the perturbation does not depend on  $J$  should play an important role here, but so far we haven't been able to exploit such a feature in

---

<sup>2</sup>As a matter of fact, this comparison was our motivation to obtain stability for the  $J$  variables.



an effective way. This discussion also shows that it is not unreasonable to believe that the exponent  $b = 1/2$  could prevail uniformly in phase space, both for the fully non-linear and the mixed case.

**1.5. Plan of the paper.** The plan of the paper is as follows. Section 2 deals in details with the geometry of resonances that was alluded above. This section contains the main technical part of the work. Section 3 deals with the analysis (construction of a normal form) and the local stability results. The analysis, and therefore the stability in the non-resonant case, is completely standard and we can simply refer to [27]. The stability in the resonant case uses convexity: our integrable Hamiltonian is just “partially” convex so we need to justify that the arguments go through, and how one can afterwards control the  $J$  variables also. The proof of our main result Theorem 1.2 will be given in Section 4, using the results of Section 2 and Section 3. The next sections contain further results that were mentioned in the Introduction: namely, we prove better stability results for solutions close to resonances in Section 5 or far away from resonances in Section 6, while in Section 7 we give a more general result assuming  $\alpha$  to be only non-resonant. The last Section 8 consists of concluding remarks.

## 2. Geometry of resonances

The purpose of this section is to study the resonant and non-resonant properties of the frequency space

$$\{(\omega, \alpha) \in \mathbb{R}^{n+m}\} \simeq \{\omega \in \mathbb{R}^n\}$$

where  $\alpha \in \mathbb{R}^m$  is a fixed vector, which will be assumed to be Diophantine, and  $\omega$  is a vector varying freely in  $\mathbb{R}^n$ . More precisely, our aim is to cover this space by neighborhoods of resonances (associated to certain submodules  $\Lambda$  of  $\mathbb{Z}^{n+m}$ ) on which non-resonant estimates can be established (for integer vectors  $k \notin \Lambda$ ).

**2.1. Admissible resonant zones and resonant blocks.** We fix a real parameter  $K \geq 1$ , and in this section,  $\alpha \in \mathbb{R}^m$  will be assumed to be simply non-resonant.

A submodule  $\Lambda$  of  $\mathbb{Z}^{n+m}$  is said to be a  $K$ -submodule if it is generated by elements  $(k, l) \in \mathbb{Z}^n \times \mathbb{Z}^m = \mathbb{Z}^{n+m}$  such that  $|(k, l)| \leq K$ , and it is said to be maximal if it is not properly contained in any other submodule of the same dimension. Given an integer  $1 \leq d \leq n + m$ , the set of all maximal  $K$ -submodules  $\Lambda$  of  $\mathbb{Z}^{n+m}$  of rank  $d$  will be denoted by  $M_{K,d}$ . For  $\Lambda \in M_{K,d}$ , we define the space of  $\Lambda$ -resonances by

$$R_\Lambda = \{\omega \in \mathbb{R}^n \mid (k, l) \cdot (\omega, \alpha) = 0, \forall (k, l) \in \Lambda\}. \quad (2.1)$$

Quite obviously, since  $\alpha \in \mathbb{R}^m$  is non-resonant,  $R_\Lambda$  will be non-empty only for certain maximal  $K$ -submodules  $\Lambda$ . Let us consider the subset  $M_{K,d}^a$  of  $M_{K,d}$  consisting of

admissible submodules: they are submodules whose intersection with  $\{0\} \times \mathbb{Z}^m \subset \mathbb{Z}^n \times \mathbb{Z}^m = \mathbb{Z}^{n+m}$  is trivial. Equivalently, given any basis  $(k^1, l^1), \dots, (k^d, l^d)$  for  $\Lambda$ , the vectors  $k^1, \dots, k^d$  in  $\mathbb{Z}^n$  are linearly independent. It is plain to check that if  $\Lambda$  is not admissible, then  $R_\Lambda$  is just the empty set. Note also that if  $\Lambda$  is admissible, its rank is at most  $n$ .

Now consider  $\Lambda \in M_{K,d}^a$  where  $1 \leq d \leq n$ . If  $\Pi : \mathbb{R}^{n+m} \mapsto \mathbb{R}^n$  is the canonical projection,  $\tilde{\Lambda} := \Pi(\Lambda)$  is a submodule of  $\mathbb{Z}^n$ , of rank  $d$ , which generates a real subspace  $\langle \tilde{\Lambda} \rangle$  of  $\mathbb{R}^n$  of dimension  $d$ . It is clear that  $\tilde{\Lambda}$  is a  $K$ -submodule, but it is not necessarily maximal. The space of  $\Lambda$ -resonances defined in (2.1) is non-empty, it is an affine subspace of  $\mathbb{R}^n$  whose associated vector subspace is the vector subspace  $\langle \tilde{\Lambda} \rangle^\perp$  orthogonal to  $\langle \tilde{\Lambda} \rangle$ .

It is not the space of resonances but rather their neighborhoods that will play a role in the construction below. To define them, given  $\Lambda \in M_{K,d}^a$  and its associated submodule  $\tilde{\Lambda}$ , we first define  $|\tilde{\Lambda}|$  as the co-volume of  $\tilde{\Lambda}$  viewed as a lattice in  $\langle \tilde{\Lambda} \rangle$ . It is the volume of the fundamental domain spanned by the vectors of any choice of basis for  $\tilde{\Lambda}$ : letting  $A$  be an  $n \times d$  matrix whose columns form a basis for  $\tilde{\Lambda}$ , then  $|\tilde{\Lambda}| = \sqrt{\det A^t A}$ , and this latter quantity is easily seen to be independent of the choice of a basis. It is worth recalling, as it will be used, that if  $S_d(A)$  denotes all square matrices of size  $d$  that can be extracted from  $A$ , then we have the equality (Cauchy–Binet formula)

$$|\tilde{\Lambda}| = \sqrt{\det A^t A} = \sqrt{\sum_{B \in S_d(A)} (\det B)^2}.$$

Then, we introduce  $n$  positive real parameters  $\lambda_1, \lambda_2, \dots, \lambda_n$  and, for  $1 \leq d \leq n$ , we define the associated resonant zone

$$Z_\Lambda := \{\omega \in \mathbb{R}^n \mid \|\omega - R_\Lambda\| < \delta_\Lambda\}, \quad \delta_\Lambda := \frac{\lambda_d}{|\tilde{\Lambda}|} \quad (2.2)$$

where

$$\|\omega - R_\Lambda\| := \inf_{\omega' \in R_\Lambda} \|\omega - \omega'\|.$$

We then define the resonant zone of multiplicity  $d$ , for  $1 \leq d \leq n+1$ , by

$$\begin{cases} Z_d := \bigcup_{\Lambda \in M_{K,d}^a} Z_\Lambda, & 1 \leq d \leq n \\ Z_{n+1} = \emptyset. \end{cases}$$

The resonant block associated to  $\Lambda \in M_{K,d}^a$ ,  $1 \leq d \leq n$ , are defined by

$$B_\Lambda := Z_\Lambda \setminus Z_{d+1}$$

and eventually the resonant block of multiplicity  $d$ , for  $1 \leq d \leq n$ , is

$$B_d := \bigcup_{\Lambda \in M_{K,d}^a} B_\Lambda.$$

Setting  $B_{\{0\}} = \mathbb{R}^n \setminus Z_1$ , we arrive at the following decomposition

$$\mathbb{R}^n = B_{\{0\}} \cup Z_1 = B_{\{0\}} \cup B_1 \cup Z_2 = \cdots = B_{\{0\}} \cup B_1 \cup \cdots \cup B_{n-1} \cup B_n \quad (2.3)$$

since  $B_n = Z_n$  as  $Z_{n+1} = \emptyset$ .

**2.2. Non-resonant domains in frequency space.** Consider a domain  $B \subset \mathbb{R}^n$ , a submodule

$$\Lambda \in M_K^a := \bigcup_{1 \leq d \leq n} M_{K,d}^a \cup \{0\}$$

and a real parameter  $\beta > 0$ . Then the domain  $B$  is said to be  $(\beta, K)$ -non resonant modulo  $\Lambda$  if for any  $(k, l) \in \mathbb{Z}^{n+m}$  such that  $|(k, l)| \leq K$  and  $(k, l) \notin \Lambda$  and any  $\omega \in B$ , we have

$$|(k, l) \cdot (\omega, \alpha)| \geq \beta.$$

Our purpose here is to show that the resonant blocks  $B_\Lambda$ , for  $\Lambda \in M_K^a$ , are  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$ , for a suitable  $\beta_\Lambda$  provided that  $\alpha$  is Diophantine and the parameters  $\lambda_1, \lambda_2, \dots, \lambda_n$  satisfies certain compatibility conditions. This is the content of the lemma below.

**Lemma 2.1.** *Let  $K \geq 1$ ,  $E > 0$  and  $F \geq E + 1$ . Assume that  $\alpha \in \mathbb{R}^m$  satisfies  $(\text{Dio}_{\gamma, \tau})$  and*

$$\begin{cases} FK\lambda_d \leq \lambda_{d+1} \leq \gamma(d+1)^{-\tau} K^{-(d+1)\tau}, & 1 \leq d \leq n-1, \\ \lambda_n \leq F^{-1}\gamma(n+1)^{-\tau} K^{-(n+1)\tau-1}. \end{cases} \quad (2.4)$$

*Then for any  $\Lambda \in M_K^a$ , the block  $B_\Lambda$  is  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$  with*

$$\begin{cases} \beta_\Lambda = EK\delta_\Lambda, & \Lambda \neq \{0\}, \\ \beta_{\{0\}} = \lambda_1. \end{cases} \quad (2.5)$$

*Proof.* Let  $\Lambda \in M_K^a$  of rank  $d$ , with  $0 \leq d \leq n$ , and  $(k, l) \notin \Lambda$  such that  $|(k, l)| \leq K$ . Let  $\Lambda_+$  be the submodule of  $\mathbb{Z}^{n+m}$  generated by  $\Lambda$  and  $(k, l)$ . Since  $\Lambda$  is maximal and does not contain  $(k, l)$ , the rank of  $\Lambda_+$  is equal to  $d + 1$ .

Let us start with the special case  $d = 0$ , that is  $\Lambda = \{0\}$ , and fix  $\omega \in B_{\{0\}}$ . Either  $\Lambda_+$  is admissible, or not. In the second situation,  $k = 0 \in \mathbb{Z}^n$  while  $l \neq 0 \in \mathbb{Z}^m$  as  $(k, l) \notin \Lambda = \{0\}$ , and we have, using the fact that  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$  and (2.4),

$$|(k, l) \cdot (\omega, \alpha)| = |l \cdot \alpha| \geq \gamma K^{-\tau} \geq \lambda_1. \quad (2.6)$$

In the first situation where  $k \neq 0 \in \mathbb{Z}^n$ , let  $\omega_+ \in R_{\Lambda_+}$  such that

$$\|\omega - \omega_+\| = \|\omega - R_{\Lambda_+}\| > \delta_{\Lambda_+}.$$



Now as  $(k, l) \cdot (\omega_+, \alpha) = 0$  we have

$$(k, l) \cdot (\omega, \alpha) = (k, l) \cdot (\omega, \alpha) - (k, l) \cdot (\omega_+, \alpha) = k \cdot (\omega - \omega_+).$$

But the vector  $\omega - \omega_+$  belongs to the line orthogonal to  $\langle \tilde{\Lambda}_+ \rangle^\perp = \langle k \rangle^\perp$  in  $\langle \tilde{\Lambda} \rangle^\perp = \mathbb{R}^n$  which is nothing but the line generated by  $k$ , so we obtain

$$|(k, l) \cdot (\omega, \alpha)| = |k \cdot (\omega - \omega_+)| = \|k\| \|\omega - \omega_+\| > \|k\| \delta_{\Lambda_+} = \|k\| |\tilde{\Lambda}_+|^{-1} \lambda_1 \geq \lambda_1 \quad (2.7)$$

where we used the fact  $|\tilde{\Lambda}_+| \leq \|k\|$ . From (2.6) and (2.7) the statement in the case  $d = 0$  follows.

Now assume  $1 \leq d \leq n$ . It is enough to prove that given any  $\omega \in R_\Lambda \setminus Z_{d+1}$  (where we recall that  $Z_{n+1} = \emptyset$ ), we have

$$|(k, l) \cdot (\omega, \alpha)| \geq FK\delta_\Lambda. \quad (2.8)$$

Indeed, for any  $\bar{\omega} \in B_\Lambda = Z_\Lambda \setminus Z_{d+1}$ , by definition there exists  $\omega \in R_\Lambda \setminus Z_{d+1}$  such that  $\|\omega - \bar{\omega}\| < \delta_\Lambda$  and thus

$$\begin{aligned} |(k, l) \cdot (\bar{\omega}, \alpha)| &\geq |(k, l) \cdot (\omega, \alpha)| - |k \cdot (\bar{\omega} - \omega)| \\ &\geq |(k, l) \cdot (\omega, \alpha)| - \|k\| \|\bar{\omega} - \omega\| \\ &\geq |(k, l) \cdot (\omega, \alpha)| - K \|\bar{\omega} - \omega\| \\ &\geq FK\delta_\Lambda - K\delta_\Lambda = (F - 1)K\delta_\Lambda \geq EK\delta_\Lambda. \end{aligned}$$

So it suffices to prove (2.8). As before, there are two possibilities for  $\Lambda_+$ : either it is admissible, or not. In the second situation (which is obviously the only possibility for  $d = n$ ), let us choose a basis  $(k^1, l^1), \dots, (k^d, l^d)$  for  $\Lambda$  such that  $|(k^j, l^j)| \leq K$  for  $1 \leq j \leq d$ . The assumption that  $\Lambda_+$  is not admissible means that  $k$  is a linear combination of  $k^1, \dots, k^d$ . For  $1 \leq j \leq d$ , let us define  $\bar{k}^j \in \mathbb{Z}^d$  by selecting the first  $d$  components of  $k^j$ , and we define  $\bar{k}$  the same way. As  $k^1, \dots, k^d$  are linearly independent (because  $\Lambda$  is admissible), the determinant  $\Delta(\bar{k}^1, \dots, \bar{k}^d)$  of the square matrix of size  $d$  whose columns are given by  $\bar{k}^1, \dots, \bar{k}^d$  can be assumed to be non-zero without loss of generality. By Cramer's rule, it then follows that  $k$  can be written as

$$k = v_1 k^1 + v_2 k^2 + \dots + v_d k^d = \frac{\Delta_1}{\Delta} k^1 + \frac{\Delta_2}{\Delta} k^2 + \dots + \frac{\Delta_d}{\Delta} k^d$$

where  $\Delta = \Delta(\bar{k}^1, \dots, \bar{k}^d)$ ,  $\Delta_1 = \Delta(\bar{k}, \bar{k}^2, \dots, \bar{k}^d)$ ,  $\Delta_d = \Delta(\bar{k}^1, \bar{k}^2, \dots, \bar{k})$  and for  $2 \leq j \leq d-1$ ,  $\Delta_j = \Delta(\bar{k}^1, \dots, \bar{k}^{j-1}, \bar{k}, \bar{k}^{j+1}, \dots, \bar{k}^d)$ . Let us denote

$$l_* = l - v_1 l^1 - \dots - v_d l^d \in \mathbb{R}^m.$$

The vector  $l_*$  is non-zero: if it were,  $(k, l)$  would be a linear combination of  $(k^1, l^1), \dots, (k^d, l^d)$ , that is  $(k, l) \in \langle \Lambda \rangle \cap \mathbb{Z}^{n+m}$ , but since  $\Lambda$  is maximal it is

equal to  $\langle \Lambda \rangle \cap \mathbb{Z}^{n+m}$  and so this would contradict  $(k, l) \notin \Lambda$ . Now using the fact that  $(k^j, l^j) \in \Lambda$  for  $1 \leq j \leq d$  and  $\omega \in R_\Lambda$  we can write

$$(k, l) \cdot (\omega, \alpha) = (k, l) \cdot (\omega, \alpha) - \sum_{j=1}^d v_j (k^j, l^j) \cdot (\omega, \alpha) = (l_*, \alpha).$$

The vector  $\Delta l_* = \Delta l - \Delta_1 l^1 - \dots - \Delta_d l^d \in \mathbb{Z}^m$ , and by Hadamard's formula, we have

$$|\Delta| \leq \|\bar{k}^1\| \cdots \|\bar{k}^d\| \leq K^d$$

and similarly  $|\Delta_j| \leq K^d$  for  $1 \leq j \leq d$ , so as a consequence

$$|\Delta l_*| \leq |\Delta||l| + |\Delta_1||l^1| + \dots + |\Delta_d||l^d| \leq (d+1)K^{d+1}.$$

Since  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , it follows that

$$|\Delta|(k, l) \cdot (\omega, \alpha) = |\Delta|(l_*, \alpha) = |(\Delta l_*, \alpha)| \geq \gamma(d+1)^{-\tau} K^{-(d+1)\tau}$$

and as a consequence of Cauchy–Binet formula,  $|\Delta| \leq |\tilde{\Lambda}|$  and hence

$$|(k, l) \cdot (\omega, \alpha)| \geq |\tilde{\Lambda}|^{-1} \gamma(d+1)^{-\tau} K^{-(d+1)\tau}.$$

Using this last inequality together with (2.4) one arrives at

$$|(k, l) \cdot (\omega, \alpha)| \geq |\tilde{\Lambda}|^{-1} \lambda_{d+1} \geq |\tilde{\Lambda}|^{-1} FK \lambda_d = FK \delta_\Lambda. \quad (2.9)$$

It remains to treat the case where  $\Lambda_+$  is admissible. Let  $\omega_+ \in R_{\Lambda_+}$  such that

$$\|\omega - \omega_+\| = \|\omega - R_{\Lambda_+}\| > \delta_{\Lambda_+}.$$

Then since  $(k, l) \cdot (\omega_+, \alpha) = 0$  we have

$$(k, l) \cdot (\omega, \alpha) = k \cdot (\omega - \omega_+).$$

The vector  $\omega - \omega_+$  belongs to the line orthogonal to  $\langle \tilde{\Lambda}_+ \rangle^\perp$  within  $\langle \tilde{\Lambda} \rangle^\perp$ , that is it belongs to  $L := \langle \tilde{\Lambda}_+ \rangle^{\perp\perp} \cap \langle \tilde{\Lambda} \rangle^\perp = \langle \tilde{\Lambda}_+ \rangle \cap \langle \tilde{\Lambda} \rangle^\perp$ . Let us decompose  $k = Pk + (\text{Id} - P)k$  where  $Pk \in L$  and  $(\text{Id} - P)k \in L^\perp$ . It follows that

$$|(k, l) \cdot (\omega, \alpha)| = |k \cdot (\omega - \omega_+)| = |Pk \cdot (\omega - \omega_+)| = \|Pk\| \|\omega - \omega_+\| > \|Pk\| \delta_{\Lambda_+}.$$

But  $L$  is also the line orthogonal to  $\langle \tilde{\Lambda} \rangle$  within  $\langle \tilde{\Lambda}_+ \rangle$ , and therefore  $|\tilde{\Lambda}_+| \leq \|Pk\| |\tilde{\Lambda}|$  so, using (2.4) again,

$$|(k, l) \cdot (\omega, \alpha)| > |\tilde{\Lambda}_+| |\tilde{\Lambda}|^{-1} \delta_{\Lambda_+} = |\tilde{\Lambda}|^{-1} \lambda_{d+1} \geq |\tilde{\Lambda}|^{-1} FK \lambda_d = FK \delta_\Lambda. \quad (2.10)$$

From (2.9) and (2.10) it follows that (2.8) holds true for  $1 \leq d \leq n$ , and this concludes the proof.  $\square$

**2.3. Covering by non-resonant domains in action space.** In Section 2.1 we obtained a covering of the frequency space  $\{(\omega, \alpha) \in \mathbb{R}^{n+m}\} \simeq \{\omega \in \mathbb{R}^n\}$  by resonant blocks associated to admissible submodules, and in Lemma 2.1 we proved that these resonant blocks satisfy some non-resonant properties.

Recall that we are given an integrable Hamiltonian  $h$  which is real-analytic on the domain  $V_{r_0}D$ , and satisfies  $(M)$  and  $(\kappa)$ . A subset of  $D$  is said to be  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $h$  if its image by the frequency map  $\nabla h$  is  $(\beta, K)$ -non resonant modulo  $\Lambda$ . It is said to be  $\delta$ -close to  $\Lambda$ -resonances for  $h$  if the Euclidean distance between its image by  $\nabla h$  and the space  $R_\Lambda$  is smaller than  $\delta$ .

Pulling back the covering (2.3) back to action space using the gradient map  $\nabla h$  and using Lemma 2.1, the following proposition will be easily obtained by carefully choosing the parameters  $\lambda_d$ , for  $1 \leq d \leq n$ .

**Proposition 2.2.** *For  $K \geq 1$  and  $\Lambda \in M_K^a$  of rank  $d$ , with  $0 \leq d \leq n$ , let us define*

$$r_\Lambda := \frac{r_0 \gamma}{|\tilde{\Lambda}| F^{n-d+1} (n+1)^\tau K^{(n+1)\tau+n-d+1}}, \quad F := 10M/\kappa$$

where we set, by convention,  $|\tilde{\Lambda}| = 1$  if  $\Lambda = \{0\}$ . Assume that  $\kappa r_0 \leq 80$ . Then there exists a covering of  $D$  by subsets  $D_\Lambda$ , where  $\Lambda \in M_K^a$ , such that each  $D_\Lambda$  is  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$  and, for  $\Lambda \neq \{0\}$ ,  $\delta_\Lambda$ -close to  $\Lambda$ -resonances with

$$\beta_\Lambda = \frac{9MKr_\Lambda}{8}, \quad \delta_\Lambda = \frac{\kappa r_\Lambda}{8}.$$

Exactly as in [27], the introduction of the parameters  $M$  and  $\kappa$  in the above statement is unnecessary (the above proposition does not depend on the assumptions  $(M)$  and  $(\kappa)$ ); these parameters are just here for later convenience.

*Proof.* Recall that (2.3) gives the decomposition

$$\mathbb{R}^n = B_{\{0\}} \cup B_1 \cup \cdots \cup B_{n-1} \cup B_n$$

that can also be written as

$$\mathbb{R}^n = \bigcup_{\Lambda \in M_K^a} B_\Lambda.$$

We now define those resonant blocks by choosing the parameters

$$\lambda_d := \frac{\kappa r_0 \gamma}{8F^{n-d+1} (n+1)^\tau K^{(n+1)\tau+n-d+1}}, \quad F = 10M/\kappa, \quad 1 \leq d \leq n.$$

With these choices, the inequalities (2.4) are satisfied (since  $K \geq 1$ ,  $F = 10M/\kappa \geq 10$  and  $\kappa r_0 \leq 80$ ), hence Lemma 2.1 can be applied with  $E := 9M/\kappa \leq F - 1 = 10M/\kappa - 1$  (since  $M/\kappa \geq 1$ ). By definition, for each non-trivial  $\Lambda$  of rank  $d$ , the block  $B_\Lambda$  is  $\delta_\Lambda$ -close to  $\Lambda$ -resonances with

$$\delta_\Lambda = \frac{\lambda_d}{|\tilde{\Lambda}|} = \frac{\kappa r_0 \gamma}{8F^{n-d+1} |\tilde{\Lambda}| (n+1)^\tau K^{(n+1)\tau+n-d+1}} = \frac{\kappa r_\Lambda}{8}.$$

Moreover, using (2.5) from Lemma 2.1,  $B_\Lambda$  is  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$  with

$$\beta_\Lambda = EK\delta_\Lambda = 9M\kappa^{-1}K\frac{\kappa r_\Lambda}{8} = \frac{9MKr_\Lambda}{8},$$

while, for  $\Lambda = \{0\}$ ,  $B_{\{0\}}$  is  $(\beta_{\{0\}}, K)$ -non resonant with

$$\begin{aligned}\beta_{\{0\}} &\geq \lambda_1 = \frac{\kappa r_0 \gamma}{8F^n(n+1)^\tau K^{(n+1)\tau+n}} \\ &\geq \frac{9MKr_0 \gamma}{8F^{n+1}(n+1)^\tau K^{(n+1)\tau+n+1}} = \frac{9MKr_{\{0\}}}{8}\end{aligned}$$

as  $F \geq 9M/\kappa$ . If we define

$$D_\Lambda := \{I \in D \mid \nabla h(I) \in B_\Lambda\}, \quad \Lambda \in M_K^a,$$

this defines a covering (up to removing such sets which are empty) of  $D$  with all the required properties, and this concludes the proof.  $\square$

### 3. Normal form and stability estimates

**3.1. Normal form.** Let us come back to our original Hamiltonian (H), and recall that the extended integrable Hamiltonian is given by

$$\bar{h}(I, J) = h(I) + \alpha \cdot J, \quad (I, J) \in \bar{D} = D \times \mathbb{R}^m.$$

Let us fix  $\Lambda \in M_K^a$ . Quite obviously, if a subset  $D_* \subset D$  is  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $h$ , then  $\bar{D}_* := D_* \times \mathbb{R}^m \subset \bar{D}$  is  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $\bar{h}$ .

Now a Hamiltonian of the form

$$\bar{h}(I, J) + g(\theta, \varphi, I, J)$$

is said to be in  $\Lambda$ -resonant normal form if

$$g(\theta, \varphi, I, J) = \sum_{(k,l) \in \Lambda} g_{k,l}(I, J) e^{i(k,l) \cdot (\theta, \varphi)}.$$

Such Hamiltonians have additional first integrals: indeed, given any vector  $(\omega, \alpha)$  which belongs to the real subspace orthogonal to  $\Lambda$ , and any solution

$$(\theta(t), \varphi(t), I(t), J(t))$$

of the system associated to  $\bar{h} + g$ , we have

$$\begin{aligned}
 (\omega, \alpha) \cdot (\dot{I}(t), \dot{J}(t)) &= -(\omega, \alpha) \cdot \partial_{(\theta, \varphi)}(\bar{h}(I(t), J(t)) + g(\theta(t), \varphi(t), I(t), J(t))) \\
 &= -(\omega, \alpha) \cdot \partial_{(\theta, \varphi)}(g(\theta(t), \varphi(t), I(t), J(t))) \\
 &= -(\omega, \alpha) \cdot \sum_{(k, l) \in \Lambda} i 2\pi(k, l) g_{k, l}(I(t), J(t)) e^{i(k, l) \cdot (\theta(t), \varphi(t))} \\
 &= -i \sum_{(k, l) \in \Lambda} (\omega, \alpha) \cdot (k, l) g_{k, l}(I(t), J(t)) e^{i(k, l) \cdot (\theta(t), \varphi(t))} \\
 &= 0.
 \end{aligned}$$

In the special case where  $\Lambda = \{0\}$ , it is straightforward to see that  $g$  is in fact independent of the angles  $(\theta, \varphi)$ , so that the resonant normal form is integrable.

We can now state the normal form lemma, which states that on a sufficiently small neighborhood of a non-resonant domain modulo  $\Lambda$ , up to a real-analytic symplectic transformation which is close to the identity, the original Hamiltonian can be written as a  $\Lambda$ -resonant normal form  $\bar{h} + g$  up to an exponentially small remainder.

**Lemma 3.1.** *Let  $H$  be as in (H), with  $h$  satisfying (M) and  $f$  satisfying ( $\varepsilon$ ), and let  $K \geq 1$ . Consider a domain  $\bar{D}_* = D_* \times \mathbb{R}^m \subset \bar{D}$  which is  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $\bar{h}$ , and given some parameter  $r > 0$ , assume that*

$$\varepsilon \leq \frac{\beta r}{2^{79} K}, \quad r \leq \frac{8\beta}{9MK}, \quad r \leq r_0, \quad (3.1)$$

and  $Ks_0 \geq 6$ . Then there exists a real-analytic symplectic embedding

$$\Phi : V_{\tilde{r}} \bar{D}_* \times V_{\tilde{s}_0} \mathbb{T}^{n+m} \rightarrow V_r \bar{D}_* \times V_{s_0} \mathbb{T}^{n+m}, \quad \tilde{r} := r/2, \quad \tilde{s}_0 := s_0/6,$$

such that

$$H \circ \Phi = \bar{h} + g + f_*$$

where  $\bar{h} + g$  is in  $\Lambda$ -resonant normal form with the estimates

$$|g + f_*|_{\tilde{r}, \tilde{s}_0} \leq 2\varepsilon, \quad |f_*|_{\tilde{r}, \tilde{s}_0} \leq e^{-Ks_0/6} \varepsilon \quad (3.2)$$

and

$$\sup_{(I, J, \theta, \varphi) \in V_{\tilde{r}} \bar{D}_* \times V_{\tilde{s}_0} \mathbb{T}^{n+m}} \|\Pi_{I, J} \Phi(I, J, \theta, \varphi) - (I, J)\| \leq \frac{18K\varepsilon}{\beta} \quad (3.3)$$

where  $\Pi_{I, J}$  denotes the projection onto the action space coordinates.

This statement is a direct consequence of the Normal Form Lemma of [27] (with the choice of the constants  $p = 9/8$  and  $q = 9$ ), to which we refer for a proof. More detailed estimates on  $g$  and on  $\Psi$  are available, but they will not be needed.

**3.2. Non-resonant stability estimates.** In the special case where the domain  $\bar{D}_*$  is  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $\bar{h}$ , with  $\Lambda = \{0\}$ , the normal form obtained in the previous section is, as we already said, integrable up to an exponentially small remainder. It is very easy to prove that in this case the action variables  $(I(t), J(t))$  remain stable for an exponentially long interval of time, and this does not require any convexity assumptions on  $h$ . Here's a precise statement.

**Proposition 3.2.** *Let  $H$  be as in (H), with  $h$  satisfying (M) and  $f$  satisfying  $(\varepsilon)$ , and let  $K \geq 1$ . Consider a domain  $\bar{D}_* = D_* \times \mathbb{R}^m \subset \bar{D}$  which is  $(\beta, K)$ -non resonant modulo  $\Lambda = \{0\}$  for  $\bar{h}$ , and given some parameter  $r > 0$ , assume that (3.1) is satisfied. Then for every solution with initial action  $(I_0, J_0) \in \bar{D}_*$  we have*

$$\|(I(t), J(t)) - (I_0, J_0)\| \leq r, \quad |t| \leq \frac{s_0 r}{5\varepsilon} e^{Ks_0/6}.$$

This is the content of Proposition 1 (Nonresonant stability estimate) of [27], to which we refer once again for a proof.

**3.3. Resonant stability estimates.** Next we study the case where the domain  $\bar{D}_*$  is  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $\bar{h}$ , with  $\Lambda$  non-trivial. The domain  $\bar{D}_* = D_* \times \mathbb{R}^m$  will be said to be  $\delta$ -close to  $\Lambda$ -resonances if  $D_*$  is  $\delta$ -close to  $\Lambda$ -resonances, as defined previously.

Assuming convexity of  $\bar{h}$ , one knows how to bound the variation of the action  $(I(t), J(t))$  using conservation of the energy and convexity arguments, as was first proved in [1] and later in [20]. However, in our situation  $\bar{h}$  is not convex, but it is convex with respect to the  $I$  variables and linear in the  $J$  variables, and we will prove below that this is sufficient to bound the variation of the action variables  $I(t)$ . This argument will just use the fact that our resonant normal form has an additional first integral; as a matter of fact, since resonances are associated to admissible submodules of rank at most  $n$ , our resonant normal form has at least  $m$  linearly independent first integrals. Using this observation, and the stability of the variables  $I(t)$ , we will obtain in turn the stability of the variables  $J(t)$  though with a worse radius of confinement.

**Proposition 3.3.** *Let  $H$  be as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$  and  $f$  satisfying  $(\varepsilon)$ , and let  $K \geq 1$ . Consider a domain  $\bar{D}_* = D_* \times \mathbb{R}^m \subset \bar{D}$  which is  $(\beta, K)$ -non resonant modulo  $\Lambda$  for  $\bar{h}$ , with  $\Lambda \in M_{K,d}^a$  non-trivial, but  $\delta$ -close to  $\Lambda$ -resonances. Given some parameter  $r > 0$ , assume that*

$$\varepsilon \leq \frac{\kappa r^2}{2^{10}}, \quad \delta = \frac{\kappa r}{8}, \quad r \leq \frac{8\beta}{9MK}, \quad r \leq r_0 \leq 24\|\alpha\|\kappa^{-1}. \quad (3.4)$$

*Then for every solution with initial action  $(I_0, J_0) \in \bar{D}_*$  we have*

$$\|I(t) - I_0\| \leq r, \quad |t| \leq \frac{\kappa s_0 r^2}{288\Omega\varepsilon} e^{Ks_0/6}$$

and

$$|J(t) - J_0|_\infty \leq K^d r, \quad |t| \leq \frac{\kappa s_0 r^2}{288 \Omega \varepsilon} e^{K s_0/6}.$$

*Proof.* First we assume that  $K s_0 \geq 6$ . Since  $\kappa \leq M$ , one easily check that (3.4) implies (3.1), and therefore Lemma 3.1 can be applied: there exists a real-analytic symplectic embedding

$$\Phi : V_{\tilde{r}, \tilde{s}_0} \bar{D}_* \rightarrow V_{r, s_0} \bar{D}_*$$

such that

$$H \circ \Phi = \bar{h} + g + f_*$$

where  $\bar{h} + g$  is in  $\Lambda$ -resonant normal form with the estimates (3.2) and (3.3). Now, from (3.3) and (3.4) we get

$$\|\Pi_{I, J} \Phi - \text{Id}\| \leq \frac{18 K \varepsilon}{\beta} \leq \frac{2^4 \varepsilon}{M r} \leq \frac{\kappa r}{2^6 M} = \frac{\delta}{8 M}.$$

Therefore the inverse image of  $\bar{D}_* \times \mathbb{T}^{n+m}$  by  $\Phi$  is contained in  $U_\rho \bar{D}_* \times \mathbb{T}^{n+m}$ , where  $\rho := \delta/(4M)$ . Recall that  $\tilde{r} = r/2$  and  $\tilde{s}_0 = s_0/6$ . We claim that for any initial action  $(\tilde{I}_0, \tilde{J}_0) \in U_\rho \bar{D}_*$ , the solution  $(\tilde{I}(t), \tilde{J}(t), \tilde{\theta}(t), \tilde{\varphi}(t))$  of the system associated to the Hamiltonian  $H \circ \Phi$  satisfies

$$\|\tilde{I}(t) - \tilde{I}_0\| \leq \tilde{r} - \rho, \quad |t| \leq \frac{\kappa \tilde{s}_0 \tilde{r}^2}{12 \Omega \varepsilon} e^{K s/6}$$

and

$$|\tilde{J}(t) - \tilde{J}_0|_\infty \leq 2 K^d (\tilde{r} - \rho), \quad |t| \leq \frac{\kappa \tilde{s}_0 \tilde{r}^2}{12 \Omega \varepsilon} e^{K s/6}.$$

Assuming this claim, for any initial action  $(I_0, J_0) \in \bar{D}_*$ , the solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to the Hamiltonian  $H$  satisfies

$$\begin{aligned} \|I(t) - I_0\| &\leq I(t) - \tilde{I}(t) + \|\tilde{I}(t) - \tilde{I}_0\| + \|\tilde{I}_0 - I_0\| \\ &\leq \rho/2 + (\tilde{r} - \rho) + \rho/2 \leq \tilde{r} \leq r \end{aligned}$$

and similarly, since  $K \geq 1$ ,

$$|J(t) - J_0|_\infty \leq \rho/2 + 2 K^d (\tilde{r} - \rho) + \rho/2 \leq 2 K^d \tilde{r} = K^d r$$

for times

$$|t| \leq \frac{\kappa s_0 r^2}{288 \Omega \varepsilon} e^{K s_0/6}$$

which is exactly the statement we want to prove. It is therefore sufficient to prove the above claim.

To simplify notations, let us drop the tildes and simply write  $(I_0, J_0) \in U_\rho \bar{D}_*$  and  $(I(t), J(t), \theta(t), \varphi(t))$  the associated solution. Let  $B$  be the ball of radius  $\tilde{r} - \rho$



around  $I_0$ , then  $B \times \mathbb{R}^m$  is contained in  $U_{\tilde{r}} \bar{D}_*$ . Let  $T_e$  be the positive time (possibly infinite) of first exit of  $(I(t), J(t))$  from  $B \times \mathbb{R}^m$ : it is then also the time of first exit of  $I(t)$  from  $B$ . Let also

$$T_* := \frac{\kappa \tilde{s}_0 \tilde{r}^2}{12\Omega\varepsilon} e^{Ks_0/6}, \quad T := \min\{T_e, T_*\}.$$

Furthermore, let us write

$$\begin{aligned} \Delta \bar{h} &:= \bar{h}(I(T), J(T)) - \bar{h}(I_0, J_0) \in \mathbb{R}, \\ \Delta(I, J) &:= (I(T), J(T)) - (I_0, J_0) = (I(T) - I_0, J(T) - J_0) \in \mathbb{R}^{n+m}, \\ \Delta I &:= I(T) - I_0 \in \mathbb{R}^n, \\ I(s) &:= I_0 + s\Delta I \in \mathbb{R}^n, \quad 0 \leq s \leq 1. \end{aligned}$$

By definition of  $\bar{h}$ , we have  $\nabla \bar{h}(I, J) = (\nabla h(I), \alpha) \in \mathbb{R}^{n+m}$  and

$$\nabla^2 \bar{h}(I, J) = \begin{pmatrix} \nabla^2 h(I) & 0 \\ 0 & 0 \end{pmatrix} \in M_{n+m}(\mathbb{R}), \quad \nabla^2 h(I) \in M_n(\mathbb{R}).$$

Using Taylor's formula with integral remainder at the point  $(I_0, J_0)$  and the special form of  $\nabla^2 \bar{h}$ , we get, letting  $\omega := \nabla h(I_0)$ ,

$$\Delta \bar{h} = (\omega, \alpha) \cdot \Delta(I, J) + \int_0^1 (1-s) \nabla^2 h(I(s)) \Delta I \cdot \Delta I \, ds.$$

Using the assumption  $(\kappa)$ , that is the convexity of  $h$ , we obtain

$$|\Delta \bar{h}| + |(\omega, \alpha) \cdot \Delta(I, J)| \geq \kappa/2 \|\Delta I\|^2. \quad (3.5)$$

By conservation of energy  $H$  and using the first part of (3.2) together with the first part of (3.4) and the definition of  $\tilde{r}$ , we have

$$|\Delta \bar{h}| \leq 2|g + f_*|_{\tilde{r}, \tilde{s}_0} \leq 4\varepsilon \leq \frac{\kappa r^2}{2^8} = \frac{\kappa \tilde{r}^2}{2^6}. \quad (3.6)$$

Then, by definition,  $(I_0, J_0) \in U_\rho \bar{D}_*$ , so there exist  $(I'_0, J'_0) \in \bar{D}_*$  which is  $\rho$ -close to  $(I_0, J_0)$ , but then by assumption  $\bar{D}_*$  is  $\delta$ -close to  $\Lambda$ -resonances, hence there exists  $\omega_* \in \mathbb{R}^n$  such that the vector  $(\omega_*, \alpha)$  belongs to the real subspace orthogonal to  $\Lambda$  and is  $\delta$ -close to  $(\nabla h(I'_0), \alpha)$ . Therefore, since  $h$  satisfies  $(M)$  and by definition of  $\rho$ , we obtain

$$\begin{aligned} \|\omega - \omega_*\| &= \|\nabla h(I_0) - \omega_*\| = \|\nabla h(I_0) - \nabla h(I'_0)\| + \|\nabla h(I'_0) - \omega_*\| \\ &\leq M\rho + \delta = 5\delta/4. \end{aligned} \quad (3.7)$$



Writing

$$\begin{aligned} (\omega, \alpha) \cdot \Delta(I, J) &= (\omega, \alpha) \cdot \Delta(I, J) - (\omega_*, \alpha) \cdot \Delta(I, J) + (\omega_*, \alpha) \cdot \Delta(I, J) \\ &= (\omega - \omega_*, 0) \cdot \Delta(I, J) + (\omega_*, \alpha) \cdot \Delta(I, J) \\ &= (\omega - \omega_*) \cdot \Delta I + (\omega_*, \alpha) \cdot \Delta(I, J) \end{aligned}$$

we obtain

$$|(\omega, \alpha) \cdot \Delta(I, J)| \leq |(\omega - \omega_*) \cdot \Delta I| + |(\omega_*, \alpha) \cdot \Delta(I, J)|. \quad (3.8)$$

Using (3.7), the second part of (3.4) and the definition of  $\tilde{r}$ , we can bound the first summand by

$$\begin{aligned} |(\omega - \omega_*) \cdot \Delta I| &\leq \|\omega - \omega_*\| \|\Delta I\| \leq (5\delta/4) \|\Delta I\| \\ &\leq \frac{5\delta^2}{2\kappa} + \frac{\kappa}{6} \|\Delta I\|^2 = \frac{5\kappa\tilde{r}^2}{32} + \frac{\kappa}{6} \|\Delta I\|^2. \end{aligned} \quad (3.9)$$

For the second summand, using the fact that  $(\omega_*, \alpha)$  belongs to the real subspace orthogonal to  $\Lambda$ , and that  $\bar{h} + g$  is in  $\Lambda$ -resonant normal form, we have

$$\begin{aligned} |(\omega_*, \alpha) \cdot \Delta(I, J)| &\leq \int_0^T |(\omega_*, \alpha) \cdot \partial_{\theta, \varphi} f_*(I(t), J(t), \theta(t), \varphi(t))| dt \\ &\leq T \|(\omega_*, \alpha)\| \sup_{(I, J, \theta, \varphi) \in B \times \mathbb{R}^n \times \mathbb{T}^{n+m}} \|\partial_{\theta, \varphi} f_*(I, J, \theta, \varphi)\|. \end{aligned}$$

Now using a Cauchy estimate and the second part of (3.2) we get

$$\|\partial_{\theta, \varphi} f_*(I, J, \theta, \varphi)\| \leq \frac{1}{e\tilde{s}_0} |f_*|_{\tilde{r}, \tilde{s}} \leq \frac{\varepsilon}{e\tilde{s}_0} e^{-Ks_0/6}.$$

Moreover, as  $T \leq T_*$  and  $\|(\omega_*, \alpha)\| \leq \Omega$  since  $h$  satisfies  $(\Omega)$ , we get

$$|(\omega_*, \alpha) \cdot \Delta(I, J)| \leq T_* \Omega \frac{\varepsilon}{e\tilde{s}_0} e^{-Ks_0/6} = \frac{\kappa\tilde{r}^2}{12e}. \quad (3.10)$$

Putting together (3.5), (3.6), (3.8), (3.9) and (3.10) we eventually arrive at

$$\kappa/3 \|\Delta I\|^2 \leq \frac{\kappa\tilde{r}^2}{64} + \frac{5\kappa\tilde{r}^2}{32} + \frac{\kappa\tilde{r}^2}{12e} \leq \frac{\kappa\tilde{r}^2}{4}.$$

Now  $\rho = \delta/(4M) = \kappa r/(32M) = \kappa\tilde{r}/(16M) \leq \tilde{r}/16$  as  $\kappa/M \leq 1$ , thus  $\tilde{r} - \rho \geq 15\tilde{r}/16$  and in particular  $\tilde{r}^2 < 4(\tilde{r} - \rho)^2/3$ . This, together with the last inequality, implies that

$$\kappa/3 \|\Delta I\|^2 < \kappa/3 (\tilde{r} - \rho)^2$$

and therefore  $\|\Delta I\| < \tilde{r} - \rho$ . This eventually proves that  $T = T_*$ , that is

$$\|\tilde{I}(t) - \tilde{I}_0\| \leq \tilde{r} - \rho, \quad 0 \leq t \leq T_* = \frac{\kappa \tilde{s}_0 \tilde{r}^2}{12\Omega_\varepsilon} e^{Ks/6}.$$

The same argument yields the same result for negative times  $-T_* \leq t \leq 0$ , and hence

$$\|\tilde{I}(t) - \tilde{I}_0\| \leq \tilde{r} - \rho, \quad |t| \leq \frac{\kappa \tilde{s}_0 \tilde{r}^2}{12\Omega_\varepsilon} e^{Ks/6}$$

which was the first part of the claim that needed to be proved.

For the second part of the claim, let us write

$$\Delta J := J(T) - J_0 = J(T_*) - J_0 \in \mathbb{R}^m$$

so that in particular

$$\Delta(I, J) = (\Delta I, \Delta J) \in \mathbb{R}^{n+m}.$$

Observe that since  $\Lambda$  is admissible of rank  $1 \leq d \leq n$ , the real subspace in  $\mathbb{R}^{n+m}$  orthogonal to  $\Lambda$  has dimension  $n + m - d \geq m$ , and given any vector  $v$  in the orthogonal of  $\Lambda$ , proceeding exactly as in (3.10), one has the inequality

$$|v \cdot \Delta(I, J)| \leq \frac{\kappa \tilde{r}^2 \|v\|}{12\Omega e} \leq \frac{\kappa \tilde{r}^2 \|v\|}{12\|\alpha\|e} \quad (3.11)$$

where the second inequality follows from  $\Omega \geq \|\alpha\|$ . We will apply this to  $m$  properly chosen such vectors  $v$ . Let  $(k^1, l^1), \dots, (k^d, l^d) \in \mathbb{Z}^{n+m}$  a basis for  $\Lambda$ , and we write

$$k^j = (k_1^j, \dots, k_n^j) \in \mathbb{Z}^n, \quad l^j = (l_1^j, \dots, l_m^j) \in \mathbb{Z}^m, \quad 1 \leq j \leq d.$$

Since  $\Lambda$  is admissible, the  $d$  vectors  $k^j \in \mathbb{Z}^n$ , for  $1 \leq j \leq d$ , are linearly independent, and hence, without loss of generality, we may assume that the  $d$  vectors in  $\mathbb{Z}^d$

$$k_j = (k_1^1, \dots, k_d^1) \in \mathbb{Z}^d, \quad 1 \leq j \leq d$$

are linearly independent. Let us further define

$$l_i = (l_i^1, \dots, l_i^d) \in \mathbb{Z}^d, \quad 1 \leq i \leq m.$$

We set  $D$  to be the determinant of the square matrix  $M$  of size  $d$  whose columns are given by  $k_j$ ,  $1 \leq j \leq d$ , and for any  $1 \leq j \leq d$  and any  $1 \leq i \leq m$ , we let  $D_{i,j}$  be the determinant of the square matrix  $M_{i,j}$  obtained by replacing the vector  $k_j$  (that is, the  $j^{\text{th}}$  column of  $M$ ) by the vector  $-l_i$ . If we denote by  $e_1, \dots, e_n, e_{n+1}, \dots, e_{n+m}$  the vectors of the canonical basis of  $\mathbb{R}^{n+m}$ , we eventually define, for  $1 \leq i \leq m$ , the vector

$$v_i := D_{i,1}e_1 + D_{i,2}e_2 + \dots + D_{i,d}e_d + De_{n+i} \in \mathbb{R}^{n+m}.$$

It is now an easy matter to check that for any  $1 \leq j \leq d$  and any  $1 \leq i \leq m$ , we have

$$(k^j, l^j) \cdot v_i = 0$$

that is the vectors  $v_i$  belongs to the real subspace orthogonal to  $\Lambda$ . Therefore the inequality (3.11) gives

$$|v_i \cdot \Delta(I, J)| \leq \frac{\kappa \tilde{r}^2 \|v_i\|}{12 \|\alpha\|_e}, \quad 1 \leq i \leq m. \quad (3.12)$$

If we write

$$\Delta(I, J) = (\Delta I, \Delta J) = (\Delta I_1, \dots, \Delta I_n, \Delta J_1, \dots, \Delta J_m)$$

then (3.12) can be written again as

$$|D_{i,1} \Delta I_1 + D_{i,2} \Delta I_2 + \dots + D_{i,d} \Delta I_d + D \Delta J_i| \leq \frac{\kappa \tilde{r}^2 \|v_i\|}{12 \|\alpha\|_e}, \quad 1 \leq i \leq m$$

and as consequence, for  $1 \leq i \leq m$

$$\begin{aligned} |D \Delta J_i| &\leq |D_{i,1} \Delta I_1 + D_{i,2} \Delta I_2 + \dots + D_{i,d} \Delta I_d| + \frac{\kappa \tilde{r}^2 \|v_i\|}{12 \|\alpha\|_e} \\ &\leq \|D_i\| \|\Delta I\| + \frac{\kappa \tilde{r}^2 \|v_i\|}{12 \|\alpha\|_e} \end{aligned}$$

where  $D_i = (D_{i,1}, \dots, D_{i,d}) \in \mathbb{R}^d$ . Recalling that  $|\Lambda|$  denotes the co-volume of the lattice  $\Lambda$ , we can use Cauchy–Binet formula and Hadamard's inequality to bound

$$\|D_i\| \leq \|v_i\| \leq |\Lambda| \leq K^d$$

and therefore

$$|D \Delta J_i| \leq K^d \left( \|\Delta I\| + \frac{\kappa \tilde{r}^2}{12 \|\alpha\|_e} \right) \leq K^d \left( \tilde{r} - \rho + \frac{\kappa \tilde{r}^2}{12 \|\alpha\|_e} \right).$$

Then, since  $\tilde{r} = r/2 \leq r_0/2 \leq 12 \|\alpha\| \kappa^{-1}$ , we have

$$\frac{\kappa \tilde{r}^2}{12 \|\alpha\|_e} \leq \frac{\tilde{r}}{e} \leq \frac{15\tilde{r}}{16} \leq \tilde{r} - \rho$$

and thus

$$|D \Delta J_i| \leq 2K^d (\tilde{r} - \rho)$$

which proves, in particular, that

$$|\Delta J|_\infty = \sup_{1 \leq i \leq m} |\Delta J_i| \leq 2K^d (\tilde{r} - \rho)$$

which was the second part of the claim that needed to be proved.

This ends the proof under the assumption  $Ks_0 \geq 6$ . But if  $Ks_0 < 6$ , the exact same argument applies to the original Hamiltonian  $H$ , by setting  $g = 0$  and  $f_* = f$ , and this concludes the proof.  $\square$

#### 4. Proof of the main result

This section is devoted to the proof of our main result, Theorem 1.2, which will be easily obtained using Proposition 2.2, Proposition 3.2 and Proposition 3.3.

*Proof of Theorem 1.2.* Recall that we are considering  $H$  as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$ ,  $f$  satisfying  $(\varepsilon)$ , and  $\alpha$  satisfying  $(\text{Dio}_{\gamma,\tau})$ . Recall also that we have defined the positive constants  $a$  and  $b$  by

$$a = \frac{1}{2(n+1)(\tau+1)}, \quad b = \frac{(n+1)\tau+1}{2(n+1)(\tau+1)}$$

and that we are assuming that

$$\kappa r_0 \leq \min\{80, 24\|\alpha\|\}. \quad (4.1)$$

For a parameter  $K \geq 1$  to be chosen below, using (4.1) we can apply Proposition 2.2 to obtain a covering of  $\bar{D} = D \times \mathbb{R}^m$  by subsets  $\bar{D}_\Lambda = D_\Lambda \times \mathbb{R}^m$ , where  $\Lambda \in M_K^a$ , such that each  $\bar{D}_\Lambda$  is  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$  for  $\bar{h}$  and, for  $\Lambda \neq \{0\}$ ,  $\delta_\Lambda$ -close to  $\Lambda$ -resonances for  $\bar{h}$  with

$$\beta_\Lambda = \frac{9MKr_\Lambda}{8}, \quad \delta_\Lambda = \frac{\kappa r_\Lambda}{8}, \quad (4.2)$$

where

$$r_\Lambda := \frac{r_0\gamma}{|\tilde{\Lambda}| F^{n-d+1} (n+1)^\tau K^{(n+1)\tau+n-d+1}}, \quad d = \text{rank } \Lambda. \quad (4.3)$$

Then, given any  $\Lambda \in M_K^a$ , we can apply either Proposition 3.2 (for  $\Lambda = \{0\}$ ) or Proposition 3.3 (for  $\Lambda \neq \{0\}$ ), with  $r = r_\Lambda$ ,  $\beta = \beta_\Lambda$  and  $\delta = \delta_\Lambda$  provided that (3.1) and (3.4) are satisfied. The inequalities (3.1) are easily seen to be implied by (3.4) as we already pointed out, hence we only need to verify (3.4), and in view of our definitions of  $r_\Lambda$ ,  $\beta_\Lambda$  and  $\delta_\Lambda$ , and using also (4.1), the latter reduces to

$$\varepsilon \leq \frac{\kappa r_\Lambda^2}{2^{10}}, \quad r_\Lambda \leq r_0. \quad (4.4)$$

Given any  $\Lambda \in M_K^a$ , we have

$$\frac{r_0\gamma}{F^{n+1}(n+1)^\tau K^{(n+1)(\tau+1)}} \leq r_\Lambda \leq \frac{r_0\gamma}{F(n+1)^\tau K^{(n+1)\tau+1}},$$

as  $F \geq 1$  and  $K \geq 1$ . Hence (4.4) is satisfied, for any  $\Lambda \in M_K^a$ , if

$$\varepsilon \leq \frac{\kappa r_0^2 \gamma^2}{2^{10} F^{2(n+1)} (n+1)^{2\tau} K^{2(n+1)(\tau+1)}}, \quad K \geq \left( \frac{\gamma}{F(n+1)^\tau} \right)^{\frac{a}{b}}. \quad (4.5)$$

Let us define

$$\varepsilon_0 := \frac{\kappa r_0^2 \gamma^2}{2^{10} F^{2(n+1)} (n+1)^{2\tau}}, \quad \varepsilon_* := \varepsilon_0 \left( \frac{F(n+1)^\tau}{\gamma} \right)^{\frac{1}{b}}$$

and

$$K := \left( \frac{\varepsilon_0}{\varepsilon} \right)^a,$$

then (4.5) eventually reduces to

$$\varepsilon \leq \min\{\varepsilon_0, \varepsilon_*\}. \quad (4.6)$$

Under this choice of  $K$  and this smallness assumption on  $\varepsilon$ , Proposition 3.2 and Proposition 3.3 apply and in the resonant case, we obtain

$$\|I(t) - I_0\| \leq \frac{r_0 \gamma}{F(n+1)^\tau K^{(n+1)\tau+1}} \leq \frac{r_0 \gamma}{F(n+1)^\tau} \left( \frac{\varepsilon}{\varepsilon_0} \right)^b$$

up to times

$$|t| \leq T_\Lambda \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^a\right), \quad T_\Lambda := \frac{\kappa s_0 r_\Lambda^2}{288 \Omega \varepsilon},$$

while in the non-resonant case, we obtain

$$\begin{aligned} \|I(t) - I_0\| &\leq \|(I(t), J(t)) - (I_0, J_0)\| \\ &\leq \frac{r_0 \gamma}{F(n+1)^\tau K^{(n+1)\tau+1}} \leq \frac{r_0 \gamma}{F(n+1)^\tau} \left( \frac{\varepsilon}{\varepsilon_0} \right)^b \end{aligned}$$

up to times

$$|t| \leq T_{\{0\}} \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^a\right), \quad T_{\{0\}} := \frac{s_0 r_{\{0\}}}{5\varepsilon}.$$

To obtain a uniform time estimate, observe that for any  $\Lambda \in M_K^a$  we have

$$T_\Lambda \geq \frac{2^{10} s_0}{288 \Omega} \geq \frac{3s_0}{\Omega}.$$

Letting

$$R_* = \frac{r_0 \gamma}{F(n+1)^\tau}, \quad T_* = \frac{3s_0}{\Omega},$$

we have just proved that

$$\|I(t) - I_0\| \leq R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^a\right),$$

provided (4.6) is satisfied, with  $\varepsilon_*$  that can be written as

$$\varepsilon_* = \varepsilon_0 \left( \frac{r_0}{R_*} \right)^{\frac{1}{b}}.$$

Recalling that  $F = 10M/\kappa$ , this was exactly the first part of the statement to be proved.

For the second part of the statement, we distinguish the case  $m = 1$  and the case  $m \geq 2$ . In the first case, recall that  $J(t) \in \mathbb{R}$  and  $\alpha \in \mathbb{R}$ . We have

$$|h(I(t)) - h(I_0)| \leq \Omega \|I(t) - I_0\| \leq \Omega R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b$$

while, using preservation of energy and our thresholds,

$$\begin{aligned} |h(I(t)) + \alpha J(t) - h(I_0) + \alpha J_0| &\leq \varepsilon \leq R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b \frac{\kappa r_0}{2^{10}} \\ &\leq R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b \frac{24|\alpha|}{2^{10}} \leq R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b |\alpha|. \end{aligned}$$

Putting the last two estimates together we obtain

$$|J(t) - J_0| \leq |\alpha|^{-1} |\alpha J(t) - \alpha J_0| \leq (\Omega/|\alpha| + 1) R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^b.$$

In the second case, observe that on  $\bar{D}_\Lambda$ , if  $\Lambda = \{0\}$ , Proposition 3.2 gives

$$|J(t) - J_0|_\infty \leq \|(I(t), J(t)) - (I_0, J_0)\| \leq r_{\{0\}}, \quad |t| \leq \frac{s_0 r}{5\varepsilon} e^{Ks_0/6}$$

whereas, if  $\Lambda$  has rank  $1 \leq d \leq n$ , Proposition 3.3 gives

$$|J(t) - J_0|_\infty \leq K^d r_\Lambda, \quad |t| \leq \frac{\kappa s_0 r^2}{288\Omega\varepsilon} e^{Ks_0/6}.$$

Therefore, along any solution we have

$$|J(t) - J_0|_\infty \leq R_* K^n \left( \frac{\varepsilon}{\varepsilon_0} \right)^b = R_* \left( \frac{\varepsilon}{\varepsilon_0} \right)^{b-na}, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_0}{\varepsilon} \right)^a \right)$$

which is the statement we wanted to prove since  $b^* = b - na$ . This proves the second part of the statement, and finished the proof.  $\square$

## 5. Improved stability close to resonances

For solutions starting close to resonances, we can obtain a better result. Consider a fixed submodule  $L$  of  $\mathbb{Z}^{n+m}$  of rank  $d$ , which is assumed to be admissible and maximal, and let  $K_L \geq 1$  such that  $L$  is a  $K_L$ -submodule. Recall that  $\tilde{L}$  denotes the

projection of  $L$  onto  $\mathbb{R}^n$ , and  $|\tilde{L}|$  denotes the co-volume of  $\tilde{L}$ . To such a  $L$  the space of  $L$ -resonances is

$$R_L = \{\omega \in \mathbb{R}^n \mid (k, l) \cdot (\omega, \alpha) = 0, \forall (k, l) \in L\}$$

and let

$$S_L := \{(I, J) \in \bar{D} \mid \nabla h(I) \in R_L\} = \{I \in D \mid \nabla h(I) \in R_L\} \times \mathbb{R}^m$$

the resonant domain in action space. Solutions with initial action close to  $S_L$  satisfy better stability estimates, as stated in the theorem below.

**Theorem 5.1.** *Let  $H$  be as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , and let us define*

$$a(d) = \frac{1}{2((n+1)\tau + n + 1 - d)}, \quad b(d) = \frac{(n+1)\tau + 1}{2((n+1)\tau + n + 1 - d)},$$

$$b^*(d) = \frac{(n+1)\tau + 1 - n}{2((n+1)\tau + n + 1 - d)}, \quad R_*(d) = \left(\frac{\kappa}{10M}\right)^{n+1-d} \frac{r_0\gamma}{(n+1)^\tau},$$

and, given a submodule  $L$  as above, let

$$\varepsilon_0(L) = \frac{1}{|\tilde{L}|^2} \frac{\kappa r_0^2 \gamma^2}{2^{10}(n+1)^{2\tau}} \left(\frac{\kappa}{10M}\right)^{2(n+1-d)}, \quad \varepsilon_*(L) = \varepsilon_0(L) \left(\frac{r_0}{R_*(d)}\right)^{\frac{1}{b(d)}}$$

$$\varepsilon_{**}(L) := \varepsilon_0(L) K_L^{-\frac{1}{a(d)}}.$$

If  $\kappa r_0 \leq \min\{80, 24\|\alpha\|\}$  and  $\varepsilon \leq \min\{\varepsilon_*(L), \varepsilon_{**}(L)\}$ , for any solution

$$(I(t), J(t), \theta(t), \varphi(t))$$

of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in U_\rho S_L \times \mathbb{T}^{n+m}$ , with  $\rho = 4M^{-1}\sqrt{\kappa\varepsilon}$ , we have

$$\|I(t) - I_0\| \leq R_*(d) \left(\frac{\varepsilon}{\varepsilon_0(L)}\right)^{b(d)}, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0(L)}{\varepsilon}\right)^{a(d)}\right)$$

with  $T_*$  as in Theorem 1.2. Moreover, in the case where  $m = 1$  and hence  $\tau = 0$ , we have

$$|J(t) - J_0| \leq (\Omega/|\alpha| + 1) R_* \left(\frac{\varepsilon}{\varepsilon_0}\right)^{b(d)}, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^{a(d)}\right)$$

whereas in the case  $m \geq 2$  and hence  $\tau \geq 1$ , we have

$$|J(t) - J_0|_\infty \leq R_* \left(\frac{\varepsilon}{\varepsilon_0}\right)^{b^*(d)}, \quad |t| \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0}{\varepsilon}\right)^{a(d)}\right).$$

Observe that in the special case where  $L = \{0\}$ ,  $S_L = D \times \mathbb{R}^m$ ,  $|\tilde{L}| = K_L = 1$  and therefore the above statement exactly reduces to Theorem 1.2.

*Proof.* Consider the subset

$$M_K^a(L) := \{\Lambda \in M_K^a \mid \Lambda \supseteq L\}$$

and assume, for any  $\Lambda \in M_K^a(L)$ ,

$$\varepsilon \leq \frac{\kappa r_\Lambda^2}{2^{10}}, \quad r_\Lambda \leq r_0, \quad K \geq K_L \quad (5.1)$$

where the last inequality is to ensure that  $M_K^a(L)$  is actually non empty.

The resonant blocks  $B_\Lambda$ , for  $\Lambda \in M_K^a(L)$ , cover the resonant zone  $Z_L$  and hence their pullbacks cover  $U_{\rho'} S_L$ , with

$$\rho' = \frac{\delta_L}{M} = \frac{\kappa r_L}{8M} \geq \frac{\sqrt{2^{10}}}{8M} \sqrt{\kappa \varepsilon} = 4M^{-1} \sqrt{\kappa \varepsilon} = \rho.$$

Moreover, for any  $\Lambda \in M_K^a(L)$ , we have

$$\frac{r_0 \gamma}{|\tilde{L}| F^{n-d+1} (n+1)^\tau K^{(n+1)\tau+n-d+1}} = r_L \leq r_\Lambda \leq \frac{r_0 \gamma}{F(n+1)^\tau K^{(n+1)\tau+1}}.$$

Using the above inequalities and proceeding exactly as in the proof of Theorem 1.2, we can define

$$K := \left( \frac{\varepsilon_0(L)}{\varepsilon} \right)^{a(d)}$$

and verify that (5.1) is implied by

$$\varepsilon \leq \varepsilon_*(L), \quad \varepsilon \leq \varepsilon_{**}(L) \leq \varepsilon_0(L).$$

The stability estimates apply uniformly to all blocks  $B_\Lambda$ , for  $\Lambda \in M_K^a(L)$ , and one easily check that the statement follows with the given constants.  $\square$

Using Theorem 5.1, we will be able to show, as in [10], how to slightly increase the value of the exponent  $a$  by greatly decreasing the value of the exponent  $b$  and loosing control on the evolution of the  $J$  variables in the case  $m \geq 2$ . Let us state precisely the results.

**Theorem 5.2.** *Let  $H$  be as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  satisfies  $(\text{Dio}_{\gamma,\tau})$ , and let us define*

$$a = \frac{1}{2(n+1)(\tau+1)}, \quad a(1) = \frac{1}{2((n+1)\tau+n)}, \quad b(1) = \frac{(n+1)\tau+1}{2((n+1)\tau+n)}$$



and

$$\varepsilon_1 := \frac{\kappa r_0^2 \gamma^2}{2^{12}(n+1)^{2\tau}} \left( \frac{\kappa}{10M} \right)^{2n}, \quad C := \frac{2\sqrt{n}\Omega}{\kappa},$$

$$R_*(1) = \frac{\kappa r_0 \gamma}{10M(n+1)^\tau}, \quad T_* = \frac{3s_0}{\Omega}.$$

Given any  $0 < \delta \leq a$ , assume that

$$\left( \frac{\varepsilon}{\varepsilon_1} \right)^\delta \leq \frac{2C}{r_0}, \quad \left( \frac{\varepsilon}{\varepsilon_1} \right)^{1-2\delta} \leq \left( \frac{r_0}{2R_*(1)} \right)^{\frac{1}{b(1)}}.$$

If  $\kappa r_0 \leq 80$ , then for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D} \times \mathbb{T}^{n+m}$ , we have

$$\|I(t) - I_0\| \leq (C + R_*(1)) \left( \frac{\varepsilon}{\varepsilon_1} \right)^\delta, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_1}{\varepsilon} \right)^{a(1)(1-2\delta)} \right)$$

and, in the case where  $m = 1$  and hence  $\tau = 0$ ,

$$|J(t) - J_0| \leq (\Omega/|\alpha| + 1)(C + R_*(1)) \left( \frac{\varepsilon}{\varepsilon_1} \right)^\delta, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_1}{\varepsilon} \right)^{a(1)(1-2\delta)} \right).$$

Moreover, if  $\kappa r_0 \leq 80$  and we assume instead that

$$\frac{\varepsilon}{\varepsilon_1} \leq \left( \frac{r_0}{2C} \right)^2 \left( \frac{r_0}{2R_*(1)} \right)^{\frac{1}{b(1)}}, \quad \frac{\varepsilon}{\varepsilon_1} \leq \left( \frac{r_0}{2C} \right)^{\frac{1}{a}}$$

then, for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D} \times \mathbb{T}^{n+m}$ , we have

$$\|I(t) - I_0\| \leq r_0, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{r_0}{2C} \right)^{2a(1)} \left( \frac{\varepsilon_1}{\varepsilon} \right)^{a(1)} \right)$$

and, in the case where  $m = 1$  and hence  $\tau = 0$ ,

$$|J(t) - J_0| \leq (\Omega/|\alpha| + 1)r_0, \quad |t| \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{r_0}{2C} \right)^{2a(1)} \left( \frac{\varepsilon_1}{\varepsilon} \right)^{a(1)} \right).$$

Let us first remark that the assumption  $\kappa r_0 \leq 24\|\alpha\|$ , which was only used to control the evolution of the  $J$  variables in the case  $m \geq 2$ , is not needed here since no control on the  $J$  variables can be obtained in this case.

In the special case  $m = 1$ ,  $\tau = 0$ , Theorem 5.2 generalizes the main result of [10]. For  $\delta > 0$  but very small, the exponent  $a(1)(1 - 2\delta)$  is very close to  $a(1)$ , which is better than  $a$  but at the same time the radius of confinement greatly deteriorates as it gets close to one. At the limit  $\delta = 0$  (the second part of the statement), we

do obtain a confinement of order one with a better time of stability given by the exponent  $a(1)$ . At the other extreme  $\delta = a$ , observe that  $a(1)(1 - 2\delta) = a$  and therefore Theorem 5.2 gives a time of stability which is equivalent to 1.2, but with a worse radius of confinement given by the exponent  $a$  instead of  $b$ .

*Proof of Theorem 5.2.* Consider an arbitrary solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D} \times \mathbb{T}^{n+m}$ . It will be sufficient to prove the statement for positive time; the exact same argument will then prove the statement for negative time. Consider a parameter  $K \geq 1$  (to be chosen) below that satisfies

$$K \geq 2r_0^{-1}. \quad (5.2)$$

Then we have the following dichotomy: either

$$\|I(t) - I_0\| < K^{-1}$$

as long as the solution is defined, or there exists  $t' > 0$  such that

$$\|I(t') - I_0\| = K^{-1}.$$

In the first case, the solution  $(I(t), J(t), \theta(t), \varphi(t))$  is then defined for all time, and as a consequence

$$\|I(t) - I_0\| < K^{-1}, \quad t \in \mathbb{R}. \quad (5.3)$$

In the second case, using  $(\kappa)$  we obtain

$$\|\nabla h(I(t')) - \nabla h(I_0)\| \geq \kappa \|I(t') - I_0\| = \kappa K^{-1}$$

and therefore

$$|\nabla h(I(t')) - \nabla h(I_0)|_\infty \geq \kappa(\sqrt{n}K)^{-1}.$$

If we denote  $\nabla h(I) = (\nabla_1 h(I), \dots, \nabla_n h(I)) \in \mathbb{R}^n$ , let  $1 \leq i \leq n$  be such that

$$|\nabla_i h(I(t')) - \nabla_i h(I_0)| = |\nabla h(I(t')) - \nabla h(I_0)|_\infty \geq \kappa(\sqrt{n}K)^{-1} \quad (5.4)$$

and also let  $1 \leq j \leq m$  be such that

$$|\alpha|_\infty = |\alpha_j| \neq 0.$$

Clearly, any closed interval in  $\mathbb{R}$  of length larger than  $l > 0$  contains an irreducible rational  $p/q$ , with a denominator  $q > 0$  bounded by  $l^{-1}$ ; therefore any closed interval of length  $l|\alpha_j|$  contains a real number of the form  $|\alpha_j|p/q$ , with an irreducible rational  $p/q$  with a denominator  $q$  bounded by  $l^{-1}$ . Applying this to the interval  $[\nabla_i h(I_0), \nabla_i h(I(t'))]$ ,  $0 \leq t \leq t'$ , and using (5.4), there exists a time  $0 \leq t^* \leq t'$  and an irreducible rational  $p/q \in \mathbb{R}$  such that

$$\nabla_i h(I(t^*)) = |\alpha_j|p/q, \quad q \leq \frac{\sqrt{n}K|\alpha_j|}{\kappa}.$$

Moreover, using  $(\Omega)$  we have

$$|\nabla_i h(I(t^*))| = |\alpha_j| |p|/q \leq \Omega$$

and therefore

$$\begin{aligned} \sqrt{q^2 + p^2} &\leq q \sqrt{1 + |\alpha_j|^{-2} \Omega^2} \leq q(1 + |\alpha_j|^{-1} \Omega) \\ &\leq \left( \frac{\sqrt{n} |\alpha_j|}{\kappa} + \frac{\sqrt{n} \Omega}{\kappa} \right) K \leq \frac{2\sqrt{n} \Omega}{\kappa} K. \end{aligned} \quad (5.5)$$

At the time  $t = t^*$  the frequency vector  $\nabla \bar{h}(I(t^*), J(t^*)) = (\nabla h(I(t^*)), \alpha) \in \mathbb{R}^{n+m}$  satisfies a resonance relation; indeed, assuming for instance  $|\alpha_j| = \alpha_j$  then for  $k = qe_i - pe_j \in \mathbb{Z}^{n+m}$  one has

$$k \cdot \nabla \bar{h}(I(t^*), J(t^*)) = q \nabla_i h(I(t^*)) - p \alpha_j = \alpha_j p - p \alpha_j = 0$$

(if  $|\alpha_j| = -\alpha_j$ , one simply replace  $qe_i - pe_j$  by  $qe_i + pe_j$ ). Letting  $L$  be the one-dimensional lattice generated by  $k$  and recalling the notations of Theorem 5.1, one has  $(I(t^*), J(t^*)) \in S_L$ , and from (5.5)

$$K_L \leq \frac{2\sqrt{n} \Omega}{\kappa} K := CK, \quad |\tilde{L}| \leq q \leq \frac{\sqrt{n} \Omega}{\kappa} \leq CK. \quad (5.6)$$

To apply Theorem 5.1 in the case  $d = 1$  with  $r_0/2$  instead of  $r_0$ , let us estimate the thresholds (replacing  $r_0$  by  $r_0/2$ ): setting

$$\varepsilon_1 := \frac{\kappa r_0^2 \gamma^2}{2^{12} (n+1)^{2\tau}} \left( \frac{\kappa}{10M} \right)^{2n}$$

we have

$$\begin{aligned} \varepsilon_0(L) &= \varepsilon_1 |\tilde{L}|^{-2} \geq \varepsilon_1 (CK)^{-2} \\ \varepsilon_*(L) &= \varepsilon_0(L) \left( \frac{r_0}{2R_*(1)} \right)^{\frac{1}{b(1)}} \geq \varepsilon_1 (CK)^{-2} \left( \frac{r_0}{2R_*(1)} \right)^{\frac{1}{b(1)}} \\ \varepsilon_{**}(L) &= \varepsilon_0(L) K_L^{-\frac{1}{a(1)}} \geq \varepsilon_1 (CK)^{-\frac{2a(1)+1}{a(1)}} = \varepsilon_1 (CK)^{-\frac{1}{a}} \end{aligned}$$

and consequently Theorem 5.1 can be applied provided

$$\varepsilon \leq \varepsilon_1 (CK)^{-2} \left( \frac{r_0}{2R_*(1)} \right)^{\frac{1}{b(1)}} \leq \varepsilon_*(L), \quad \varepsilon \leq \varepsilon_1 (CK)^{-\frac{1}{a}} \leq \varepsilon_{**}(L). \quad (5.7)$$

To prove the first part of the statement, we choose

$$K := \frac{1}{C} \left( \frac{\varepsilon_1}{\varepsilon} \right)^\delta, \quad 0 < \delta \leq a. \quad (5.8)$$

Since  $\delta \leq a$ , the second part of (5.7) is, by definition, satisfied whereas since  $\delta > 0$  and  $\delta < 1/2$ , the condition (5.2) and the first part of (5.7) are satisfied provided

$$\left(\frac{\varepsilon}{\varepsilon_1}\right)^\delta \leq \frac{2C}{r_0}, \quad \left(\frac{\varepsilon}{\varepsilon_1}\right)^{1-2\delta} \leq \left(\frac{r_0}{2R_*(1)}\right)^{\frac{1}{b(1)}}. \quad (5.9)$$

Under these conditions, we obtain

$$\|I(t) - I(t^*)\| \leq R_*(1) \left(\frac{\varepsilon}{\varepsilon_0(L)}\right)^{b(1)}, \quad t^* \leq t \leq t^* + T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0(L)}{\varepsilon}\right)^{a(1)}\right)$$

with  $T_*$  as in Theorem 1.2. But then

$$R_*(1) \left(\frac{\varepsilon}{\varepsilon_0(L)}\right)^{b(1)} \leq R_*(1) \left(\frac{\varepsilon}{\varepsilon_1}\right)^{b(1)(1-2\delta)}$$

and

$$T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_0(L)}{\varepsilon}\right)^{a(1)}\right) \geq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_1}{\varepsilon}\right)^{a(1)(1-2\delta)}\right)$$

and therefore

$$\|I(t) - I(t^*)\| \leq R_*(1) \left(\frac{\varepsilon}{\varepsilon_1}\right)^{b(1)(1-2\delta)}, \quad t^* \leq t \leq t^* + T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_1}{\varepsilon}\right)^{a(1)(1-2\delta)}\right).$$

But recalling that  $t^* \leq t'$ , we also have

$$\|I(t^*) - I_0\| \leq K^{-1} = C \left(\frac{\varepsilon}{\varepsilon_1}\right)^\delta, \quad 0 \leq t \leq t^*$$

and since  $a(1) \leq b(1)$ ,

$$\delta \leq a = \frac{a(1)}{2a(1) + 1} \leq \frac{b(1)}{2a(1) + 1} \leq (1 - 2\delta)b(1)$$

and as a consequence, we do have

$$\begin{aligned} \|I(t) - I_0\| &\leq \|I(t) - I(t^*)\| + \|I(t^*) - I_0\| \leq C \left(\frac{\varepsilon}{\varepsilon_1}\right)^\delta + R_*(1) \left(\frac{\varepsilon}{\varepsilon_1}\right)^{b(1)(1-2\delta)} \\ &\leq (C + R_*(1)) \left(\frac{\varepsilon}{\varepsilon_1}\right)^\delta \end{aligned}$$

for times

$$0 \leq t \leq T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_1}{\varepsilon}\right)^{a(1)(1-2\delta)}\right) \leq t^* + T_* \exp\left(\frac{s_0}{6} \left(\frac{\varepsilon_1}{\varepsilon}\right)^{a(1)(1-2\delta)}\right).$$

To summarize, from the dichotomy we either have

$$\|I(t) - I_0\| \leq K^{-1} \leq C \left( \frac{\varepsilon}{\varepsilon_1} \right)^\delta, \quad 0 \leq t < +\infty$$

or

$$\|I(t) - I_0\| \leq (C + R_*(1)) \left( \frac{\varepsilon}{\varepsilon_1} \right)^\delta, \quad 0 \leq t \leq T_* \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon_1}{\varepsilon} \right)^{a(1)(1-2\delta)} \right)$$

therefore in both cases the last estimate is satisfied, thus proving the first part of the statement.

For the second part of the statement (the case  $\delta = 0$ ), one can proceed exactly the same way but choosing

$$K := \frac{2}{r_0}$$

instead.

To conclude, in both cases, the estimate on the evolution of the  $J$  variables (for  $m = 1, \tau = 0$ ) can be obtained exactly as in the proof of Theorem 1.2.  $\square$

## 6. Improved stability far away from resonances

We now investigate solutions which start far away from resonances. Results of this section do not depend on  $h$  being convex or its gradient being bounded, that is  $(\kappa)$  and  $(\Omega)$  are unnecessary. It will be sufficient to assume the existence of  $\bar{\kappa} > 0$  such that for any Lebesgue measurable subset  $U \subset \mathbb{R}^n$ , we have the measure estimate

$$\text{Leb}(\nabla h^{-1}(U) \cap D) \leq \bar{\kappa}^{-1} \text{Leb}(U) \quad (\bar{\kappa})$$

where  $\text{Leb}$  denotes the Lebesgue measure on  $\mathbb{R}^n$ . Certainly, the convexity assumption  $(\kappa)$  implies  $(\bar{\kappa})$  with  $\bar{\kappa} = \kappa$ , but the latter is more general: in particular, it holds true if  $h$  is Kolmogorov non-degenerate, that is if the determinant of the Hessian  $\nabla h^2(I)$  is uniformly bounded away from zero for any  $I \in D$ .

Now consider  $\Lambda \in M_{K,1}^a$  an admissible maximal  $K$ -submodule of  $\mathbb{Z}^{n+m}$  of rank 1. Then  $\Lambda$  contains a unique vector  $(k, l) \in \mathbb{Z}^n \setminus \{0\} \times \mathbb{Z}^m$  such that  $|(k, l)| \leq K$  and such that its components are relatively primes. The submodule  $\tilde{\Lambda}$  of  $\mathbb{Z}^n$  is then generated by  $k$ , and  $|\tilde{\Lambda}| = \|k\|$ . In the sequel, we shall write  $\Lambda = \Lambda(k, l)$ .

With these notations, we recall that the completely non-resonant block  $B_{\{0\}} \subset \mathbb{R}^n$ , introduced in Section 2, can be defined by

$$B_{\{0\}} := \left\{ \omega \in \mathbb{R}^n \mid \|\omega - R_{\Lambda(k,l)}\| \geq \frac{\lambda_1}{\|k\|} \right\},$$

$$\lambda_1 = \frac{r_0 \gamma}{8F^n(n+1)^\tau K^{(n+1)\tau+n}}, \quad F = 10M/\bar{\kappa}.$$

Therefore, from Proposition 7.1, this set is  $(\beta_{\{0\}}, K)$ -non resonant with

$$\beta_{\{0\}} = \frac{9MKr_{\{0\}}}{8}, \quad r_{\{0\}} = \frac{r_0\gamma}{F^{n+1}(n+1)^\tau K^{(n+1)(\tau+1)}}.$$

If we further define

$$D_{\{0\}} := \{I \in D \mid \nabla h(I) \in B_{\{0\}}\}, \quad \bar{D}_{\{0\}} := D_{\{0\}} \times \mathbb{R}^m,$$

we arrive at the following statement.

**Theorem 6.1.** *Let  $H$  be as in (H), with  $h$  satisfying  $(M)$  and  $(\bar{\kappa})$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  satisfies  $(\text{Dio}_{\gamma,\tau})$ , and let us define*

$$\bar{R}_* := \left(\frac{\bar{\kappa}}{10M}\right)^{n+1} \frac{r_0\gamma}{(n+1)^\tau},$$

$$\bar{\varepsilon}_0 := \frac{Mr_0^2\gamma^2}{2^{10}(n+1)^{2\tau}} \left(\frac{\bar{\kappa}}{10M}\right)^{2(n+1)}, \quad \bar{\varepsilon}_* = \bar{\varepsilon}_0 \left(\frac{r_0}{\bar{R}_*}\right)^2.$$

If  $\varepsilon \leq \min\{\bar{\varepsilon}_0, \bar{\varepsilon}_*\}$ , for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D}_{\{0\}} \times \mathbb{T}^{n+m}$ , we have

$$\|(I(t), J(t)) - (I_0, J_0)\| \leq \bar{R}_* \left(\frac{\varepsilon}{\bar{\varepsilon}_0}\right)^{\frac{1}{2}}, \quad |t| \leq \frac{s_0 \bar{R}_*}{5\bar{\varepsilon}_0} \left(\frac{\bar{\varepsilon}_0}{\varepsilon}\right)^{\frac{1}{2}} \exp\left(\frac{s_0}{6} \left(\frac{\bar{\varepsilon}_0}{\varepsilon}\right)^a\right).$$

Moreover, the complement of  $D_{\{0\}}$  in  $\mathbb{R}^n$  has a measure of order  $\varepsilon^b$ , with  $a$  and  $b$  as in Theorem 1.2.

Observe that the measure estimate on the complement of  $D_{\{0\}}$  is better than the one obtained in [27] for  $\tau = 0$ , as in the latter reference it is only of order one. Observe also that it is only for this measure estimate that  $(\bar{\kappa})$  is needed; the first part of the statement holds true without this assumption.

*Proof.* Proposition 3.2 can be applied with  $r = r_{\{0\}}$ , provided that

$$\varepsilon \leq \frac{\beta_{\{0\}} r_{\{0\}}}{2^{79}K} = \frac{Mr_{\{0\}}^2}{2^{10}}, \quad r_{\{0\}} \leq r_0, \tag{6.1}$$

holds true. Choosing

$$K := \left(\frac{\bar{\varepsilon}_0}{\varepsilon}\right)^a,$$

we have

$$r_{\{0\}} = \bar{R}_* \left(\frac{\varepsilon}{\bar{\varepsilon}_0}\right)^{\frac{1}{2}}$$

and the inequalities (6.1) are satisfied if  $\varepsilon \leq \min\{\bar{\varepsilon}_0, \bar{\varepsilon}_*\}$ , where  $\bar{\varepsilon}_0$ ,  $\bar{\varepsilon}_*$  and  $\bar{R}_*$  are the constants given in the statement. It then follows from Proposition 3.2 that

$$\|(I(t), J(t)) - (I_0, J_0)\| \leq r_{\{0\}}, \quad |t| \leq \frac{s_0 r_{\{0\}}}{5\varepsilon} e^{Ks_0/6},$$

that is

$$\|(I(t), J(t)) - (I_0, J_0)\| \leq \bar{R}_* \left( \frac{\varepsilon}{\bar{\varepsilon}_0} \right)^{\frac{1}{2}}, \quad |t| \leq \frac{s_0 \bar{R}_*}{5\bar{\varepsilon}_0} \left( \frac{\bar{\varepsilon}_0}{\varepsilon} \right)^{\frac{1}{2}} \exp \left( \frac{s_0}{6} \left( \frac{\bar{\varepsilon}_0}{\varepsilon} \right)^a \right),$$

which proves the first part of the statement.

Concerning the second part of the statement, we follow [27]. First, in view of  $(\bar{\kappa})$ , it suffices to show that the complement of  $B_{\{0\}}$  in  $\mathbb{R}^n$  has a relative Lebesgue measure of order  $\varepsilon^b$ . But by construction, the latter set is  $Z_1$ , the resonant zone of multiplicity 1 defined in Section 2, which has a relative measure of order

$$\sum_{\Lambda(k,l) \in M_{K,1}^a} \frac{\lambda_1}{\|k\|} \sim \frac{1}{K^{(n+1)\tau+n}} \sum_{\substack{k \in \mathbb{Z}^n, \\ 0 < |k| \leq K}} \frac{1}{\|k\|} \sim \frac{1}{K^{(n+1)\tau+1}} \sim \varepsilon^{a((n+1)\tau+1)} = \varepsilon^b.$$

□

Next we look at a different, and in some sense more natural, non-resonant set. Choose  $0 < \gamma' \leq \gamma$  and  $\tau'$  such that  $\tau' > n + m - 1$  and  $\tau' \geq \tau$ . We define

$$B_{\gamma', \tau'} := \left\{ \omega \in \mathbb{R}^n \mid |(k, l) \cdot (\omega, \alpha)| \geq \gamma'(|k| + |l|)^{-\tau'} \right\}$$

and

$$D_{\gamma', \tau'} := \{I \in D \mid \nabla h(I) \in B_{\gamma', \tau'}\}, \quad \bar{D}_{\gamma', \tau'} := D_{\gamma', \tau'} \times \mathbb{R}^m.$$

Those sets are clearly  $(\beta', K)$ -non resonant, with  $\beta' := \gamma' K^{-\tau'}$ , and we obtain the following result.

**Theorem 6.2.** *Let  $H$  be as in (H), with  $h$  satisfying (M) and  $(\bar{\kappa})$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , and let us define*

$$a' := \frac{1}{2(\tau' + 1)}, \quad R'_* := \frac{8r_0\gamma'}{9M}, \quad \varepsilon'_0 := \frac{r_0^2\gamma'^2}{768M}, \quad \varepsilon'_* = \varepsilon'_0 \left( \frac{r_0}{R'_*} \right)^2.$$

*If  $\varepsilon \leq \min\{\varepsilon'_0, \varepsilon'_*\}$ , for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D}_{\gamma', \tau'} \times \mathbb{T}^{n+m}$ , we have*

$$\|(I(t), J(t)) - (I_0, J_0)\| \leq R'_* \left( \frac{\varepsilon}{\varepsilon'_0} \right)^{\frac{1}{2}}, \quad |t| \leq \frac{s_0 R'_*}{5\varepsilon'_0} \left( \frac{\varepsilon'_0}{\varepsilon} \right)^{\frac{1}{2}} \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon'_0}{\varepsilon} \right)^{a'} \right).$$

*Moreover, the complement of  $D_{\gamma', \tau'}$  in  $\mathbb{R}^n$  has a measure of order  $\gamma'$ .*



As we already said, when  $m \geq 2$ , the set of vectors  $\alpha \in \mathbb{R}^m$  that satisfy  $(\text{Dio}_{\gamma, \tau})$  with  $\tau = m - 1$  has zero measure, and when  $\tau > m - 1$ , we can choose  $\tau' = n + \tau$  and thus

$$a' = \frac{1}{2(n + \tau + 1)}$$

in the statement above.

*Proof.* Let us define

$$r' := \frac{8r_0\beta'}{9MK} = \frac{8r_0\gamma'}{9MK^{\tau'+1}}.$$

Proposition 3.2 can be applied with  $r = r'$ , provided

$$\varepsilon \leq \frac{\beta' r'}{2^{79} K} = \frac{M r'^2}{2^{10}}, \quad r' \leq r_0 \quad (6.2)$$

holds true. Choosing

$$K := \left( \frac{\varepsilon'_0}{\varepsilon} \right)^{a'}$$

we have

$$r' = R'_* \left( \frac{\varepsilon}{\varepsilon'_0} \right)^{\frac{1}{2}}$$

and the inequalities (6.2) are satisfied if  $\varepsilon \leq \min\{\varepsilon'_0, \varepsilon'_*\}$ , where  $a'$ ,  $\varepsilon'_0$ ,  $\varepsilon'_*$  and  $R'_*$  are the constants given in the statement. Then, exactly as before, Proposition 3.2 yields

$$\|(I(t), J(t)) - (I_0, J_0)\| \leq R'_* \left( \frac{\varepsilon}{\varepsilon'_0} \right)^{\frac{1}{2}}, \quad |t| \leq \frac{s_0 R'_*}{5\varepsilon'_0} \left( \frac{\varepsilon'_0}{\varepsilon} \right)^{\frac{1}{2}} \exp \left( \frac{s_0}{6} \left( \frac{\varepsilon'_0}{\varepsilon} \right)^{a'} \right),$$

which gives the first part of the statement.

Concerning the second part of the statement, as before it is enough to prove that the complement of  $B_{\gamma', \tau'}$  has a relative measure of order  $\gamma'$ . The complement of  $B_{\gamma', \tau'}$  is

$$\left\{ \omega \in \mathbb{R}^n \mid \exists (k, l) \in \mathbb{Z}^{n+m}, \quad |(k, l) \cdot (\omega, \alpha)| < \gamma'(|k| + |l|)^{-\tau'} \right\},$$

but since  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , and since  $\gamma' \leq \gamma$  and  $\tau' \geq \tau$ , this set is also equal to

$$\left\{ \omega \in \mathbb{R}^n \mid \exists (k, l) \in \mathbb{Z}^n \setminus \{0\} \times \mathbb{Z}^m, \quad |(k, l) \cdot (\omega, \alpha)| < \gamma'(|k| + |l|)^{-\tau'} \right\}.$$

But now the above set is known to have a relative measure of order  $\gamma'$ : this is exactly the content of Lemma 2.12 in [19].  $\square$

## 7. A more general stability result

Let us finally give a more general result, where the Diophantine assumption  $(\text{Dio}_{\gamma,\tau})$  is removed. Assuming  $\alpha \in \mathbb{R}^m$  to be simply non-resonant, we can define a function  $\Psi = \Psi_\alpha$  by

$$\Psi(K) = \max \{ |k \cdot \alpha|^{-1} \mid k \in \mathbb{Z}^m, 0 < |k| \leq K \}, \quad K \geq 1. \quad (7.1)$$

Then we define  $\Delta = \Delta_\alpha$  by

$$\Delta(x) = \sup \{ K \geq 1 \mid K\Psi(K) \leq x \}, \quad x \geq \Psi(1) = |\alpha|_\infty^{-1}, \quad |\alpha|_\infty = \max_{1 \leq j \leq m} |\alpha_j|. \quad (7.2)$$

If  $\alpha$  satisfies  $(\text{Dio}_{\gamma,\tau})$ , then the functions  $\Psi$  and  $\Delta$  defined above satisfy

$$\Psi(K) \leq \gamma^{-1} K^\tau, \quad \Delta(x) \geq (\gamma x)^{\frac{1}{\tau+1}}. \quad (7.3)$$

The only place where  $(\text{Dio}_{\gamma,\tau})$  was used was in Lemma 2.1. But using the function  $\Psi$  instead, the exact same proof yields the following more general lemma.

**Lemma 7.1.** *Let  $K \geq 1$ ,  $E > 0$  and  $F \geq E + 1$ . Assume that  $\alpha \in \mathbb{R}^m$  is non-resonant and*

$$\begin{cases} FK\lambda_d \leq \lambda_{d+1} \leq \Psi((d+1)K^{d+1})^{-1}, & 1 \leq d \leq n-1, \\ \lambda_n \leq F^{-1}\Psi((n+1)K^{n+1})^{-1}. \end{cases}$$

*Then for any  $\Lambda \in M_K^a$ , the block  $B_\Lambda$  is  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$  with*

$$\begin{cases} \beta_\Lambda = EK\delta_\Lambda, & \Lambda \neq \{0\}, \\ \beta_{\{0\}} = \lambda_1. \end{cases}$$

Using this Lemma instead of Lemma 2.1, we arrive at the following proposition which generalizes Proposition 2.2.

**Proposition 7.2.** *For  $K \geq 1$  and  $\Lambda \in M_K^a$  of rank  $d$ , with  $0 \leq d \leq n$ , let us define*

$$r_\Lambda := \frac{r_0}{|\tilde{\Lambda}| F^{n-d+1} \Psi((n+1)K^{n+1}) K^{n-d+1}}, \quad F := 10M/\kappa$$

*where we set, by convention,  $|\tilde{\Lambda}| = 1$  if  $\Lambda = \{0\}$ . Assume that  $\kappa r_0 \leq 80$ . Then there exists a covering of  $D$  by subsets  $D_\Lambda$ , where  $\Lambda \in M_K^a$ , such that each  $D_\Lambda$  is  $(\beta_\Lambda, K)$ -non resonant modulo  $\Lambda$  and, for  $\Lambda \neq \{0\}$ ,  $\delta_\Lambda$ -close to  $\Lambda$ -resonances with*

$$\beta_\Lambda = \frac{9MKr_\Lambda}{8}, \quad \delta_\Lambda = \frac{\kappa r_\Lambda}{8}.$$

Then, using Proposition 7.2 instead of Proposition 2.2, together with Proposition 3.2 and Proposition 3.3, and proceeding exactly as in the Proof of Theorem 1.2, we obtain the following statement.

**Theorem 7.3.** *Let  $H$  be as in (H), with  $h$  satisfying (M),  $(\Omega)$  and  $(\kappa)$ , and  $f$  satisfying  $(\varepsilon)$ . Assume also that  $\alpha$  is non-resonant,  $m \geq 2$  and that*

$$\varepsilon \leq \frac{(n+1)^2 \kappa r_0^2 |\alpha|_\infty^2}{2^{10} F^{2(n+1)}}$$

so that we can define

$$\Delta_\varepsilon := \Delta \left( \frac{(n+1)r_0}{2^5 F^{n+1}} \sqrt{\frac{\kappa}{\varepsilon}} \right),$$

$$K_\varepsilon := \left( \frac{\Delta_\varepsilon}{n+1} \right)^{\frac{1}{n+1}}, \quad R_\varepsilon := \frac{1}{F K_\varepsilon \Psi((n+1)K_\varepsilon^{n+1})}.$$

Then, if  $\kappa r_0 \leq \min\{80, 24\|\alpha\|\}$ ,  $K_\varepsilon \geq 1$  and  $R_\varepsilon \leq r_0$ , for any solution  $(I(t), J(t), \theta(t), \varphi(t))$  of the system associated to  $H$  with initial condition  $(I_0, J_0, \theta_0, \varphi_0) \in \bar{D} \times \mathbb{T}^{n+m}$ , we have

$$\|I(t) - I_0\| \leq R_\varepsilon, \quad |t| \leq T_* \exp\left(\frac{s_0 K_\varepsilon}{6}\right).$$

Moreover, we have

$$|J(t) - J_0|_\infty \leq K_\varepsilon^n R_\varepsilon, \quad |t| \leq T_* \exp\left(\frac{s_0 K_\varepsilon}{6}\right).$$

When  $\alpha$  is Diophantine, that is when  $\alpha$  satisfies  $(\text{Dio}_{\gamma, \tau})$ , then (7.3) holds true and the above statement exactly reduces to Theorem 1.2. Observe also that for  $m = 1$ , any non zero vector  $\alpha \in \mathbb{R}$  satisfies  $(\text{Dio}_{\gamma, \tau})$  with  $\gamma = |\alpha|$  and  $\tau = 0$ , so Theorem 7.3 gives new information only when  $m \geq 2$ .

Now for  $m \geq 2$ , if  $\alpha$  does not satisfy any Diophantine condition,  $\Delta_\varepsilon$  and hence  $K_\varepsilon$  cannot grow as a power of  $\sqrt{\varepsilon}^{-1}$  and therefore the stability time is not exponentially large with respect to (some power of)  $\varepsilon^{-1}$ . Yet it is always exponentially large with respect to some other function of  $\sqrt{\varepsilon}^{-1}$ , namely  $K_\varepsilon$ , and this can give some non-trivial stability results even if  $\alpha$  is Liouville (that is, when  $\alpha$  is not Diophantine).

But first let us observe that, unfortunately, in full generality this stability estimate might not be better than the trivial stability estimate. To simplify the discussion here, we will assume that  $\Psi$  is in fact continuous so that  $\Delta$  is nothing but the functional inverse of the increasing continuous map  $K \mapsto K\Psi(K)$ ; observe that in fact only the values of  $\Psi(K)$  at integers  $K \in \mathbb{N}$  are interesting so it is always possible to modify  $\Psi$  into a continuous without affecting the Diophantine properties of  $\alpha$ . Dropping constants which are independent of  $\varepsilon$ , it is easy to check that  $R_\varepsilon \simeq K_\varepsilon^n \sqrt{\varepsilon}$  so that the stability estimate reads

$$\|I(t) - I_0\| \lesssim K_\varepsilon^n \sqrt{\varepsilon}, \quad |t| \lesssim \exp K_\varepsilon.$$

Such estimate is better than the trivial estimate if

$$\sqrt{\varepsilon}^{-1} K_\varepsilon^n \leq \exp K_\varepsilon \quad (7.4)$$

but the latter inequality is not always satisfied. Indeed, the growth of  $\Psi(K)$ , as  $K$  goes to infinity, can be arbitrarily fast; for instance, when  $m = 2$  and assuming  $\alpha = (1, \alpha_1)$  with  $\alpha_1 \in [0, 1]$  irrational, letting  $(p_j/q_j)_{j \geq 1}$  be the convergents of  $\alpha_1$ , the growth of  $\Psi$  is nothing but the growth of the denominators  $q_j$ , and this growth can be arbitrarily fast (given any sufficiently increasing function  $\phi : [1, +\infty[ \rightarrow [1, +\infty[$ , one can always construct  $\alpha_1$  such that  $q_{j+1} \simeq \phi(q_j)$ ). As a consequence, the growth of  $\Delta(x)$ , as  $x$  goes to infinity can be arbitrarily slow, and since  $K_\varepsilon \simeq \Delta(\sqrt{\varepsilon}^{-1})^{\frac{1}{n+1}}$ , the inequality (7.4) is not necessarily satisfied.

As a side remark, it is because of the confinement by convexity and energy preservation that the time of stability is of the form  $\exp K_\varepsilon$  and not of the form  $\sqrt{\varepsilon}^{-1} \exp K_\varepsilon$ ; the latter is indeed the stability time one can reach in the general steep case where this convexity argument cannot be used (see for instance [18]). If we could replace  $\exp K_\varepsilon$  in (7.4) by  $\sqrt{\varepsilon}^{-1} \exp K_\varepsilon$ , then (7.4) would be always satisfied for  $\varepsilon$  sufficiently small, and therefore the stability would be non-trivial for any Liouville vector. But here we do rely on convexity and so our result do not give anything interesting for very Liouville vectors.

Let us now give somehow more concrete examples. If

$$\Psi(K) \sim K^{-1} \exp \left( K^{\frac{1}{n+1}} \right)$$

then the stability result is not interesting: indeed, in this case one has

$$\Delta(x) \sim (\log(x))^{n+1}, \quad K_\varepsilon \sim \log(\sqrt{\varepsilon}^{-1})$$

and so the result is

$$\|I(t) - I_0\| \lesssim \sqrt{\varepsilon} \log \left( \frac{1}{\sqrt{\varepsilon}} \right), \quad |t| \lesssim \frac{1}{\sqrt{\varepsilon}}.$$

However, if

$$\Psi(K) \sim K^{-1} \exp \left( c^{-1} K^{\frac{1}{n+1}} \right), \quad (7.5)$$

for some constant  $c > 1$ , then

$$\|I(t) - I_0\| \lesssim \sqrt{\varepsilon} \log \left( \frac{1}{\sqrt{\varepsilon}} \right)^c, \quad |t| \lesssim \left( \frac{1}{\sqrt{\varepsilon}} \right)^c$$

so we have a non-trivial stability estimate for an interval of time which is a power of  $\sqrt{\varepsilon}^{-1}$ . In particular, if (7.5) is satisfied for all  $c > 1$ , for instance if

$$\Psi(K) \sim K^{-1} \exp \left( K^{\frac{\alpha}{n+1}} \right), \quad 0 < \alpha < 1$$

or even

$$\Psi(K) \sim K^{-1} \exp\left(\log(K)^{-\beta} K^{\frac{1}{n+1}}\right), \quad \beta > 1,$$

then the stability time is better than any fixed power of  $\sqrt{\varepsilon}^{-1}$  (and consequently any fixed power of  $\varepsilon^{-1}$ ). For these class of vectors, the stability time, even though not exponential, is actually very large.

To conclude, let us also point out that there are also analogues of Theorem 5.1 and Theorem 6.1 in the above context that can be stated and proved the same way; we leave the details to the reader.

## 8. Concluding remarks

We conclude this paper with some remarks. We first discuss the possibility of extending our results to less regular Hamiltonians, then to more general classes of integrable Hamiltonians and finally to more general time-dependence.

Let us first comment on the regularity assumption in the results we proved. The assumptions that  $h$  and  $f$  are real-analytic were only used in the normal form Lemma 3.1, which is taken from [27]. So in order to have results for Hamiltonians which are not real-analytic, for instance Hamiltonians which are only Gevrey regular or finitely differentiable, one just needs a version of Lemma 3.1 in those settings, and this appears to be only a problem of technical nature. One can find in [8] and [9] normal form results for non-analytic Hamiltonians in the spirit of Lemma 3.1, even though those statements do not recover Lemma 3.1. Let us also recall that in the autonomous case (or time-periodic case when  $h$  is convex), Nekhoroshev type estimates are known for Gevrey or finitely differentiable Hamiltonians [4, 5, 24] but all those proofs are based on the Lochak method.

Next let us discuss the more interesting question of whether our results extend to steep, or  $S$ -steep or  $P$ -steep, integrable Hamiltonians, the original classes of integrable Hamiltonians considered by Nekhoroshev. One should first recall that in the general steep case, unlike what happens in the convex case, motions near resonances are not necessarily confined: if they are not, the steepness property ensures that they evolve towards a less resonant domain and, eventually, end up in a non-resonant domain on which an integrable normal form (up to a small remainder) can be constructed, leading to the stability of the action variables. First is not hard to see that one cannot obtain such a stability result for a quasi-periodic time-dependent or even a periodic time-dependent perturbation of a steep integrable Hamiltonian. Indeed, the simplest example of steep integrable Hamiltonian is given by the quasi-convex Hamiltonian

$$h(I_1, \dots, I_n) = \frac{1}{2}(I_1^2 + \dots + I_{n-1}^2) + I_n.$$

After a periodic time-dependent perturbation the extended integrable Hamiltonian then reads

$$\bar{h}(I_1, \dots, I_n, J) = \frac{1}{2}(I_1^2 + \dots + I_{n-1}^2) + I_n + J$$

and obviously the latter cannot be stable: adding the perturbation  $\varepsilon \cos(\theta_n - \varphi)$ , where  $(\theta_n, \varphi)$  are the angles conjugated to  $(I_n, J)$ , it is easy to see that the evolution of  $(I_n(t), J(t))$  is unbounded. The correct question to be asked is whether the stability result holds true for a quasi-periodic time-dependent perturbation of a  $S$ -steep or  $P$ -steep integrable Hamiltonian. What we have shown is that one does have stability in the simplest class of  $S$ -steep and  $P$ -steep integrable Hamiltonians, namely convex Hamiltonians. If  $h$  is  $S$ -steep, then it is also  $P$ -steep and it follows from the work of Nekhoroshev that one has stability for periodic time-dependent perturbations of  $S$ -steep Hamiltonians, but the question we asked is whether one has stability for quasi-periodic time-dependent perturbations of  $S$ -steep integrable Hamiltonian. We can also the same question for the larger class of  $P$ -steep integrable Hamiltonians, and more generally, the question is what could be the class of  $QP$ -steep (“quasi-periodically steep”) integrable Hamiltonians for which one has stability after a quasi-periodic time-dependent perturbation.

To conclude, let us mention that it seems quite unlikely to have a non-trivial stability result for an arbitrary time-dependent perturbation, unless the time depends on the small parameter, in which case the conjugated action variable can be considered as degenerate (see [16] or [7]). But it may be possible to extend our results for a class of perturbation whose Fourier transform (with respect to time) has suitable localization properties.

**Acknowledgements.** This work was written while the author was visiting IMPA; I would like to thank this institute for its great hospitality. I would like also to thank an anonymous referee for his constructive remarks that were important in improving the results of this paper (in particular, the fact that the  $J$  variables are also stable was obtained after one of his remarks).

## References

- [1] G. Benettin and G. Gallavotti, Stability of motions near resonances in quasi-integrable Hamiltonian systems, *J. Stat. Phys.*, **44** (1986), 293–338. Zbl 0636.70018 MR 0857061
- [2] U. Bessi, An approach to Arnold’s diffusion through the calculus of variations, *Nonlinear Anal. Theory Methods Appl.*, **26** (1996), no. 6, 1115–1135. Zbl 0867.70013 MR 1375654
- [3] U. Bessi, Arnold’s example with three rotators, *Nonlinearity*, **10** (1997), no. 3, 763–781. Zbl 0912.70015 MR 1448586



- [4] A. Bounemoura, Nekhoroshev theory for finitely differentiable quasi-convex Hamiltonians, *Journal of Differential Equations*, **249** (2010), no. 11, 2905–2920. Zbl 1206.37029 MR 2718671
- [5] A. Bounemoura, Effective stability for Gevrey and finitely differentiable prevalent Hamiltonians, *Communication in Mathematical Physics*, **307** (2011), no. 1, 157–183. Zbl 1253.37057 MR 2835876
- [6] A. Bounemoura, Optimal stability and instability for near-linear hamiltonians, *Annales Henri Poincaré*, **13** (2012), no. 4, 857–868. Zbl 1320.70014 MR 2913624
- [7] A. Bounemoura, Effective stability for slow time-dependent near-integrable Hamiltonians and application, *C. R. Acad. Sci. Paris, Ser. I*, **351** (2013), no. 17–18, 673–676. Zbl 1288.37018 MR 3124324
- [8] A. Bounemoura, Normal forms, stability and splitting of invariant manifolds, I. Gevrey Hamiltonians, *Regul. Chaotic Dyn.*, **18** (2013), no. 3, 237–260. Zbl 06197380 MR 3061808
- [9] A. Bounemoura, Normal forms, stability and splitting of invariant manifolds, II. Finitely differentiable Hamiltonians, *Regul. Chaotic Dyn.*, **18** (2013), no. 3, 261–276. Zbl 1325.37034 MR 3061809
- [10] A. Bounemoura and J.-P. Marco, Improved exponential stability for near-integrable Hamiltonian systems, *Nonlinearity*, **24** (2011), no. 1, 97–112. Zbl 1213.37090 MR 2739833
- [11] A. Bounemoura and L. Niederman, Generic Nekhoroshev theory without small divisors, *Ann. Inst. Fourier*, **62** (2012), no. 1, 277–324. Zbl 1257.37036 MR 2986272
- [12] B. Chirikov, A universal instability of many-dimensional oscillator systems, *Phys. Reports*, **52** (1979), 263–379. MR 0536429
- [13] B. Chirikov and V. Vecheslavov, *How fast is the Arnold diffusion?*, preprint INP, 89–72, Novosibirsk, 1989. <http://www.quantware.ups-tlse.fr/chirikov/publi/binp/chiinp1989e.pdf>
- [14] B. Chirikov and V. Vecheslavov, *Arnold diffusion in large systems*, preprint Budker INP, 96–44, Novosibirsk, 1996. <http://www.quantware.ups-tlse.fr/chirikov/publi/binp/chiinp1996ad.pdf>
- [15] G. Gallavotti, Twistless KAM tori, *Commun. Math. Phys.*, **164** (1994), no. 1, 145–156. Zbl 0805.58054 MR 1288156
- [16] A. Giorgilli and E. Zehnder, Exponential stability for time dependent potentials, *Z. Angew. Math. Phys.*, **43** (1992), no. 5, 827–855. Zbl 0766.58032 MR 1182784
- [17] M. Guzzo, L. Chierchia and G. Benettin, The steep Nekhoroshev's theorem and optimal stability exponents (an announcement), *Rend. Lincei Mat. Appl.*, **25** (2014), 293–299. Zbl 1326.37040 MR 3256210



- [18] M. Guzzo, L. Chierchia and G. Benettin, The steep Nekhoroshev's theorem, *Commun. Math. Phys.*, **342** (2016), no. 2, 569–601. Zbl 1337.37038 MR 3459160
- [19] A. Jorba, *On Quasi-Periodic Perturbations of Ordinary Differential Equations*, PhD thesis, Universitat de Barcelona, 1991. MR 1202451
- [20] P. Lochak, Canonical perturbation theory via simultaneous approximation, *Russ. Math. Surv.*, **47** (1992), no. 6, 57–133. Zbl 0795.58042 MR 1209145
- [21] P. Lochak, Tores invariants à torsion évanescence dans les systèmes hamiltoniens proches de l'intégrable, *C.R. Acad. Sci. Paris Sér. I Math.*, **327** (1998), no. 9, 833–836. Zbl 0917.58011 MR 1663746
- [22] P. Lochak, J.-P. Marco and D. Sauzin, On the splitting of invariant manifolds in multidimensional near-integrable Hamiltonian systems, *Mem. Am. Math. Soc.*, **163** (2003), no. 775, 145pp. Zbl 1038.70001 MR 1964346
- [23] P. Lochak and A. Neishtadt, Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian, *Chaos*, **2** (1992), no. 4, 495–499. Zbl 1055.37573 MR 1195881
- [24] J.-P. Marco and D. Sauzin, Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems, *Publ. Math. Inst. Hautes Études Sci.*, **96** (2002), 199–275. Zbl 1086.37031 MR 1986314
- [25] N. Nekhoroshev, An exponential estimate of the time of stability of nearly integrable Hamiltonian systems, *Russian Math. Surveys*, **32** (1977), no. 6, 1–65. Zbl 0389.70028 MR 0501140
- [26] N. Nekhoroshev, An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. II, *Trudy Sem. Petrovs*, **5** (1979), 5–50. Zbl 0473.34021 MR 0549621
- [27] J. Pöschel, Nekhoroshev estimates for quasi-convex Hamiltonian systems, *Math. Z.*, **213** (1993), 187–216. Zbl 0857.70009 MR 1221713
- [28] J. Pöschel, A lecture on the classical KAM theory, in Katok, Anatole (ed.) et al., *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, 707–732, Proc. Symp. Pure Math., 69, Providence, RI, Amer. Math. Soc. (AMS), 2001. Zbl 0999.37053 MR 1858551
- [29] M. B. Sevryuk, The classical KAM theory at the dawn of the Twenty-First Century, *Mosc. Math. J.*, **3** (2003), no. 3, 1113–1144. Zbl 1121.37045 MR 2078576
- [30] K. Zhang, Speed of Arnold diffusion for analytic Hamiltonian systems, *Inventiones mathematicae*, **186** (2011), no. 2, 255–290. Zbl 1251.37063 MR 2845619

Received September 09, 2015

A. Bounemoura, CNRS-IMPA UMI 2924, Rio de Janeiro, RJ 22460-320, Brazil

E-mail: abedbou@gmail.com

