IMPROVED STABILITY FOR ANALYTIC QUASI-CONVEX NEARLY INTEGRABLE SYSTEMS AND OPTIMAL SPEED OF ARNOLD DIFFUSION

JIANLU ZHANG† AND KE ZHANG‡

ABSTRACT. We improve the global Nekhoroshev stability for analytic quasi-convex nearly integrable Hamiltonian systems. The new stability result is optimal, as it matches the fastest speed of Arnold diffusion.

1. Introduction

We consider a real analytic Hamiltonian

\[ H(\theta, I) = h(I) + f(\theta, I), \quad I \in \mathbb{R}^n, \quad \theta \in \mathbb{T}^n = (\mathbb{R}/2\pi\mathbb{Z})^n, \]

with \( |f| < \epsilon \ll 1 \). It is a classical result of Nekhoroshev (9, 11) that when \( h(I) \) satisfies a non-degeneracy condition known as steepness (see also the modern treatments of 11, 5), the system enjoys a global stretched exponential stability, of the type

\[ \|I(t) - I(0)\| \leq Ce^{b}, \quad \text{for } |t| \leq \exp \left(-C^{-1}e^{-a}\right). \]

In the case when the integrable Hamiltonian is quasi-convex (see definition below), the system enjoys the largest stability exponent \( b \). Lochak and Neishtadt, also Pöschel (see for example 6, 8, 12) obtained the exponents

\[ a = b = \frac{1}{2n}. \]

Lochak also discovered the remarkable phenomenon known as “stability by resonance”, that if the initial condition is close to a \( d \)-resonance of low order, then one expects the stability exponents \( a = b = \frac{1}{2(n-d)} \). By taking advantage of this fact, and that 1–resonances divide the space, in 4, Bounemoura and Marco obtained larger stability.
exponent $a$ by allowing larger stability region (i.e. smaller $b$). The exponents obtained are

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

where $\sigma > 0$ can be arbitrarily small. The exponent $a$ can be taken to be $\frac{1}{2(n-1)}$ if one allows stability region of order 1.

On the flip side, one is interested in the instability question known as Arnold diffusion. This research was started by the nominal work of Arnold ([1]), where he discovered the first mechanism for instability for nearly integrable Hamiltonian systems. Bessi ([2], [3]) proved that for $n = 3, 4$, there exists diffusion orbits $(\theta, I)(t)$, for which there exists $t > 0$ such that

$$\|I(t) - I(0)\| \geq C^{-1}, \quad |t| \leq C \exp \left( C^{-1} e^{-\frac{1}{2(n-2)}} \right).$$

This result was then generalized to arbitrary $n \geq 5$ by the second author of this paper ([14], see also related work in [7]). The reason for the exponent $\frac{1}{2(n-2)}$ is due to restriction of Arnold’s mechanism: the orbit constructed using Arnold’s idea must always cross a double resonance, therefore the exponents obtained are the best allowed in that class.

Up to now, there was still a gap between the best lower bound and upper bound of the stability exponent $a$:

$$\frac{1}{2(n - 1)} - \sigma \leq a < \frac{1}{2(n - 2)}.$$ 

In this paper, we close this gap by improving the stability exponents to

$$b = \frac{n - 2}{4\sigma}, \quad a = \frac{1}{2(n - 2)} - \sigma. \quad (1)$$

Thus, the stability exponent $a$ can be arbitrary close to $\frac{1}{2(n-2)}$, and for Arnold diffusion, the exponent $\frac{1}{2(n-2)}$ is optimal.

We obtain the improvements by separating the frequency space into two sets, one is close to resonances of order up to $|\log \epsilon|$, and the complement which is sufficiently non-resonant. In the non-resonant region we provide an improved stability result using first a normal form, then applying the Nekhoroshev’s theory. In the resonant region, we apply an argument similar to the one in [3], to show that the fast diffusion orbit has to be close to a double resonance.
The paper is structured as follows: in section 2 we introduce notations and formulate the result. We also reduce the main theorem to two stability results, in the non-resonant and resonant regions. These results are proven in sections 3 and 4.

2. Formulation of the main result

For $D \subset \mathbb{R}^n$ and $r > 0$ define:

$$V_r D = \{ I \in \mathbb{C}^n : d(I, B) < r \}, \quad U_r D = V_r D \cap \mathbb{R}^n,$$

and

$$V_{r,s} D = \{ (\theta, I) \in \mathbb{C}^n \times (\mathbb{C}/2\pi \mathbb{Z})^n : d(I, B) < r, \quad |\Im(\theta)| < s. \}$$

where $d(x, y) = \max_i |x_i - y_i|$ is induced by the sup-norm in $\mathbb{C}^n$. Consider the space $A_{r,s}(D)$ of real analytic functions $\varphi(\theta, I)$ that is complex analytic on $V_{r,s} D$. The norm on this space is the sup-norm $|\varphi|_{D,r,s} = \sup_{(\theta, I) \in V_{r,s} D} |\varphi(\theta, I)|$.

Let $R, r_0, s_0, m, M > 0$ be parameters let $B(0, R) \subset \mathbb{R}^n$ be the ball of radius $R$, we assume the following conditions for $h$:

- $h \in A_{r_0,s_0}(B(0, R))$.
- $h$ is $l, m$-quasi-convex on $U_{r_0}B(0, R)$, namely, for all $I \in U_{r_0}B(0, R)$, $\nabla h(I) \neq 0$, and
  $$\nabla^2 h(I) v \cdot v \geq m\|v\|^2, \quad \text{if } |v \cdot \nabla h(I)| \leq l\|v\|^1.$$
- $|\nabla h(I)|, |\nabla^2 h(I)| \leq M$ for all $I \in U_{r_0}B(0, R)$.

Let us denote $\mathcal{M} = (R, r_0, s_0, l, m, M)$ the ensemble of parameters, and we reserve the notation $C = C(\mathcal{M})$ or $C_k = C_k(\mathcal{M})$ for unspecified positive constants depending only on $\mathcal{M}$. The following is our main theorem.

**Theorem 2.1.** Under the standard assumptions on $h$, for any $0 < \delta < \frac{1}{2n+4}$, there is $C = C(\mathcal{M}) > 1$, and $\epsilon_0 = \epsilon_0(\delta, \mathcal{M}) > 0$ such that if

$$|f|_{D,r_0,s_0} < \epsilon \leq \epsilon_0,$$

then for all solutions $(\theta, I)(t)$ of $H$ with $I(0) \in B = B(0, R/2)$, we have

$$|I(t) - I(0)| < Ce^{\delta}, \quad \text{for } |t| < C^{-1} \exp \left( C^{-1}e^{-\frac{1-8\delta}{2n+4}} \right).$$

**Remark.** (1) follows by taking $\sigma = \frac{8\delta}{2(n-2)}$. 
The theorem is proven by dividing the $I$-space into two regions: neighborhood of lower order 1-resonance, and the complement. We produce a stability result on each region.

Let $\Lambda \subset \mathbb{Z}^n$ be a submodule, the space of $\Lambda$ resonant frequencies is defined by

$$R_\Lambda = \{ \omega \in \mathbb{R}^n : k \cdot \omega = 0 \text{ for all } k \in \Lambda \}.$$

The associated resonance surface is

$$S_{\Lambda} = \{ I \in \mathbb{R}^n : \omega(I) \in R_\Lambda \}.$$

We say that $\Lambda$ has rank $d$ if there is linearly independent $\{k_1, \cdots, k_d\} \subset \mathbb{Z}^n$ such that $\Lambda = \text{Span}_\mathbb{Z}\{k_1, \cdots, k_d\}$. In this case, we also write $R_\Lambda = R_{k_1, \cdots, k_d}$. $\Lambda$ is called maximal if it’s not contained by a larger module of the same rank. Following Pöschel, we say that $\Lambda$ is a $K$-module if is generated by $|k_i| \leq K$, for all $i = 1, \cdots, d$.

Given a parameter $0 < \beta < 1$, we define

$$L = 12s_0, \quad K(\epsilon) = -L \log \epsilon, \quad r(\epsilon) = \beta^{-1} \epsilon^{\frac{1}{2}}, \quad \alpha(\epsilon) = \beta^{-1} r(\epsilon) K(\epsilon), \quad (2)$$

and

$$\mathcal{N}(\epsilon) = \{ I \in B(0, R) : d(\omega(I), \bigcup_{0<|k| \leq K} R_k) < \alpha(\epsilon) \}, \quad \mathcal{D}(\epsilon) = B(0, R) \setminus \mathcal{N}(\epsilon).$$

Let $\Lambda \subset \mathbb{Z}^n$ be a maximal submodule, then a set $D \subset \mathbb{R}^n$ is called $\alpha, K$ non-resonant modulo $\Lambda$ if $|k \cdot \omega(I)| > \alpha$ for all $k \in \mathbb{Z}^n_K \setminus \Lambda$. $D$ is called $\alpha, K$ fully non-resonant if $\Lambda = \{0\}$. Then the set $\mathcal{N}(\epsilon)$ is $\alpha(\epsilon)$ close to $K$-strong resonances, while the set $\mathcal{D}(\epsilon)$ sufficiently non-resonant.
The main observation is that orbits in the fully non-resonant region are much more stable than expected.

**Proposition 2.2.** Let $N = L/(6s_0)$. Under the our standing assumptions, there is $\epsilon_0 = \epsilon_0(\mathcal{M}), \beta = \beta_0(\mathcal{M}) > 0$, $C_1 = C_1(\mathcal{M}) > 1$ such that for $\mathcal{D}(\epsilon)$ defined using (2), if $|f|_{B(0,R),r_0,s_0} < \epsilon \leq \epsilon_0$, the following hold for $H = h + f$: Suppose $(\theta(t), I(t))$ be an orbit of $H$ starting with $I(0) \in \mathcal{D}(\epsilon)$, then

$$|I(t) - I(0)| \leq C_1 \epsilon^2, \quad \text{for } |t| \leq C_1^{-1} \exp(C_1^{-1} \epsilon - \frac{\delta}{2n}).$$

**Remark.** Choosing $L = 12s_0$ implies $N = 2$, and the stability time is $C_1 \exp(C_1^{-1} \epsilon - \frac{\delta}{2n})$, which is much longer than what’s claimed in Theorem 2.1. The reason is the stability time is completely determined by what’s happening in the resonant region.

**Proposition 2.3.** For $0 < \delta < \frac{1}{2n+4}$, there exists $C_2 = C_2(\mathcal{M}) > 1$ and $\epsilon_0 = \epsilon_0(\mathcal{M}, \delta) > 0$ such that for

$$T = C_2^{-1} \exp(C_2^{-1} \epsilon^{-\frac{1-8\delta}{2n+4}}),$$

and $0 < \epsilon \leq \epsilon_0$, the following hold:

Let $(\theta(t), I(t)), t \in [0,T]$ be an orbit of $H$ with $I(0) \in B(0,R/2)$ and $I(t) \in \mathcal{N}(\epsilon)$ for all $t \in [0,T]$. Then

$$|I(t) - I(0)| \leq C_2 \epsilon^\delta \text{ for all } t \in [0,T].$$

**Proof of Theorem 2.1** Let $T$ be as in (3), and assume that $\epsilon_0$ is small enough so that

$$T \leq C_1^{-1} \exp(C_1^{-1} \epsilon - \frac{\delta}{2n}).$$

Consider any orbit $(\theta, I)(t), t \in [0,T]$ such that $I(0) \in B(0,R/2)$. If there is $t_* \in [0,T]$ such that $I(t_*) \in \mathcal{D}(\epsilon)$, then by Proposition 2.2,

$$|I(t) - I(t_*)| \leq C_1 \epsilon^\delta, \quad t \in [t_* - T, t_* + T] \supset [0,T].$$

Alternatively, $I(t) \in B(0,R/2) \cap \mathcal{N}(\epsilon)$ for all of $[0,T]$, then Proposition 2.3 applies, and the theorem follows.

3. Stability in the non-resonant region

Let $\Lambda \subset \mathbb{Z}^n$ be a maximal submodule, define the projection operator $T_K\varphi$ for $\varphi(\theta, I) = \sum_{k \in \mathbb{Z}^n} \varphi_k(I)e^{(k-\theta)i}$ as follows:

$$T_K\varphi = \sum_{|k| \leq K} \varphi_k(I)e^{(k-\theta)i}, \quad P_{\Lambda}\varphi = \sum_{k \in \Lambda} \varphi_k(I)e^{(k-\theta)i}.$$
We have the following resonant normal form lemma:

**Lemma 3.1** ([12], Normal Form Lemma, page 192). *Suppose \( D \subset B(0,R) \) is \( \alpha, K \)-nonresonant modulo \( \Lambda \), and \( h \) satisfies the standing assumptions. There is \( C_3 = C_3(\mathcal{M}) > 1 \) such that, if \( 0 < r \leq r_0, 0 < s \leq s_0 \) and \( f \in \mathcal{A}_{r,s}D \) satisfies

\[
\|f\|_{D,r,s} \leq \epsilon \leq C_3^{-1} \frac{\alpha r}{K}, \quad r \leq C_3^{-1} \frac{\alpha}{K},
\]

and \( Ks \geq 6 \), then there exists a real analytic coordinate change \( \Phi : V_{r_1,s_1}D \to V_{r,s}D \) with \( r_1 = r/2, s_1 = s/6 \) such that \( H \circ \Phi = h + g_1 + f_1 \) with

\[
\|g_1 - g_0\|_{D,r_1,s_1} \leq C_3 \frac{K}{\alpha r} \epsilon^2, \quad \|f_1\|_{D,r_1,s_1} \leq e^{-Ks/6} \epsilon,
\]

where \( g_0 = P_\Lambda T_K f \) and \( P_\Lambda g = g \). Moreover, \( \|\Pi_I \Phi - I\| \leq C_3 \frac{K}{\alpha} \epsilon \) uniformly on \( V_{r_1,s_1}D \), where \( \Pi_I \) denote the projection \((\varphi, I) \mapsto I\).

We apply Lemma 3.1 to the fully non-resonant case \( \Lambda = \{0\} \), then \( g_0, g \) depends only on \( I \).

**Corollary 3.2.** Assume the standing assumptions for \( h \), and let \( \alpha(\epsilon), r(\epsilon), K(\epsilon) \) be chosen as in [2].

Write \( r_1(\epsilon) = r(\epsilon)/2, r_2(\epsilon) = r(\epsilon)/4 \) and \( s_1 = s_0/6 \). Then there exists \( \epsilon_0 = \epsilon_0(\mathcal{M}), \beta_0(\mathcal{M}) \) such that if \( \epsilon < \epsilon_0 \) and \( \beta < \beta_0 \), there exists

\[
\Phi : V_{r_1(\epsilon),s_1}D(\epsilon) \to V_{r(\epsilon),s_0}D(\epsilon),
\]

such that \( H \circ \Phi = h_1 + f_1 \), with

1. \( h_1 \) is \( l/2, m/2 \)-quasi-convex on \( U_{r_2(\epsilon)}D(\epsilon) \).
2. \( |\nabla h_1(I)|, |\nabla^2 h_1(I)| \leq 2M \) for all \( I \in U_{r_2(\epsilon)}D(\epsilon) \).
3. \( |f_1|_{D(\epsilon),r_1(\epsilon),s_1} \leq \epsilon^{1+N} \).
4. \( |\Pi_I \Phi - I| \leq \epsilon^2 \).

**Proof.** Let \( D = D(\epsilon), r = r(\epsilon), \alpha = \alpha(\epsilon), K = K(\epsilon), \Lambda = \{0\} \). Then for \( \beta < C_3^{-1} \),

\[
C_3^{-1} \frac{\alpha r}{K} = C_3^{-1} \beta^{-1} r^2 = C_3^{-1} \beta^{-3} \epsilon > \epsilon, \quad C_3^{-1} \frac{\alpha}{K} = C_3^{-1} \beta^{-1} r > r.
\]

Therefore, Lemma 3.1 applies. It follows that \( H \circ \Phi = h + g_1 + f_1 \), with

\[
\|g_1 - g_0\|_{D,r_1,s_1} \leq C_3 \beta r^{-2} \epsilon^2 = C_3 \beta^3 \epsilon < \epsilon, \quad \|f_1\|_{D,r_1,s_1} \leq e^{-Ks/6} \epsilon = \epsilon^{N+1}.
\]
Since $\Lambda$ is the trivial module, $g_1, g_0$ depends only on $I$, and $\|g_0\|_{D,r_1,s_1} \leq \|f\|_{D,r,s_0} = \epsilon$. Define

$$h_1 = h + g_1,$$

using Cauchy estimates we have

$$\|\nabla h_1 - \nabla h\|_{U_{r_2}} \leq (r_1/2)^{-1} \|g_1\|_{U_{r_1}} \leq 8\beta \epsilon^{\frac{1}{2}} < 8\epsilon^{\frac{1}{2}},$$

$$\|\nabla^2 h_1 - \nabla^2 h\|_{U_{r_2}} \leq (r_1/2)^{-2} \|g_1\|_{U_{r_1}} \leq 32\beta^2.$$

Choose $\epsilon_0, \beta_0$ such that

$$8\epsilon^{\frac{1}{2}} < \min\{M, l/2\}, \quad 32\beta^2 < \min\{m/2, M\}$$

Then $\|\nabla h_1\|, \|\nabla^2 h_1\| \leq 2M$ on $U_{r_2}$. To prove quasi-convexity, note that one of the following holds for all $\|v\| = 1$:

$$\|\nabla^2 h(I)v \cdot v\| \geq m, \text{ or } \|\nabla h(I) \cdot v\| > l.$$

Our estimates imply one of the following always hold:

$$\|\nabla^2 h_1(I)v \cdot v\| \geq m/2, \text{ or } \|\nabla h_1(I) \cdot v\| > l/2,$$

implying $l/2, m/2$-semi-concavity. \hfill \Box

We then apply the following global stability theorem, which we apply to the normal form system. It’s important to note that $h_1$ does not satisfy our standing assumption, and special care needs to paid to which parameters the constants depends on.

**Theorem 3.3** ([12], Theorem 1). Suppose $H = h_1 + f_1 \in A_{r,s}D$, $h_1$ is $l, m$-quasi-convex and

$$\|\nabla^2 h_1(I)\|_{D,r,s} \leq M.$$ 

There is $C_4 > 1$ depending on $s, l, m, M$ such that the following hold. For $r \leq s$, let $f_1 \in A_{r,s}(D)$ satisfy

$$\|f_1\|_{D,r,s} \leq \epsilon \leq \epsilon_0 = C_s^{-1} r^2.$$

Then for every orbit of $H$ with $(\theta(0), I(0)) \in \mathbb{T}^n \times D$, one has

$$\|I(t) - I_0\| \leq C_4r \left(\frac{\epsilon}{\epsilon_0}\right)^{\frac{1}{2n}}, \quad \text{for } |t| \leq C_4^{-1} \exp\left(C_4^{-1} \left(\frac{\epsilon_0}{\epsilon}\right)^{\frac{1}{2n}}\right).$$
Proof of Proposition 2.2. Let \( r_3 = r_3(\epsilon) = r_2(\epsilon)/2 \).

Let \( (\theta, I)(t) \) be an orbit of \( H \) with \( I(0) \in D(\epsilon) \), then \( (\theta', I')(t) = \Phi^{-1}(\theta, I)(t) \) is an orbit of \( H \circ \Phi \) as long as \( I'(t) \in U_{r_1,s_1}D(\epsilon) \). Note that according to item 4 of Corollary 3.2, \( |I'(0) - I(0)| < \epsilon^{\frac{1}{2}} < r_3 = \beta^{-1}\epsilon^{\frac{1}{2}}/8 \), so \( I'(0) \in U_{r_3}D \).

Consider \( H \circ \Phi = h_1 + f_1 \), \( D = U_{r_3}D(\epsilon) \), then Theorem 3.3 applies with parameters \( r_3, s_1 \) since \( \epsilon_1 = \epsilon^{N+1} \leq \epsilon_0 = C_4^{-1}r_3^2 = (64C_4)^{-1}r^2 = (64C_4)\beta^{-2}\epsilon \).

Therefore
\[
\|I'(t) - I'(0)\| \leq C_4r_3 \left( \frac{\epsilon_1}{\epsilon_0} \right)^{\frac{1}{2n}} \leq C_4\beta^{-1}\epsilon^{\frac{1}{2}} \left( \frac{\epsilon_1^{1+N}}{\epsilon} \right)^{\frac{1}{2n}} \leq C_4\beta^{-1}\epsilon^{\frac{1}{2}} + \frac{\epsilon}{2n},
\]

since \( \epsilon_0 > \epsilon \); for the time interval
\[
|t| \leq C_4^{-1}\exp \left( C_4\epsilon^{\frac{N}{2n}} \right) \leq C_4^{-1}\exp \left( C_4^{-1} \left( \frac{\epsilon_0}{\epsilon_1} \right)^{\frac{1}{2n}} \right),
\]

which includes the time interval
\[
|t| \leq C_4^{-1}\exp \left( C_4^{-1}\epsilon^{-\frac{N}{2n}} \right).
\]

Using \( I'(t) \in U_{r_1,s_1}D(\epsilon) \), and \( |I(t) - I'(t)| < \epsilon^{\frac{1}{2}} \) we obtain our proposition. \( \square \)

4. Stability near strong 1-resonances

Suppose \( \Lambda \subset \mathbb{Z}^n \) is a maximal submodule, and let \( k_1, \ldots, k_d \in \mathbb{Z}^n \) be linearly independent and generates \( \Lambda \) over \( \mathbb{Z} \). The volume \( |\Lambda| \) of \( \Lambda \) is defined as
\[
|\Lambda|^2 = \det \begin{bmatrix} k_1^T \\ \vdots \\ k_d^T \end{bmatrix} \begin{bmatrix} k_1 & \cdots & k_d \end{bmatrix}.
\]

This definition is independent of the basis \( k_1, \ldots, k_d \). \( \Lambda \) is called a \( K \)-lattice if \( |k_1|, \ldots, |k_d| \leq K \).

**Theorem 4.1** ([12], Theorem 3). Suppose \( h \) satisfies the standing assumption, and consider a \( K_\Lambda \)-lattice \( \Lambda \) of dimension \( d \). Then there exist \( C_5 = C_5(\mathcal{M}) > 1 \) such that if
\[
|f|_{B,r,s} \leq \epsilon \leq \frac{\epsilon_\Lambda}{K_\Lambda^{2(n-\delta)}}, \quad \epsilon_\Lambda = C_5^{-1}|\Lambda|^{-2},
\]
Figure 2. Any curve of sufficient length in $\mathcal{N}(\epsilon)$ must pass close to a double resonance

where $|\Lambda|$ is the volume of $\Lambda$. Then for every orbit $(\theta, I)(t)$ such that

$$d(\omega(I(0)), R_{\Lambda}) < C_5^{-1}\sqrt{\epsilon},$$

one has

$$\|I(t) - I_0\| \leq C_5 r\left(\frac{1}{\epsilon_{\Lambda}}\right)^{\frac{1}{2(n-d)}}, \text{ for } |t| \leq C_5^{-1}\exp\left(\frac{C_5^{-1}}{\epsilon_{\Lambda}}\right).$$

The stability in the resonant area follows by two steps. First, by geometric consideration, we show that any orbit which drifts a large enough distance, in the neighborhood of strong $1$-resonance must be close to a $2$-resonance $R_{k_1,k_2}$ with estimates on $|k_1|, |k_2|$. We then apply Theorem 4.1.

**Lemma 4.2.** Let $(\theta, I)(t)$ be an orbit of $H$ with $\|I(T) - I(0)\| > \epsilon^\delta$, and $I(t) \in \mathcal{N}(\epsilon)$ for all $t \in [0,T]$. Then there exists $C_6 = C_6(M)$, $t_* \in [0,T]$ and

$$k_1, k_2 \in \mathbb{Z}^n \setminus \{0\}, \quad |k_1| \leq K, \quad |k_2| \leq C_6\epsilon^{-\delta},$$

such that

$$d(\omega(I(t_*)), R_{k_1,k_2}) < C_6\beta^{-2}\epsilon^{\frac{1}{2}-\delta}K^2(\epsilon).$$

First we have the following lemma, which is a modified version of Lemma 3.4 from [4].

**Lemma 4.3.** Let $I \subset [-1,1]$ be a closed interval of length $l > 0$. Suppose $0 < K^2 < 2l^{-1}$, then there is an irreducible rational number $p/q \in I \cap \mathbb{Q}$ such that

$$K < q < 3l^{-1}.$$
Proof. Let $Q = \lfloor 3q^{-1} \rfloor > 2q^{-1}$, then there is $m \in \mathbb{Z}$ such that $m, \frac{m+1}{q} \in I$. We now show at least one of them satisfies the conclusion of the lemma. Indeed, if $\frac{p_1}{q_1}, \frac{p_2}{q_2} \in \mathbb{Q}$ are distinct and $|q_1|, |q_2| \leq K$, then $\frac{|p_1 - p_2|}{|q_1q_2|} > K^{-2} > Q^{-1}$, therefore at most one of $\frac{m}{q}$ and $\frac{m+1}{q}$ can have denominator bounded by $K$ when reduced. \hfill \square

Proof of Lemma 4.2. The proof is inspired by Lemma 3.3 of [4]. Consider the map

$$\Psi_h : (I, \lambda) \mapsto (h(I), \lambda \omega(I)),$$

then $\Psi_h$ is a local diffeomorphism. Therefore, there exists $\rho_0, C > 0$ depending on $M$ such that

$$|\Psi_h(I_1, \lambda_1) - \Psi_h(I_2, \lambda_2)| \geq C^{-1} |(I_1 - I_2, \lambda_1 - \lambda_2)|, \quad \text{if} \quad |(I_1 - I_2, \lambda_1 - \lambda_2)| < \rho_0.$$

Suppose $\epsilon_0$ is small enough that $\epsilon_0^2 < \rho_0$. Write $\omega(t) = \omega(I(t))$ and $t_0$ be the first time the curve $(I(t), |\omega(t)|^{-1})$ leaves the $\epsilon^2$ neighborhood of $(I(0), |\omega(0)|^{-1})$, with $0 < \epsilon < \epsilon_0$. Then the above observation implies

$$\left| \frac{\omega(t_0)}{|\omega(t_0)|} - \frac{\omega(0)}{|\omega(0)|} \right| \geq C^{-1} \epsilon^2.$$

Since energy conservation implies $|h(I(t_0)) - h(I(0))| < 2\epsilon$, we obtain

$$\left| \frac{\omega(t_0)}{|\omega(t_0)|} - \frac{\omega(0)}{|\omega(0)|} \right| > C^{-1} \epsilon^2,$$

which implies for some index $i \in \{1, \cdots, n\}$, the interval $\{\omega_i(t)/|\omega(t)| : t \in [0, t_0]\} \subset [-1, 1]$ has length at least $C^{-1} \epsilon^2$. Then according to Lemma 4.3 there exists irreducible $p/q \in \mathbb{Q}$ with $K < |q| < 3C\epsilon^{-2}$ and $t_* \in [0, t_0]$, such that $\omega_i(t_*)/|\omega(t_*)| = p/q$. Let $j \in \{1, \cdots, n\}$ such that $\omega_j(t_*)/|\omega(t_*)| = 1$, (which exists since $|\omega| = \sup_j |\omega_j|$), it follows that for $k_2 = qe_i + pe_j \in \mathbb{Z}^n$, where $e_i$ denotes the coordinate vectors, we have

$$k_2 \cdot \omega(t_*) = 0, \quad |k_2| \leq 6C\epsilon^{-2}.$$

Moreover, since $|q| > K$ and $k_2$ is irreducible, $k_2$ cannot be generated by any vector with $|k| \leq K$, therefore $\{k_1, k_2\}$ is linearly independent.

Since $I(t_*) \in \mathcal{N}(\epsilon)$, there exists $0 < |k_1| \leq K$ such that dist $(\omega(t_*) , R_{k_1}) < \alpha(\epsilon)$. Let $\tilde{\omega}$ be the projection of $\omega(t_*)$ to the hyperplane $R_{k_1} \cap R_{k_2}$, we first note

$$\sin \angle(R_{k_1}, R_{k_2}) = \sin \angle(k_1, k_2) = \frac{\sqrt{||k_1||^2 ||k_2||^2 - (k_1 \cdot k_2)^2}}{||k_1|| ||k_2||} \geq \frac{1}{||k_1|| ||k_2||},$$
then
\[ |\omega(t_s) - \bar{\omega}| \leq \frac{d(\omega(t_s), R_{k_1})}{\sin \angle(R_{k_1}, R_{k_1})} \leq \alpha(\epsilon) |k_1||k_2| \leq 6C\alpha(\epsilon)\epsilon^{-\delta}K(\epsilon) \]
and the lemma follows from taking \( C_6 = 6C \), and plugging in \( \alpha(\epsilon) = \beta^{-1}r(\epsilon)K(\epsilon) \) and \( r(\epsilon) = \beta^{-1}\epsilon^{\frac{1}{2}} \).

According to our definition, \( R_{k_1, k_2} \) is generated by the module \( \text{Span}_{\mathbb{Z}}\{k_1, k_2\} \), which is not necessarily maximal. In order to apply Theorem 4.1, we need the following lemma.

**Lemma 4.4.** Suppose \( \Lambda \) is the maximal module containing \( k_1, k_2 \), namely
\[ \Lambda = \text{Span}_{\mathbb{R}}\{k_1, k_2\} \cap \mathbb{Z}^n \]
where \( \{k_1, k_2\} \) is linearly independent. Then \( \Lambda \) is a \(|k_1| + |k_2|\)-lattice, and \(|\Lambda| \leq |k_1||k_2|\).

**Proof.** The lemma is non-trivial because \( k_1, k_2 \) does not necessarily generate \( \Lambda \) over \( \mathbb{Z} \).

We first derive a relation for arbitrary number of generators. Suppose \( \Lambda \) is the maximal module containing \( k_1, \cdots, k_d \), let \( k'_1, \cdots, k'_d \) generate \( \Lambda \) over \( \mathbb{Z} \). Let \( A \) be the matrix with columns \( k_1, \cdots, k_d \), and \( B \) with columns \( k'_1, \cdots, k'_d \). Then there exist invertible \( d \times d \) integer matrix \( G \) such that
\[ A = BG. \]

Then
\[ \det(A^T A) = \det(G^T B^T BG) = \det(G^T) \det(B^T B) \det(G) \geq \det(B^T B) = |\Lambda|^2. \]

We now go to the case \( d = 2 \), we have \(|\Lambda|^2 = \|k_1\|^2\|k_2\|^2 - (k_1 \cdot k_2)^2 \leq \|k_1\|^2\|k_2\|^2\), and the estimate follows from \( \|k\| \leq |k| \).

To prove \( \Lambda \) is a \(|k_1| + |k_2|\) lattice, we claim there exists \( k'_1, k'_2 \) generating \( \Lambda \) with \(|k'_1|, |k'_2| \leq |k_1| + |k_2|\). The argument presented here is based on the more general argument in [13], Theorem 18. Define
\[ s_2 = \min\{t_2 > 0 : \mathbb{R}k_1 + t_2k_2 \cap \Lambda \neq \emptyset\}, \quad s_1 = \min\{t_1 \geq 0 : t_1k_1 + s_2k_2 \in \Lambda\}, \]
and \( k'_2 = s_1k_1 + s_2k_2 \). We now show \( k'_1, k'_2 \) generates \( \Lambda \) over \( \mathbb{Z} \). For any \( k \in \Lambda \), there exists \( t_1, t_2 \in \mathbb{R} \) such that \( k = t_1k'_1 + t_2k'_2 \). Assume that \( t_2 \notin \mathbb{Z} \), then there exists \( n \in \mathbb{Z} \) such that \( 0 < a = t_2 + n < 1 \). We have
\[ k + nk'_2 = t_1k'_1 + ak'_2 = (t_1 + as_1)k_1 + as_2k_2 \in \Lambda. \]
Since $0 < as_2 < s_2$, this contracts with the minimality of $s_2$. As a result $t_2 \in \mathbb{Z}$. We can show $t_1 \in \mathbb{Z}$ by the same argument. Since $0 \leq s_1 < 1$ and $0 < s_2 \leq 1$ by definition, we know $|k_2'| < |k_1| + |k_2|$.

**Proof of Proposition 2.5** Suppose $(\theta, I)(t), t \in [0, T]$ is an orbit satisfying $I(t) \in \mathcal{N}(\epsilon)$ for all $t$. Arguing by contradiction, suppose $|I(t) - I(0)| > \epsilon\delta$ for some $t \in [0, T]$. We apply Lemma 4.2 to obtain that there exists $t_* \in [0, T]$, and $|k_1| \leq K(\epsilon), |k_2| \leq C\epsilon^{-\delta}$, such that

$$d(\omega(t_*), R_{k_1,k_2}) < C_6\beta^{-2}\epsilon^{\frac{1}{2} - \delta}K^2(\epsilon).$$

We will pick $\epsilon_0$ depending on $\delta$ such that for all $\epsilon < \epsilon_0$, we have

$$K(\epsilon) = -L \log \epsilon \leq \epsilon^{-\delta}.$$

Let $\Lambda = \text{Span}_{\mathbb{R}} \{k_1, k_2\} \cap \mathbb{Z}^n$ be the maximal lattice generated by $k_1, k_2$. According to Lemma 4.4

$$K_{\Lambda} \leq K(\epsilon) + C_6\epsilon^{-\delta} \leq 2C_6\epsilon^{-\delta}, \quad 1 \leq |\Lambda| \leq C_6K(\epsilon)\epsilon^{-\delta} \leq C_6\epsilon^{-2\delta}.$$

We attempt to apply Theorem 4.1 near the resonance $R_{\Lambda}$. Set

$$\epsilon_2 = C_5 \left( C_6\beta^{-2}\epsilon^{\frac{1}{2} - \delta}K^2 \right) = C_5C_6^2\beta^{-4}\epsilon^{1-2\delta}K^4 \leq C_5C_6^2\beta^{-4}\epsilon^{1-6\delta}.$$

Then $d(\omega(t_*), R_{k_1,k_2}) < C_5^{-1}\sqrt{\epsilon_2}$. Then

$$\epsilon_{\Lambda} = C_5^{-1}|\Lambda|^{-2} \geq (C_5C_6)^{-1}\epsilon^{2\delta}, \quad \frac{\epsilon_{\Lambda}}{K_{\Lambda}^{2(n-2)}} \geq \frac{(C_5C_6)^{-1}\epsilon^{2\delta}}{(2C_6)^{2n-2}\epsilon^{-\delta(2n-4)}} \geq C_7^{-1}\epsilon^{(2n-2)\delta}$$

for $C_7 = (C_5C_6)(2C_6)^{2n-2}$. Then if $(2n+4)\delta < 1$ and $\epsilon$ small enough depending on $C_5, C_6, C_7, \beta$ and $\delta$, we have

$$\epsilon_2 \leq C_5C_6^2\beta^{-4}\epsilon^{1-6\delta} < C_7^{-1}\epsilon^{(2n-2)\delta} \leq \frac{\epsilon_{\Lambda}}{K_{\Lambda}^{2(n-2)}},$$

and

$$\epsilon_2/\epsilon_{\Lambda} \leq C_5C_6^2\beta^{-4}\epsilon^{1-8\delta}.$$
for $C_2 = C_5 (C_6 C_0)^{1 \over 2n - 4}$, as long as

$$|t - t_*| \leq C_5^{-1} \exp \left( C_5^{-1} \left( \frac{\epsilon A}{\epsilon_2} \right)^{1 \over 2n - 4} \right)$$

which is implied by

$$|t - t_*| \leq C_2^{-1} \exp \left( C_2^{-1} \epsilon^{1 \over 2n - 4} \frac{1-8\delta}{2n-4} \right).$$

This in particular would imply $|I(T) - I(0)| \leq |I(T) - I(t_*)| + |I(0) - I(t_*)| \leq 2C_2 \epsilon^{1 \over 2n - 4}$. Since $(2n + 4)\delta < 1$, then $\frac{1-8\delta}{2n-4} > \delta$, therefore for $\epsilon$ small enough $|I(T) - I(0)| \leq 2C_2 \epsilon^{1 \over 2n - 4} < \epsilon^\delta$ which is a contradiction. □

Acknowledgments. The authors thank Abed Bounemoura for valuable discussions. K. Zhang was supported by the NSERC Discovery grant, reference number 436169-2013.

References

[1] V. I. Arnol’d. Instability of dynamical systems with many degrees of freedom. Dokl. Akad. Nauk SSSR, 156:9–12, 1964.
[2] Ugo Bessi. An approach to Arnol’d’s diffusion through the calculus of variations. Nonlinear Anal., 26(6):1115–1135, 1996.
[3] Ugo Bessi. Arnold’s example with three rotators. Nonlinearity, 10(3):763–781, 1997.
[4] Abed Bounemoura and Jean-Pierre Marco. Improved exponential stability for near-integrable quasi-convex Hamiltonians. Nonlinearity, 24(1):97, 2011.
[5] Abed Bounemoura and Laurent Niederman. Generic Nekhoroshev theory without small divisors. Ann. Inst. Fourier (Grenoble), 62(1):277–324, 2012.
[6] P. Lochak and A. I. Ne˘ıshhtadt. Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian. Chaos, 2(4):495–499, 1992.
[7] Pierre Lochak and Jean-Pierre Marco. Diffusion times and stability exponents for nearly integrable analytic systems. Cent. Eur. J. Math., 3(3):342–397 (electronic), 2005.
[8] P. Loshak. Canonical perturbation theory: an approach based on joint approximations. Uspekhi Mat. Nauk, 47(6(288)):59–140, 1992.
[9] N. N. Nekhoroshev. An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. Uspehi Mat. Nauk, 32(6(198)):5–66, 287, 1977.
[10] N. N. Nekhoroshev. An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. II. Trudy Sem. Petrovsk., (5):5–50, 1979.
[11] Laurent Niederman. Prevalence of exponential stability among nearly integrable Hamiltonian systems. Ergodic Theory Dynam. Systems, 27(3):905–928, 2007.
[12] Jürgen Pöschel. Nekhoroshev estimates for quasi-convex Hamiltonian systems. Math. Z., 213(2):187–216, 1993.
[13] Carl Ludwig Siegel. *Lectures on the geometry of numbers*. Springer-Verlag, Berlin, 1989. Notes by B. Friedman, Rewritten by Komaravolu Chandrasekharan with the assistance of Rudolf Suter, With a preface by Chandrasekharan.

[14] Ke Zhang. Speed of Arnold diffusion for analytic Hamiltonian systems. *Invent. Math.*, 186(2):255–290, 2011.

† Institute of Theoretical Studies, ETH Zürich, CH-8092 Zürich, Switzerland

E-mail address: jianlu.zhang@math.ethz.ch

‡ Department of Mathematics, University of Toronto, Toronto, Canada

E-mail address: kzhang@math.toronto.edu