Nekhoroshev theory and discrete averaging

V. Gelfreich[1] and A. Vieiro[2,3]

[1] Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

v.gelfreich@warwick.ac.uk

[2] Departament de Matemàtiques i Informàtica,

Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain

[3] Centre de Recerca Matemàtica (CRM),

Campus Bellaterra, 08193 Bellaterra, Spain

vieiro@maia.ub.es

October 11, 2024

Abstract

This paper contains a proof of the Nekhoroshev theorem for quasi-integrable symplectic maps. In contrast to the classical methods, our proof is based on the discrete averaging method and does not rely on transformations to normal forms. At the centre of our arguments lies the theorem on embedding of a near-the-identity symplectic map into an autonomous Hamiltonian flow with exponentially small error.

Keywords— Symplectic maps, Hamiltonian systems, Nekhoroshev theorem, stability, near-the-identity maps, discrete averaging

1 Introduction

If a quasi-integrable Hamiltonian system satisfies suitable non-degeneracy assumptions, the KAM theorem states that invariant tori occupy the major part of the phase space and the Lebesgue measure of the complement to the set of the tori converges to zero when a perturbation parameter vanishes [2, 6, 19, 24]. This complement is dense in the phase space. Moreover, if the number of degrees of freedom is three or larger, the complement is a connected set. Therefore, in contrast to the case of two degrees of freedom, the KAM theory does not prevent existence of trajectories which may travel large distances inside an energy level. This phenomenon is known under the name of Arnold diffusion [1]. The maximal speed of Arnold diffusion is bounded by Nekhoroshev estimates [21], which state that action variables oscillate near their initial values for exponentially long times.

Hamiltonian perturbation theory studies both systems of Hamiltonian equations and symplectic maps. A $2d$-dimensional symplectic quasi-integrable map can be seen as an isoenergetic Poincaré section of a Hamiltonian flow with $d+1$ degrees of freedom [13], and a Nekhoroshev theorem for maps is usually derived from a Nekhoroshev theorem for flows, for example, from the results of [15], although direct proofs for maps are also available (see e.g. [11]).

In the literature one can find two different strategies for proving a Nekhoroshev type estimate. The first strategy, originally proposed by Nekhoroshev [21, 22], relies on a careful study of normal forms for resonances of all possible multiplicities. An alternative strategy was proposed by Lochak [14] who showed that the analysis can be restricted to resonances of the highest multiplicity only. In the convex case both strategies lead to optimal stability coefficients [15, 23].

In this paper we provide a new proof of the Nekhoroshev theorem for quasi-integrable symplectic analytic maps under the convexity assumption, without reducing the problem to a Hamiltonian flow. A part of our proof follows the Lochak-Neishtadt approach [15] but, in contrast to the traditional approach, our method does not rely on transformations to normal forms. Instead we propose a direct reconstruction of slow observables from iterates of the quasi-integrable map in original coordinates with the help of a discrete averaging method. Our construction can be applied to an individual map and provides explicit values for constants in the estimates. Therefore, our method can be used for numerical analysis of Arnold diffusion and for developing new tools for visualisation of the dynamics [10].

Our proof of the Nekhoroshev theorem includes a refined version of classical Neishtadt's theorem [20] which may be of independent interest. Neishtadt proved that if a member of a smooth near-the-identity family of analytic symplectic maps is sufficiently close to the identity, then it can be approximated by an autonomous Hamiltonian flow with an exponentially small error. Neishtadt's proof relies on representing the maps as time-one maps of time-periodic flows and classical averaging. Alternatively the theorem can be proved using Moser's analysis of the formal interpolating flow [4, 18]. By contrast, our construction is applicable to individual maps. We will show that the approximation error can be controlled by the ratio of two natural parameters: one characterises the size of a complex neighbourhood where the map is close to the identity and the second one is the distance from the map to the identity in a suitably chosen norm. Our construction is based on the discrete averaging and has an additional important advantage: the construction is explicit in terms of iterates of the map and can be easily implemented numerically [10].

The paper is organized in the following way. Section 2 presents the statement of the Nekhoroshev theorem and explains its derivation from the analysis of stability near fully resonant tori. Section 3 contains a proof of some elementary bounds and explains the strategy of our proof. Section 4 contains the necessary details of the theory of interpolating vector fields. In Section 5 we prove our refined version of Neishtadt's theorem, the embedding theorem of a near-the-identity map into an autonomous Hamiltonian flow up to an exponentially small error. The exponential bounds for stability time are proved in Section 6. In this way Sections 4, 5 and 6 represent a self-contained proof of the Nekhoroshev theorem stated in Section 2. Finally, in Section 7 we consider a small region around a resonance, that we called a nucleus of resonance, and discuss some improved bounds for stability times in these regions of the phase space. In particular we discuss how the stability times scale when ε approaches zero. Our final comments and conclusions are given in Section 8.

2 Nekhoroshev theorem for a near-integrable map

The main object of this paper is a one-parameter family $F_\varepsilon : (I,\varphi) \mapsto (\bar I, \bar\varphi)$ of real-analytic exact symplectic maps of the form

$$\begin{cases} \bar I = I + \varepsilon a(I,\varphi), \\ \bar\varphi = \varphi + \omega(I) + \varepsilon b(I,\varphi) \pmod 1, \end{cases} \tag{1}$$

where the functions a, b are periodic in $\varphi \in \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, I is a vector in $\mathbb{R}^d$ and $\varepsilon \geq 0$ is a perturbative parameter. At $\varepsilon = 0$ the map is integrable and $F_0 : (I,\varphi) \mapsto (\bar I, \bar\varphi)$ is simply given by

$$\begin{cases} \bar I = I, \\ \bar\varphi = \varphi + \omega(I) \pmod 1. \end{cases} \tag{2}$$

Obviously the variable I remains constant along trajectories of F_0 and the equation $I = I_0$ defines an invariant torus with the frequency vector $\omega(I_0)$. Trajectories of F_ε with very close initial conditions are capable of separating from each other exponentially fast with stability times $T_L \sim \varepsilon^{-1/2}$, a natural Lyapunov time for the system. The Nekhoroshev estimates address substantially longer timescales. Given an initial condition (I_0, φ_0) let $(I_k, \varphi_k) = F_\varepsilon^k(I_0,\varphi_0)$, $k \in \mathbb{Z}$, be the corresponding trajectory. A Nekhoroshev estimate states that, for $|\varepsilon| \leq \varepsilon_0$,

$$|I_k - I_0| \leq R(\varepsilon) \quad \text{for } |k| \leq T_\varepsilon,$$

where the radius of confinement $R(\varepsilon) = \mathcal{O}(\varepsilon^\beta)$ and the stability time $T_\varepsilon \sim \exp(c/\varepsilon^\alpha)$, for suitable stability exponents $0 < \alpha, \beta \leq 1$.

It is well known that the long time behaviour of trajectories is sensitive to the smoothness class of the map F_ε [3, 7, 8, 17]. In this paper we assume that F_ε is real-analytic although some of our arguments are applicable to other smoothness classes. Without loosing in generality we assume that F_ε is real-analytic in $B_R \times \mathbb{T}^d$, where $B_R \subset \mathbb{R}^d$ is a ball of radius $R > 0$ centered at some point of $\mathbb{R}^d$. We assume that for all ε with $0 \le \varepsilon \le \varepsilon_0$, the lift of the map have an analytic continuation onto the complex neighbourhood $\mathcal{D}_F$ of $B_R \times \mathbb{R}^d$ given by

$$\mathcal{D}_F = \left\{ \, (I, \varphi) \in \mathbb{C}^{2d} \ : \ \operatorname{dist}(I, B_R) \le \sigma, \ |\operatorname{Im}(\varphi)| \le r \, \right\} \tag{3}$$

for some $\sigma, r > 0$ independent of ε. We hide the dependence of a and b on ε to simplify notations. Our proof is not sensitive to the nature of this dependence and it is sufficient to assume that supremum norms of a and b are bounded uniformly in ε.

It is also well known that the long time stability of action variables I depends on the properties of the unperturbed frequency map ω [4, 12, 21, 22]. Since the map is symplectic, ω is a gradient of a scalar function. We assume that $\omega = h_0'$ where the function h_0 is strongly convex on the intersection of $\mathcal{D}_F$ with the real subspace, i.e. there is $\nu > 0$ such that for every real I in the domain of h_0 one has

$$h_0''(I)(v, v) = (\omega'(I)v) \cdot v \ge \nu \, v \cdot v \tag{4}$$

for all $v \in \mathbb{R}^d$. The convexity assumption means that ν is a lower bound for the spectrum of the Hessian matrices $h_0''(I)$ for real values of I. Note that a function h_0 is strongly convex on a convex set iff for any I, J

$$(h_0'(I) - h_0'(J)) \cdot (I - J) \ge \nu \, (I - J) \cdot (I - J) \tag{5}$$

with the same convexity constant ν.

Under these assumptions we will prove the Nekhoroshev estimates with optimal exponents $\alpha = \beta = \frac{1}{2(d+1)}$.

Theorem 2.1 (Nekhoroshev theorem)**.** *If a real-analytic exact symplectic map F_ε satisfies the assumptions stated above and h_0 is strongly convex on a real neighbourhood of B_R, then there are positive constants c_1, c_2, c_3 such that for every initial condition $(I_0, \varphi_0) \in B_R \times \mathbb{T}^d$ one has*

$$|I_k - I_0| < c_1 \varepsilon^{1/2(d+1)} \quad \textit{for} \quad 0 \le k \le T_\varepsilon = c_2 \exp\left(c_3 \varepsilon^{-1/2(d+1)} \right).$$

We derive the Nekhoroshev theorem from a statement which provides more detailed information about the stability of actions.

Theorem 2.2 (long term stability of actions)**.** *Under the assumptions of the Nekhoroshev theorem, there are constants $\gamma_0 > 0$ and $r_0 \in (0, 1)$ with the following property. For every $\gamma \ge \gamma_0$ there are positive constants ε_0, c_2, c_3 such that* *if* $0 < \varepsilon \le \varepsilon_0$*,* $n < \varepsilon^{-d/2(d+1)}$*, and* $I_* \in B_R$ *corresponds to a fully resonant unperturbed torus with* $n\omega(I_*) \in \mathbb{Z}^d$*, then any trajectory with initial conditions satisfying* $|I_0 - I_*| < r_0\rho_n$ *and* $\varphi_0 \in \mathbb{T}^d$ *satisfies the inequality*

$$|I_k - I_*| < \rho_n \qquad \text{for} \quad 0 \le k \le T_\varepsilon = c_2 \exp\left(c_3 \varepsilon^{-1/2(d+1)}\right),$$

where $\rho_n = \gamma n^{-1} \varepsilon^{1/2(d+1)}$*.*

Note that we will provide explicit expressions for γ_0 and r_0.

In order to derive Theorem 2.1 from Theorem 2.2 we use the ideas of Lochak covering to show that the balls $|I_0 - I_*| < r_0\rho_n$ cover all initial conditions. Our arguments use the Dirichlet Theorem on simultaneous approximations in a way similar to the papers [14, 15, 16].

Theorem 2.3 (Dirichlet [9])**.** *For any* $\omega \in \mathbb{R}^d$ *and any* $N > 1$ *there are* $\omega_* \in \mathbb{Q}^d$ *and* $n \in \mathbb{N}$ *such that* $n < N$*,* $n\omega_* \in \mathbb{Z}^d$ *and* $|\omega - \omega_*| < \frac{1}{nN^{1/d}}$*.*

If the frequency map $\omega : I \mapsto h_0'(I)$ is defined by a strongly convex function h_0, we can prove a similar result in the space of actions.

Lemma 2.4. *Let a convex set* $U_\delta \subset \mathbb{R}^d$ *be a* δ*-neighbourhood of a set* $U \subset \mathbb{R}^d$*. If* h_0 *is strongly convex in* U_δ *with parameter* ν*, then there is* $N_0 = N_0(\nu, \delta)$ *such that for any* $N > N_0$ *and any* $I_0 \in U$ *there is a point* $I_* \in U_\delta$ *and* $n \in \mathbb{N}$ *such that* $n < N$*,* $n\omega(I_*) \in \mathbb{Z}^d$ *and*

$$|I_0 - I_*| < \frac{\sqrt{d}}{\nu n N^{1/d}}.$$

Proof. Using the strong convexity of the function h_0 in the form (5) we get

$$|\omega(I_1) - \omega(I_2)|_2\, |I_1 - I_2|_2 \ge (\omega(I_1) - \omega(I_2)) \cdot (I_1 - I_2) \ge \nu |I_1 - I_2|_2^2$$

for all $I_1, I_2 \in U_\delta$ and consequently $|\omega(I_1) - \omega(I_2)|_2 \ge \nu |I_1 - I_2|_2$. For the sake of convenience we use the Euclidean norm in this bound. It follows that $I_1 \ne I_2$ implies $\omega(I_1) \ne \omega(I_2)$ and consequently the map $\omega : U_\delta \to \omega(U_\delta)$ is bijective.

Now let $I_0 \in U$ and $N \in \mathbb{N}$. The Dirichlet theorem implies that there is $\omega_* \in \mathbb{Q}^d$ such that $n\omega_* \in \mathbb{Z}^d$ for some $n < N$ and $|\omega(I_0) - \omega_*| < n^{-1}N^{-1/d}$. We note that if $|I_0 - I_*| < \delta$ then

$$|\omega(I_0) - \omega(I_*)| \ge \frac{1}{\sqrt{d}} |\omega(I_0) - \omega(I_*)|_2 \ge \frac{\nu}{\sqrt{d}} |I_0 - I_*|_2 \ge \frac{\nu}{\sqrt{d}} |I_0 - I_*|.$$

Consequently, if $n^{-1}N^{-1/d} < \nu d^{-1/2}\delta$ then $\omega_* = \omega(I_*)$ for some I_* with $|I_0 - I_*| < \delta$. Let $N_0^{-1/d} = \nu d^{-1/2}\delta$. Then for any $N > N_0$ there is $I_* \in U_\delta$ such that $n\omega(I_*) \in \mathbb{Z}^d$ for $n < N$ and

$$|I_0 - I_*| < \frac{\sqrt{d}}{\nu} |\omega(I_0) - \omega_*| < \frac{\sqrt{d}}{\nu n N^{1/d}}. \qquad \square$$

We conclude that if $\varepsilon_0^{d/2(d+1)} < 1/N_0$ and $\gamma_0 r_0 > \sqrt{d}\nu^{-1}$, then the balls $B(I_*, r_0\rho_n)$ with n and I_* such that $n < \varepsilon^{-d/2(d+1)}$ and $n\omega(I_*) \in \mathbb{Z}^d$ cover B_R. Consequently every initial condition belongs to a neighbourhood of a resonant torus $I = I_*$ where the stability bounds of Theorem 2.2 are applicable and Theorem 2.1 follows immediately. We note that there is no claim of uniqueness for I_* and some initial conditions may belong to several zones of stability.

3 A priori bounds and strategy of the proof

The n^{th} iterate of the map F_ε can be written explicitly in the form

$$\begin{cases} I_n = I_0 + \varepsilon \sum_{k=0}^{n-1} a(I_k, \varphi_k), \\ \varphi_n = \varphi_0 + \sum_{k=0}^{n-1} \omega(I_k) + \varepsilon \sum_{k=0}^{n-1} b(I_k, \varphi_k), \end{cases} \tag{6}$$

where $(I_k, \varphi_k) = F_\varepsilon(I_{k-1}, \varphi_{k-1})$ denote points on the trajectory with initial conditions (I_0, φ_0). We can slightly overload our notation by assuming that the angle component in this formula is computed without taking the angle modulo one. We hope that this will not create too much confusion as the functions a and b are periodic in φ. The following simple lemma implies that trajectories of F_ε follow rather closely trajectories of the unperturbed integrable map F_0 for times much shorter than $T_L \sim \varepsilon^{-1/2}$.

Lemma 3.1 (a priori bounds)**.** *Suppose that the map F_ε has an analytic continuation onto the complex domain $\mathcal{D}_F$ and $n \in \mathbb{N}$. If $(I_k, \varphi_k) \in \mathcal{D}_F$ for $0 \le k < n$, then*

$$|I_n - I_0| \le C_1 n\varepsilon, \qquad |\varphi_n - \varphi_0 - n\omega(I_0)| \le C_2 n^2 \varepsilon. \tag{7}$$

where $C_1 = \|a\|$ and $C_2 = \frac{1}{2}\|\omega'\|\|a\| + \|b\|$.

Proof. Since the iterates of the initial point belong to $\mathcal{D}_F$ the triangle inequality implies that $|I_n - I_0| \le n\varepsilon\|a\|$. Then

$$\begin{aligned} |\varphi_n - \varphi_0 - n\omega(I_0)| &\le \sum_{k=0}^{n-1} |\omega(I_k) - \omega(I_0)| + n\varepsilon\|b\| \\ &\le \|w'\| \sum_{k=0}^{n-1} |I_k - I_0| + n\varepsilon\|b\| \\ &\le \|w'\| \sum_{k=0}^{n-1} k\varepsilon\|a\| + n\varepsilon\|b\| \le \frac{n^2\varepsilon}{2}\|\omega'\|\,\|a\| + n\varepsilon\|b\| \end{aligned}$$

and the desired estimate follows immediately as $n^2 \ge n$. □

In a way similar to Lochak-Neishtadt's proof of the Nekhoroshev theorem [15], we analyse dynamics in carefully chosen neighbourhoods of unperturbed tori bearing periodic motions. Let $n\,\omega(I_*) \in \mathbb{Z}^d$ for some $n \in \mathbb{N}$ and $I_* \in \mathbb{R}^d$. The equation $I = I_*$ defines a torus filled with periodic orbits of the integrable map F_0. The point I_* corresponds to a resonance of the maximal multiplicity because the set $\{r \in \mathbb{Z}^d : r \cdot \omega(I_*) = 0 \,(\mathrm{mod}\ 1)\}$ contains the d-dimensional sublattice $n\mathbb{Z}^d$.

Since $n\omega(I_*) \in \mathbb{Z}^d$ we can consider another lift of F_ε^n defined by the equation

$$f_\varepsilon^n : (I_0, \varphi_0) \mapsto (I_n, \varphi_n - n\omega(I_*)). \tag{8}$$

Of course, the maps F_ε^n and f_ε^n define the same trajectories when the angle variables are considered modulo one. In order to prove Theorem 2.2 we restrict our attention to $n < N_\varepsilon$ with

$$N_\varepsilon = \varepsilon^{-d/2(d+1)}, \tag{9}$$

and study the dynamics of f_ε^n on the domain

$$\mathcal{D}_0(I_*) = B(I_*, \rho_n) \times \mathbb{R}^d \tag{10}$$

where the radius of the ball $\rho_n = \rho_\varepsilon / n$ with

$$\rho_\varepsilon = \gamma N_\varepsilon^{-1/d} = \gamma \varepsilon^{1/2(d+1)} \tag{11}$$

and γ is a constant independent of n and ε.

For $\varepsilon = 0$ the map takes the form

$$f_0^n : (I, \varphi) \mapsto (I, \varphi + n\omega(I) - n\omega(I_*)).$$

Consequently the set defined by $I = I_*$ consists of fixed points. It is also easy to see that f_0^n coincides with the time-one map of the integrable flow defined by the Hamiltonian function

$$h_n(I) = n\big(h_0(I) - h_0(I_*) - \omega(I_*) \cdot (I - I_*)\big).$$

Using Lemma 3.1 we will check that in $\mathcal{D}(I_*)$, a suitable complex neighbourhood of $\mathcal{D}_0(I_*)$, the lift f_ε^n is close to the identity.

In Section 5 we prove a refined version of Neishtadt's theorem which establishes explicit bounds for the error of approximation of a near-the-identity symplectic map by an autonomous Hamiltonian flow. In Section 6.1 we will check that this theorem can be used to show that f_ε^n is exponentially close to the time-one map of a Hamiltonian flow $\hat{X}_m = \mathrm{J}\nabla H_m$ where J is the standard symplectic matrix. Here the subscript m refers to the fact that H_m is obtained from an interpolating vector field based on m consecutive iterates of the map f_ε^n. We will use $m \sim 1/\epsilon_n$ where ϵ_n is the distance from f_ε^n to the identity map in $\mathcal{D}(I_*)$. Then in Section 6.2 we will show that

$$H_m(I, \varphi) = h_n(I) + w_{n,m}(I, \varphi),$$

where the perturbative term $w_{n,m}$ is negligible for the purpose of the stability analysis. The convexity of h_0 and the Taylor formula imply that

$$\tfrac{1}{2}\nu n\,|I-I_*|^2 \le h_n(I) \le \tfrac{d}{2}\|h''\|n\,|I-I_*|^2.$$

These inequalities imply that there is $r_0 \in (0,1)$ such that if $I \in B(I_*, r_0\rho_n)$ then the corresponding energy level set of H_m is located inside $B(I_*, \rho_n)$. Consequently, the trajectories of the Hamiltonian flow which start in the smaller ball will never leave the larger one. Then we can easily conclude that a trajectory of the map is trapped in such a neighbourhood for exponentially long times as one iterate of f_ε^n changes the energy by an exponentially small quantity. So we will need exponentially many iterates to reach a change in the energy needed to leave the domain. While the action coordinates remain in $B(I_*, \rho_n)$ the oscillations of I will not exceed the diameter $2\rho_n = 2\gamma\varepsilon^{1/2(d+1)}/n$. This completes the sketch of proof of Theorem (2.2).

Remark 3.2. *The choices of N_ε and ρ_ε originate from the following reasoning. The distance to the identity in the angle component of f_ϵ^n is proportional to $n\rho_n = \rho_\varepsilon$. Using Theorem 5.1 we will get interpolation error of the order of $O(\exp(-c/\rho_\varepsilon))$. In order to achieve longer stability times we would like to reduce ρ_ε. Then we also get sharper estimates for changes in actions. On the other hand, there are two factors which limit our ability to decrease ρ_ε.*

(1) In the Hamiltonian H_m the integrable part (represented by h_n which depends on I only) is to dominate $w_{n,m} \sim n\varepsilon$. The integrable part is approximately quadratic in actions, i.e. $h_n \sim n\rho_n^2$. Therefore we need $n\rho_n^2 \gg n\varepsilon$ for all $n < N_\varepsilon$ or equivalently

$$\rho_\varepsilon \gg \varepsilon^{1/2} N_\varepsilon.$$

(2) The sizes of ρ_n are to be sufficiently large to ensure that the balls $B(I_*, \rho_n)$ cover all actions. It is sufficient to assume

$$\rho_\varepsilon \gg N_\varepsilon^{-1/d}.$$

The sharpest bounds are achieved when these two restrictions are of the same order in ε. In particular we can choose $N_\varepsilon = \varepsilon^{-d/2(d+1)}$.

4 Interpolating vector fields

The interpolating vector fields were originally introduced in [10] and used to approximate dynamics of a near-the-identity map by the flow of a vector field X_m obtained by taking a weighted average of several consecutive iterates of the map. In this paper we use a similar construction based on the Newton interpolation scheme which uses the forward orbit $x_0, \ldots, x_m$ for the construction of $X_m(x_0)$.

We will show that this interpolation scheme is sufficiently accurate for our proof of the Nekhoroshev theorem.

Let us describe the construction of an interpolating vector field X_m. Let $U \subset \mathbb{R}^s$ be an open domain, $f : U \to \mathbb{R}^s$ a real analytic function and $m \in \mathbb{N}$. Suppose that there is a subset $U_0 \subset U$ such that $f^k(U_0) \subset U$ for $0 \le k \le m$. Then the iterates $x_k = f^k(x_0)$ are defined for all $k \le m$ and $x_0 \in U_0$. There is a unique polynomial $P_m(t; x_0)$ of degree m in t such that $P_m(k; x_0) = x_k$ for $0 \le k \le m$. The interpolating vector field is defined by

$$X_m(x_0) = \frac{\partial P(0; x_0)}{\partial t}.$$

The polynomial P_m can be obtained with the help of the Newton finite-difference interpolation scheme. Consider the following finite differences:

$$\Delta_0(x) = x, \qquad \Delta_k(x) = \Delta_{k-1}(f(x)) - \Delta_{k-1}(x), \ k \ge 1. \tag{12}$$

Then the Newton interpolating polynomial with equally spaced data points and step $h = 1$ takes the form

$$P_m(t; x_0) = x_0 + \sum_{k=1}^{m} \frac{\Delta_k(x_0)}{k!} t(t-1)\dots(t-k+1).$$

Differentiating $P_m(t; x_0)$ with respect to t at $t = 0$, we get

$$X_m(x_0) = \sum_{k=1}^{m} \frac{(-1)^{k-1}}{k} \Delta_k(x_0). \tag{13}$$

Remark 4.1. *We say that X_m is obtained by application of a* discrete averaging *procedure to the map f as the sum in* (13) *is a weighted average of $x_0, \dots, x_m$. Namely*

$$X_m(x_0) = \sum_{k=0}^{m} p_{mk} f^k(x_0) \tag{14}$$

where the coefficients p_{mk} do not depend on the map and can be found explicitly: p_{m0} is the harmonic number and for $k > 1$

$$p_{mk} = (-1)^{k+1} \frac{m+1-k}{k(m+1)} \binom{m+1}{k}.$$

We skip the derivation of these coefficients.

This construction can be applied not only to a single map f but also to a family of maps. In the case of a tangent to the identity family, we can use the interpolation procedure to recover coefficients of a formal embedding into a formal vector field.

Let us state our claim more formally. Let $U \subset \mathbb{R}^s$ be an open set. Suppose that $f_\mu : U \to \mathbb{R}^s$ is an analytic family tangent to the identity. In other words, the maps are defined on U for $|\mu| \leq \mu_0$ and f_0 is the identity map, $f_0(x) = x$ for all $x \in U$. Since the set U is open, for every $x_0 \in U$ and every $m \in \mathbb{N}$ we can find $\mu_m(x_0) > 0$ such that $x_k = f_\mu^k(x_0) \in U$ for all $|\mu| \leq \mu_m(x_0)$ and $|k| \leq m$. Then the interpolating vector field X_m is an analytic function of x and μ in a neighbourhood of $x = x_0$ and $\mu = 0$. The following lemma shows that the Taylor expansion of the right-hand side of (13) in powers of μ coincides with the formal vector field up to the order m.

Lemma 4.2 (formal interpolation)**.** *There is a unique sequence of analytic functions $g_k : U \to \mathbb{R}^s$, $k \in \mathbb{N}$, such that for every $m \in \mathbb{N}$ and every $x \in U$*

$$f_\mu(x) - \Phi^1_{G_m}(x) = O(\mu^{m+1})$$

where $\Phi^1_{G_m}$ is the time one-map of the vector field $G_m = \sum_{k=1}^m \mu^k g_k$. Moreover, the interpolating vector field (13) *of order m satisfies*

$$X_m(x) = G_m(x) + O(\mu^{m+1}).$$

Proof. Let $x \in U$. Then $f_\mu(x) = x + \sum_{k=1}^\infty \mu^k f_k(x)$. The radius of convergence may depend on x.

For any sequence of coefficients g_k, the time t map of the vector field G_m is an analytic function of μ in a neighbourhood of x provided t is sufficiently small,

$$\Phi^t_{G_m}(x) = \sum_{k=0}^\infty \mu^k a_{m,k}(x,t).$$

The flow is a solution of the initial value problem

$$\partial_t \Phi^t_{G_m}(x) = G_m(\Phi^t_{G_m}(x)), \qquad \Phi^0_{G_m}(x) = x.$$

The initial condition implies $a_0(x,0) = x$. Since the series G_m starts with $k = 1$ the differential equation implies that $\partial a_0(x,t) = 0$ for all t. Consequently, $a_0(x,t) = x$ for all t.

The initial condition implies $a_k(x,0) = 0$ for $k \geq 1$, and the differential equation with $t = 0$ implies that $\partial_t a_k(x,0) = g_k(x)$ for $1 \leq k \leq m$. Collecting the terms of the first order in μ we get

$$\partial_t a_1(x,t) = g_1(x).$$

Consequently $a_1(x,t) = g_1(x)t$. Using induction in k and collecting Taylor coefficients of order k in μ, it is not too difficult to prove that $a_k(x,t)$ are defined uniquely and are polynomial in t of order k with coefficients depending on x.

Moreover for $k \leq m$ we get $a_k(x,t) = tg_k(x) + t^2 b_k(x,t)$, where b_k depends on $g_1, \dots, g_{k-1}$ only. Therefore for $k \leq m$ we can set

$$g_k(x) = f_k(x) - b_k(x,1)$$

and get $a_k(x,1) = f_k(x)$.

It is easy to check that if we repeat the procedure with m replaced by $m+1$ the values of a_k with $k \leq m$ are not affected. Consequently, the series g_k are defined uniquely for all k.

The smooth dependence of a flow on its vector field implies that, for any $x \in U$,

$$\Phi^t_{G_m}(x) = \sum_{j=0}^{m} \mu^j a_j(x,t) + \mu^{m+1} r_m(x,t,\mu)$$

where r_m is a bounded function in $\mathcal{V} = B_r(x) \times [0,m] \times \{|\mu| \leq \mu_m(x)\}$ and $r > 0$ depends on x and m. Consequently, for all $0 \leq k \leq m$, $\Phi^k_{G_m}$ and f^k_μ have a common m-jet in μ at the point x and

$$x_k = f^k_\mu(x) = \Phi^k_{G_m}(x_0) + \mu^{m+1} q_{m,k}(x_0,\mu)$$

where $q_{m,k}$ is a bounded function on $\mathcal{V}$. Combining these two bounds we obtain

$$x_k = \sum_{j=0}^{m} \mu^j a_j(x_0,k) + \mu^{m+1} \left(q_{m,k}(x_0,\mu) + r_m(x_0,k,\mu) \right).$$

The interpolation by a polynomial of degree m is exact on polynomials of degree m. Consequently,

$$P_m(t) = \sum_{j=0}^{m} \mu^j a_j(x_0,t) + \mu^{m+1} R_m(x_0,t)$$

where $R_m(x_0,t)$ is the polynomial of degree m in t which interpolates the points $q_{m,k}(x_0,\mu)+r_m(x_0,k,\mu)$ with the node $t=k$ and $0 \leq k \leq m$. Taking the derivative at $t=0$ we get

$$X_m(x) = \sum_{j=0}^{m} \mu^j \dot a_j(x,0) + \mu^{m+1} \dot R_m(x_0,0) = G_m(x) + \mu^{m+1} \dot R_m(x_0,0),$$

where, by (14), we get

$$\dot R_m(x,0) = \sum_{k=0}^{m} p_{mk} \left(q_{m,k}(x,\mu) + r_m(x,k,\mu) \right).$$

Hence $\dot R_m(x,0)$ is bounded on $\mathcal{V}$ and $X_m(x) = G_m(x) + O(\mu^{m+1})$. □

5 Embedding a symplectic near-the-identity map into an autonomous Hamiltonian flow

Suppose that a symplectic map f is ϵ-close to the identity on a complex δ-neighbourhood of $D_0 \subset \mathbb{C}^{2d}$. The following theorem shows that if the ratio δ/ϵ is sufficiently large then an interpolating vector field of optimal order $m \sim \delta/\epsilon$ provides exponentially accurate approximation for the map f. In contrast to the classical result of Neishtadt [20] our theorem provides explicit expressions for all constants. Therefore our theorem can be applied not only to members of a near-the-identity family (where we are able to decrease ϵ when necessary) but to an individual map as well. This subtle difference will play the key role in our proof of the Nekhoroshev theorem.

Let $D_0 \subset \mathbb{C}^{2d}$ and D be a δ-neighbourhood of D_0. Suppose that a symplectic map $f : (p, q) \mapsto (P, Q)$ admits a generating function of the form

$$G(P, q) = Pq + S(P, q), \tag{15}$$

i.e., the map is defined implicitly by the equations

$$p = P + \frac{\partial S}{\partial q}(P, q), \quad Q = q + \frac{\partial S}{\partial P}(P, q).$$

We assume that S has an analytic continuation onto D and we use

$$\epsilon = \|\nabla S\|_D = \sup_{(P,q)\in D} \max\{\, |P - p|, |Q - q| \,\}$$

to characterise the closeness of f to the identity. Our definition is slightly different from the traditional one where the supremum is taken over the domain of the map while we use the domain of its generation function. Our choice slightly simplifies analysis of transitions between a symplectic map and its generating function.

In the following theorem we use the infinity norm for vectors and supremum norms for functions. We use $\lfloor\cdot\rfloor$ to denote the integer part of a number.

Theorem 5.1. *If* $m = \left\lfloor \dfrac{\delta}{6\mathrm{e}\,\epsilon} - d \right\rfloor \geq 1$ *and* X_m *is the interpolating vector field* (13) *of order* m*, then* $\|X_m\|_{D_1} \leq 2\epsilon$ *and*

$$\|\Phi_{X_\mathrm{m}} - f\|_{D_0} \leq 3\, \mathrm{e}^{d+1} \epsilon \exp\left(-\delta/(6\mathrm{e}\,\epsilon)\right),$$

where D_1 *is the* $\frac{\delta}{2}$*-neighbourhood of* D_0*. Moreover there is a Hamiltonian vector field* $\hat{X}_m$ *such that*

$$\|\hat{X}_m - X_m\|_{D_1} \leq 4\, \mathrm{e}^{d+1} \epsilon \exp\left(-\delta/(6\mathrm{e}\epsilon)\right),$$

and

$$\|\Phi_{\hat{X}_\mathrm{m}} - f\|_{D_0} \leq 5\, \mathrm{e}^{d+1} \epsilon \exp\left(-\delta/(6\mathrm{e}\,\epsilon)\right).$$

Moreover, $\|\hat{X}_m\|_{D_1} \le 4\epsilon$ *and*

$$\|\hat{X}_m - \mathrm{J}\nabla S\|_{D_1} \le \frac{c\epsilon^2}{\delta} \tag{16}$$

where $c = 17(d+3)^2$ *and* J *is the standard symplectic matrix.* [1]

Remark 5.2. *We also prove the following statements about interpolating vector fields. Under the assumption of the theorem for every* m *such that*

$$1 \le m < \frac{\delta}{6\epsilon} - d$$

the following inequalities hold: $\|X_m\|_{D_1} \le 2\epsilon$, $\|\hat{X}_m\|_{D_1} \le 4\epsilon$,

$$\|\Phi_{X_m} - f\|_{D_0} \le 3C_m^m \epsilon^{m+1},$$
$$\|\Phi_{\hat{X}_m} - f\|_{D_0} \le 5C_m^m \epsilon^{m+1},$$
$$\|\hat{X}_m - X_m\|_{D_1} \le 4C_m^m \epsilon^{m+1},$$

where $C_m = \frac{6(m+d)}{\delta}$. *These bounds show that* X_m *provides an embedding of the map into a flow with* $O(\epsilon^{m+1})$ *error. The exponential bound is obtained by choosing* m *to minimize the error bound. The best approximation of the map by an interpolating flow is achieved when* $m \approx \delta/6\mathrm{e}\epsilon$. *This step is possible due to the explicit control of the constants in the error bounds.*

Proof. We consider the map f as a member of a family of symplectic maps f_μ defined implicitly by the generating function

$$G_\mu(P,q) = Pq + \mu S(P,q)$$

where μ is a complex parameter. When $\mu = 1$ the map f_μ coincides with f. When $\mu = 0$ the map f_μ is the identity. Therefore this family interpolates between f and the identity map $\xi : (p,q) \mapsto (p,q)$. Obviously the function G_μ is analytic in the same domain D as the function S.

First we are going to prove that if $x_0 \in D_1$, $k \in \mathbb{N}$ and

$$|\mu| \le \mu_k = \frac{\delta}{2\epsilon(k+d)},$$

then $x_k = f_\mu^k(x_0) \in D$ and $\left|x_k - x_{k-1}\right| \le |\mu|\epsilon$. Indeed, let $x_k = (p_k, q_k)$, the trajectory is defined by the system

$$\begin{cases} p_{k-1} = p_k + \mu \dfrac{\partial S}{\partial q}(p_k, q_{k-1}), \\ \quad q_k = q_{k-1} + \mu \dfrac{\partial S}{\partial P}(p_k, q_{k-1}). \end{cases} \tag{17}$$

[1] The constant in the estimate (16) is not optimal. It is obtained using the interpolation of the first order and can be improved using a more accurate approximation for H_m.

In order to find p_k we need to solve the first equation. Then we substitute the solution into the second one. The Implicit Function Theorem A.1 implies that the system with $k=1$ has a solution $(p_1,q_1)\in D$. We continue with the help of finite induction in k. Suppose that $k\in\mathbb{N}$ and the first $k-1$ iterates of x_0 belong to D provided $|\mu|\le\mu_{k-1}$. Then the system implies that

$$|x_{k-1}-x_0|\le\sum_{j=1}^{k-1}|x_j-x_{j-1}|\le(k-1)|\mu|\epsilon.$$

Let $r_{k-1}=\frac{1}{2}\delta-(k-1)|\mu|\epsilon$. Since $x_0\in D_1$, the set D contains the $\frac{\delta}{2}$-neighbourhood of x_0 and, consequently, the ball $B_{r_{k-1}}(x_{k-1})\subset D$. Taking into account the definition of μ_k we get that for $|\mu|\le\mu_k$

$$r_{k-1}=\tfrac{1}{2}\delta-(k-1)|\mu|\epsilon=(k+d)\mu_k\epsilon-(k-1)|\mu|\epsilon\ge(d+1)|\mu|\epsilon.$$

We see that the assumptions of the Implicit Function Theorem A.1 are satisfied by the first line of the system and consequently it defines p_k as a function of (p_{k-1},q_{k-1}). It is not too difficult to check that $x_k=(p_k,q_k)\in D$.

Now we can study interpolating vector fields for the map f_μ. First we are to find upper bounds for the finite differences. We introduce the following notation: for a function g let $\mathrm{T}_f(g)=g\circ f$ and $\mathrm{I}(g)=g$. We note that

$$(\mathrm{I}-\mathrm{T}_{f_\mu})^k\xi=(\mathrm{I}-\mathrm{T}_{f_\mu})^{k-1}(\xi-f_\mu)=\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j(f_\mu^j-f_\mu^{j+1})$$

(recall that ξ stands for the identity map) and consequently

$$\left\|(\mathrm{I}-\mathrm{T}_{f_\mu})^k\xi\right\|_{D_1}\le\sum_{j=0}^{k-1}\binom{k-1}{j}\left\|f_\mu^j-f_\mu^{j+1}\right\|_{D_1}\le 2^{k-1}|\mu|\epsilon. \tag{18}$$

Next we recall that the operator $\mathrm{T}_{f_\mu}-\mathrm{I}$ increases valuation in μ, consequently $\mathrm{val}_\mu((\mathrm{I}-\mathrm{T}_{f_\mu})^k\xi)\ge k$. Applying the MMP[2] for each $x\in D_1$ fixed, we obtain

$$\left\|(\mathrm{I}-\mathrm{T}_{f_\mu})^k\xi\right\|_{D_1}\le\frac{|\mu|^k}{\mu_k^k}\sup_{|\mu|=\mu_k}\left\|(\mathrm{I}-\mathrm{T}_{f_\mu})^k\xi\right\|_{D_1}\le\frac{|\mu|^k2^{k-1}\epsilon}{\mu_k^{k-1}}.$$

Now we let $m\ge1$ and consider the interpolating vector field (13) written in the form

$$X_{m,\mu}=-\sum_{k=1}^{m}\frac{1}{k}(\mathrm{I}-\mathrm{T}_{f_\mu})^k\xi.$$

[2]The degree of the first non-zero monomial of the Taylor expansion in a variable ζ defines a valuation, that will be denoted by val_ζ, of the ring of formal series $\mathbb{C}[[\zeta]]$. We will use the following simple statement of complex analysis: if g is analytic in $\{\zeta\in\mathbb{C},|\zeta|\le\zeta_0\}$, $\zeta_0>0$, and $\mathrm{val}_\zeta(g)\ge k$ then it follows from the maximum modulus principle (MMP) that $|g(\zeta)|\le(|\zeta|/\zeta_0)^k\max_{|\zeta|=\zeta_0}|g(\zeta)|$.

If $|\mu| \le \mu_m$, it is analytic in D_1 and admits the following upper bound

$$\|X_{m,\mu}\|_{D_1} \le \sum_{k=1}^{m} \frac{1}{k} \left\| (\mathrm{I} - \mathrm{T}_{f_\mu})^k \xi \right\|_{D_1} \le \epsilon |\mu| \sum_{k=1}^{m} \frac{1}{k} \left(\frac{2|\mu|}{\mu_k} \right)^{k-1} .$$

Using that $\mu_k \ge \mu_m$ for $k \le m$, we get that

$$\|X_{m,\mu}\|_{D_1} \le \epsilon |\mu| \sum_{k=1}^{m} \frac{1}{k} \left(\frac{2}{3} \right)^{k-1} < \epsilon |\mu| \, \frac{3}{2} \log 3 < 2\epsilon |\mu| \tag{19}$$

for $|\mu| \le \mu_m/3$. If $\mu_m \ge 3$, the domain of validity of the upper bound includes $\mu = 1$. Since $X_m = X_{m,1}$ we conclude that

$$\|X_m\|_{D_1} < 2\epsilon.$$

Now we consider $\Phi_{X_{m,\mu}}$, the time-one map of the vector field $X_{m,\mu}$. Equation (19) implies that

$$\|X_{m,\mu}\|_{D_1} < \frac{2\epsilon \mu_m}{3} = \frac{\delta}{3(m+d)} \le \frac{\delta}{6}.$$

Then the orbit of every point in D_0 remains in D_1 during one unit of time and

$$\|\Phi_{X_{m,\mu}} - \xi\|_{D_0} \le \|X_{m,\mu}\|_{D_1}.$$

Then

$$\|\Phi_{X_{m,\mu}} - f_\mu\|_{D_0} \le \|\Phi_{X_{m,\mu}} - \xi\|_{D_0} + \|\xi - f_\mu\|_{D_0} \le \|X_{m,\mu}\|_{D_1} + \epsilon|\mu| = 3\epsilon|\mu|.$$

Lemma 4.2 states that the Taylor expansion in μ of $\Phi_{X_{m,\mu}}$ matches the Taylor expansion of f_μ up to the order m. Then MMP implies that

$$\|\Phi_{X_{m,\mu}} - f_\mu\|_{D_0} \le \epsilon \mu_m \, \frac{3^{m+1} |\mu|^{m+1}}{\mu_m^{m+1}}. \tag{20}$$

Substituting $\mu = 1$ we obtain

$$\|\Phi_{X_m} - f\|_{D_0} \le 3\epsilon \, \frac{3^m}{\mu_m^m} = 3\epsilon \left(\frac{6(m+d)\epsilon}{\delta} \right)^m . \tag{21}$$

The right hand side depends on m and takes the smallest values somewhere near $m = \lfloor M_\epsilon \rfloor$ where

$$M_\epsilon = \frac{\delta}{6\mathrm{e}\epsilon} - d. \tag{22}$$

We note that for $m \le M_\epsilon$ we get that

$$\mu_m = \frac{\delta}{2\epsilon(m+d)} \ge \frac{\delta}{2\epsilon(M_\epsilon + d)} = 3\mathrm{e} > 6$$

and consequently the inequality (21) holds for all these values of m. We notice that for $m = \lfloor M_\epsilon \rfloor$ we have

$$\frac{6(m+d)\epsilon}{\delta} \le \mathrm{e}^{-1}$$

and consequently

$$\|\Phi_{X_m} - f\|_{D_0} \le 3\epsilon\,\mathrm{e}^{-m} \le 3\epsilon\,\mathrm{e}^{1+d}\exp\left(-\frac{\delta}{6\mathrm{e}\epsilon}\right).$$

The interpolating vector fields $X_{m,\mu}$ are rarely Hamiltonian. On the other hand, the formal interpolating vector field is Hamiltonian. Although this property is known, we give a simple proof in Appendix B. Then Lemma 4.2 implies that $\hat{X}_{\mu,m}$, the Taylor polynomial in μ of degree m for $X_{m,\mu}$, is Hamiltonian. For example, the interpolating vector field of the first order is given by

$$X_{1,\mu}(p,q) = f_\mu(p,q) - (p,q).$$

In general, there is no reason for this vector field to be Hamiltonian. On the other hand, its Taylor polynomial of degree one,

$$\hat{X}_{1,\mu} = \left.\frac{\partial X_{1,\mu}}{\partial \mu}\right|_{\mu=0} \mu = \left.\frac{\partial f_\mu}{\partial \mu}\right|_{\mu=0} \mu,$$

is Hamiltonian. In order to find the corresponding Hamiltonian function we recall that the map $f_\mu : (p,q) \mapsto (p_1, q_1)$ is defined implicitly by the system (17) with $k = 1$ (we assume $(p_0, q_0) = (p,q)$ are independent of μ). Differentiating the system with respect to μ at $\mu = 0$ and using that $p_1 = q$ and $q_1 = q$ for $\mu = 0$, we get

$$\begin{cases} 0 = \dfrac{\partial p_1}{\partial \mu} + \dfrac{\partial S}{\partial q}(p,q), \\ \dfrac{\partial q_1}{\partial \mu} = 0 + \dfrac{\partial S}{\partial p}(p,q). \end{cases}$$

We conclude that

$$\hat{X}_{1,\mu}(p,q) = \mu\left(-\frac{\partial S}{\partial q}(p,q), \frac{\partial S}{\partial p}(p,q)\right). \tag{23}$$

We see that the vector field $\hat{X}_{1,\mu}$ is Hamiltonian with the Hamiltonian function $H_{1,\mu} = \mu S$.

In order to estimate $\hat{X}_{m,\mu}$ for $m \le M_\epsilon$, we notice that equation (19) implies that $\|X_{m,\mu}\|_{D_1} \le 2\epsilon\mu_m/3$ for $|\mu| \le \mu_m/3$. Then for $k \le m$

$$\frac{1}{k!}\left\|\partial_\mu^k X_m\big|_{\mu=0}\right\|_{D_1} \le 2\epsilon\left(\frac{3}{\mu_m}\right)^{k-1}$$

and

$$\|\hat{X}_{m,\mu}\|_{D_1} \le \sum_{k=1}^{m} \frac{1}{k!} \left\| \partial_\mu^k X_{m,\mu}\big|_{\mu=0} \right\|_{D_1} |\mu|^k \le 2\epsilon|\mu| \sum_{k=1}^{m} \left(\frac{3|\mu|}{\mu_m}\right)^{k-1}.$$

For $|\mu| \le \mu_m/6$ we get

$$\|\hat{X}_{m,\mu}\|_{D_1} \le 2\epsilon|\mu| \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} \le 4\epsilon|\mu|. \tag{24}$$

Since $\mu_m > 6$ we can substitute $\mu = 1$ to obtain

$$\|\hat{X}_m\|_{D_1} \le 4\epsilon.$$

Then we repeat the previous arguments using $\hat{X}_{m,\mu}$ instead of $X_{m,\mu}$ and the upper bound (24) instead of (19). The equation (24) implies that $\|\hat{X}_{m,\mu}\|_{D_1} \le 2\epsilon\mu_m/3 = \frac{\delta}{3(m+d)} < \frac{\delta}{6}$ for all $|\mu| \le \mu_m/6$. Then repeating the previous arguments we get $\|\Phi_{\hat{X}_{m,\mu}} - f_\mu\|_{D_0} \le 5\epsilon|\mu|$. Then using the MMP we get

$$\|\Phi_{\hat{X}_{m,\mu}} - f_\mu\|_{D_0} \le \frac{5\epsilon\mu_m}{6} \frac{6^{m+1}|\mu|^{m+1}}{\mu_m^{m+1}}.$$

Substituting $\mu = 1$ we get

$$\|\Phi_{\hat{X}_m} - f\|_{D_0} \le 5\epsilon \left(\frac{6(m+d)\epsilon}{\delta}\right)^m.$$

In particular, for $m = \lfloor M_\epsilon \rfloor$

$$\|\Phi_{\hat{X}_\mathrm{m}} - f\|_{D_0} \le 5\,\mathrm{e}^{d+1}\epsilon \exp\left(-\delta/(6\mathrm{e}\,\epsilon)\right).$$

Since $\hat{X}_{m,\mu}$ is the Taylor polynomial of $X_{m,\mu}$ of order m we can use the standard bound for the remainder (the radius of convergence is $\mu_m/3$, the bound is for $\mu = 1$, $\mu_m > 6$):

$$\left\|\hat{X}_m - X_m\right\|_{D_1} \le \frac{(2\epsilon\mu_m/3)(3/\mu_m)^{m+1}}{1 - 3/\mu_m} \le 4\epsilon(3/\mu_m)^m = 4\epsilon\left(\frac{6\epsilon(m+d)}{\delta}\right)^m.$$

Substituting $m = \lfloor M_\epsilon \rfloor$ we get

$$\|\hat{X}_m - X_m\|_{D_1} \le 4\epsilon\,\mathrm{e}^{1+d} \exp\left(-\frac{\delta}{6\mathrm{e}\epsilon}\right). \tag{25}$$

We see that the interpolating vector field X_m is exponentially close to a Hamiltonian one.

In order to complete the proof we have to show that the vector fields $\hat{X}_m = \hat{X}_{m,1}$ are close to S for all $m \le M_\epsilon$. Using the upper bounds for the derivatives we get

$$\|\hat{X}_m - \hat{X}_1\|_{D_1} \le \sum_{k=2}^{m} \frac{1}{k!} \left\| \partial_\mu^k X_m\big|_{\mu=0} \right\|_{D_1} \le 2\epsilon \sum_{k=2}^{m} \left(\frac{3}{\mu_k}\right)^{k-1}.$$

It is not too difficult to check that the sequence $\left(\frac{3}{\mu_k}\right)^{k-1}$ is monotone decreasing for $k \le m \le M_\epsilon$. Therefore

$$\|\hat{X}_m - \hat{X}_1\|_{D_1} \le \frac{6\epsilon}{\mu_2} + 2\epsilon \sum_{k=3}^{m} \left(\frac{3}{\mu_k}\right)^{k-1} \le \frac{6\epsilon}{\mu_2} + \frac{18\epsilon\, M_\epsilon}{\mu_3^2}.$$

Recalling the definitions of μ_k we get

$$\|\hat{X}_m - \hat{X}_1\|_{D_1} \le \frac{12\epsilon^2}{\delta}(d+2) + \frac{12\epsilon^2}{\mathrm{e}\delta}(d+3)^2 \le \frac{17\epsilon^2}{\delta}(d+3)^2.$$

We get the desired estimate as $\hat{X}_1 = \mathrm{J}\nabla S$. □

6 Interpolating flow near a fully resonant torus

This section contains the proof of Theorem 2.2. We recall that $I_* \in B_R$ corresponds to a fully resonant torus, i.e., $n\omega(I_*) \in \mathbb{Z}^d$ for some natural $n < N_\varepsilon$. For the rest of this section we assume that

$$\gamma_0^2 = \frac{18d\|a\|}{\nu} \qquad \text{and} \qquad r_0^2 = \frac{\nu}{6d\|h_0''\|}\,.$$

We also fix $\gamma \ge \gamma_0$. We will reduce ε_0 when necessary.

6.1 Exponentially accurate interpolation

Theorem 2.2 establishes estimates for stability times for real initial conditions. On the other hand, in order to use Theorem 5.1 we need a bound of the map f_ε^n in a complex neighbourhood of its real domain. As a first step we check that the map satisfies the assumptions of Theorem 5.1 in

$$\mathcal{D}(I_*) = \Big\{ (I,\varphi) \in \mathbb{C}^{2d} : |I - I_*| < 2\rho_n, |\operatorname{Im}(\varphi)| < r/2 \Big\}, \tag{26}$$

a complex neighbourhood of the domain $\mathcal{D}_0(I_*)$ defined in (10). Let $(I_0,\varphi_0) \in \mathcal{D}(I_*)$. Applying Lemma 3.1 recursively to check that the previous iterates do not leave the domain $\mathcal{D}_F$, we conclude that

$$\begin{aligned} |I_k - I_*| &\le |I_0 - I_*| + C_1 k\varepsilon \le 2\rho_n + C_1 k\varepsilon < \sigma, \\ |\operatorname{Im}(\varphi_k - \varphi_0 - k\omega(I_0))| &\le C_2 k^2 \varepsilon < \frac{r}{4}, \end{aligned}$$

while k is not too large. Then

$$|\operatorname{Im}(\varphi_k - \varphi_0)| \le \frac{r}{4} + k|\operatorname{Im}(\omega(I_0))| \le \frac{r}{4} + k\|\omega'\|\,|\operatorname{Im}(I_0)| \le \frac{r}{4} + C_3 k\rho_n < \frac{r}{2}.$$

Here we use the constant $C_3 = 2\|\omega'\|$ and assume that

$$C_1 k\varepsilon < \frac{\sigma}{2}, \qquad C_2 k^2\varepsilon < \frac{r}{4}, \qquad C_3 k\rho_n < \frac{r}{4}, \qquad 4\rho_n < \sigma. \tag{27}$$

Recalling our choice of $\rho_n = \rho_\varepsilon/n$, $\rho_\varepsilon = \gamma\varepsilon^{1/2(d+1)}$ and $n < N_\varepsilon = \varepsilon^{-d/2(d+1)}$, see (11) and (9), it is easy to check that the inequalities (27) are satisfied for $k \le n$ provided $\varepsilon < \varepsilon_0$ with a sufficiently small ε_0 (independent of n). Then the first n iterates $(I_n, \varphi_n) = F_\varepsilon^n(I_0, \varphi_0)$ are well defined for initial conditions in $\mathcal{D}(I_*)$. Since $|\omega(I_0) - \omega(I_*)| \le \|\omega'\| 2\rho_n$ we get from Lemma 3.1 that

$$|I_n - I_0| \le C_1 n\varepsilon, \qquad |\varphi_n - \varphi_0 - n\omega(I_*)| \le C_2 n^2\varepsilon + C_3 n\rho_n. \tag{28}$$

To study the dynamics in $\mathcal{D}(I_*)$ we introduce translated and scaled actions J with the help of the equality

$$I = I_* + \rho_n J. \tag{29}$$

Let $\hat f_\varepsilon^n$ denote the map (8) expressed in the new coordinates. It can be written in the form $\hat f_\epsilon^n : (J, \varphi) \mapsto (\bar J, \bar\varphi)$,

$$\begin{cases} \bar J = J + \rho_n^{-1}\varepsilon \displaystyle\sum_{k=0}^{n-1} a(I_k, \varphi_k), \\ \bar\varphi = \varphi + \displaystyle\sum_{k=0}^{n-1} (\omega(I_k) - \omega(I_*)) + \varepsilon \sum_{k=0}^{n-1} b(I_k, \varphi_k), \end{cases} \tag{30}$$

and $(I_k, \varphi_k) = F_\varepsilon^k(I_0, \varphi_0)$ denote iterates of $(I_0, \varphi_0) = (I_* + \rho_n J, \varphi)$ under the original map (1). The map $\hat f_\varepsilon^n$ is ϵ_n-close to the identity on $\mathcal{D}(I_*)$ where

$$\begin{aligned} \epsilon_n &= \sup_{|J|<2, |\operatorname{Im}(\varphi)|<r/2} \max\{\, |\bar J - J|, |\bar\varphi - \varphi| \,\} \\ &= \sup_{(I_0,\varphi_0)\in\mathcal{D}(I_*)} \max\{\, \rho_n^{-1}|I_n - I_0|, |\varphi_n - \varphi_0 - n\omega(I_*)| \,\}. \end{aligned}$$

Then, the estimates (28) show that

$$\epsilon_n \le \max\left\{C_1\rho_n^{-1} n\varepsilon,\, C_3 n\rho_n + C_2 n^2\varepsilon\right\} = n\rho_n \max\left\{C_1 \frac{\varepsilon}{\rho_n^2},\, C_3 + C_2\frac{n\varepsilon}{\rho_n}\right\}. \tag{31}$$

Since

$$\frac{\varepsilon}{\rho_n^2} = \frac{n^2\varepsilon}{\rho_\varepsilon^2} = \frac{n^2\varepsilon^{d/(d+1)}}{\gamma^2} < \frac{N_\varepsilon^2\varepsilon^{d/(d+1)}}{\gamma^2} = \frac{1}{\gamma^2}$$

and

$$\frac{n\varepsilon}{\rho_n} = \frac{n^2\varepsilon}{\rho_\epsilon} < \frac{\rho_\varepsilon}{\gamma^2} = \frac{\varepsilon^{1/2(d+1)}}{\gamma},$$

there is a positive constant C_4 such that

$$\epsilon_n \leq C_4 n \rho_n = \gamma C_4 \varepsilon^{1/2(d+1)}. \tag{32}$$

The Implicit Function Theorem A.1 can be applied to the first component of (30) to show that J is an analytic function of $\bar{J}$ and φ on the set

$$\hat{D} = \Big\{ (J, \varphi) \in \mathbb{C}^d : |J| \leq 2 - (d+1)\epsilon_n, \ |\operatorname{Im}(\varphi)| < r/2 \Big\}.$$

Then the expression for J can be substituted into the second component of (30) to express $(J, \bar{\varphi})$ as a function of $(\bar{J}, \varphi)$. Since the map is symplectic, then according to Appendix C there is a function S_n such that

$$\bar{J} - J = -\frac{\partial S_n}{\partial \varphi}(\bar{J}, \varphi), \qquad \bar{\varphi} - \varphi = \frac{\partial S_n}{\partial \bar{J}}(\bar{J}, \varphi). \tag{33}$$

The definition of ϵ_n implies

$$\|\nabla S_n\|_{\hat{D}} \leq \epsilon_n.$$

Let

$$\delta = \tfrac{1}{2} \min\{1, r\} \qquad \text{and} \qquad c_3 = \frac{\delta}{C_4 \gamma 6\mathrm{e}}. \tag{34}$$

If $\epsilon_n \leq 1/2(d+1)$, then $\hat{D}$ contains a complex δ-neighbourhood of the real set

$$\hat{D}_0 = B(0,1) \times \mathbb{R}^d.$$

If $\epsilon_n \leq \delta/6\mathrm{e}(d+1)$, then the map $\hat{f}_\varepsilon^n$ satisfies the assumptions of Theorem 5.1 which states that there is $m = m(\epsilon_n) \sim \epsilon_n^{-1}$ such that the time-one map of the Hamiltonian vector field $\hat{X}_m$ approximates $\hat{f}_\varepsilon^n$ with exponential accuracy:

$$\left\| \hat{f}_\varepsilon^n - \Phi_{\hat{X}_m} \right\|_{B(0,1)\times\mathbb{R}^d} \leq 5\mathrm{e}^{d+1} \epsilon_n \exp\left(-c_3 \, \varepsilon^{-1/2(d+1)} \right). \tag{35}$$

According to the theorem $\hat{X}_m$ is close to the vector field with the Hamiltonian function S_n. We will analyse the Hamiltonian function of $\hat{X}_m$ in the next subsection.

Remark 6.1. *In this section we use the linear scaling* (29) *of the original action coordinate I by the factor ρ_n. This scaling is useful to enable a direct application of Theorem 5.1. However, we can compute X_m using the iterates of f_ε^n in the original coordinates (I, φ) as the interpolation procedure commutes with linear changes of variables.*

6.2 Long term stability of actions

In the previous section we have established that f_ε^n, the lift of the map F_ε^n, is exponentially close to the time-one map of the autonomous Hamiltonian vector field $\hat{X}_m$ in a small neighbourhood of the fully resonant torus. In this section we will derive an approximation for the corresponding Hamiltonian function H_m and use its properties to establish which trajectories of the map are trapped inside this neighbourhood for exponentially long times.

According to (16), the interpolating Hamiltonian H_m is close to S_n, the generating function of the map $\hat{f}_\varepsilon^n : (J,\varphi) \mapsto (\bar{J},\bar{\varphi})$. As a first step we derive the leading order approximation for S_n. Comparing (30) with (33) we conclude that S_n is a solution of the system of equations

$$\frac{\partial S_n}{\partial \varphi}(\bar{J},\varphi) = -\rho_n^{-1}\varepsilon \sum_{k=0}^{n-1} a(I_k,\varphi_k),$$

$$\frac{\partial S_n}{\partial \bar{J}}(\bar{J},\varphi) = n(\omega(I_n)-\omega(I_*)) + \sum_{k=0}^{n-1}(\omega(I_k)-\omega(I_n)) + \varepsilon \sum_{k=0}^{n-1} b(I_k,\varphi_k).$$

Note that the right hand side is expressed in terms of $(I_k,\varphi_k) = F_\varepsilon^k(I_*+\rho_n J,\varphi)$. Since the value of J can be expressed in terms of $(\bar{J},\varphi)$, (I_k,φ_k) can also be expressed in terms of $(\bar{J},\varphi)$. In particular $I_n = I_* + \rho_n \bar{J}$. The symplecticity of $\hat{f}_\varepsilon^n$ implies existence of a solution. We write it in the form

$$S_n(\bar{J},\varphi) = h_n(\bar{J}) + w_n(\bar{J},\varphi). \tag{36}$$

The first term

$$h_n(\bar{J}) = n\rho_n^{-1}\left(h_0(I_*+\rho_n\bar{J}) - h_0(I_*) - \rho_n\,\omega(I_*)\cdot\bar{J}\right)$$

represents the part independent of φ and comes from the explicit calculation performed with the help of the equality $\omega(I) = h_0'(I)$. The second term is expressed as an integral

$$w_n(\bar{J},\varphi) = \int_{(0,0)}^{(\bar{J},\varphi)} \sum_{l=1}^{d}(v_l\, d\bar{J}_l - u_l\, d\varphi_l)$$

where the index l refers to components of the vectors, the integral does not depend on the path connecting the end points and

$$\begin{aligned} u(\bar{J},\varphi) &= \rho_n^{-1}\varepsilon\sum_{k=0}^{n-1} a(I_k,\varphi_k),\\ v(\bar{J},\varphi) &= \sum_{k=0}^{n-1}(\omega(I_k)-\omega(I_n)) + \varepsilon\sum_{k=0}^{n-1} b(I_k,\varphi_k). \end{aligned} \tag{37}$$

The equation (36) suggests that S_n has the form of an integrable part plus a perturbative term, denoted by h_n and w_n respectively. The integrable part is approximately quadratic. Indeed, $h_n'(0) = 0$ and the strong convexity and smoothness of h_0 imply that

$$\frac{\nu\, n\rho_n}{2} |\bar J|^2 \le h_n(\bar J) \le \frac{\nu_2 n\rho_n}{2} |\bar J|^2. \tag{38}$$

where $\nu_2 = d\|h_0''\|$. In order to see that h_n dominates w_n outside a small neighbourhood of the origin, we look for an explicit bound for w_n paying attention to the uniformity for all resonances. The triangle inequality and (37) imply that

$$|u| \le C_1 n\varepsilon \rho_n^{-1}.$$

Then using arguments of Lemma 3.1 we get

$$|v| \le \sum_{k=0}^{n-1} \|\omega'\|\, |I_n - I_k| + \varepsilon n \|b\| \le \frac{\varepsilon n^2}{2} \|\omega'\|\, \|a\| + \varepsilon n\|b\| \le C_2 n^2 \varepsilon.$$

Taking into account the periodicity arguments it is sufficient to consider w_n on the set $\hat D_0 = B(0,1) \times [-1,1]^d$. Then

$$\|w_n\|_{\hat D_0} \le dC_2 n^2 \varepsilon + dC_1 n\varepsilon \rho_n^{-1}. \tag{39}$$

With our choice of $\rho_n = \rho_\varepsilon n^{-1}$, $\rho_\varepsilon = \gamma \varepsilon^{1/2(d+1)}$ and $n < N_\varepsilon = \varepsilon^{-d/2(d+1)}$, see (11) and (9), we get

$$\frac{\|w_n\|_{\hat D_0}}{n\rho_n} \le dC_2 n\varepsilon \rho_n^{-1} + dC_1 \varepsilon \rho_n^{-2} \le dC_2 \varepsilon^{1/2(d+1)} \gamma^{-1} + dC_1 \gamma^{-2},$$

where we used the bounds presented after equation (31). With our choice of $\gamma \ge \gamma_0$, the second term in the sum does not exceed $\nu/18$. Decreasing ε_0 (if necessary) we get

$$\frac{\|w_n\|_{\hat D_0}}{n\rho_n} \le \frac{\nu}{9}. \tag{40}$$

In this way we have got upper bounds for both terms in (36).

According to Theorem 5.1 the interpolating vector field is Hamiltonian, $\hat X_m = \mathrm{J}\nabla H_m$, and the Hamiltonian H_m is close to the generating function S_n due to the bound (16). The Hamiltonian H_m can be obtained by integrating the vector field $\hat X_m = \mathrm{J}\nabla H_m$. Since the map $\hat f_\varepsilon^n$ is periodic in angles the vector field is also periodic. Moreover, since the map is exact symplectic and in Theorem 5.1 $\hat X_m$ is obtained from a truncated expansion, Appendix C shows that the Hamiltonian is periodic. Then we can restrict the integration to a fundamental domain in the angle variables to get from (16) the inequality

$$\|H_m - S_n\|_{\hat D_0} \le \frac{2cd\epsilon_n^2}{\delta} \le \frac{2cdC_4^2 n^2 \rho_n^2}{\delta} = C_5 n^2 \rho_n^2\ .$$

Decreasing ε_0 (if necessary) we get $C_5 n\rho_n = C_5\rho_\varepsilon < \frac{\nu}{18}$. Then the equation (36) implies that

$$\|H_m - h_n\|_{\hat{D}_0} \le C_5 n^2\rho_n^2 + \|w_n\|_{\hat{D}_0} < \frac{\nu}{6} n\rho_n\,.$$

With the help of the bounds (38) we get that

$$\tfrac{1}{2}\nu|J|^2 - \tfrac{1}{6}\nu \le \frac{H_m(J,\varphi)}{n\rho_n} \le \tfrac{1}{2}\nu_2|J|^2 + \tfrac{1}{6}\nu\,. \tag{41}$$

Suppose that at some point the Hamiltonian $H_m(J,\varphi) < n\rho_n E_{\max}$ where $E_{\max} = \frac{1}{3}\nu$. The first inequality of (41) implies that

$$\frac{\nu}{2}|J|^2 < \frac{\nu}{3} + \frac{\nu}{6} = \frac{\nu}{2}\,.$$

Since the Hamiltonian flow preserves H_m, the whole trajectory of (J,φ) is inside the domain $|J| < 1$.

Now let $E_0 = \frac{1}{4}\nu$. Any point with $|J| < r_0$ belongs to the set of initial conditions which satisfy the inequality $H_m(J,\varphi) < n\rho_n E_0$. Indeed, our choice of r_0 implies that

$$\frac{H_m(J,\varphi)}{n\rho_n} < \frac{\nu_2}{2} r_0^2 + \frac{\nu}{6} = \frac{\nu}{4}.$$

Unlike the Hamiltonian flow, the map does not preserve the energy. Fortunately the change in the energy after a single iterate is exponentially small:

$$\begin{aligned} M_\varepsilon &= \left\| H_m \circ \hat{f}_\varepsilon^n - H_m \right\|_{\hat{D}_0} = \left\| H_m \circ \hat{f}_\varepsilon^n - H_m \circ \Phi_{\hat{X}_m} \right\|_{\hat{D}_0} \\ &\le \|H_m'\|_{\hat{D}_1} \left\| \hat{f}_\varepsilon^n - \Phi_{\hat{X}_m} \right\|_{\hat{D}_0} = \|\hat{X}_m\|_{\hat{D}_1} \left\| \hat{f}_\varepsilon^n - \Phi_{\hat{X}_m} \right\|_{\hat{D}_0} \\ &\le 20\epsilon_n^2 \mathrm{e}^{d+1} \exp\left(-c_3\varepsilon^{-1/2(d+1)}\right) \end{aligned}$$

where we use (35) and the bound $\|\hat{X}_m\|_{\hat{D}_1} \le 4\epsilon_n$ of Theorem 5.1. Here $\hat{D}_1$ is the $\frac{\delta}{2}$-neighbourhood of $\hat{D}_0$. If we take an initial condition with $|J| < r_0$ then the initial energy is below $n\rho_n E_0$. We can be sure that the point remains inside the domain $|J| < 1$ while the energy does not exceed $n\rho_n E_{\max}$. In this case we can use the telescopic sum to see that

$$\begin{aligned} H_m \circ \hat{f}_\varepsilon^{kn}(J,\varphi) - H_m(J,\varphi) &= \sum_{j=1}^{k} (H_m \circ \hat{f}_\varepsilon^{jn}(J,\varphi) - H_m \circ \hat{f}_\varepsilon^{(j-1)n}(J,\varphi)) \\ &\le kM_\varepsilon. \end{aligned}$$

Then $H_m \circ \hat{f}_\varepsilon^{kn}(J,\varphi) < n\rho_n E_{\max}$ for all $k \le n\rho_n(E_{\max} - E_0)/M_\varepsilon$. Consequently, the minimal number of iterates of F_ε needed to start with an energy below $n\rho_n E_0$ and finish above $n\rho_n E_{\max}$ is larger than

$$T_\varepsilon = \frac{n^2\rho_n\nu}{12M_\varepsilon} \ge \frac{n^2\rho_n\nu}{240\epsilon_n^2 \mathrm{e}^{d+1}} \exp\left(c_3\,\varepsilon^{-1/2(d+1)}\right) \ge c_2 \exp\left(c_3\,\varepsilon^{-1/2(d+1)}\right).$$

We have proved that a trajectory with an initial condition (I_0, φ_0) such that $|I_0 - I_*| < r_0\rho_n$ has the property $|I_{kn} - I_*| < \rho_n$ for $0 \le kn \le T_\varepsilon$. We complete the proof of Theorem 2.2 by noting that between multiples of n the changes in action variables are controlled by Lemma 3.1 and do not exceed $C_1 n\varepsilon$. Consequently $|I_k - I_*| < \rho_n$ for all $k < T_\varepsilon$. This argument completes the proof of Theorem 2.2.

7 Nucleus of a resonance

Our proof of the exponential estimates for the stability times of the action variables uses a covering of the phase space by ρ_n-neighbourhoods of unperturbed fully resonant tori. Each of these tori is characterised by its frequency ω_* such that $n\omega_* \in \mathbb{Z}^d$ for some $n < N_\varepsilon = \varepsilon^{-d/2(d+1)}$. Therefore a fully resonant torus of period n is included into the analysis when ε becomes smaller than $n^{-2(d+1)/d}$ and eventually every fully resonant torus is used. In this section we show that every fully resonant torus has a small neighbourhood, which we call a *nucleus of the resonance*, where the stability times are much longer than in the Nekhoroshev theorem. Moreover the difference in stability times grows as ε decreases due to the presence of the factor $\varepsilon^{-1/2}$ instead of $\varepsilon^{-1/2(d+1)}$ in the exponent.

For the purpose of this analysis it is convenient to rewrite the map $F_\varepsilon : (I, \varphi) \mapsto (\bar I, \bar\varphi)$ with the help of a generating function

$$S(\bar I, \varphi) = \bar I \cdot \varphi + h_0(\bar I) + \varepsilon s(\bar I, \varphi)$$

where the function s depends periodically on the angles φ. Then the map is defined implicitly by the system

$$\begin{cases} \bar I = I - \varepsilon \dfrac{\partial s}{\partial \varphi}(\bar I, \varphi), \\ \bar\varphi = \varphi + \omega(\bar I) + \varepsilon \dfrac{\partial s}{\partial \bar I}(\bar I, \varphi) \pmod 1, \end{cases} \tag{42}$$

where $\omega(\bar I) = h_0'(\bar I)$. When $\varepsilon = 0$, these equations can be easily solved explicitly. On the other hand, the geometric arguments of Appendix C and Implicit Function Theorem A.1 can be used to show that every quasi-integrable map F_ε can be represented in this form. The n-th iterate of the map takes the form

$$\begin{cases} I_n = I_0 - \varepsilon \displaystyle\sum_{k=0}^{n-1} \partial_2 s(I_{k+1}, \varphi_k), \\ \varphi_n = \varphi_0 + \displaystyle\sum_{k=1}^{n} h_0'(I_k) + \varepsilon \sum_{k=0}^{n-1} \partial_1 s(I_{k+1}, \varphi_k). \end{cases} \tag{43}$$

The subsequent analysis is motivated by the application of the standard scaling near the resonant torus $I = I_*$ with the scaled action J defined by the equation

$I = I_* + \sqrt{\varepsilon} J$. In the scaled variables, the map

$$(\varphi, J) \mapsto (\bar{\varphi}, \bar{J}) = (\varphi_n - n\omega_*, (I_n - I_*)/\sqrt{\varepsilon})$$

takes the form

$$\begin{cases} \bar{J} = J - \varepsilon^{1/2} \sum_{k=0}^{n-1} \partial_2 s(I_{k+1}, \varphi_k), \\ \bar{\varphi} = \varphi + \sum_{k=1}^{n} (h_0'(I_k) - \omega_*) + \varepsilon \sum_{k=0}^{n-1} \partial_1 s(I_{k+1}, \varphi_k), \end{cases} \tag{44}$$

where (I_k, φ_k) denote the trajectory of the point $(\varphi_0, I_0) = (\varphi, I_* + \sqrt{\varepsilon} J)$ under the original map. It is not too difficult to see that on a bounded domain the scaled map is $O(n\sqrt{\varepsilon})$-close to the identity. The interpolating vector field of order one is explicitly represented by the formula above as $X_1 = (\bar{J} - J, \bar{\varphi} - \varphi)$. Expanding X_1 into Taylor series in powers of $\sqrt{\varepsilon}$ we see that the leading term is of the order of $\sqrt{\varepsilon}$ and, in agreement with the general theory of Lemma 4.2, it is Hamiltonian with the Hamiltonian function

$$\hat{H}_1(J, \varphi) = \sqrt{\varepsilon} n \big(K(J) + V_*(\varphi) \big)$$

where

$$K(J) = \frac{1}{2} (h_0''(I_*) J) \cdot J \qquad \text{and} \qquad V_*(\varphi) = \frac{1}{n} \sum_{k=0}^{n-1} s(I_*, \varphi + k\omega_*). \tag{45}$$

The function $K(J)$ comes from the quadratic part of the Taylor expansion of h_0 around $I = I_*$ while the linear part of the expansion vanishes due to the equality $h_0'(I_*) = \omega_*$. We see that the potential V_* coincides with the average of the generating function s over the unperturbed periodic orbit on the resonant torus. The strong convexity of h_0 provides a lower bound for K so we have that

$$\frac{\nu}{2} |J|^2 \le K(J) \le \frac{\nu_2}{2} |J|^2 = \frac{d}{2} \|h_0''\| \, |J|^2.$$

The function $E(J, \varphi) = \frac{1}{n\sqrt{\varepsilon}} \hat{H}_1(J, \varphi)$ defines a slow variable in a neighbourhood of the resonance, it is constant along orbits of the flow of $\hat{H}_1$ and it changes slowly under iterates of the map (44). If $E_0 > \max V_*$, then the set $E \le E_0$ contains a ball $|J| \le \hat{r}_0$ provided $\frac{1}{2} \nu_2 \hat{r}_0^2 \le E_0 - \max V_*$. Let $E_1 > E_0$. The set $E \le E_1$ is contained in the ball $|J| \le r_1$ provided $\frac{1}{2} \nu r_1^2 \ge E_1 - \min V_*$. Consequently, if the initial point satisfies $|J_0| \le \hat{r}_0$ then its energy $E(J_0, \varphi_0) \le E_0$ and if some iterate satisfies $|J_k| > r_1$ then its energy $E(J_k, \varphi_k) \ge E_1$.

Since $|V_*| \le \|s\|$ we can choose $E_0 = 2\|s\|$, $E_1 = 4\|s\|$, $\hat{r}_0^2 = 2\nu_2^{-1} \|s\|$ and $r_1^2 = 10\nu^{-1} \|s\|$. Then we can conclude that a trajectory with initial condition satisfying $|J_0| \le r_0$ remains in the ball $|J| \le r_1$ while the changes in E do not exceed $2\|s\|$.

In order to achieve exponential estimates for the stability times we need to consider the optimal approximation for the map by an autonomous Hamiltonian flow instead of the leading order approximation discussed above. For this purpose we can use Theorem 5.1 to get an embedding into a Hamiltonian flow with exponentially small error. We can repeat the arguments of Section 6.1 replacing in the definition of the domain $\mathcal{D}(I_*)$ the radius ρ_n by $\hat\rho_n = R_*\sqrt{\varepsilon}$. We choose a constant $R_* > r_1$ and $\gamma > R_*$. Then $\hat\rho_n \le \rho_n$ and consequently we already know that the scaled map is analytic and the a-priori bounds (28) remain valid. Using Cauchy estimates for the derivatives of the function s we get from (44)

$$|\bar J - J| \le \frac{2\|s\| n\sqrt{\varepsilon}}{r} \quad \text{and} \quad |\bar\varphi - \varphi| \le 2\|h_0''\| R_* n\sqrt{\varepsilon} + \frac{2\|s\| n\varepsilon}{\sigma}.$$

Taking $R_*^2 = 11\nu^{-1}\|s\|$ and using that $\varepsilon \le 1$ we get

$$|\bar J - J| \le C_0 n\sqrt{\varepsilon} \quad \text{and} \quad |\bar\varphi - \varphi| \le C_0 n\sqrt{\varepsilon}$$

where C_0 can be easily expressed in terms of $\|s\|$, r, σ, ν and $\|h_0''\|$.

Then using arguments similar to the previous section we arrive to the following theorem.

Theorem 7.1. *Under the assumptions of Theorem 2.2, there are constants* c_4, c_5 *independent of the resonance such that if* $|I_0 - I_*|^2 \le 2\nu_2^{-1}\|s\|\varepsilon$ *then*

$$|I_k - I_*|^2 \le 11\nu^{-1}\|s\|\varepsilon \quad \textit{for} \quad 0 \le k \le \hat T_\varepsilon = c_4 \exp(c_5/\sqrt{n^2\varepsilon}).$$

It should be noted that the proof of this theorem is a refinement of the proof of Theorem 2.2. Both theorems cover the same set of fully resonant tori and for each one Theorem 7.1 provides a nucleus, a smaller stability zone with longer stability times. For a fixed n the difference becomes more prominent as ε decreases. The estimate suggests that Arnold diffusion slows down substantially in a neighbourhood of resonances of maximal multiplicity provided the period n is not too high for a given ε.

In Theorem 7.1 the constant c_5 is chosen to be the same for all resonances. It should be noted that for some resonances the bounds for the stability times can be substantially improved (note that doubling c_5 is equivalent to squaring a very large number $\hat T_\varepsilon$). Indeed, at the centre of our proofs are the upper bounds for the sums in the right-hand side of the equation (44) which are used to control the distance of the map from the identity. These sums can be interpreted as average values of functions taken over a finite segment of a trajectory of the map and we used elementary but not always optimal bounds. For example, we used that $|V_*| \le \|s\|$, which does not take into account that the average value of a periodic function can be much smaller. A sharper bound can be obtained if we take into account properties of the frequency vector ω_*. We notice that the function V_*

inherits periodicity in φ from the function s and in addition $n\omega_* \in \mathbb{Z}^d$ implies that for all φ

$$V_*(\varphi + \omega_*) = V_*(\varphi).$$

It follows easily that all non-resonant Fourier coefficients of V_* must vanish, i.e., if $j \cdot \omega_* \notin \mathbb{Z}$ for some $j \in \mathbb{Z}^d$ then the Fourier expansion of V_* does not have a term proportional to $\exp(2\pi i\, j \cdot \varphi)$. In terms of Fourier expansions we can write

$$V_*(\varphi) = \sum_{j \cdot \omega_* \in \mathbb{Z}} s_j(I_*) \mathrm{e}^{2\pi i\, j \cdot \varphi}.$$

Since s is an analytic function of φ, its Fourier coefficients $s_j(I_*)$ decay exponentially fast when $|j|_1$ grows. Therefore the amplitude of V_* can be substantially smaller than $\|s\|$.

This observation suggests that in the absence of low order resonances the stability times should be much larger than the general lower bound $\hat{T}_\varepsilon$. This situation can arise either due to the properties of the frequency vector ω_* or due to the absence of the resonant terms in the Fourier expansion of the generating function. In particular, we expect that in the latter case the lower bound for the stability time scales as $\exp(c\varepsilon^{-\alpha}/n)$ with $\alpha > \frac{1}{2}$ (a phenomenon similar to [25]). On the other hand, in the case when a full spectrum condition is satisfied the lower bound for the stability time scales as $\exp(c_5 \varepsilon^{-1/2}/n)$ with a constant $c_5 \sim \mathrm{e}^{\pi r j_0}$, i.e. the constant becomes very large for larger values of j_0, the order of the lowest order resonance of ω_*.

8 Final comments and conclusions

Our proof of the Nekhoroshev estimates is based on discrete averaging. The weighted averages of iterates of the near-integrable map are explicitly computed to produce the interpolating vector field X_m. This vector field is not necessarily Hamiltonian but Theorem 5.1 states the existence of a Hamiltonian vector field $\hat{X}_m = J\nabla H_m$ very close to X_m. In a neighbourhood of a fully resonant torus the time-one maps of X_m and $\hat{X}_m$ are both exponentially close to f_ε^n for $n < N_\varepsilon$ and, consequently, the map preserves H_m up to an exponentially small error. The convexity arguments are used to show that level lines of H_m present obstacles for the drift of action variables.

The analytical tools developed in this paper rely on explicit constructions and provide a useful tool for analytical and numerical exploration of long term dynamics.

Computing a slow variable from iterates of the map in original variables. The value of H_m is a natural slowly moving observable which provides a useful instrument for studying long time stability and Arnold diffusion. Our method provides an explicit expression for this slow variable in terms of weighted averages of the iterates of the map, hence avoiding transformations of coordinates traditionally used to reduce the system to a normal form. In particular, we may construct the interpolating vector field X_m using the iterates of f_ε^n in the original coordinates (I, φ), see Remark (6.1).

Formal embedding of a near-the-identity map into an autonomous flow. Our method provides a new algorithm for constructing the formal embedding of a near-the-identity map into an autonomous flow. For example, let $f_\mu : (p, q) \mapsto (p_1, q_1)$ be defined with the help of a generating function

$$G_\mu(p_1, q) = p_1 q + \mu S(p_1, q).$$

We can get an explicit expression for $\hat{X}_{m,\mu}$ for any m by differentiating m times with respect to μ the system

$$\begin{cases} p_k = p_{k-1} - \mu \dfrac{\partial S}{\partial q}(p_k, q_{k-1}), \\ q_k = q_{k-1} + \mu \dfrac{\partial S}{\partial p}(p_k, q_{k-1}) \end{cases}$$

for $k = 1, \ldots, m$, and evaluating at $\mu = 0$. The derivatives depend in a polynomial way on partial derivatives of S and can be computed explicitly. Then the Hamiltonian H_m can be restored from the vector field. For example, the second order interpolating Hamiltonian for f_μ is

$$H_{2,\mu} = \mu S - \mu^2 \frac{1}{2} \frac{\partial S}{\partial p} \cdot \frac{\partial S}{\partial q}$$

where all functions are evaluated at a point (p, q). This argument can also be applied to an individual map with a generating function $G(p_1, q) = p_1 q + S(p_1, q)$. This map is approximated by the time one map of the flow defined by

$$H_2 = S - \frac{1}{2} \frac{\partial S}{\partial p} \cdot \frac{\partial S}{\partial q}$$

with the error cubic in $\epsilon = \|\nabla S\|$ and explicitly computable constants according to Remark 5.2.

Numerical evaluation of H_m. In numerical computations it is usually not convenient to rely on algebraic manipulations and instead one can evaluate $H_m(x)$ from integrals of X_m along continuous paths connecting a base point p and the point x, see [10]. Note that this procedure typically produces a Hamiltonian which is not periodic in the angle variables even when X_m is periodic. The periodicity of the Hamiltonian can be restored by adding a small correction.

The choice of the interpolation scheme. In this paper we have used the Newton interpolation scheme to obtain X_m. This scheme uses the forward orbit

$x_0, \ldots, x_m$ for the construction of X_m and simplifies some analytical expressions involved in the proof. But any other interpolation scheme will lead to similar results.

From the numerical point of view, higher accuracy of interpolation is expected when interpolation nodes are located symmetrically around x_0. This can be useful for numerical studies of concrete examples when relatively large values of ε are to be used in order to observe Arnold diffusion on a time scale accessible to the computer.

For example we can use the Gauss forward formula

$$P_m(t; x_0) = x_0 + \Delta_1(x_0)t + \frac{\Delta_2(x_{-1})}{2!} t(t-1) + \frac{\Delta_3(x_{-1})}{3!}(t+1)t(t-1) + \ldots$$

If $m = 2j$ is even, then the interpolation is based on a symmetrical piece of the orbit, $x_{-j}, \ldots, x_0, \ldots, x_j$. Differentiating with respect to t at $t = 0$ we obtain the interpolating vector field

$$X_m(x_0) = \sum_{k=1}^{j} (-1)^{k-1} \left(\frac{(k-1)!^2 \Delta_{2k-1}(x_{-k+1})}{(2k-1)!} - \frac{(k-1)!k!\Delta_{2k}(x_{-k})}{(2k)!} \right) \tag{46}$$

instead of (13). This interpolating vector field can be used in Theorem 5.1 provided two steps in the proof are modified. First, the constant in the bound (19) depends on the coefficients of P_m. Using (46) and the bound (18) for the finite differences it is easy to check that for the Gaussian symmetric scheme the same upper bound holds. Second, given a number m of iterates of the original map, the Gauss symmetric scheme allows us to double the value of μ_m in the proof of Theorem 5.1. This leads to better accuracy of the embedding of the map into a flow with the error being of the order $\sim \exp(-\delta/3e\varepsilon)$, i.e. we get the error term approximately squared.

Nucleus of resonances. The fact that the methodology to obtain the estimates on the long term dynamics of the map does not depend on changes of coordinates leads to a description of the leading order dynamics near the nucleus of the resonances, that is, in a ball of radius $\mathcal{O}(\sqrt{\varepsilon})$ near a resonant torus $I = I_*$. The corresponding energy preservation leads to much larger stability times for initial conditions in the nucleus of the resonances. The construction is explicit. In particular, the potential part of the Hamiltonian is given by the average of the generating function $s(I, \varphi)$ of f_ε along the unperturbed periodic orbit corresponding to the resonant torus $I = I_*$.

Finally we note that we have used the convexity assumption for the generating function of the unperturbed map. At the present time it is not clear up to which extent the convexity assumption can be relaxed in our proof. Nevertheless we expect that our method can be useful for studying systems without the convexity assumption. In this case our method produces a very slow variable near a resonance but the corresponding level lines are not necessarily an obstacle for the movements of actions. Nevertheless, the slow variables may provide useful information on possible directions of Arnold diffusion. We also hope that our method can be applied to study dynamics of near integrable systems without references to action-angle variables for the integrable part, opening potential applications to study dynamics of maps in neighbourhoods of totally elliptic fixed points.

A Implicit function theorem

In the proof we switch between a symplectic map and the corresponding generating function. This transition relies on the following version of the implicit function theorem.

Theorem A.1 (Implicit Function Theorem)**.** *Let $A, B \subset \mathbb{C}^d$ be open sets, $f : A \times B \to \mathbb{C}^d$ an analytic function, $x_0 \in A$ and $y_0 \in B$. If there is $R > 0$ such that $B_R(y_0) \subset B$ and*

$$M = \sup_{y \in B_R(y_0)} |f(x_0, y)| < \frac{R}{d+1}$$

then the equation

$$y = y_0 + f(x_0, y)$$

has a unique solution $y \in B_R(y_0)$. Moreover, this solution depends analytically on x_0, y_0.

Proof. We use the contracting mapping theorem (the ∞-norm is used for vectors in $\mathbb{C}^d$). The closed ball $\overline{B_M(y_0)}$ is invariant under the map

$$g : y \mapsto y_0 + f(x_0, y).$$

In order to check that g is contracting we take $u, v \in \overline{B_M(y_0)}$, then

$$\begin{aligned}
|g_j(u) - g_j(v)| &= |f_j(x_0, u) - f_j(x_0, v)| \\
&= \left| \int_0^1 \sum_{k=1}^d \partial_{y_k} f_j(x_0, us + (1-s)v)(u_k - v_k) ds \right| \\
&\le \int_0^1 \sum_{k=1}^d |\partial_{y_k} f_j(x_0, us + (1-s)v)| \; |u_k - v_k| ds \\
&\le \sum_{k=1}^d \frac{\|f_j\|_{B_R}}{R - M} |u_k - v_k| = \frac{Md}{R - M} |u - v|
\end{aligned}$$

where we used the Cauchy bound for the derivatives. The inequality $M < \frac{R}{d+1}$ implies $\frac{Md}{R-M} < 1$. Therefore the map g is contracting and it has a unique fixed point which depends analytically on the parameters. □

B Formal interpolation of a symplectic family

Let B denote a ball (or a simply connected domain) in $\mathbb{R}^d$ and let F_μ be an analytic family of exact symplectic maps defined in $B \times \mathbb{T}^d$ with $F_0 = \mathrm{Id}$. The following theorem represents a generally known statement (see e.g. [5]). Here we provide a more direct proof of the statement in the form needed for the proof of our main theorem.

Theorem B.1. *If X_μ is the formal vector field on $B \times \mathbb{T}^d$ such that its formal time one map coincides with the Taylor expansion of F_μ, then there is a formal Hamiltonian H_μ with coefficients defined on $B \times \mathbb{T}^d$ such that $X_\mu = \mathrm{J}\nabla H_\mu$.*

Proof. We have already proved existence of the formal vector field

$$X_\mu = \sum_{k \geq 1} \mu^k X^{(k)}(p,q)$$

where the coefficients are smooth functions independent of μ and periodic in q. We want to show that if F_μ is symplectic for every μ then the formal vector field is Hamiltonian, i.e. for every k, $X^{(k)} = (-\partial_q h^{(k)}, \partial_p h^{(k)})$ for some function $h^{(k)} : B \times \mathbb{R}^d \to \mathbb{R}$. If in addition F_μ are exact symplectic, then $h^{(k)}$ are periodic in q. The proof is based on analysis of loop actions.

Let $\gamma : [0,t] \to \mathbb{R}^{2d}$ be a smooth curve inside the domain of the map such that $\gamma(1) - \gamma(0) \in \{0\} \times \mathbb{Z}^d$, i.e. γ is a lift of a loop from $\mathbb{R}^d \times \mathbb{T}^d$. Let

$$\mathcal{I}_F(\gamma) = \int_{F(\gamma)} p\, dq - \int_\gamma p\, dq.$$

If F is symplectic, Stokes' theorem implies that $\mathcal{I}_F(\gamma)$ depends only on the homotopy class of γ and $\mathcal{I}_F(\gamma) = 0$ for any contractile loop γ. If F is exact symplectic map then $\mathcal{I}_F(\gamma) = 0$ for any lift γ of a loop from $\mathbb{R}^d \times \mathbb{T}^d$. For a flow defined by a vector field $X = (X_p, X_q)$ we write $\gamma_t(s) = (p(t,s), q(t,s)) := \Phi_X^t \gamma(s)$. Then

$$\begin{aligned}
\frac{d}{dt}\mathcal{I}_{\Phi_X^t}(\gamma) &= \frac{d}{dt}\int_{\Phi_X^t(\gamma)} p\, dq \;=\; \frac{d}{dt}\int_0^1 p(t,s)\,\frac{\partial q}{\partial s}(t,s) ds \\
&= \int_0^1 \left(\frac{\partial p}{\partial t}(t,s)\,\frac{\partial q}{\partial s}(t,s) + p(t,s)\,\frac{\partial^2 q}{\partial s \partial t}(t,s) \right) ds \\
&= \int_0^1 \left(\frac{\partial p}{\partial t}(t,s)\frac{\partial q}{\partial s}(t,s) - \frac{\partial p}{\partial s}(t,s)\,\frac{\partial q}{\partial t}(t,s) \right) ds \\
&= \int_{\Phi_X^t\gamma(0)}^{\Phi_X^t\gamma(1)} \left(X_p dq - X_q dp \right).
\end{aligned}$$

The right hand sides vanishes for every γ with $\gamma(0) = \gamma(1)$ iff X is Hamiltonian, i.e $X_p = -\partial_q H$ and $X_q = \partial_p H$ for some function $H : B \times \mathbb{R}^d \to \mathbb{R}$. In this case the integral can be evaluated explicitly for any lift γ,

$$\frac{d}{dt}\mathcal{I}_{\Phi_X^t}(\gamma) = H(\Phi_X^t(\gamma(0))) - H(\Phi_X^t(\gamma(1))) = H(\gamma(0)) - H(\gamma(1)).$$

Since $\mathcal{I}_{\Phi_X^0}(\gamma) = 0$ we get

$$\mathcal{I}_{\Phi_X^1}(\gamma) = H(\gamma(0)) - H(\gamma(1)).$$

Recall that the coefficients of the formal vector field are defined form the following requirement: for every $m \in \mathbb{N}$, a partial sum of the formal series

$$X_{m,\mu}(p,q) = \sum_{k=1}^{m} \mu^k X^{(k)}(p,q)$$

defines the flow $\Phi^1_{X_{m,\mu}} = F_\mu + O(\mu^{m+1})$. If F_μ are symplectic, then $\mathcal{I}_{F_\mu}(\gamma) = 0$ for all contractible loops γ and, consequently, for $k \le m$

$$\oint_\gamma \left(X_p^{(k)} dq - X_q^{(k)} dp \right) = 0.$$

Therefore X_μ is Hamiltonian.

If F_μ are exact symplectic, then $\mathcal{I}_{\Phi^1_{X_{m,\mu}}} = \mathcal{I}_{F_\mu}(\gamma) + O(\mu^{m+1}) = O(\mu^{m+1})$ for all lifts γ. Since the Hamiltonian of $X_{m,\mu}$ is polynomial of degree m in μ we conclude that all coefficients $h^{(k)}$ are periodic in q. □

C Generating functions of exact symplectic maps

In this paper we need to find a generating function for a near-the-identity exact symplectic map defined on a subset of $\mathbb{R}^d \times \mathbb{T}^d$.

We recall that a map is called symplectic if it preserves the standard symplectic form

$$\omega = \sum_{l=1}^{d} dp_l \wedge dq_l.$$

The map is exact if it preserves the loop action

$$A(\gamma) = \oint_\gamma \sum_{l=1}^{d} p_l \, dq_l$$

for all loops inside its domain. A symplectic map automatically preserves loop actions for contractible loops.

Suppose that $f : (p, q) \mapsto (\bar p, \bar q)$ is a lift of a symplectic map. We assume that the map is analytic in a neighbourhood of $B_\rho \times \mathbb{R}^d$. Suppose that we can rewrite f in the cross form

$$\begin{aligned} p &= \bar p + u(\bar p, q),\\ \bar q &= q + v(\bar p, q), \end{aligned} \tag{47}$$

where the functions $u = (u_1, \ldots, u_d)$ and $v = (v_1, \ldots, v_d)$ are periodic in q. Since the map is symplectic we get that $\sum_{l=1}^d d\bar p_l \wedge d\bar q_l = \sum_{l=1}^d dp_l \wedge dq_l$ and we get

$$\begin{aligned} \sum_{l=1}^d d(u_l dq_l + v_l d\bar p_l) &= \sum_{l=1}^d (du_l \wedge dq_l + dv_l \wedge d\bar p_l)\\ &= \sum_{l=1}^d ((dp_l - d\bar p_l) \wedge dq_l + (d\bar q_l - dq_l) \wedge d\bar p_l)\\ &= \sum_{l=1}^d dp_l \wedge dq_l - \sum_{l=1}^d d\bar p_l \wedge d\bar q_l = 0. \end{aligned}$$

Then we choose a base point $(\bar p_0, q_0)$ and define

$$s(\bar p, q) = \int_{(\bar p_0, q_0)}^{(\bar p, q)} \sum_{l=1}^d (u_l dq_l + v_l d\bar p_l). \tag{48}$$

The previous argument implies that for a symplectic f the value of the integral is independent of the path connecting the end points as the domain is simply connected. Differentiating the integral we see that

$$u_l = \frac{\partial s}{\partial q_l}, \qquad v_l = \frac{\partial s}{\partial \bar p_l}.$$

Consequently the map f can be defined with the help of the generating function $\bar p \cdot q + s(\bar p, q)$. Let e_l denote a vector of the canonical basis in $\mathbb{R}^d$. Then

$$s(\bar p, q + e_l) - s(\bar p, q) = \int_{(\bar p_0, q_0)}^{(\bar p_0, q_0 + e_l)} \sum_{l=1}^d (u_l dq_l + v_l d\bar p_l).$$

Now suppose that f is homotopic to the identity and consider a smooth curve $\gamma_l = (p(t), q(t))$ such that $q(1) = q(0) + e_l$. Let $\bar\gamma_l = (\bar p(t), \bar q(t))$ be the image of this curve. Since the map is homotopic to the identity we have $\bar q(1) = \bar q(0) + e_l$.

We compute the difference of the loop actions:

$$
\begin{aligned}
A(\bar{\gamma}_l) - A(\gamma_l) &= \int_0^1 \sum_{l=1}^d (\bar{p}_l(t)\, d\bar{q}_l(t) - p_l(t)\, dq_l(t)) \\
&= \int_0^1 \sum_{l=1}^d (\bar{p}_l(t)\, dv_l(\bar{p}(t), q(t)) - u_l(\bar{p}(t), q(t))\, dq_l(t)) \\
&= \int_0^1 \sum_{l=1}^d \big(-v_l(\bar{p}(t), q(t)) d\bar{p}_l(t) - u_l(\bar{p}(t), q(t))\, dq_l(t)\big) \\
&= -\int_{(\bar{p}(0),q(0))}^{(\bar{p}(0),q(0)+e_l)} \sum_{l=1}^d \big(v_l d\bar{p}_l + u_l dq_l\big) = -s(\bar{p}(0), q(0)+e_l) + s(\bar{p}(0), q(0)).
\end{aligned}
$$

We see that the conservation of loop actions is equivalent to the periodicity of s.

Acknowledgements

A.V. is supported by the Spanish grant PID2021-125535NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU. He also acknowledges the Catalan grant 2021-SGR-01072 and the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D (CEX2020-001084-M).

References

[1] V.I. Arnol'd. Instability of dynamical systems with several degrees of freedom. *Soviet Math. Dokl.*, 5:581–585, 1964.

[2] V.I. Arnol'd. Proof of a theorem of a. n. kolmogorov on the invariance of quasi-periodic motions under small perturbations of the hamiltonian. *Russian Math. Surveys*, 18(5):9–36, 1963.

[3] D. Bambusi and B. Langella. A C^∞ Nekhoroshev theorem. *Math. Eng.*, 3(2):Paper No. 019, 17, 2021.

[4] G. Benettin, L. Galgani, and A. Giorgilli. A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems. *Celestial Mech.*, 37:1–25, 1985.

[5] G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms. *J. Statist. Phys.*, 74(5-6):1117–1143, 1994.

[6] L. Biasco and L. Chierchia. Explicit estimates on the measure of primary KAM tori. *Ann. Mat. Pura Appl. (4)*, 197(1):261–281, 2018.

[7] A. Bounemoura. Nekhoroshev estimates for finitely differentiable quasi-convex Hamiltonians. *J. Differential Equations*, 249(11):2905–2920, 2010.

[8] A. Bounemoura and L. Niederman. Generic Nekhoroshev theory without small divisors. *Ann. Inst. Fourier (Grenoble)*, 62(1):277–324, 2012.

[9] J.W.S. Cassels. *An introduction to Diophantine approximation*, volume No. 45 of *Cambridge Tracts in Mathematics and Mathematical Physics.* Cambridge University Press, New York, 1957.

[10] V. Gelfreich and A. Vieiro. Interpolating vector fields for near identity maps and averaging. *Nonlinearity*, 31(9):4263–4289, 2018.

[11] M. Guzzo. A direct proof of the Nekhoroshev theorem for nearly integrable symplectic maps. *Ann. Henri Poincaré*, 5(6):1013–1039, 2004.

[12] M. Guzzo, L. Chierchia, and G. Benettin. The steep Nekhoroshev's theorem. *Comm. Math. Phys.*, 342(2):569–601, 2016.

[13] S. Kuksin and J. Pöschel. On the inclusion of analytic symplectic maps in analytic Hamiltonian flows and its applications. In *Seminar on Dynamical Systems (St. Petersburg, 1991)*, volume 12 of *Progr. Nonlinear Differential Equations Appl.*, pages 96–116. Birkhäuser, Basel, 1994.

[14] P. Lochak. Canonical perturbation theory via simultaneous approximation. *Russian Math. Surveys*, 47(6):57–133, 1992.

[15] P. Lochak and A.I. Neishtadt. Estimates of stability time for nearly integrable systems with a quasiconvex hamiltonian. *Chaos*, 2:495, 1992.

[16] P. Lochak, A.I. Neishtadt, and L. Niederman. Stability of nearly integrable convex Hamiltonian systems over exponentially long times. In *Seminar on Dynamical Systems (St. Petersburg, 1991)*, volume 12 of *Progr. Nonlinear Differential Equations Appl.*, pages 15–34. Birkhäuser, Basel, 1994.

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

[18] J.K. Moser. Lectures on Hamiltonian systems. *Mem. Amer. Math. Soc.*, 81:60, 1968.

[19] A.I. Neishtadt. Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions. *J. Appl. Math. Mech.*, 45(6):766–772, 1981.

[20] A.I. Neishtadt. The separation of motions in systems with rapidly rotating phase. *J. Appl. Math. Mech.*, 48:133 – 139, 1984.

[21] N.N. Nekhoroshev. An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. *Russian Math. Surveys*, 32:1–65, 1977.

[22] N.N. Nekhoroshev. An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. II. *Trudy Sem. Petrovsk.*, No. 5:5–50, 1979.

[23] J. Pöschel. Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Math. Z.*, 213(2):187–216, 1993.

[24] J. Pöschel. Integrability of hamiltonian systems on cantor sets. *Communications on Pure and Applied Mathematics*, 35(5):653–696, 1982.

[25] C. Simó. Averaging under fast quasiperiodic forcing. *Hamiltonian Mechanics: Integrability and Chaotic Behavior*, Springer US:13–34, 1994.