Evolution of slow variables in *a priori* unstable Hamiltonian systems

D. Treschev *

Abstract

We study diffusion phenomena in *a priori* unstable (initially hyperbolic) Hamiltonian systems. These systems are perturbations of integrable ones, which have a family of hyperbolic tori. We prove that in the case of two and a half degrees of freedom the action variable generically drifts (i.e., changes on a trajectory by a quantity of order one). Moreover, there exists a trajectory such that the velocity of this drift is $\varepsilon/\log\varepsilon$, where ε is the parameter of the perturbation.

1 Introduction

In [1] Arnold proposed an example of a near-integrable Hamiltonian system

$$\dot{x} = \partial H/\partial y, \quad \dot{y} = -\partial H/\partial x, \quad H = H_0(y) + \varepsilon H_1(x, y, t, \varepsilon), \qquad (1.1)$$
$$x \in \mathbf{T}^n = \mathbf{R}^n/\mathbf{Z}^n, \quad y \in \mathbf{R}^n, \quad t \in \mathbf{T}, \quad \varepsilon \in (-\varepsilon_0, \varepsilon_0)$$

with convex in the actions y unperturbed Hamiltonian H_0, where the variables y can change by a quantity of order 1 on a trajectory. Later this phenomenon was called the Arnold diffusion. Arnold [1] also presented a mechanism for such a diffusion. It is well-known: a chain of hyperbolic tori (a transition chain) should exist in the perturbed problem. The corresponding chain of stable and unstable asymptotic surfaces should be connected by heteroclinic trajectories. Then the drift of the actions appears on a trajectory, which follows this chain of asymptotic surfaces.

The main problem associated to the Arnold diffusion is its genericity for systems (1.1). It is very important what smoothness assumptions are imposed on H. Real-analytic case is usually regarded as the most interesting.

According to the Nekhoroshev theory [29], in real-analytic systems, satisfying a rather weak steepness condition, average velocity of the action drift along a trajectory is estimated from above by an exponentially small quantity $e^{-\alpha\varepsilon^{-\beta}}$ with some positive α, β. Actually, the exponentially small effects present the main difficulty in the analysis of the phenomenon. Another difficulty is connected with the fact that tori in a transition chain generically form not continuous, but Cantorian family. In particular, such a family contains gaps. It is clear that these gaps should not be too large. Otherwise it is very difficult for asymptotic surfaces of two tori, separated by a gap, to reach each other. Estimates for the width of gaps must use KAM technics. Because of this they should be cumbersome and nontrivial.

When constructing his example, Arnold did not overcome these difficulties, but went around them by using some tricks. In the Arnold example transition chain is formed by a smooth family of hyperbolic tori. The problem of exponentially small effects dissapears, because the perturbation depends on two small parameters: ε and $\delta \sim e^{-c/\sqrt{\varepsilon}}$, $c > 0$.

*The work was partially supported by Russian Foundation of Basic Research grants 02-01-00400 and 00-15-99269, and by INTAS grant 00-221

The parameter ε is responsible for the appearence of hyperbolic tori, and δ for the existence of heteroclinic connections between them. Since δ is exponentially small with respect to ε, existence of these connections can be detected by the Poincaré–Melnikov method.

Douady [18] established in C^∞-smooth Hamiltonian systems with 3 degrees of freedom the existence of formaly stable elliptic equilibriums which are Lyapunov unstable. The instability is generated by the Arnold diffusion. Slightly smoother situation (Gevrey-α case, $\alpha > 1$) is considered in [25]. Velocity of the diffusion is estimated.

Generalizing the Arnold's ideas, Fontich and Martin [20] constructed a large class of real-analytic systems (still highly non-generic) with diffusion trajectories. Perturbations, depending on two parameters are considered in [21]. However, unlike the original Arnold paper, the second small parameter, δ, is not exponentially small with respect to the first one, ε, i.e., $\delta = \varepsilon^p$, where p is sufficiently large. Variational approach for the Arnold example was proposed in [4]. Simó and Valls [31] performed numerical analysis of the problem in the Arnold example with $\varepsilon = \delta$.

Recently Mather [27] announced variational proof of genericity of the Arnold diffusion for systems (1.1) with positive definite Hessian $\partial^2 H_0/\partial y^2$ in the case of two and a half degrees of freedom.

There are several simpler problems, where diffusion-like phenomena (i.e., drift of slow variables along transition chains) occur without exponentially small effects.

First, we mention the Mather problem [26]. Consider a 2-torus with some Riemannian metric and a perturbation of a free motion of a particle on the torus by a time-periodic potential. Then generically there exists a trajectory on which the energy tends to infinity.[1] The geometric mechanism of this on a first glance quite unexpected phenomenon is as follows. Let γ be a minimal geodesic in some homotopy class of closed curves on the torus. It is well-known that generically this geodesic is hyperbolic i.e., exponentially unstable. Since the perturbation is nonautonomous, it is natural to regard time as a phase variable. Then γ generates in the unperturbed system a 1-parametric family of hyperbolic 2-dimensional tori. Projection of each torus of the family to the configuration space is $\gamma \times \mathbf{T}_t$, where $\mathbf{T}_t$ is the circle of time. The family is parametrized by the velocity of the particle motion on γ. After the perturbation majority of the tori survive, heteroclinic intersections of their asymptotic surfaces appear, and therefore, we obtain a transition chain. The coresponding proofs are contained in [10] and [15]. Note that the original Mather's proof, based on variational methods, is still not published. A proof, partially based on Mather's variational ideas is presented in [23].

An unusual set up (quite far from the original Arnold's one) was proposed by Easton, Meiss and Roberts [19]. They consider diffusion in a system, which is not near-integrable. On the contrary, the system is close to the anti-integrable limit [2, 3, 24]. Because of a strong hyperbolic properties of the system the construction of diffusion trajectories turns out to be simple and elegant.

In this paper we deal with the so-called *a priori* unstable systems. The terminology is taken from [12], where the systems (1.1) are called *a priori* stable.

As usual, it is more convenient to consider the case of a non-autonomous system with periodic in time Hamiltonian function. The general form of the Hamiltonian of a near-integrable non-autonomous *a priori* unstable system with two and a half degrees of freedom is as follows:

$$H(y, x, v, u, t, \varepsilon) = H_0(y, v, u) + \varepsilon H_1(y, x, v, u, t) + O(\varepsilon^2). \qquad (1.2)$$

Here $x \bmod 1$ is a point of the torus $\mathbf{T} = \mathbf{R}/\mathbf{Z}$, $y \in \mathcal{D}_0 \subset \mathbf{R}$, where $\mathcal{D}_0$ is an interval, the pair (v, u) belongs to an open domain $D \subset \mathbf{R}^2$. The pairs x and y, u and v are canonically conjugate i.e., the symplectic structure

[1] The torus can be replaced by any compact 2-dimensional manifold of positive genus. There are also multidimensional generalizations [10].

is as follows: $\omega = dy \wedge dx + dv \wedge du$. The Hamiltonian equations have the form

$$\dot y = -\partial H/\partial x, \quad \dot x = \partial H/\partial y, \quad \dot v = -\partial H/\partial u, \quad \dot u = \partial H/\partial v. \qquad (1.3)$$

We assume that H is 1-periodic in t and the parameter ε is small. Below we regard t as a point of the torus $\mathbf{T}$. It is convenient to assume that $\varepsilon \geq 0$.

The system with Hamiltonian H_0 is integrable. We call it unperturbed. The Hamiltonian (1.2) satisfies several hypotheses.

$\mathbf{H_0 1}$. *The function H is $C^{\mathbf{r}}$-smooth in all arguments, $\mathbf{r} \in \{6, 7, \ldots, \infty, \omega\}$ and H_0 is real-analytic.*

In fact, the smoothness assumptions can be considerably weakened: it is sufficient to have a finite large enough smoothness for H_0 as well.

$\mathbf{H_0 2}$. *For any $y^0 \in \mathcal{D}_0$ the function $H_0(y^0, v, u)$ has a nondegenerate saddle point $(v, u) = (v^0, u^0)$, smoothly depending on y^0. Any point (v^0, u^0) belongs to a compact connected component of the set*

$$\widehat\gamma(y^0) = \{(v, u) \in D : H_0(y^0, v, u) = H_0(y^0, v^0, u^0)\}.$$

In dynamical terminology (v^0, u^0) is a hyperbolic equilibrium of the system with one degree of freedom and Hamiltonian $H_0(y^0, v, u)$, and the corresponding separatrices $\widehat\gamma$ are doubled. In particular, topologically these separatrices form a figure-eight-like structure.

A priori unstable systems with a larger number of degrees of freedom are defined similarly, but x and y variables become multi-dimensional.

We denote the loops of the "eight" by $\widehat\gamma^{\pm}(y^0)$, where $\widehat\gamma^{+}(y^0)$ is called the upper loop and $\widehat\gamma^{-}(y^0)$ the lower one, $\widehat\gamma(y^0) = \widehat\gamma^{+}(y^0) \cup \widehat\gamma^{-}(y^0)$. The loops $\widehat\gamma^{\pm}(y^0)$ have the natural orientation generated by the flow of the system. We can define an orientation on D by the coordinate system u, v.

$\mathbf{H_0 2^*}$. *For any $y^0 \in \mathcal{D}_0$ the natural orientation of $\widehat\gamma^{\pm}(y^0)$ coincides with the orientation of the domain D i.e., the motion on the separatrices is counterclockwise (see Fig 1).*

This hypothesis is not restrictive. Indeed, if it does not hold, we perform the change $t \mapsto -t$, $H \mapsto -H$.

In this paper we assume that two other hypotheses hold:

$\mathbf{H_0 3}$. *In the unperturbed Hamiltonian the variables y are separated from u and v i.e., $H_0(y, v, u) = F(y, f(v, u))$.*

Because of $\mathbf{H_0 3}$ location of an equilibrium (v^0, u^0) does not depend on y. Below without loss of generality we assume that $v^0 = u^0 = 0$.

In this paper we answer two important questions concerning the diffusion in *a priori* unstable systems.

(I) Is this phenomenon generic i.e., takes place for generic H_0, satisfying $\mathbf{H_0 1}$–$\mathbf{H_0 3}$ and generic H_1 for some reasonable meaning of genericity?

(II) What is the maximal velocity of the drift of y?

We present positive answer to question (I) and give the lower estimate for the velocity of the diffusion $\mathrm{const} \cdot \varepsilon/|\log \varepsilon|$.

We put

$$E(y) = H_0(y, 0, 0), \quad \nu = E_y : \mathcal{D}_0 \to \mathbf{R}.$$

The following hypothesis depends on a positive parameter c_*.

$\mathbf{H_0 4}(c_*)$. *The function $\nu'(y) = d\nu(y)/dy$ does not vanish at strong resonances i.e., at the points, where $\nu = p/q$, $p \in \mathbf{Z}$, $q \in \mathbf{N}$, $q < 1/c_*$. Furthermore, for any $y \in D_0$ we have: $|\nu'(y)| + |\nu''(y)| > 2c_\nu$ for some $c_\nu > 0$.*

Let $\mathcal{D} \subset \mathcal{D}_0$ be an interval such that its closure $\overline{\mathcal{D}}$ also belongs to $\mathcal{D}_0$.

Theorem 1 *Suppose that H_0 satisfies* $\mathbf{H_0 1}$–$\mathbf{H_0 3}$. *Then for generic H_1 there exist $c_*, \varepsilon_0 > 0$ such that if $\mathbf{H_0 4}(c_*)$ holds and $\varepsilon \in (0, \varepsilon_0)$, the perturbed system has a trajectory*

$$\big(x(t), y(t), v(t), u(t)\big), \qquad t \in [0, T] \tag{1.4}$$

such that
(i) $\overline{\mathcal{D}} \subset [y(0), y(T)]$,
(ii) $c_1 < \frac{\varepsilon}{|\log \varepsilon|} T < c_2$ for some $c_1, c_2 > 0$.

Remarks. **1** The words "generic H_1" mean that H_1 belongs to an open dense set in $C^{\mathbf{r}}(\mathcal{D}_0 \times \mathbf{T} \times D \times \mathbf{T}, \mathbf{R})$. More precisely, H_1 must satisfy hypotheses $\mathbf{H_1 1}$–$\mathbf{H_1 3}$ (see Section 5). Conditions $\mathbf{H_1 1}$–$\mathbf{H_1 3}$ are constructive and, in principle, can be checked for a given H_1.

It is possible to formulate the theorem in a slightly another way. We can put $\varepsilon = 1$ and say that H_1 is small enough in the $C^{\mathbf{r}}$-norm. But then instead of the genericity condition for H_1 we have to say that H_1 belongs to a cusp-residual subset of a small ball in $C^{\mathbf{r}}$ (compare with [27]). These two possibilities are obviously equivalent.

2 Assertion (i) means that (1.4) is a diffusion trajectory, passing through $\overline{\mathcal{D}}$, and according to (ii), the average velocity of the diffusion is of order $\varepsilon/|\log \varepsilon|$. The same estimate for the diffusion velocity was obtained in [5] for quasi-periodic perturbation of a pendulum, in [14], under the assumption that a transition chain is given, and in [7, 6] in a "non-resonant" domain of a multi-dimensional *a priori* unstable version of the Arnold example. Note that perturbations in these papers are trigonometric polynomials i.e., non-degeneric.

3 Variational approach to the problem of diffusion in *a priori* unstable systems is developed in [36, 37]. More traditional geometric methods combined with new ideas in the vicinity of resonances are used in [16]. Note that in these papers speed of the diffusion is not estimated.

3 Theorem 1 remains true if equations (1.3) have non-conservative terms of order $O(\varepsilon^2)$.

4 It is clear that Theorem 1 in fact, gives the existence of infinitely many trajectories with properties (i)–(ii). The obvious lower estimate for the measure μ of the corresponding initial conditions $\mu \geq e^{-cT}$, $c > 0$ is apparently, quite realistic in the sense that the upper estimate should be similar.

Example. Consider the system (1.3) with $H = H_0 + \varepsilon H_1$,

$$\begin{aligned} H_0 &= \frac{y^2}{2} + \frac{v^2}{2} + \Omega^2 \cos(2\pi u), \quad H_1 = \cos(2\pi u) f(x, t), \qquad (1.5) \\ f(x, t) &= a\big(\cos(2\pi x) + \cos(2\pi t) + b\cos(2\pi(x - t)) + b\cos(2\pi(x + t))\big). \end{aligned}$$

The unperturbed system is a direct product of a rotator on the cylinder $(x \bmod 1, y)$ and a pendulum on the cylinder $(u \bmod 1, v)$.

We put $\mathcal{D}_0 = (-3/2, 3/2)$. Since we are intrested in the dynamics near the separatrices of the pendulum, we can regard D as a neighborhood of the separatrices. Then D is homeomorphic to a neighborhood of an eight figure on a plane. The unperturbed Hamiltonian H_0 obviously satisfies Hypotheses $\mathbf{H_0 1}$–$\mathbf{H_0 4}(c_*)$ with arbitrary c_*.

Below for convenience we put

$$\Omega = \pi/2, \quad a = \pi^2/8.$$

Theorem 2 *There exists a positive constant b_0 such that for any $b \in (0, b_0)$ the perturbation H_1 is generic in the sense of Theorem 1.*

Corollary 1.1 *Suppose that $b \in (0, b_0)$. Then for sufficiently small $\varepsilon > 0$ the system (1.3), (1.5) has a trajectory whose y-component passes the interval $(-1, 1)$ during a time interval of order $\log \varepsilon / \varepsilon$.*

Proof of Theorem 1 uses perturbative methods, but it does not follow the traditional strategy (construction of a transition chain and a "shadowing" orbit). Our proof is based on the analysis of the corresponding separatrix map. Due to the dynamical nature of our technics, we do not use cumbersome methods of KAM-theory.

We do not need the existence of perturbed hyperbolic tori. Hence, formally we do not have the "large gap" problem i.e., the problem to construct a heteroclinic connection between two tori, located on a "large" distance from each other. (See detailed discussion of this problem and some methods to solve it in [16]). However, the "large gap" problem does not disappear completely. Its trace (in another language) can be found in the part of our paper dealing with crossing resonances. Our method is similar to symbolic dynamics. We construct codes which push a trajectory in a given direction. By the Attachment lemma (Section 4) a code generates a trajectory. There are some strong restrictions for a choice of a code. In a non-resonance domain a proper code can be constructed easily. At strong resonances the situation is more complicated: sometimes we even have to return back a little on y-line, but fortunately, succeed to preserve an evolution in a proper direction with a proper average velocity.

Our proof consists basically of 3 steps.

I. We obtain formulas for the separatrix map [34].

II. We use these formulas to construct a kind of symbolic dynamics for the perturbed system [35]. Due to this it is possible to produce a large class of trajectories.

III. In the present paper we construct diffusion trajectories.

This paper is a continuation of [34]-[35]. Below we present all necessary information from [34]-[35].

The plan of the paper is as follows. In Section 2 we define the separatrix map. In Section 3 we present explicit formulas for the separatrix map and explain how constants and functions from these formulas can be expressed in terms of H_0 and H_1. In the end of Section 3 we calculate these functions for system defined by (1.5).

Section 4 contains our main technical tool, the Attachment lemma. This lemma establishes in the system a kind of symbolic dynamics. Unlike the usual situation we are not able to construct an infinite code in advance. We extend the code and the corresponding trajectory step by step. Arbitrariness in the choice of a code is used in the next sections to construct a diffusion trajectory.

In Section 5 we present genericity hypotheses for H_1 and check them for H_1, satisfying (1.5). Here we also divide $\overline{\mathcal{D}}$ into 3 subsets: $\overline{\mathcal{D}} = \mathcal{D}_N \cup \mathcal{D}_{CR} \cup \mathcal{D}_V$ (nonresonant, clear resonant and vague sets). Each set consists of a finite number of intervals. Dynamics differs considerably when y belongs to different sets. So we have to deal with different sets separately. We construct a trajectory with η passing through $\overline{\mathcal{D}}$ in 3 steps. First (Section 6) we show how to pass through nonresonant intervals. Then (Section 7) we deal with a clear resonant interval. Finally (Sections 8–11) we consider the case of a vague interval. Sections 13-14 contain proofs of some technical statements.

In the paper we denote by c all constants which do not play an essential role. These constants are assumed to be positive and small. The angular brackets $\langle \, , \rangle$ denote the standard inner product in $\mathbf{R}^s$, where usually $s = 2$. If a subscript of a function is a real variable, it denotes the partial derivative. For example, $\Theta_\xi \equiv \partial\Theta/\partial\xi$. For convenience we assume that $\varepsilon \geq 0$. Below we assume that hypotheses $\mathbf{H_0 1}$–$\mathbf{H_0 4}$ as well as $\mathbf{H_0 2^*}$ are satisfied.

2 The separatrix map: construction

The 2-dimensional torus $N(y^0) = \{(y, x, v, u, t) : y = y^0, v = u = 0\}$ is invariant with respect to the unperturbed system and is called hyperbolic [22, 39] (see also [11], where an invariant definition of a hyperbolic torus

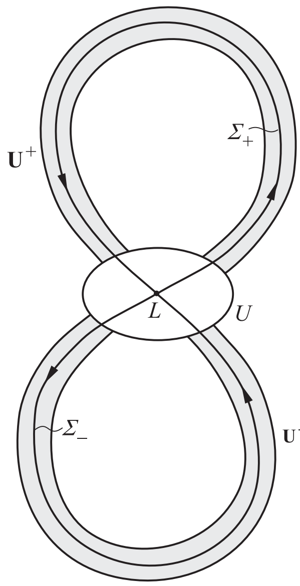

Figure 1:

is presented). There are two asymptotic surfaces

$$\widehat{\Gamma}^+(y^0), \widehat{\Gamma}^-(y^0) \subset \{(y,x,v,u,t) : y = y^0, H_0(y^0,v,u) = H_0(y^0,0,0)\},$$
$$\widehat{\Gamma}^\pm(y^0) = \{y^0\} \times \mathbf{T} \times \widehat{\gamma}^\pm(y^0) \times \mathbf{T}.$$

The surfaces $\widehat{\Gamma}^\pm(y^0)$ consist of unperturbed solutions which tend to $N(y^0)$ as $t \to \pm\infty$.

We consider the dynamics of the perturbed system in a neighborhood of the set

$$\widehat{\Gamma} = \cup_{y\in\overline{\mathcal{D}}}\big(\widehat{\Gamma}^+(y) \cup \widehat{\Gamma}^-(y)\big).$$

To study this dynamics, it is convenient to pass to the time-one map T_ε which assigns to any point (y,x,v,u) the point $(y(1),x(1),v(1),u(1))$, where $(y(t),x(t),v(t),u(t))$ is the solution of system (1.3) with the initial conditions $(y(0),x(0),v(0),u(0)) = (y,x,v,u)$.

The map T_0 has the 1-dimensional hyperbolic tori $L(y) = \mathrm{pr}(N(y))$, where pr is the projection $(y,x,v,u,t) \mapsto (y,x,v,u)$. We denote by $\Sigma^\pm(y)$ the asymptotic surfaces $\Sigma^\pm(y) = \mathrm{pr}\,(\widehat{\Gamma}^\pm(y))$.

Now we are going to define the separatrix map SPM_ε corresponding to T_ε in the vicinity of $\widehat{\Sigma}$,

$$\widehat{\Sigma} = \cup_{y\in\overline{\mathcal{D}}}\big(\widehat{\Sigma}^+(y) \cup \widehat{\Sigma}^-(y)\big).$$

Let U be a small neighborhood of $\cup_{y\in\overline{\mathcal{D}}}L(y)$ and $\mathbf{U}$ a neighborhood of $\widehat{\Sigma}$ (see Fig. 1). If $\mathbf{U}$ is sufficiently small, $\mathbf{U}\setminus U$ breaks into two connected components $\mathbf{U}^+$ and $\mathbf{U}^-$ such that $\Sigma^\pm \subset \mathbf{U}^\pm \cup U$.

Consider a point $z \in \mathbf{U}^+ \cup \mathbf{U}^-$. Let $k_1 = k_1(z)$ be the minimal natural number such that $T_\varepsilon^{k_1}(z) \notin \mathbf{U}^+ \cup \mathbf{U}^-$ and let $k_2 = k_2(z)$ be the minimal natural number such that $k_2 > k_1$ and $T_\varepsilon^{k_2}(z) \in \mathbf{U}^+ \cup \mathbf{U}^-$. So, for $k = k_1$ the trajectory $T_\varepsilon^k(z)$ leaves the domain $\mathbf{U}^+ \cup \mathbf{U}^-$. For $k = k_2$ the trajectory returns to $\mathbf{U}^+ \cup \mathbf{U}^-$. We call the point z good if $k_2 < \infty$ and $T_\varepsilon^{k_1}(z), \dots, T_\varepsilon^{k_2-1}(z) \in U$. Putting

$$\mathbf{U}_\varepsilon = \{z \in \mathbf{U}^+ \cup \mathbf{U}^- : z \text{ is good}\},$$

we obtain the maps

$$SPM_\varepsilon(\cdot, k_2(\cdot)+k) : \mathbf{U}_\varepsilon \to \mathbf{U}^+\cup\mathbf{U}^-, \quad SPM_\varepsilon(z, k_2(z)+k) = T_\varepsilon^{k_2(z)+k}(z),$$

where we assume that $T_\varepsilon^{k_2}(z), \dots, T_\varepsilon^{k_2(z)+k}(z) \in \mathbf{U}$.

Below we will put $t_+ = k_2(z) + k$.

In [34] explicit formulas for the map $SPM_\varepsilon(z, t_+)$ are obtained. Note that the construction of the separatrix map does not rely on any result of KAM type. In particular, some of the unperturbed tori $L(y)$, $N(y)$ are in general resonant. Actually, we construct the map SPM_ε as a perturbation of SPM_0.

In [35] we showed that the separatrix map can be combined with the method of anti-integrable limit [2, 24, 33] (see also [8]). The small parameter due to which the system is close to the anti-integrable limit is $\varepsilon^{-1}e^{-\lambda t_+}$. Here $\lambda = \lambda(y)$ is the positive characteristic number of the hyperbolic equilibrium of the system $(D, du\wedge dv, H_0(y, v, u))$, where y is regarded as a parameter. In fact, $\varepsilon^{-1}e^{-\lambda t_+}$ is not always small in the separatrix map. However it is small on the trajectories we construct.

3 The separatrix map: formulas

In this section we present formulas for the separatrix map (an explicit part + a small error). Formally speaking, below we use only definitions of the functions $\mathbf{H}, \lambda, \kappa, \Theta$, and the fact that $\varepsilon\rho$ is close to y i.e., the diffusion problem can be regarded as the problem of evolution of $\varepsilon\rho$. However, for understanding of dynamical meaning of constructions, presented in Section 4 we believe that it is reasonable at least to look through all this section.

The function $H_1(y, x, 0, 0, t)$ has the following Fourier expansion:

$$H_1(y, x, 0, 0, t) = \sum_{(k,k_0)\in\mathbf{Z}^2} H_1^{k,k_0}(y) e^{2\pi i(kx+k_0t)}.$$

Let $\phi : \mathbf{R}\to[0,1]$ be an even C^∞-smooth function such that $\phi(r) = 0$ for any $|r|\ge 1$, and $\phi(r) = 1$ for any $|r|\le 1/2$. We put

$$\overline{\mathbf{H}}(y, x, t) = \sum_{(k,k_0)\in\mathbf{Z}^2} \phi\Big(\frac{k\nu(y)+k_0}{\varepsilon^{1/4}}\Big) H_1^{k,k_0}(y) e^{2\pi i(kx+k_0t)}, \tag{3.1}$$

$$\mathbf{H}(y, x) = \overline{\mathbf{H}}(y, x, 0). \tag{3.2}$$

The function $\overline{\mathbf{H}}$ is a smoothed in y average of $H_1(y, x, 0, 0, t)$ along the unperturbed trajectories on the tori $N(y)$. It tends pointwise to the actual average

$$\sum_{k\nu(y)+k_0=0} H_1^{k,k_0}(y) e^{2\pi i(\langle k,x\rangle+k_0t)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T H_1(y, x+\nu(y)s, 0, 0, t+s)\, ds \tag{3.3}$$

as $\varepsilon\to 0$. We prefer to deal with $\mathbf{H}$ and $\overline{\mathbf{H}}$, because the function (3.3) is discontinuous at resonant y. Smoothness properties of $\overline{\mathbf{H}}, \mathbf{H}$ are anisotropic: the variables y are of a special sort in the sense that j-th derivative with respect to y is of order $\varepsilon^{-j/4}$.

Proposition 3.1 *Suppose that $H_1(y, x, 0, 0, t)\in C^j(\mathcal{D}_0\times\mathbf{T}^2, \mathbf{R})$ and*

$$\nu(y^0) = p/q + \delta, \quad p\in\mathbf{Z}, \quad q\in\mathbf{N}, \quad q < \varepsilon^{-1/4}/2.$$

Then for sufficiently small ε

$$|\mathbf{H}(y^0, x + p/q) - \mathbf{H}(y^0, x)| < C\delta^j q^j. \tag{3.4}$$

Corollary 3.1 *Consider the function $\widetilde{\mathbf{H}}(y^0, x) = \frac{1}{q}\sum_{l=1}^q \mathbf{H}(y^0, x + l/q)$. Then $\widetilde{\mathbf{H}}$ is $1/q$-periodic in x and*

$$|\mathbf{H}(y^0, x) - \widetilde{\mathbf{H}}(y^0, x)| < C\delta^j q^{j+1}.$$

Proof of Proposition 3.1 is contained in Section 14.

In [34]-[35] we construct canonical coordinates $(\rho, \zeta, r, \tau, \sigma)$. Their main properties are as follows.

(1) $dy \wedge dx + dv \wedge du = \varepsilon(d\rho \wedge d\zeta + dr \wedge d\tau)$.

(2) For some function $f(y, u, v, \varepsilon)$ such that $f(y, 0, 0, 0) = 0$

$$\varepsilon\rho = y + O(\varepsilon^{3/4}, H_0 - E(y)), \quad \zeta = x + f(y, u, v, \varepsilon),$$
$$\varepsilon r = H_0 - E(y) + O(\varepsilon^{3/4}, H_0 - E(y)),$$

where $H_0 = H_0(y, u, v)$.

(3) The variable $\tau \in (-c^{-1}, c^{-1})$ has physical meaning of time. Because of some technical reasons we have to separate the time variable in the system in, roughly speaking, the fractional part τ and the integer part $t \in \mathbf{N}$. Below t is of order $|\log \varepsilon|$

(4) We need also the discrete variable $\sigma \in \{-1, 1\}$ which denotes the loop of the separatrix (upper = 1, lower = −1), near which the point lies.

Let $(\rho_k, \zeta_k, r_k, \tau_k, t_k, \sigma_k)$, $k \in \mathbf{Z}$ be a trajectory of SPM_ε. In [35] we present formulas (an explicit part + small errors) which express $(\rho_{k+1}, \zeta_{k+1}, r_{k+1}, \tau_{k+1}, t_{k+1}, \sigma_{k+1})$ in terms of $(\rho_k, \zeta_k, r_k, \tau_k, t_k, \sigma_k)$. To construct trajectories, we need a certain special version of these equations: Hamiltonian in ρ, ζ and Lagrangian in τ. We eliminate the r variable from the equations. Then the dynamical system does not contain r, but becomes of second order in τ. To present formulas, we need another discrete variable

$$\vartheta \in \{-1, 1\}, \qquad \vartheta_{k+1} = \operatorname{sign}(r_{k+1} - \mathbf{H}(\varepsilon\rho_{k+1}, \zeta_k)).$$

Suppose that Hypotheses $\mathbf{H_0 1}$–$\mathbf{H_0 3}$ hold, $K_0 > 0$ is a constant independent of ε and $\overline{K}$ is a constant, satisfying the inequality $\overline{K} + K_0 < -\frac{1}{9}\log\varepsilon$.

Then there exist $C^{\mathbf{r}}$-smooth functions

$$\lambda, \kappa^\pm : \overline{\mathcal{D}} \to \mathbf{R}, \quad \Theta^\pm : \overline{\mathcal{D}} \times \mathbf{T}^2 \to \mathbf{R}$$

and coordinates $(\rho, \zeta, \tau, t, \sigma, \vartheta)$ such that for any trajectory $(\rho_k, \zeta_k, \tau_k, t_k, \sigma_k, \vartheta_k)$ of the separatrix map, where

$$\frac{1}{9}\log\varepsilon \le -\overline{K} - K_0 \le -\lambda t_{k+1} - \log\varepsilon \le -K_0, \tag{3.5}$$

the following equations hold:

$$\begin{aligned}
&\rho_{k+1} = \rho_k - \widehat{\Theta}^{\sigma_k}_\zeta(\varepsilon\rho_{k+1}, \zeta_k, \tau_k) + (\tau_{k+1} - \tau_k - t_{k+1})\mathbf{H}_\zeta(\varepsilon\rho_{k+1}, \zeta_k) + \mathcal{R}_1,\\
&\zeta_{k+1} = \zeta_k + \nu t_{k+1} - (\tau_{k+1} - \tau_k - t_{k+1})\mathbf{H}_\rho(\varepsilon\rho_{k+1}, \zeta_k) + \mathcal{R}_2,\\
&\widehat{\Theta}^{\sigma_k}_\tau(\varepsilon\rho_{k+1}, \zeta_k, \tau_k) = \tfrac{\lambda}{\varepsilon}\Big(\tfrac{\vartheta_k}{\kappa^{\sigma_{k-1}}} e^{\lambda(\tau_k - \tau_{k-1} - t_k)} - \tfrac{\vartheta_{k+1}}{\kappa^{\sigma_k}} e^{\lambda(\tau_{k+1} - \tau_k - t_{k+1})}\Big) + \mathcal{R}_3,\\
&\sigma_{k+1} = \sigma_k\vartheta_{k+1},
\end{aligned} \tag{3.6}$$

In these equations $\lambda, \nu, \kappa^\sigma$ are functions of $\varepsilon\rho_{k+1}$ and $\widehat{\Theta}(\varepsilon\rho, \zeta, \tau) = \Theta(\varepsilon\rho, \zeta - \nu(\varepsilon\rho)\tau, \tau)$. Exact statements on smallness of the error terms $\mathcal{R}_{1,2,3}$ are presented in [35]. We do not use them in this paper.

The functions $\lambda > 0$ and $\kappa^\pm > 0$ are determined by the unperturbed system, and $\Theta^\pm$ turn out to be certain integrals of the Poincaré-Melnikov type. Now we present explicit formulas for λ and $\Theta^\pm$. The proofs are presented in [34]–[35].

The function λ. According to $\mathbf{H_0 2}$, both eigenvalues of the matrix

$$\Lambda(y) = \begin{pmatrix} -(H_0)_{uv}(y, 0, 0), & -(H_0)_{uu}(y, 0, 0) \\ (H_0)_{vv}(y, 0, 0), & (H_0)_{vu}(y, 0, 0) \end{pmatrix}$$

are distinct and real. Their sum equals $\operatorname{tr}\Lambda = 0$. The function $\lambda = \lambda(y)$ is the positive eigenvalue of Λ.

The functions $\Theta^\pm$. Let $\gamma^\pm(y, \cdot) : \mathbf{R} \to \{(v, u) : H_0(y, v, u) = H_0(y, 0, 0)\}$ be natural parametrizations of the upper and lower separatrix loops $\widehat{\gamma}^\pm(y)$:

$$\dot\gamma^\pm(y, t) = \big(-(H_0)_u(y, \gamma^\pm(t)), (H_0)_v(y, \gamma^\pm(t))\big).$$

These parametrizations are defined up to a shift of the time: $t \mapsto t + t_0(y)$.

Definition 3.1 *We call natural parametrizations of $\widehat{\gamma}^+$ and $\widehat{\gamma}^-$ compartible if they depend smoothly on y and*

$$\lim_{t\to-\infty}\frac{\langle a_+(y),\gamma^+(y,t)\rangle}{\langle a_+(y),\gamma^-(y,t)\rangle}=-1.$$

Obviously, compartible parametrizations are defined uniquely up to a simultaneous rigid shift: if $\gamma^+(y,t^+(y,t))$, $\gamma^-(y,t^-(y,t))$ is another pair of compartible parametrizations then $t^+(y,t)=t^-(y,t)=t-t_0(y)$.

Any solution of the unperturbed system lying on $\widehat{\Gamma}^{\pm}(y)$ has the form

$$\begin{aligned}(y,x,v,u)(t) &= \Gamma^\sigma(y,\xi,\tau+t), \qquad \xi\in\mathbf{T},\ \tau\in\mathbf{R},\ \sigma\in\{+,-\},\\ \Gamma^\sigma(y,\xi,\tau) &= (y,\xi+\nu(y)\tau,\gamma^\sigma(y,\tau)).\end{aligned}$$

Let us put

$$H_*^\sigma(y,\xi,\tau,t)=H_1(\Gamma^\sigma(y,\xi,t),t-\tau)-H_1(y,\xi+\nu t,0,0,t-\tau).$$

Note that $H_*^\sigma(y,\xi,\tau,t)$ exponentially tends to zero as $t\to\pm\infty$.

Suppose that the parametrizations $\gamma^\pm$ are compartible. Then

$$\widehat{\Theta}^\sigma(y,\zeta,\tau)=\Theta^\sigma(y,\zeta-\nu(y)\tau,\tau),\quad \Theta^\sigma(y,\xi,\tau)=-\int_{-\infty}^{+\infty}H_*^\sigma(y,\xi,\tau,t)\,dt.$$

The functions Θ^σ are called the splitting potentials. They are obviously 1-periodic in ξ and τ. In the case $y=\mathrm{const}$ analogous functions were introduced in [30], see also [17] for the case of arbitrary dimension of x,y and Diophantine $\nu(y^0)$. Note that in [34] we present a more general construction, where $\mathbf{H_0 3}$ is not assumed.

Some calculations for Hamiltonian (1.5). Now we give explicit formulas for Θ^+ and $\mathbf{H}$, corresponding to the system, defined by (1.5). Since in this case $\nu(y)=y$, we get:

$$\begin{aligned}\mathbf{H}(y,x) &= a\Big(\phi(y/\varepsilon^{1/4})\cos(2\pi x)+\phi(1/\varepsilon^{1/4})\cos(2\pi t)\\ &\quad +b\phi\big((y-1)/\varepsilon^{1/4}\big)\cos(2\pi(x-t))\\ &\quad +b\phi\big((y+1)/\varepsilon^{1/4}\big)\cos(2\pi(x+t))\Big).\end{aligned}$$

On the unperturbed separatrix $\gamma^+=\{(v,u):v^2/2+\Omega^2\cos(2\pi u)=\Omega^2, v>0\}$ the natural parametrization is defined by the equation

$$\cos\big(2\pi i u(t)\big)=1+\frac{i}{\sinh(2\pi\Omega t)-i}-\frac{i}{\sinh(2\pi\Omega t)+i}.$$

The function $H_*^+(y,\xi,\tau,t)$ is as follows:

$$\begin{aligned}&a\big(\cos(2\pi u(t))-1\big)\Big(\cos(2\pi(\xi+yt))+\cos(2\pi(t-\tau))\\ &\qquad +b\cos(2\pi(\xi+\tau+yt-t))+b\cos(2\pi(\xi-\tau+yt+t))\Big).\end{aligned}$$

Direct calculation yields,

$$\begin{aligned}\Theta^+(y,\xi,\tau) &= \Big(\frac{y}{\sinh y}\cos(2\pi\xi)+\frac{1}{\sinh y}\cos(2\pi\tau)\\ &\quad +b\frac{y-1}{\sinh(y-1)}\cos(2\pi(\xi+\tau))+b\frac{y+1}{\sinh(y+1)}\cos(2\pi(\xi-\tau))\Big).\end{aligned}$$

Here we have used that $\Omega=\pi/2$ and $a=\pi^2/8$. The function $\frac{q}{\sinh q}$ in the above equation is assumed to be defined at $q=0$ by continuity.

4 Construction of a trajectory

In this section we present a method of the construction of trajectories for the separatrix map. The essence of the method is as follows (see the details below). Suppose that we have a finite piece of an "admissible" trajectory an the corresponding quasi-trajectory (a code). Suppose that the quasi-trajectory can be extended i.e., it is possible to attach to it two more points (one from the left and another from the right). The extension should be compatible in the sense of Definition 4.3. Then by Lemma 4.1 the extended quasi-trajectory can be regarded as a code for a larger trajectory of the separatrix map. Since the extension of a quasi-trajectory can be performed with a certain arbitrariness, it is possible to use Lemma 4.1 as a tool for the construction of diffusion trajectories.

Below we construct trajectories of the separatrix map with $\sigma \equiv 1$. Hence, ϑ also identically equals 1. Due to this we will skip the variables σ and ϑ, assuming that they always equal 1. Analogous trajectories generically exist with non-constant σ and ϑ.

Below together with variables ρ, ζ, τ we use variables η, ξ, τ defined by the map $\pi : \mathbf{R} \times \mathbf{T} \times \mathbf{R} \to \mathbf{R} \times \mathbf{T}^2$,

$$\pi(\rho, \zeta, \tau) = (\eta, \xi, \tau) = (\varepsilon\rho, \zeta - \nu(\varepsilon\rho)\tau, \tau),$$

The main motivation for this is that the potential $\widehat{\Theta}$ in the new variables turns to Θ. The last one is periodic in ξ and τ. This periodicity makes some arguments simpler.

We put

$$\overline{\nu} = \begin{pmatrix} \nu \\ -1 \end{pmatrix} \in \mathbf{R}^2, \quad \overline{\partial} = \nu\partial/\partial\xi - \partial/\partial\tau.$$

Hence, $\overline{\partial}$ is the differentiation on $\mathbf{T}^2 = \{\xi, \tau\}$ along the constant vector field $\overline{\nu}$. Consider the sets

$$J_0 = \{(\eta, \xi, \tau) \in \overline{\mathcal{D}} \times \mathbf{T}^2 : \overline{\partial}\Theta(\eta, \xi, \tau) = 0, \, \overline{\partial}^2\Theta(\eta, \xi, \tau) \neq 0\}.$$

Generically for any $\eta = \eta^0$ the set $J_0|_{\eta=\eta^0}$ contains a nonempty collection of curves on $\mathbf{T}^2$.

We put

$$\mathcal{J}_0 = \pi^{-1}(J_0) \cap \{(\rho, \zeta, \tau) : -1 < \tau < 1\} \subset \frac{1}{\varepsilon}\overline{\mathcal{D}} \times \mathbf{T} \times \mathbf{R}.$$

The condition $-1 < \tau < 1$ is introduced because below we need τ in $\mathcal{J}_0$ to be bounded. Here is another definition of the set $\mathcal{J}_0$:

$$\mathcal{J}_0 = \big\{(\rho, \zeta, \tau) \in \frac{1}{\varepsilon}\overline{\mathcal{D}} \times \mathbf{T} \times (-1, 1) : \widehat{\Theta}_\tau(\varepsilon\rho, \zeta, \tau) = 0, \widehat{\Theta}_{\tau\tau}(\varepsilon\rho, \zeta, \tau) \neq 0\big\}. \tag{4.1}$$

Consider the equation

$$\widehat{\Theta}_\tau(\varepsilon\rho, \zeta, \tau) = z, \qquad z \in \mathbf{R}. \tag{4.2}$$

It can be solved with respect to τ for small $|z|$ near any point $(\rho_0, \zeta_0, \tau_0) \in \mathcal{J}_0$. The solution is a smooth function $\Psi^{\rho_0,\zeta_0,\tau_0}(\varepsilon\rho, \zeta, z)$ with values on $(-2, 2)$. We put

$$\begin{aligned}
\mathcal{J}_{c',c''} &= \big\{(\rho_0, \zeta_0, \tau_0) \in \mathcal{J}_0 : \Psi = \Psi^{\rho_0,\zeta_0,\tau_0}(\varepsilon\rho, \zeta, z) \quad \text{is smooth for} \\
&\qquad \varepsilon^{3/4}|\rho - \rho_0| < c', \, |\zeta - \zeta_0| < c', \, |z| < c', \text{ where} \\
&\qquad |\Psi| < 2, \, |\Psi_\rho| < \varepsilon^{3/4}/c'', \, |\Psi_\zeta| < 1/c'', \, |\Psi_z| < 1/c''\big\}, \\
J_{c',c''} &= \big\{(\eta, \xi, \tau) \in J_0 : (\varepsilon^{-1}\eta, \xi + \nu(\eta)\tau, \tau) \in \mathcal{J}_{c',c''}\big\}.
\end{aligned}$$

Obviously,

$$\cup_{c'>0,c''>0}\mathcal{J}_{c',c''} = \mathcal{J}_0, \quad \cup_{c'>0,c''>0}J_{c',c''} = J_0.$$

Below we fix sufficiently small c', c''.

Our construction of trajectories for the separatrix map is based on a certain inductive procedure. Below we use the notation $\Omega = (\rho, \zeta, \tau, t)$. A finite sequence of vectors Ω is denoted by $\mathcal{O}$:

$$\mathcal{O} = (\Omega_{-k}, \dots, \Omega_k), \quad \Omega_j = (\rho_j, \zeta_j, \tau_j, t_j), \qquad -k \le j \le k.$$

where we always assume that Ω_{-k} does not contain the coordinate t i.e., $\Omega_{-k} = (\rho_{-k}, \zeta_{-k}, \tau_{-k})$. For briefness we admit by default that sub- and superscripts of $\mathcal{O}$ (respectively, of Ω) are automatically transmitted to the corresponding sequence of Ω_j (respectively, to the corresponding vector (ρ, ζ, τ, t)). For example, $\widehat{\mathcal{O}}^{(l)} = (\widehat{\Omega}^{(l)}_{-k}, \dots, \widehat{\Omega}^{(l)}_k)$, and $\Omega'_1 = (\rho'_1, \zeta'_1, \tau'_1, t'_1)$.

For some big constant C and a positive constant $b < \min\{1/3, c'/2\}$ we define

$$b_\rho = \frac{b^5}{60C^3} e^{K_0}, \quad b_\tau = \frac{b^4}{3C^2} e^{K_0}, \quad b_\zeta = \frac{b^5}{48C^3} e^{K_0} \varepsilon^{-3/4}.$$

Here we assume that K_0 is chosen (after C and b) so large that b_ρ and b_τ are large constants. Below it is just important that b_ρ, b_τ are large and b_ζ is of order $\varepsilon^{-3/4}$.

For any two sequences Ω', Ω'' we define

$$\operatorname{dist}(\Omega', \Omega'') = \begin{cases} +\infty & \text{if} \quad t' \ne t'', \\ \max\{b_\rho|\rho' - \rho''|,\, b_\zeta|\zeta' - \zeta''|,\, b_\tau|\tau' - \tau''|\} & \text{otherwise,} \end{cases}$$

Here $|\rho' - \rho''|$, $|\zeta' - \zeta''|$, and $|\tau' - \tau''|$ are taken with respect to the standard metric of $\mathbf{R}$, $\mathbf{T}$, and $\mathbf{R}$ respectively.

Definition 4.1 *We call a sequence* $\overline{\mathcal{O}} = (\overline{\Omega}_{-k}, \overline{\Omega}_{-k+1}, \dots, \overline{\Omega}_k)$, $k \ge 0$ *a quasi-trajectory if*

(i) $(\overline{\rho}_j, \overline{\zeta}_j, \overline{\tau}_j) \in \mathcal{J}_{c',c''}$, $\qquad -k \le j \le k$,
(ii) $\overline{t}_j \in \mathbf{N}$, $|\overline{\tau}_j| < 1$, $\qquad -k \le j \le k$,
(iii) $K_0 \le \lambda \overline{t}_j + \log \varepsilon \le K_0 + \overline{K}$, $\qquad -k < j \le k$.

In particular, $\overline{\mathcal{O}}$ is a quasi-trajectory for $k = 0$ provided

$$\overline{\mathcal{O}} = \Omega_0 = (\overline{\rho}_0, \overline{\zeta}_0, \overline{\tau}_0) \in \mathcal{J}_{c',c''}, \qquad |\overline{\tau}_0| < 1.$$

Definition 4.2 *We call a trajectory* $\mathcal{O}$ *admissible if a quasi-trajectory (a code)* $\overline{\mathcal{O}}$ *exists such that*

$$\operatorname{dist}(\Omega_j, \overline{\Omega}_j) < b(2 - b^{1+k-|j|}) \quad \text{for any integer } j \in [-k, k], \tag{4.3}$$
$$(\rho_{\pm k}, \zeta_{\pm k}, \tau_{\pm k}) \in \mathcal{J}_{c',c''}. \tag{4.4}$$

Since $2b < c'$, inclusions (4.4) can be replaced by the equations

$$\tau_{\pm k} = \Psi^{\overline{\rho}_{\pm k}, \overline{\zeta}_{\pm k}, \overline{\tau}_{\pm k}}(\varepsilon\rho_{\pm k}, \zeta_{\pm k}, 0).$$

Definition 4.3 *Let* $\mathcal{O}$ *be an admissible trajectory with the code* $\overline{\mathcal{O}}$. *We say that a quasi-trajectory* (Ω, Ω_+) *is compartible with* $\mathcal{O}$ *from the right if*

$$\Omega = \overline{\Omega}_k,$$
$$\big|\rho_+ - \rho_k + \widehat{\Theta}_\zeta(\varepsilon\rho_+, \zeta_k, \tau_k) - (\tau_+ - \tau_k - t_+)\mathbf{H}_\zeta(\varepsilon\rho_+, \zeta_k)\big| < \frac{b^2}{2b_\rho}, \tag{4.5}$$
$$\big|\zeta_+ - \zeta_k - \nu(\varepsilon\rho_+)t_+ + (\tau_+ - \tau_k - t_+)\mathbf{H}_\rho(\varepsilon\rho_+, \zeta_k)\big| < \frac{b^2}{2b_\zeta}.$$

A quasi-trajectory (Ω_-, Ω) *is compartible with* $\mathcal{O}$ *from the left if*

$$\rho = \overline{\rho}_{-k}, \quad \zeta = \overline{\zeta}_{-k}, \quad \tau = \overline{\tau}_{-k},$$
$$\big|\rho_{-k} - \rho_- + \widehat{\Theta}_\zeta(\varepsilon\rho_{-k}, \zeta_-, \tau_-) - (\tau_{-k} - \tau_- - t_{-k})\mathbf{H}_\zeta(\varepsilon\rho_{-k}, \zeta_-)\big| < \frac{b^2}{2b_\rho},$$
$$\big|\zeta_{-k} - \zeta_- - \nu(\varepsilon\rho_{-k})t_{-k} + (\tau_{-k} - \tau_- - t_{-k})\mathbf{H}_\rho(\varepsilon\rho_{-k}, \zeta_-)\big| < \frac{b^2}{2b_\zeta}.$$

Lemma 4.1 *(The Attachment Lemma.) Let $\mathcal{O}^0 = (\Omega^0_{-k}, \dots, \Omega^0_k)$ be an admissible trajectory with the code $\overline{\mathcal{O}}$, and let quasi-trajectories $(\Omega_-, \overline{\Omega}_{-k})$, $(\overline{\Omega}_k, \Omega_+)$ be compartible with $\mathcal{O}^0$ from the left and from the right respectively. Then there exists an admissible trajectory $\widehat{\mathcal{O}} = (\widehat{\Omega}_{-k-1}, \dots, \widehat{\Omega}_{k+1})$ with the code*

$$\overline{\Omega}^* = (\overline{\Omega}_{-k-1}, \overline{\Omega}_{-k}, \dots, \overline{\Omega}_k, \overline{\Omega}_{k+1}), \qquad \overline{\Omega}_{-k-1} = \Omega_-, \quad \overline{\Omega}_{k+1} = \Omega_+.$$

This lemma is the main result of [35] and the main technical tool in the present paper. Indeed, we can start from some admissible trajectory with $k = 0$. If there exist both left and right compartible quasi-trajectories, by using Lemma 4.1, we can extend the admissible trajectory $\mathcal{O}^0$, extend again, and so on. Hence it is possible to obtain a large class of trajectories, following the codes we prescribe.

Unfortunately, unlike the usual situation in symbolic dynamics we are not able to choose an infinite code in advance. Indeed, according to Definition 4.3, compartible trajectories are defined in terms of both the trajectory $\mathcal{O}$ and the corresponding code $\overline{\mathcal{O}}$. We are not able to define compartible trajectories only in terms of the code $\overline{\mathcal{O}}$, because the trajectories we construct are only partially hyperbolic. At least we do not use their total hyperbolicity even if it really exists.

5 Clear and vague sets

In this section we present hypotheses $\mathbf{H_1 1}$–$\mathbf{H_1 3}$ and divide $\mathcal{D}$ into 3 sets $\mathcal{D}_N, \mathcal{D}_{CR}$, and $\mathcal{D}_V$, which will be studied separately. Consider the sets

$$j_\pm(\eta) = \{(\xi, \tau) \in \mathbf{T}^2 : (\eta, \xi, \tau) \in J_0, \ \pm\Theta_\xi(\eta, \xi, \tau) > 0\}. \tag{5.1}$$

Let $g^{\eta,t} : \mathbf{T}^2 \to \mathbf{T}^2$ be the phase flow of the vector field $\overline{\nu}$:

$$g^{\eta,t}(\xi, \tau) = (\xi, \tau) + \overline{\nu}(\eta)t = (\xi + \nu(\eta)t, \tau - t).$$

Consider the map $\mathrm{pr}_{\overline{\nu}} : \mathbf{T}^2 \to \mathbf{T}_0 = \{(\xi, \tau) \in \mathbf{T}^2 : \tau = 0\}$,

$$\mathrm{pr}_{\overline{\nu}}(\xi, \tau) = g^{\eta,\{\tau\}}(\xi, \tau) \equiv (\xi + \nu(\eta)\{\tau\}, 0),$$

where $\{\tau\} \in [0, 1)$ is the fractional part of the real number τ. This map is the "projection" on the circle $\mathbf{T}_0$ along the vector field $\overline{\nu}(\eta)$. This projection is discontinuous on $\mathbf{T}_0$.

For $0 \le \alpha < 1/2$ and a set $S \subset \mathbf{T}$ we put

$$S - \alpha = \{\xi \in S : (\xi - \alpha, \xi + \alpha) \subset S\}.$$

$\mathbf{H_1 1}$. *There exists $c_* > 0$ independent of $\eta \in \overline{\mathcal{D}}$ such that the sets $\mathrm{pr}_{\overline{\nu}}\, j_\pm(\eta) - 2c_*$ are not empty.*

Proposition 5.1 *The set of C^j-smooth functions $\Theta(\eta, \xi, \tau)$, $2 \le j \le \omega$, satisfying $\mathbf{H_1 1}$, is open and dense in C^j topology.*

Proof of Proposition 5.1 is contained in Section 13.

Definition 5.1 *Suppose that $\mathbf{H_1 1}$ holds and c_* is the corresponding constant. We call $\eta \in \overline{\mathcal{D}}$ K-clear if*

$$\bigcup_{0 \le t \le K-1} g^{\eta,t}(pr_{\overline{\nu}} j_\pm - c_*) = \mathbf{T}^2.$$

Otherwise η is called K-vague.

Proposition 5.2 *Suppose that $\eta \in \overline{\mathcal{D}}$ is K-vague and $I \subset pr_{\overline{\nu}} j_\pm(\eta) - c_*$ is an interval of length $|I|$. Then there exist $p \in \mathbf{Z}$, $q \in \mathbf{N}$, such that*

$$q \le K - 1, \quad q < \frac{2}{|I|}, \quad \text{and} \quad \Big|\nu(\eta) - \frac{p}{q}\Big| < \frac{1}{qK}. \tag{5.2}$$

Corollary 5.1 *If* $\mathbf{H_1 1}$ *holds then* $q < 1/c_*$ *in (5.2). Hence, according to Proposition 5.2, vague values of* η *are gathered into intervals around resonances of low orders. The number of these intervals is finite and independent of* ε*.*

Proof of Proposition 5.2 is contained in Section 14.

Below

$$K = \overline{K}/\lambda - 4, \quad \overline{K} = |\log \varepsilon|/10. \tag{5.3}$$

Definition 5.2 *We call* $\eta \in \overline{\mathcal{D}}$ *essential if*

$$\frac{|\log \varepsilon|}{\lambda} \max_{\zeta \in \mathbf{T}} |\mathbf{H}_\zeta(\eta, \zeta)| > b.$$

Essential points $\eta \in \overline{\mathcal{D}}$ form intervals, on which $\nu(\eta)$ are close to resonances of not very large orders. More precisely, the following statement holds.

Proposition 5.3 *Let* $\eta \in \overline{\mathcal{D}}$ *be essential and let* H_1 *be* C^j*-smooth,* $j \geq 2$*. Then there exist* $k, k_0 \in \mathbf{Z}$*,* $|k| < c^{-1}(|\log \varepsilon|/b)^{1/(j-1)}$ *such that*

$$|k\nu(\eta) + k_0| \leq \varepsilon^{1/4}. \tag{5.4}$$

Proof of Proposition 5.3 follows from the estimate $|H_1^{l,l_0}| < c_1^{-1}(|l| + |l_0|)^{-j}$ for the Fourier coefficients of H_1 and from the definition of $\mathbf{H}$.

Let $\mathcal{E}$ be the set of essential points and let $\mathcal{D}_C$ be the set of K-clear points of $\overline{\mathcal{D}}$ with K, satisfying (5.3). We call the sets

$$\mathcal{D}_N = \mathcal{D}_C \setminus \mathcal{E}, \quad \mathcal{D}_{CR} = \mathcal{D}_C \cap \mathcal{E}, \quad \mathcal{D}_V = \overline{\mathcal{D}} \setminus \mathcal{D}_C$$

the nonresonant set, the clear resonant set, and the vague set respectively. Then $\overline{\mathcal{D}} = \mathcal{D}_N \cup \mathcal{D}_{CR} \cup \mathcal{D}_V$. The sets $\mathcal{D}_N$, $\mathcal{D}_{CR}$, and $\mathcal{D}_V$ are finite collections of intervals.

Consider a resonance $\eta = \eta^0$, generating a vague interval $\mathbf{I} \subset \mathcal{D}_V$ (see Corollary 5.1). Denote $\nu_0 = \nu(\eta^0) = p_0/q_0$, where $p_0 \in \mathbf{Z}$, $q_0 \in \mathbf{N}$, $q_0 < 2/c_*$, and the fraction p_0/q_0 is irreducible. Consider the two functions:

$$\begin{aligned} F_+(\eta^0, \zeta) &= \max_{s \in [-q_0, 0]} \{\Theta_\xi(\eta^0, \zeta - \nu_0 s, s) : (\eta^0, \zeta - \nu_0 s, s) \in \overline{J}_0\}, \\ F_-(\eta^0, \zeta) &= \min_{s \in [-q_0, 0]} \{\Theta_\xi(\eta^0, \zeta - \nu_0 s, s) : (\eta^0, \zeta - \nu_0 s, s) \in \overline{J}_0\}. \end{aligned}$$

Obviously, if F_+ is defined at a point ζ then F_- is defined at ζ (and vice versa) and moreover, $F_+(\eta^0, \zeta) \geq F_-(\eta^0, \zeta)$. Generically $F_\pm$ are defined at any ζ (see Proposition 5.4 below).

The function $\Theta(\eta^0, \zeta - \nu_0 s, s)$ is q_0-periodic in s. Hence, $F_\pm$ can be defined as maximum and minimum with respect to $s \in \mathbf{R}$. The functions $F_\pm(\eta^0, \zeta)$ are $1/q_0$-periodic in ζ. We define the average:

$$\langle F_\pm \rangle(\eta^0) = q_0 \int_0^{1/q_0} F_\pm(\eta^0, \zeta)\, d\zeta.$$

Definition 5.3 *We call a function* $f : \mathbf{T} \to \mathbf{R}$ *piecewise smooth if there exist a finite number of points*[2] $\varphi_1, \ldots, \varphi_k \in \mathbf{T}$ *and continuous functions* $f_j : [\varphi_j, \varphi_{j+1}] \to \mathbf{R}$ *such that*

(1) f_j *are smooth on the open intervals* $(\varphi_j, \varphi_{j+1}) \subset \mathbf{T}$*,*

(2) $f|_{(\varphi_j, \varphi_{j+1})} = f_j$*.*

Note that according to this definition a piecewise smooth function can be discontinuous.

[2] We assume that $\varphi_1, \ldots, \varphi_k$ are well-ordered on the circle i.e., moving from φ_j to φ_{j+1} in the positive direction, we meet no other points φ_s (here $\varphi_{k+1} = \varphi_1$).

Proposition 5.4 *For C^j-generic Θ*

(1) $\pm\langle F_\pm\rangle > 0$.

(2) $F_+ = F_-$ *at most at a finite number of points* $\zeta_1, \dots, \zeta_k \in \mathbf{T}$, *where for sufficiently small* $|\delta|$ *the following transversality condition holds:*

$$F_+(\eta^0, \zeta_j + \delta) - F_-(\eta^0, \zeta_j + \delta) > 2c_F|\delta|, \qquad c_F > 0, \quad j = 1, \dots, k.$$

(3) $F_\pm(\eta^0, \zeta) = \Theta_\xi(\eta^0, \zeta - \nu_0\gamma_\pm(\zeta), \gamma_\pm(\zeta))$, *where for some* $c'_F > 0$

$$\overline{\partial}\Theta(\eta^0, \zeta - \nu_0\gamma_\pm(\zeta), \gamma_\pm(\zeta)) = 0, \quad |\overline{\partial}^2\Theta(\eta^0, \zeta - \nu_0\gamma_\pm(\zeta), \gamma_\pm(\zeta))| \geq 2c'_F. \tag{5.5}$$

(4) The functions $\gamma_\pm$ *and* $F_\pm$ *are piecewise* C^{j-1}*-smooth in* ζ.

The genericity is understood here in the sense that the set of Θ, satisfying (1)–(4), contains a subset which is open and dense in the C^j topology.

Proof of Proposition 5.4 is contained in Section 13.

$\mathbf{H_1 2}$. *For any* η^0, *generating a vague interval, the function* $\Theta(\eta^0, \zeta, \tau)$ *is generic in the sense of Proposition 5.4.*

$\mathbf{H_1 3}$. *For any* η^0, *generating a vague interval the function* $\mathbf{H}(\eta^0, \zeta)$, $\zeta = \zeta \bmod \frac{1}{q_0}$ *takes global maximum at a unique point* $\zeta_* = \zeta_*(\eta^0)$ *and this maximum is nondegenerate.*

Hypotheses $\mathbf{H_1 1}$–$\mathbf{H_1 3}$ and system (1.5). Now we outline the proof of Theorem 2. We need to check that the functions $\mathbf{H}$ and $\Theta = \Theta^+$, calculated in the end of Section 3, satisfy Hypotheses $\mathbf{H_1 1}$–$\mathbf{H_1 3}$. First, we put $b = 0$. Since $\overline{\partial} = \eta\partial/\partial\xi - \partial/\partial\tau$, we have:

$$\begin{aligned}
\overline{\partial}\Theta(\eta, \xi, \tau) &= -2\pi\Big(\frac{\eta^2}{\sinh\eta}\sin(2\pi\xi) - \frac{1}{\sinh 1}\sin(2\pi\tau)\Big), \\
\overline{\partial}^2\Theta(\eta, \xi, \tau) &= -4\pi^2\Big(\frac{\eta^3}{\sinh\eta}\cos(2\pi\xi) + \frac{1}{\sinh 1}\cos(2\pi\tau)\Big).
\end{aligned}$$

For any η^0 the set $J_0 \cap \{\eta = \eta^0\}$ consists of curves, determined by the equation $\overline{\partial}\Theta = 0$ without a finite set, where $\overline{\partial}^2\Theta = 0$. There are only two exceptions: $\eta^0 = \pm 1$. For example, for $\eta^0 = 1$ we have to remove the curve $\{\xi + \tau = 1/2\}$, where both equations hold. However, another curve $\{\xi = \tau\}$ without two points $\xi = \tau = \pm 1/4$ still lies in $J_0 \cap \{\eta = 1\}$. For $\eta = -1$ the situation is analogous.

For $|\eta^2/\sinh\eta| \neq 1/\sinh 1$ the set $\{(\xi, \tau) : \overline{\partial}\Theta(\eta, \xi, \tau) = 0\}$ consists of two curves, passing through four points $(\xi, \tau) \in \mathbf{Z}^2/2 \bmod 1$ (two points on each curve). Both curves are homotopic to either $\{\tau = 0\}$ or $\{\xi = 0\}$ depending on what inequality $|\eta^2/\sinh\eta| > 1/\sinh 1$ or $|\eta^2/\sinh\eta| < 1/\sinh 1$ holds.

Since $\Theta_\xi = -\frac{2\pi\eta}{\sinh\eta}\sin(2\pi\xi)$, we see that

$$j_\pm(\eta^0) = J_0 \cap \{\eta = \eta^0\} \cap \{\mp\sin(2\pi\xi) > 0\}.$$

After these observations Hypothesis $\mathbf{H_1 1}$ can be easily checked.

Vague intervals are neighborhoods of integer η. Hence, we need to check $\mathbf{H_2 2}$ only for $\eta = \pm 1$ and $\eta = 0$. For $\eta = \pm 1$ the corresponding functions $F_\pm$ satisfy Proposition 5.4. In the case $\eta^0 = 0$ we have:

$$F_+(0, \zeta) = F_-(0, \zeta) = \sin\zeta$$

i.e., a degeneracy takes place. This degeneracy disappears for small b.

Hypothesis $\mathbf{H_1 3}$ at $\eta^0 = \pm 1, 0$ is checked easily for $b \neq 0$.

6 Passage through a nonresonant interval

We denote $\lambda_{\min} = \min_{y \in \overline{\mathcal{D}}} \lambda(y)$. According to $\mathbf{H_0 2}$, $\lambda_{\min} > 0$.

Consider a connected component (a_1, a_2) of $\mathcal{D}_N$ and an admissible trajectory $\mathcal{O} = (\Omega_{-k}, \dots, \Omega_k)$ which enters the interval i.e., $a_1 \leq \varepsilon\rho_k \leq a_1 + O(\varepsilon|\log\varepsilon|)$.

Lemma 6.1 *Suppose that* $\mathbf{H_1 1}$ *holds. Then it is possible to continue the trajectory to an admissible trajectory which leaves* (a_1, a_2) *through another end:* $a_2 \le \varepsilon\rho_{k'} \le a_2 + O(\varepsilon|\log\varepsilon|)$ *for some* $k' > k$. *Moreover,*

$$\sum_{j=k}^{k'} t_j < \frac{(a_2 - a_1)|\log\varepsilon|}{2b\varepsilon\lambda_{\min}}. \tag{6.1}$$

Corollary 6.1 *The average velocity of the transition through* (a_1, a_2)

$$\frac{|\eta_2 - \eta_1|}{\sum_{j=k}^{k'} t_j} = \varepsilon \frac{|\rho_2 - \rho_1|}{\sum_{j=k}^{k'} t_j}$$

is of order $\varepsilon/|\log\varepsilon|$.

Proof of Lemma 6.1. Let $\overline{\mathcal{O}} = (\overline{\Omega}_{-k}, \dots, \overline{\Omega}_k)$ be the code of the trajectory $\mathcal{O}$. We discuss the continuation of the trajectory forward with respect to time. (The continuation backward is not essential). We will use the following

Lemma 6.2 [3] *Suppose that* $\mathbf{H_1 1}$ *holds and* $\varepsilon\rho_k$ *is* K*-clear with* K, *satisfying (5.3). Then there exists a right compartible quasi-trajectory* $(\overline{\Omega}_k, \overline{\Omega}_{k+1})$ *and another right compartible quasi-trajectory* $(\overline{\Omega}'_k, \overline{\Omega}'_{k+1})$ *with*

$$\widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_{k+1}, \overline{\zeta}_{k+1}, \overline{\tau}_{k+1}) < -7b, \quad \widehat{\Theta}_\zeta(\varepsilon\overline{\rho}'_{k+1}, \overline{\zeta}'_{k+1}, \overline{\tau}'_{k+1}) > 7b. \tag{6.2}$$

Now Lemma 6.1 follows from the Attachment Lemma. Indeed, according to the Attachment lemma, the new trajectory $\widehat{\mathcal{O}} = (\Omega_{-k-1}, \dots, \Omega_{k+1})$ will be such that $\mathrm{dist}(\Omega_{k+1}, \overline{\Omega}_{k+1}) < 2b$. In particular, $b_\rho|\rho_{k+1} - \overline{\rho}_{k+1}| < 2b$. This inequality preserves when we extend the trajectory by using the Attachment lemma repeatedly. We deal with a nonresonant interval. Therefore, $|(\overline{\tau}_{k+1} - \tau - \overline{t}_{k+1})\mathbf{H}_\zeta(\varepsilon\overline{\rho}_{k+1}, \zeta)| < b$. This implies (see (3.6)) that $\overline{\rho}_{k+1} - \overline{\rho}_k > 5b$. Hence, on all extensions of the trajectory $\mathcal{O}$

$$|\rho_{k+1} - \rho_k| \ge -|\rho_{k+1} - \overline{\rho}_{k+1}| + \overline{\rho}_{k+1} - \overline{\rho}_k - |\overline{\rho}_k - \rho_k| > 5b - 4b/b_\rho > 4b$$

which means that on the trajectory the variable ρ increases at least by $4b$ on each step. Since each step takes time less than $2|\log\varepsilon|/\lambda_{\min}$, we obtain estimate (6.1).

6.1 Proof of Lemma 6.2

We define

$$\widetilde{j}_\pm(\eta, c, c', c'') = \{(\xi, \tau) \in \mathbf{T}^2 : (\eta, \xi, \tau) \in J_{c',c''},\ \pm\Theta_\xi(\eta, \xi, \tau) \ge c\}. \tag{6.3}$$

Since $\cup_{c,c',c''>0}\widetilde{j}_\pm(\eta, c, c', c'') = j_\pm(\eta)$, for any $\eta \in \overline{\mathcal{D}}$ there exists $\widetilde{c} > 0$ such that

$$\mathrm{pr}_{\overline{\mathcal{D}}}\widetilde{j}_\pm(\eta, \widetilde{c}, \widetilde{c}, \widetilde{c}) \subset \mathrm{pr}_{\overline{\mathcal{D}}} j_\pm - c_*.$$

Since $\overline{\mathcal{D}}$ is a compact, positive $\widetilde{c}$ can be chosen independent of $\eta \in \overline{\mathcal{D}}$.

Hence for any K-clear η we have:

$$\bigcup_{0 \le t \le K+1} g^{\eta,t}\Big(\mathrm{pr}_{\overline{\mathcal{D}}}\widetilde{j}_\pm(\eta, 8b, 2c', 2c'')\Big) = \mathbf{T}^2 \tag{6.4}$$

with small enough b, c', c''.

Now we construct a right compartible quasi-treajectory $(\overline{\Omega}_k, \overline{\Omega}_{k+1})$.

Below for briefness we put $\Omega_+ = \overline{\Omega}_{k+1}$ and deal with the first inequality (6.2). According to the definition of a right compartible trajectory, we need to define $\rho_+, \zeta_+, \tau_+, t_+$ so that

$$\big|\rho_+ - \rho_k + \widehat{\Theta}_\zeta(\varepsilon\rho_+, \zeta_k, \tau_k) - (\tau_+ - \tau_k - t_+)\mathbf{H}_\zeta(\varepsilon\rho_+, \zeta_k)\big| < \frac{b^2}{2b_\rho}, \tag{6.5}$$

$$\big|\zeta_+ - \zeta_k - \nu(\varepsilon\rho_+)t_+ + (\tau_+ - \tau_k - t_+)\mathbf{H}_\rho(\varepsilon\rho_+, \zeta_k)\big| < \frac{b^2}{2b_\zeta}, \tag{6.6}$$

$$(\rho_+, \zeta_+, \tau_+) \in \mathcal{J}_{c',c''}, \tag{6.7}$$

$$K_0 \le \lambda t_+ + \log\varepsilon \le K_0 + \overline{K}, \qquad |\tau_+| < 1,\ t_+ \in \mathbf{N}. \tag{6.8}$$

[3] A multi-dimensional analog of this lemma is presented in [35], Section 5.

Moreover, because of the condition $\widehat{\Theta}_\zeta(\varepsilon\rho_+, \zeta_+, \tau_+) < -7b$, instead of (6.7) we need the following stronger condition:

$$(\eta_+, \xi_+, \tau_+) \equiv (\varepsilon\rho_+, \zeta_+ - \nu(\eta_+)\tau_+, \tau_+) \in \tilde{j}_-(\eta_+, 7b, c', c''). \qquad (6.9)$$

Condition (6.8) means that t_+ can be chosen from a certain interval of length $\overline{K}/\lambda$.

(1) First, we find t_+ and some auxiliary τ_*, ζ_*. Consider the condition

$$(\varepsilon\rho_k, \zeta_k + \nu(\varepsilon\rho_k)t, -t) \in \tilde{j}_-(\eta_k, 8b, 2c', 2c''). \qquad (6.10)$$

According to (6.4) and (5.3), for any point $(\xi_0, \tau_0) \in \mathbf{T}^2$ any interval $I \subset \mathbf{R}$ of length greater than or equal to $\overline{K}/\lambda - 1$ contains a point t with $(\eta_k, \xi_0 + \nu(\varepsilon\rho_k)t, \tau_0 - t)$ lying in $\tilde{j}_-(\eta_k, 8b, 2c', 2c'')$. Hence, (6.10) has a solution

$$t = \hat{t}, \quad K_0 + \lambda/2 \le \lambda\hat{t} + \log\varepsilon \le K_0 + \overline{K} - \lambda/2.$$

We define t_+, τ_*, ζ_* as follows:

$$t_+ - \tau_* = \hat{t}, \quad t_+ \in \mathbf{Z}, \quad \tau_* \in (-1/2, 1/2], \quad \zeta_* = \zeta_k + \nu(\varepsilon\rho_k)t_+. \quad (6.11)$$

Then (6.8) holds. Inclusion (6.10) implies that

$$(\varepsilon\rho_k, \zeta_* - \nu(\varepsilon\rho_k)\tau_*, \tau_*) \in \tilde{j}_\pm(\eta_k, 8b, 2c', 2c'').$$

(2) Now we define ρ_+, ζ_+:

$$\begin{aligned} \rho_+ &= \rho_k - \widehat{\Theta}_\zeta(\varepsilon\rho_k, \zeta_k, \tau_k) + (\tau_* - \tau_k - t_+)\mathbf{H}_\zeta(\varepsilon\rho_k, \zeta_k), \\ \zeta_+ &= \zeta_k + \nu(\varepsilon\rho_k)t_+ - (\tau_* - \tau_k - t_+)\mathbf{H}_\rho(\varepsilon\rho_k, \zeta_k). \end{aligned} \qquad (6.12)$$

(3) We find τ_+ from the equation $\overline{\partial}\Theta(\varepsilon\rho_+, \zeta_+ - \nu(\varepsilon\rho_+)\tau_+, \tau_+) = 0$ which is equivalent to $\widehat{\Theta}_\tau(\varepsilon\rho_+, \zeta_+, \tau_+) = 0$. Hence,

$$\tau_+ = \Psi^{\rho_k, \zeta_*, \tau_*}(\varepsilon\rho_+, \zeta_+, 0).$$

Since $\tau_* = \Psi^{\rho_k, \zeta_*, \tau_*}(\varepsilon\rho_k, \zeta_*, 0)$, $|\rho_+ - \rho_k| = O(\log\varepsilon)$, and $|\zeta_+ - \zeta_*| = O(\varepsilon^{3/4}\log\varepsilon)$ we have:

$$|\tau_+ - \tau_*| = O(\varepsilon^{3/4}|\log\varepsilon|).$$

Therefore, (6.9) holds, (6.5)–(6.6) can be checked easily. ∎

7 Passage through a clear essential resonance

Let $\mathbf{I}$, be a connected component of $\mathcal{D}_{CR}$. In this section we discuss passage of the trajectory through $\mathbf{I}$. According to Proposition 5.3, $\mathbf{I}$ corresponds to some "near-resonance" (5.4). We will assume that there exists $\eta^0 \in \mathbf{I}$ such that $k\nu_0 + k_0 = 0$, $\nu_0 = \nu(\eta^0)$. The case when $\nu(\eta)|_{\mathbf{I}} \ne \nu_0$ is simpler and can be considered analogously. For brevity we put $\eta^0 = 0$. Then $\nu_0 = \nu(0)$.

Lemma 7.1 *Suppose that* $\mathbf{H_1 1}$ *holds. Then the trajectory can be continued through* $\mathbf{I}$. *This takes time* $t = O(|I|\,|\log\varepsilon|/\varepsilon)$.

Proof. Our construction of right compartible trajectories is analogous to the one presented in Section 6. Unlike the situation with the passage through a nonresonant interval the term $(\tau_+ - \tau - t_+)\mathbf{H}_\zeta(\varepsilon\rho_+, \zeta)$ can prevent the variable ρ from increasing if $\mathbf{H}_\zeta(\varepsilon\rho_+, \zeta)$ is positive. So, it is important to look at the evolution of the variable ζ.

We denote

$$\nu_0 = \nu(0) = -k_0/k, \quad \nu_0' = \frac{d\nu}{d\eta}(0), \quad \nu_0'' = \frac{d^2\nu}{d\eta^2}(0),$$

where by $\mathbf{H_04}$, $|\nu_0'| + |\nu_0''| > 2c_\nu$. We put

$$c_\Theta = \sup_{\eta\in\overline{\mathcal{D}},(\xi,\tau)\in\mathbf{T}^2} \Theta_\xi(\eta,\xi,\tau).$$

Now we define one dynamical problem with control and noice: Problem **C** (**C** is from "clear").

Problem C. *Consider the discrete dynamical system*

$$\mu_{n+1} - \mu_n = \sqrt{\varepsilon}\big(u_n - t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n)\big), \tag{7.1}$$

$$\zeta_{n+1} - \zeta_n = \nu(\sqrt{\varepsilon}\mu_n)t_{n+1} + \beta_n. \tag{7.2}$$

Here for any n

$$K_0 < \lambda t_n + \log\varepsilon < K_0 + \overline{K}, \quad |\beta_n| < c^{-1}\varepsilon^{3/4}|\log\varepsilon|. \tag{7.3}$$

The variable u_n is a control, which according to our needs can be chosen from one of the intervals:

$$(-c_\Theta - b, -5b) \quad or \quad (5b, c_\Theta + b) \tag{7.4}$$

Values of u_n from these intervals and values of t_n,β_n, satisfying (7.3), are arbitrary (i.e., their precise values are unknown). This arbitrariness can be regarded as a noice. The goal is to construct a trajectory which starts at a point (μ_0,ζ_0), and finishes at a point (μ_N,ζ_N), where

$$\widehat{\mathbf{I}} = \{\mu\in\mathbf{R} : \sqrt{\varepsilon}\mu\in\mathbf{I}\} \subset [\mu_0,\mu_N].$$

Note that $\widehat{\mathbf{I}}$ is a connected component of the set $\{\mu\in\mathbf{R} : |k\nu(\sqrt{\varepsilon}\mu) + k_0| \le \varepsilon^{1/4}\}$.

Proposition 7.1 *Problem **C** has a solution with $N = O(|\mathbf{I}|/\varepsilon)$.*

Proposition 7.2 *Suppose that Problem **C** has a solution. Then Lemma 7.1 holds with $t = O(N\log\varepsilon)$.*

7.1 Proof of Proposition 7.1

We begin with some auxiliary estimates. For some large constant C_1 consider the set

$$\widehat{\mathbf{I}}^{cr} = \big\{\mu\in\widehat{\mathbf{I}} : |\nu(\sqrt{\varepsilon}\mu) - \nu_0| \le C_1\varepsilon^{3/4}\big\}.$$

Let $|I|$ be the length of an interval $I\subset\mathbf{R}$.

Proposition 7.3 *Suppose that $|\nu_0'| + |\nu_0''| \ge 2c_\nu$. Then*

(a) $|\widehat{\mathbf{I}}| = O(\varepsilon^{-3/8})$,

(b) $\widehat{\mathbf{I}}^{cr}$ consists of one or two intervals $\widehat{\mathbf{I}}_j^{cr}$, $j = 1,2$ and $|\widehat{\mathbf{I}}_j^{cr}| = O(\varepsilon^{-1/8})$.

Proof of Proposition 7.3. The following statement is crusial for the proof.

Claim. *Let u,v,c_u,c_v,s be constants such that c_u,c_v,s are positive, $|u|/c_u + |v|/c_v \ge 2$, and let $I\subset\mathbf{R}$ be an interval. Suppose that*

$$|ux + vx^2| < s \quad \textit{for any } x\in I.$$

Then the length of the interval I does not exceed $\max\{4s/c_u, \sqrt{4s/c_v}\}$.

We omit the straightforward proof of the claim. Now to prove Proposition 7.3, it is sufficient to note that

$$\widehat{\mathbf{I}} \subset \big\{\mu\in\mathbf{R} : \big|k(\nu_0'\mu + \frac{1}{2}\nu_0''\sqrt{\varepsilon}\mu^2 + O(\varepsilon\mu^3))\big| \le 2\varepsilon^{-1/4}\big\},$$

$$\widehat{\mathbf{I}}^{cr} \subset \big\{\mu\in\mathbf{R} : \big|\nu_0'\mu + \frac{1}{2}\nu_0''\sqrt{\varepsilon}\mu^2 + O(\varepsilon\mu^3)\big| \le C_1\varepsilon^{1/4}\big\}.$$

∎

Consider the system

$$\begin{cases} \dot{\mu} = v - \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu, \zeta), \\ \dot{\zeta} = \frac{1}{\sqrt{\varepsilon}}(\nu(\sqrt{\varepsilon}\mu) - \nu_0) + \sqrt{\varepsilon}\mathbf{H}_\eta(\sqrt{\varepsilon}\mu, \zeta), \end{cases} \qquad v = \frac{2b\lambda}{|\log \varepsilon|}. \tag{7.5}$$

The constant v is small and positive. This system is Hamiltonian with symplectic structure $d\mu \wedge d\zeta$ and Hamiltonian function

$$\chi(\mu, \zeta) = \frac{1}{\varepsilon}\big(E(\sqrt{\varepsilon}\mu) - E(0) - \nu_0\sqrt{\varepsilon}\mu\big) + \mathbf{H}(\sqrt{\varepsilon}\mu, \zeta) - v\zeta. \tag{7.6}$$

We use this system to compare its trajectories with trajectories of Problem **C**.

Hamiltonian (7.6) is multi-valued on the cylinder

$$Z = \{(\mu, \zeta) : \mu \in \widehat{\mathbf{I}}, \zeta \bmod 1\}.$$

However, the corresponding Hamiltonian vector field is 1-periodic in ζ and therefore, single-valued on Z. To make χ single-valued, we assume that $\chi : Z \to \mathbf{R}/(v\mathbf{Z})$. In other words, we take $\chi \bmod v$. Note also that due to Proposition 3.1 the 1-periodic in ζ function $\mathbf{H}$ is approximately $1/k$-periodic in ζ:

$$\mathbf{H}(\eta, \zeta + 1/k) - \mathbf{H}(\eta, \zeta) = O(\eta^j \log \varepsilon) \quad \text{for } C^j\text{-smooth } H_1. \tag{7.7}$$

Let $\partial_- Z$ and $\partial_+ Z$ be connected components of the boundary ∂Z:

$$\partial_- Z \cup \partial_+ Z = \partial Z, \quad \mu|_{\partial_- Z} < 0 < \mu|_{\partial_+ Z}.$$

System (7.5) is integrable and its dynamics is trivial. In particular, any trajectory which begins on $\partial_- Z$ either reaches $\partial_+ Z$ or tends asymptotically to an equilibrium of (7.5).

If $\widehat{\mathbf{I}}^{cr}$ consists of two connected components, let μ_0 be some number, separating them and

$$S_0 = \{(\mu, \zeta) \in Z : \mu = \mu_0\}.$$

Proposition 7.4 *Suppose that $\widehat{\mathbf{I}}^{cr}$ is connected (respectively, disconnected). Then system (7.5) has a trajectory*

$$\gamma(t) = (\mu(t), \zeta(t)), \quad t \in [0, t_*], \quad \chi|_\gamma = \chi_*$$

(respectively, trajectories $\gamma^\pm(t) = (\mu^\pm(t), \zeta^\pm(t))$, $t \in [0, t_*^\pm]$, $\chi|_{\gamma^\pm} = \chi_*^\pm$) such that

(1) $\mu(0) \in \partial_- Z$, $\mu(t_*) \in \partial_+ Z$ (respectively, $\mu^-(0) \in \partial_- Z$, $\mu^-(t_*^-) \in S_0$ $\mu^+(0) \in S_0$, $\mu^+(t_*^+) \in \partial_+ Z$),

(2) $t_* < |\widehat{\mathbf{I}}|/(cv)$ (respectively, $t_*^-, t_*^+ < |\widehat{\mathbf{I}}|/(cv)$) for some constant c,

(3) assertions (1) and (2) hold for any trajectory with energy from the interval $[\chi_* - c_1 v^2, \chi_* + c_1 v^2]$ (respectively, from the intervals $[\chi_*^\pm - c_1 v^2, \chi_*^\pm + c_1 v^2]$).

Remark 7.1 *Due to (7.7) the quantity χ_* in (3) can be taken* $\bmod v/k$.

Proof. We consider only the case of connected $\widehat{\mathbf{I}}^{cr}$, For disconnected $\widehat{\mathbf{I}}^{cr}$ the proof is the same. Let

$$M = \max_{\zeta \in \mathbf{T}} \mathbf{H}(0, \zeta), \quad A = \max_{\zeta \in \mathbf{T}} \mathbf{H}_{\zeta\zeta}(0, \zeta).$$

Claim. For any $h \in (M - v^2/(8A), M)$ any $\widehat{\zeta}$ such that $\mathbf{H}(0, \widehat{\zeta}) = h$ satisfies $|\mathbf{H}_\zeta(0, \widehat{\zeta})| \le v/2$.

Proof of the claim. Take any $\widehat{\zeta}$ such that $\mathbf{H}(0, \widehat{\zeta}) = h \in (M - v^2/(8A), M)$. Suppose that $\mathbf{H}_\zeta(0, \widehat{\zeta}) > v/2$ (the case $\mathbf{H}_\zeta(0, \widehat{\zeta}) < -v/2$ is analogous).

For $\zeta > \widehat{\zeta}$ we have:

$$\mathbf{H}(0, \zeta) - \mathbf{H}(0, \widehat{\zeta}) - \mathbf{H}_\zeta(0, \widehat{\zeta})(\zeta - \widehat{\zeta}) + \frac{A}{2}(\zeta - \widehat{\zeta})^2 \ge 0.$$

Indeed, the function in the left-hand side and its first derivative in ζ vanish at $\zeta = \widehat{\zeta}$ while its second derivative is non-negative for any ζ. Taking $\zeta = \widehat{\zeta} + \mathbf{H}_\zeta(0,\widehat{\zeta})/A$, we get:

$$\mathbf{H}(0,\zeta) \geq \mathbf{H}(0,\widehat{\zeta}) + \mathbf{H}_\zeta^2(0,\widehat{\zeta})/(2A) > M - \frac{v^2}{8A} + \frac{v^2}{8A}.$$

This contradicts to the assumption that M is the global maximum of $\mathbf{H}(0,\zeta)$.

Now let us return to the proof of Proposition 7.4. Let $\gamma(t)$ be the trajectory with initial conditions

$$(\mu,\zeta) = (0,\zeta^0), \quad \mathbf{H}(0,\zeta^0) = h \in \Big(M - \frac{3v^2}{32A}, M - \frac{v^2}{32A}\Big).$$

According to Claim, while

$$\gamma(t) \in \widehat{\mathbf{I}}^{cr} \cap \{(\mu,\zeta) : \mathbf{H}(0,\zeta) \in (M - \frac{v^2}{8A}, M)\}, \tag{7.8}$$

we have: $\dot{\mu} > v/2$, $\dot{\zeta} = O(\varepsilon^{1/4})$. Recall that by Proposition 7.3 (b) $|\widehat{\mathbf{I}}^{cr}| = O(\varepsilon^{-1/8})$. Hence, $\gamma(t)$ passes $\widehat{\mathbf{I}}^{cr}$ through the ractangle (7.8) with μ-velocity greater than $v/2$.

Outside $\widehat{\mathbf{I}}^{cr}$ $\dot{\zeta}$ is separated from zero. Therefore, in average, μ increases with velocity v. ■

Proposition 7.5 *For C^6-smooth H_1 and $\mu \in \mathbf{I}$*

$$\begin{aligned} & \chi(\mu_{n+1},\zeta_{n+1}) - \chi(\mu_n,\zeta_n) + v\nu_0 t_{n+1} \\ = \ & \big(\nu(\sqrt{\varepsilon}\mu_n) - \nu_0\big)(u_n - vt_{n+1}) + |\log\varepsilon| O\big(\varepsilon^{3/4} + (\nu(\sqrt{\varepsilon}\mu) - \nu_0)^2\big) \end{aligned} \tag{7.9}$$

Proof of Proposition 7.5. We have by (7.6):

$$\begin{aligned} & \chi(\mu_{n+1},\zeta_{n+1}) - \chi(\mu_n,\zeta_n) + v\nu_0 t_{n+1} = \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}, \\ & \mathbf{A} = \frac{1}{\varepsilon}(E(\sqrt{\varepsilon}\mu_{n+1}) - E(\sqrt{\varepsilon}\mu_n)), \\ & \mathbf{B} = -\frac{1}{\sqrt{\varepsilon}}\nu_0(\mu_{n+1} - \mu_n), \\ & \mathbf{C} = \mathbf{H}(\sqrt{\varepsilon}\mu_{n+1},\zeta_{n+1}) - \mathbf{H}(\sqrt{\varepsilon}\mu_n,\zeta_n), \\ & \mathbf{D} = -v(\zeta_{n+1} - \nu_0 t_{n+1} - \zeta_n). \end{aligned}$$

By using (7.1)–(7.2), we get:

$$\begin{aligned} \mathbf{A} &= \nu(\sqrt{\varepsilon}\mu_n)\big(u_n - t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n)\big) + O(\varepsilon\log^2\varepsilon), \\ \mathbf{B} &= -\nu_0(u_n - t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n)), \\ \mathbf{D} &= -v\big(\nu(\sqrt{\varepsilon}\mu_n) - \nu_0\big)t_{n+1} + O(\varepsilon^{3/4}). \end{aligned}$$

To deal with the term $\mathbf{C}$, we use the estimates

$$\mu_{n+1} - \mu_n = O(\sqrt{\varepsilon}\log\varepsilon), \quad \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu,\zeta+\nu_0 t_{n+1}) - \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu,\zeta) = O(\varepsilon^{3/2}).$$

The last estimate follows from Proposition 3.1 for C^6-smooth H_1, because for $\sqrt{\varepsilon}\mu \in \mathbf{I}$ we have $\delta = |\nu(\sqrt{\varepsilon}\mu) - \nu_0| < \varepsilon^{1/4}$. Now we get:

$$\mathbf{C} = \big(\nu(\sqrt{\varepsilon}\mu_n) - \nu_0\big)t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n) + |\log\varepsilon|\big(O(\varepsilon^{3/4}) + O(\nu(\sqrt{\varepsilon}\mu) - \nu_0)^2)\big).$$

These estimates for $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}$ imply (7.9). ■

Since $\lambda t_{n+1} = |\log\varepsilon| + O(1)$, we have:

$$\begin{aligned} & u_n - vt_{n+1} > 2b \qquad \text{for } u_n \in [5b, c_\Theta + b], \\ & u_n - vt_{n+1} < -6b \quad \text{for } u_n \in [-c_\Theta - b, -5b]. \end{aligned}$$

Hence, with the help of u_n we can increase or decrease $\chi \bmod v/k$ on the trajectory. In particular, we can use the control $u_0, u_1, \dots$ first, to make χ close to χ_*:

$$\chi(\mu_n,\zeta_n) = (\chi_* + O(\sqrt{\varepsilon})) \bmod v/k, \qquad n = n_0 \leq v\varepsilon^{-1/2},$$

and then to preserve this condition for $n > n_0$ provided

$$|\nu(\sqrt{\varepsilon}\mu_n) - \nu_0| > C\varepsilon^{3/4}|\log\varepsilon| \quad \text{i.e.,}\ \mu \notin \widehat{\mathbf{I}}^{cr}. \tag{7.10}$$

Since the sequence $\chi(\mu_k, \zeta_k)$, $k > n_0$ just weakly oscillates, the trajectory (μ_n, ζ_n) really follows γ from Proposition 7.4 provided (7.10) holds. To enter the strip

$$S = \{(\mu, \zeta) : \mu \in \mathbf{I}^{cr}\},$$

we need

$$N_1 \leq O(|\widehat{\mathbf{I}}^{cr}|/\sqrt{\varepsilon}) \tag{7.11}$$

steps. Indeed, (7.11) follows from the fact that average increment of μ on this part of the trajectory we construct is $O(\sqrt{\varepsilon})$.

When the trajectory (μ_n, ζ_n) enters the strip S, error terms in (7.9) can dominate. In this case we choose the control $u \in [5b, c_\Theta + b]$. On $\gamma \cap S$ equation (7.5) implies that

$$\dot{\mu} = v - \mathbf{H}_\zeta(\eta^0, \zeta) > v/2. \tag{7.12}$$

Moreover, according to Proposition 7.4, this inequality remains true while

$$(\mu, \zeta) \in \{|\chi(\mu, \zeta) - \chi_*| < c_1 v^2\} \cap S, \qquad c_1 = 1/(32A). \tag{7.13}$$

Now turn to the trajectory (μ_n, ζ_n). By using (7.1) and (7.12), we see that (7.13) implies

$$\mu_{n+1} - \mu_n > \sqrt{\varepsilon}\Big(5b - \frac{2|\log\varepsilon|}{\lambda}\mathbf{H}_\zeta(\eta^0, \zeta_n)\Big) > \frac{3}{2}\sqrt{\varepsilon}b.$$

Hence, μ_k increases while $(\mu, \zeta) = (\mu_k, \zeta_k)$ satisfies (7.13). To pass S, we need not more than

$$N_2 = \frac{C\varepsilon^{-1/8}}{\sqrt{\varepsilon}b} = \frac{C}{\varepsilon^{5/8}b}$$

steps. During this time χ changes not more than by $N_2\, O(\varepsilon^{3/4}|\log\varepsilon|) = O(\varepsilon^{1/8}\log\varepsilon)$. This means that if the trajectory enters S with $\chi = (\chi_* + O(\sqrt{\varepsilon})) \bmod v/k$, it will remain in the domain (7.13) when leaving S.

Finally, for the total amount N of steps on the trajectory we have:

$$N = 2N_1 + N_2 = O(|\widehat{\mathbf{I}}|/\sqrt{\varepsilon}) = O(|\mathbf{I}|/\varepsilon).$$

■

7.2 Proof of Proposition 7.2

By using Lemma 6.2, we can construct a trajectory such that on the corresponding sequence of right compatible trajectories

$$(\overline{\Omega}_j, \overline{\Omega}_{j+1}),\ (\overline{\Omega}_{j+1}, \overline{\Omega}_{j+2}),\ \ldots$$

according to our needs $\widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n, \overline{\tau}_n)$, $n = j, j+1, \ldots$ belong either to $[6b, c_\Theta]$ or to $[-c_\Theta, -6b]$. For $|\varepsilon\overline{\rho}_n - \eta^0| < c\varepsilon^{1/4}$ (recall that we assume that $\eta^0 = 0$), combining (4.5) and the estimates

$$|\overline{\rho}_n - \rho_n| < 2b/b_\rho, \quad |\overline{\zeta}_n - \zeta_n| < 2b/b_\zeta, \quad |\overline{\tau}_n - \tau_n| < 2b/b_\tau,$$

we get:

$$\begin{aligned} \overline{\rho}_{n+1} &= \overline{\rho}_n - \widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n, \overline{\tau}_n) + (\overline{\tau}_{n+1} - \overline{\tau}_n - \overline{t}_{n+1})\mathbf{H}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n) + \widetilde{\alpha}_n, \\ \overline{\zeta}_{n+1} &= \overline{\zeta}_n + \nu(\varepsilon\overline{\rho}_n)\overline{t}_{n+1} + \varepsilon(\overline{\tau}_{n+1} - \overline{\tau}_n - \overline{t}_{n+1})\mathbf{H}_\eta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n) + \widetilde{\beta}_n, \end{aligned} \tag{7.14}$$

where $|\widetilde{\alpha}_n| < b/2$, $\quad |\widetilde{\beta}_n| < c^{-1}\varepsilon^{3/4}|\log\varepsilon|/2$.

We denote

$$\varepsilon\overline{\rho}_n = \sqrt{\varepsilon}\mu_n + \varepsilon\overline{\tau}_n\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n, \overline{\zeta}_n).$$

Then

$$\begin{aligned}
\mu_{n+1} &= \mu_n - \widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n, \overline{\tau}_n) - \overline{t}_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n, \overline{\zeta}_n, \overline{\tau}_n) + \alpha_n,\\
\alpha_n &= \widetilde{\alpha}_n - (\overline{\tau}_n + \overline{t}_{n+1})\delta_1 + \overline{\tau}_{n+1}\delta_2,\\
\delta_1 &= \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n + \varepsilon\overline{\tau}_n\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n, \overline{\zeta}_n), \overline{\zeta}_n) - \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n, \overline{\zeta}_n),\\
\delta_2 &= \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n + \varepsilon\overline{\tau}_n\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n, \overline{\zeta}_n), \overline{\zeta}_n) - \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_{n+1}, \overline{\zeta}_{n+1}).
\end{aligned}$$

We have: $\delta_1 = O(\varepsilon^{3/4})$, $\delta_2 = O(\varepsilon^{1/4}\log\varepsilon) + O(\sqrt{\varepsilon}\mu\log\varepsilon)$. Indeed, the first estimate is obvious. The second one follows from the equations

$$\mu_{n+1} - \mu_n = O(\log\varepsilon), \quad \overline{\zeta}_{n+1} - \overline{\zeta}_n = -k_0/k + O(\sqrt{\varepsilon}\mu\log\varepsilon),$$
$$\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu, \zeta + k_0/k) - \mathbf{H}_\zeta(\sqrt{\varepsilon}\mu, \zeta) = O(\varepsilon).$$

The last equation follows from Proposition 3.1 for C^j-smooth H_1, $j > 5$.

Thus $|\alpha_n| < b$ and the quantity $u_n = -\widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n, \overline{\tau}_n) + \alpha_n$ belongs to one of intervals (7.4) and we get (7.1). Since

$$\nu(\varepsilon\overline{\rho}_n) = \nu(0) + \nu_0'\sqrt{\varepsilon}\mu_n + \frac{\nu_0''}{2}\varepsilon\mu_n^2 + O(\varepsilon) + \varepsilon^{3/2}O(\mu_n^3),$$

we get (7.1)–(7.2). ∎

8 Passage through a vague interval

We fix a resonance $\eta = \eta^0$, generating a vague interval $\mathbf{I} \subset \mathcal{D}_V$. Below without loss of generality we assume that $\eta^0 = 0$. Denote $\nu_0 = \nu(0) = p_0/q_0$, where $p_0 \in \mathbf{Z}$, $q_0 \in \mathbf{N}$, $q_0 < 2/c_*$, and the fraction p_0/q_0 is irreducible.

Lemma 8.1 *Suppose that* $\mathbf{H_1 2}$ *holds for the vague resonance* $\eta = 0$. *Then an admissible trajectory can be continued through the corresponding vague interval* $\mathbf{I} \subset \mathcal{D}_V$. *This takes time* $t = O(|\log\varepsilon|/\varepsilon)$.

Due to $\mathbf{H_1 2}$, the functions $F_\pm$ satisfy conditions (1)–(4) of Proposition 5.4. For sufficiently small c', c'' the functions

$$\begin{aligned}
F_{c',c'',+}(\zeta) &= \max_{s\in[-q_0,0]}\{\Theta_\xi(0, \zeta - \nu_0 s, s) : (0, \zeta - \nu_0 s, s) \in J_{2c',2c''}\},\\
F_{c',c'',-}(\zeta) &= \min_{s\in[-q_0,0]}\{\Theta_\xi(0, \zeta - \nu_0 s, s) : (0, \zeta - \nu_0 s, s) \in J_{2c',2c''}\}
\end{aligned}$$

also satisfy these conditions.

Let $\gamma_0^\pm = \gamma_0^\pm(\zeta)$ be such that $F_{c',c'',\pm}(\zeta) = \Theta_\xi(0, \zeta - \nu_0\gamma_0^\pm, \gamma_0^\pm)$. Since $(0, \zeta - \nu_0\gamma_0^\pm, \gamma_0^\pm) \in J_{2c',2c''}$, we have: $\gamma_0^\pm(\zeta) = \Psi^{0,\zeta,\gamma_0^\pm(\zeta)}(0, \zeta, 0)$. The functions $\gamma_0^\pm(\zeta)$ are piecewise C^{j-1}-smooth.

For small $|\eta|$ we put

$$\widehat{F}_{c',c'',\pm}(\eta, \zeta) = \Theta_\xi(\eta, \zeta - \nu(\eta)\gamma^\pm, \gamma^\pm),$$
$$\gamma^\pm = \gamma^\pm(\eta, \zeta) = \Psi^{0,\zeta,\gamma_0^\pm(\zeta)}(\eta, \zeta, 0).$$

Hence, $\widehat{\Theta}_\tau(\eta, \zeta, \gamma^\pm) = 0$ and $(\eta, \zeta - \nu(\eta)\gamma^\pm, \gamma^\pm) \in J_{c',c''}$. The the functions $\widehat{F}_{c',c'',\pm}$ are smooth in η and for any small $|\eta|$ satisfy conditions (1)–(4) of Proposition 5.4 with $c_f/2$ instead of c_f.

We will assume that $c', c'', |\eta|$ we deal with are sufficiently small in this sense. For brevity we will skip the subscripts c', c''. In partiqular we write $\widehat{F}_\pm$ instead of $\widehat{F}_{c',c'',\pm}$.

Suppose that an admissible trajectory $\mathcal{O} = (\Omega_{-k}, \dots, \Omega_k)$ with the code $\overline{\mathcal{O}} = (\overline{\Omega}_{-k}, \dots, \overline{\Omega}_k)$ reaches the vague interval i.e., $\varepsilon\overline{\rho}_k \in \mathbf{I}$, $\overline{\rho}_k < 0$. Note that according to Proposition 5.2, $\mathbf{I} \subset [-C_\nu/(q_0 K), C_\nu/(q_0 K)]$, where C_ν is the Lipschitz constant for $\nu^{-1}(\eta)$ near the point $\eta = 0$, and K satisfies (5.3).

Below the functions $F_\pm$ will play the role of Θ_ζ in the dynamical equations. In fact we use the following

Lemma 8.2 *Suppose that* $\mathbf{H_1 2}$ *holds for the vague* $\eta = 0$. *Then for any* $|\varepsilon\rho_k| < C_\nu/(q_0 \overline{K})$ *and* $* \in \{+,-\}$ *there exists a right compartible quasi-trajectory* $(\overline{\Omega}_k, \overline{\Omega}_{k+1})$ *and integer* $\mathbf{t}$

$$\mathbf{t} - \Big[\frac{1}{\lambda}(K_0 - \log\varepsilon)\Big] \in \{1, 2, \dots, 2q_0\},$$

where $[\cdot] : \mathbf{R} \to \mathbf{Z}$ *is the integer part of the argument, such that*

$$\widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_{k+1}, \overline{\zeta}_{k+1}, \overline{\tau}_{k+1}) = \widehat{F}_*(\varepsilon\rho_k, \zeta_k + \nu(\varepsilon\rho_k)\mathbf{t}) + \alpha_*, \tag{8.1}$$
$$|\alpha_*| < d, \quad d \text{ is a small constant.}$$

We prove Lemma 8.2 in Section 9, and now we define one more dynamical problem: Problem V (V is from "vague").

Problem V. *Consider the discrete dynamical system*

$$\mu_{n+1} - \mu_n = -\sqrt{\varepsilon}\big(v_n + t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n, \zeta_n) + \alpha_n\big), \tag{8.2}$$
$$\zeta_{n+1} - \zeta_n = \nu(\sqrt{\varepsilon}\mu_n)t_{n+1} + \beta_n. \tag{8.3}$$

Here for any n

$$\mathbf{t}_n \le t_n \le \mathbf{t}_n + 2q_0, \qquad t_n \in \mathbf{Z}, \tag{8.4}$$
$$\mathbf{t}_n - \big[\tfrac{1}{\lambda}(K_0 - \log\varepsilon)\big] \in \{1, 2, \dots, 2q_0\}, \tag{8.5}$$
$$|\beta_n| < C\varepsilon^{3/4}|\log\varepsilon|, \quad |\alpha_n| < d, \quad d = O(\varepsilon^{3/4}\log\varepsilon). \tag{8.6}$$

The variables $\mathbf{t}_n, v_n$ *should be regarded as a control, where according to our needs* $\mathbf{t}_n$, *satisfying (8.5), can be chosen arbitrarily, and* v_n *can be chosen equal to one of the following two quantities:*

$$\widehat{F}_\pm(\sqrt{\varepsilon}\mu_{n-1}, \zeta_{n-1} + \nu(\sqrt{\varepsilon}\mu_{n-1})\mathbf{t}_n). \tag{8.7}$$

The small terms α_n, β_n *are regarded as a noice. They and* t_n *take arbitrary values in the correspondent intervals. The goal is to construct a trajectory which starts at an arbitrary point* (μ_0, ζ_0), $\sqrt{\varepsilon}\mu_0 > -1/(q_0\overline{K})$ *and finishes at a point* (μ_N, ζ_N), $\sqrt{\varepsilon}\mu_N > 1/(q_0\overline{K})$ *with* N *as small as possible.*

Lemma 8.3 *Problem V always has a solution with* $N \le 1/(c\varepsilon\overline{K})$.

Lemma 8.4 *Suppose that Problem V has a solution. Then Lemma 8.1 holds with* $t = O(N\log\varepsilon)$.

9 Proof of Lemma 8.2

The argument is analogous to the one we used in Section 6.1. We prove the lemma for *= +.

We need to satisfy (6.5)–(6.8) and condition (8.1). Again for brevity we put $\Omega_+ = \overline{\Omega}_{k+1}$.

(1) First, we consider instead of (8.1) the equation

$$\Theta_\xi(\varepsilon\rho_k, \zeta_k + \nu(\varepsilon\rho_k)t, -t) = \widehat{F}_+(\varepsilon\rho_k, \zeta_k + \nu(\varepsilon\rho_k)\mathbf{t}). \tag{9.1}$$

with unknown integer $t \in [\mathbf{t}, \mathbf{t} + 2q_0]$.

According to definition of $\widehat{F}_+$, there exists $t = \widehat{t} \in [\mathbf{t}, \mathbf{t} + 2q_0]$ such that

$$(\varepsilon\rho_k, \zeta_k + \nu(\varepsilon\rho_k)\widehat{t}, -\widehat{t}) \in J_{2c', 2c''} \quad \text{and (9.1) holds.} \tag{9.2}$$

(2) We define t_+, τ_*, ζ_* by (6.11). Then t_+ satisfies (6.8). Conditions (9.2) imply that

$$(\varepsilon\rho_k, \zeta_* - \nu(\varepsilon\rho_k)\tau_*, \tau_*) \in J_{c',c''}, \quad \widehat{\Theta}_\zeta(\varepsilon\rho_k, \zeta_*, \tau_*) = \widehat{F}_+(\varepsilon\rho_k, \zeta_k + \nu(\varepsilon\rho_k)\mathbf{t}).$$

(3) We define ρ_+, ζ_+ by (6.12).

(4) We find τ_+ from the equation $\overline{\partial}\Theta(\varepsilon\rho_+, \zeta_+ - \nu(\varepsilon\rho_+)\tau_+, \tau_+) = 0$ which is equivalent to $\widehat{\Theta}_\tau(\varepsilon\rho_+, \zeta_+, \tau_+) = 0$. Hence,

$$\tau_+ = \Psi^{\rho_k, \zeta_*, \tau_*}(\varepsilon\rho_+, \zeta_+, 0).$$

Since $\tau_* = \Psi^{\rho_k,\zeta_*,\tau_*}(\varepsilon\rho_k,\zeta_*,0)$ and $|\zeta_+ - \zeta_*| = O(\varepsilon^{3/4}\log\varepsilon)$, we have:

$$|\tau_+ - \tau_*| = O(\varepsilon^{3/4}|\log\varepsilon|).$$

Therefore, $|\tau_+| < 1$. We have:

$$\big|\widehat{\Theta}_\zeta(\varepsilon\rho_+,\zeta_+,\tau_+) - \widehat{\Theta}_\zeta(\varepsilon\rho_k,\zeta_*,\tau_*)\big| = O(\varepsilon^{3/4}\log\varepsilon).$$

Now (6.5)–(6.6) can be checked easily. ∎

10 Proof of Lemma 8.3

Suppose that the trajectory $\mathcal{O}$ of the Problem V reaches some point $(\mu,\zeta) = (\mu_s,\zeta_s)$ with $-1/(q_0\overline{K}) \le \sqrt{\varepsilon}\mu_s \le 1/(q_0\overline{K})$.

10.1 Step 1 (far from the resonance)

First, consider the case

$$\varepsilon^{1/6} < |\sqrt{\varepsilon}\mu| < 1/(q_0\overline{K}). \tag{10.1}$$

Proposition 10.1 *Let H_1 be C^j-smooth. Then for sufficiently small ε (10.1) implies that $\big|t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n)\big| < C|\log\varepsilon|^{1-j}$.*

Proof. Let c_ν be a Lipschitz constant for $\nu(y)$ and $c'_\nu = \frac{d\nu}{dy}(0)$. According to $\mathbf{H_0 4}$, $c'_\nu > 0$. Inequality (10.1) implies that

$$q_0\varepsilon^{1/6}c'_\nu/2 < |q_0\nu(\sqrt{\varepsilon}\mu) - p_0| < c_\nu/\overline{K}. \tag{10.2}$$

For any resonance $\nu(\sqrt{\varepsilon}\mu) = p/q$, $p \in \mathbf{Z}$, $q \in \mathbf{N}$, satisfying (10.2), we have:

$$1 \le |q_0 p - p_0 q| < c_\nu q/\overline{K}.$$

Since $\overline{K}$ is of order $|\log\varepsilon|$, q is big: $q > \overline{K}/c_\nu$.

Now let us estimate $\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu,\zeta)$ for μ, satisfying (10.1). According to (3.1)–(3.2), we have:

$$|\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu,\zeta)| \le C_{\mathbf{H}}q^{-j} \le C_{\mathbf{H}}c_\nu^j\overline{K}^{-j}$$

for some constant $C_{\mathbf{H}}$. Here we have used the standard estimate for Fourier coefficients for C^j-smooth functions. Finally we have:

$$|t_{n+1}\mathbf{H}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n)| < \frac{2|\log\varepsilon|c_\nu^j C_{\mathbf{H}}}{\lambda\overline{K}^j} < \frac{C}{|\log\varepsilon|^{j-1}}.$$

∎

We take

$$v_n = \widehat{F}_-(\sqrt{\varepsilon}\mu_{n-1},\zeta_{n-1} + \nu(\sqrt{\varepsilon}\mu_{n-1})\mathbf{t}_n).$$

Hence, (8.2)–(8.3) can be replaced by the system

$$\begin{aligned} \mu_{n+1} - \mu_n &= -\sqrt{\varepsilon}\Big(\widehat{F}_-(\sqrt{\varepsilon}\mu_{n-1},\zeta_{n-1} + \nu(\sqrt{\varepsilon}\mu_{n-1})\mathbf{t}_n) + \widehat{\alpha}_n\Big), && (10.3)\\ \zeta_{n+1} - \zeta_n &= \nu(\sqrt{\varepsilon}\mu_n)t_{n+1} + \beta_n && (10.4)\end{aligned}$$

with $|\beta_n| < C\varepsilon^{3/4}|\log\varepsilon|$ and $|\widehat{\alpha}_n| < 2d$. Below we skip hats over α_n.

Our idea is as follows. Starting from the point (μ_s,ζ_s) for natural k we will have:

$$\begin{aligned} &\zeta_{s+k-1} + \nu(\sqrt{\varepsilon}\mu_{s+k-1})\mathbf{t}_{s+k} \approx \widehat{\zeta} + k\Delta,\\ &\text{with}\quad \Delta = \nu(\sqrt{\varepsilon}\mu_s)(\mathbf{t} + 2q_0), && (10.5)\\ &\mathbf{t} \in \mathbf{N},\quad \mathbf{t} - \Big[\frac{1}{\lambda}(K_0 - \log\varepsilon)\Big] \in \{1,2,\dots,2q_0\}, && (10.6)\\ &\text{and some}\quad \widehat{\zeta} \in \mathbf{T}^1. \end{aligned}$$

Hence, if Δ satisfies good arithmetic properties, for some $\mathbf{q} \in \mathbf{N}$ the points $\zeta_s, \dots, \zeta_{s+\mathbf{q}-1}$ more or less uniformly cover the circle $\{\zeta \bmod 1\}$. Therefore,

$$\mu_{s+\mathbf{q}+1} - \mu_{s+1} \approx -\sqrt{\varepsilon} \sum_{k=0}^{\mathbf{q}-1} \widehat{F}_-\big(\sqrt{\varepsilon}\mu_s, \zeta_s + k\Delta\big) - \sqrt{\varepsilon} \sum_{k=0}^{\mathbf{q}-1} \alpha_{s+k}.$$

Here the first sum approximately equals $-\mathbf{q}\sqrt{\varepsilon}\langle \widehat{F}_-(0)\rangle$, while the second one does not exceed $2\mathbf{q}\sqrt{\varepsilon}d$. Hence, if ε is sufficiently small, $\mu_{s+\mathbf{q}+1} - \mu_{s+1} > \frac{1}{2}\langle \widehat{F}_-(0)\rangle\sqrt{\varepsilon}\mathbf{q}$, because d can be chosen less than $\frac{1}{4}\langle \widehat{F}_-(0)\rangle$.

More precisely, we choose $\mathbf{t}$ such that (10.5)–(10.6) hold and Δ satisfies the following weak resonant condition (WRC):

WRC. *Given a large constant $C > 0$ and a small constant $c > 0$.*
For any μ_s satisfying (10.1) there exist $\mathbf{p}, \mathbf{q} \in \mathbf{Z}$ such that
(a) $C \le \mathbf{q} \le c(\varepsilon \log^2 \varepsilon)^{-1/6}$,
(b) the fraction $\mathbf{p}/\mathbf{q}$ is irreducible,
(c) there exists Δ, determined by (10.5)–(10.6), satisfying the inequality $|\mathbf{q}\Delta - \mathbf{p}| < \mathbf{q}^{-1}$.

Proposition 10.2 *For any $\mu = \mu_s$, satisfying (10.1) there exists $\mathbf{t}$, satisfying (10.6) such that for the corresponding Δ WRC holds.*

Proof of Proposition 10.2 is contained in Section 12.

We choose the sequence $\mathbf{t}_{s+k}$, such that for any $k = 0, \dots, \mathbf{q}$

$$\begin{aligned} &\mathbf{t} \le \mathbf{t}_{s+k} \le \mathbf{t} + 2q_0, \\ &|\zeta_{s+k-1} + \nu(\sqrt{\varepsilon}\mu_s)\mathbf{t}_{s+k} - \widehat{\zeta} - k\Delta| \le c^{-1}k^2\sqrt{\varepsilon}|\log \varepsilon|, \\ &\widehat{\zeta} = \zeta_{s-1} + \nu(\sqrt{\varepsilon}\mu_s)\mathbf{t}_s. \end{aligned} \qquad (10.7)$$

These inequalities can be satisfied. Indeed, for $k = 0$ (10.7) holds. Now we apply induction argument. Suppose that (10.7) holds for some $k > 0$. Then by (10.3),

$$\mu_{s+k} = \mu_s + O(k\sqrt{\varepsilon}). \qquad (10.8)$$

Using (10.8) and (10.4), we get:

$$\begin{aligned} &\zeta_{s+k} + \nu(\sqrt{\varepsilon}\mu_s)\mathbf{t}_{s+k+1} \\ = \;&\zeta_{s+k-1} + \nu(\sqrt{\varepsilon}\mu_s)(t_{s+k} + \mathbf{t}_{s+k+1}) + \psi_k, \end{aligned} \qquad (10.9)$$

where $|\psi_k| < c^{-1}k\sqrt{\varepsilon}|\log \varepsilon|$ and the constant $c > 0$ does not depend on k. The quantity (10.9) equals

$$\zeta_{s+k-1} + \nu(\sqrt{\varepsilon}\mu_s)\mathbf{t}_{s+k} + \nu(\sqrt{\varepsilon}\mu_s)(t_{s+k} - \mathbf{t}_{s+k} + \mathbf{t}_{s+k+1}) + \psi_k.$$

Since $t_{s+k} - \mathbf{t}_{s+k} \in [0, 2q_0] \cap \mathbf{Z}$ then $\mathbf{t}_{s+k+1} \in [\mathbf{t}, \mathbf{t} + 2q_0] \cap \mathbf{Z}$ can be chosen such that $t_{s+k} - \mathbf{t}_{s+k} + \mathbf{t}_{s+k+1} = \mathbf{t} + 2q_0$. This finishes the induction.

Proposition 10.3 *Let $\mathbf{t}$ be such that (10.5)–(10.6) hold and the corresponding Δ satisfies WRC. Suppose that the sequence (μ_n, ζ_n) satisfies (10.3)–(10.4), and (10.7). Then*

$$\Sigma \equiv -\sum_{k=0}^{\mathbf{q}-1} \widehat{F}_-\big(\sqrt{\varepsilon}\mu_{s+k}, \zeta_{s+k} + \nu(\sqrt{\varepsilon}\mu_{s+k})\mathbf{t}_{s+k+1}\big) > \frac{2\mathbf{q}}{3}\big|\langle \widehat{F}_-(0)\rangle\big|. \qquad (10.10)$$

Proof of Proposition 10.3. By definition (see Section 8), the function $\widehat{F}_-(\eta, \zeta) \equiv \widehat{F}_{c',c'',-}(\eta, \zeta)$ is smooth in η provided η is close to zero.

By using the estimate (10.8), we get:

$$\begin{aligned} \Sigma &= \Sigma_1 + O(\mathbf{q}^2\sqrt{\varepsilon}), \\ \Sigma_1 &= -\sum_{k=0}^{\mathbf{q}-1} \widehat{F}_-\big(\sqrt{\varepsilon}\mu_s, \zeta_{s+k} + \nu(\sqrt{\varepsilon}\mu_{s+k})\mathbf{t}_{s+k+1}\big). \end{aligned} \qquad (10.11)$$

Estimates (10.7)–(10.8) imply that

$$\zeta_{s+k}+\nu(\sqrt{\varepsilon}\mu_{s+k})\mathbf{t}_{s+k+1}=\widehat{\zeta}+(k+1)\Delta+\delta_k, \qquad |\delta_k|<2c^{-1}k^2\sqrt{\varepsilon}|\log\varepsilon|.$$

Therefore by WRC Σ_1 can be regarded as an integral sum for

$$\int_0^1 \widehat{F}_-(\sqrt{\varepsilon}\mu_s,\zeta)\,d\zeta=\langle\widehat{F}(0)\rangle+O(\sqrt{\varepsilon}\mu_s).$$

Provided $\mathbf{q}$ is sufficiently large (to have a good approximation of the integral), but less than $c(\varepsilon\log^2\varepsilon)^{-1/6}$ (to have $|\delta_k|<1/\mathbf{q}$), the following estimate holds:

$$\big|\Sigma_1-\mathbf{q}|\langle\widehat{F}(0)\rangle|\big|<\frac{1}{4}\mathbf{q}|\langle\widehat{F}(0)\rangle|.$$

Combined with (10.11), this implies (10.10). ∎

According to (10.3),

$$\mu_{s+\mathbf{q}+1}-\mu_{s+1}=\sqrt{\varepsilon}\Sigma+\sqrt{\varepsilon}\sum_{k=0}^{\mathbf{q}-1}\widehat{\alpha}_{s+k+1}, \qquad |\widehat{\alpha}_n|<2d.$$

Proposition 10.3 implies that

$$\mu_{s+\mathbf{q}+1}-\mu_{s+1}>\sqrt{\varepsilon}\mathbf{q}\Big(\frac{2}{3}|\langle\widehat{F}_-(0)\rangle|-2d\Big)>\frac{\sqrt{\varepsilon}\mathbf{q}}{2}\big|\langle\widehat{F}_-(0)\rangle\big|$$

provided $d<\frac{1}{12}\langle\widehat{F}_-(0)\rangle$.

We proceed in this way while (10.1) holds.

10.2 Step 2 (near the resonance)

Now turn to the case

$$\widehat{C}\sqrt{\varepsilon}/|\log\varepsilon|<|\sqrt{\varepsilon}\mu|<\varepsilon^{1/6}. \tag{10.12}$$

We use the equations

$$\nu(\sqrt{\varepsilon}\mu)=\nu_0+\nu_0'\sqrt{\varepsilon}\mu+O(\varepsilon\mu^2).$$

Due to (10.12) the function $\mathbf{H}(\sqrt{\varepsilon}\mu,\zeta)$ is close to a $1/q_0$-periodic in ζ function $\widetilde{\mathbf{H}}$ (see Corolary 3.1):

$$|\mathbf{H}(\sqrt{\mu},\zeta)-\widetilde{\mathbf{H}}(\sqrt{\mu},\zeta)|=O(\varepsilon^{j/6}q_0^{j+1}).$$

We replace $\mathbf{H}$ by $\widetilde{\mathbf{H}}$ in (8.2). Then the order of the error term α_n remains the same.

Since $\nu_0=p_0/q_0$, the functions $\widehat{F}_\pm(0,\zeta)$ are $\frac{1}{q_0}$-periodic in ζ, and the sequences t_n, $\mathbf{t}_n$ are integer, we will consider in (8.2)–(8.3) $\zeta=\zeta \bmod \frac{1}{q_0}$. Therefore we can replace (8.3) by the equation $\zeta_{n+1}-\zeta_n=(\nu(\sqrt{\varepsilon}\mu_n)-\nu_0)t_{n+1}+\beta_n$. Equations (8.2)–(8.3) take the form

$$\begin{aligned}\mu_{n+1}-\mu_n &= -\sqrt{\varepsilon}\big(\widehat{v}_n+t_{n+1}\widetilde{\mathbf{H}}_\zeta(\sqrt{\varepsilon}\mu_n,\zeta_n)+\widehat{\alpha}_n\big), && (10.13)\\ \zeta_{n+1}-\zeta_n &= \nu_0'\sqrt{\varepsilon}\mu_n t_{n+1}+\beta_n'+\beta_n'', && (10.14)\end{aligned}$$

$$|\widehat{\alpha}_n|<2d, \quad |\beta_n'|\le C\varepsilon\mu_n^2|\log\varepsilon|, \quad |\beta_n''|\le C\varepsilon^{3/4}|\log\varepsilon|,$$

where $\widehat{v}_n$ equals to one of the following two quantities:

$$\widehat{F}_\pm(0,\widehat{\zeta}_n), \qquad \widehat{\zeta}_n=\zeta_{n-1}+\nu_0'\sqrt{\varepsilon}\mu_{n-1}\mathbf{t}_n+O(\varepsilon\mu_{n-1}^2|\log\varepsilon|). \tag{10.15}$$

Note that according to (10.14)–(10.15)

$$\zeta_n=\widehat{\zeta}_n+O(\sqrt{\varepsilon}\mu_n)+O(\varepsilon^{3/4}|\log\varepsilon|). \tag{10.16}$$

We choose

$$\mathbf{t}\equiv\Big[\frac{1}{\lambda}\big(K_0-\log\varepsilon\big)\Big]+1$$

and the sequence $\mathbf{t}_{s+k}$, $k = 0, 1, \dots$ such that

$$\begin{aligned}
&\mathbf{t} \le \mathbf{t}_{s+k} \le \mathbf{t} + 2q_0,\\
&\widehat{\zeta}_{s+k+1} = \widehat{\zeta}_{s+k} + \nu_0'\sqrt{\varepsilon}\mu_{s+k}\mathbf{t}_* + \delta_{s+k}, \qquad \mathbf{t}_* = \mathbf{t} + 2q_0, \quad (10.17)\\
&|\delta_{s+k}| < C_1\big(\varepsilon\mu_{s+k}^2|\log\varepsilon| + \varepsilon^{3/4}|\log\varepsilon|\big).
\end{aligned}$$

The sequence $\mathbf{t}_{s+k}$ can be chosen as follows. By (10.15) and (10.14)

$$\begin{aligned}
\widehat{\zeta}_{s+k+1} &= \zeta_{s+k} + \nu_0'\sqrt{\varepsilon}\mu_{s+k}\mathbf{t}_{s+k+1} + O(\varepsilon\mu_{s+k}^2|\log\varepsilon|)\\
&= \zeta_{s+k-1} + \nu_0'\sqrt{\varepsilon}(\mu_{s+k-1}t_{s+k} + \mu_{s+k}\mathbf{t}_{s+k+1})\\
&\qquad +\beta_{s+k-1}' + \beta_{s+k-1}'' + O(\varepsilon\mu_{s+k}^2|\log\varepsilon|)\\
&= \widehat{\zeta}_{s+k} + \nu_0'\sqrt{\varepsilon}(\mu_{s+k-1}t_{s+k} - \mu_{s+k-1}\mathbf{t}_{s+k} + \mu_{s+k}\mathbf{t}_{s+k+1})\\
&\qquad +\beta_{s+k-1}' + \beta_{s+k-1}'' + O(\varepsilon\mu_{s+k}^2|\log\varepsilon|)
\end{aligned}$$

We choose $\mathbf{t}_{s+k+1} \in [\mathbf{t}, \mathbf{t} + 2q_0] \cap \mathbf{Z}$ such that $t_{s+k} - \mathbf{t}_{s+k} + \mathbf{t}_{s+k+1} = \mathbf{t} + 2q_0$. Then (10.17) holds.

Lemma 8.3 on step **2** follows from

Proposition 10.4 *For any*

$$\mu_s, \zeta_s, \qquad \textit{with} \quad \sqrt{\varepsilon}\mu_s > -\varepsilon^{1/6}, \quad \zeta_s = \zeta_s \bmod \frac{1}{q_0}$$

there exists a trajectory

$$(\mu_n, \zeta_n), \qquad n = s, \dots, s + N \tag{10.18}$$

of Problem $\mathbf{V}$ *such that*

(a) $\mu_{s+N} > -\widehat{C}/|\log\varepsilon|$*,*

(b) $N < C\varepsilon^{-5/6}$*.*

Remark 10.1 *By using Proposition 10.4, it is possible to pass through the interval* $-\varepsilon^{1/6} \le \sqrt{\varepsilon}\mu \le -\widehat{C}\sqrt{\varepsilon}/|\log\varepsilon|$*. The interval* $\widehat{C}\sqrt{\varepsilon}/|\log\varepsilon| \le \sqrt{\varepsilon}\mu \le \varepsilon^{1/6}$ *can be passed analogously.*

To construct such a trajectory, we consider the systems

$$\begin{cases} \dot{\mu} = -\frac{1}{\mathbf{t}_*}(\mathcal{F}_\pm)_\zeta(\zeta) - \widetilde{\mathbf{H}}_\zeta(\sqrt{\varepsilon}\mu, \zeta), \\ \dot{\zeta} = \nu_0'\mu + \varepsilon^{1/4}\mathbf{H}_\eta(\sqrt{\varepsilon}\mu, \zeta), \end{cases} \tag{10.19}$$

$$(\mathcal{F}_\pm)_\zeta(\zeta) = \widehat{F}_\pm(0, \zeta) - \mathbf{F}_\pm, \quad \mathbf{F}_\pm = \int_0^1 \widehat{F}_\pm(0, \zeta)\, d\zeta. \tag{10.20}$$

This system is Hamiltonian with symplectic structure $d\mu \wedge d\zeta$ and Hamiltonian function

$$\begin{aligned}
\chi_\pm(\mu, \zeta) &= \frac{1}{2}\nu_0'\mu^2 + \widetilde{\mathbf{H}}(\sqrt{\varepsilon}\mu, \zeta) + \frac{1}{\mathbf{t}_*}F_\pm(\zeta) + h_\pm, && (10.21)\\
h_\pm &= -\max_\zeta\Big(\widetilde{\mathbf{H}}(\sqrt{\varepsilon}\mu, \zeta) + \frac{1}{\mathbf{t}_*}F_\pm(\zeta)\Big). && (10.22)
\end{aligned}$$

Below in this section we concentrate on properties of the system with Hamiltonian χ_-. The case of the function χ_+ can be studied analogously. System (10.19) is integrable and has phase portrait analogous to that of a pendulum.

We note that because of (10.13)–(10.14)

$$\begin{aligned}
&\chi_-(\mu_{n+1}, \zeta_{n+1}) - \chi_-(\mu_n, \zeta_n) = \mathbf{d}_n^\pm - \sqrt{\varepsilon}\nu_0'\mu_n\widehat{\alpha}_n + \widetilde{\beta}_n' + \widetilde{\beta}_n'',\\
&\mathbf{d}_n^\pm = -\sqrt{\varepsilon}\nu_0'\mu_n\widehat{F}_\pm(0, \widehat{\zeta}_n) + \frac{1}{\mathbf{t}_*}(\mathcal{F}_-(\zeta_{n+1}) - \mathcal{F}_-(\zeta_n)), \qquad (10.23)\\
&|\widetilde{\beta}_n'| \le \widetilde{C}\varepsilon\mu_n^2|\log\varepsilon|, \quad |\widetilde{\beta}_n''| \le \widetilde{C}\varepsilon^{3/4}|\log\varepsilon|. \qquad (10.24)
\end{aligned}$$

The superscript $+$ or $-$ in $\mathbf{d}_n^\pm$ corresponds to the choice of $\widehat{v}_n$ in (10.15).

Remark 10.2 *According to (10.17), the difference* $\widehat{\zeta}_{n+1} - \widehat{\zeta}_n \in \mathbf{R}/(q_0^{-1}\mathbf{Z})$ *is close to zero. Below we will regard* $\widehat{\zeta}_{n+1} - \widehat{\zeta}_n$ *(respectively* $\zeta_{n+1} - \zeta_n$*) as a real number which belongs to a small neighborhood of zero. Due to this convention we regard* $\widehat{\zeta}_{s+k+1} - \widehat{\zeta}_s$ *(respectively* $\zeta_{s+k+1} - \zeta_s$*) as the sum* $\sum_{n=s}^{s+k}(\widehat{\zeta}_{n+1} - \widehat{\zeta}_n)$ *(respectively* $\sum_{n=s}^{s+k}(\zeta_{n+1} - \zeta_n)$*).*

For $s \le k' \le k''$ we put $\mathbf{D}^{\pm}_{k',k''} = \chi_-(\mu_{k''}, \zeta_{k''}) - \chi_-(\mu_{k'}, \zeta_{k'})$, where the superscript of $\mathbf{D}$ is the same as the one of $\mathbf{d}^{\pm}_n$ in (10.23).

Proposition 10.5 *Suppose that the sequence $\zeta_{k'}, \zeta_{k'+1}, \ldots, \zeta_{k''}$ is monotonous. Then there exists a constant $K \in \mathbf{N}$ such that for $k'' - k' \ge K$ and for $|\widehat{\zeta}_{s+k+1} - \widehat{\zeta}_s| < 1$*

$$\mathbf{D}^{\pm}_{k',k''} = -\frac{1}{\mathbf{t}_*} a_\pm + a'_\pm + a''_\pm,$$

where

$$\begin{aligned} a_\pm &= -\int_{\zeta_{k'}}^{\zeta_{k''}} \widehat{F}_\pm(0,\zeta)\, d\zeta + \mathcal{F}(\zeta_{k''}) - \mathcal{F}(\zeta_{k'}), && (10.25) \\ |a'_\pm| &< \frac{2d}{\mathbf{t}_*} |\zeta_{k''} - \zeta_{k'}|, \\ |a''_\pm| &< C\big(\sqrt{\varepsilon}\mu_{k',k''} + (k''-k')\varepsilon\mu^2_{k',k''}|\log\varepsilon| + (k''-k')\varepsilon^{3/4}|\log\varepsilon|\big), \\ \mu_{k',k''} &= \max_{k' \le n \le k''} |\mu_n|. && (10.26) \end{aligned}$$

Proof of Proposition 10.5 is presented in Section 12.

Note that according to (10.20) and (10.25),

$$a_- = -(\zeta_{k''} - \zeta_{k'})\mathbf{F}_-, \quad a_- - a_+ = \int_{\zeta_{k'}}^{\zeta_{k''}} \big(\widehat{F}_+(0,\zeta) - \widehat{F}_-(0,\zeta)\big) d\zeta.$$

We construct the trajectory (10.18) as follows. We take in (10.13) $\widehat{v}_n = \widehat{F}_-(0, \widehat{\zeta}_n)$.

Consider a constant K satisfying Proposition 10.5. For $K \le n - s \le c\frac{\mu_n}{\sqrt{\varepsilon}|\log\varepsilon|}$

$$|\mu_s - \mu_n| = |O(K\sqrt{\varepsilon}\log\varepsilon)| < \mu_n/2. \tag{10.27}$$

The quantity $\zeta_n - \zeta_{n-K}$ is negative. Moreover, by (10.27)

$$\zeta_n - \zeta_{n-K} < \frac{1}{3}\nu'_0\sqrt{\varepsilon}\mu_n K \frac{|\log\varepsilon|}{\lambda}. \tag{10.28}$$

Therefore we have: $\chi_-(\mu_n, \zeta_n) - \chi_-(\mu_{n-K}, \zeta_{n-K}) = \mathbf{D}^-_{n-K,n}$,

$$\begin{aligned} \mathbf{D}^-_{n-K,n} &< -\frac{(\zeta_n - \zeta_{n-K})\mathbf{F}_-}{\mathbf{t}_*} + \frac{2d|\zeta_n - \zeta_{n-K}|}{\mathbf{t}_*} + 2C\sqrt{\varepsilon}\mu_n + 2CK\varepsilon^{3/4}|\log\varepsilon| \\ &< -\frac{F_- + 2d}{4}\sqrt{\varepsilon}\mu_n\nu'_0 K. && (10.29) \end{aligned}$$

provided the constant K is sufficiently big compared with C. This means that proceeding in this way, we reach a point (μ_{n_0}, ζ_{n_0}) such that

$$\mu_{n_0-1} \le -\widehat{C}/|\log\varepsilon| \le \mu_{n_0}. \tag{10.30}$$

Moreover, since on this part of the trajectory average increment of μ is of order $\sqrt{\varepsilon}$, we obtain the estimate

$$n_0 - s \le C\varepsilon^{-5/6}. \tag{10.31}$$

Construction of the trajectory for $\mu > \widehat{C}/|\log\varepsilon|$, $\chi_+(\mu,\zeta) > 0$ is analogous: for $\widehat{v} = \widehat{F}_+(0,\widehat{\zeta})$ the energy χ_+ increases.

10.3 Step 3 (at the resonance)

In this section we lead the trajecory through the cylinder

$$\mathbf{S} = \{(\mu,\zeta) : |\mu| \le \widehat{C}/|\log\varepsilon|\}.$$

Due to (10.22) and $\mathbf{H_1 3}$, the equation $\chi = 0$ determines separatrices. These separatrices are outer in the sense that all critical points of χ_- are in the domain $\{\chi_- \le 0\}$. There can be also inner separatrices, corresponding to smaller values of the energy, but below we will call by separatrices only

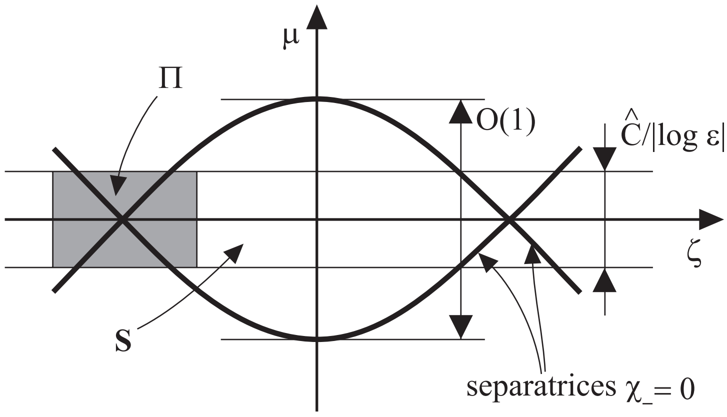

Figure 2:

the outer ones. We call the curve $\{\chi = 0, \mu < 0\}$ by the lower separatrix and $\{\chi = 0, \mu < 0\}$ by the upper one.

According to $\mathbf{H_1 3}$, for ε small enough the maximums (10.22) are taken at unique poins $\zeta_\pm = \zeta_\pm(\varepsilon) = \zeta_* + O(1/\log\varepsilon)$.

Consider the ractangle

$$\Pi = \{(\mu, \zeta \bmod 1/q^0) : |\mu| \le \widehat{C}/|\log\varepsilon|, |\zeta - \zeta_*| \le \widehat{C}_1/|\log\varepsilon|\},$$

where the constants $\widehat{C}_1$, $\widehat{C}$ are such that for small ε

(1) the energy χ_- is negative at the four points

$$(\mu, \zeta) = (\pm\widehat{C}/|\log\varepsilon|, \zeta_* \pm \widehat{C}_1/|\log\varepsilon|)$$

(see Fig. 2)

(2) $|\zeta_\pm(\varepsilon) - \zeta_*| < \widehat{C}_1/2$.

In Section 10.2 we continued a trajectory of Problem **V** up to a point $(\mu_{n_0}, \zeta_{n_0}) \in \mathbf{S}$, $\mu_{n_0} = -\widehat{C}/|\log\varepsilon| + O(\sqrt{\varepsilon}\log\varepsilon)$. There are two cases. The point (μ_{n_0}, ζ_{n_0}) can be outside and inside Π. In the first case $\chi(\mu_{n_0}, \zeta_{n_0}) < 0$. Therefore, the trajectory has crossed the lower separatrix. In the last case we can loose control on the trajectory. To avoid such a possibility, we return a little back. More precisely, starting from some point $(\mu_{n'}, \zeta_{n'})$, $n' < n_0$, we construct the trajectory as in step **2**, using $\widehat{v} = \widehat{f}_+(0, \widehat{\zeta})$ instead of $\widehat{v} = \widehat{f}_-(0, \widehat{\zeta})$.

We denote the new trajectory by

$$(\mu_{n'}, \zeta_{n'}), (\mu^+_{n'+1}, \zeta^+_{n'+1}), (\mu^+_{n'+2}, \zeta^+_{n'+2}), \dots, (\mu^+_{n''}, \zeta^+_{n''}). \qquad (10.32)$$

Now by Proposition 10.5 the energy χ on the trajectory decreases slower or does not decrease at all. We choose $(\mu_{n'}, \zeta_{n'})$ so that

$$\int_{\zeta_{n_0}}^{\zeta_{n'}} \big(\widehat{F}_+(0,\zeta) - \widehat{F}_-(0,\zeta)\big)\, d\zeta = c_0/|\log\varepsilon| + O(\sqrt{\varepsilon}\log\varepsilon).$$

Then we take the point $(\mu^+_{n''}, \zeta^+_{n''})$, $n'' > n'$ such that $\zeta^+_{n''} = \zeta_{n_0} + O(\sqrt{\varepsilon}\log\varepsilon)$. According to Proposition 10.5,

$$\chi_-(\mu^+_{n''}, \zeta^+_{n''}) - \chi_-(\mu_{n_0}, \zeta_{n_0}) = \frac{c_0 + Y_1}{\mathbf{t}_*|\log\varepsilon|},$$

where $|Y_1| < 3d$ for sufficiently small ε. Therefore,

$$\frac{1}{2}\nu_0'(\mu^+_{n''})^2 - \frac{1}{2}\nu_0'\mu^2_{n_0} + O(\sqrt{\varepsilon}\log\varepsilon) = \frac{c_0 + Y_1}{\mathbf{t}_*|\log\varepsilon|},$$

$$(\mu^+_{n''})^2 = \mu^2_{n_0} + \frac{2(c_0 + Y_2)}{\nu_0'\mathbf{t}_*|\log\varepsilon|}, \qquad |Y_2| < 4d.$$

(Recall that both $\mu^+_{n''}$ and μ_{n_0} are negative.)

For some c_0, continuation of the trajectory (10.32) with $\widehat{v}_n = \widehat{F}_-(0, \widehat{\zeta}_n)$ will cross the lower separatrix outside Π. Hence, redenoting (μ_n^+, ζ_n^+) by (μ_n, ζ_n) and n'' by n_0, we can return to our assumption that (10.30) holds and $(\mu_{n_0}, \zeta_{n_0}) \in \mathbf{S} \setminus \Pi$.

Now we just need to pass the resonance itself, crossing the upper separatrix outside Π. Consider the energy $\chi_+(\mu_{n_0}, \zeta_{n_0})$ (instead of χ_-). There are two cases.

(1) $\chi_+(\mu_{n_0}, \zeta_{n_0}) > -c_\bullet/|\log\varepsilon|$, where $c_\bullet > 0$ is sufficiently small.

In this case we pass $\mathbf{S}$ by using $\widehat{v}_n = \widehat{F}_-(0, \widehat{\zeta}_n)$. (this is possible since (μ_{n_0}, ζ_{n_0}) is outside Π). This takes $O(\varepsilon^{-1/2}|\log\varepsilon|^{-2})$ steps, and ζ changes by a quantity $O(\log^{-2}\varepsilon)$. Hence, the trajectory leaves $\mathbf{S}$ with

$$\chi_+(\mu_{n_0}, \zeta_{n_0}) > -c_\bullet/|\log\varepsilon| + O(\log^{-2}\varepsilon). \tag{10.33}$$

Now we switch to $\widehat{v}_n = \widehat{F}_+(0, \widehat{\zeta}_n)$. Because of (10.33) the trajectory crosses the upper separatrix outside Π.

(2) $\chi_+(\mu_{n_0}, \zeta_{n_0}) < -c_\bullet/|\log\varepsilon|$. In this case we return back. Using $\widehat{v}_n = \widehat{F}_+(0, \widehat{\zeta}_n)$ beginning from some $n' < n_0$, we obtain a trajectory, where χ_+ is larger than before. Choosing a proper n' and $n'' > n'$, where again $\widehat{v}_n = \widehat{F}_+(0, \widehat{\zeta}_n)$, we reduce the situation to case (1).

11 Proof of Lemma 8.4

By using Lemma 8.2, we can construct a trajectory such that on the corresponding sequence of right

compartible trajectories $(\overline{\Omega}_k, \overline{\Omega}_{k+1})$, $(\overline{\Omega}_{k+1}, \overline{\Omega}_{k+2})$, ... according to our needs $v_n = \widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n, \overline{\tau}_n)$, $n = k, k+1, \dots$ equals

either to $\widehat{F}_+^{\sigma_n}(\varepsilon\rho_{n-1}, \zeta_{n-1} + \nu(\varepsilon\rho_{n-1})\mathbf{t}_n)$ or to $\widehat{F}_-^{\sigma_n}(\varepsilon\rho_{n-1}, \zeta_{n-1} + \nu(\varepsilon\rho_{n-1})\mathbf{t}_n)$

(see (8.7)).

We need to show that the codes $(\overline{\Omega}_k, \overline{\Omega}_{k+1})$ can be chosen so that on the corresponding trajectory $\varepsilon\rho$ passes through the interval $\mathbf{I}$. To this end we reduce the problem to Problem $\mathbf{V}$.

Indeed, for $\varepsilon\overline{\rho}_n \in \mathbf{I}$, combining (4.5) and the estimates

$$|\overline{\rho}_n - \rho_n| < b/b_\rho, \quad |\overline{\zeta}_n - \zeta_n| < b/b_\zeta, \quad |\overline{\tau}_n - \tau_n| < b/b_\tau,$$

we get:

$$\begin{aligned} \overline{\rho}_{n+1} &= \overline{\rho}_n - \widehat{\Theta}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n, \overline{\tau}_n) + (\overline{\tau}_{n+1} - \overline{\tau}_n - \overline{t}_{n+1})\mathbf{H}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n) + \widetilde{\alpha}_n, \\ \overline{\zeta}_{n+1} &= \overline{\zeta}_n + \nu(\varepsilon\overline{\rho}_n)\overline{t}_{n+1} + \widetilde{\beta}_n, \end{aligned}$$

where for small $d > 0$ we have $|\widetilde{\alpha}_n| < d/2$, $|\widetilde{\beta}_n| < C\varepsilon^{3/4}|\log\varepsilon|/2$.

We denote

$$\begin{aligned} \mu_n &= \sqrt{\varepsilon}(\overline{\rho}_n - \overline{\tau}_n\mathbf{H}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n)), \\ \alpha_n &= \widetilde{\alpha}_n + \overline{\tau}_{n+1}(\mathbf{H}_\zeta(\varepsilon\overline{\rho}_n, \overline{\zeta}_n) - \mathbf{H}_\zeta(\varepsilon\overline{\rho}_{n+1}, \overline{\zeta}_{n+1})). \end{aligned}$$

Then we get (8.2). Since $\nu(\varepsilon\overline{\rho}_n) = \nu(\sqrt{\varepsilon}\mu_n) + O(\varepsilon)$, we get (8.3). ∎

12 Addendum to Section 10

Proof of Proposition 10.2. Take arbitrary $\mathbf{t}$, satisfying (10.6) (except the greatest one) and the corresponding $\Delta = \Delta(\mathbf{t})$, determined by (10.5). Take $N = [c(\varepsilon\log^2\varepsilon)^{-1/6}]$, where c is the same as in condition (1) of WRC. According to the Dirichlet theorem there exists a natural $q \le N$ and $p \in \mathbf{Z}$ such that the fraction p/q is irreducible and $|q\Delta - p| < 1/N$. If $q \ge C$, the proof is complete. Otherwise we put

$$\delta = \nu(\sqrt{\varepsilon}\mu_s) - p_0/q_0$$

and consider two cases:

$$(I) \qquad \frac{\lambda}{4q_0 C|\log\varepsilon|} < |\delta| < \frac{2\nu_0'}{q_0\overline{K}}, \qquad \nu_0' = \frac{d\nu}{dy}(0), \tag{12.1}$$

$$(II) \qquad \frac{2\nu_0'}{q_0\overline{K}} < |\delta| < \frac{1}{2}\nu_0'\varepsilon^{1/6}. \tag{12.2}$$

According to (10.1) and to the inequality $\nu_0' \neq 0$ (see $\mathbf{H_0 4}$), δ satisfies one of these conditions.

(I) In this case take $\Delta' = \Delta(\mathbf{t}')$, $\mathbf{t}' = \mathbf{t} + 1$. Again apply the Dirichlet theorem and assume that for the corresponding $p' \in \mathbf{Z}$ and $q' \in \mathbf{N}$ we have $q' < C$. Then

$$\Delta' - \Delta = \nu(\sqrt{\varepsilon}\mu_s) = \frac{p_0}{q_0} + \delta = \frac{p'}{q'} + \frac{\xi'}{q'} - \frac{p}{q} + \frac{\xi}{q},$$

where $|\xi| \le 1/N$, $|\xi'| \le 1/N$.

We have the equation

$$\Big|\frac{p_0}{q_0} - \frac{p'}{q'} + \frac{p}{q}\Big| = \Big| - \delta + \frac{\xi'}{q'} - \frac{\xi}{q}\Big|. \tag{12.3}$$

Since q', q do not exceed C, the left-hand side of (12.3) can not belong to the interval $(0, 1/(q_0 C^2))$. The right-hand side belongs to $(\delta/2, 3\delta/2) \subset (0, 1/(q_0 C^2))$ (the last inclusion follows for small ε from (12.1)). We come to a conradiction.

(II) This case is simpler. The equation

$$\Delta = (\mathbf{t} + 2q_0)\big(\frac{p_0}{q_0} + \delta\big) = \frac{p}{q} + \frac{\xi}{q}, \qquad |\xi| < 1/N$$

implies that

$$\Big|(\mathbf{t} + 2q_0)\frac{p_0}{q_0} - \frac{p}{q}\Big| = \Big|\frac{\xi}{q} - (\mathbf{t} + 2q_0)\delta\Big|. \tag{12.4}$$

The left-hand side of (12.4) can not belong to the interval $(0, 1/(q_0 C))$. For sufficiently small ε we have: $2q_0 < \mathbf{t}/2$. Hence, according to (10.6) and (12.2),

$$\frac{\nu_0'\varepsilon^{1/6}}{2}\frac{|\log\varepsilon|}{2\lambda} \le |(t + 2q_0)\delta| \le \frac{\lambda}{4q_0 C|\log\varepsilon|}\frac{2|\log\varepsilon|}{\lambda}.$$

For small ε we have: $|\xi| < |(t + 2q_0)\delta|$. Therefore, the right-hand side of (12.4) belongs to $(0, 1/(q_0 C))$. Again we obtain a contradiction.

■

Proof of Proposition 10.5. We have:

$$\mathbf{D}^{\pm}_{k',k''} = \Sigma'_{\pm} + \Sigma''_{\pm} + \Sigma'''_{\pm},$$

$$\Sigma'_{\pm} = \sum_{n=k'}^{k''-1} \mathbf{d}^-_n, \quad \Sigma''_{\pm} = -\sqrt{\varepsilon}\nu_0' \sum_{n=k'}^{k''-1} \mu_n \widehat{\alpha}_n, \quad \Sigma'''_{\pm} = \sum_{n=k'}^{k''-1} (\beta'_n + \beta''_n).$$

Proposition 10.5 follows from the estimates

$$\Sigma'_{\pm} = \frac{1}{\mathbf{t}_*}\Big(- \int_{\zeta_{k'}}^{\zeta_{k''}} \widehat{F}_{\pm}(0,\zeta)\, d\zeta + \mathcal{F}(\zeta_{k''}) - \mathcal{F}(\zeta_{k'}) + \delta'\Big),$$

$$|\delta'| < C\big(\sqrt{\varepsilon}\mu_{\max} + k\varepsilon\mu^2_{\max}|\log\varepsilon| + k\varepsilon^{3/4}|\log\varepsilon|\big), \tag{12.5}$$

$$|\Sigma''| < \frac{2d}{\mathbf{t}_*}|\zeta_{k''} - \zeta_{k'}| + \frac{C(k'' - k')}{\Delta}(\varepsilon\mu^2_{\max} + \varepsilon^{3/4})|\log\varepsilon|. \tag{12.6}$$

First, we check (12.5):

$$\Sigma'_{\pm} = - \sum_{n=k'}^{k''-1} \sqrt{\varepsilon}\nu_0'\mu_n \widehat{F}_{\pm}(0, \widehat{\zeta}_n) + \frac{1}{\mathbf{t}_*}\big(\mathcal{F}(\zeta_{k''}) - \mathcal{F}(\zeta_{k'})\big).$$

Due to (10.17) the multiplier $\sqrt{\varepsilon}\nu_0'\mu_n$
in the last sum equals $\frac{1}{\mathbf{t}_*}\big(\widehat{\zeta}_{n+1} - \widehat{\zeta}_n - \delta_n\big)$. Therefore

$$\begin{aligned}
\Sigma'_{\pm} &= \frac{1}{\mathbf{t}_*}\Big(\Sigma'_{1\pm} + \Sigma'_{2\pm} + \mathcal{F}(\zeta_{k''}) - \mathcal{F}(\zeta_{k'})\Big), \\
\Sigma'_{1\pm} &= -\sum_{n=k'}^{k''-1} \widehat{F}_{\pm}(0, \widehat{\zeta}_n)\big(\widehat{\zeta}_{n+1} - \widehat{\zeta}_n\big), \\
\Sigma'_{2\pm} &= \sum_{n=k'}^{k''-1} \widehat{F}_{\pm}(0, \widehat{\zeta}_n)\delta_n.
\end{aligned}$$

The functions $\widehat{F}_{\pm}$ are piecewise smooth. Hence by (10.16) we have:

$$\Big|\Sigma'_{1\pm} + \int_{\zeta_{k'}}^{\zeta_{k''}} \widehat{F}_{\pm}(0,\zeta)\, d\zeta\Big| < C\sqrt{\varepsilon}\mu_{\max}.$$

The sum $\Sigma'_{2\pm}$ can be estimated as follows:

$$|\Sigma'_{2\pm}| \le C_1 \sum_{n=k'}^{k''} |\delta_n| \le C(k'' - k')(\mu_{\max}^2\varepsilon + \varepsilon^{3/4})|\log\varepsilon|.$$

This implies (12.5).

To check (12.6), we note that $|\Sigma''| \le \frac{2d}{\mathbf{t}_*}|\Sigma_{n=k'}^{k''-1}(\widehat{\zeta}_{n+1} - \widehat{\zeta}_n - \delta_n)|$. Hence,

$$|\Sigma''| \le \frac{2d}{\mathbf{t}_*}|\widehat{\zeta}_{k''} - \widehat{\zeta}_{k'}| + \frac{2d}{\mathbf{t}_*}\Sigma_{n=k'}^{k''-1}|\delta_n|.$$

The last sum can be estimated analogously to $\Sigma'_{2\pm}$. ∎

13 Genericity of the hypotheses

Proof of Proposition 5.1. First, let us fix $\eta = \eta^0$ and put $\Theta = \Theta(\eta^0, \xi, \tau)$. The function $\Theta : \mathbf{T}^2 \to \mathbf{R}$ is C^2-smooth. Let (ξ^0, τ^0) be a nondegenerate critical point of Θ. We put

$$A = \frac{\partial^2\Theta}{\partial(\xi,\tau)^2}(\xi^0, \tau^0), \qquad \det A \ne 0. \tag{13.1}$$

Since $\overline{\partial}\Theta(\xi^0, \tau^0) = 0$ and $\overline{\partial}^2\Theta(\xi^0, \tau^0) = \frac{1}{\lambda^2}\langle A\overline{\nu}, \overline{\nu}\rangle$, we see that $(\eta^0, \xi^0, \tau^0) \in J_0$ provided

$$\langle A\overline{\nu}, \overline{\nu}\rangle \ne 0. \tag{13.2}$$

Moreover, (13.2) implies that in a neighborhood of the point (ξ^0, τ^0) the set $j_0 = J_0|_{\eta=\eta^0}$ is a smooth curve. The normal vector to j_0 at (ξ^0, τ^0) equals $A\overline{\nu}$.

On the other hand since

$$\Theta_\xi(\xi^0, \tau^0) = 0 \quad \text{and} \quad \Theta_{\xi\xi}(\xi^0, \tau^0) = \frac{1}{\lambda^2}\langle A\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\rangle,$$

the set $S = \{(\xi, \tau) \in \mathbf{T}^2 : \Theta_\xi(\xi, \tau) = 0\}$
near (ξ^0, τ^0) is a smooth curve provided

$$\langle A\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\rangle \ne 0. \tag{13.3}$$

Moreover, (13.3) implies that locally near (ξ_0, τ_0) the curve S separates the sets $\{\Theta_\xi > 0\}$ and $\{\Theta_\xi < 0\}$. The normal vector to S at (ξ^0, τ^0) equals $A\begin{pmatrix}1\\0\end{pmatrix}$.

Since $\det A \ne 0$, the two normals are not parallel: $A\overline{\nu} \not\parallel A\begin{pmatrix}1\\0\end{pmatrix}$. Hence, the function $\Theta_\xi|_{j_0}$ is positive locally on one side with respect to (ξ_0, τ_0) and negative on another.

According to (13.2), $\overline{\nu}$ and the tangent vector to j_0 at the point (ξ^0, τ^0) are not parallel. Therefore, for $\eta = \eta^0$ and some $c_* > 0$ $\mathbf{H_1 1}$ holds.

The set of C^j-smooth functions $\Theta : \mathbf{T}^2 \to \mathbf{R}$, satisfying (13.1)–(13.3) at at least one critical point, is obviously open and dense.

To finish the proof, it remains to recall that any function on $\mathbf{T}^2$ has at least 3 critical points and to use compactness of $\overline{D}$. ■

Proof of Proposition 5.4. Given a C^j-smooth, $2 \le j \le \infty$ function $\vartheta : \mathbf{T}^2 \to \mathbf{R}$ we put

$$f_+(\varphi_1) = \max_{\varphi_2}\{\vartheta_{\varphi_1} : \vartheta_{\varphi_2} = 0\}, \quad f_-(\varphi_1) = \min_{\varphi_2}\{\vartheta_{\varphi_1} : \vartheta_{\varphi_2} = 0\}.$$

Proposition 13.1 *For C^j-generic ϑ*

(1) $\pm\int_0^1 f_\pm(\varphi_1)\, d\varphi_1 > 0$.

(2) $f_+ = f_-$ at most at a finite number of points $x_1, \dots, x_k \in \mathbf{T}$, where for sufficiently small $|\delta|$ the following transversality condition holds:

$$f_+(x_j + \delta) - f_-(x_j + \delta) > 2c_f|\delta|, \qquad c_f > 0, \quad j = 1, \dots, k. \tag{13.4}$$

(3) $f_\pm(\varphi_1) = \vartheta_{\varphi_1}(\varphi_1, q_\pm(\varphi_1))$, where for some $c'_f > 0$

$$\vartheta_{\varphi_2}(\varphi_1, q_\pm(\varphi_1)) = 0, \quad |\vartheta_{\varphi_2\varphi_2}(\varphi_1, q_\pm(\varphi_1))| > 2c'_f.$$

(4) The functions $q_\pm$ and $f_\pm$ are piecewise C^{j-1}-smooth in φ_1.

The genericity is understood here in the sense that the set of ϑ, satisfying (1)–(4), contains a subset which is open and dense in the C^j topology.

Proposition 13.1 implies Proposition 5.4. To see this, it is sufficient to put $\vartheta(\zeta, s) = \Theta(\eta^0, \zeta - \nu_0 q_0 s, q_0 s)$.

Proof of Proposition 13.1. First, note that for generic ϑ the set $\Lambda = \{\varphi \in \mathbf{T}^2 : \vartheta_{\varphi_2} = 0\}$ contains only a finite number of points $\varphi^{1*}, \dots, \varphi^{l*}$, $\varphi^{j*} = (\varphi_1^{j*}, \varphi_2^{j*})$, where $\vartheta_{\varphi_2\varphi_2} = 0$. Moreover, for generic ϑ

$$\vartheta_{\varphi_2\varphi_2\varphi_2}(\varphi) \ne 0 \quad \text{for any} \varphi \in \{\varphi^{1*}, \dots, \varphi^{l*}\}. \tag{13.5}$$

We define the projection

$$\mathrm{pr} : \Lambda \to \mathbf{T}^1, \qquad \mathrm{pr}\,(\varphi_1, \varphi_2) = \varphi_1.$$

Then $\varphi^{1*}, \dots, \varphi^{l*}$ are folding points for pr.

Remark 13.1 *Due to (13.5) for any folding point $(\varphi_1^0, \varphi_2^0) \in \Lambda$ the function $g(\alpha) = \vartheta(\varphi_1^0, \alpha)$ has no extremum at $\alpha = \varphi_2^0$. Hence, for any $\varphi_1 \in \mathbf{T}$ the preimage $pr^{-1}(\varphi_1)$ consists of at least 2 points different from $\varphi^{1*}, \dots, \varphi^{l*}$ (maximum and minimum of $\tilde{g}(\varphi_2) = \vartheta(\varphi_1, \varphi_2)$).*

The set $\Lambda^* = \Lambda \setminus \{\varphi^{1*}, \dots, \varphi^{l*}\}$ breaks into a finite number of connected components Λ_m, $1 \le m \le M$ which are graphs of C^{j-1}-smooth functions $\varphi_2 = \varphi_2(\varphi_1)$.

For any connected piece $\lambda = \Lambda_j \subset \Lambda^*$ we define its orientation to be compartible with positive orientation on $\mathrm{pr}(\lambda) \subset \mathbf{T}^1$. Let λ begin at φ' and finish at φ'' i.e., the motion along λ from φ' to φ'' takes place in the positive direction. Then

$$\vartheta(\varphi'') - \vartheta(\varphi') = \int_\lambda d\vartheta = \int_\lambda \vartheta_{\varphi_1}\, d\varphi_1 \le \int_{\mathrm{pr}(\lambda)} f_+(\varphi_1)\, d\varphi_1. \tag{13.6}$$

Take $\varphi_1^{(0)} \in \mathbf{T}^1$ such that

$$g^0(\alpha) = \vartheta(\varphi_1^{(0)}, \alpha), \qquad \alpha \in \mathbf{T}^1.$$

has a unique global minimum at some point $\alpha = \varphi_2^{(0)}$. We put $\varphi^{(0)} = (\varphi_1^{(0)}, \varphi_2^{(0)}) \in \Lambda_{m_0}$. Consider the longest connected piece $\lambda^{(0)} \subset \Lambda^*$, beginning at $\varphi^{(0)}$.

Now there are two cases:

(a) $\lambda^{(0)}$ has an end point $\psi^{(0)} \in \{\varphi^{1*}, \dots, \varphi^{l*}\}$,

(b) $\lambda^{(0)}$ is a circle.

Consider these possibilities separately.

(a). If $\mathrm{pr}(\lambda^{(0)}) = \mathbf{T}^1$, we stop. Otherwise consider $\varphi^{(1)} \subset \Lambda^*$ such that $\varphi_1^{(1)} = \psi_1^{(0)}$ and $\vartheta(\varphi_1^{(1)}, \varphi_2^{(1)}) = \min_{\varphi_2} \vartheta(\varphi_1^{(1)}, \varphi_2)$. According to Remark 13.1, $\vartheta(\varphi^{(1)}) < \vartheta(\psi^{(0)})$ and $\varphi^{(1)} \neq \psi^{(0)}$.

We regard $\varphi^{(1)}$ as the beginning point for the curve $\lambda^{(1)} \subset \Lambda^*$ which is again the longest connected piece of Λ^*. If $\mathrm{pr}(\lambda^{(0)}) \cup \mathrm{pr}(\lambda^{(1)}) \neq \mathbf{T}^1$, we continue in the same way. At last we get a finite collection of curves $\lambda^{(n)}$, $n = 0, \dots, N$ beginning at $\varphi^{(n)}$ and ending at $\psi^{(n)}$. The curves project one-to-one on $\mathbf{T}^1$ and $\cup_{n=0}^N \lambda^{(n)} = \mathbf{T}^1$.

Instead of the last curve $\lambda^{(N)}$ we take a shorter piece, $\hat{\lambda}^{(N)} \subset \lambda^{(N)}$, with the same beginning having the end point $\widehat{\psi}^{(N)}$, where $\widehat{\psi}_1^{(N)} = \varphi_1^{(0)}$. Below we skip the hats. According to the definition of $\varphi^{(0)}$ (more precisely, to the minimality property of $\varphi^{(0)}$), $\vartheta(\varphi^{(0)}) \leq \vartheta(\psi^{(N)})$. Moreover,

$$\vartheta(\varphi^{(n+1)}) < \vartheta(\psi^{(n)}) \quad \text{for any } n \in \{0, N-1\}. \tag{13.7}$$

By using (13.6), we get:

$$\int_0^1 f_+(\varphi_1)\, d\varphi_1 = \sum_{n=0}^N \int_{\mathrm{pr}(\lambda^{(n)})} f_+(\varphi_1)\, d\varphi_1 \geq \sum_{n=0}^N \vartheta(\psi^{(n)}) - \sum_{n=0}^N \vartheta(\varphi^{(n)}).$$

The last sum is positive if $N \geq 1$ because of (13.7). If $N = 0$, it is also positive because in this case $\varphi^{(0)} \neq \psi^{(N)}$ and therefore, $\vartheta(\psi^{(N)}) - \vartheta(\varphi^{(0)}) > 0$.

(b) First, note that in the arguments from part (a) we could assume that $\varphi^{(j)}$ are maximums instead of minimums and construct the curves $\lambda^{(n)}$ to the negative direction from the points $\varphi^{(n)}$. This means that the proof we presented in part (a) does not work only in the case when Λ^* contains at least 2 smooth curves projecting one-to-one onto $\mathbf{T}^1 = \{\varphi_1\}$. Let these curves be $\varphi_2 = q_n(\varphi_1)$, $n = 1, 2$. Since $f_+(\varphi_1) \geq \vartheta(\varphi_1, q_n(\varphi_1))$ for any $n \in \{1, 2\}$ and

$$\int_0^1 f_+(\varphi_1)\, d\varphi_1 \geq \int_0^1 \vartheta(\varphi_1, q_n(\varphi_1))\, d\varphi_1 = \vartheta(1, q_n(1)) - \vartheta(0, q_n(0)) = 0,$$

it is sufficient to show that generically $f_+(\varphi_1) \neq \vartheta(\varphi_1, q_n(\varphi_1))$ for at least one $n \in \{1, 2\}$. The contrary would imply that $\vartheta(\varphi_1, q_1(\varphi_1)) = \vartheta(\varphi_1, q_2(\varphi_1))$, and this does not hold for generic ϑ.

Now turn to assertion 2. The set of functions, where this condition holds, is obviously open. Let us show that it is dense. Generically the equation $f_+(\varphi_1) = f_-(\varphi_1)$ has only a finite number of solutions.

Let $x \in \mathbf{T}$ be one of them. Since $\mathrm{pr}^{-1}(x)$ contains at least two points $\varphi', \varphi'' \in \Lambda^*$, we have: $\vartheta_{\varphi_1}(\varphi') = \vartheta_{\varphi_1}(\varphi'')$. Let $\varphi' \in \lambda' \subset \Lambda^*$, $\varphi'' \in \lambda'' \subset \Lambda^*$ and the curves λ', λ'' have the form $\varphi_2 = f'(\varphi_1)$, $\varphi_2 = f''(\varphi_1)$. If condition (13.4) is violated, we have:

$$\frac{d}{d\varphi_1}\Big|_{\varphi_1 = x} \big[\vartheta_{\varphi_1}(\varphi_1, \lambda'(\varphi_1)) - \vartheta_{\varphi_1}(\varphi_1, \lambda''(\varphi_1))\big] = 0. \tag{13.8}$$

Adding to ϑ a small perturbation, we can either remove the solution of the equation $\vartheta_{\varphi_1}(\varphi_1, \lambda'(\varphi_1)) = \vartheta_{\varphi_1}(\varphi_1, \lambda''(\varphi_1))$ from a neighborhood of the point x or make this solution non-degenerate in the sense that (13.8) does not hold. Then the equation $f_+(\varphi_1) = f_-(\varphi_1)$ will have at most one solution x_j in a neighborhood of x and for small δ (13.4) holds. Doing the same near any solution of the equation $f_+(\varphi_1) = f_-(\varphi_1)$, we obtain ϑ satisfying assertion (2).

Assertions (3) and (4) follow from the fact that the graphs of the functions $q_\pm(\varphi_1)$ are subsets of Λ. ■

14 Auxiliary statements

Proof of Proposition 3.1. The difference $\mathbf{H}(y^0, x+p/q) - \mathbf{H}(y^0, x)$ equals

$$\sum_{(k,k_0)\in\mathbf{Z}^2} \phi\Big(\frac{kp/q + k_0 + k\delta}{\varepsilon^{1/4}}\Big) H_1^{k,k_0}(y^0) e^{2\pi i k x}(e^{2\pi i k p/q} - 1).$$

Consider nonzero terms in this sum. First, note that

$$|kp/q + k_0 + k\delta| < \varepsilon^{1/4}, \tag{14.1}$$

otherwise $\phi = 0$. Secondly, $kp/q \notin \mathbf{Z}$, otherwise $e^{2\pi i k p/q} - 1 = 0$. Therefore, we have: $|kp/q + k_0| \geq 1/q > 2\varepsilon^{1/4}$. Hence by (14.1) $|k\delta| > 1/(2q)$ i.e., $|k| > 1/(2q\delta)$. Now inequality (3.4) follows from the standard estimate $|H_1^{k,k_0}| < C_1(|k| + |k_0|)^{-j}$. ∎

Proof of Proposition 5.2. According to the Dirichlet theorem there exists $q \leq K - 1$, $p \in \mathbf{Z}$ such that the fraction p/q is irreducible and $|q\nu - p| \leq \frac{1}{K}$ i.e., $\nu = \frac{p}{q} + \xi$, $|\xi| < \frac{1}{qK}$.

Take $I \subset \mathrm{pr}_{\overline{\nu}} j_+^\sigma(\eta)$. Suppose that $q > 2/|I|$. Then Proposition 5.2 will be proven if

$$\cup_{0\leq t\leq K}\, g^{\eta,t}(I) = \mathbf{T}^2. \tag{14.2}$$

Equation (14.2) is a consequence of the following one:

$$\cup_{n=0}^{K-1} g^{\eta,n}(I) = \mathbf{T}_0 \equiv \mathbf{T}^2 \cap \{\tau = 0\}. \tag{14.3}$$

Equation (14.3) is equivalent to the following one:

$$\mathbf{T} = \big(\cup_{n=0}^{K-1}\,[n\nu, |I| + n\nu]\big) \bmod 1.$$

We replace the last set by a smaller one:

$$U = \Big(\cup_{n=0}^{q-1}\Big[\frac{np}{q} + n\xi, |I| + \frac{np}{q} + n\xi\Big]\Big) \bmod 1.$$

If $\xi = 0$, U covers "at least twice" the torus $\mathbf{T}$. This follows from the assumption $|I| > 2/q$. Since $|n\xi| < 1/K < 1/q$ for $0 < n < q-1$, U still coincides with $\mathbf{T}_0$. ∎

References

[1] Arnold V.I., Instability of dynamical systems with several degrees of freedom. Sov. Math. Dokl. **5**, 581–585, 1964.

[2] Aubry S., Abramovici G., Chaotic trajectories in the standard map: the concept of anti-integrability. Physica 43 D, (1990) 199–219.

[3] Aubry S., The concept of anti-integrability: Definition, theorems and applications to the standard map. *Twist mappings and their applications, Ed R.McGehee, K.R.Meyer*, P. 7–54, 1992.

[4] Bessi U., An approach to Arnold's diffusion through the calculus of variations. *Nonlin. Anal. TMA* **20**, 1303–1318 (1996)

[5] Berti M, Bolle P, A functional analysis approach to Arnold Diffusion, Annales Institute Henri Poincare', analyse non-lineaire, 19, 4, 2002, pp. 395-450.

[6] Berti B., Biasco L., Bolle P., Optimal stability and instability results for a class of nearly integrable Hamiltonian systems, Rend. Mat. Acc. Naz. Lincei, s. 9, vol 13, 2002.

[7] Berti M., Biasco L., and Bolle P., Drift in phase space: A new variational mechanism with optimal diffusion time. Journal de mathematiques pures et appliquees, 82/6 pp. 613-664, 2003.

[8] Bolotin S., MacKay R., Multibump orbits near the anti-integrable limit for Lagrangian systems. Nonlinearity, **10**, (1997) 1015–1029.

[9] Bolotin S., Infinite number of homoclinic orbits to hyperbolic invariant tori of Hamiltonian systems. Regular and Chaotic Dynamics (2000).

[10] Bolotin S. and Treschev D., Unbounded grouth of energy in nonautonomus Hamiltonian systems. Nonlinearity **12** (2), 365–388 (1999).

[11] Bolotin S. and Treschev D., Remarks on definition of hyperbolic tori of Hamiltonian systems. Regular and Chaotic Dynamics, V. 5, No. 4, 2000, 401-412.

[12] Chierchia L. and Gallavotti G., Drift and diffusion in phase space. Ann. Inst. Henri Poincaré, **60** (1), 1994 1–144.

[13] Chirikov B. V., A universal instability of many-dimensional oscillation systems. Phys. Rep., **52**, No. 5, (1979) 263–379.

[14] Cresson J., Guillet C., Periodic orbits and Arnold diffusion. Preprint 2001.

[15] Delshams A., de la Llave R. and Seara T., A geometric approach to the existence of orbits with unbounded energy in generic periodic perturbations by a potential of generic geodesic flows of $\mathbf{T}^2$. Commun. Math. Phys. 2000, **209**, 353–392.

[16] Delshams A., de la Llave R. and Seara T., A geometric mechanism for diffusion in Hamiltonian systems. Preprint 2001 http://www.ma.utexas.edu/mp_arc/e/03-182.ps

[17] Delshams A., Gutiérrez, Splitting potential and Poincaré-Melnikov method for whiskered tori in Hamiltonian systems. to appear in J.Nonlinear Sci.

[18] Douady R., Stabilité ou instabilité des points fixes elliptiques, *Ann. Sci. Ec. Norm. Super.* IV Ser. **21**, 1–46 (1988)

[19] Easton R.W., Meiss J.D., Roberts G., Drift by coupling to an anti-integrable limit. Nonlinearity 2000.

[20] Fontich E. and Martin P. Arnold diffusion in perturbations of analytic integrable Hamiltonian systems

[21] Gallavotti G., Gentile G., Mastropietro V. Hamilton-Jacobi equation and existence of heteroclinic chains in three time scales systems. Nonlinearity 13 (2000) 323-340

[22] Graff S. M., On the conservation of hyperbolic tori for Hamiltonian systems. J. Differ. Equat., **15**, No. 1, (1974) 1–69.

[23] Kaloshin V. Geometric proofs of Mather's connecting and accelerating theorems Katsiveli, Proceedings 2000 Cambridge University Press.

[24] MacKay R. S., Meiss J. D., Cantori for symplectic maps near the anti-integrable limit. Nonlinearity **5**, (1992) 1–12.

[25] Marco J.-P. and Sausin D. Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems. Institut de Mé canique Cé leste et de Calcul des É phé mé rides (2002) Preprint No 0201.

[26] Mather J., Talk given at the conference in dynamical systems, Oberwolfach (1997)

[27] Mather J., Lectures given at School and Workshop on Dynamical Systems ICTP, August (2001)

[28] Moser J., The analytical invariants of an area-preserving mapping near hyperbolic point. Comm. Math. Phys., **9**, No. 4, (1956) 673–692.

[29] Nekhoroshev N., An exponential estimate of the time of stability in Hamiltonian systems close to integrable, *Uspekhi Mat. Nauk*, 1977, V. 32, No. 6, P. 5–66 (in Russian). English transl. in *Russian Math. Surveys* 1977, V. 32, No. 6, P. 1–65.

[30] Poincaré H. *Les métodes nouvelles de la mécanique céleste.* V. 1–3. Paris: Gauthier–Villars, 1892, 1893, 1899.

[31] Simó C. and Valls C., A formal approximation of the splitting of separatrices in the classical Arnold's example of diffusion with two equal parameters, Nonlinearity **14** (2001) 1707–1760.

[32] Treschev D. Width of stochastic layers in near-integrable two-dimensional symplectic maps. Phys. D 116, no. 1-2, (1998) 21–43.

[33] Treschev D. Introduction to the perturbation theory of Hamiltonian systems. Phasis: Moscow (in Russian).

[34] Treschev D. Multidimensional symplectic separatrix maps. J. Nonlin. Sci. V. 12, No 1, (2002) 27–58. (see also www-ma1.upc.es/intas/preprints)

[35] Treschev D. Trajectories in a neighborhood of asymptotic surfaces of a priori unstable Hamiltonian systems. Nonlinearity Vol. 15 (2002) 2033–2052. (see also www-ma1.upc.es/intas/preprints/preprints.html)

[36] Xia Z. Arnold Diffusion: a Variational Construction. Documenta Mathematica, Extra Volume ICM, II, (1998), 867–877.

[37] Xia, J. Arnold Diffusion and Instabilities in Hamiltonian dynamics, preprint, 34pp. www.math.neu.edu/ xia/preprint/arndiff.ps.

[38] Zaslavsky G. M. and Filonenko N. N., Stochastic instability of trapped particles and the conditions of applicability of the quasi-linear approximation, Zh. Eksp. Teor. Fiz. **54** (1968), 1590–1602, (in Russian)

[39] Zehnder E., Generalized implicit function theorem with applications to some small divisor problems I,II. Comm. Pure Appl. Math., **28**, No. 1, (1975) 91–140; **29**, No. 1, (1976) 49–111.