

REPULSION FROM RESONANCES.

DMITRY DOLGOPYAT

ABSTRACT. We consider slow-fast systems with periodic fast motion and integrable slow motion in the presence of both weak and strong resonances. Assuming that the initial phases are random and that appropriate non-degeneracy assumptions are satisfied we prove that the effective evolution of the adiabatic invariants is given by a Markov process. This Markov process consists of the motion along the trajectories of a vector field with occasional jumps. The generator of the limiting process is computed from the dynamics of the system near strong resonances.

CONTENTS

| | |
|---|----|
| 1. Introduction. | 3 |
| 2. The Problem. | 5 |
| 3. Results. | 10 |
| 3.1. Dynamics before capture. | 10 |
| 3.2. Life after capture. | 12 |
| 3.3. Formal solution of the averaged equation. | 14 |
| 4. Examples. | 16 |
| 4.1. Motion in rapidly oscillating force field. | 16 |
| 4.2. Motion on narrow cylinder in the presence of magnetic field. | 17 |
| 5. Idea of the proof. | 18 |
| 6. Plan of the proof. | 23 |
| 6.1. One passage. | 23 |
| 6.2. Standard pairs. | 26 |
| 6.3. Short time evolution. | 28 |
| 7. Convergence. | 29 |
| 8. Short time dynamics. | 30 |
| 9. Free returns. | 34 |
| 9.1. Definitions. | 34 |
| 9.2. Free orbits form standard pairs. | 35 |
| 9.3. Measure estimates. | 38 |
| 10. Equidistribution. | 42 |
| 10.1. Equidistribution on a unit scale. | 42 |

| | | |
|-------|--|-----|
| 10.2. | Isolated returns. | 42 |
| 10.3. | Getting stuck at isolated returns. | 43 |
| 10.4. | Proof of Proposition 6.5(c) for returns of the second kind. | 44 |
| 10.5. | Isolated returns are rare. | 44 |
| 10.6. | Passages near saddle points. | 48 |
| 11. | Theorem 2. | 48 |
| 11.1. | Preliminaries. | 48 |
| 11.2. | Proof of Theorem 2. | 50 |
| 12. | Open problems. | 52 |
| 12.1. | Weakening of conditions (A)-(K) | 53 |
| 12.2. | Separatrix crossings. | 54 |
| 12.3. | Systems without strong resonances. | 55 |
| 12.4. | Non integrable averaged motion. | 57 |
| 12.5. | Higher dimensions. | 57 |
| | Appendix A. Asymptotics of the Poincare map. | 59 |
| A.1. | Size of the jump. | 59 |
| A.2. | Passage time. | 60 |
| A.3. | Proof of Lemma 6.3. | 60 |
| | Appendix B. Derivatives of the Poincare map. Outline of the proof | 61 |
| | Appendix C. Derivatives of the inner map. | 64 |
| C.1. | Some classes of maps invariant under the compositions. | 64 |
| C.2. | Step by step analysis of the inner map. | 66 |
| | Appendix D. Derivatives of the outer map. | 73 |
| | Appendix E. Dynamics near the separatrix. | 80 |
| E.1. | Normal form. | 80 |
| E.2. | Some consequences of volume preservation. | 82 |
| E.3. | Analysis of passages near the origin. | 86 |
| E.4. | Derivative bounds of Proposition 6.2. | 91 |
| E.5. | Measure bounds of Proposition 6.2. | 92 |
| | Appendix F. Captured points. | 95 |
| F.1. | Dividing the trajectory | 95 |
| F.2. | Entrance phase. | 97 |
| F.3. | Middle phase. | 99 |
| F.4. | Estimates of the derivatives. | 101 |
| F.5. | Proof of Proposition 6.2* | 106 |
| | Appendix G. Examples. | 106 |
| G.1. | Example 1. | 107 |
| G.2. | Example 2. | 108 |
| | Appendix H. Distortion bound. | 110 |
| | References | 111 |

1. INTRODUCTION.

Averaging method is one of the most classical and most effective tools in dynamics. The basic idea is very simple. Consider a two scale system

$$\begin{aligned} \dot{y} &= Y(x, y, \varepsilon) \\ \dot{x} &= \frac{X(x, y, \varepsilon)}{\varepsilon}, \text{ where } \varepsilon \ll 1. \end{aligned}$$

If we are interested in the evolution of the slow variables y the direct computations are costly since y changes at times $\mathcal{O}(1)$ while any numerical procedure should have step $o(\varepsilon)$ which is the natural time scale for the change of fast variables. In fact, the instabilities of the fast system can make computations unreliable. Therefore it is natural to approximate y by the solution of the effective equation

$$(1.1) \quad \dot{\bar{y}} = \bar{Y}(\bar{y}) \text{ where } \bar{Y}(\bar{y}) = \langle Y(\bar{y}, x, 0) \rangle$$

and $\langle \dots \rangle$ denotes the averaging with respect to an invariant measure for the frozen system

$$\dot{x} = X(x, \bar{y}, 0)$$

(finding the correct measure for $\langle \dots \rangle$ is part of the problem).

While the averaging method itself was invented about quarter of the millennium ago motivated by the needs of the Celestial Mechanics (see [58] for a historical survey) the work on its rigorous justification started much later. The first results were limited to the case where the fast motion is periodic [20, 38, 10, 36] or more generally uniquely ergodic [9]. In the uniquely ergodic case there is only one invariant measure so the meaning of $\langle \dots \rangle$ in (1.1) is clear. However the uniquely ergodic setting is insufficient even for describing small perturbation of nearly integrable systems (because in this case the unperturbed system contains resonant tori which possess many invariant measures). The justification of averaging method in the general case is more subtle since in that case the actual trajectory is close to the averaged one not everywhere but only on a large measure set of initial conditions. The work on the justification of the averaging method in the general setting was undertaken in the second half of 20th century mostly by the Soviet School ([2, 3, 4, 21, 32, 33, 35, 39, 41, 42]). By now the averaging principle is justified under quite general conditions (see e.g. [5, 24, 34, 37]).

In the case the fast motion is chaotic (for example an Anosov system or a Markov process) one can also obtain the limiting distribution for

the difference between the actual and the averaged trajectory (see review [34]). Such estimates are not yet available in the classical setting of the quasiperiodic fast motion. One situation where such deviations are important is when the averaged system has first integrals (which are called *adiabatic invariants* for the original system). The change of the adiabatic invariants occurs only due to deviations from the averaged motion. It is well known that in the quasiperiodic case the main source of deviations from the averaged motion happens due to passages through resonances. The contribution of the individual resonance has been computed by several authors [14, 26, 27, 40, 41, 48]. There are many examples of the systems where multiple resonance passages can lead to destruction of adiabatic invariants on the appropriate time scale (see the above cited papers as well as [30, 31, 43, 47, 57, 59, 60, 61, 62]). Since the quasiperiodic case remains the prevalent source of application of the averaging techniques the development of the statistical theory of adiabatic invariants is one of the most important problems in the averaging theory. The goal of the present paper is to make a step in this direction by considering the simplest case of periodic fast motion.

The basic idea is that the passage through resonances makes the dynamics hyperbolic on most of the phase space [47]. In fact the hyperbolicity is created due to the combination of strong shearing away from resonances with the destruction of the shear-invariant foliation near the resonances. (For the general discussion of the sheared induced stochasticity we refer the reader to [1, 28, 56, 63].) Therefore the methods developed to treat hyperbolic fast motion ([7, 13, 17, 18, 19]) can be applied. The difference between our approach and the other papers on quasiperiodic averaging is that rather than computing C^0 -norm of the deviation with a very high precision we get more coarse information about C^2 -norms and exploit the properties which are shared by our system and its C^2 -small perturbations. Unfortunately this shift of the point of view leads to the increased size of the paper. Indeed the C^2 -estimates required for our method were not readily available in the literature (even though their derivation proceeds similarly to the C^0 -bounds). For completeness we provide the required estimates in the appendices.

We hope that this new point of view can be useful in the general quasiperiodic case. However new ideas will be required to handle the overwhelming growth of complexity coming from the fact that in the quasiperiodic case there are infinitely many resonances.

2. THE PROBLEM.

Consider the simplest three scale system.

$$(2.1) \quad \begin{aligned} \dot{I} &= \alpha_1(I, \phi, \theta) + \varepsilon \alpha_2(I, \phi, \theta, \varepsilon) \\ \dot{\phi} &= p(I) + \beta_1(I, \phi, \theta) + \varepsilon \beta_2(I, \phi, \theta, \varepsilon) \\ \dot{\theta} &= \frac{1}{\varepsilon} \omega(I, \phi) + \eta(I, \phi, \theta, \varepsilon) \end{aligned}$$

where α_1 and β_1 satisfy

$$(2.2) \quad \int_0^1 \alpha_1(I, \phi, \theta) d\theta = \int_0^1 \beta_1(I, \phi, \theta) d\theta = 0.$$

Here I varies over an interval $[I_1, I_2]$ and ϕ and θ vary over the circle \mathbb{R}/\mathbb{Z} .

The averaging principle guarantees that away from resonant surfaces $\{\omega = 0\}$ the effective dynamics of slow variables is given by the averaged equation

$$(2.3) \quad \begin{aligned} \dot{\bar{I}} &= 0 \\ \dot{\bar{\phi}} &= p(\bar{I}) \end{aligned}$$

In particular I is an adiabatic invariant of (2.1). We are interested in evolution of this invariant on a longer time scale. Before formulating our results let us review known facts (see [5], [37]).

The case where $\omega \neq 0$ is quite well understood. Namely, we can introduce an improved invariant

$$(2.4) \quad J = I - \frac{\varepsilon}{\omega} A_1$$

where $\frac{\partial A_1}{\partial \theta} = \alpha_1$. Then $\dot{J} = \mathcal{O}(\varepsilon)$. So if we are interested in the dynamics of I on a time scale shorter than $\frac{1}{\varepsilon}$ then all changes happen in a small neighborhood of resonances $\omega = 0$. Let us study the dynamics near the resonances more closely. If the resonance is non-degenerate in the sense that $\frac{\partial \omega}{\partial \phi} \neq 0$ then it is convenient to make the following change of variables

$$\tau = \frac{t}{\sqrt{\varepsilon}}, \quad r = \frac{\omega}{\sqrt{\varepsilon}}.$$

Then (2.1) takes the following form.

$$(2.5) \quad \begin{aligned} \theta' &= r + \sqrt{\varepsilon} \bar{\eta}(I, \sqrt{\varepsilon} r, \theta, \varepsilon) \\ r' &= a(I) p(I) + g(I, \theta) + \sqrt{\varepsilon} r \bar{\beta}(I, \sqrt{\varepsilon} r, \theta, \varepsilon) + \varepsilon \hat{\beta}(I, \theta, \varepsilon) \\ I' &= \sqrt{\varepsilon} \bar{\alpha}(I, \sqrt{\varepsilon} r, \theta, \varepsilon) \end{aligned}$$

Here we let $(I, \phi(I))$ be the parametrization of the resonance curve,

$$(2.6) \quad a(I) = \frac{\partial \omega}{\partial \phi}(I, \phi(I)),$$

$$(2.7) \quad g(I, \theta) = \frac{\partial \omega}{\partial \phi}(I, \phi(I))\beta_1(I, \phi(I), \theta, 0) + \frac{\partial \omega}{\partial I}(I, \phi(I))\alpha_1(I, \phi(I), \theta, 0)$$

and $\bar{\beta}, \hat{\beta}$ represent the corrections to the main term coming from the fact that in the RHS of (2.5) $\phi \neq \phi(I)$ and $\varepsilon \neq 0$ respectively. Note that

$$\int_0^1 g(I, \theta) d\theta = 0.$$

(2.5) has a limit at $\varepsilon = 0$ where I does not move and the dynamics of the other variables is given by

$$(2.8) \quad \begin{aligned} \theta' &= r \\ r' &= L(I) + g(I, \theta) \end{aligned}$$

where $L(I) = a(I)p(I)$. (2.8) is Hamiltonian with the Hamiltonian function

$$(2.9) \quad H_I(\theta, r) = \frac{r^2}{2} - L(I)\theta - G_I(\theta)$$

where

$$(2.10) \quad \begin{aligned} G_I(\theta) &= \int_0^\theta g(I, s) ds = \\ &= a(I) \int_0^\theta \beta_1(I, \phi(I), s) ds + \frac{\partial \omega}{\partial I}(I, \phi(I)) \int_0^\theta \alpha_1(I, \phi(I), s) ds. \end{aligned}$$

It is convenient to introduce a variable

$$(2.11) \quad E = \frac{H_I(\theta, r)}{L(I)}$$

which allows us to consider the dynamics of (2.8) on the cylinder by identifying points whose E values differ by an integer.

Observe that in (2.5)

$$(2.12) \quad \bar{\alpha}(I, \sqrt{\varepsilon}r, \theta, 0) = \alpha_1(I, \phi(I), \theta) + \mathcal{O}(\sqrt{\varepsilon}).$$

Hence (2.5) and (2.8) suggest that the main contribution to the change of I due to resonance crossing equals $\sqrt{\varepsilon}\sigma(E, I)$ where E is the value of the energy when the orbit crosses the resonance, and

$$(2.13) \quad \sigma(E, I) = \int_{-\infty}^{\infty} \alpha_1(I, \phi(I), \theta(s)) ds$$

where $(r(s), \theta(s))$ is the solution of (2.8) with energy E .

For computations it is more convenient to restate this formula using θ as integration variable. Thus if $L > 0$ then

$$(2.14) \quad \sigma(E, I) = 2 \int_{\theta^*(E)}^{\infty} \frac{\alpha_1(I, \phi(I), \theta)}{\sqrt{2(L\theta + G)}} d\theta$$

and if $L < 0$ then

$$(2.15) \quad \sigma(E, I) = 2 \int_{-\infty}^{\theta^*(E)} \frac{\alpha_1(I, \phi(I), \theta)}{\sqrt{2(L\theta + G)}} d\theta$$

where $G_I(\theta^*(E)) = -L(E + \theta^*(E))$.

To see to what extent (2.13) can be justified we need to look more closely at the dynamics of (2.8). We distinguish two cases. (Below in order to fix our notation we assume that $L > 0$.)

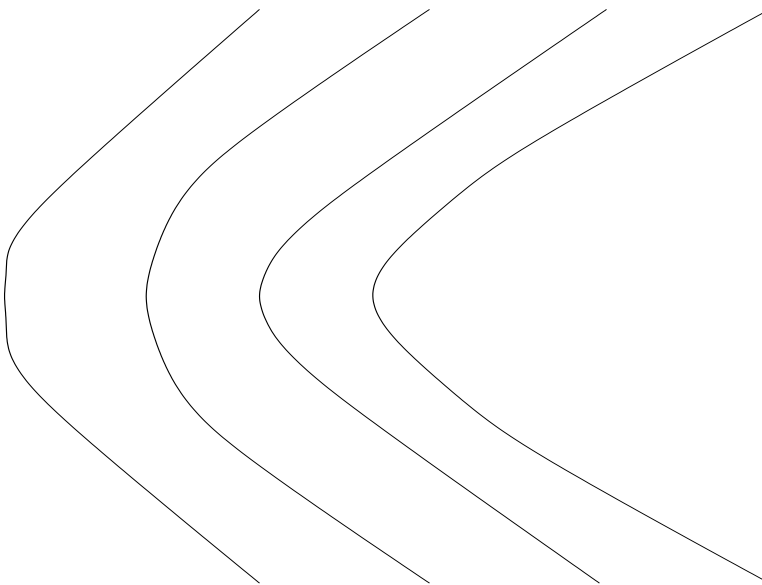


FIGURE 1. Motion near a weak resonance

(I) **Weak resonance.** If $\min_{\theta} g(I, \theta) > -L$ then the phase portrait of (2.8) is topologically the same as the phase portrait of the averaged system (2.3) (which is obtained from (2.8) by dropping g). In this case r is a monotone function of s and the amount of time an orbit spends near the resonance is uniformly bounded from above. Hence for $\varepsilon \neq 0$ the phase portrait is also similar to the phase portrait of the averaged system and our formal asymptotics for the change of I can be easily justified.

(II) **Strong resonance.** Suppose that $\min_{\theta} g(I, \theta) < -L$. The phase portrait of (2.8) in this case is shown on fig. 2. In this case points can spend arbitrary long time near the saddle. As a consequence, for $\varepsilon \neq 0$ the dynamics is qualitatively different. Namely given a saddle $\theta = \theta(I)$ let Ω denote the set of points inside the separatrix loop of this saddle and let Γ be the boundary of this set. Let

$$\hat{H}_I(\theta, r) = \frac{r^2}{2} - L(I)\theta - G_I(\theta) + K(I)$$

where $K(I)$ is chosen so that \hat{H} equals 0 on Γ . Let

$$(2.16) \quad M(I) = - \oint_{\Gamma} \left[r^2 \bar{\beta}(I, 0, \theta, 0) - (L(I) + g(I, \theta)) \bar{\eta}(I, 0, \theta, 0) + \frac{\partial \hat{H}}{\partial I} \bar{\alpha}(I, 0, \theta, 0) \right] ds.$$

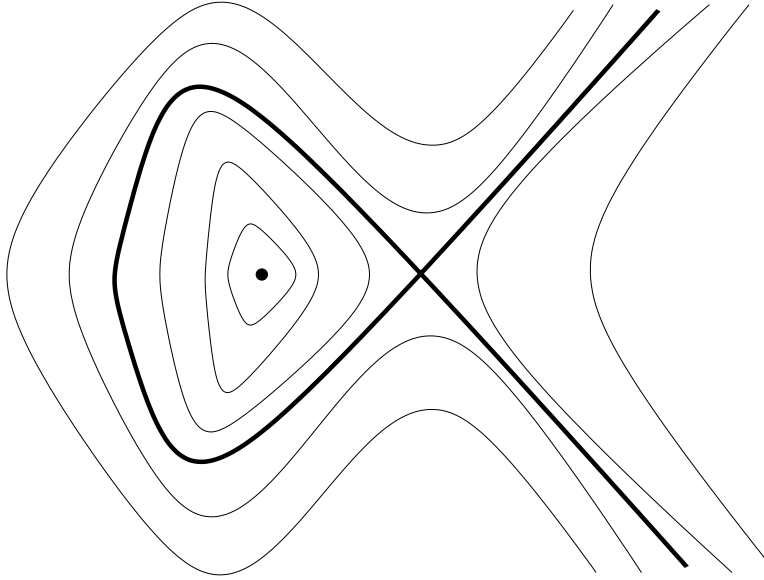


FIGURE 2. The limiting system for strong resonance

(To see that this integral converges note that

$$(2.17) \quad \frac{\partial \hat{H}}{\partial I}(I, 0, \theta(I)) + \frac{\partial \hat{H}}{\partial \theta}(I, 0, \theta(I)) \frac{d\theta}{dI} = 0.$$

Since $\frac{\partial \hat{H}}{\partial \theta}(I, 0, \theta(I)) = 0$ it follows that $\frac{\partial \hat{H}}{\partial I}(I, 0, \theta(I)) = 0$.)

A typical behaviour of the trajectories for $M > 0$ is depicted on fig. 3. In particular the initial conditions starting on the thick segment get captured into resonance and so I experience a jump. A captured

point moves along the resonance. The motion along the resonance is also a slow-fast system with slow variables \hat{H} and I and fast variable θ so it can also be described by the averaging principle, see (3.3) below. At some point the orbit can enter a region where $M(I) < 0$. In this case dynamics near the saddle $(0, \theta(I))$ looks like a mirror image of fig. 3 and the orbit can escape from the resonance so its motion again can be described by the averaged system (2.3). According to [47] (see also Appendix E.5) during each passage a set of points of measure about $\sqrt{\varepsilon}M(I)_+$ gets captured. (Here and below we use the notation $a_+ = \max(a, 0)$.) Therefore we expect that after $\mathcal{O}(1/\sqrt{\varepsilon})$ passages a set of measure $\mathcal{O}(1)$ gets captured. Hence $1/\sqrt{\varepsilon}$ is the natural time scale for this problem. At this scale there are two mechanisms responsible for the change of I .

(I) **Capture into resonance.** This phenomenon has been described above. It is only relevant if $M > 0$ for one of the saddles of (2.8).

The second mechanism is important regardless of the sign of M .

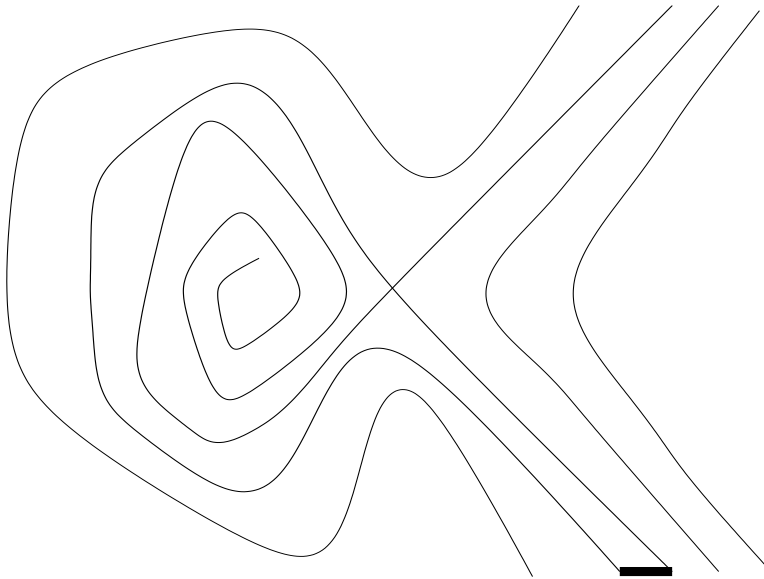


FIGURE 3. Motion near a strong resonance. A projection of a set of orbits starting with a fixed values of $I = I_0$ and $r = -r_0$ where $r_0 \gg 1$ to the (r, θ) -plane is shown.

(II) **Repulsion from resonances** ([51]). To describe this phenomenon, suppose for a moment that (2.1) preserves the Lebesgue measure $dId\phi d\theta$. For each time t the phase space of (2.1) is the union of two sets. First, there is a set of measure $\mathcal{O}(\sqrt{\varepsilon})$ consisting of points which has been captured into resonance and at time t move inside one of

the separatrix domains. For these points I changes with unit speed. This creates a flux of measure of order $\sqrt{\varepsilon}$. Secondly there are points of measure $1 - \mathcal{O}(\sqrt{\varepsilon})$ which follow the orbits of the averaged system (2.3). Therefore in order to maintain the balance of measure most of these points which are not captured move with speed $\sqrt{\varepsilon}$ in the opposite direction. In general Lebesgue measure is not preserved. However the divergence of the flow is given by

$$(2.18) \quad D(I, \phi, \theta) = \frac{\partial \alpha_1}{\partial I} + \frac{\partial \beta_1}{\partial \phi} + \frac{\partial \eta}{\partial \theta} + \mathcal{O}(\varepsilon)$$

Now (2.2) and the periodicity of η imply

$$\int_0^1 D(I, \phi, \theta) d\theta = \mathcal{O}(\varepsilon).$$

So in average the measure is preserved on time scale $o(1/\varepsilon)$ and the previous argument applies.

The flux of orbits moving inside the separatrix domain Ω equals to

$$(2.19) \quad \Psi_\Omega(I) = - \left(\iint_\Omega \alpha_1(I, \phi_0(I), \theta) dr d\theta \right) / L(I),$$

M and Ψ play an important role in the description of the effective evolution of the adiabatic invariant I given in the next section.

3. RESULTS.

3.1. Dynamics before capture. We are now ready to formulate our main results. They require certain non-degeneracy conditions for the system (2.1). Assume that for each $I \in [I_1, I_2]$ the following conditions are satisfied.

(A) $p(I) \neq 0$. This means that I is the only invariant of the averaged system (2.3).

(B) Resonances are non-degenerate. Namely if $(I, \phi_1(I)), (I, \phi_2(I)) \dots (I, \phi_q(I))$ are resonances counted counterclockwise then

$$a_j(I) = \frac{\partial \omega}{\partial \phi}(I, \phi_j(I)) \neq 0.$$

(C) (Twist condition) $\frac{\partial w_j}{\partial I} \neq 0$ where

$$w_j(I) = \int_{\phi_j(I)}^{\phi_{j+1}(I)} \frac{\omega(I, \phi)}{p(I)} d\phi.$$

(Here $j + 1$ is understood mod q .)

(D) For each j , the function $\theta \rightarrow L_j(I)\theta + G_I(\theta)$ is Morse. Namely, there is at most one critical point on each level set and the critical

points are non-degenerate. (Here and below we ignore the dependence of G on j in order to simplify the notation).

We also impose a non-degeneracy condition in the I -direction, namely

(E) If $(0, \theta_{jk}(I))$ is a saddle of (2.8) then for each (j, k) we have $\alpha_1(I, \phi_j(I), \theta_{jk}(I)) \neq 0$. (Here j counts the resonances and k counts the saddles of the given resonance).

(F) For each I the critical points of the function $E \rightarrow \sigma(I, E)$ are non-degenerate.

Suppose that $I = I_0$ and that $(\phi(0), \theta(0))$ are chosen uniformly with respect to the Lebesgue measure. Let $\mathcal{I}_\varepsilon(\mathbf{t}) = I(\mathbf{t}/\sqrt{\varepsilon})$. Since $(\phi(0), \theta(0))$ are random, $\mathcal{I}_\varepsilon(\mathbf{t})$ is a random process. Stop it when it leaves $[I_1, I_2]$ or get captured. Let

$$\Psi_j(I) = \sum_k \Psi_{jk}(I), \quad \Psi(I) = \sum_j \Psi_j(I)$$

where $\Psi_{jk}(I) = \Psi_{\Omega_{jk}}(I)$ (see (2.19)) and Ω_{jk} denotes the domain bounded by the separatrix loop of the saddle $(0, \theta_{jk})$. We denote

$$\lambda(\mathcal{I}) = \begin{cases} \sum_{jk} \frac{M_{jk}(\mathcal{I}) + p(\mathcal{I})}{|L_j(\mathcal{I})|} & \text{if } \mathcal{I} \in (I_1, I_2) \\ \infty & \text{otherwise.} \end{cases}$$

Theorem 1. *Under conditions (A)–(F) the process $\mathcal{I}_\varepsilon(\mathbf{t})$ converges weakly as $\varepsilon \rightarrow 0$ to the solution of*

$$(3.1) \quad \frac{d\mathcal{I}}{d\mathbf{t}} = \Psi(\mathcal{I})p(\mathcal{I}), \quad \mathcal{I}(0) = I_0$$

killed with intensity $\lambda(\mathcal{I})$.

In other words let $\mathcal{I}(I_0, s)$ denote the solution of (3.1). Then the probability that the orbit is not stopped until time \mathbf{t} equals

$$(3.2) \quad \mathbf{p}(\mathbf{t}) = \exp \left[- \int_0^{\mathbf{t}} \lambda(\mathcal{I}(I_0, s)) ds \right].$$

In case of survival $\mathcal{I}(\mathbf{t}) = \mathcal{I}(I_0, \mathbf{t})$.

Remark. In fact Theorem 1 as well as Theorem 2 below are valid for a larger class of initial distribution. Namely it is enough to assume only that the fast variable θ has a Hölder density with respect to Lebesgue measure. The precise class of allowable measures consists of the convex hull of the measures corresponding to the standard pairs. See subsection 6.2 for details.

3.2. Life after capture. In this section we extend Theorem 1 to a fixed size time interval. During that time a positive measure set of orbits is going to be captured, some of them several times, and we need to extend our analysis beyond the time of the capture.

To this end we introduce *entrance–exit maps* $Q_{jk}(I)$ ([47, 5]). To define these maps we consider the system (2.8) on the domain Ω_{jk} . We assume that

(G) There are no saddle points of (2.8) inside Ω_{jk} .

In this case all orbits in the interior of Ω_{jk} are closed so that (2.5) is a slow–fast system with two slow variables I and H_{jk} where

$$H_{jk} = \frac{r^2}{2} - L_j(I)\theta - G_I(\theta) + K_{jk}(I)$$

and $K_{jk}(I)$ is chosen so that H_{jk} vanishes on Ω_{jk} .

Let $\theta_{jk}(H, I, s)$ denote the solution of (2.8) with Hamiltonian H and action I and let $T_{jk}(H, I)$ denote the period of this solution. Consider the averaged quantities

$$X_{jk}(H, I) = \int_0^{T_{jk}(H, I)} \left[r^2 \bar{\beta}(I, 0, \theta_{jk}(s), 0) - \right. \\ \left. (L_j(I) + g(I, \theta_{jk}(s))) \bar{\eta}(I, 0, \theta_{jk}(s), 0) + \frac{\partial H_{jk}}{\partial I} \bar{\alpha}(I, 0, \theta_{jk}(s), 0) \right] ds,$$

$$Y_{jk}(H, I) = \int_0^{T_{jk}(H, I)} \bar{\alpha}(I, 0, \theta_{jk}(s), 0) ds.$$

(In principle we need to divide those quantities by $T_{jk}(H, I)$ but we can avoid doing this by rescaling the time. Observe that the entrance–exit maps Q_{jk} defined below depend only on the orbit of (3.3), not on its time parametrization.)

We observe that as $H \rightarrow 0$, $X_{jk} \sim -M_{jk}(I)$. Hence $X_{jk} < 0$ for small H since a capture is possible near the saddle θ_{jk} . Also $Y_{jk} \sim \bar{\alpha}(I, 0, \theta_{jk}(I), 0) T_{jk}(H, I)$. Thus Y_{jk} is non-zero for small H due to condition (E). Consider *inner averaged equation*

$$(3.3) \quad \frac{d\bar{H}}{ds} = X_{jk}(\bar{H}, \bar{I}), \quad \frac{d\bar{I}}{ds} = Y_{jk}(\bar{H}, \bar{I})$$

with initial condition $(\bar{H}, \bar{I})(0) = (0, I)$. Then for small positive \mathbf{s} we have $\bar{H}(\mathbf{s}) < 0$. We make two assumptions

(H) There exists $\mathbf{s}_{jk}(I)$ such that $\bar{H}(\mathbf{s}) < 0$ for $s \in [0, \mathbf{s}_{jk}(I)]$, $\bar{H}(\mathbf{s}_{jk}(I)) = 0$. Moreover $M_{jk}(I) > 0$ and $M_{jk}(\bar{I}(\mathbf{s}_{jk}(I))) < 0$.

(I) (3.3) is not overtwisted at I (the notion of overtwisting is defined in Section 3.3.)

To formulate our last assumptions consider inner averaged equation (3.3) near an elliptic rest point. By assumption (G) there is a unique such point $(\theta_{jk}^*(I), 0, I)$ inside each domain Ω_{jk} . It is convenient to use variable $\tilde{H} = \frac{r^2}{2} - L_j(I)\theta - G_I(\theta) + \tilde{K}(I)$ where \tilde{K} is chosen so that $\tilde{H}(\theta_{jk}^*(I), 0, I) = 0$. Arguing similarly to (2.17) we get

$$\nabla \tilde{H}(\theta_{jk}^*(I), 0, I) = 0.$$

Considering Taylor series of $(\bar{\eta}, \bar{\beta}, \bar{\alpha})$ in $(r, \theta - \theta_{jk}^*)$ and observing that the averaging kills odd degree terms we can write the averaged equation for \tilde{H} near $\{\tilde{H} = 0\}$ as

$$(3.4) \quad \frac{d\tilde{H}}{ds} = \Lambda(\bar{I}, \tilde{H})\tilde{H}.$$

Let $\mathbf{S}_{jk} = (\theta_{jk}^*(I), 0, I)$ denote the curve of elliptic fixed points. (3.4) means that \mathbf{S}_{jk} is invariant under the averaged dynamics. The restriction of the inner averaged equation to \mathbf{S}_{jk} is

$$(3.5) \quad \frac{d\bar{I}}{ds} = 2\pi \frac{\bar{\alpha}(\theta_{jk}^*(\bar{I}), 0, \bar{I}, 0)}{\omega_{jk}(\bar{I})}$$

where $i\omega_{jk}(I)$ is the eigenvalue of the rest point. We assume that

(J) The zeroes of $\bar{\alpha}$ are non-degenerate.

Thus \mathbf{S}_{jk} is a union of fixed points and their stable manifolds.

Let I_{jkl} be the set of zeroes of the RHS of (3.5). Our last assumption is

(K) $\Lambda(I_{jkl}, 0) \neq 0$.

In other words all fixed points on \mathbf{S}_{jk} are hyperbolic.

We let

$$Q_{jk}(I) = \bar{I}(\mathbf{s}_{jk}(I)).$$

We are now ready to generalize Theorem 1. Let \mathbf{G} be a finite union of closed intervals such that the conditions (A)–(K) are satisfied for all $I \in \mathbf{G}$. Moreover we assume that for each $I \in \mathbf{G}$ for each pair (j, k) the assumptions (G), (I) and (K) are satisfied for all I along the orbit of inner averaged equation (3.3) from I to $\mathbf{s}_{jk}(I)$. We now relax the stopping rules as follows. We stop our process when the orbit is less than $\sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}$ away from the resonance surface and $\mathcal{I}_\varepsilon(t)$ is not in \mathbf{G} . (It will follow from Theorem 2 below that there are two reasons a typical orbit is stopped. Either it reaches the boundary of \mathbf{G} while moving along the trajectory of (3.1) or it gets captured at some I and $Q_{jk}(I) \notin \mathbf{G}$.) Unlike Theorem 1 we do not require \mathbf{G} to be connected since entrance-exit maps Q_{jk} can move orbits between the different

components. In particular we do not insist that $I \in \mathbf{G}$ while it moves along the resonance surface.

As in Theorem 1 we assume that $I(0) = I_0$ is fixed and $(\phi(0), \theta(0))$ are uniformly distributed.

Theorem 2. *As $\varepsilon \rightarrow 0$, $\mathcal{I}_\varepsilon(t)$ converges to the Markov process with generator*

$$\mathcal{L}(A) = \Psi(I) \frac{dA}{dI} + \sum_{jk} \left(\frac{M_{jk}(I)_+ p(I)}{L_j(I)} [A(Q_{jk}(I)) - A(I)] \right).$$

In other words (3.1) and (3.2) have to be supplemented by the following

- Given that the point got captured at time t the conditional probability that it get captured near θ_{jk} is

$$\frac{M_{jk}(\mathcal{I}(t))_+ / L_j(\mathcal{I}(t))}{\sum_{jk} [M_{jk}(\mathcal{I}(t))_+ / L_j(\mathcal{I}(t))]}$$

- If the point gets captured at time t near θ_{jk} then it moves instantly to $Q_{jk}(\mathcal{I}(t))$.

Remark. Since the limiting process has jumps the convergence is understood in the space of functions without jumps of second kind with Skorokhod topology ([25], Section VI.5). Namely two functions are close if their discontinuities are close and the values of the functions are close away from the discontinuities. More precisely, the topology is given by the distance

$$d(\mathcal{I}_1, \mathcal{I}_2) = \inf_h \left(\sup_{t \in [0, T]} |\mathcal{I}_1(t) - \mathcal{I}_2(h(t))| + \sup_{t \in [0, T]} |t - h(t)| \right)$$

where $[0, T]$ is the common domain of \mathcal{I}_1 and \mathcal{I}_2 and the infimum is taken over all homeomorphisms $h : [0, T] \rightarrow [0, T]$.

Remark. Assumptions (A)-(K) state that certain functions are different from 0. Therefore for a typical system we can expect them to hold in a neighborhood of a typical point. However we can not expect them to hold globally so in Theorems 1 and 2 we stop the process when it escapes the region of validity of assumptions (A)-(K). See Section 12 for the discussion of the ways to relax the assumptions (A)-(K).

3.3. Formal solution of the averaged equation. To complete the formulation of Theorem 2 it remains to explain the condition (I). To do so we discuss the properties of the inner averaged equation (3.3). In this section we shall omit the subscripts j and k in order to simplify

the notation. Thus we shall write θ_{cr} instead of θ_{jk} and let U be the potential of the Hamiltonian H_{jk} . Define

$$(3.6) \quad \mathbf{c}(I) = \frac{\alpha_1(I, \phi(I), \theta_{cr}(I))}{\sqrt{\left| \frac{\partial^2 U}{\partial \theta^2}(\theta_{cr}(I), I) \right|}}.$$

The next result is proven in Appendix F (see also [46]).

Lemma 3.1. *As H tends to 0*

$$\begin{aligned} X(H, I) &\rightarrow -M(I), \quad Y(H, I) \sim \mathbf{c}(I) |\ln |H||. \\ \frac{\partial X}{\partial H} &\sim \mathcal{O}(|\ln |H||), \quad \frac{\partial X}{\partial I} = \mathcal{O}(1), \\ \frac{\partial Y}{\partial H} &\sim \frac{\mathbf{c}(I)}{|H|}, \quad \frac{\partial Y}{\partial I} \sim \mathbf{c}'(I) |\ln |H||, \\ \frac{\partial^2 X}{\partial H^2} &\sim \mathcal{O}\left(\frac{1}{|H|}\right), \quad \frac{\partial^2 X}{\partial H \partial I} = \mathcal{O}(|\ln |H||), \quad \frac{\partial^2 X}{\partial I^2} = \mathcal{O}(1), \\ \frac{\partial^2 Y}{\partial H^2} &\sim \frac{\mathbf{c}(I)}{H^2}, \quad \frac{\partial^2 Y}{\partial H \partial I} = \frac{\mathbf{c}'(I)}{|H|}, \quad \frac{\partial^2 Y}{\partial I^2} = \mathbf{c}''(I) |\ln |H||. \end{aligned}$$

Using this lemma we analyze the variational equation of (3.3). Consider a solution to (3.3) defined on the interval $[0, \mathbf{t}]$ such that $(H, I)(0) = (H_0, I_0)$ and $(H, I)(\mathbf{t}) = (H_f, I_f)$. We assume that H_0 and H_f are small so that \mathbf{t} is close to $\mathbf{s}(I_0)$. We have

$$\begin{aligned} \delta \dot{H} &= \frac{\partial X}{\partial H} \delta H + \frac{\partial X}{\partial I} \delta I, \\ \delta \dot{I} &= \frac{\partial Y}{\partial H} \delta H + \frac{\partial Y}{\partial I} \delta I. \end{aligned}$$

Let $\Delta(t) = Y \delta H - X \delta I$. Then $\dot{\Delta} = \left(\frac{\partial X}{\partial H} + \frac{\partial Y}{\partial I} \right) \Delta$.

Observe that due to assumption (H) the RHS has an integrable singularity. Let

$$A(t) = \exp \left(\int_0^t \left[\frac{\partial X}{\partial H} + \frac{\partial Y}{\partial I} \right] (s) ds \right).$$

Then $\Delta(t) = A(t) \Delta_0$. Accordingly

$$\delta \dot{H} = \left(\frac{\partial X}{\partial H} + \frac{Y}{X} \frac{\partial X}{\partial I} \right) \delta H - \frac{\frac{\partial X}{\partial I} A(t) \Delta_0}{X}.$$

Let

$$B(t) = \exp \left(\int_0^t \left[\frac{\partial X}{\partial H} + \frac{Y}{X} \frac{\partial X}{\partial I} \right] (s) ds \right).$$

Then

$$\delta H(t) = B(t) \left(\delta H(0) - \Delta_0 \int_0^t \frac{A(s) \frac{\partial X}{\partial I}}{B(s)X} ds \right).$$

Denoting

$$(3.7) \quad C^*(I) = B(\mathbf{s}(I)) \int_0^{\mathbf{s}(I)} \frac{A(s) \frac{\partial X}{\partial I}}{B(s)X} ds$$

we obtain the following asymptotics of the solutions as $H_0, H_f \rightarrow 0$

$$(3.8) \quad \frac{\partial H_f}{\partial H_0} \sim -C^*(I_0) \mathbf{c}(I_0) |\ln |H_0||, \quad \frac{\partial I_f}{\partial H_0} \sim \frac{-C^*(I_0) \mathbf{c}(I_0) \mathbf{c}(I_f) |\ln |H_0|| |\ln |H_f||}{M(I_f)},$$

$$(3.9) \quad \frac{\partial H_f}{\partial I_0} \sim -C^*(I_0) M(I_0), \quad \frac{\partial I_f}{\partial I_0} \sim \frac{-C^*(I_0) M(I_0) \mathbf{c}(I_f) |\ln |H_f||}{M(I_f)}.$$

In the proof of Theorem 2 it is convenient to have the leading terms given by the above equations non-vanishing. We say that (3.3) is *over-twisted at I_0* if $C^*(I_0) = 0$.

Condition (I) provides additional hyperbolicity for our system since it implies that for orbits passing closer to a saddle point of (2.8) expansion is stronger than for points staying far from the saddles. See Appendix F for details.

4. EXAMPLES.

Here we illustrate typical applications of our main results. In this section we restrict ourselves to demonstrating how to reduce these examples to the form (2.1) required to apply our results relegating straightforward but lengthy computations of the limiting process to Appendix G. We leave it to the reader to check that the nondegeneracy conditions needed in Theorems 1 and 2 are satisfied on an open set of values of I for a typical values of the parameters involved in our examples.

4.1. Motion in rapidly oscillating force field.

Example 1. Consider a particle in a potential field which is subjected to forcing which rapidly changes both in space and time.

$$\ddot{x} + U'(x) = \sin \left(\frac{x-t}{\varepsilon} \right).$$

Introduce $\theta = \frac{x-t}{\varepsilon}$. Then we have $\dot{\theta} = \frac{v-1}{\varepsilon}$. The averaged equation therefore takes form

$$(4.1) \quad \ddot{x} + U'(x) = 0$$

and so its energy

$$I = \frac{v^2}{2} + U - \frac{1}{2} \text{ where } v = \dot{x}$$

is an adiabatic invariant for original system. (Here the normalization by subtracting $\frac{1}{2}$ is made in order to simplify the formulas below). Observe that

$$\dot{I} = v \sin \theta.$$

The resonant curve takes form $\{v = 1\}$. Introducing $r = \frac{v-1}{\sqrt{\varepsilon}}$ we obtain the inner system

$$\theta'' = r' = -U'(x) + \sin \theta.$$

The relation

$$I = U(x) + \frac{(1 + \sqrt{\varepsilon}r)^2}{2} - \frac{1}{2}$$

gives

$$U(x) = I - \sqrt{\varepsilon}r + \mathcal{O}(\varepsilon).$$

Let Z denote the local inverse of U , $U(Z(u)) = u$ then $Z' = \frac{1}{U'}$. The inner system takes form

$$(4.2) \quad \theta'' = -U'(Z(I)) + \sin \theta + \frac{U''}{U'}(Z(I))r\sqrt{\varepsilon} + \mathcal{O}(\varepsilon).$$

The computation of the limiting process is given in Appendix G.1. We see that at time intervals of order 1 we can ignore the oscillating force and so the motion of the particle appears to follow the averaged system (4.1). The influence of the forcing is felt at time intervals of order $\varepsilon^{-1/2}$. Namely it causes the changes of the particle's energy by a slow drift according to the equation (G.3) and occasional captures into resonances which happen with intensity given by (G.4). The entrance-exit function is computed using inner averaged equation (G.2).

4.2. Motion on narrow cylinder in the presence of magnetic field.

Example 2. Consider a particle moving on a narrow cylinder in the presence of a magnetic field

$$\begin{aligned} \dot{z} &= v, & \dot{v} &= z(N + y)u, \\ \dot{\theta} &= \frac{u}{\varepsilon}, & \dot{u} &= -z(N + y)v \end{aligned}$$

where $y = \sin \theta$. The kinetic energy is preserved and we assume that it is equal to 2, $v^2 + u^2 = 1$. Introducing a variable ψ such that

$$v = \cos \psi, \quad u = \sin \psi$$

we rewrite our system as

$$\begin{aligned}\dot{\theta} &= \frac{\sin \psi}{\varepsilon}, \\ \dot{z} &= \cos \psi, \\ \dot{\psi} &= z(N + y).\end{aligned}$$

After averaging over the fast variable θ the last equation becomes $\dot{\psi} = Nz$ and so the averaged system

$$(4.3) \quad \dot{z} = \cos \psi, \quad \dot{\psi} = Nz$$

has a first integral

$$I = \frac{Nz^2}{2} - \sin \psi.$$

For the actual system we have

$$\dot{I} = -z \cos \psi \sin \theta.$$

The resonance curves are $\{\psi = 0\}$ and $\{\psi = \pi\}$. Introducing $r = \frac{\sin \psi}{\sqrt{\varepsilon}}$ we get

$$r' = z \cos \psi (N + \sin \theta) = \pm \sqrt{\frac{2(I + \sqrt{\varepsilon}r)}{N}} (N + \sin \theta) + \mathcal{O}(\varepsilon).$$

So the inner system takes form

$$(4.4) \quad \theta'' = \pm \left(\sqrt{\frac{2I}{N}} (N + \sin \theta) + \frac{\sqrt{\varepsilon}}{\sqrt{2NI}} r (N + \sin \theta) \right) + \mathcal{O}(\varepsilon).$$

The computation of the limiting process is given in Appendix G.2. We see that most of the time particle makes rapid rotations around the cylinder with its vertical coordinate changing according to a pendulum equation (4.3). At time intervals of order $1/\sqrt{\varepsilon}$ the energy of the pendulum experiences abrupt changes due to captures in resonances which manifest themselves by particle moving vertically for times of order 1. The time before capture is an exponential random variable with parameter $\lambda(\mathcal{I})$ given by formula (G.7) and the entrance-exit function is computed using equations (G.5)–(G.6).

5. IDEA OF THE PROOF.

In this section we present main ideas of the proof of Theorem 1. The proof of Theorem 2 proceeds along similar lines. The necessary modifications are presented in Section 11.

We shall use freely formal manipulations and heuristic arguments. Rigorous justifications (of somewhat weaker results which are still sufficient for the proof of Theorem 1) will be given later.

We need some notation. Given a resonance $\phi = \phi_j(I)$ define two surfaces \mathcal{S} and $\tilde{\mathcal{S}}$ as follows. Let c be a small number. (The precise conditions on c will be given later but they are not important in the discussion to follow.) If $a_j(I) > 0$ let

$$(5.1) \quad \mathcal{S} = \left\{ \frac{\omega}{\sqrt{\varepsilon}} = - \left(c\varepsilon^{-1/4} + \frac{G_I(\theta)}{c\varepsilon^{-1/4}} \right) \right\},$$

$$(5.2) \quad \tilde{\mathcal{S}} = \left\{ \frac{\omega}{\sqrt{\varepsilon}} = c\varepsilon^{-1/4} + \frac{G_I(\theta)}{c\varepsilon^{-1/4}} \right\}.$$

If $a_j(I) < 0$ then we define \mathcal{S} by (5.2) and $\tilde{\mathcal{S}}$ by (5.1). Thus \mathcal{S} is a section immediately before the resonance and $\tilde{\mathcal{S}}$ is a section immediately after resonance. One motivation for this choice of the sections is that (5.1) and (5.2) have quite simple expressions in terms of the improved adiabatic invariants, see Appendix B. Another motivation is that in (r, θ) coordinates

$$\tilde{\mathcal{S}} = \left\{ r = \mp \left(\bar{R} + \frac{G}{\bar{R}} \right) \right\}$$

where $\bar{R} = c/\varepsilon^{-1/4}$. Hence

$$(5.3) \quad \frac{\partial E}{\partial \theta} \sim \pm 1.$$

Let x_n be consecutive visits to \mathcal{S} s corresponding to different resonances. (2.3) suggests that it takes time about $\frac{1}{p(I)}$ to make a complete circle. Since there are q resonances per period we expect that the time of the n th visit is $t_n = \frac{n}{p(I)q}(1 + o(1))$. Let $A(x_n)$ denote the change of I between t_n and t_{n+1} . To establish Theorem 1 we need to show that if $N \ll \frac{1}{\sqrt{\varepsilon}}$ and $m \leq \text{Const}/\sqrt{\varepsilon}$ then for orbits which has not been captured up to time m

(I) The probability of capture on the segment $[m, m + N]$ is about $\frac{\lambda(I)N\sqrt{\varepsilon}}{pq}$

(II) For non-captured points

$$\frac{1}{N} \sum_{n=m}^{m+N} A(x_n) \sim \frac{\Psi(I)N}{pq}.$$

The first statement is nothing but the Poisson Limit Theorem for our system and the second is the Law of Large Numbers. Both results are well understood for independent or weakly dependent random variables so we need to show that x_n are weakly dependent. To understand

where this independence comes from consider first few iterations of the Poincare map. Suppose that x_0 has a smooth density on

$$\mathcal{S}_{I_0} := \mathcal{S} \cap \{I = I_0\}.$$

We want to describe the distribution of x_1 . The passage to the next resonant surface $\bar{\mathcal{S}}$ consists of two parts.

- (1) Passage of the resonant zone (from \mathcal{S} to $\tilde{\mathcal{S}}$).
- (2) Motion far from resonance (from $\tilde{\mathcal{S}}$ to $\bar{\mathcal{S}}$).

The considerations of Section 2 suggest that during the first part I changes as follows

$$(5.4) \quad \tilde{I} \approx I + \sqrt{\varepsilon}\sigma(E, I), \quad \text{so} \quad \frac{\partial \tilde{I}}{\partial E} \approx \sqrt{\varepsilon} \frac{\partial \sigma}{\partial E}.$$

On the other hand (2.1) and (2.3) suggest that

$$\bar{\theta} - \tilde{\theta} \approx \frac{1}{\varepsilon} w(\tilde{I}) = \frac{1}{\varepsilon} \int_{\phi}^{\bar{\phi}} \frac{\omega(\tilde{I}, \phi)}{p(\tilde{I})} d\phi.$$

Formal differentiation gives

$$(5.5) \quad \frac{\partial \bar{\theta}}{\partial \tilde{\theta}} \approx \frac{1}{\varepsilon} \frac{\partial w}{\partial I}(I).$$

Combining this with (5.3) we obtain

$$\frac{\partial \bar{E}}{\partial E} \approx \frac{1}{\sqrt{\varepsilon}} \frac{\partial w}{\partial I} \frac{\partial \sigma}{\partial E}.$$

Now Assumption (F) implies that for most points $\frac{\partial \sigma}{\partial E} \neq 0$. In this case Assumption (C) tells us that a preimage of a unit interval has length $\mathcal{O}(\sqrt{\varepsilon})$ so $\frac{\partial \sigma}{\partial E}$ is approximately constant at this preimage. That is, on such intervals $E \rightarrow \bar{E}$ can be approximated by a linear map. For linear maps of slope $\frac{1}{\sqrt{\varepsilon}}$ the images of the segments of length $\gg \sqrt{\varepsilon}$ are uniformly distributed. In other words we can prescribe the phase of x_0 with a good precision and still the phase of x_1 is uniformly distributed. In this sense we can regard the phases of x_1 and x_0 as weakly dependent. To summarize the image of \mathcal{S}_{I_0} consists of finitely many segments consisting of captured points, $\mathcal{O}(1/\sqrt{\varepsilon})$ almost linear segments and finitely many 'parabolic' segments coming from neighborhoods of the critical points of σ . Assumption (F) makes it plausible that if E_0 is a critical point of σ then for E near E_0

$$\bar{E} - \bar{E}_0 \sim \frac{(E - E_0)^2}{\sqrt{\varepsilon}}$$

so the total measure in parabolic pieces is $\mathcal{O}(\varepsilon^{1/4})$ (by definition a parabolic piece consists of points which are within a unit distance from \bar{E}_0).

Now let us see what happens for large n . The non-captured part of the image of \mathcal{S}_{I_0} consists of

- (a) almost linear segments;
- (b) parabolic pieces;
- (c) curves of more complicated geometry appearing when a parabolic piece comes near the critical points of σ .

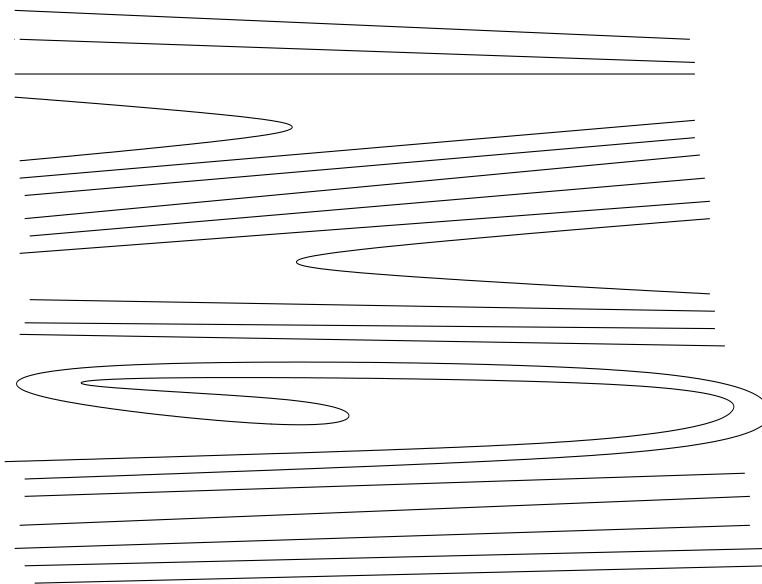


FIGURE 4. Large time image of \mathcal{S}_{I_0} .

To apply the argument used for x_0 we need to prove that most of the measure is concentrated in the linear segments. To this end we have to show that once an orbit finds itself in a parabolic piece it is much more likely to escape (that is to get farther than a unit distance from the tip) rather than stay close to the tip or even enter into a curve of type (c). The reason is the following. Since $\frac{\partial\sigma}{\partial E} \sim 1$ for most points we can expect that

$$(5.6) \quad \frac{\partial\sigma}{\partial E}(\bar{E}_0) \sim 1$$

in which case for the second iteration we would get

$$\bar{\bar{E}} - \bar{E}_0 \sim \frac{(E - E_0)^2}{\varepsilon}$$

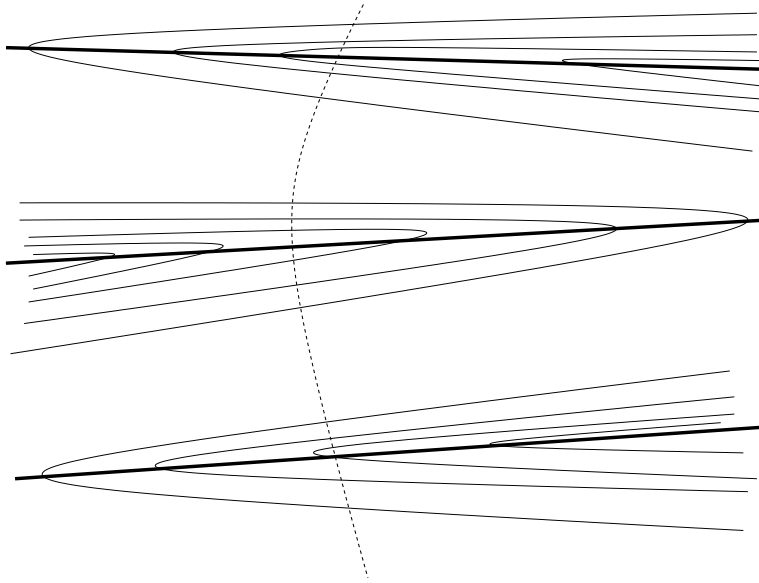


FIGURE 5. Parabolic pieces are close to the image of \mathcal{C} (thick line). Since this image is transversal to \mathcal{C} (dashed line) most of the returns are isolated.

so the set of points staying close to the tip for two iterations in a row would have measure $\mathcal{O}(\sqrt{\varepsilon})$ (there are also secondary parabolas appearing at the intersection of the primary one with the critical curves but if (5.6) holds then their measures are $\mathcal{O}(\sqrt{\varepsilon})$ as well). Continuing this reasoning it is not difficult to convince oneself that the set of points staying in (b) or (c) segments three times in a row have measure $o(\sqrt{\varepsilon})$. Thus if $n < \text{Const}/\sqrt{\varepsilon}$ then the set of non-linear pieces is small as required. The problem with this argument is that it is unreasonable to expect (5.6) for all initial values of I_0 . Let $\mathcal{C} = \{\frac{\partial \sigma}{\partial E} = 0\}$. Then (5.5) implies that the image of \mathcal{C} also consist of almost linear segments and so we can not keep the image of \mathcal{C} from intersecting itself. Dynamics near such intersection could be quite complicated. In particular elliptic islands could be formed. For this reason we can not guarantee for the first and the second iteration of the Poincare map that only a set of measure $o(\sqrt{\varepsilon})$ contributes to the nonlinear part of the image. Fortunately this problem is confined to the first two iterates only. Indeed even for the first iteration the required $o(\varepsilon)$ bound can be obtained if we *assume* that I_0 has *isolated returns* that is, the images of $\mathcal{C} \cap \mathcal{S}_{I_0}$ are far from \mathcal{C} . Starting from the third iteration we can actually prove that most of the returns are isolated. Indeed (5.5) tells us that the points of intersection of \mathcal{C} and the image of \mathcal{C} form a lattice of step

$\mathcal{O}(\varepsilon)$. On the other hand most of the second image of \mathcal{S}_{I_0} is in linear segments. There are $\mathcal{O}(1/\varepsilon)$ such segments. By (5.4) the variation of their I values is $\mathcal{O}(\sqrt{\varepsilon})$ so the distance between consecutive segments is $\mathcal{O}(\varepsilon^{3/2})$. Thus most of the segments are far from the critical lattice. Hence most of the returns are isolated (see fig. 5). So the total measure of the non-linear segments can be bounded by

$$o(n\sqrt{\varepsilon}) + \mathcal{O}(\varepsilon^{1/4})$$

where the second term estimates the measure of bad points formed during the first two iterations. Thus most of the measure is in linear segments giving the required independence.

6. PLAN OF THE PROOF.

Here we present the main steps of the proof of Theorem 1.

6.1. One passage. Our analysis in Section 5 was based on heuristic formulas for the derivatives of the Poincare map. Here we present the precise results. Let us emphasize that the parts of the statements not dealing with derivative estimates are well known. Still we provide sketches of proofs in Appendix A in order to make our exposition self contained.

Let \mathcal{S}_j denote the preresonance surface of the resonance $(I, \phi_j(I))$. Denote $\mathbb{S} = \bigcup_j \mathcal{S}_j$. We want to study the Poincare map $P : \mathbb{S} \rightarrow \mathbb{S}$. Let \mathcal{S} and $\bar{\mathcal{S}}$ be the surfaces corresponding to two consecutive resonances $(I, \phi(I))$ and $(I, \bar{\phi}(I))$. Below we present some information about $P : \mathcal{S} \rightarrow \bar{\mathcal{S}}$. Observe that if we are interested in the derivatives bounds then it is not convenient to work with I and θ since they are rapidly oscillating and a slight change of the surface produces big changes of the derivatives. For this reason we shall work with the variables J and E defined by (2.4) and (2.11) respectively. Denote $P(J, E) = (\bar{J}, \bar{E})$.

In the statements below $o(\dots)$ means the limit then both c and ε tend to 0 but c goes to 0 much slower than ε . That is there exists a function $c_0 = c_0(\varepsilon)$ such that the asymptotics below are uniform for c, ε small, $c \geq c_0(\varepsilon)$.

In Appendix B we prove the following.

Proposition 6.1. *If $(I, \phi(I))$ is a weak resonance then*

- (a) *The passage time $t(J, E) \sim \frac{\bar{\phi} - \phi}{p(I)}$.*
- (b)

$$\bar{I} - I = \bar{J} - J + o(\sqrt{\varepsilon}) = \sqrt{\varepsilon}\sigma(J, E) + o(\sqrt{\varepsilon}).$$

(c) We have

$$(6.1) \quad \frac{\partial \bar{E}}{\partial E} = \frac{1}{\sqrt{\varepsilon}} (\Lambda(J, E) + o(1)) + \mathcal{O}(1), \quad \frac{\partial \bar{E}}{\partial J} = \frac{1}{\varepsilon} \left(\frac{\partial w}{\partial I} + o(1) \right),$$

where

$$\Lambda(J, E) = \frac{\partial w}{\partial I} \frac{\partial \sigma}{\partial E}, \quad \text{and}$$

$$(6.2) \quad \frac{\partial^2 \bar{E}}{\partial E^2} = \frac{1}{\sqrt{\varepsilon}} \frac{\partial \Lambda(J, E)}{\partial E} + \mathcal{O} \left(\varepsilon^{1/4} \left| \frac{\partial \bar{E}}{\partial E} \right|^2 \right).$$

$$(d) \quad \frac{\partial \bar{J}}{\partial J} = \mathcal{O}(1), \quad \frac{\partial \bar{J}}{\partial E} = \mathcal{O}(\sqrt{\varepsilon}), \quad \frac{\partial^2 \bar{J}}{\partial E^2} = \mathcal{O}(\sqrt{\varepsilon}) + \mathcal{O} \left(\varepsilon^{5/4} \left| \frac{\partial \bar{E}}{\partial E} \right|^2 \right),$$

$$\frac{\partial^2 \bar{E}}{\partial E \partial J} = \mathcal{O} \left(\frac{1}{\varepsilon} \right) + \mathcal{O} \left(\varepsilon^{-3/4} \left| \frac{\partial \bar{E}}{\partial E} \right| \right), \quad \frac{\partial^2 \bar{J}}{\partial E \partial J} = \mathcal{O}(1) + \mathcal{O} \left(\varepsilon^{1/4} \left| \frac{\partial \bar{E}}{\partial E} \right| \right),$$

$$\frac{\partial^2 \bar{E}}{\partial J^2} = \mathcal{O} \left(\frac{1}{\varepsilon^{-7/4}} \right), \quad \frac{\partial^2 \bar{J}}{\partial J^2} = \mathcal{O}(\varepsilon^{-3/4}).$$

For strong resonance the formulations are more complicated because as we explained above the heuristic formulas of Proposition 6.1 are not valid for all orbits but on the other hand we do not want to exclude too large a set.

Fix arbitrary $\delta > 0$. For $\varepsilon = 0$ the union of the saddles of (2.8)

$$\mathcal{N} = \bigcup_{jk} \{ \phi = \phi_j(I), \quad \theta = \theta_{jk}(I) \}$$

is a normally hyperbolic invariant set of the flow (2.5). Hence for small ε there is a normally hyperbolic invariant set \mathcal{N}_ε near \mathcal{N} (see [29]). Let d be the distance of the orbit from \mathcal{N}_ε and τ be the time the orbit spends in a δ neighborhood of \mathcal{N}_ε .

In Appendix E we prove the following.

Proposition 6.2. *If $(I, \phi(I))$ is a strong resonance then the following statements hold.*

(a) *If $d > \delta$ then the estimates of Proposition 6.1 are valid.*

(b) *If $\sqrt{\frac{\varepsilon}{|\ln \varepsilon|}} < d \leq \delta$ but the orbit is not captured, then the statements of Proposition 6.1 should be modified as follows. Part (a) remains valid, part (b) has to be replaced by*

$$(6.3) \quad \bar{J} - J \sim \sqrt{\varepsilon} \tau \alpha_1(I, \phi(I), \theta_{jk}(I))$$

where $(0, \theta_{jk}(I))$ is the saddle of (2.8) the orbit is passing near. Part (c) has to be replaced by

$$(6.4) \quad \begin{aligned} \frac{\partial \bar{E}}{\partial E} &\geq \frac{\text{Const}}{d\sqrt{\varepsilon}}, \\ \frac{\partial \bar{E}}{\partial J} &\leq \frac{\text{Const}}{d\varepsilon}. \end{aligned}$$

Also for part (d) in the estimates for the first derivative the RHS has to be multiplied by $1/d$ and in the estimates for the second derivative the RHS has to be multiplied by $|\ln^{\mathbf{m}} d|/d^2$ where \mathbf{m} is a sufficiently large number.

Let γ be the graph of $J = g(E)$, with $|g'| < \varepsilon^{1/2+\delta}$,

(c) Let $\xi > \varepsilon^{1/2+\delta}$. Then measure of points with $d \leq \xi$ is less than $\text{Const}\xi$.

(d) Let $[E_-^{(jk)}, E_+^{(jk)}]$ be the segment of points captured while passing near the resonance $(I, \phi_j(I), \theta_{jk}(I))$. Then

$$E_+^{(jk)} - E_-^{(jk)} = \sqrt{\varepsilon} \frac{M_{jk}(I)}{L_j(I)} + o(\sqrt{\varepsilon}).$$

Comparing the statements above with the heuristic estimates of Section 5 we see the following. The first derivative bounds can be obtained by the formal differentiation of the formulas suggested by the averaged equation (after replacing I by J). This is not true however for the second derivatives. The reason is that our analysis suggested that the map $E \rightarrow \bar{E}$ was close to linear. Now if we perturb a linear map such as $E \rightarrow \frac{E}{\sqrt{\varepsilon}}$ by a small nonlinearity, say

$$E \rightarrow E \rightarrow \frac{E}{\sqrt{\varepsilon}} + \delta h\left(\frac{E}{\sqrt{\varepsilon}}\right)$$

then the effect of the nonlinear term is $\delta h''(E/\sqrt{\varepsilon})/\varepsilon$ which can be larger than $1/\sqrt{\varepsilon}$ even if δ is small. However the second derivative bounds are still sufficient to derive the conclusions we want. Namely, for most points

$$\frac{\partial \bar{E}}{\partial E} \sim \frac{1}{\sqrt{\varepsilon}} \quad \text{whereas} \quad \frac{\partial^2 \bar{E}}{\partial E^2} = \mathcal{O}(\varepsilon^{-3/4}) \ll \left| \frac{\partial \bar{E}}{\partial E} \right|^2$$

which is enough to conclude that the image looks like a linear map (cf. Lemma H.1). On the other hand if $\frac{\partial \bar{E}}{\partial E} \ll \frac{1}{\sqrt{\varepsilon}}$ then the formal arguments of Section 5 remain valid (since we have quadratic decrease in the terms coming from the nonlinearity!) and we still are able to prove quadratic bounds near the parabolic pieces.

We conclude this section by computing the average for our Law of Large Numbers. The proof is given in Appendix A.

Lemma 6.3. *The average value of the jump at resonance $(I, \phi_j(I))$ is*

$$\int_0^1 \sigma(I, E) dE = \Psi_j(I).$$

6.2. Standard pairs. Now we state more precisely what we mean by “almost linear” segments. The most obvious requirement is that we want to control the curvature. However we need two more conditions. First we want not only the geometry of the image to be close to linear but also the image of the initial measure to be uniform since the uniformity of the phase plays a key role in our argument. Secondly, to control the geometry of high iterates inductively the way we analyzed the first iteration we want to assert that the images are close to \mathcal{S}_I curves, namely that they go roughly in E -direction. The precise definition is the following.

Let c_1, c_2, C_1, C_2, C_3 be constants whose precise values will be specified later and δ be a small number.

Definition. *A standard pair is a pair $\ell = (\gamma, \rho)$ where γ is a curve in some \mathcal{S} and ρ is a probability density on γ such that*

- (a) $c_1 \leq \text{length}(\gamma) \leq c_2$
- (b) γ is a graph of a map $J = g(E)$ with
 - (b1) $|g'(E)| \leq C_1 \varepsilon$,
 - (b2) $|g''(E)| \leq C_2 \varepsilon^{1/2+\delta}$,
- (c) $|\frac{d}{dE} \ln \rho| \leq C_3$

If ℓ is a standard pair and A is a function we write

$$\mathbb{E}_\ell(A) = \int_\gamma A(x) \rho(x) dx.$$

Standard pairs had been applied to derive the averaging results in case the fast motion is hyperbolic ([7, 13, 19]). Here we shall use them in the periodic setting.

We want to show that most of the image of \mathcal{S}_{I_0} consists of standard curves. The problem is that an orbit can pass close to either region where $\frac{\partial \sigma}{\partial E} = 0$ where expansion coefficient of (6.1) is small or it can pass close to the separatrix of (2.8) in which case it spends a long time near resonance and the perturbation terms can become too large invalidating (6.1). Let γ be a standard curve. If either of the above problems happen we say that the orbit has a close return. We call passages near the separatrix close returns of the first kind and passages near critical

region close returns of the second kind. We postpone the precise definition of these notions till the later sections since it is quite technical. For close returns of the first kind, the time spent near the separatrix is logarithmic function of the minimal distance which implies that returns where the deviations from the unperturbed inner system are significant have small probability, so they can be ignored. However close returns of the second kind should be treated more carefully. Namely in this case we consider few consecutive iterations of the Poincare map. If they also give close returns when we say that the orbit got stuck in the critical region. Likewise for the returns of the first kind we say that the orbit got stuck near the separatrix. Also observe that whatever return is close or not depends not only on the point itself but on the curve under consideration because we want to have expansion in the tangent direction of this curve. However we will not emphasize this dependence in order to simplify the notation. In this section we shall only use the properties of the close returns described in the next two propositions. Let $\ell = (\gamma, \rho)$ be a standard pair. Denote by $\bar{P}(x)$ the first free return to some resonant surface and let $n(x)$ be the number of the first free return. We shall arrange that $1 \leq n(x) \leq 3$.

We claim that most of the image of a standard curve consists of standard curves. The precise statements are given below. As it was explained in Section 5 we get more precise bounds for higher iterates since we are able to avoid non-isolated returns.

Proposition 6.4. (Invariance) (a) $\bar{P}(\gamma) = \bigcup_j \gamma_j$ where n is continuous on $\bar{P}^{-1}\gamma_j$ and $(\gamma_j, \bar{\rho}/c_j)$ is a standard pair where $\bar{\rho}$ is the induced density on $\bar{P}\gamma$ and $c_j = P_\ell(\bar{P}^{-1}\gamma_j)$
 (b) $\mathbb{P}_\ell(n(x) > 1) \leq \text{Const}\varepsilon^{1/4}$.
 (c) The time before the first free return is

$$\sum_{j=0}^{n(x)-1} \frac{\phi(P^{j+1}x) - \phi(P^jx)}{p(I)} + o(1)$$

(of course, the angle difference is measured counterclockwise).

\bar{P} is not defined on a set $Z_1 \cup Z_2$ where Z_1 consists of the points which got stuck and Z_2 consists of captured points.

(d) $\mathbb{P}_\ell(Z_1) \leq \varepsilon^{7/16-\delta/4}$

(e) $\mathbb{P}_\ell(Z_2) \leq \text{Const}\sqrt{\varepsilon}$

(f) **(Hyperbolicity)** There exists $\tilde{\delta} > 0$ such that $\|d\bar{P}^{-1}\| \leq \text{Const}\varepsilon^{\tilde{\delta}}$.

Observe that *a priori* bounds (d) and (e) look unsatisfactory since they do not preclude that all points get eliminated during the first

$1/\sqrt{\varepsilon}$ iterations. However already $\bar{P}^2\gamma$ is sufficiently well distributed to improve these bounds.

Proposition 6.5. (Equidistribution)

(a) Let A be a C^1 function and $\ell = (\gamma, \rho)$ be a standard pair. Let (\tilde{J}, \tilde{E}) be a point on γ . Then

$$\mathbb{E}_\ell(A \circ \bar{P}) = \int_0^1 A(\tilde{J}, E) dE + \mathcal{O}(\varepsilon^{\tilde{\delta}})$$

(b) $\mathbb{P}_\ell(\bar{P}^2x \text{ gets captured before the next free return to } S) = \sqrt{\varepsilon}M(\tilde{J})(1+o(1))$.

(c) $\mathbb{P}_\ell(\bar{P}^2x \text{ gets stuck before the next free return to } S) \leq \text{Const} \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}$.

Proposition 6.4 is proved in Section 9 and Proposition 6.5 is proved in Section 10.

6.3. Short time evolution. As we explained in Section 5 Proposition 6.5 allows us to obtain the Poisson Limit Theorem and the Law of Large Numbers (statements (I) and (II) of Section 5) if the number of iterations is much smaller than $1/\sqrt{\varepsilon}$. Let us formulate the precise result used in the proof of Theorem 1.

We define inductively a sequence of curves $\{\gamma_{kj}\}$ such that $\bigcup_j \gamma_{kj}$ lie on the part of the orbit of γ not captured in the resonance after k steps. Put $\gamma_{01} = \gamma$. If $\{\gamma_{kj}\}$ is already defined up to some k , decompose

$$\bar{P}\gamma_{kj} = \bigcup_l \gamma_{kjl} \bigcup Z_{1kj} \bigcup Z_{2kj}$$

as in Proposition 6.4 and let $\{\gamma_{(k+1)j}\}$ be some reindexing of $\{\gamma_{kjl}\}$. Let $\Gamma_k \subset \gamma$ be the set of points having representatives $\bigcup_j \gamma_{kj}$. For $x \in \Gamma_k$ let $x_k = (J_k, E_k)$ denote the representative of $\gamma_{kj}(x)$. Let Γ'_k denote the set of points captured before time k and Γ''_k denote the set of points which got stuck at m -th free return for some $m \leq k$. Fix a small number δ_1 . Let $N = \lceil \delta_1 q / \sqrt{\varepsilon} \rceil$.

Proposition 6.6. *There exist a function $\delta_2(\delta_1)$ such that $\lim_{\delta_1 \rightarrow 0} \delta_2 = 0$ and a subset $\Gamma'''_N \subset \gamma$ such that*

$$(a) \quad \mathbb{P}_\ell(\Gamma'''_N) = o(\delta_1),$$

$$(b) \quad |I_N - I_0 - \delta_1 \Psi(I_0)| \leq \delta_1 \delta_2$$

for all $x \in \Gamma_N - \Gamma'''_N$,

(c) Let $t^{(0)}(x)$ be the time between x_0 and x_N then

$$\left| \sqrt{\varepsilon} t^{(0)}(x) - \frac{\delta_1}{p(I_0)} \right| \leq \delta_1 \delta_2$$

for all $x \in \Gamma_N - \Gamma_N'''$,

$$(d) \quad \left| \mathbb{P}_\ell(\Gamma_N') - \frac{\delta_1 \lambda(I_0)}{p(I_0)} \right| \leq \delta_1 \delta_2,$$

$$(e) \quad \mathbb{P}_\ell(\Gamma_N'') \leq \delta_1 \delta_2.$$

Theorem 1 is derived from Proposition 6.6 in Section 7. In Section 8 we explain how Proposition 6.6 follows from Propositions 6.4, 6.5 and Lemma 6.3.

7. CONVERGENCE.

Proof of Theorem 1. Let us first prove this result in the case then $(I, \phi, \theta)(0)$ are distributed according to some standard pair ℓ . Let $\Gamma^{(1)} = \Gamma_N - \Gamma_N'''$. Let $\{\gamma_j^{(1)}\}$ be the union of Γ_{N_j} not falling into Γ_N''' . Repeat the procedure described above with each $\gamma_j^{(1)}$ instead of γ and let $\Gamma^{(2)}$ be union of the resulting sets. Continue inductively to obtain a nested sequence

$$\gamma \supset \Gamma^{(1)} \supset \dots \supset \Gamma^{(n-1)} \supset \Gamma^{(n)} \supset \dots$$

Let $\{\gamma_j^{(n)}\}$ denote the set of the resulting curves. For $x \in \Gamma^{(n)}$ let $x^{(n)} = (I^{(n)}, E^{(n)})$ be the point from the orbit of x on $\gamma_j^{(n)}$ and let $t^{(n-1)}$ denote the time between the returns of $x^{(n-1)}$ and $x^{(n)}$. Set $\tau_n(x) = \sum_{m=0}^{n-1} t^{(m)}(x)$. By Proposition 6.6

$$|I^{(n+1)} - I^{(n)} - \delta_1 \Psi(I^{(n)})| \leq \delta_1 \delta_2, \quad \left| \sqrt{\varepsilon} [\tau_{n+1} - \tau_n] - \frac{\delta_1}{p(I^{(n)})} \right| \leq \delta_1 \delta_2.$$

Let $(\tilde{J}, \tilde{\tau})$ be the solution of

$$\tilde{J}' = \Psi(\tilde{J}), \quad \tau' = 1/p.$$

Then

$$(I^{(\lfloor t/\delta_1 \rfloor)}, \tau^{(\lfloor t/\delta_1 \rfloor)}) = (\tilde{J}, \tilde{\tau})(t) + o(1), \text{ as } \delta_1, \delta_2 \rightarrow 0.$$

This describes the dynamics of the points from $\Gamma^{(\lfloor T/\delta_1 \rfloor)}$. Next define $\Gamma^{(n)'}$, $\Gamma^{(n)''}$ and $\Gamma^{(n)'''}$ similarly to Γ_N' , Γ_N'' and Γ_N''' . It remains to show that Γ'' and Γ''' have small probability and to compute the asymptotics of Γ' (captured trajectories). First, by induction

$$\mathbb{P}_\ell(\Gamma^{(n)''} - \Gamma^{(n-1)'}) \leq \delta_1 \delta_2 n, \quad \mathbb{P}_\ell(\Gamma^{(n)'''} - \Gamma^{(n-1)''}) \leq \delta_1 \delta_2 n$$

which implies that $\Gamma^{(n)''}$ and $\Gamma^{(n)'''}$ have small measure for $n \leq T/\delta_1$. Denote $\mathbf{p}_n = \mathbb{P}_\ell(\Gamma^{(n)'})$. Then the foregoing discussion and Proposition 6.6(d) imply

$$(7.1) \quad \mathbf{p}_{n+1} - \mathbf{p}_n = \delta_1(1 - \mathbf{p}_n)\lambda(\tilde{J}(n\delta_1))(1 + o_{\delta_1 \rightarrow 0}(1)).$$

Letting $\delta_1 \rightarrow 0$ we obtain (3.2). This completes the proof of Theorem 1 for (I, ϕ, θ) being chosen according to \mathbb{P}_ℓ . Next, let γ_α be the first visit of

$$\sigma_\alpha = \{I = I_0, \phi = \alpha\}$$

to \mathbb{S} and let ρ_α denote the measures on γ_α which is the image of the uniform measure on σ_α . Write $\ell_\alpha = (\gamma_\alpha, \rho_\alpha)$. Applying the foregoing discussion to

$$\mu = \int_0^1 \mathbb{P}_{\ell_\alpha} d\alpha$$

we obtain the original statement of Theorem 1. \square

Remark. In fact, the proof gives a stronger result. Namely, randomness in θ alone is sufficient to obtain Markovian evolution.

8. SHORT TIME DYNAMICS.

Proof of Proposition 6.6. Given γ let

$$b(x) = \frac{I(Px) - I(x)}{\sqrt{\varepsilon}},$$

$$\bar{b}_\gamma(x) = \sum_{j=0}^{n(x)-1} b(P^j x), \quad \bar{b}(x_k) = \bar{b}_{\gamma_{k_j(x)}}(x_k).$$

We extend $\bar{b}(x_k) = \Psi_k$ (here k is understood mod q) if the trajectory of x gets stuck or captured for some $m \leq k$. Likewise we define $n(x_k) = n_{\gamma_{k_j(x)}}(x_k)$ and extend it to 1 if x_k is not defined. Define

$$\tilde{\lambda}_j(I) = \sum_k \frac{M_{jk}(I)_+}{|L_j(I)|}$$

where the sum is over all saddles of (2.8) on \mathcal{S}_j . Write $\bar{\Psi}(x) = \Psi_j$, $\tilde{\lambda}(x) = \tilde{\lambda}_j$ if $x \in \mathcal{S}_j$.

Let $\tilde{\ell} = (\tilde{\gamma}, \tilde{\rho})$ be a standard pair such that with $\tilde{\gamma} \subset \bar{P}^k \gamma$ and $\tilde{\rho}$ is a normalized induced density. Then by Proposition 6.4 (b)

$$\mathbb{E}_{\tilde{\ell}}(|n(x) - 1|) = \mathcal{O}(\varepsilon^{1/4}).$$

Next, let ℓ_{kj} denote the pair $(\gamma_{kj}, \rho_\gamma(x_0)/c_{kj})$ where $c_{kj} = \mathbb{P}_\ell(x_k \in \gamma_{kj})$. Then

$$\mathbb{E}_\ell(|n(x_k) - 1|) \leq \sum_j c_{kj} \mathbb{E}_{\ell_{kj}}(|n(x) - 1|) \leq \text{Const} \varepsilon^{1/4} \sum_j c_{kj} \leq \text{Const} \varepsilon^{1/4}.$$

Therefore

$$\mathbb{E}_\ell \left(\left| \sum_{k=0}^{N-1} n(x_k) - N \right| \right) = \mathcal{O}(N \varepsilon^{1/4}).$$

Hence

$$\mathbb{P}_\ell \left(\left| \sum_{k=0}^{N-1} n(x_k) - N \right| > \frac{\delta_2 N}{2} \right) \leq \frac{2 \varepsilon^{1/4}}{\delta_2}.$$

Since $t_k = n(x_k)/p(I_k) + o(\delta_2)$ we get using Proposition 6.1(a)

$$\mathbb{P}_\ell \left(\left| \sum_{k=0}^{N-1} t_k(x) - \frac{N}{p(I_0)} \right| > \delta_2 N \right) \leq \frac{\text{Const} \varepsilon^{1/4}}{\delta_2}.$$

Including the part of γ_{Nj} where the above sum is larger than $\delta_2 N$ in Γ_N''' we obtain part (c) of Proposition 6.6. Next let

$$\hat{I}_n = I_0 + \frac{\sqrt{\varepsilon} n}{q} \Psi(I_0), \quad \Delta_n = I_n - \hat{I}_n.$$

We have

$$\begin{aligned} \Delta_{nq} &= \sqrt{\varepsilon} \left[\sum_{j=0}^{n-1} \sum_{r=0}^{q-1} (\bar{b}(x_{jq+r}) - \bar{\Psi}(x_{jq+r})) \right. \\ &\quad \left. + \left(\sum_{j=0}^{n-1} (\Psi(I_{jq}) - \Psi(\hat{I}_{jq})) \right) + \left(\sum_{j=0}^{n-1} (\Psi(\hat{I}_{jq}) - \Psi(I_0)) \right) \right] \\ (8.1) \quad &= \sqrt{\varepsilon} \left[\sum_{j=0}^{n-1} \sum_{r=0}^{q-1} (\bar{b}(x_{jq+r}) - \bar{\Psi}(x_{jq+r})) + \left(\sum_{j=0}^{n-1} \mathcal{O}(|I_{jq} - \hat{I}_{jq}|) \right) + \mathcal{O}(n^2 \sqrt{\varepsilon}) \right]. \end{aligned}$$

Lemma 8.1.

$$(8.2) \quad \mathbb{E}_\ell \left(\left(\sum_{k=0}^{n-1} [\bar{b}(x_k) - \bar{\Psi}(x_k)] \right)^2 \right) \leq \text{Const} n^2 \delta_1^{20}.$$

Therefore

$$\mathbb{E}_\ell \left(\left| \sum_{k=0}^{n-1} [\bar{b}(x_k) - \bar{\Psi}(x_k)] \right| \right) \leq \text{Const} n \delta_1^{10}.$$

Proof. We have

$$\begin{aligned} & \mathbb{E}_\ell \left(\left(\sum_k [\bar{b}(x_k) - \bar{\Psi}(x_k)] \right)^2 \right) \\ &= \mathbb{E}_\ell \left(\sum_{k_2 - k_1 > 2} ([\bar{b}(x_{k_1}) - \bar{\Psi}(x_{k_1})][\bar{b}(x_{k_2}) - \bar{\Psi}(x_{k_2})]) \right) + \mathcal{O}(n). \end{aligned}$$

Let $k_3 = \lfloor \frac{k_1 + k_2}{2} \rfloor$. By Proposition 4(f) the map $x_{k_1} \rightarrow x_{k_3}$ expands distances by at least $\varepsilon^{-\bar{\delta}}$. We claim that this implies that for each standard pair $\ell_{k_3j} = (\gamma_{k_3j}, \rho_{k_3j})$ such that $\gamma_{k_3j} \subset \bar{P}^{k_3}\gamma$ there exists a number ζ_{k_3j} such that

$$(8.3) \quad |\bar{b}(x_{k_1}) - \bar{\Psi}(x_{k_1}) - \zeta_{k_3j}| \leq \delta_1^{20}$$

uniformly on each γ_{k_3j} . Indeed since the preimages of γ_{k_3j} under \bar{P} have lengths $\mathcal{O}(\varepsilon^{\bar{\delta}})$ and since b is well approximated by σ which is continuous away from the separatrix of the inner system we only have to establish (8.3) in case x_{k_1} passes near the separatrix. To analyze this case we assume that $n(x_{k_1}) = 1$ (other possibilities are similar). Then by Proposition 6.2

$$\left| \frac{d\sigma(x_k)}{dE_k} \right| \leq \frac{\text{Const}}{d}$$

whereas

$$\left| \frac{dE_{k+1}}{dE_k} \right| \geq \frac{\text{Const}}{d\sqrt{\varepsilon}}$$

so the preimage of $\gamma_{k+1,j(x)}$ has E_k -length $\mathcal{O}(d\sqrt{\varepsilon})$ and the oscillations of b on such interval are $\mathcal{O}(\sqrt{\varepsilon}) \ll 1$ verifying (8.3).

On the other hand applying Proposition 6.5(a) for each j we have

$$\mathbb{E}_{\ell_{k_3j}} (\zeta_{k_3} (\bar{b}(x_{k_2}) - \bar{\Psi}_{k_2}(x_{k_2}))) = \mathcal{O}(\varepsilon^{\bar{\delta}}).$$

Summation over j gives

$$\mathbb{E}_\ell (\zeta_{k_3} (\bar{b}(x_{k_2}) - \bar{\Psi}_{k_2}(x_{k_2}))) = \mathcal{O}(\varepsilon^{\bar{\delta}}).$$

Combining this with (8.3) we obtain

$$\mathbb{E}_\ell (([\bar{b}(x_{k_1}) - \bar{\Psi}(x_{k_1})][\bar{b}(x_{k_2}) - \bar{\Psi}(x_{k_2})])) = \mathcal{O}(\delta_1^{20}).$$

proving (8.2). □

Next, since $\mathbb{E}_\ell(|\bar{b}(x_k)|) = \mathcal{O}(1)$ we get an a priori bound

$$(8.4) \quad \mathbb{E}_\ell(|I_k - I_0|) \leq \text{Const}k\sqrt{\varepsilon}$$

and hence

$$(8.5) \quad \mathbb{E}_\ell(|I_k - \hat{I}_k|) \leq \text{Const} k \sqrt{\varepsilon}.$$

Plugging (8.5) into (8.1) we get

$$(8.6) \quad \mathbb{E}_\ell(|\Delta_k|) = \mathcal{O}(\delta_1^2 \sqrt{\varepsilon} + k^2 \varepsilon)$$

for $k \leq N$. Plugging (8.6) back into (8.1) we get

$$\Delta_n = \mathcal{O}(\delta_1^2) + \tilde{\Delta}_n$$

where

$$\mathbb{E}_\ell(|\tilde{\Delta}_N|) = \mathcal{O}(\delta_1^3).$$

Adding to Γ_N'''' the set of points where $|\tilde{\Delta}_1| > \delta_1^{3/2}$ we obtain parts (a) and (b) of Proposition 6.6.

Next, using Proposition 6.5(c) we get by induction that for $m \geq 3$

$$\mathbb{P}_\ell(\Gamma_m'' - \Gamma_{m-1}'') \leq \text{Const} \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}.$$

Combining this with parts (d) and (e) of Proposition 6.4 for $m < 3$ we get (e).

Let now $Q_k = \mathbb{P}_\ell(\Gamma_k')$. Then

$$Q_{k+1} - Q_k = \sum_j c_{k-2,j} \mathbb{P}_{\ell_{k-1,j}}(\Gamma_{k+1}' - \Gamma_k')$$

where $c_{k,j} = \mathbb{P}_\ell(\bar{P}^{-k} \gamma_{kj})$. Using Proposition 6.5(b) we get

$$Q_{k+1} - Q_k = \left(\sum_j c_{k-2,j} \right) \sqrt{\varepsilon} \mathbb{E}_\ell(\tilde{\lambda}(x_k))(1 + o(1)).$$

By already proven parts (a) and (e) of Proposition 6.6

$$\sum_j c_{k-2,j} = 1 - Q_{k-2} + o(\delta_1)$$

so

$$(8.7) \quad Q_{k+1} - Q_k = (1 - Q_{k-2}) \sqrt{\varepsilon} \mathbb{E}_\ell(\tilde{\lambda}(x_k))(1 + o(\delta_1)).$$

This gives *a priori* bounds

$$(8.8) \quad Q_{k+1} - Q_k = \mathcal{O}(\sqrt{\varepsilon}) \quad \text{and} \quad Q_k = \mathcal{O}(\delta_1).$$

Combining (8.7) and (8.8) we get

$$Q_{k+1} - Q_k = \sqrt{\varepsilon} \tilde{\lambda}(x_k)(1 + \mathcal{O}(\delta_1)).$$

Thus given r we obtain after the summation over the period

$$\sum_{k=r}^{r+q-1} Q_{k+1} - Q_k = \sqrt{\varepsilon} \lambda(I_0) (1 + \mathcal{O}(\delta_1)).$$

This proves part (d) of Proposition 6.6. \square

9. FREE RETURNS.

9.1. Definitions. Here we prove Proposition 6.4. First we need to give precise definitions of close returns and also to explain when we consider an orbit stuck. Let $(J, E) \rightarrow (\bar{J}, \bar{E})$ denote the Poincare map. Given a standard curve we write

$$k = \frac{dE}{dJ}, \quad q = \frac{d^2 E}{dJ^2}, \quad r = \frac{d}{dE} \ln \rho, \quad \bar{k} = \frac{d\bar{E}}{d\bar{J}}, \quad \bar{q} = \frac{d^2 \bar{E}}{d\bar{J}^2}, \quad \bar{r} = \frac{d}{d\bar{E}} \ln \bar{\rho}.$$

Close returns of the first kind are defined by the condition that the orbit of (J, E) is within distance $\sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}$ from \mathcal{N}_ε . Close returns of the second kind are more complicated to define. Indeed we need to satisfy several conditions (conditions (a)–(c) of the definition of the standard pair). We begin with expansion. Observe that

$$\frac{d\bar{E}}{dE} = \frac{\partial \bar{E}}{\partial E} + k \frac{\partial \bar{E}}{\partial J} = \frac{\partial \bar{E}}{\partial E} + \mathcal{O}(1)$$

by Proposition 6.1 and part (b1) of the definition of the standard pair. Hence expansion may fail only near zeroes of $\frac{\partial \bar{E}}{\partial E}$. By Proposition 6.1 if c is sufficiently small then these zeroes are close to the critical points of σ . Proposition 6.1 and Assumption (F) now imply that if $\left| \frac{\partial \bar{E}}{\partial E} \right| \leq \delta \varepsilon^{-3/8}$

then $\left| \frac{\partial^2 \bar{E}}{\partial E^2} \right| \geq \text{Const}/\sqrt{\varepsilon}$. Also

$$\frac{d^2 \bar{E}}{dE^2} = \frac{\partial^2 \bar{E}}{\partial E^2} + 2k \frac{\partial^2 \bar{E}}{\partial E \partial J} + k^2 \frac{\partial^2 \bar{E}}{\partial J^2} + q \frac{\partial \bar{E}}{\partial J} = \frac{\partial^2 \bar{E}}{\partial E^2} + \mathcal{O}(\varepsilon^{-(1/2-\delta)}).$$

Hence $\left| \frac{d^2 \bar{E}}{dE^2} \right| \geq \frac{\text{Const}}{\sqrt{\varepsilon}}$ near possible zeroes of $\frac{d\bar{E}}{dE}$. We now define close returns of the second kind by the condition $\left| \frac{d\bar{E}}{dE} \right| < \varepsilon^{-1/4}$. If a point has a close return of the first kind we declare it stuck and remove it from consideration. For the close returns of the second kind we track the orbit for two more iterations to see if it recovers the lost hyperbolicity. Namely if

$$\left| \frac{d\bar{E}}{dE} \right| < \varepsilon^\delta$$

we declare the point stuck. Otherwise let (\bar{J}, \bar{E}) and $(\bar{\bar{J}}, \bar{\bar{E}})$ be the images of (\bar{J}, \bar{E}) . If

$$\left| \frac{d\bar{E}}{dE} \right| < \varepsilon^{-1/16}$$

we declare the orbit stuck if

$$\left| \frac{d\bar{E}}{d\bar{E}} \right| < \varepsilon^{-(1/4+\delta)} \text{ or } \left| \frac{d\bar{\bar{E}}}{d\bar{\bar{E}}} \right| < \varepsilon^{-(1/4+\delta)}$$

or if (\bar{J}, \bar{E}) or $(\bar{\bar{J}}, \bar{\bar{E}})$ experience close returns of the first kind. Otherwise we declare the orbit free. If

$$(9.1) \quad \varepsilon^{-1/16} \leq \left| \frac{d\bar{E}}{dE} \right| < \varepsilon^{-1/8}$$

we declare the orbit stuck if

$$\left| \frac{d\bar{E}}{dE} \right| < \varepsilon^{-1/8} \text{ or } \left| \frac{d\bar{\bar{E}}}{d\bar{\bar{E}}} \right| < \varepsilon^{-(1/4+\delta)}$$

or if (\bar{J}, \bar{E}) or $(\bar{\bar{J}}, \bar{\bar{E}})$ experience close returns of the first kind. Otherwise we declare the orbit free. Finally if

$$\left| \frac{d\bar{E}}{dE} \right| \geq \varepsilon^{-1/8}$$

then we set an orbit free if

$$\left| \frac{d\bar{\bar{E}}}{d\bar{\bar{E}}} \right| > \varepsilon^{-1/4},$$

otherwise we declare it stuck.

9.2. Free orbits form standard pairs. We now show that these rules allow us to preserve standard pairs. First we consider the case when all the traversed resonances are weak and then describe the modifications needed to treat strong resonances. We begin with points without close returns. We have

$$(9.2) \quad \bar{k} = \frac{\frac{\partial \bar{J}}{\partial E} + k \frac{\partial \bar{J}}{\partial J}}{\frac{\partial \bar{E}}{\partial E} + k \frac{\partial \bar{E}}{\partial J}}.$$

Combining the identity

$$\frac{\partial \bar{J}}{\partial E} \frac{\partial \bar{E}}{\partial J} = \frac{\partial \bar{J}}{\partial J} \frac{\partial \bar{E}}{\partial E} + \mathcal{O}(1)$$

(valid by (2.18)) with Proposition 6.1 we get

$$(9.3) \quad \frac{\partial \bar{J}}{\partial E} = \mathcal{O}(\varepsilon) \frac{\partial \bar{E}}{\partial E} + \mathcal{O}(\varepsilon).$$

Thus on the set where $\left| \frac{\partial \bar{E}}{\partial E} \right|^{-1} = \mathcal{O}(1)$ we have

$$(9.4) \quad \frac{\frac{\partial \bar{J}}{\partial E} + k \frac{\partial \bar{J}}{\partial J}}{\frac{d\bar{E}}{dE}} = \mathcal{O}(\varepsilon).$$

so $\bar{k} = \mathcal{O}(\varepsilon)$ and (b1) follows if C_1 is large enough. Next

$$(9.5) \quad \bar{q} = \frac{\frac{\partial^2 \bar{J}}{\partial E^2} + 2k \frac{\partial^2 \bar{J}}{\partial J \partial E} + k^2 \frac{\partial^2 \bar{J}}{\partial J^2}}{\left(\frac{d\bar{E}}{dE} \right)^2} - \frac{\left(\frac{\partial \bar{J}}{\partial E} + k \frac{\partial \bar{J}}{\partial J} \right) \left(\frac{\partial^2 \bar{E}}{\partial E^2} + 2k \frac{\partial^2 \bar{E}}{\partial J \partial E} + k^2 \frac{\partial^2 \bar{E}}{\partial J^2} \right)}{\left(\frac{d\bar{E}}{dE} \right)^3} \\ + q \left[\frac{\frac{\partial \bar{J}}{\partial J}}{\left(\frac{d\bar{E}}{dE} \right)^2} - \frac{\frac{d\bar{J}}{dE} \frac{\partial \bar{E}}{\partial J}}{\left(\frac{d\bar{E}}{dE} \right)^3} \right] = (I) + (II) + (III).$$

Now

$$|I| = \frac{\mathcal{O}(\sqrt{\varepsilon})}{\left| \frac{\partial \bar{E}}{\partial E} \right|^2} + \frac{\mathcal{O}(\varepsilon^{5/4})}{\left| \frac{\partial \bar{E}}{\partial E} \right|} + \mathcal{O}(\varepsilon^{5/4}) = \mathcal{O}(\varepsilon).$$

Next by (9.4)

$$(II) = \mathcal{O} \left(\varepsilon \times \frac{\frac{\partial^2 \bar{E}}{\partial E^2} + 2k \frac{\partial^2 \bar{E}}{\partial J \partial E} + k^2 \frac{\partial^2 \bar{E}}{\partial J^2}}{\left(\frac{d\bar{E}}{dE} \right)^2} \right).$$

The second factor here is

$$\frac{\mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)}{\left| \frac{\partial \bar{E}}{\partial E} \right|^2} + \frac{\mathcal{O}(\varepsilon^{1/4})}{\left| \frac{\partial \bar{E}}{\partial E} \right|} + \mathcal{O}(\varepsilon^{1/4}) = \mathcal{O}(1)$$

so $II = \mathcal{O}(\varepsilon)$. Finally using (9.3) we get

$$(III) = q \times \left[\mathcal{O} \left(\frac{1}{\left| \frac{\partial \bar{E}}{\partial E} \right|^2} \right) + \mathcal{O} \left(\frac{\frac{\partial \bar{E}}{\partial J}}{\left| \frac{\partial \bar{E}}{\partial E} \right|^2} \varepsilon \right) \right] = \mathcal{O}(\varepsilon^{1+\delta}).$$

Thus

$$(9.6) \quad \bar{q} = \mathcal{O}(\varepsilon).$$

Condition (a) is automatic since we delete points which come close to critical set and the distance between those points is of order 1. Next,

$$\bar{\rho}(\bar{E}) = \left(\frac{\rho(E)}{\frac{d\bar{E}}{dE}} \right) / \bar{c}$$

where \bar{c} is the normalization constant. Hence

$$(9.7) \quad \bar{r} = \frac{r}{\left(\frac{d\bar{E}}{dE}\right)} - \frac{\frac{d^2\bar{E}}{dE^2}}{\left(\frac{d\bar{E}}{dE}\right)^2} = \frac{r}{\left(\frac{d\bar{E}}{dE}\right)} + \mathcal{O}(\varepsilon^{1/4}) + \mathcal{O}\left(\frac{1/\sqrt{\varepsilon}}{\left(\frac{d\bar{E}}{dE}\right)^2}\right).$$

The first term is $\mathcal{O}(\varepsilon^{1/4})$ and the last is

$$\mathcal{O}\left(\frac{1/\sqrt{\varepsilon}}{(\varepsilon^{-1/4})^2}\right) = \mathcal{O}(1)$$

so (c) follows if C_3 is large enough.

We now come to the close returns of the second kind. Consider the case $\left|\frac{d\bar{E}}{dE}\right| \geq \varepsilon^{-1/8}$ first. In this case (9.4) gives $\bar{k} = \mathcal{O}(\varepsilon)$. Next (9.5) and (9.4) give $\bar{q} = \mathcal{O}(\varepsilon^{3/4})$ and (9.7) implies $\mathcal{O}(\bar{r}) = \varepsilon^{-1/4}$. Thus if the orbit brakes free on the second step then we have

$$\bar{k} = \mathcal{O}(\varepsilon), \quad \bar{q} = \mathcal{O}(\varepsilon), \quad \bar{r} = \mathcal{O}(1)$$

verifying (a)–(c).

Next if $\varepsilon^{-1/16} \leq \left|\frac{d\bar{E}}{dE}\right| \leq \varepsilon^{-1/8}$ then we get

$$\bar{k} = \mathcal{O}(\varepsilon), \quad \bar{q} = \mathcal{O}(\varepsilon^{5/8}), \quad \bar{r} = \mathcal{O}(\varepsilon^{-3/8})$$

If the orbit survives the next steps we have

$$(9.8) \quad \bar{k} = \mathcal{O}(\varepsilon), \quad \bar{q} = \mathcal{O}(\varepsilon^{3/4}), \quad \bar{r} = \mathcal{O}(\varepsilon^{-1/4}),$$

$$\bar{\bar{k}} = \mathcal{O}(\varepsilon), \quad \bar{\bar{q}} = \mathcal{O}(\varepsilon), \quad \bar{\bar{r}} = \mathcal{O}(1).$$

In case $\varepsilon^\delta < \frac{d\bar{E}}{dE} < \varepsilon^{-1/16}$ these estimates take the following form

$$(9.9) \quad \bar{k} = \mathcal{O}(\varepsilon^{1-2\delta}), \quad \bar{q} = \mathcal{O}(\varepsilon^{1/2-4\delta}), \quad \bar{r} = \mathcal{O}(\varepsilon^{-(1/2+2\delta)})$$

$$\bar{k} = \mathcal{O}(\varepsilon), \quad \bar{q} = \mathcal{O}(\varepsilon^{1-2\delta}), \quad \bar{r} = \mathcal{O}(\varepsilon^{-(1/4+\delta)}),$$

$$\bar{\bar{k}} = \mathcal{O}(\varepsilon), \quad \bar{\bar{q}} = \mathcal{O}(\varepsilon), \quad \bar{\bar{r}} = \mathcal{O}(1).$$

This proves part (a) of Proposition 6.4. Also, part (f) follows because at each step we have $\left|\frac{d\bar{E}}{dE}\right| > \varepsilon^\delta$ and on the last step the expansion is at least $\varepsilon^{-1/4}$. Part (c) follows from Proposition 6.1(a). This concludes the proof of parts (a)–(c) of Proposition 6.5 in case all three resonances γ has to traverse are weak. The case where some of the resonances are

strong requires minimal modifications. First we have to remove the points coming too close to \mathcal{N}_ε . Since the distance between the saddles of (2.8) and the zeroes of $\frac{\partial \bar{E}}{\partial E}$ is of order 1 by (6.1) and assumption (E) we still have uniform lower bounds on the image lengths. For the estimates of k , q and r we claim that if at some point $|\frac{\partial \bar{E}}{\partial E}| > \varepsilon^{-1/4}$ then the worst case is still when $|\frac{\partial \bar{E}}{\partial E}| \sim \varepsilon^{-1/4}$. Indeed (9.4) holds near \mathcal{N}_ε by Proposition 6.2. Thus comparing what happens near \mathcal{N}_ε with points near parabolic pieces we see that for terms having $\frac{d\bar{E}}{dE}$, $(\frac{d\bar{E}}{dE})^2$ and $(\frac{d\bar{E}}{dE})^3$ in the denominator the denominator is multiplied by $\frac{1}{d\varepsilon^{1/4}}$, $\frac{1}{d^2\varepsilon^{1/2}}$ and $\frac{1}{d^3\varepsilon^{3/4}}$ respectively whereas the numerator increases by $\frac{|\ln^m d|}{d}$, $\frac{\ln^{2m} d}{d^2}$ and $\frac{|\ln^{3m} d|}{d^3}$ respectively so the situation is much better. It remains to prove parts (d) and (e) of Proposition 6.4.

9.3. Measure estimates. Here we complete the proof of Proposition 6.4. Again we first analyze the case when all resonances are weak and describe the modifications needed for strong resonances at the end of this section. In this section we shall make a frequent use of two estimates. The first is a standard distortion bound (see Lemma H.1 of Appendix H). The second is (9.13) below.

Let E_0 be a point where $\frac{d\bar{E}}{dE} = 0$. Then near E_0

$$(9.10) \quad \frac{\hat{c}_1}{\sqrt{\varepsilon}} \leq \left| \frac{d^2 \bar{E}}{dE^2} \right| \leq \frac{\hat{c}_2}{\sqrt{\varepsilon}}.$$

Hence in the region where

$$(9.11) \quad \left| \frac{d\bar{E}}{dE} \right| \leq \delta \varepsilon^{-3/8}$$

we get

$$(9.12) \quad \frac{c'_1}{\sqrt{\varepsilon}} |E - E_0| \leq \left| \frac{d\bar{E}}{dE} \right| \leq \frac{c'_2}{\sqrt{\varepsilon}} |E - E_0|.$$

For these values of E we have

$$(9.13) \quad \frac{c''_1}{\sqrt{\varepsilon}} (E - E_0)^2 \leq |\bar{E} - \bar{E}_0| \leq \frac{c''_2}{\sqrt{\varepsilon}} (E - E_0)^2.$$

It follows from (9.12) that the close returns of the second kind have measure $\mathcal{O}(\varepsilon^{1/4})$. We now estimate the measure of Z_1 . It consists of several parts. Points coming too close on the first step have measure

$$(9.14) \quad \mathbb{P}_\ell(Z_I) \leq \text{Const} \varepsilon^{1/2+\delta}.$$

by (9.12) (since ρ is uniformly bounded above and below by parts (a) and (c) of the definition of the standard pair). Next there are points

which satisfy $\varepsilon^\delta < \frac{d\bar{E}}{dE} < \varepsilon^{-1/16}$. In the worst case scenario all those points will be removed at the second step so we can not do better than estimate the measure of points removed from this part by their total measure, that is

$$(9.15) \quad \mathbb{P}_\ell(Z_{II}) \leq \text{Const}\varepsilon^{7/16}$$

(see (9.12)). Next, consider points satisfying (9.1). They satisfy

$$|\bar{E} - \bar{E}_0| \leq \text{Const}$$

and if these points are removed on the second step then (9.12) and (9.13) show that \bar{E} should be close to the zeroes of $\partial\bar{E}/\partial\bar{E}$. Hence on the second step we remove finitely many intervals of \bar{E} -length $\mathcal{O}(\varepsilon^{3/8})$ (till the end of this section we use the phrase 'finitely many' to mean that the corresponding number is uniformly bounded as $\varepsilon \rightarrow 0$). Next (9.12) and (9.13) show that the induced density (which we denote $\tilde{\rho}$ to distinguish it from the conditional density $\bar{\rho}$) satisfies

$$\tilde{\rho}(E) = \frac{\rho(E)}{\left|\frac{d\bar{E}}{dE}\right|} \sim \frac{\varepsilon^{1/4}}{\sqrt{|\bar{E} - \bar{E}_0|}}.$$

Since $\frac{1}{\sqrt{|\bar{E} - \bar{E}_0|}}$ is increasing towards \bar{E}_0 the worst case is when one of whose intervals contains E_0 . In this case the probability Δ of getting stuck can be estimated with the help of (9.13). We get

$$\frac{\Delta^2}{\sqrt{\varepsilon}} = \text{Const}\varepsilon^{3/8}$$

that is

$$(9.16) \quad \mathbb{P}_\ell(Z_{III}) \leq \text{Const}\varepsilon^{7/16}$$

On the third step we have to remove finitely many critical intervals (where $\left|\frac{\partial\bar{E}}{\partial E}\right| \leq \varepsilon^{-1/4}$) and several non-critical intervals.

Let us first estimate how much we remove from the critical intervals. Reasoning as above we see that on each interval we have to remove a set of \bar{E} -length $\mathcal{O}(\varepsilon^{1/4-\delta})$ hence its \bar{E} -length is $\mathcal{O}(\varepsilon^{3/8-\delta/2})$ and E -length is $\mathcal{O}(\varepsilon^{7/16-\delta/4})$. Thus

$$(9.17) \quad \mathbb{P}_\ell(Z_{IV}) \leq \text{Const}\varepsilon^{\frac{7}{16}-\frac{\delta}{4}}.$$

Now consider non-critical intervals. Recall constants c_1, c_2 from the definition of the standard pair. Decompose

$$\left\{ \varepsilon^{-1/16} < \left|\frac{\partial\bar{E}}{\partial E}\right| < \varepsilon^{-1/8}, \quad \left|\frac{\partial\bar{E}}{\partial E}\right| \geq \varepsilon^{-1/4} \right\} = \bigcup_{\alpha} Y_{\alpha}$$

where $\text{length}(PY_\alpha) < c_2$ and either $\text{length}(PY_\alpha) > c_1$ or Y_α is adjacent to one of the interval removed at the first two steps. If the first alternative holds we call Y_α complete, otherwise we call it incomplete. Let $Z_\alpha = P(Y_\alpha)$.

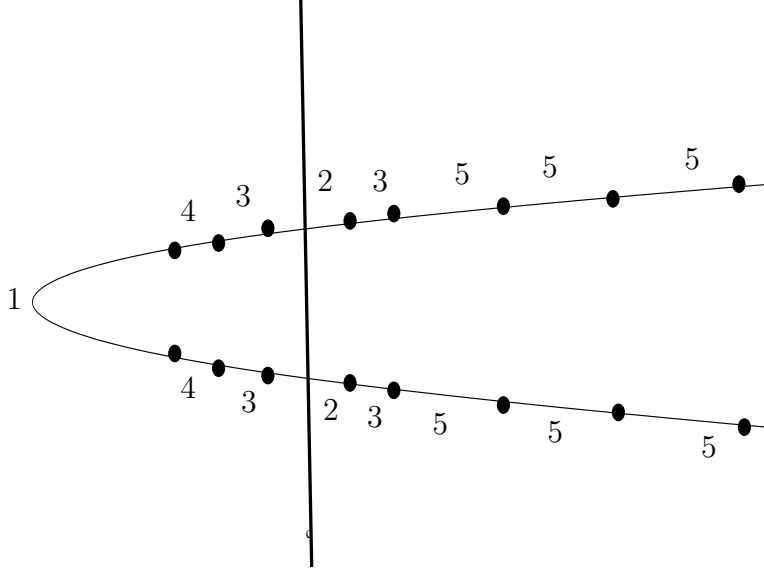


FIGURE 6. Thick line stands for \mathcal{C} . 1 is the interval removed at the first step, 2 are the intervals removed at the second step, 3 are critical intervals, 4 are incomplete intervals, 5 are complete intervals.

By the estimates of Section 9.2, P has bounded distortion on Y_α (since Z_α are non-critical). We remove a set of \bar{E} -measure $\mathcal{O}(\varepsilon^{1/4-\delta})$ from each Z_α . This set has \bar{E} -measure $\mathcal{O}(\varepsilon^{1/4-\delta}|Y_\alpha|)$. Taking the union of all α s we get a set of \bar{E} -measure $\mathcal{O}(\varepsilon^{1/4-\delta}|\bar{E}_1 - \bar{E}_0|)$ where $\frac{\partial \bar{E}}{\partial E}(E_1) = \varepsilon^{-1/8}$. Thus the total \bar{E} -measure of the deleted set is $\mathcal{O}(\varepsilon^{3/8}\varepsilon^{1/4-\delta}) = \mathcal{O}(\varepsilon^{5/8-\delta})$. Since P expands by at least $\varepsilon^{-1/16}$ on the set where (9.1) holds the total measure deleted from the complete intervals is

$$(9.18) \quad \mathbb{P}_\ell(Z_V) \leq \text{Const}\varepsilon^{11/16-\delta}.$$

There are also finitely many incomplete intervals. Since the minimal extension on the second step is $\text{Const}\varepsilon^{-1/4}$ for non-critical intervals we conclude that \bar{E} -length of the removed set is $\mathcal{O}(\varepsilon^{1/2-\delta})$ and since the expansion of $E \rightarrow \bar{E}$ is at least $\varepsilon^{-1/16}$ we get

$$(9.19) \quad \mathbb{P}_\ell(Z_{VI}) \leq \text{Const}\varepsilon^{9/16-\delta}.$$

Finally we consider points with $\left| \frac{d\bar{E}}{dE} \right| > \varepsilon^{-1/8}$. We remove intervals of \bar{E} -length $\mathcal{O}(\varepsilon^{3/8})$ and since we have expansion of at least $\varepsilon^{-1/8}$ the bound we obtain is

$$(9.20) \quad \mathbb{P}_\ell(Z_{VII}) \leq \text{Const} \sqrt{\varepsilon}.$$

This completes the proof of Proposition 6.4(d) for weak resonances. In the presence of strong resonances there are additional complications.

(a) There are points which are removed on the first step due to returns of the first kind.

(b) Returns of the first kind could appear after the returns of the second kind.

(c) There are points which come close to \mathcal{N}_ε on the second step without getting stuck and we need to check that this does not destroy our estimates.

Case (a) contributes a set of measure $\mathcal{O}\left(\sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}\right)$ by Proposition 6.2(d).

To handle (b) observe that the probability to have a close return of the first kind on the second step is $\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/2})$ since we have to remove an interval of \bar{E} -length $\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/2})$ and E -length of the preimage of any such interval is $\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/4})$. On the third step we have to distinguish critical and non-critical cases. In the critical case we have a bound $\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/8})$ by the argument used for the second step. In the non-critical case we see using the bounded distortion property that we remove a set of \bar{E} -measure

$$\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/2} |\bar{E}_1 - \bar{E}_0|) = \mathcal{O}(\varepsilon^{3/4} |\ln \varepsilon|^{-1/2})$$

so the contribution of non-critical intervals is $\mathcal{O}(\varepsilon^{5/8} |\ln \varepsilon|^{-1/4})$.

Finally (c) is of no concern since the expansion near \mathcal{N}_ε is better than the expansion away from \mathcal{N}_ε and the bounded distortion property still holds so our estimates only become better.

It remains to prove Proposition 6.4(e). Again capture can occur either on the first step or after a return of the second kind. For immediate capture we use a bound $\mathcal{O}(\sqrt{\varepsilon})$ of Proposition 6.2(c). Captures on the second and the third steps are analyzed similarly to close returns of the first kind. The difference is that now we use Proposition 6.2(d) instead of 6.2(c) so we do not have powers of δ in the corresponding estimates. So instead of $\mathcal{O}(\varepsilon^{1/2+\delta/4})$ bound we get $\mathcal{O}(\varepsilon^{1/2})$ bound.

We also observe that this bound comes from possible captures in critical intervals. Namely for captures from non-critical intervals we

have $\mathcal{O}(\varepsilon^{5/8})$ bound corresponding to $\mathcal{O}(\varepsilon^{5/8+\delta/2})$ bound for the first kind returns.

Proposition 6.4 follows.

10. EQUIDISTRIBUTION.

10.1. Equidistribution on a unit scale. Here we prove Proposition 6.5. To prove (a) observe that if $|E - E_0| > \varepsilon^{1/4-\delta}$ then $\bar{r} < C\varepsilon^{2\delta}$. Divide the part of γ where $|E - E_0| > \varepsilon^{1/4-\delta}$ into segments $\{Y_\alpha\}$ so that the image of each segment has \bar{E} -length 1. Then by Lemma H.1 on each Y_α the map $E \rightarrow \bar{E}$ is $\varepsilon^{2\delta}$ close to linear. Since linear maps have required equidistribution properties part (a) of Proposition 6.5 follows.

10.2. Isolated returns. Our next task is to prove part (c) of Proposition 6.5 following the outline given in Section 5. We say that a standard curve γ has an isolated return if for all E_0 such that $\frac{d\bar{E}}{dE}(E_0) = 0$ we have

$$\left| \frac{\partial \bar{\bar{E}}}{\partial \bar{E}}(\bar{E}_0) \right| > \varepsilon^{-[(1/4)+4\delta]}.$$

Let us derive some consequences of the isolation. Let \mathcal{X} be a component consisting of points which have a close return of the second kind but are not removed at the first step. \mathcal{X} consists of three segments

$$\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_3 \text{ where } \mathcal{X}_1 = \left\{ \varepsilon^\delta \leq \left| \frac{d\bar{E}}{dE} \right| \leq \varepsilon^{-1/16} \right\},$$

$$\mathcal{X}_2 = \left\{ \varepsilon^{-1/16} < \left| \frac{d\bar{E}}{dE} \right| \leq \varepsilon^{-1/8} \right\}, \quad \mathcal{X}_3 = \left\{ \varepsilon^{-1/8} < \left| \frac{d\bar{E}}{dE} \right| \leq \varepsilon^{-1/4} \right\}.$$

Let (E_1, J_1) be the endpoint of \mathcal{X} in \mathcal{X}_1 . Then by (9.12) $|E_1 - E_0| \leq \text{Const}\varepsilon^{1/2+\delta}$. So by property (b1) of standard pairs the J -coordinate satisfies $|J_1 - J_0| \leq \text{Const}\varepsilon^{3/2+\delta}$. It follows from (9.13) that $|\bar{E}_1 - \bar{E}_0| \leq \text{Const}\varepsilon^{1/2+2\delta}$ and Proposition 6.1(d) gives $|\bar{J}_1 - \bar{J}_0| \leq \varepsilon^{1+\delta}$. It follows from the bounds on the second derivatives of \bar{E} given by Proposition 6.1 that

$$(10.1) \quad \left| \frac{\partial \bar{\bar{E}}}{\partial \bar{E}} \right| (E_1) \geq \frac{1}{2} \varepsilon^{-[1/4+4\delta]}.$$

Next let E_2 be the closest point to E_1 in \mathcal{X} such that $|\frac{d\bar{\bar{E}}}{d\bar{E}}(\bar{E}_2)| = \varepsilon^{-1/4}$. Then (9.9) implies that the arguments of Section 9.3 are available giving

$$(10.2) \quad |\bar{E}_2 - \bar{E}_1| \sim C\varepsilon^{1/4-4\delta}.$$

Now (9.13) implies that $|E_2 - E_1| \sim \text{Const}\varepsilon^{3/8-2\delta}$ and (9.12) gives

$$(10.3) \quad \left| \frac{d\bar{E}}{dE}(E_2) \right| \geq \text{Const}\varepsilon^{-1/8-2\delta}.$$

10.3. Getting stuck at isolated returns. Let \hat{Z}_1 denote the event that the orbit gets stuck due to the second kind return.

Lemma 10.1. *If γ has an isolated return then*

$$\mathbb{P}_\ell(\hat{Z}_1) < \varepsilon^{1/2+\delta}.$$

Proof. We need to show how to improve (9.15), (9.16), (9.17) and (9.20) for isolated returns.

We begin with \mathcal{X}_1 . An argument similar to the proof of (10.1) shows that on $P\mathcal{X}_1$

$$(10.4) \quad \left| \frac{d\bar{E}}{dE} \right| \geq \frac{1}{10} \varepsilon^{-(1/4+4\delta)}$$

so nothing is removed on the second step. Also (10.4) means that P has bounded distortion on $P\mathcal{X}_1$. Consider two cases.

(1) $|P^2\mathcal{X}_1| > 1$. By the bounded distortion \bar{E} -measure of the removed set is at most $\mathcal{O}(\varepsilon^{1/4-\delta}|P\mathcal{X}_1|) = \mathcal{O}(\varepsilon^{5/8-\delta})$ and since $|\frac{d\bar{E}}{dE}| > \varepsilon^\delta$ on \mathcal{X}_1 the E -measure of the removed set is $\mathcal{O}(\varepsilon^{5/8-2\delta})$.

(2) $|P^2\mathcal{X}_1| \leq 1$. Since \bar{E} -measure of the removed set is $\mathcal{O}(\varepsilon^{1/4-\delta})$, by (10.4) \bar{E} -measure of the removed set is at most $\mathcal{O}(\varepsilon^{1/2+3\delta})$ and its E -measure is at most

$$\mathbb{P}(\tilde{Z}_{II}) = \mathcal{O}(\varepsilon^{1/2+2\delta}).$$

(Here the tilde is used to emphasize that this inequality is valid only for isolated returns).

Case III is analyzed similarly to case I giving

$$\mathbb{P}(\tilde{Z}_{III}) = 0.$$

In case IV we have the following changes comparing with case II. Due to (9.12) and (9.13)

$$\frac{\varepsilon^{1/4}}{C} \leq |P(\mathcal{X}_2)| \leq C\varepsilon^{1/4}$$

and (10.2) shows that (10.4) still holds so only case (1) can happen. Now \bar{E} -length of the removed set is $\mathcal{O}(\varepsilon^{1/4-\delta}|P\mathcal{X}_2|) = \mathcal{O}(\varepsilon^{1/2-\delta})$. But the expansion on the first step is at least $\varepsilon^{-1/16}$ so

$$P(\tilde{Z}_{IV}) = \mathcal{O}(\varepsilon^{9/16-\delta}).$$

Case *VII* is analyzed as before except that due to (10.3) for points having close return on the second step the minimal expansion on the first step is now $\mathcal{O}(\varepsilon^{-1/8-2\delta})$ giving much needed improvement of 2δ . Thus

$$\mathbb{P}(\tilde{Z}_{VII}) = \mathcal{O}(\varepsilon^{1/2+\delta}).$$

This completes the proof of Lemma 10.1. \square

10.4. Proof of Proposition 6.5(c) for returns of the second kind. In the next section we show that non-isolated returns are rare. More precisely we establish the following result.

Lemma 10.2. $\mathbb{P}(\text{component containing } \bar{P}^2x \text{ has a non-isolated return}) = \mathcal{O}(\varepsilon^{1/8-4\delta} |\ln \varepsilon|).$

This Lemma implies the estimate of part (c) of Proposition 6.5 for returns of the second kind since the contribution of isolated returns is $\mathcal{O}(\varepsilon^{1/2+\delta})$ by Lemma 10.1 and the contribution of the non-isolated returns is

$$\mathcal{O}(\varepsilon^{1/8-4\delta} |\ln \varepsilon| \times \varepsilon^{7/16-\delta/4}) = \mathcal{O}(\varepsilon^{9/16-(17/4)\delta} |\ln \varepsilon|)$$

where the first factor is the probability of a non-isolated returns and the second factor is the probability of getting stuck in a non-isolated return.

10.5. Isolated returns are rare. Here we prove Lemma 10.2. Let

$$C_\varepsilon = \left\{ \frac{\partial \bar{E}}{\partial E} = 0 \right\}, \quad U_\varepsilon = \left\{ \left| \frac{\partial \bar{E}}{\partial E} \right| < \varepsilon^{-3/8} \right\}, \quad \mathcal{H}_\varepsilon = \left\{ \left| \frac{\partial \bar{E}}{\partial E} \right| \geq \varepsilon^{-3/8} \right\}.$$

Thus C_ε is the critical set, U_ε is a small neighborhood of the critical set and on \mathcal{H}_ε our system is strongly hyperbolic.

We want to follow the outline of Section 5 but we need to address two issues.

- (i) The isolated returns are defined in terms of $\frac{d\bar{E}}{dE}$, not $\frac{\partial \bar{E}}{\partial E}$.
- (ii) We know that

$$\frac{\partial \bar{E}}{\partial E} \sim \frac{1}{\sqrt{\varepsilon}} \left(\frac{\partial w}{\partial I} \frac{\partial \sigma}{\partial E} + o(1) \right)$$

but we need to check that $o(1)$ term does not invalidate our arguments.

To address (i) let E_0 be a point such that $\frac{d\bar{E}}{dE}(E_0) = 0$. Then

$$\frac{\partial \bar{E}}{\partial E} = \mathcal{O}(1)$$

whereas

$$\left| \frac{d}{dE} \frac{\partial \bar{E}}{\partial E} \right| = \left| \frac{\partial^2 \bar{E}}{\partial E^2} + \mathcal{O}(1) \right| \geq \frac{\text{Const}}{\sqrt{\varepsilon}}.$$

Therefore either there exists E'_0 such that $|E'_0 - E_0| \leq \text{Const}\sqrt{\varepsilon}$ and $(E'_0, I'_0) \in C_\varepsilon$ or E_0 is near a boundary of its standard curve.

In the first case we have $|\bar{E}'_0 - \bar{E}_0| \leq \text{Const}\sqrt{\varepsilon}$ so if the return is non-isolated then

$$(10.5) \quad \left| \frac{\partial \bar{\bar{E}}}{\partial \bar{E}}(\bar{E}'_0) \right| \leq 2\varepsilon^{-(1/4+4\delta)}.$$

Therefore we have the following statement.

Corollary 10.3. *If a subcurve $\tilde{\gamma} \subset \bar{P}^2\gamma$ has a non-isolated return then either there exists $x \in \tilde{\gamma}$ such that $x \in C_\varepsilon$ and $d(Px, C_\varepsilon) \leq \text{Const}\varepsilon^{1/4-4\delta}$ or $d(\partial\tilde{\gamma}, C_\varepsilon) \leq \text{Const}\sqrt{\varepsilon}$.*

Now in our inductive construction we have a freedom of how to break $P\gamma$ into pieces. If we avoid putting the endpoints in U_ε then the second possibility would mean that $\tilde{\gamma}$ experienced a close return and this has probability $\varepsilon^{1/4}$. Thus we can ensure isolation by excluding (10.5). This takes care of (i).

To handle (ii) we show that C_ε has properties similar to \mathcal{C} . Namely on C_ε , $\frac{\partial}{\partial \bar{E}}(\frac{\partial \bar{E}}{\partial E}) \neq 0$, so C_ε is a graph of a function $E = F(J)$ with

$$(10.6) \quad \frac{\partial F}{\partial J} = -\frac{\frac{\partial^2 \bar{E}}{\partial E \partial J}}{\frac{\partial^2 \bar{E}}{\partial E^2}} = \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right).$$

Let (\hat{E}, \hat{J}) be the image of C_ε . Then

$$(10.7) \quad \frac{1}{C^*\varepsilon} \leq \left| \frac{d\hat{E}}{dJ} \right| = \left| \frac{\partial \bar{E}}{\partial J} \right|(F(J), J) \leq \frac{C^*}{\varepsilon}$$

by twist condition.

The next lemma shows what we can disregard points with small expansion.

Lemma 10.4. $\mathbb{P}(\exists j \leq n(x) + n(\bar{P}(x)) : P^j x \in U_\varepsilon) = \mathcal{O}(\varepsilon^{1/8})$.

Proof. Since close returns have probability $\mathcal{O}(\varepsilon^{1/4}) \ll \varepsilon^{1/8}$ we can consider only points avoiding close returns. In this case we have bounded distortion property so the probabilities are the same for image and preimage giving the result. \square

Now consider $P\gamma \cap \mathcal{H}_\varepsilon$. It has many components.

Lemma 10.5. *J-distance between consecutive components is $\mathcal{O}(\varepsilon)$.*

Proof. If we lift the picture to a strip in (E, J) -plane then the E -distance between the components is $\mathcal{O}(1)$. On the other hand by (9.3)

$$\frac{d\bar{J}}{dE} = \mathcal{O}\left(\left|\frac{\partial\bar{E}}{\partial E}\right|\right)\varepsilon.$$

Since $\ln\left|\frac{\partial\bar{E}}{\partial E}\right|$ oscillates by $\mathcal{O}(1)$ on components from \mathcal{H}_ε the result follows. \square

Now consider $P^{-1}C_\varepsilon$. Since $D = \frac{\partial\bar{E}}{\partial E}\frac{\partial\bar{J}}{\partial J} - \frac{\partial\bar{E}}{\partial J}\frac{\partial\bar{J}}{\partial E} = 1 + o(1)$ we have

$$(10.8) \quad \frac{\partial J}{\partial\bar{J}} \sim \frac{\partial\bar{E}}{\partial E}, \quad \frac{\partial E}{\partial\bar{J}} \sim -\frac{1}{\varepsilon}\frac{\partial w}{\partial I},$$

$$(10.9) \quad \frac{\partial J}{\partial\bar{E}} = \mathcal{O}(\sqrt{\varepsilon}), \quad \frac{\partial E}{\partial\bar{E}} = \mathcal{O}(1).$$

Lemma 10.6. *Take $(\bar{E}_1, \bar{J}_1) \in C_\varepsilon$. Let $P^{-1}(\bar{E}_1, \bar{J}_1) = (E_1, J_1)$ belong to one of the components of $P\gamma \cap \mathcal{H}_\varepsilon$. Then $\exists(\bar{E}_1^*, \bar{J}_1^*) \in C_\varepsilon$ such that*

$$(10.10) \quad |\bar{J}_1^* - \bar{J}_1| \sim \frac{\text{Const}\varepsilon}{|(\partial J/\partial\bar{J})(\bar{J}_1)|}$$

and $P^{-1}(\bar{E}_1^*, \bar{J}_1^*)$ belongs to either the consecutive component of $P\gamma \cap \mathcal{H}_\varepsilon$ or to U_ε .

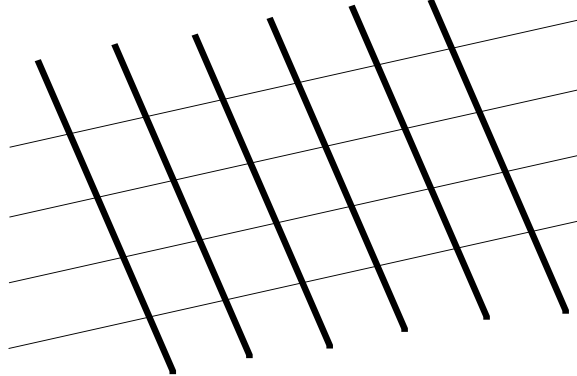


FIGURE 7. Components of $P\gamma$ (thin lines) and $P^{-1}C_\varepsilon$ (thick lines) have many crossings

Proof. Suppose to fix our notation that $\frac{\partial J}{\partial\bar{J}}(\bar{J}_1) = \frac{(\partial\bar{E}/\partial E)(E_1, J_1)}{D} > 0$. Let $J = f_1(E)$ and $J = f_2(E)$ be equations of consecutive components of $P\gamma$. We claim that $P^{-1}C_\varepsilon$ can not be squeezed between the graphs of f_1 and f_2 . Indeed by property (b1) of standard pairs $f_j(E)$ have slopes $\mathcal{O}(\varepsilon)$. Also the distance between the components is $\mathcal{O}(\varepsilon)$ by

Lemma 10.5. On the other hand $P^{-1}C_\varepsilon$ has slope at least $c\varepsilon^{5/8}$ by (10.6), (10.8) and (10.9). This shows the existence of the intersection. It remains to establish (10.10).

We have by the Intermediate Value Theorem

$$|J_1 - J_1^*| = |\bar{J}_1 - \bar{J}_1^*| \frac{dJ}{d\bar{J}}(\bar{E}_\diamond, \bar{J}_\diamond).$$

Next

$$\frac{dJ}{d\bar{J}} = \frac{\partial J}{\partial \bar{J}}(1 + o(1)) = \frac{\partial \bar{E}}{\partial E}(E_\diamond, J_\diamond)(1 + o(1)).$$

Using the definition of standard pairs and Proposition 6.1 we see that on $P\gamma$

$$\left| \frac{d}{dE} \frac{\partial \bar{E}}{\partial E} \right| \leq \frac{\text{Const}}{\sqrt{\varepsilon}}.$$

Next if the image of $[E_\diamond, E_1]$ belongs to \mathcal{H}_ε then $|E_1 - E_\diamond| \leq \varepsilon^{3/8}$ so that

$$\left| \frac{\partial \bar{E}}{\partial E}(E_1) - \frac{\partial \bar{E}}{\partial E}(E_\diamond) \right| \leq \varepsilon^{-1/8}$$

and hence

$$\frac{\partial \bar{E}}{\partial E}(E_\diamond, J_\diamond) = \frac{\partial \bar{E}}{\partial E}(E_1, J_1)(1 + o(1)).$$

This completes the proof of (10.10). Lemma 10.6 is proved. \square

Let now $E_2 < E_1 < E_3$ be such that the interval $[E_2, E_3]$ inside $P\gamma$ is the preimage of the standard component containing (\bar{E}_1, \bar{J}_1) . Let $\Delta E = E_3 - E_2$, $\overline{\Delta J} = |\bar{J}_1 - \bar{J}_1^*|$. By bounded distortion, (10.8) and Lemma 10.6

$$|\Delta E| \leq \frac{\text{Const}}{\frac{\partial \bar{E}}{\partial E}(E_1, J_1)} \leq \frac{\text{Const}}{\varepsilon} |\overline{\Delta J}|.$$

Let $\mathcal{R}_\varepsilon \subset C_\varepsilon$ denote the subset of points having non-isolated returns. Let $(\bar{E}_\alpha, \bar{J}_\alpha)$ be the points of the intersection $\mathcal{R}_\varepsilon \cap P^2\gamma \cap P\mathcal{H}_\varepsilon$ and consider corresponding $\Delta E_\alpha, \overline{\Delta J}_\alpha$. It follows that

$$\sum_\alpha |\Delta E_\alpha| \leq \frac{\text{Const}}{\varepsilon} \sum_\alpha |\overline{\Delta J}_\alpha| \leq \frac{\text{Const}}{\varepsilon} |\mathcal{R}_\varepsilon \cap \mathcal{K}_\varepsilon|$$

where \mathcal{K}_ε is the interval of feasible values of J after two passages. Observe that $|\mathcal{K}_\varepsilon| \leq \text{Const}\sqrt{\varepsilon} \ln \varepsilon$ since J jumps by at most $\mathcal{O}(\sqrt{\varepsilon} \ln \varepsilon)$. By (10.7) each component of \mathcal{R}_ε has J -length $\mathcal{O}(\varepsilon^{5/4-4\delta})$. Since it takes at least the length of $\text{Const}\varepsilon$ to wind around the cylinder there are at

most $\mathcal{O}\left(\frac{|\ln \varepsilon|}{\sqrt{\varepsilon}}\right)$ such components intersecting $\bar{P}^2\tilde{\gamma}$. Their total measure is $\mathcal{O}(\varepsilon^{3/4-4\delta}|\ln \varepsilon|)$. Thus

$$|\mathcal{R}_\varepsilon \cap \mathcal{K}_\varepsilon| \leq \text{Const} \varepsilon^{3/4-4\delta} |\ln \varepsilon|.$$

Therefore

$$\sum_{\alpha} |\Delta E_{\alpha}| \leq \text{Const} \varepsilon^{-1/4-4\delta} |\ln \varepsilon|.$$

We need to estimate the measure of preimages of these intervals. Since P expands by at least $\text{Const} \varepsilon^{-3/8}$ by the definition of \mathcal{H}_ε this measure is at most $\varepsilon^{1/8-4\delta} |\ln \varepsilon|$ as claimed.

10.6. Passages near saddle points. We are now ready to complete the proof of Proposition 6.5. Divide γ into subintervals $\{Z_\alpha\}$ such that $|\bar{P}^2 Z_\alpha| \sim 1$. Again for most segments the map $Z_\alpha \rightarrow \bar{P}^2 Z_\alpha$ has distortion $\mathcal{O}(\varepsilon^\delta)$ and so Proposition 6.2(d) gives

(10.11)

$$\mathbb{P}_\ell(\bar{P}^2 x \text{ gets captured before the next return } |x \in Z_\alpha) = \sqrt{\varepsilon} \lambda_{j(x)+2}(x).$$

To prove part (b) of Proposition 6.5 we need to get this estimate with *next return* replaced by *next free return*. The problem is that near the parabolic tip the preimage of a set of \bar{E} -measure $\mathcal{O}(\sqrt{\varepsilon})$ can have measure $\mathcal{O}(\sqrt{\varepsilon})$ (see (9.13)). Therefore on the second step a set of measure $\mathcal{O}(\sqrt{\varepsilon})$ can be removed so that (10.11) could fail if we replace \bar{P}^2 by \bar{P} . In other words we need to show that most of the tips of $\bar{P}^2\gamma$ do not get captured immediately. To do so we can extend the definition of isolation to require $d(\bar{E}_0, \mathcal{N}_\varepsilon) \geq \varepsilon^{1/4}$. Arguing as in subsection 10.5 we can show that most returns are isolated in the sense of this new definition. Therefore captures of the second step are much less likely than the captures at the first step so that (10.11) gives the main contribution in Proposition 6.5(b). This completes the proof of part (b).

The new notion of isolation also allows to improve the estimate for the measure of points getting stuck due to a return of the first kind to $\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/2})$ (even though the bound $\mathcal{O}(\varepsilon^{1/2} |\ln \varepsilon|^{-1/8})$ established in Section 9 would suffice for the proof of Theorem 1). This completes the proof of part (c) of Proposition 6.5.

11. THEOREM 2.

11.1. Preliminaries. We shall follow the main ideas of the proof of Theorem 1. We need to supplement Propositions 6.1 and 6.2 by the description of the dynamics of captured orbits. In the Proposition below d is the closest distance the orbit comes to \mathcal{N}_ε during the time of

capture (this distance is realized at either entrance into the resonance or the exit from it).

Proposition 6.2*. *The Poincaré map $P : \mathbb{S} \rightarrow \mathbb{S}$ satisfies the following on the set of captured orbits. Let orbit of (J, E) be captured near the saddle θ_{jk} and $(\bar{J}, \bar{E}) = P(J, E)$. Then for $d > \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}$ the following holds*

(a) $\sup_{d < \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}} |\bar{J} - Q_{jk}(J)| \rightarrow 0.$

(b) *There exists a function $\bar{\mathbf{c}}(I)$ such that*

$$\frac{\partial \bar{E}}{\partial E} \sim \bar{\mathbf{c}}(I) \frac{\ln^2 \varepsilon}{\varepsilon}, \quad \frac{\partial \bar{E}}{\partial J} = \mathcal{O}\left(\frac{\ln^2 \varepsilon}{\varepsilon^{3/2}}\right),$$

$$\frac{\partial \bar{J}}{\partial J} = \mathcal{O}\left(\frac{\ln^2 \varepsilon}{\sqrt{\varepsilon}}\right), \quad \frac{\partial \bar{J}}{\partial E} = \mathcal{O}(\ln^2 \varepsilon).$$

(c) *For any $\bar{\delta} > 0$ the estimates for the second derivatives are worse than the estimate of Proposition 6.2 by a factor of $\mathcal{O}(\varepsilon^{-(1+\bar{\delta})})$. Thus*

$$\frac{\partial^2 \bar{J}}{\partial E^2} = \mathcal{O}\left(\varepsilon^{-(3/4+\bar{\delta})}\right), \quad \frac{\partial^2 \bar{E}}{\partial E^2} = \mathcal{O}\left(\varepsilon^{-(7/4+\bar{\delta})}\right),$$

$$\frac{\partial^2 \bar{J}}{\partial J \partial E} = \mathcal{O}\left(\varepsilon^{-(5/4+\bar{\delta})}\right), \quad \frac{\partial^2 \bar{E}}{\partial J \partial E} = \mathcal{O}\left(\varepsilon^{-(9/4+\bar{\delta})}\right),$$

$$\frac{\partial^2 \bar{J}}{\partial J^2} = \mathcal{O}\left(\varepsilon^{-(7/4+\bar{\delta})}\right), \quad \frac{\partial^2 \bar{E}}{\partial J^2} = \mathcal{O}\left(\varepsilon^{-(11/4+\bar{\delta})}\right).$$

(d) *The time of capture is bounded by $\text{Const} \ln \varepsilon$.*

(e) $\text{Prob}_\ell\left(d < \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}\right) = \mathcal{O}\left(\sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}\right).$

The proof of Proposition 6.2* is given in the Appendix F. The explicit expression for $\bar{\mathbf{c}}(I)$ is (recall the notation of Section 3)

$$\bar{\mathbf{c}}(I) = \frac{-C^* \mathbf{c}(I) \mathbf{c}(Q(I))}{4M(Q(I))} \frac{\partial w}{\partial I}(Q(I)).$$

Next we extend \bar{P} to captured points as follows. If the orbit gets captured we let $\bar{P} = P$ unless $d < \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}$ in which case we declare the orbit stuck. Propositions 6.4 and 6.5 have to be modified as follows.

Proposition 6.4*. *Proposition 6.4 remains valid with the new definition of \bar{P} .*

The proof of Proposition 6.4* for captured points is similar but easier than the proof (of Proposition 6.4) for non captured points. Indeed for captured points the expansion is stronger while distortion is still under control. We leave the details for the reader.

Proposition 6.5* *Proposition 6.5 remains valid with the new definition of \bar{P} .*

Proposition 6.5* follows immediately from Proposition 6.5 since we now exclude a smaller set.

Comparing to subsection 6.3 we now enlarge Γ''_k since some captured points get stuck afterwards. We let Γ'_k to be the set of points which have been captured but have not been stuck so we have $\Gamma'_k \subset \Gamma_k$ in the new definition. Let Γ_{Njk}^* denote the set of points experiencing exactly one capture which happens near θ_{jk} and let $\hat{\Gamma}_N = \Gamma'_N - \left(\bigcup_{jk} \Gamma_{Njk}^* \right)$.

Proposition 6.6* *There exist a function $\delta_2(\delta_1)$ such that $\lim_{\delta_1 \rightarrow 0} \delta_2 = 0$ and a subset $\Gamma_N''' \subset \gamma$ such that*

$$(a) \quad \mathbb{P}_\ell(\Gamma_N''') = o(\delta_1),$$

$$(b) \quad |I_N - I_0 - \delta_1 \Psi(I_0)| \leq \delta_1 \delta_2$$

for all $x \in \Gamma_N - (\Gamma'_N \cup \Gamma_N''')$,

(c) Let $t^{(0)}(x)$ be the time between x_0 and x_N then

$$\left| \sqrt{\varepsilon} t^{(0)}(x) - \frac{\delta_1}{p(I_0)} \right| \leq \delta_1 \delta_2$$

for all $x \in \Gamma_N - (\Gamma'_N \cup \Gamma_N''')$,

$$(d1) \quad \left| \mathbb{P}_\ell(\Gamma_{Njk}^*) - \frac{\delta_1 M_{jk}(I_0)_+}{L_j(I_0)} \right| \leq \delta_1 \delta_2 \text{ and for } x \in \Gamma_{Njk}^*$$

$$(d2) \quad I_N - Q_{jk}(I_0) = o(1),$$

$$(d3) \quad t^0(x) \leq \text{Const}N,$$

$$(e) \quad \mathbb{P}_\ell(\Gamma_N'') \leq \delta_1 \delta_2,$$

$$(f) \quad \mathbb{P}_\ell(\hat{\Gamma}_N) \leq \delta_1 \delta_2.$$

The proof of Proposition 6.6* is similar to the proof of Proposition 6.6. The only new issue is to show that the set of points experiencing more than one collision has measure $o_{\delta_1 \rightarrow 0}(N\sqrt{\varepsilon})$. This is done in two steps. First we use Proposition 6.5(b) to show that the set of points experiencing at least one collision by the time of the n -th return has measure at most $Cn\sqrt{\varepsilon}$. Secondly we use this bound to deduce that the set of points experiencing at least two collisions by the time of the n -th return has measure at most $(Cn\sqrt{\varepsilon})^2$.

11.2. Proof of Theorem 2. We follow the proof of Theorem 1. Again it is enough to obtain the result for $(I, \phi)(0)$ being distributed according to a standard pair ℓ .

We define $\Gamma^{(n)}$, $\Gamma^{(n)'''}$, $\Gamma^{(n)''}$, $\hat{\Gamma}^{(n)}$ and $\gamma_j^{(n)}$, similarly to the definitions of Section 7. As before we show that if $n \leq T/\delta_1$ then the measures of $\Gamma^{(n)'''}$, $\Gamma^{(n)''}$ and $\hat{\Gamma}^{(n)}$ are small. Also Proposition 6.6*(b) shows that $I^{(n)}$ changes little between τ_n and τ_{n+1} except for a small measure set

so it is enough to restrict our attention to $(I^{(n)}, \tau_n)$. We want to show that as $\varepsilon \rightarrow 0$ and $\delta_1(\varepsilon) \rightarrow 0$ at appropriate rate $(I^{(t/\delta_1)}, \tau_{t/\delta_1})$ converges to a Markov process with generator

$$(\mathbb{L}A)(I, \tau) = \Psi \frac{\partial A}{\partial I} + \frac{1}{p} \frac{\partial A}{\partial \tau} + \sum_{jk} \frac{M_{jk}(I)_+}{L_j(I)} [A(Q_{jk}(I), \tau) - A(I, \tau)].$$

We first show that the family $\{(I^{(t/\delta_1)}, \tau_{t/\delta_1})\}$ is tight. Let T_0 be a small number. Since by Proposition 6.6* this family is uniformly Lipschitz apart from the jumps by [25], Section VI.5 it suffices to show that given $\kappa > 0$ there exists $T_0, \varepsilon_0, \delta_0$ such that

$$(11.1)$$

$$\text{Prob}(\exists t_0 \in [0, T-T_0] : I^{(t/\delta)} \text{ experience at least two jumps on } [t_0, t_0+T_0]) \leq \kappa$$

provided that $\delta_1 \leq \delta_0, \varepsilon \leq \varepsilon_0$.

To this end we show that there are constants $C_1, C_2 > 0$ such that

$$(11.2)$$

$$\mathbb{P}_\ell(\text{there are at least } m \text{ jumps for by time } T_0/\delta_1) \leq C_1(C_2 T_0)^m.$$

(11.2) with $m = 2$ implies (11.1) since the interval $[0, T]$ can be subdivided into $\mathcal{O}(T_0^{-1})$ subintervals of size T_0 . So it remains to demonstrate (11.2). Proposition 6.6*(f) allows us to neglect points having at most two jumps between τ_n and τ_{n+1} for some n . Let

$$w_{mn} = \mathbb{P}_\ell(\mathcal{J}_{mn}) \quad \text{where}$$

$\mathcal{J}_{mn} = \{\text{there are at least } m \text{ jumps by time } n \text{ but there is at most 1 jump on } \tau_k, \tau_{k+1} \text{ for each } k \leq n\}$.

We wish to show that

$$(11.3) \quad w_{mn} \leq C_1 (C_2 n \delta_1)^m$$

Proposition 6.6*(d1) gives a recursive relation

$$w_{m+1, n+1} \leq w_{m+1, n} + K \delta_1 w_{mn}$$

from which (11.3) follows by induction. (11.3) implies (11.2) proving the tightness.

Let $(\mathbf{I}(t), \boldsymbol{\tau}(t))$ be a limit point. We shall show that for any functions of compact support $B_1, B_2 \dots B_m, A$ for any moments $s_1 < s_2 < \dots < s_m < t_1 < t_2$

$$(11.4)$$

$$\mathbb{E} \left(\prod_k B_k(\mathbf{I}(s_k), \boldsymbol{\tau}(s_k)) \left[A(\mathbf{I}(t_2), \boldsymbol{\tau}(t_2)) - A(\mathbf{I}(t_1), \boldsymbol{\tau}(t_1)) - \int_{t_1}^{t_2} (\mathbb{L}A)(\mathbf{I}(s), \boldsymbol{\tau}(s)) ds \right] \right) = 0.$$

To establish (11.4) we let n_q ($q = 1, 2$) be the numbers such that $n_q \delta_1$ are closest to t_q . Take some $\gamma_j^{(n_1)}$ and let ℓ_j be the corresponding

standard pair. We show that

$$(11.5) \quad \mathbb{E}_{\ell_j} \left(A(I^{(n_2)}, \tau_{n_2}) - A(I^{(n_1)}, \tau_{n_1}) - \delta_1 \sum_{s=n_1}^{n_2} (\mathbb{L}A)(I^{(s)}, \tau_s) \right) \rightarrow 0$$

as $\varepsilon \rightarrow 0$, $\delta(\varepsilon) \rightarrow 0$.

Let $\gamma_r^{(s)}$ be standard curves such that $\bar{P}^{-(s-n_1)N} \gamma_r^{(s)} \subset \gamma_j^{(n_1)}$ and let $\ell_r^{(s)}$ be the associated standard pairs. Let (I^*, τ^*) be values of action and time for a point from $\gamma_r^{(s)}$. Then on $\gamma_r^{(s)}$ we have $(I, \tau) = (I^*, \tau^*) + o(1)$. We shall show that

$$(11.6) \quad \mathbb{E}_{\ell_r^{(s)}} (A(I^{(n_2)}, \tau_{n_2}) - A(I^{(n_1)}, \tau_{n_1}) - \delta_1 (\mathbb{L}A)(I^*, \tau^*)) = o(\delta_1).$$

Then the summation over r and s will give (11.5). Next, since the product of B s in (11.4) is almost constant on $\gamma_j^{(n_1)}$ we have

$$\mathbb{E}_{\ell_j} \left(\mathbf{B} \left[A(I^{(n_2)}, \tau_{n_2}) - A(I^{(n_1)}, \tau_{n_1}) - \delta_1 \sum_{s=n_1}^{n_2} (\mathbb{L}A)(I^{(s)}, \tau_s) \right] \right) \rightarrow 0$$

as $\varepsilon \rightarrow 0$, $\delta(\varepsilon) \rightarrow 0$ where $\mathbf{B} = \prod_k B_k(I^{-(n_1-s_j\delta_1)}, \tau_{-(n_1-s_j\delta_1)})$.

Summing over j and passing to the limit as $\varepsilon \rightarrow 0$ we obtain (11.4) as claimed.

To establish (11.6) we consider the contribution of several terms to $\mathbb{E}_{\ell_r^{(s)}} (A(I_N, t^{(0)}(x)) - A(I^*, \tau^*))$.

(I) The contributions of Γ''' , Γ'' and $\hat{\Gamma}$ are $o(\delta_1)$ due to parts (a) and (f) of Proposition 6.6*.

(II) The contribution of $\Gamma - (\Gamma' \cup \Gamma'' \cup \Gamma''')$ is

$$\delta_1 \left[\Psi \frac{\partial A}{\partial I} + \frac{1}{p} \frac{\partial A}{\partial \tau} \right] (I^*, \tau^*)$$

due to parts (b) and (c) of Proposition 6.6*.

(III) The contribution of Γ_{Njk}^* is

$$\delta_1 \frac{M_{jk}(I^*)_+}{L_j(I^*)} [A(Q_{jk}(I^*), \tau^*) - A(I^*, \tau^*)]$$

due to part (d) of Proposition 6.6*.

Combining (I), (II) and (III) we obtain (11.6) completing the proof of Theorem 2.

12. OPEN PROBLEMS.

In this paper we described effective evolution of slow-fast systems with periodic fast motion and integrable slow motion in the presence of both weak and strong resonances. This is a first step in developing the

statistical theory of adiabatic invariants. Below we list open questions motivated by our work.

12.1. Weakening of conditions (A)-(K). The theorems of this paper are valid under nondegeneracy assumptions (A)-(K). For a typical system one can expect these assumptions to hold in a neighborhood of a typical point. However they may not be valid globally. For example, assumption (E) says that a certain function of I is non-zero. While one can expect this functions to be non-zero near a given point I_0 , on the whole interval it may have zeroes which can not be removed by a small perturbation. As a result Theorems 1 and 2 describe the evolution of the system only locally in time, that is, until the orbit leaves the region where the assumptions (A)-(I) hold. By, contrast, the assumptions (J) and (K) can be expected to hold globally since they ask that certain maps $\mathbb{R} \rightarrow \mathbb{R}^2$ are non-zero. It is desirable to relax the other conditions as well so that they would hold globally for a typical system. Below we discuss possible weakening of conditions (A)-(I).

Question 1. *Is it possible to extend Theorem 1 and 2 to the following cases*

- (a) *there is an orbit of the averaged system which is tangent to a resonance curve;*
- (b) *two resonances cross?*

The first step in answering these questions is to obtain the analogue the normal form ((2.8), (2.13)) for systems with degeneracies described above. The estimates of errors for averaging of slow-fast systems with single frequency fast motion under the assumption of general position singularities were obtained in [6] but we need a more precise information to deal with the problem of adiabatic invariants.

Question 2. *Are Theorems 1 and 2 still valid if the twist condition is replaced by the assumption that the critical points of the function $I \rightarrow w_j(I)$ are non-degenerate?*

Heuristic arguments of Section 5 indicate that in case all w_j are constant the system may have many elliptic islands since the hyperbolicity is lost (cf. [52]). On the other hand if $\frac{\partial w_j}{\partial I}$ have only isolated zeroes then there is still some hyperbolicity even though it is much weaker than in the case where the twist condition holds.

Question 3. *Extend Theorems 1 and 2 to the case where assumptions (D)-(F) hold except for finitely many values of I .*

In order to achieve such an extension one needs to do two things. First, an asymptotics of Propositions 6.1 and 6.2 should be improved

since lower order terms will play a role near the degeneracies. Secondly, one needs to modify the definition of stuck orbits. Indeed the assumptions (D)-(F) are used to establish hyperbolicity so near the points where (D)-(F) fail the hyperbolicity is weaker and more time is needed to recover good estimates near the parabolic tips.

Assumption (G) is discussed in Section 12.2.

Finally we come to assumptions (H) and (I). They can be weakened as follows.

(H') There exists $\mathbf{s}_{jk}(I)$ such that $\bar{H}(\mathbf{s}) < 0$ for $s \in [0, \mathbf{s}_{jk}(I)]$, $\bar{H}(\mathbf{s}_{jk}(I)) = 0$. The functions $I \rightarrow M_{jk}(I)$ and $I \rightarrow M_{jk}(\bar{I}(\mathbf{s}_{jk}(I)))$ have only isolated zeroes $\{I'_{jkl}\}$ and $\{I''_{jkl}\}$ respectively. Moreover $\Psi(I''_{jkl}) \neq 0$.

(I') The inner averaged equation is not overtwisted apart from finitely many points $\{\tilde{I}_{jkl}\}$. Moreover $\Psi(\tilde{I}_{jkl}) \neq 0$.

Theorem 2'. Theorem 2 remains valid if assumptions (H) and (I) are replaced by (H') and (I').

Proof. Take $h(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$ sufficiently slowly (for example $h(\varepsilon) = \varepsilon^\delta$ where $\delta = 0.001$ would do). Let $\Delta_{jkl} = [I_{jkl} - h, I_{jkl} + h]$. (I_{jkl} can be either I'_{jkl} , I''_{jkl} , or \tilde{I}_{jkl} .) On the complement of Δ_{jkl} 's the argument of Theorem 2 remain valid since Proposition 6.2* remains valid (indeed assumptions (H) and (I) is satisfied on $\mathbf{G} - \cup_{jkl} \Delta_l$). However inside Δ_l standard pairs need not be preserved after the capture since the estimates of Appendix F are no longer valid. However by Theorem 1 the probability that the orbit is captured inside one of Δ_{jkl} near the bad saddle is $\mathcal{O}(h)$ so it tends to 0 as $\varepsilon \rightarrow 0$. (In case of I''_{jkl} and \tilde{I}_{jkl} this is true because a typical orbit spends time $\mathcal{O}(h)$ on Δ_{jkl} . In case if I'_{jkl} the estimates of Appendix F can only be violated for the orbit captured near the saddle $\theta_{jk}(I)$ but the probability of such capture is small since M_{jk} is small on Δ_{jk} .) Therefore bad captures inside Δ_{jkl} can be disregarded. \square

12.2. Separatrix crossings. A typical situation where the results of this paper apply is small perturbations of integrable systems. In order to bring the system to the form 2.1 one needs to pass to action-angle variables of the integrable system. However the integrable systems usually admit action-angle coordinates only locally. In fact, in 1 degree of freedom systems the resonances typically happen at the separatrices separating the regions with different action-angle coordinates (since the vanishing of frequency is equivalent to the period being infinite). A well known example of this situation is one (and a half) degree of freedom systems with slowly changing Hamiltonian $H(p, q, \varepsilon t)$. Currently the

results of our paper do not apply to this setting since action-angle coordinates are singular near the separatrices and so the smoothness assumptions of our paper are not satisfied. The C^0 expansions for the change of quantities of interest are obtained in [11, 12, 44, 53], but to apply our arguments one needs to supplement them by the C^2 -bounds. We hope that the analysis of Appendices E and F can be helpful in obtaining such bounds.

Question 4. *Extend Theorems 1 and 2 to the case where the vanishing of the frequency ω happens on the separatrices of the integrable system.*

This question deals with separatrices of the fast system. However separatrix crossing appear at other stages of our analysis as well.

Question 5. *Extend Theorems 1 and 2 to the case where the averaged system (2.3) has separatrices.*

For systems with separatrices it is natural to consider the limiting Markov process not on the segment but on the graph whose vertexes correspond to the separatrices. The motion inside each edge could be analyzed by the method of our paper, but this analysis should be supplemented by the boundary conditions describing the probability of the orbit to enter different action-angle domains after the separatrix crossing. We refer the reader to [22, 23, 48, 49] for surveys of problems where similar limiting processes on graphs appear.

Question 6. *Extend Theorem 2 to the case where the inner unperturbed system (2.8) has both centers and saddles.*

In this case the entrance-exit maps become random since the domain which the trajectory chooses after crossing the separatrix of the inner averaged equation becomes random. The probability of choosing a particular domain is computed in [43] (see also [5, 8]).

12.3. Systems without strong resonances.

Question 7. *Describe the effective evolution of the adiabatic invariants in case all resonances are weak.*

If all resonances are weak then Theorem 2 tells us that most orbits do not move at time $t \sim \varepsilon^{-1/2}$. Since the Law of Large Numbers gives a trivial description of the dynamics it is natural to conjecture that the main contribution comes from deviations which are described by the Central Limit Theorem. That is, one expects [50] that the adiabatic invariants evolve so that

$$I(t) - I(0) \sim \sqrt{\varepsilon} \sqrt{t}$$

and hence the correct scaling is $\tau = t\varepsilon^{-1}$ and the limiting process should be a diffusion. That is the generator should be

$$\mathcal{L} = \mathbf{a}(I)\partial_I + \frac{1}{2}\mathbf{b}(I)\partial_I^2.$$

A comparison with systems with chaotic fast motion [18] suggests that the diffusion matrix should be determined by the leading terms in the change of I . Namely,

$$\mathbf{b}(I) = \int \sigma^2(I, E)dE.$$

On the other hand in order to compute the drift one needs to take into account subleading terms as well. To understand this consider a simple recurrence relation

$$I_{n+1} - I_n = \sqrt{\varepsilon}\sigma_n + \varepsilon\tilde{\sigma}_n$$

where $(\sigma_n, \tilde{\sigma}_n)$ are independent and $\mathbb{E}(\sigma_n) = 0$. In this case $n \sim \varepsilon^{-1}$ steps are needed for I to change by $\mathcal{O}(1)$. During this time the systematic contribution of the subleading term $\tilde{\sigma}_n$ is $\sim \varepsilon n$ which is of the same order as the fluctuations $\mathcal{O}(\sqrt{\varepsilon}\sqrt{n})$ of the leading term σ_n . One case where drift computations can be simplified is when the invariant measure of the limiting process is known. For example if the original slow-fast system has smooth invariant measure this measure could be projected into the space of adiabatic invariants yielding the invariant measure for the limiting process. In this case the condition that a given measure is invariant yields a relation between drift and diffusion coefficients (sometimes called *Einstein relation*).

In the general case in order to compute the drift one needs to improve the asymptotics of Proposition 6.1. It is likely that higher order improved adiabatic invariants can be helpful but the computations would be more involved than the computations presented in the appendices of the present paper.

Another problem is that one needs to relax the definition of stuck orbits since otherwise all the orbits will be removed from consideration by the time $t \sim \varepsilon^{-1}$. Some progress in this direction has been obtained in [15].

After the systems with weak resonances only are understood one can pass to the generic systems where the segment of possible I values is divided into several regions, some of which admit only weak resonances while the others have the resonances of both types. In particular, one should study the transition between different regions. The questions discussed in Section 12.1 should be of particular relevance here in describing the motion near the boundary.

In case the space of I s is higher dimensional the regions also are higher dimensional and so their boundaries have positive dimension. In the analysis of these systems one is likely to encounter diffusions with non standard boundary conditions. The study of such diffusions is of independent interest.

12.4. Non integrable averaged motion. Our paper deals with the case where the averaged system is integrable. This is a special instance of the general problem in averaging theory which can be stated as follows.

Suppose that the long time behavior of the averaged system is well understood. Describe the long time behavior of the actual system.

While this problem is too general we would like to formulate the following question.

Question 8. *What happens if the averaged system is Morse-Smale?*

In other words we suppose that the limit set of the averaged system consists of finitely many periodic (and fixed) orbits. Therefore each trajectory of the averaged system eventually settles close to one of these orbits. However the capture into a resonance may eventually move the actual trajectory into the attraction domain of another orbit. So the actual system will exhibit *metastability* ([49]). The limiting process should be a finite state Markov chain describing near which orbit the trajectory is located. Question 8 makes sense for any dimension of the slow variables but it is especially relevant in two dimensions since the Morse-Smale property is generic among dissipative two dimensional flows [54].

12.5. Higher dimensions.

Question 9. *Extend Theorem 2 to higher dimensional systems.*

It seems likely that the methods of the present paper are sufficient to handle the case of higher dimensional slow variables. In fact, the inner unperturbed system (2.8) is integrable in any dimensions ([5, 37]) and many results of Appendices C and D admit straightforward generalizations to the case I and ϕ have arbitrary dimensions.

The case of higher dimensional fast variables (that is of quasiperiodic fast motion) is quite different. It is possible that our approach can be adapted to the case when there are only finitely many resonances (that is α s and β s in (2.1) are trigonometric polynomials in θ) once an appropriate definition of the standard pair is found in this case. However the general case when infinitely many resonances are present requires new ideas. In fact, it is not even clear how to generalize our

assumptions (A)–(K) to the quasiperiodic case. In the periodic case these conditions require that certain expressions should be different from zero. Since there are only finitely many resonances we could utilize compactness to get uniform bounds from below. It is unrealistic to expect such uniform bounds in case of infinitely many resonances. On the other hand merely requiring the corresponding quantities to be different from zero is likely to be insufficient for our arguments to work. For example, recall that in the periodic case if there are no resonances then the KAM theory is applicable. However mere irrationality of rotation number is insufficient to guarantee the existence of invariant tori. One has to require some qualitative non resonance estimate as given for example by the Diophantine condition. It is not clear how to formulate qualitative extensions of conditions (A)–(K) which would be sufficient for our approach to work but would not be too restrictive so that they could be verified for systems of interest.

As a first step in handling quasiperiodic fast motion one can try to prove that the jumps of adiabatic invariants at different resonances are independent for times of order 1. (For periodic fast motion such results have been obtained in [55] in a model problem).

Acknowledgment. I am grateful to Anatoly Neishtadt for many useful comments on the preliminary version of this paper.

APPENDIX A. ASYMPTOTICS OF THE POINCARÉ MAP.

A.1. Size of the jump. Here we prove Proposition 6.1(b). First we check the convergence of (2.13). To this end we observe that by (2.14) we have to establish the convergence of

$$\int \frac{\alpha_1(I, \phi(I), \theta) d\theta}{\sqrt{2(LE + L\theta + G)}} = \int \frac{dA_1(I, \phi(I), \theta)}{\sqrt{2(LE + L\theta + G)}} =$$

$$\frac{A_1(I, \phi(I), \theta)}{\sqrt{2(LE + L\theta + G)}} + \frac{1}{2} \int \frac{A_1(2L + g)}{\left(\sqrt{2(LE + L\theta + G)}\right)^3} d\theta$$

and the last integral converges at ∞ since the denominator behaves as $\theta^{3/2}$.

Next, we estimate the change of J in three different regions.

- (1) $\{|\omega| \geq K\varepsilon^{1/4}\}$
- (2) $\{R\sqrt{\varepsilon} \leq |\omega| \leq K\varepsilon^{1/4}\}$
- (3) $\{|\omega| \leq R\sqrt{\varepsilon}\}$

where K and R are parameters. Let Δ_i denote the jump of J in the region (i). Since $|\dot{J}| \leq \text{Const} \frac{\varepsilon}{\omega^2}$ we have $|\Delta_1| \leq \text{Const} \sqrt{\varepsilon}/K^2$. To bound Δ_2 we have to estimate $\int \frac{\varepsilon}{\omega^2} dt$ where the integral is over region (2). Recall that $r = \frac{\omega}{\sqrt{\varepsilon}}$, $\frac{d\theta}{dt} = \sqrt{\varepsilon}r$, $r^2 \sim \theta$. Therefore using θ as the integrand we get

$$\left| \int \frac{\varepsilon}{\omega^2} dt \right| \leq \text{Const} \int \frac{\sqrt{\varepsilon}}{|r|^3} d\theta \leq \text{Const} \frac{\sqrt{\varepsilon}}{\sqrt{\theta}} \Big|_{\theta_1}^{\theta_2} \leq \text{Const} \frac{\sqrt{\varepsilon}}{r} \Big|_R^{K\varepsilon^{-1/4}} \leq \text{Const} \frac{\sqrt{\varepsilon}}{R}.$$

To estimate Δ_3 we observe that by (2.4) $\Delta_3 = \tilde{\Delta}_3 + \mathcal{O}(\frac{\sqrt{\varepsilon}}{R})$ where $\tilde{\Delta}_3$ is the change of I in region (3). To estimate $\tilde{\Delta}_3$ change variables in (2.5): $\hat{I} = \frac{I - \tilde{I}}{\sqrt{\varepsilon}}$, where \tilde{I} is the value of I at the time our trajectory crosses $\{|\omega| = \sqrt{\varepsilon}R\}$. Then

$$\hat{I}' = \bar{\alpha}(\tilde{I} + \sqrt{\varepsilon}\hat{I}, \sqrt{\varepsilon}r, \theta, \varepsilon)$$

and we also replace I by $\tilde{I} + \sqrt{\varepsilon}\hat{I}$ in the RHS of (2.5). Now as $\varepsilon \rightarrow 0$ the equation (2.5) converges to

$$\theta' = r, \quad r' = L + g, \quad \hat{I}' = \alpha_1(\tilde{I}, 0, \theta)$$

It follows that $\tilde{\Delta}_3 \sim \sqrt{\varepsilon}\sigma_R(\tilde{I}, \tilde{E})$ where \tilde{E} is the value of E at the time our trajectory crosses $\{|\omega| = \sqrt{\varepsilon}R\}$ and σ_R denotes the integral (2.13) taken between the limits $r = \mp R$.

Next we claim that

$$(A.1) \quad \tilde{E} = E + o(1), \quad c \rightarrow 0$$

Observe that the evolution of E is given by the following equation (see (2.5))

$$(A.2) \quad E' = \sqrt{\varepsilon} \frac{r^2 L'}{2L^2} \alpha_1 + \sum_{m=1}^M \sqrt{\varepsilon}^m r^{m+1} \tilde{\gamma}_1^{(m)} + \sum_{m=1}^M \sqrt{\varepsilon}^{m+1} r^m \tilde{\gamma}_2^{(m)} + HOT$$

where $\tilde{\gamma}_1^{(m)}$ have zero mean in θ . Thus $\tilde{\gamma}_1^{(m)} = \frac{d\Gamma_1^{(m)}}{d\theta}$. Make a change of variables

$$\bar{E} = E - \left[\sqrt{\varepsilon} \frac{r^2 L'}{2L^2} A_1 + \sum_m \varepsilon^m r^m \Gamma_1^{(m)} \right].$$

Then (A.2) and the fact that $\theta' = \mathcal{O}(r)$ imply that $\frac{d\bar{E}}{d\theta} = \mathcal{O}(\sqrt{\varepsilon})$. Since θ changes on the interval of order $c^2/\sqrt{\varepsilon}$ (A.1) follows. Combining our bounds for Δ_1 , Δ_2 and Δ_3 we get the following asymptotics for the total jump

$$\Delta \sim \sqrt{\varepsilon} \left[\mathcal{O} \left(\frac{1}{K^2} + \frac{1}{\sqrt{r}} \right) + \sigma_R(I, E) \right].$$

Since K and R are arbitrary we can let them go to ∞ getting $\Delta \sim \sqrt{\varepsilon} \sigma(I, E)$.

A.2. Passage time. To obtain Proposition 6.1(a) we observe that by argument of Section A.1 implies that

$$(I, \phi)(t) = (\bar{I}, \bar{\phi})(t) + \mathcal{O}(\sqrt{\varepsilon})$$

where $(\bar{I}, \bar{\phi})(t)$ is the solution of (2.3) with the same initial condition. Now the result is obvious.

A.3. Proof of Lemma 6.3.

Proof. Observe that (2.8) preserves Liouville measure $drd\theta = L(I)dEdt$. Let t_z be the first moment when the solution of (2.8) has $r(t) = z$. Then uniformly in θ, I

$$a(\theta, I) = \lim_{z \rightarrow \infty} \int_{-t_z}^{t_z} \alpha_1(I, \phi_j(I), \theta(I, E, s)) ds.$$

Since the union of orbits starting on $\left\{ r = - \left(R + \frac{G_I(\theta)}{R} \right) \right\}$ covers whole cylinder $[0, 1] \times \mathbb{R}$ except for the loops Ω_{jk} we get

$$\begin{aligned} & \int \left[\int_{-t_z}^{t_z} \alpha_1(I, \phi_j(I), \theta(I, E, s)) ds \right] dE \\ &= \iint_{[0,1] \times [-z,z] - \bigcup_k \Omega_{jk}} \alpha_1(I, \phi_j(I), \theta) \frac{drd\theta}{L(I)} \end{aligned}$$

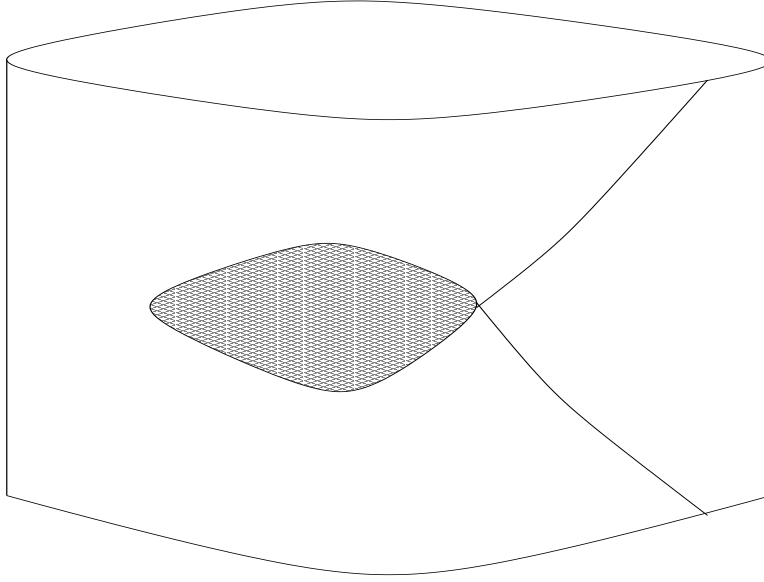


FIGURE 8. Proof of Lemma 6.3. The fluxes through the clear and filled domains have opposite signs.

$$= - \sum_k \iint_{\Omega_{jk}} \alpha_1(I, \phi_j(I), \theta) \frac{dr d\theta}{L(I)} = \Psi_j(I)$$

(the second identity follows from the fact that

$$\iint_{[0,1] \times [-z,z]} \alpha_1(I, \phi_j(I), \theta) dr d\theta = 0$$

by (2.2).)

□

APPENDIX B. DERIVATIVES OF THE POINCARÉ MAP. OUTLINE OF THE PROOF

Here we describe the asymptotics of Poincaré maps between sections corresponding to different resonances. We assume first that the orbit avoids δ_0 neighborhood of the separatrix and then show how to remove this restriction. Let $\bar{R} = c\varepsilon^{-1/4}$. Given a resonance we let $r = \omega/\sqrt{\varepsilon}$,

$$\mathcal{S} = \left\{ r = - \left(\bar{R} + \frac{G}{\bar{R}} \right) \right\}, \quad \tilde{\mathcal{S}} = \left\{ r = \bar{R} + \frac{G}{\bar{R}} \right\}.$$

To simplify the formulas we use H instead of E variable. Then the asymptotics of Proposition 6.1 take form

$$\frac{\partial \bar{H}}{\partial H} = \frac{1}{\sqrt{\varepsilon}} \bar{L} \frac{\partial w}{\partial I} \frac{\partial \sigma}{\partial H} + o(1), \quad \frac{\partial \bar{H}}{\partial J} = \frac{\bar{L}}{\varepsilon} \frac{\partial w}{\partial I} + o(1),$$

$$\frac{\partial^2 \bar{H}}{\partial H^2} = \frac{1}{\sqrt{\varepsilon}} \bar{L} \frac{\partial w}{\partial I} \frac{\partial^2 \sigma}{\partial H^2} + \mathcal{O} \left(\varepsilon^{1/4} \left| \frac{\partial \bar{H}}{\partial H} \right|^2 \right).$$

where \bar{L} is the value of L at the new resonance.

To estimate the derivatives we decompose the Poincare map into several pieces by cutting the orbit between the sections into several parts.

Below we use the following terminology. Given a surface S and a point x_0 let $\tau(x_0)$ be the first time the orbit of x_0 visits S . We call time $\tau(x_0)$ map of our differential equation *the hit map* of S (for x_0). If instead of fixing time we project orbits near x_0 to S along the flow lines we shall call the result *landing (to S) map*.

- (1) $(J, H) \rightarrow (I, H)$
- (2) Landing to $\{\theta = \theta_0\}$. For steps (3) and (5) we use θ as the time variable
- (3) Hit of $\{r = -R\}$
- (4) Passage of resonance (from $\{r = -R\}$ to $\{r = R\}$)
- (5) Hit of \tilde{S}_1
- (6) Landing to \tilde{S}_1
- (7) $(I, H) \rightarrow (J, H)$
- (8) $(J, H) \rightarrow (J, \theta)$
- (9) Hit of S_2
- (10) Landing to S_2
- (11) $(J, \theta) \rightarrow (J, H)$

In the computations below we always assume that $\varepsilon \rightarrow 0$, $c \rightarrow 0$, $R \rightarrow \infty$ so that $\varepsilon \ll c$, $\varepsilon \ll 1/R$. That is first, we choose c as small and R as large as needed and then let $\varepsilon \leq \bar{\varepsilon}(c, R)$.

We shall use subscripts j for the variables appearing at step j . Thus the total Poincare map takes $(J_0, H_0) \rightarrow (J_{11}, H_{11})$.

Steps (1)–(7) constitute the passage through the resonance. We call the map $(J_0, H_0) \rightarrow (J_7, H_7)$ the *inner map*. It is analyzed in Appendix C. The estimates of Appendix C can be summarized as follows.

Proposition B.1.

$$(B.1) \quad \frac{\partial J_7}{\partial H_0} \sim \sqrt{\varepsilon} \frac{\partial \sigma}{\partial H},$$

$$\frac{\partial J_7}{\partial J_0} \sim 1, \quad \frac{\partial H_7}{\partial H_0} \sim 1, \quad \frac{\partial H_7}{\partial J_0} \leq \frac{\text{Const}}{\sqrt{\varepsilon}},$$

$$(B.2) \quad \frac{\partial^2 J_7}{\partial H_0^2} = \sqrt{\varepsilon} \frac{\partial^2 \sigma}{\partial H^2} + o(\sqrt{\varepsilon}),$$

$$\begin{aligned} \frac{\partial^2 J_7}{\partial H_0 \partial J_0} &= \mathcal{O}(1), & \frac{\partial^2 J_7}{\partial J_0^2} &= \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right), \\ \frac{\partial^2 H_7}{\partial H_0^2} &= \mathcal{O}(1), & \frac{\partial^2 H_7}{\partial H_0 \partial J_0} &= \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right), & \frac{\partial^2 H_7}{\partial J_0^2} &= \mathcal{O}\left(\frac{1}{\varepsilon}\right). \end{aligned}$$

Steps (8)-(11) describe the motion far from resonance. We call the map $(J_7, H_7) \rightarrow (J_{11}, H_{11})$ the *outer map*. It is analyzed in Appendix D. The upshot is the following.

Proposition B.2.

$$(B.3) \quad \begin{aligned} \frac{\partial H_{11}}{\partial J_7} &\sim \frac{1}{\varepsilon} \bar{L} \frac{\partial w}{\partial I}, \\ \frac{\partial H_{11}}{\partial H_7} &= \mathcal{O}(1), & \frac{\partial J_{11}}{\partial H_7} &= \mathcal{O}(\varepsilon), & \frac{\partial J_{11}}{\partial J_7} &\sim 1. \\ \frac{\partial^2 H_{11}}{\partial J_7^2} &= \mathcal{O}(\varepsilon^{-7/4}), & \frac{\partial^2 H_{11}}{\partial J_7 \partial H_7} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{\partial^2 H_{11}}{\partial H_7^2} &= \mathcal{O}(1), \\ \frac{\partial^2 J_{11}}{\partial J_7^2} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{\partial^2 J_{11}}{\partial J_7 \partial H_7} &= \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial^2 J_{11}}{\partial H_7^2} &= \mathcal{O}(\varepsilon). \end{aligned}$$

Proof of Proposition 6.1. The first derivative estimates follow immediately from the identity

$$\begin{pmatrix} A & B\varepsilon \\ C/\varepsilon & D \end{pmatrix} \begin{pmatrix} a & b\sqrt{\varepsilon} \\ c/\sqrt{\varepsilon} & d \end{pmatrix} = \begin{pmatrix} Aa & Ab\sqrt{\varepsilon} \\ Ca/\varepsilon & Cb/\sqrt{\varepsilon} \end{pmatrix} + HOT$$

and Propositions B.1 and B.2.

Also the estimates of $(\frac{\partial}{\partial J_0})^2$ follow directly from the above propositions. For other derivatives we obtain using Propositions B.1 and B.2

$$(B.4) \quad \begin{aligned} \frac{\partial^2 H_{11}}{\partial H_0^2} &= \frac{\partial H_{11}}{\partial J_7} \frac{\partial^2 J_7}{\partial^2 H_0^2} + \mathcal{O}(1) + \mathcal{O}\left(\varepsilon^{-7/4} \left| \frac{\partial J_7}{\partial H_0} \right|^2\right), \\ \frac{\partial^2 J_{11}}{\partial H_0^2} &= \mathcal{O}(\sqrt{\varepsilon}) + \mathcal{O}\left(\varepsilon^{-3/4} \left| \frac{\partial J_7}{\partial H_0} \right|^2\right), \\ \frac{\partial^2 H_{11}}{\partial H_0 \partial J_0} &= \mathcal{O}\left(\frac{1}{\varepsilon}\right) + \mathcal{O}\left(\varepsilon^{-7/4} \left| \frac{\partial J_7}{\partial H_0} \right|\right), & \frac{\partial^2 J_{11}}{\partial H_0 \partial J_0} &= \mathcal{O}(1) + \mathcal{O}\left(\varepsilon^{-3/4} \left| \frac{\partial J_7}{\partial H_0} \right|\right). \end{aligned}$$

Next, using Propositions B.1 and B.2 once more we get

$$(B.5) \quad \frac{\partial H_{11}}{\partial H_0} = \frac{\partial H_{11}}{\partial J_7} \frac{\partial J_7}{\partial H_0} + \frac{\partial H_{11}}{\partial H_7} \frac{\partial H_7}{\partial H_0} \sim \frac{L}{\varepsilon} \frac{\partial w}{\partial I} \frac{\partial J_7}{\partial H_0} + \mathcal{O}(1).$$

Thus Assumption (C) gives

$$\frac{\partial J_7}{\partial H_0} = \mathcal{O}(\varepsilon) \frac{\partial H_{11}}{\partial H_0} + \mathcal{O}(\varepsilon).$$

Plugging this into last four inequalities we obtain the second derivative bounds.

The asymptotic formula for $\frac{\partial J_{11}}{\partial H_0}$ follows from (B.1) and (B.5). The asymptotics for $\frac{\partial^2 H_{11}}{\partial H_0^2}$ follows from (B.2), (B.3) and (B.4). \square

APPENDIX C. DERIVATIVES OF THE INNER MAP.

C.1. Some classes of maps invariant under the compositions.

We say that a family of maps $(a, b) \rightarrow (A, B)$ depending on a parameter ε is in class \mathcal{T} if $D = \det \left(\frac{\partial(A, B)}{\partial(a, b)} \right)$ is uniformly bounded from (above and) below,

$$\begin{aligned} \left\| \frac{\partial A}{\partial a} \right\| &\leq \text{Const}, & \left\| \frac{\partial A}{\partial b} \right\| &\leq \text{Const} \sqrt{\varepsilon}, & \left\| \frac{\partial B}{\partial a} \right\| &\leq \frac{\text{Const}}{\sqrt{\varepsilon}}, & \left\| \frac{\partial B}{\partial b} \right\| &\leq \text{Const}, \\ \left\| \frac{\partial^2 A}{\partial b^2} \right\| &\leq \text{Const} \sqrt{\varepsilon}, & \left\| \frac{\partial^2 A}{\partial a \partial b} \right\| &\leq \text{Const} & \left\| \frac{\partial^2 A}{\partial a^2} \right\| &\leq \frac{\text{Const}}{\sqrt{\varepsilon}}, \\ \left\| \frac{\partial^2 B}{\partial b^2} \right\| &\leq \text{Const} & \left\| \frac{\partial^2 B}{\partial a \partial b} \right\| &\leq \frac{\text{Const}}{\sqrt{\varepsilon}} & \left\| \frac{\partial^2 B}{\partial a^2} \right\| &\leq \frac{\text{Const}}{\varepsilon}. \end{aligned}$$

We further say that this family is in \mathcal{T}_+ if addition

$$\|A\|_{C^2} \leq \text{Const}, \quad \|B\|_{C^2} \leq \text{Const}$$

(that is powers of $(\sqrt{\varepsilon})^{-1}$ are replaced by constants).

Lemma C.1. *Classes \mathcal{T} and \mathcal{T}_+ are closed with respect to compositions and inverses.*

Proof. Let $(a_0, b_0) \rightarrow (a_1, b_1)$ and $(a_1, b_1) \rightarrow (a_2, b_2)$ belong to \mathcal{T} . Then

$$\frac{\partial a_2}{\partial b_0} = \frac{\partial a_2}{\partial a_1} \frac{\partial a_1}{\partial b_0} + \frac{\partial a_2}{\partial b_1} \frac{\partial b_1}{\partial b_0}$$

and both terms are $\mathcal{O}(\sqrt{\varepsilon})$. Also

$$\frac{\partial^2 a_2}{\partial b_0^2} = \frac{\partial^2 a_2}{\partial a_1^2} \left(\frac{\partial a_1}{\partial b_0} \right)^2 + 2 \frac{\partial^2 a_2}{\partial a_1 \partial b_1} \frac{\partial a_1}{\partial b_0} \frac{\partial b_1}{\partial b_0} + \frac{\partial^2 a_2}{\partial b_1^2} \left(\frac{\partial b_1}{\partial b_0} \right)^2 + \frac{\partial a_2}{\partial a_1} \frac{\partial^2 a_1}{\partial b_0^2} + \frac{\partial a_2}{\partial b_1} \frac{\partial^2 b_1}{\partial b_0^2}$$

and each term here is $\mathcal{O}(\sqrt{\varepsilon})$. Now each time we replace a_2 by b_2 or b_0 by a_0 the estimates worsen by a factor $(\sqrt{\varepsilon})^{-1}$. This proves the estimates for other derivatives, so \mathcal{T} is closed with respect to compositions.

Next let $(a, b) \rightarrow (A, B)$ be in \mathcal{T} . We have

$$\frac{\partial a}{\partial B} = -\frac{\frac{\partial A}{\partial b}}{D} = \mathcal{O}(\sqrt{\varepsilon})$$

and the rest of the first derivatives can be estimated similarly. Next

$$\frac{\partial^2 a}{\partial B^2} = -\frac{\partial}{\partial B} \left[\frac{\frac{\partial A}{\partial b}}{D} \right] = \frac{\partial}{\partial b} \left[\frac{\frac{\partial A}{\partial b}}{D} \right] \mathcal{O}(1) + \frac{\partial}{\partial a} \left[\frac{\frac{\partial A}{\partial b}}{D} \right] \mathcal{O}(\sqrt{\varepsilon}).$$

Furthermore,

$$\frac{\partial}{\partial b} \left[\frac{\frac{\partial A}{\partial b}}{D} \right] = \frac{\frac{\partial^2 A}{\partial b^2} D - \frac{\partial A}{\partial b} \frac{\partial D}{\partial b}}{D^2} = I + II$$

where $I = \mathcal{O}(\sqrt{\varepsilon})$ and

$$\begin{aligned} II &= \mathcal{O}(\sqrt{\varepsilon}) \frac{\partial}{\partial b} \left[\frac{\frac{\partial A}{\partial a} \frac{\partial B}{\partial b} - \frac{\partial A}{\partial b} \frac{\partial B}{\partial a}}{D} \right] \\ &= \mathcal{O}(\sqrt{\varepsilon}) \left[\frac{\partial^2 A}{\partial a \partial b} \frac{\partial B}{\partial b} + \frac{\partial A}{\partial a} \frac{\partial^2 B}{\partial b^2} - \frac{\partial^2 A}{\partial b^2} \frac{\partial B}{\partial a} - \frac{\partial A}{\partial b} \frac{\partial^2 B}{\partial a \partial b} \right] = \mathcal{O}(\sqrt{\varepsilon}). \end{aligned}$$

Similarly

$$\frac{\partial}{\partial a} \left[\frac{\frac{\partial A}{\partial b}}{D} \right] = \mathcal{O}(1),$$

so $\frac{\partial^2 a}{\partial B^2} = \mathcal{O}(\sqrt{\varepsilon})$. The rest of the derivatives can be estimated using the same reasoning as for compositions. This completes the proof for \mathcal{T} . Since \mathcal{T}_+ is the intersection of \mathcal{T} with C^2 bounded maps the result follows. \square

The maps we consider also depend on two other parameters c and R . We use \mathcal{T}_0 to indicate maps in \mathcal{T} which for c sufficiently small, R sufficiently large and $\varepsilon \ll c$, $\varepsilon \ll 1/R$ satisfy

$$\begin{aligned} \left\| \frac{\partial A}{\partial a} \right\| &\sim 1, & \left\| \frac{\partial A}{\partial b} \right\| &= o(\sqrt{\varepsilon}), & \left\| \frac{\partial B}{\partial a} \right\| &= o\left(\frac{1}{\sqrt{\varepsilon}}\right), & \left\| \frac{\partial B}{\partial b} \right\| &\sim 1, \\ \left\| \frac{\partial^2 A}{\partial b^2} \right\| &= o(\sqrt{\varepsilon}), & \left\| \frac{\partial^2 A}{\partial a \partial b} \right\| &= o(1), & \left\| \frac{\partial^2 A}{\partial a^2} \right\| &= o\left(\frac{1}{\sqrt{\varepsilon}}\right), \\ \left\| \frac{\partial^2 B}{\partial b^2} \right\| &= o(1), & \left\| \frac{\partial^2 B}{\partial a \partial b} \right\| &= o\left(\frac{1}{\sqrt{\varepsilon}}\right), & \left\| \frac{\partial^2 B}{\partial a^2} \right\| &= o\left(\frac{1}{\varepsilon}\right). \end{aligned}$$

Lemma C.2. \mathcal{T}_0 is closed with respect to compositions and inverses.

Proof. The statement about the compositions is clear since in all terms we get $o(1)$ improvement compared with \mathcal{T} . For inverses observe that for maps in \mathcal{T}_0 , $D \sim 1$ so the first derivative bounds are straightforward. For the second derivatives all terms in the numerator contain second derivative of either A or B so again we get $o(1)$ improvement against \mathcal{T} . \square

C.2. Step by step analysis of the inner map.

Lemma C.3. *The map of step (7) belongs to \mathcal{T}_0 .*

Proof. We have

$$(C.1) \quad \delta J_7 = \left(1 - \varepsilon \frac{\partial}{\partial I} \left(\frac{A_1}{\omega}\right)\right) \delta I_6 - \varepsilon \frac{\partial}{\partial \phi} \left(\frac{A_1}{\omega}\right) \delta \phi_6 - \frac{\varepsilon \alpha_1}{\omega} \delta \theta_6.$$

Next, let $S = \omega - \frac{\sqrt{\varepsilon}}{R} G$. Then on $\tilde{\mathcal{S}}_1$

$$(C.2) \quad 0 = dS = \left(\frac{\partial \omega}{\partial I} - \frac{\sqrt{\varepsilon}}{R} \frac{\partial G}{\partial I}\right) \delta I_6 + \frac{\partial \omega}{\partial \phi} \delta \phi_6 + \frac{\sqrt{\varepsilon}}{R} g \delta \theta_6.$$

It follows that

$$(C.3) \quad \delta \phi_6 = - \frac{\frac{\partial \omega}{\partial I} - \frac{\sqrt{\varepsilon}}{R} \frac{\partial G}{\partial I}}{\frac{\partial \omega}{\partial \phi}} \delta I_6 - \frac{\sqrt{\varepsilon} g}{R \frac{\partial \omega}{\partial \phi}} \delta \theta_6.$$

Also $H = \frac{\omega^2}{2\varepsilon} - \theta L - G$ so

$$\delta H_6 = \frac{\omega}{\varepsilon} \left(\frac{\partial \omega}{\partial I} \delta I_6 + \frac{\partial \omega}{\partial \phi} \delta \phi_6\right) - \left(L' \theta - \frac{\partial G}{\partial I}\right) \delta I_6 - (L + g) \delta \theta_6.$$

Expressing the first term through (C.2) we get

$$\delta H_6 = \left(\frac{\omega}{\sqrt{\varepsilon} R} \frac{\partial G}{\partial I} - L' \theta - \frac{\partial G}{\partial I}\right) \delta I_6 + \left(\frac{\omega}{\varepsilon R} g - g - L\right) \delta \theta_6$$

Since on $\tilde{\mathcal{S}}_1$ $\frac{\omega}{\sqrt{\varepsilon} R} - 1 = \frac{G}{R^2}$ we have

$$\delta H_6 = \left(\frac{G \frac{\partial G}{\partial I}}{R^2} - L' \theta\right) \delta I_6 + \left(\frac{G g}{R^2} - L\right) \delta \theta_6.$$

Thus

$$(C.4) \quad \delta \theta_6 = - \frac{\delta H_6}{L - \frac{gG}{R^2}} - \frac{L' \theta - \frac{G \frac{\partial G}{\partial I}}{R^2}}{L - \frac{Gg}{R^2}} \delta I_6.$$

Plugging (C.3) into (C.1) we get

$$\delta J_7 = \left(1 - \varepsilon \frac{\partial}{\partial I} \left(\frac{A_1}{\omega}\right) + \varepsilon \frac{\partial}{\partial \phi} \left(\frac{A_1}{\omega}\right) \frac{\frac{\partial \omega}{\partial I} - \frac{\sqrt{\varepsilon}}{R} \frac{\partial G}{\partial I}}{\frac{\partial \omega}{\partial \phi}}\right) \delta I_6 - \left(\frac{\varepsilon \alpha_1}{\omega} + \frac{\sqrt{\varepsilon} g}{R \frac{\partial \omega}{\partial \phi}}\right) \delta \theta_6.$$

Using (C.4) we get

$$\begin{aligned} \delta J_7 = & \left(1 - \varepsilon \frac{\partial}{\partial I} \left(\frac{A_1}{\omega} \right) + \varepsilon \frac{\partial}{\partial \phi} \left(\frac{A_1}{\omega} \right) \frac{\frac{\partial \omega}{\partial I} - \frac{\varepsilon}{R} G}{\frac{\partial \omega}{\partial \phi}} + \left(\frac{\varepsilon \alpha_1}{\omega} + \frac{\sqrt{\varepsilon} g}{R \frac{\partial \omega}{\partial \phi}} \right) \left(\frac{L' \theta - \frac{G \frac{\partial G}{\partial I}}{R^2}}{L - \frac{gG}{R^2}} \right) \right) \delta I_6 \\ & + \left(\frac{\varepsilon \alpha_1}{\omega} + \frac{\sqrt{\varepsilon} g}{R \frac{\partial \omega}{\partial \phi}} \right) \frac{\delta H_6}{L - \frac{gG}{R^2}}. \end{aligned}$$

This proves the first derivative estimates.

To estimate the second derivatives we must differentiate these expressions once more. Observe that

$$\frac{d}{dH} = \frac{\partial \theta}{\partial H} \frac{\partial}{\partial \theta}, \quad \frac{d}{dI} = \frac{\partial}{\partial I} + \frac{\partial \theta}{\partial I} \frac{\partial}{\partial \theta},$$

so, by (C.4), $\frac{d}{dH}$ does not change ε -powers and $\frac{d}{dI}$ decreases the terms containing ω in the denominator by $(\bar{R}\sqrt{\varepsilon})^{-1}$. The result follows. \square

Corollary C.4. *The map of step (1) is in \mathcal{T}_0 .*

Proof. Similarly to Lemma C.3 we obtain that the inverse of this map is in \mathcal{T} so the result follows from Lemma C.1. \square

Lemma C.5. *The map of step (2) is in \mathcal{T}_0 .*

Proof. Write our equations as

$$(C.5) \quad \dot{I} = U, \quad \dot{H} = V, \quad \dot{\theta} = W.$$

Then we have

$$\begin{aligned} I_2 &= \hat{I}(I_1, H_1, \theta_1, \tau) \\ H_2 &= \hat{H}(I_1, H_1, \theta_1, \tau) \\ \theta_2 &= \hat{\theta}(I_1, H_1, \theta_1, \tau) \end{aligned}$$

where τ is the hit time and $(\hat{I}, \hat{H}, \hat{\theta})$ denotes the time τ map of (C.5) and θ_1 is a function of I_1 and H_1 . Thus

$$\frac{\partial H_2}{\partial H_1} = \frac{\partial \hat{H}}{\partial H_1} + \frac{\partial \hat{H}}{\partial \theta_1} \frac{\partial \theta_1}{\partial H_1} + V \frac{\partial \tau}{\partial H_1}.$$

At $\tau = 0$ we get

$$\frac{\partial H_2}{\partial H_1} = 1 + V \frac{\partial \tau}{\partial H_1}.$$

Since θ_2 is constant we get

$$(C.6) \quad 0 = \frac{\partial \theta_2}{\partial H_1} = \frac{\partial \hat{\theta}}{\partial H_1} + \frac{\partial \hat{\theta}}{\partial \theta_1} \frac{\partial \theta_1}{\partial H_1} + W \frac{\partial \tau}{\partial H_1}$$

$$= \frac{\partial \theta_1}{\partial H_1} + W \frac{\partial \tau}{\partial H_1}.$$

Therefore

$$\frac{\partial \tau}{\partial H_1} = -\frac{\frac{\partial \theta_1}{\partial H_1}}{W} = \frac{1}{W \left(L - \frac{gG}{R^2} \right)}$$

where the last equality follows similarly to (C.4). Therefore

$$\frac{\partial H_2}{\partial H_1} = 1 + \frac{V}{W \left(L - \frac{gG}{R^2} \right)}.$$

Likewise

$$\begin{aligned} \frac{\partial H_2}{\partial I_1} &= \frac{V}{W} \left(\frac{L'\theta - \frac{g\frac{\partial G}{\partial I}}{R^2}}{L - \frac{gG}{R^2}} \right), \\ \frac{\partial I_2}{\partial H_1} &= \frac{U}{W \left(L - \frac{gG}{R^2} \right)}, \\ \frac{\partial I_2}{\partial I_1} &= 1 + \frac{U}{W} \left(\frac{L'\theta - \frac{g\frac{\partial G}{\partial I}}{R^2}}{L - \frac{gG}{R^2}} \right). \end{aligned}$$

Observe that

$$(C.7) \quad U = \sum_m \sqrt{\varepsilon}^{m+1} r^m \tilde{\alpha}_1^{(m)} + \sum_m \sqrt{\varepsilon}^{m+2} r^m \tilde{\alpha}_2^{(m)},$$

$$(C.8) \quad V = \sum_m \sqrt{\varepsilon}^{m+1} r^{m+1} \tilde{\gamma}_1^{(m)} + \sum_m \sqrt{\varepsilon}^{m+1} r^m \tilde{\gamma}_2^{(m)}$$

where $\tilde{\alpha}_1^{(m)}$ and $\tilde{\gamma}_1^{(m)}$ have zero mean in θ and $W = r + \sqrt{\varepsilon}\eta$. This proves the first derivative estimates.

Next, let us show how to bound $\frac{\partial^2 I_2}{\partial H_1^2}$, other derivative bounds being similar. We begin with the identity

$$\frac{\partial I_2}{\partial H_1} = \frac{\partial \hat{I}}{\partial H_1} + \frac{\partial \hat{I}}{\partial \theta_1} \frac{\partial \theta_1}{\partial H_1} + U \frac{\partial \tau}{\partial H_1}.$$

Differentiating once more with respect to H_1 and discarding the terms vanishing at $\tau = 0$ we get

$$\begin{aligned} \frac{\partial^2 I_2}{\partial H_1^2} &= \frac{\partial}{\partial \tau} \left(\frac{\partial \hat{I}}{\partial H_1} \right) \frac{\partial \tau}{\partial H_1} + \frac{\partial}{\partial \tau} \left(\frac{\partial \hat{I}}{\partial \theta_1} \right) \frac{\partial \tau}{\partial H_1} \frac{\partial \theta_1}{\partial H_1} + \left(\frac{\partial U}{\partial I} \frac{\partial I_2}{\partial H_1} + \frac{\partial U}{\partial H} \frac{\partial H_2}{\partial H_1} \right) \frac{\partial \tau}{\partial H_1} + U \frac{\partial^2 \tau}{\partial H_1^2} \\ (C.9) \quad &= \frac{\partial U}{\partial H} \frac{\partial \tau}{\partial H_1} + \frac{\partial U}{\partial \theta} \frac{\partial \tau}{\partial H_1} \frac{\partial \theta_1}{\partial H_1} + \left(\frac{\partial U}{\partial I} \frac{\partial I_2}{\partial H_1} + \frac{\partial U}{\partial H} \frac{\partial H_2}{\partial H_1} \right) \frac{\partial \tau}{\partial H_1} + U \frac{\partial^2 \tau}{\partial H_1^2}. \end{aligned}$$

Observe that to compute the differential of U with respect to (I, H, θ) variables we can compute the differential of U with respect to (I, r, θ) variables and then replace

$$\delta r = \frac{\delta H + \left(\frac{\partial G}{\partial I} + L'\theta\right) \delta I + (g + L)\delta\theta}{r}.$$

From this it is easy to see that all terms in (C.9) except possibly the last one are $\mathcal{O}(\varepsilon^{3/4})$. It remains to bound $\frac{\partial^2 \tau}{\partial H_1^2}$. Differentiating (C.6) we obtain

$$\begin{aligned} 0 &= \frac{\partial}{\partial \tau} \left(\frac{\partial \hat{\theta}}{\partial H_1} \right) \frac{\partial \tau}{\partial H_1} + \frac{\partial}{\partial \tau} \left(\frac{\partial \hat{\theta}}{\partial \theta_1} \right) \frac{\partial \tau}{\partial H_1} \frac{\partial \theta_1}{\partial H_1} + \frac{\partial \hat{\theta}}{\partial \theta_1} \frac{\partial^2 \theta_1}{\partial H_1^2} \\ &\quad + \left(\frac{\partial W}{\partial I} \frac{\partial I_2}{\partial H_1} + \frac{\partial W}{\partial H} \frac{\partial H_2}{\partial H_1} \right) \frac{\partial \tau}{\partial H_1} + W \frac{\partial^2 \tau}{\partial H_2} \\ &= \frac{\partial W}{\partial H} \frac{\partial \tau}{\partial H_1} + \frac{\partial W}{\partial \theta} \frac{\partial \tau}{\partial H_1} \frac{\partial \theta_1}{\partial H_1} + \frac{\partial^2 \theta_1}{\partial H_1^2} + \left(\frac{\partial W}{\partial I} \frac{\partial I_2}{\partial H_1} + \frac{\partial W}{\partial H} \frac{\partial H_2}{\partial H_1} \right) \frac{\partial \tau}{\partial H_1} + W \frac{\partial^2 \tau}{\partial H_2}. \end{aligned}$$

It follows that $\frac{\partial^2 \tau}{\partial H_1^2} = \mathcal{O}(1/r)$ completing the estimate of $\frac{\partial^2 I_2}{\partial H_1^2}$. \square

Corollary C.6. *The map of step (6) is in \mathcal{T}_0 .*

Proof. This follows from Lemmas C.1 and C.5. \square

Lemma C.7. *The map of step (3) is in $\mathcal{T}_+ \cap \mathcal{T}_0$.*

Proof. From (C.7), (C.8) we get

$$(C.10) \quad \frac{dI}{d\theta} = \sum_m \sqrt{\varepsilon}^{m+1} r^{m-1} \alpha_1^{(m)} + \sum_m \sqrt{\varepsilon}^{m+2} r^{m-1} \alpha_2^{(m)}$$

$$(C.11) \quad \frac{dH}{d\theta} = \sum_m \sqrt{\varepsilon}^{m+1} r^m \gamma_1^{(m)} + \sum_m \sqrt{\varepsilon}^{m+1} r^{m-1} \gamma_2^{(m)}$$

where $\alpha_1^{(m)}$ and $\gamma_1^{(m)}$ have zero mean in θ . Observe that

$$(C.12) \quad \delta r = \frac{\delta H + \left(L'\theta + \frac{\partial G}{\partial I}\right) \delta I}{r}.$$

Hence the variational equation takes form

$$\begin{aligned} \frac{d}{d\theta} \delta H &= A \delta H + B \delta I, \\ \frac{d}{d\theta} \delta I &= C \delta H + D \delta I, \end{aligned}$$

where

$$A = \mathcal{O}(\sqrt{\varepsilon}), \quad B = \mathcal{O}(\sqrt{\varepsilon}), \quad C = \mathcal{O}\left(\varepsilon + \frac{\sqrt{\varepsilon}}{r^3}\right), \quad D = \mathcal{O}(\sqrt{\varepsilon}).$$

Let Q be the fundamental solution of

$$(C.13) \quad \frac{dQ}{d\theta} = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} Q.$$

Then $Q = \mathcal{O}(1)$, $Q^{-1} = \mathcal{O}(1)$. Substituting $\begin{pmatrix} \delta H \\ \delta I \end{pmatrix} = QZ$ we obtain

$$(C.14) \quad \frac{dZ}{d\theta} = Q^{-1} \begin{pmatrix} 0 & 0 \\ \bar{C} & 0 \end{pmatrix} QZ = \mathcal{O} \left(\varepsilon + \frac{\sqrt{\varepsilon}}{r^3} \right).$$

It follows that $(\delta H, \delta I) = (\hat{X}, \hat{Y}) + \mathcal{O}(\sqrt{\varepsilon})$ where (\hat{X}, \hat{Y}) is a solution of (C.13). But those solutions are of the form

$$(C.15) \quad \begin{pmatrix} \mathcal{O}(1)\delta H(0) + \mathcal{O}(1)\delta \hat{I}(0) \\ \mathcal{O}(1)\delta \hat{I}(0) \end{pmatrix}.$$

This proves the first derivative estimates required for \mathcal{T}_+ . For \mathcal{T}_0 estimates observe that time changes on the interval of size $o(1/\sqrt{\varepsilon})$ so the integral of (C.14) becomes $o(\sqrt{\varepsilon})$ instead of $\mathcal{O}(\sqrt{\varepsilon})$ and (C.15) becomes

$$\begin{pmatrix} 1 + o(1) & o(1) \\ 0 & 1 + o(1) \end{pmatrix}.$$

To estimate the second derivatives we begin with $(\frac{\partial}{\partial I})^2$. We have

$$(C.16) \quad \begin{aligned} \frac{d}{d\theta} \begin{pmatrix} \delta^2 H \\ \delta^2 I \end{pmatrix} &= \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \delta^2 H \\ \delta^2 I \end{pmatrix} \\ &+ \left[\frac{\partial}{\partial I} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right] \begin{pmatrix} \delta H \\ \delta I \end{pmatrix} \delta I + \left[\frac{\partial}{\partial H} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right] \begin{pmatrix} \delta H \\ \delta I \end{pmatrix} \delta H. \end{aligned}$$

Let M be the solution of

$$\frac{dM}{d\theta} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} M$$

then

$$(C.17) \quad M = \mathcal{O}(1)$$

Now arguing as before using (C.17) we get $(\delta^2 H, \delta^2 I) = \mathcal{O}(1)$. The same argument applies to $\frac{\partial}{\partial I} \frac{\partial}{\partial H}$ and $(\frac{\partial}{\partial H})^2$. However for $\frac{\partial^2 I}{\partial H^2}$ we want stronger bounds $\mathcal{O}(\sqrt{\varepsilon})$ for \mathcal{T}_+ and $o(\sqrt{\varepsilon})$ for \mathcal{T}_0 . To this end observe that by the first derivative bounds M and M^{-1} have entries

$$\begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(\sqrt{\varepsilon}) & \mathcal{O}(1) \end{pmatrix}$$

Observe that if we multiply the RHS of (C.16) by M^{-1} then the second row is

$$(C.18) \quad \mathcal{O}(\varepsilon) + \mathcal{O}(\sqrt{\varepsilon}/r^3).$$

Now the bounds of $\delta^2 I$ follows easily (for \mathcal{T}_0 we use again the fact that time changes on the interval of size $o(1/\sqrt{\varepsilon})$). \square

Corollary C.8. *The map of step (5) is in $\mathcal{T}_+ \cap \mathcal{T}_0$.*

Lemma C.9. *The map of step (4) satisfies the following.*

(a) *For fixed R we have $(I_4, E_4) = (E_3, I_3) + \mathcal{O}(\sqrt{\varepsilon})$ where ‘ \mathcal{O} ’ bound holds in C^2 topology.*

(b) *As $R \rightarrow \infty$*

$$\frac{\partial I_4}{\partial H_3} \sim \sqrt{\varepsilon} \frac{\partial \sigma}{\partial H}(H_3), \quad \frac{\partial^2 I_4}{\partial H_3^2} \sim \sqrt{\varepsilon} \frac{\partial^2 \sigma}{\partial H^2}(H_3).$$

Proof. Part (a) follows from the theorem on differentiability of solutions of ODEs with respect to parameters. To establish part (b) we prove three statements

(i) For fixed R we have

$$\frac{\partial I_4}{\partial H_3} \sim \sqrt{\varepsilon} \frac{\partial \sigma_R}{\partial H}(H_3), \quad \frac{\partial^2 I_4}{\partial H_3^2} \sim \sqrt{\varepsilon} \frac{\partial^2 \sigma_R}{\partial H^2}(H_3).$$

where σ_R stands for integral (2.13) taken between the limits s_- and s_+ where $r(s_{\mp}) = \mp R$.

(ii) As $R \rightarrow \infty$

$$\frac{\partial \sigma_R}{\partial H} \rightarrow \frac{\partial \sigma}{\partial H}, \quad \frac{\partial^2 \sigma_R}{\partial H^2} \rightarrow \frac{\partial^2 \sigma}{\partial H^2},$$

that is we can interchange differentiation and $R \rightarrow \infty$ limit.

(iii) If R_1 and R_2 are sufficiently large then

$$\frac{1}{\sqrt{\varepsilon}} \frac{\partial I_4}{\partial H_3}(R_1) \sim \frac{1}{\sqrt{\varepsilon}} \frac{\partial I_4}{\partial H_3}(R_2), \quad \frac{1}{\sqrt{\varepsilon}} \frac{\partial^2 I_4}{\partial H_3^2}(R_1) \sim \frac{1}{\sqrt{\varepsilon}} \frac{\partial^2 I_4}{\partial H_3^2}(R_2).$$

To establish (i) change variables in (2.5): $\hat{I} = \frac{I - I_4}{\sqrt{\varepsilon}}$. Then

$$\hat{I}' = \bar{\alpha}(I_4 + \sqrt{\varepsilon} \hat{I}, \sqrt{\varepsilon} r, \theta, \varepsilon)$$

and we also replace I by $I_4 + \sqrt{\varepsilon} \hat{I}$ in the RHS of (2.5). Now as $\varepsilon \rightarrow 0$ the equation (2.5) converges to

$$\theta' = r, \quad r' = L + g, \quad \hat{I}' = \alpha_1(I_4, 0, \theta)$$

so the result follows by differentiable dependence of solutions on parameters.

To get (ii) rewrite the expression for σ using θ as the time variable (see (2.14)). Since $r = \pm\sqrt{2(H + G + L\theta)}$ we need to estimate the H -derivatives of

$$\sigma(I, H) = \int_{\theta(R_1)}^{\theta(R_2)} \frac{\alpha_1(I, 0, \theta)}{\sqrt{2(H + G + L\theta)}} d\theta$$

where $\theta(R) = (H + G - \frac{R^2}{2})/L$. Now the first (second) derivative of the integrand decays as $\theta^{-3/2}$ ($\theta^{-5/2}$) so

$$\begin{aligned} \left| \frac{\partial}{\partial H}(\sigma - \sigma_R) \right| &\leq \frac{\text{Const}}{\theta^{-1/2}} && (\sim \frac{\text{Const}}{R}) \\ \left| \frac{\partial^2}{\partial H^2}(\sigma - \sigma_R) \right| &\leq \frac{\text{Const}}{\theta^{-3/2}} && (\sim \frac{\text{Const}}{R^3}) \end{aligned}$$

Thus (ii) follows. The proof of (iii) is similar to the proof of Lemma C.7. (b) is proven. \square

Proof of Proposition B.1. Combining Lemma C.1–Lemma C.9 we get that the map $(J_0, H_0) \rightarrow (J_7, H_7)$ is in \mathcal{T} . This gives the inequalities claimed in Proposition B.1. To get the asymptotic formulas observe that each of $\frac{\partial J_7}{\partial J_0}$, $\frac{\partial H_7}{\partial H_0}$, $\frac{\partial J_7}{\partial H_0}$ is a sum of monomials in matrix elements computed at steps (1)–(7). Since the composition is in \mathcal{T} we know that each monomial of $\frac{\partial J_7}{\partial J_0}$ and $\frac{\partial H_7}{\partial H_0}$ is $\mathcal{O}(1)$ and monomial of $\frac{\partial J_7}{\partial H_0}$ is $\mathcal{O}(\sqrt{\varepsilon})$. To avoid an extra $o(1)$ factor coming from \mathcal{T}_0 the factors should stay on the diagonal except for step (4) since all off-diagonal terms in steps (1)–(3) and (5)–(7) have the extra $o(1)$ factors. Thus

$$\frac{\partial J_7}{\partial J_0} \sim \frac{\partial I_1}{\partial J_0} \frac{\partial I_2}{\partial I_1} \frac{\partial I_3}{\partial I_2} \frac{\partial I_4}{\partial I_3} \frac{\partial I_5}{\partial I_4} \frac{\partial I_6}{\partial I_5} \frac{\partial J_7}{\partial I_6},$$

$$\frac{\partial H_7}{\partial H_0} \sim \prod_{j=0}^6 \frac{\partial H_{j+1}}{\partial H_j},$$

$$\frac{\partial J_7}{\partial H_0} \sim \frac{\partial H_1}{\partial H_0} \frac{\partial H_2}{\partial H_1} \frac{\partial H_3}{\partial H_2} \frac{\partial I_4}{\partial H_3} \frac{\partial I_5}{\partial I_4} \frac{\partial I_6}{\partial I_5} \frac{\partial J_7}{\partial I_6}.$$

Together with (A.1) this proves the results about the first derivatives. A similar reasoning gives

$$\frac{\partial^2 J_7}{\partial H_0^2} \sim \frac{\partial J_7}{\partial I_6} \frac{\partial I_6}{\partial I_5} \frac{\partial I_5}{\partial I_4} \frac{\partial^2 I_4}{\partial H_3^2} \left(\frac{\partial H_3}{\partial H_2} \frac{\partial H_2}{\partial H_1} \frac{\partial H_1}{\partial H_0} \right)^2.$$

\square

APPENDIX D. DERIVATIVES OF THE OUTER MAP.

In order to analyze the map of step (9) we first consider a more general setting of equations

$$(D.1) \quad \dot{x} = a_1(x, \theta) + a_2(x) + \varepsilon a_3(x, \theta, \varepsilon)$$

$$(D.2) \quad \dot{\theta} = \frac{\omega(x)}{\varepsilon} + \eta(x, \theta, \varepsilon)$$

where a_1 has zero mean in θ . Introduce the improved variables $y = x - \frac{\varepsilon}{\omega} A_1$ where $\frac{\partial A_1}{\partial \theta} = a_1$. We want to study time t maps in a region where $|\omega| > c\varepsilon^{1/4}$. Let (y, θ) denote the original variables and $(\bar{y}, \bar{\theta})$ denote the final variables.

Lemma D.1. (a) *We have*

$$\frac{\partial \bar{y}}{\partial y} = \mathcal{O}(1), \quad \frac{\partial \bar{y}}{\partial \theta} = \mathcal{O}(\varepsilon), \quad \frac{\partial \bar{\theta}}{\partial y} = \mathcal{O}(1/\varepsilon), \quad \frac{\partial \bar{\theta}}{\partial \theta} = \mathcal{O}(1).$$

$$\frac{\partial^2 \bar{y}^2}{\partial y^2} = \mathcal{O}(\varepsilon^{-3/4}), \quad \frac{\partial^2 \bar{y}^2}{\partial y \partial \theta} = \mathcal{O}(\varepsilon^{1/4}), \quad \frac{\partial^2 \bar{y}^2}{\partial \theta^2} = \mathcal{O}(\varepsilon^{5/4}).$$

$$\frac{\partial^2 \bar{\theta}^2}{\partial y^2} = \mathcal{O}(\varepsilon^{-7/4}), \quad \frac{\partial^2 \bar{\theta}^2}{\partial y \partial \theta} = \mathcal{O}(\varepsilon^{-3/4}), \quad \frac{\partial^2 \bar{\theta}^2}{\partial \theta^2} = \mathcal{O}(\varepsilon^{1/4}).$$

(b) *If x is a pair (I, ϕ) from the equation (2.1) then*

$$\frac{\partial \bar{\theta}}{\partial J} \sim \frac{1}{\varepsilon} \frac{\partial w}{\partial I}, \quad \frac{\partial \bar{\theta}}{\partial \psi} = o\left(\frac{1}{\varepsilon}\right), \quad \frac{\partial \bar{J}}{\partial J} = 1 + o(1), \quad \frac{\partial \bar{J}}{\partial \psi} = o(1).$$

Proof. (a) We have

$$\dot{y} = a_2(x) + \varepsilon \left[a_3 - \frac{\partial}{\partial x} \left(\frac{A_1}{\omega} \right) \right].$$

Denote $q = a_3 - \frac{\partial}{\partial x} \left(\frac{A_1}{\omega} \right)$. Then the variational equation takes form

$$\dot{\delta y} = \left(\frac{\partial a_2}{\partial x} + \varepsilon \frac{\partial q}{\partial x} \right) \delta x + \varepsilon \frac{\partial q}{\partial \theta} \delta \theta,$$

$$\dot{\delta \theta} = \left(\frac{1}{\varepsilon} \frac{\partial \omega}{\partial x} + \frac{\partial \eta}{\partial x} \right) \delta x + \frac{\partial \eta}{\partial \theta} \delta \theta.$$

Observe that

$$\delta y = \left(1 - \varepsilon \frac{\partial}{\partial x} \left(\frac{A_1}{\omega} \right) \right) \delta x - \frac{\varepsilon a_1}{\omega} \delta \theta$$

so that

$$(D.3) \quad \delta x = \left(1 - \varepsilon \frac{\partial}{\partial x} \left(\frac{A_1}{\omega} \right) \right)^{-1} \left(\delta y + \frac{\varepsilon a_1}{\omega} \delta \theta \right).$$

Introducing $Z = \varepsilon \delta \theta$ we get

$$\dot{\delta y} = \frac{\partial a_2}{\partial x} \delta y + \left(\frac{\partial a_2}{\partial x} \frac{a_1}{\omega} + \frac{\partial q}{\partial \theta} \right) Z + HOT,$$

$$\dot{Z} = \frac{\partial \omega}{\partial x} \delta y + \left[\frac{\partial \eta}{\partial \theta} + \frac{\partial \omega}{\partial x} \frac{a_1}{\omega} \right] Z + HOT.$$

Consider

$$\tilde{Y} = \delta y - \frac{\varepsilon}{\omega} \left[\frac{\partial a_2}{\partial x} \frac{A_1}{\omega} + q \right] Z, \quad \tilde{Z} = \left(1 - \frac{\varepsilon}{\omega} \left[\eta + \frac{\partial \omega}{\partial x} \frac{A_1}{\omega} \right] \right) Z$$

We obtain an equation

$$(D.4) \quad \frac{d}{dt} \begin{pmatrix} \tilde{Y} \\ \tilde{Z} \end{pmatrix} = \mathcal{R} \begin{pmatrix} \tilde{Y} \\ \tilde{Z} \end{pmatrix}.$$

where $R = \mathcal{O}(1 + \frac{\varepsilon}{\omega^4})$ where ω^4 appears in the denominator due to the differentiation with respect to x (recall that due to (D.4) $\delta x = \delta y + \frac{Z}{\omega} + HOT$). Hence to establish the statement about the first derivatives it is enough to show that

$$\int |\mathcal{R}| dt = \mathcal{O}(1).$$

But indeed

$$(D.5) \quad \int \frac{\varepsilon}{\omega^4} dt = \int \frac{\varepsilon^2}{\omega^5} d\theta \\ = \mathcal{O} \left(\frac{1}{\sqrt{\varepsilon}} \int \frac{d\theta}{\theta^{5/2}} \right) = \mathcal{O} \left((\sqrt{\varepsilon} \theta^{3/2})^{-1} \right) = \mathcal{O} \left(\frac{\varepsilon}{\omega^3} \right) = \mathcal{O}(\varepsilon^{1/4}).$$

Since the solution to the variational equation are obtained from the solutions of (D.4) by conjugation by $\begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}$ the result follows.

Let us now estimate $(\frac{\partial}{\partial \theta})^2$. The second variational equation takes form

$$\delta^2 \dot{y} = \left(\frac{\partial a_2}{\partial x} + \varepsilon \frac{\partial q}{\partial x} \right) \delta^2 x + \varepsilon \frac{\partial q}{\partial \theta} \delta^2 \theta \\ + \frac{\partial}{\partial x} \left(\frac{\partial a_2}{\partial x} + \varepsilon \frac{\partial q}{\partial x} \right) (\delta x)^2 + 2\varepsilon \frac{\partial^2 q}{\partial x \partial \theta} (\delta x)(\delta \theta) + \varepsilon \frac{\partial^2 q}{\partial \theta^2} (\delta \theta)^2, \\ \delta^2 \dot{\theta} = \left(\frac{1}{\varepsilon} \frac{\partial \omega}{\partial x} + \frac{\partial \eta}{\partial x} \right) \delta^2 x + \frac{\partial \eta}{\partial \theta} \delta^2 \theta \\ + \frac{\partial}{\partial x} \left(\frac{1}{\varepsilon} \frac{\partial \omega}{\partial x} + \frac{\partial \eta}{\partial x} \right) (\delta x)^2 + 2 \frac{\partial^2 \eta}{\partial x \partial \theta} (\delta x)(\delta \theta) + 2 \frac{\partial^2 \eta}{\partial \theta^2} (\delta \theta)^2$$

where δx is related to δy by (D.3) and $\delta^2 x$ is related to $\delta^2 y$ by

$$\begin{aligned} \delta^2 y &= \left(1 - \varepsilon \frac{\partial}{\partial x} \left(\frac{A_1}{\omega}\right)\right) \delta^2 x - \frac{\varepsilon a_1}{\omega} \delta^2 \theta \\ &\quad - \varepsilon \left(\frac{\partial}{\partial x}\right)^2 \left(\frac{A_1}{\omega}\right) (\delta x)^2 - 2\varepsilon \frac{\partial}{\partial x} \left(\frac{a_1}{\omega}\right) (\delta x)(\delta \theta) - \frac{\varepsilon}{\omega} \frac{\partial a_1}{\partial \theta} (\delta \theta)^2. \end{aligned}$$

Substituting $X = \varepsilon \delta^2 \theta$ and keeping in mind that $\delta y = \mathcal{O}(\varepsilon)$ by the first derivative estimate we obtain an inhomogenous system whose fundamental solution is uniformly bounded and the inhomogenous terms are $\mathcal{O}(\frac{\varepsilon^2}{\omega^4})$ except for $\varepsilon \left[\frac{1}{\omega} \frac{\partial a_2}{\partial x} \frac{\partial a_1}{\partial \theta} + \frac{\partial^2 q}{\partial \theta^2}\right] (\delta \theta)^2$ in the first equation and $\varepsilon \frac{\partial}{\partial \theta} \left(\frac{\partial \eta}{\partial \theta} + \frac{\partial \omega}{\partial x} \frac{a_1}{\omega}\right) (\delta \theta)^2$ in the second equation. Introducing

$$\hat{Y} = \delta^2 y - \varepsilon^2 \left[\frac{a_1}{\omega} \frac{\partial a_2}{\partial x} + \frac{\partial q}{\partial \theta}\right] \frac{(\delta \theta)^2}{\omega}, \quad \hat{X} = X - \varepsilon^2 \left(\frac{\partial \eta}{\partial \theta} + \frac{\partial \omega}{\partial x} \frac{a_1}{\omega}\right) \frac{(\delta \theta)^2}{\omega}$$

we get

$$\frac{d\hat{Y}}{dt} = \mathcal{O}\left(\frac{\varepsilon^2}{\omega^4}\right), \quad \frac{d\hat{X}}{dt} = \mathcal{O}\left(\frac{\varepsilon^2}{\omega^4}\right).$$

By (D.5) $\hat{Y} = \mathcal{O}(\varepsilon^{5/4})$, $\hat{X} = \mathcal{O}(\varepsilon^{5/4})$. Observe that $\hat{Y} - \delta^2 y = \mathcal{O}(\varepsilon^{5/4})$ due to $\frac{1}{\omega} \frac{\partial q}{\partial \theta}$ term and $\hat{X} - X = \mathcal{O}(\varepsilon^{3/2})$ due to $\frac{1}{\omega^2}$ term. This proves the estimate for $(\frac{\partial}{\partial \theta})^2$. Now if we replace $\frac{\partial}{\partial \theta}$ by $\frac{\partial}{\partial y}$ the estimates are similar except that each replacement increases the RHS by a factor of ε^{-1} due to the first derivative estimates. This completes the proof of (a).

Next, (D.4) reads in the setting of (b) as follows (we put

$$\delta y = (Y_J, Y_\psi))$$

$$\dot{Y}_J = \dots$$

$$\dot{Y}_\psi = \frac{\partial p}{\partial I} Y_J + \dots$$

$$\dot{Z} = \frac{\partial \omega}{\partial I} Y_J + \frac{\partial \omega}{\partial \phi} Y_\psi + \dots$$

where \dots denote lower order terms. Hence

$$Y_J(t) \sim Y_J(0), \quad Y_\psi(t) = Y_\psi(0) + s \frac{\partial p}{\partial I} Y_J(0),$$

$$Z(t) \sim Z(0) + \left[\int_0^t \frac{\partial \omega}{\partial I} ds + \int_0^t \frac{\partial p}{\partial I} \frac{\partial \omega}{\partial \phi} s ds \right] Y_J(0) + \left[\int_0^t \frac{\partial \omega}{\partial \phi} ds \right] Y_\psi(0).$$

In terms of our original variables this says $\delta J(t) \sim \delta J(0)$,

$$(D.6) \quad \delta \theta(t) - \delta \theta(0)$$

$$\sim \frac{1}{\varepsilon} \left(\left[\int_0^t \frac{\partial \omega}{\partial I} ds + \int_0^t \frac{\partial p}{\partial I} s \frac{\partial \omega}{\partial \phi} ds \right] \delta J(0) + \left[\int_0^t \frac{\partial \omega}{\partial \phi} ds \right] \delta \psi(0) \right).$$

We need to apply this formula with t being the time it takes to pass from one resonance to the next. Observe that the integrals in (D.6) can be approximated by the corresponding integrals for the averaged system (2.3). But in this case $ds = \frac{d\bar{\phi}}{p(\bar{I})}$. Hence

$$\int_0^t \frac{\partial \omega}{\partial \phi}(\bar{I}(s), \bar{\phi}(s)) ds = \frac{1}{p(\bar{I})} \int_{\phi_-}^{\phi_+} \frac{\partial \omega}{\partial \phi}(\bar{I}, \bar{\phi}) d\bar{\phi} = 0$$

since

$$(D.7) \quad \omega(\phi_-) = \omega(\phi_+) = 0.$$

Similarly integrating by parts and using (D.7) we get

$$\frac{\partial p}{\partial I}(\bar{I}) \int_0^t s \frac{\partial \omega}{\partial \phi}(\bar{I}(s), \bar{\phi}(s)) ds = -\frac{\partial p}{\partial I} \int_{\phi_-}^{\phi_+} \omega(\bar{I}, \bar{\phi}) d\bar{\phi}.$$

Finally the first term in (D.6) can be rewritten as

$$\frac{1}{p} \int_{\phi_-}^{\phi_+} \omega(\bar{I}, \bar{\phi}) d\phi$$

so we get

$$\delta\theta(t) - \delta\theta(0) \sim \frac{1}{\varepsilon} \left(\int_{\phi_-}^{\phi_+} \frac{\frac{\partial \omega}{\partial I} p - \omega \frac{\partial p}{\partial I}}{p^2} d\phi \right) \delta J(0) = \frac{1}{\varepsilon} \frac{\partial w}{\partial I} \delta J(0)$$

as claimed. \square

Lemma D.2. *The map of step (10) satisfies*

$$\frac{\partial J_{10}}{\partial J_9} = 1 + \mathcal{O}(\sqrt{\varepsilon}), \quad \frac{\partial J_{10}}{\partial \psi_9} = \mathcal{O}(\sqrt{\varepsilon}), \quad \frac{\partial J_{10}}{\partial \theta_9} = \mathcal{O}(\varepsilon^{3/2}),$$

$$\frac{\partial \theta_{10}}{\partial y_9} = \mathcal{O}(\varepsilon^{-3/4}), \quad \frac{\partial \theta_{10}}{\partial \theta_9} = 1 + \mathcal{O}(\varepsilon^{1/4}).$$

$$\frac{\partial^2 J_{10}}{\partial y_9^2} = \mathcal{O}(\varepsilon^{-1/4}), \quad \frac{\partial^2 J_{10}}{\partial y_9 \partial \theta_9} = \mathcal{O}(\varepsilon^{3/4}), \quad \frac{\partial^2 J_{10}}{\partial \theta_9^2} = \mathcal{O}(\varepsilon^{3/2}).$$

$$\frac{\partial^2 \theta_{10}}{\partial y_9^2} = \mathcal{O}(\varepsilon^{-3/2}), \quad \frac{\partial^2 \theta_{10}}{\partial y_9 \partial \theta_9} = \mathcal{O}(\varepsilon^{-1/2}), \quad \frac{\partial^2 \theta_{10}}{\partial \theta_9^2} = \mathcal{O}(\varepsilon^{1/4}).$$

Proof. We have

$$J_{10} = \hat{J}(y_9, \theta_9, \tau), \quad \theta_{10} = \hat{\theta}(y_9, \theta_9, \tau)$$

where τ is the hit time and \hat{J} and $\hat{\theta}$ are the components of time τ map of our system which we denote by

$$(D.8) \quad \dot{J} = U(y, \theta), \quad \dot{\theta} = V(y, \theta).$$

Observe that

$$(D.9) \quad U = \mathcal{O}(\sqrt{\varepsilon}), \quad V = \mathcal{O}(\varepsilon^{-3/4}).$$

(D.8) gives

$$(D.10) \quad \frac{\partial J_{10}}{\partial y_9} = \frac{\partial \hat{J}}{\partial y_9} + U \frac{\partial \tau}{\partial y_9}, \quad \frac{\partial J_{10}}{\partial \theta_9} = \frac{\partial \hat{J}}{\partial \theta_9} + U \frac{\partial \tau}{\partial \theta_9}.$$

$$(D.11) \quad \frac{\partial \theta_{10}}{\partial y_9} = \frac{\partial \hat{\theta}}{\partial y_9} + V \frac{\partial \tau}{\partial y_9}, \quad \frac{\partial \theta_{10}}{\partial \theta_9} = \frac{\partial \hat{\theta}}{\partial \theta_9} + V \frac{\partial \tau}{\partial \theta_9}.$$

To find the partial derivatives of τ let $S = \omega + \frac{\sqrt{\varepsilon}G}{R} + \sqrt{\varepsilon}\bar{R}$. Then

$$\frac{dS}{dt} = \frac{\partial \omega}{\partial I} \alpha_1 + \frac{\partial \omega}{\partial \phi} (p + \beta_1) + \frac{g\omega}{R\sqrt{\varepsilon}} + o(1) = L(I) + \frac{Sg}{\sqrt{\varepsilon}R} + o(1).$$

On S_2 we have $S = 0$ and so its partial derivatives vanish as well. Thus

$$\frac{\partial S}{\partial y_9} + \frac{dS}{dt} \frac{\partial \tau}{\partial y_9} = 0$$

and so

$$\frac{\partial \tau}{\partial y_9} = - \frac{\frac{\partial S}{\partial y_9}}{L(I) + \frac{Sg}{\sqrt{\varepsilon}R} + o(1)}.$$

Likewise

$$\frac{\partial \tau}{\partial \theta_9} = - \frac{\frac{\partial S}{\partial \theta_9}}{L + \frac{Sg}{\sqrt{\varepsilon}R} + o(1)}.$$

A direct computation gives $\frac{\partial S}{\partial y_9} = \mathcal{O}(1)$. We also claim that on S_2

$$(D.12) \quad \frac{\partial S}{\partial \theta_9} = \mathcal{O}(\varepsilon).$$

Indeed

$$dS = \left(\frac{\partial \omega}{\partial I} + \mathcal{O}(\varepsilon^{3/4}) \right) \delta I + \left(\frac{\partial \omega}{\partial \phi} + \mathcal{O}(\varepsilon^{3/4}) \right) \delta \phi + \frac{\varepsilon g}{R} \delta \theta.$$

Recalling the definitions of J, ψ and g we get we see that

$$\delta I = \delta J + \frac{\varepsilon \alpha_1}{\omega} \delta \theta + HOT, \quad \delta \phi = \delta \psi + \frac{\varepsilon \beta_1}{\omega} \delta \theta + HOT$$

and hence

$$(D.13) \quad \frac{\partial S}{\partial \theta_9} = \varepsilon g \left(\frac{1}{\omega} + \frac{1}{\sqrt{\varepsilon}R} \right) + \mathcal{O}(\varepsilon).$$

On S_2 the first term equals to $-\frac{\varepsilon g G}{R^2 \omega} = \mathcal{O}(\varepsilon^{5/4})$ proving (D.12). Now the first derivative estimates follow easily from (D.10) and (D.11).

To obtain the second derivative estimates we differentiate (D.10) and (D.11) once more. We have

$$(D.14) \quad \frac{\partial}{\partial \theta_9} = \frac{\partial \theta_{10}}{\partial \theta_9} \frac{\partial}{\partial \theta_{10}} + \frac{\partial y_{10}}{\partial \theta_9} \frac{\partial}{\partial y_{10}}$$

$$(D.15) \quad \frac{\partial}{\partial y_9} = \frac{\partial \theta_{10}}{\partial y_9} \frac{\partial}{\partial \theta_{10}} + \frac{\partial y_{10}}{\partial y_9} \frac{\partial}{\partial y_{10}}$$

Therefore $\frac{\partial}{\partial \theta_9}$ does not worsen the first derivative estimates while taking $\frac{\partial}{\partial y_9}$ we loose $\varepsilon^{-3/4}$ due to the first term in (D.15). \square

Lemma D.3. *We have*

$$\delta J_{10} \sim \delta J_8 + o(1)\delta \phi_8 + \mathcal{O}(\varepsilon)\delta \theta_8,$$

$$\delta \theta_{10} \sim \frac{1}{\varepsilon} \frac{\partial w}{\partial I} \delta I_8 + o\left(\frac{1}{\varepsilon}\right) \delta \phi_8 + \mathcal{O}(1)\delta \theta_8.$$

$$\begin{aligned} \frac{\partial^2 J_{10}}{\partial y_8^2} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{\partial^2 J_{10}}{\partial y_8 \partial \theta_8} &= \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial^2 J_{10}}{\partial \theta_8^2} &= \mathcal{O}(\varepsilon^{5/4}). \\ \frac{\partial^2 \theta_{10}}{\partial y_8^2} &= \mathcal{O}(\varepsilon^{-7/4}), & \frac{\partial^2 \theta_{10}}{\partial y_8 \partial \theta_8} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{\partial^2 \theta_{10}}{\partial \theta_8^2} &= \mathcal{O}(\varepsilon^{1/4}). \end{aligned}$$

Proof. This follows from Lemmas D.1 and D.2 by direct computation. \square

Lemma D.4. *The composition $(J_8, \theta_8) \rightarrow (J_{10}, \theta_{10})$ satisfies*

$$\frac{dJ_{10}}{dJ_8} = 1 + o(1), \quad \frac{dJ_{10}}{d\theta_8} = \mathcal{O}(\varepsilon), \quad \frac{d\theta_{10}}{dJ_8} = \frac{1}{\varepsilon} \frac{\partial w}{\partial I} + o(\varepsilon^{-1}), \quad \frac{d\theta_{10}}{d\theta_8} = \mathcal{O}(1).$$

$$\begin{aligned} \frac{d^2 J_{10}}{dJ_8^2} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{d^2 J_{10}}{dJ_8 d\theta_8} &= \mathcal{O}(\varepsilon^{1/4}), & \frac{d^2 J_{10}}{d\theta_8^2} &= \mathcal{O}(\varepsilon). \\ \frac{d^2 \theta_{10}}{dJ_8^2} &= \mathcal{O}(\varepsilon^{-7/4}), & \frac{d^2 \theta_{10}}{dJ_8 d\theta_8} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{d^2 \theta_{10}}{d\theta_8^2} &= \mathcal{O}(1). \end{aligned}$$

Proof. The difference with lemma D.3 is that now ψ_8 is a function of J_8 and θ_8 so

$$\frac{d}{d\theta_8} = \frac{\partial}{\partial \theta_8} + \frac{\partial \psi_8}{\partial \theta_8} \frac{\partial}{\partial \psi_8}, \quad \frac{d}{dJ_8} = \frac{\partial}{\partial J_8} + \frac{\partial \psi_8}{\partial J_8} \frac{\partial}{\partial \psi_8}.$$

The same computations as in Lemmas C.3 and D.2 give

$$(D.16) \quad \frac{\partial \psi_8}{\partial J_8} = \mathcal{O}(1), \quad \frac{\partial^2 \psi_8}{\partial J_8^2} = \mathcal{O}(1).$$

Let $\tilde{S} = \omega - \sqrt{\varepsilon}\bar{R} - \frac{\sqrt{\varepsilon}G}{R}$. Arguing as in (D.13) we get

$$(D.17) \quad \frac{\partial\psi_8}{\partial\theta_8} = -\frac{\frac{\partial\tilde{S}}{\partial\theta_8}}{\frac{\partial\tilde{S}}{\partial\psi_8}} = \frac{\varepsilon g \left(\frac{1}{\omega} - \frac{1}{\sqrt{\varepsilon}\bar{R}} \right) + \mathcal{O}(\varepsilon)}{\frac{\partial\tilde{S}}{\partial\psi_8}} = \mathcal{O}(\varepsilon),$$

$\frac{\partial^2\psi_8}{\partial\theta_8\partial J_8} = \mathcal{O}(\sqrt{\varepsilon})$ due to $\varepsilon\frac{d}{dJ_8}(1/\omega)$ term, and $\frac{\partial^2\psi_8}{\partial\theta_8^2} = \mathcal{O}(\varepsilon)$. (D.16) and (D.17) immediately imply the estimates for the first derivatives. Next,

$$\begin{aligned} \frac{d^2\theta_{10}}{d\theta_8^2} &= \frac{\partial^2\theta_{10}}{\partial\theta_8^2} + 2\frac{\partial\psi_8}{\partial\theta_8}\frac{\partial^2\theta_{10}}{\partial\theta_8\partial\psi_8} + \left(\frac{\partial\psi_8}{\partial\theta_8}\right)^2\frac{\partial^2\theta_{10}}{\partial\psi_8^2} + \frac{\partial^2\psi_8}{\partial\theta_8^2}\frac{\partial\theta_{10}}{\partial\psi_8} = \mathcal{O}(1), \\ \frac{d^2J_{10}}{d\theta_8^2} &= \frac{\partial^2J_{10}}{\partial\theta_8^2} + 2\frac{\partial\psi_8}{\partial\theta_8}\frac{\partial^2J_{10}}{\partial\theta_8\partial\psi_8} + \left(\frac{\partial\psi_8}{\partial\theta_8}\right)^2\frac{\partial^2J_{10}}{\partial\psi_8^2} + \frac{\partial^2\psi_8}{\partial\theta_8^2}\frac{\partial J_{10}}{\partial\psi_8} = \mathcal{O}(\varepsilon) \end{aligned}$$

the leading term in both cases being the last one. Other derivatives are easier since now $\frac{\partial^2\psi_8}{\partial\theta_8\partial\psi_8}$ does not spoil the main term. \square

Lemma D.5. *The map of step (11) satisfies*

$$\delta J_{11} = \delta J_{10},$$

the second derivatives of J_{11} vanish,

$$\begin{aligned} \frac{\partial H_{11}}{\partial\theta_{10}} &= \bar{L} + \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial H_{11}}{\partial J_{10}} &= \mathcal{O}(\varepsilon^{-3/4}), \\ \frac{\partial^2 H_{11}}{\partial\theta_{10}^2} &= \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial^2 H_{11}}{\partial\theta_{10}\partial J_{10}} &= \mathcal{O}(\varepsilon^{-1/4}), & \frac{\partial^2 H_{11}}{\partial J_{10}^2} &= \mathcal{O}(\varepsilon^{-1}). \end{aligned}$$

Proof. We have

$$H = \frac{\omega^2}{2\varepsilon} - L\theta - G.$$

Direct differentiation implies all first derivative estimates except for $\frac{\partial H}{\partial\theta}$. To obtain this last estimate we rewrite using (2.10)

$$\omega^2(I, \phi) = \omega^2(J, \psi) + 2\varepsilon G + \mathcal{O}(\varepsilon^{5/4}).$$

Thus

$$H = \frac{\omega^2(J, \psi)}{2\varepsilon} - L\theta + \mathcal{O}(\varepsilon^{1/4}).$$

Now the rest of the proof proceeds as in Lemma C.3. \square

Lemma D.6.

$$\begin{aligned} \frac{\partial H_{11}}{\partial\theta_8} &= \mathcal{O}(1), & \frac{\partial H_{11}}{\partial J_8} &= \frac{\bar{L}}{\varepsilon} \left(\frac{\partial w}{\partial I} + o(1) \right), & \frac{\partial J_{11}}{\partial\theta_8} &= \mathcal{O}(\varepsilon), & \frac{\partial J_{11}}{\partial J_8} &\sim 1. \\ \frac{\partial^2 H_{11}}{\partial J_8^2} &= \mathcal{O}(\varepsilon^{-7/4}), & \frac{\partial^2 H_{11}}{\partial\theta_8\partial J_8} &= \mathcal{O}(\varepsilon^{-3/4}), & \frac{\partial^2 H_{11}}{\partial\theta_8^2} &= \mathcal{O}(1), \end{aligned}$$

$$\frac{\partial^2 J_{11}}{\partial J_8^2} = \mathcal{O}(\varepsilon^{-3/4}), \quad \frac{\partial^2 J_{11}}{\partial \theta_8 \partial J_8} = \mathcal{O}(\varepsilon^{1/4}), \quad \frac{\partial^2 J_{11}}{\partial \theta_8^2} = \mathcal{O}(\varepsilon).$$

Proof. Direct computation. \square

Lemma D.7. *The map of step (8) satisfies*

$$\delta J_8 = \delta J_7,$$

the second derivatives of J_8 vanish,

$$\begin{aligned} \frac{\partial \theta_8}{\partial H_7} &= \mathcal{O}(1), & \frac{\partial \theta_8}{\partial J_7} &= \mathcal{O}(\varepsilon^{-3/4}). \\ \frac{\partial^2 \theta_8}{\partial H_7^2} &= \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial^2 \theta_8}{\partial H_7 \partial J_7} &= \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right), & \frac{\partial^2 \theta_8}{\partial J_7^2} &= \mathcal{O}(\varepsilon^{-5/4}). \end{aligned}$$

Proof. Similarly to Lemma D.5 we obtain

$$\begin{aligned} \frac{\partial H_7}{\partial \theta_8} &= L + \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial H_7}{\partial J_8} &= \mathcal{O}(\varepsilon^{-3/4}), \\ \frac{\partial^2 H_7}{\partial \theta_8^2} &= \mathcal{O}(\varepsilon^{1/4}), & \frac{\partial^2 H_7}{\partial \theta_8 \partial J_8} &= \mathcal{O}(\varepsilon^{-1/4}), & \frac{\partial^2 H_7}{\partial J_8^2} &= \mathcal{O}(\varepsilon^{-1}). \end{aligned}$$

Next

$$\begin{aligned} \frac{\partial \theta_8}{\partial H_7} &= \left(\frac{\partial H_7}{\partial \theta_8} \right)^{-1} = \mathcal{O}(1), \\ \frac{\partial \theta_8}{\partial J_7} &= -\frac{\frac{\partial H_7}{\partial J_8}}{\frac{\partial H_7}{\partial \theta_8}} = \mathcal{O}(\varepsilon^{-3/4}). \end{aligned}$$

Now the result follows by direct computation using the formulas for derivatives of the inverse mapping. \square

Proof of Proposition B.2. The result follows from Lemmas D.6 and D.7 by direct computation. \square

APPENDIX E. DYNAMICS NEAR THE SEPARATRIX.

E.1. Normal form. In this section we describe the dynamics near the separatrix of the inner map.

Lemma E.1. *There exist functions $x(r, \theta, I, \varepsilon)$, $y(r, \theta, I, \varepsilon)$, such that in coordinates (x, y, I) the following holds*

$$(E.1) \quad \begin{aligned} \dot{x} &= a(x, y, I, \varepsilon) \\ \dot{y} &= b(x, y, I, \varepsilon) \\ \dot{I} &= \sqrt{\varepsilon} c(x, y, I, \varepsilon) \end{aligned}$$

where

$$(E.2) \quad a(0, y, I, \varepsilon) = b(x, 0, I, \varepsilon) = 0,$$

$$(E.3) \quad \frac{\partial a}{\partial x}(x, y, I, 0) + \frac{\partial b}{\partial y}(x, y, I, 0) = 0$$

and

$$(E.4) \quad a(x, 0, I, \varepsilon) = \lambda_1(I, \varepsilon)x, \quad b(0, y, I, \varepsilon) = -\lambda_2(I, \varepsilon)y,$$

Denote $\lambda(I) = \lambda_1(I, 0) = \lambda_2(I, 0)$. Due to assumption (D) there exists λ_0 such that $\lambda(I) > \lambda_0$.

Proof. We first consider the case where I is fixed and $\varepsilon = 0$. In this case equations (E.2)–(E.4) mean that

- (i) the origin is fixed;
- (ii) the stable manifold of the origin has coordinates $\{y = 0\}$ and the unstable manifold of the origin has coordinates $\{x = 0\}$;
- (iii) The flow restricted to the invariant manifolds is linear;
- (iv) The area form $\Omega := drd\theta$ equals to $dx dy$.

To satisfy this conditions we first choose an arbitrary coordinate system (\bar{x}, \bar{y}) satisfying (i) and (ii). This is possible since the invariant manifolds are smooth ([29, Theorem 4.1(d)]).

Next we further change coordinates $\hat{x} = \hat{x}(\bar{x})$, $\hat{y} = \hat{y}(\bar{y})$, to satisfy (iii). To fix our ideas consider the unstable manifold. The flow restricted to it has form $\dot{\hat{y}} = \bar{g}(\bar{y})$ for some function \bar{g} . We need our change to satisfy $\frac{d\hat{y}}{d\bar{y}}g(\bar{y}) = \lambda(I)\hat{y}$ that is

$$\frac{\ln \hat{y}}{\lambda(I)} = C + \int_{y^*}^{\bar{y}} \frac{ds}{\bar{g}(s)}.$$

Using the fact that

$$\frac{1}{\bar{g}(s)} = \frac{1}{\lambda(I)s} + g^*(s)$$

where g^* is a smooth function we get $\hat{y} = \bar{y}e^{G^*(\bar{y})}$ where G^* is a smooth function near the origin (in fact, G^* is a rescaled antiderivative of g^*).

Now we need one last change $(\hat{x}, \hat{y}) \rightarrow (x, y)$ to satisfy (iv). In coordinates (\hat{x}, \hat{y}) we have $\Omega = z(\hat{x}, \hat{y})d\hat{x}d\hat{y}$ for some function z . We obtain the point with coordinates (x, y) by starting from the point with coordinate $(0, y)$ (recall that our coordinates has been already defined on the unstable manifold of the fixed point) and moving for time x along the flowlines of the vectorfield (\mathbb{X}, \mathbb{Y}) . We need to satisfy the following conditions

$$\operatorname{div}_\Omega(\mathbb{X}, \mathbb{Y}) = 0 \text{ and } (\mathbb{X}, \mathbb{Y})(0, x) = (1, 0).$$

We take $\mathbb{X} = 1$ and then obtain \mathbb{Y} solving

$$\frac{\partial \mathbb{Y}}{\partial y} + \mathbb{Y} \frac{\partial z}{\partial y} + \frac{\partial z}{\partial x} = 0, \quad \mathbb{Y}(0, x) = 0.$$

Next we claim that our coordinates depend smoothly on I . First we note that the manifolds $\{x = 0\}$ and $\{y = 0\}$ are smooth. For example, $\{y = 0\}$ is normally hyperbolic (in fact transversally we have just contraction so the claim follows from the smoothness of normally hyperbolic manifolds [29, Theorem 4.1(d)]). Now the smoothness of $x(r, \theta, I, 0)$ and $y(r, \theta, I, 0)$ follows from the fact that the solutions of ODEs depend smoothly on initial conditions.

Next we have that the manifolds $\{x = 0\}$, $\{y = 0\}$ and $\{x = y = 0\}$ are normally hyperbolic and by [29, Theorem 4.1(f)] these structures survive for small non-zero ε . In fact the set of points $(r, \theta, I, \sqrt{\varepsilon})$ such that $\{x = 0\}$ is also normally hyperbolic for the equation (2.5) supplemented by $\dot{\varepsilon} = 0$, and so it is smooth. Arguing as in $\varepsilon = 0$ we obtain that the functions $(r, \theta, I, \sqrt{\varepsilon}) \rightarrow (x, y)$ are smooth (again we need to introduce an additional coordinate change in order to satisfy (E.4)). Since (x, y) are smooth we obtain (E.1) with smooth a, b and c . \square

E.2. Some consequences of volume preservation. Consider the Poincaré map between the sections $\{y = \delta\}$ and $\{x = \delta\}$ for small δ . In order to study its derivatives we decompose this map into two parts

- (1) Hit of $\{x = \delta\}$;
- (2) Landing to $\{x = \delta\}$.

Motivated by Proposition 6.2* we assume that

$$(E.5) \quad x_0 > \sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}.$$

The following estimates will be helpful in our analysis. Let Z denote the set $Z = \{x = 0 \text{ or } y = 0\}$. Denote $\pi : \mathbb{R}^3 \rightarrow Z$ denote the map

$$\pi(x, y, I) = \begin{cases} x & \text{if } x < y \\ y & \text{if } x \geq y \end{cases}.$$

Lemma E.2. *Let $t_1 < t_2$ be an intervals such that $x(t) < \delta, y(t) < \delta$ for all $t_1 < t < t_2$. Let $p(t) = \pi((x(t), y(t), z(t)))$. There exist functions $A, B : [0, \delta] \times Z \rightarrow \mathbb{R}$ such that*

- (a) $\int_{t_1}^{t_2} x(t) dt < \text{Const} x(t_2)$;
- (b) $\int_{t_1}^{t_2} y(t) dt < \text{Const} y(t_1)$;

(c) Let $z(t) = x(t)y(t)$. Then for $t^2 \ll \frac{1}{\sqrt{\varepsilon}}$

$$\left| \frac{z(t)}{z(0)} - 1 \right| \leq \text{Const} [\sqrt{\varepsilon} + z(0)] \tau$$

(d) $x(t) \sim A(x(0), \pi(t))e^{\lambda(I_0)t}x(0)$;

(e) $y(t) \sim B(x(0), \pi(t))e^{-\lambda(I_0)t}y(0)$.

Proof. We have $a(x, y, I, \varepsilon) = x\tilde{a}(x, y, I, \varepsilon)$, $b(x, y, I, \varepsilon) = x\tilde{b}(x, y, I, \varepsilon)$ and if $x(t) < \delta$, $y(t) < \delta$ then $\tilde{a} > \lambda_0/2$, $\tilde{b} < -\lambda_0/2$. It follows that for $t < t_2$

$$(E.6) \quad x(t) < x(t_2) \exp(\lambda_0(t-t_2)/2), \quad y(t) < y(t_1) \exp(-\lambda_0(t-t_1)/2).$$

This implies (a) and (b). Next

$$\dot{z} = \dot{x}y + x\dot{y} = ay + bx.$$

We have

$$a(x, y) = a(0, y) + \frac{\partial a}{\partial x}(0, y)x + \frac{1}{2} \frac{\partial^2 a}{\partial x^2}(\xi x, y)x^2 \text{ for some } \xi < 1.$$

By (E.3) and (E.4)

$$\frac{\partial a}{\partial x}(0, y) = -\frac{\partial b}{\partial y}(x, 0) = \lambda(I) + \mathcal{O}(\sqrt{\varepsilon}).$$

Therefore

$$a(x, y) = \lambda(I)x + \mathcal{O}(\sqrt{\varepsilon}x) + \mathcal{O}(x^2y).$$

Likewise

$$b(x, y) = -\lambda(I)y + \mathcal{O}(\sqrt{\varepsilon}y) + \mathcal{O}(xy^2).$$

Thus

$$\dot{z} = z \times \mathcal{O}(\sqrt{\varepsilon} + z)$$

Now (c) follows easily.

Next from the equation $\dot{x} = \tilde{a}x$ we conclude that

$$\begin{aligned} x(t) &= x(0) \exp \left(\int_0^t \tilde{a}(x(s), y(s), I(s), \varepsilon) ds \right) \\ &= x(0) \exp(\lambda(I_0)t) \exp \left(\int_0^t [\tilde{a}(x(s), y(s), I(s), \varepsilon) - \tilde{a}(0, 0, I_0, 0)] ds \right). \end{aligned}$$

To estimate the last integral we split

$$\tilde{a}(x(s), y(s), I(s), \varepsilon) - \tilde{a}(0, 0, I_0, 0) =$$

$$[\tilde{a}(x(s), y(s), I(s), \varepsilon) - \tilde{a}(x(s), y(s), I_0, 0)] + [\tilde{a}(x(s), y(s), I_0, 0) - \tilde{a}(0, 0, I_0, 0)].$$

The first term here is $\mathcal{O}(\sqrt{\varepsilon}t)$. To estimate the integral of the second term we split it into three parts. Fix $t^* > 0$. Then on the interval $[t-t^*, t]$ y is exponentially small while $x(t)$ is well approximated by the

solution of $\dot{\bar{x}} = a(\bar{x}, 0, I_0, 0)$ with the boundary condition $\bar{x}(t) = x(t)$. Denoting $p(s, t) = (\bar{x}(s), 0, I_0, 0)$ we have

$$\int_{t-t^*}^t [\tilde{a}(x(s), y(s), I_0, 0) - \tilde{a}(0, 0, I_0, 0)] ds \sim \int_{t-t^*}^t [\tilde{a}(p(s, t)) - \tilde{a}(0, 0, I_0, 0)] ds.$$

Likewise

$$\int_0^{t^*} [\tilde{a}(x(s), y(s), I_0, 0) - \tilde{a}(0, 0, I_0, 0)] ds \sim \int_0^{t^*} [\tilde{a}(p(s, 0)) - \tilde{a}(0, 0, I_0, 0)] ds.$$

Here $p(s, 0) = (\bar{x}(s), \bar{y}(s), I_0, 0)$ where \bar{x}, \bar{y} denotes the solution of

$$\dot{\bar{x}} = a(\bar{x}, \bar{y}, I_0, 0) \quad \dot{\bar{y}} = b(\bar{x}, \bar{y}, I_0, 0)$$

with initial condition $\bar{x}(0) = \delta, \bar{y}(0) = y(0)$.

Finally due to parts (a), (b) and (E.6)

$$\begin{aligned} & \left| \int_{t^*}^{t-t^*} [\tilde{a}(x(s), y(s), I_0, 0) - \tilde{a}(0, 0, I_0, 0)] ds \right| \\ & \leq \text{Const} \int_{t^*}^{t-t^*} [x(s) + y(s)] ds \leq \text{Const} [x(t-t^*) + y(t^*)] \leq \text{Const} \delta e^{-\lambda_0 t^*/2}. \end{aligned}$$

Now (d) and (e) follow easily. \square

Lemma E.3. *Suppose that $(\delta x, \delta y, \delta I)(0) = \mathcal{O}(1)$. Then*

- (a) $\|(\delta x, \delta y)\| = \mathcal{O}\left(\frac{\delta}{y(t)}\right)$.
- (b) $\|\delta I(t) - \delta I(0)\| \leq \frac{\text{Const} \sqrt{\varepsilon} \delta}{y(t)}$.
- (c) Denote

$$\Delta = b\delta x - a\delta y.$$

Then for all $t \in [0, \tau]$

$$\Delta(t) - \Delta(0) = \mathcal{O}(\sqrt{\varepsilon\tau} + x_0^2)\tau.$$

Proof. Write the variational equation as

$$\frac{d}{dt} \begin{pmatrix} \delta x \\ \delta y \\ \delta I \end{pmatrix} = \mathcal{R} \begin{pmatrix} \delta x \\ \delta y \\ \delta I \end{pmatrix}.$$

Then by the argument of Lemma E.2

$$\int_0^t \|\mathcal{R}(s)\| ds = \lambda t + \mathcal{O}(1).$$

Combining this with Lemma E.2(e) gives part (a). Plugging the estimate of part (a) into the equation for δI proves part (b).

To prove part (c) write

$$(E.7) \quad \dot{\Delta} = \left(\frac{\partial a}{\partial x} + \frac{\partial b}{\partial y} \right) \Delta + \sqrt{\varepsilon} c \left(\frac{\partial b}{\partial I} - \frac{\partial a}{\partial I} \right) \delta I + \left(\frac{\partial a}{\partial I} b - \frac{\partial b}{\partial I} a \right) \delta I.$$

The first term is $\mathcal{O}(\sqrt{\varepsilon}\Delta)$ due to (E.3), the second term is $\mathcal{O}(\sqrt{\varepsilon}\tau)$ by part (b) and (E.5). The contribution of the last term will be split into two parts according to the bound on $|\delta I(t)|$ provided by part (b).

Let $\Sigma(t)$ be the surface spanned by the trajectories starting from

$$\{0 \leq x \leq x_0, y = \delta, I = I_0\}$$

and terminating at $x = x(t)$. Let $\gamma(t) = \partial\Sigma(t)$. Consider the contribution of

$$\delta I(0) \int_0^t \left(\frac{\partial a}{\partial I} b - \frac{\partial b}{\partial I} a \right) (s) ds.$$

The second factor here equals to

$$(E.8) \quad \oint_{\gamma(t)} \left(\frac{\partial a}{\partial I} dy - \frac{\partial b}{\partial I} dx \right) + e(t)$$

where the error term $e(t)$ can be estimated as follows.

$$\begin{aligned} e(t) &= \int_0^{x_0} \frac{\partial b}{\partial I}(x, \delta) dx + \int_0^{y(t)} \frac{\partial a}{\partial I}(x(t), y) dy \\ &= \lambda(I)(x_0 y_0 - x(t)y(t)) + \mathcal{O}(\sqrt{\varepsilon} + x_0^2) = \mathcal{O}(\sqrt{\varepsilon} + x_0^2)\tau \end{aligned}$$

by Lemma E.2(c). The main term in (E.8) equals

$$\oint_{\gamma(t)} \left(\frac{\partial a}{\partial I} dy - \frac{\partial b}{\partial I} dx \right) = \iint_{\Sigma(t)} \frac{\partial}{\partial I} \left(\frac{\partial b}{\partial y} + \frac{\partial a}{\partial x} \right) dx dy = \mathcal{O}(\sqrt{\varepsilon}).$$

Next by (E.2)

$$\left(\frac{\partial a}{\partial I} b - \frac{\partial b}{\partial I} a \right) \leq \text{Const} x(t)y(t)$$

so by Lemma E.2 the contribution of

$$\int_0^t \left(\frac{\sqrt{\varepsilon}}{y(s)} \right) \left(\frac{\partial a}{\partial I} b - \frac{\partial b}{\partial I} a \right) (s) ds$$

is $\mathcal{O}(\sqrt{\varepsilon})$. Now the result follows easily. \square

E.3. Analysis of passages near the origin. We are now ready to estimate first derivatives of the map of step (1).

Lemma E.4.

$$\frac{\partial x_1}{\partial x_0} \sim \frac{\delta}{x_0}, \quad \left| \frac{\partial y_1}{\partial x_0} \right| \leq \text{Const} \delta, \quad \left| \frac{\partial I_1}{\partial x_0} \right| \leq \frac{\text{Const} \delta \sqrt{\varepsilon}}{x_0},$$

$$\frac{\partial x_1}{\partial y_0} = \mathcal{O}(1), \quad \frac{\partial y_1}{\partial y_0} = \mathcal{O}(x_0), \quad \frac{\partial I_1}{\partial y_0} = \mathcal{O}(\sqrt{\varepsilon}).$$

$$\frac{\partial x_1}{\partial I_0} = \mathcal{O}(\tau^2), \quad \frac{\partial y_1}{\partial I_0} = \mathcal{O}(\sqrt{\varepsilon} \tau + x_0 \tau), \quad \frac{\partial I_1}{\partial I_0} - 1 = \mathcal{O}(\sqrt{\varepsilon} \tau^2).$$

Proof. Substituting the identity

$$(E.9) \quad \delta x = \frac{a \delta y + \Delta(t)}{b}$$

into the equation for δy we get

$$(E.10) \quad \dot{\delta y} = \left(\frac{\frac{\partial b}{\partial x} a}{b} + \frac{\partial b}{\partial y} \right) \delta y + \tilde{\Delta}(t)$$

where

$$\tilde{\Delta}(t) = \frac{\frac{\partial b}{\partial x} \Delta(t)}{b} + \frac{\partial b}{\partial I} \delta I.$$

Let $\Xi(s, t)$ denote the fundamental solution of the homogeneous equation

$$\dot{\Xi} = \left(\frac{\frac{\partial b}{\partial x} a}{b} + \frac{\partial b}{\partial y} \right) \Xi, \quad \Xi(s, s) = 1.$$

Observe that on $\{x = y = 0\}$ the expression in parenthesis equals to $-\lambda(I_0) + \mathcal{O}(\sqrt{\varepsilon})$ (due to (E.2) and (E.3)). Therefore arguing as in the proof of Lemma E.2 we see that

$$C_1 e^{-\lambda(I_0)(t-s)} \leq \Xi(s, t) \leq C_2 e^{-\lambda(I_0)(t-s)}.$$

In particular

$$\int_0^t \Xi(s, t) ds = \mathcal{O}(1).$$

We have

$$\delta y(t) = \Xi(0, t) \delta y(0) + \int_0^t \Xi(s, t) \tilde{\Delta}(s) ds.$$

Since

$$\Delta(t) = \Delta(0) + \mathcal{O}((\sqrt{\varepsilon} \tau + x_0^2) \tau), \quad \frac{\partial b}{\partial x} = \mathcal{O}(1)$$

We have

$$\int_0^t \Xi(s, t) \Delta(s) ds = \mathcal{O}(\Delta(0) + (\sqrt{\varepsilon\tau} + x_0^2)\tau).$$

Since

$$\frac{\partial b}{\partial I}(s) \Xi(s, t) = \mathcal{O}(e^{-\lambda(I_0)t}), \quad \frac{\frac{\partial b}{\partial I}(s)}{y(s)} = \mathcal{O}(1)$$

we have

$$\begin{aligned} \int_0^t \Xi(s, t) \frac{\partial b}{\partial I}(s) \delta I_0 ds &= \mathcal{O}(x_0 \tau \delta I_0), \\ \int_0^t \Xi(s, t) \frac{\partial b}{\partial I}(s) \frac{\sqrt{\varepsilon}}{x(s)} ds &= \mathcal{O}(\sqrt{\varepsilon}). \end{aligned}$$

Therefore

$$\delta y_1 = \mathcal{O}(\Delta(0) + (\sqrt{\varepsilon\tau} + x_0^2)\tau + x_0 \tau \delta I_0).$$

The estimates for δx_1 are now obtained from (E.9). In particular

$$\begin{aligned} \frac{\partial x_1}{\partial x_0} &= \frac{\Delta(0)}{b_1} + \mathcal{O}\left(\frac{\delta^2}{x_0}\right) = \frac{b_0}{b_1} + \mathcal{O}\left(\frac{\delta^2}{x_0}\right) = \frac{y_0}{y_1} + \mathcal{O}\left(\frac{\delta^2}{x_0}\right) = \frac{x_1}{x_0} + \mathcal{O}\left(\frac{\delta^2}{x_0}\right) \\ &= \frac{\delta}{x_0} + \mathcal{O}\left(\frac{\delta^2}{x_0}\right). \end{aligned}$$

To get the estimates for δI_1 we plug the bounds for δx and δy into the equation

$$\delta \dot{I} = \sqrt{\varepsilon} \frac{\partial c}{\partial x} \delta x + \sqrt{\varepsilon} \frac{\partial c}{\partial y} \delta y + \sqrt{\varepsilon} \frac{\partial c}{\partial I} \delta I$$

and observe that the main contribution comes from the first term and that $\int_0^t |\delta x|(s) ds = \mathcal{O}(|\delta x|(t))$. This gives the required estimates for x_0 - and I_0 -derivatives. However for y_0 derivatives we get slightly weaker bounds

$$\frac{\partial y_1}{\partial y_0} = \mathcal{O}(x_0 + \sqrt{\varepsilon\tau\tau}), \quad \frac{\partial x_1}{\partial y_0} = \mathcal{O}\left(1 + \frac{\sqrt{\varepsilon\tau\tau}}{x_0}\right) = \mathcal{O}(\tau^2)$$

$$(E.11) \quad \delta I = \mathcal{O}(\sqrt{\varepsilon\tau^2}).$$

Substituting (E.11) into (E.7) we get

$$(E.12) \quad \Delta(t) = \Delta_0 + \mathcal{O}(\sqrt{\varepsilon}x_0 t^3 + \varepsilon t^3).$$

Substituting (E.11) and (E.12) into (E.10) we get

$$\frac{\partial y_1}{\partial y_0} = \mathcal{O}(x_0).$$

Now the bounds for other y_0 -derivatives follow easily. \square

Next, we estimate the second derivatives.

Lemma E.5.

$$\begin{aligned}\frac{\partial^2 x_1}{\partial x_0^2} &= \mathcal{O}\left(\frac{1}{x_0}\right), & \frac{\partial^2 y_1}{\partial x_0^2} &= \mathcal{O}\left(\frac{1}{x_0}\right), & \frac{\partial^2 I_1}{\partial x_0^2} &= \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_0^2}\right). \\ \frac{\partial^2 x_1}{\partial x_0 \partial I_0} &= \mathcal{O}\left(\frac{1}{x_0}\right), & \frac{\partial^2 y_1}{\partial x_0 \partial I_0} &= \mathcal{O}\left(\frac{1}{x_0}\right), & \frac{\partial^2 I_1}{\partial x_0 \partial I_0} &= \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_0}\right). \\ \frac{\partial^2 x_1}{\partial I_0^2} &= \mathcal{O}\left(\frac{1}{x_0}\right), & \frac{\partial^2 y_1}{\partial I_0^2} &= \mathcal{O}(\tau^2), & \frac{\partial^2 I_1}{\partial I_0^2} &= \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_0}\right).\end{aligned}$$

Proof. We will show how to estimate $\frac{\partial^2}{\partial x_0^2}$. Other derivatives are similar. Consider the second variational equation

$$\frac{d}{dt} \begin{pmatrix} \delta^2 x \\ \delta^2 y \\ \delta^2 I \end{pmatrix} = \mathcal{R} \begin{pmatrix} \delta^2 x \\ \delta^2 y \\ \delta^2 I \end{pmatrix} + \mathcal{Q}$$

where \mathcal{Q} denotes the quadratic part. Using Lemma E.4 and the fact that due to (E.4)

$$\frac{\partial^2 a}{\partial x^2} = \mathcal{O}(y), \quad \frac{\partial^2 b}{\partial x^2} = \mathcal{O}(y)$$

we get the following bounds for the components of \mathcal{Q}

$$\mathcal{Q}_x(s) = \mathcal{O}(e^{\lambda(I_0)s}), \quad \mathcal{Q}_y(s) = \mathcal{O}(e^{\lambda(I_0)s}), \quad \mathcal{Q}_I(s) = \mathcal{O}(\sqrt{\varepsilon}e^{2\lambda(I_0)s}).$$

On the other hand if $\Gamma(s, t)$ denote the fundamental solution of $\dot{\Gamma} = \mathcal{R}\Gamma$ then by the estimates of Lemma E.4 we have

$$\Gamma(s, t) = \mathcal{O} \left(\begin{pmatrix} e^{\lambda(I_0)(t-s)} & 1 & (t-s)^2 \\ 1 & e^{-\lambda(I_0)(t-s)} & e^{-\lambda(I_0)(t-s)}(t-s) + \sqrt{\varepsilon}(t-s)^{3/2} \\ \sqrt{\varepsilon}e^{\lambda(I_0)(t-s)} & \sqrt{\varepsilon} & 1 \end{pmatrix} \right).$$

This implies the bounds for $\frac{\partial^2}{\partial x_0^2}$ derivatives. \square

Lemma E.6. *The map of step (2) satisfies*

$$\begin{aligned}\frac{\partial y_2}{\partial x_1} &= -\frac{b}{a}, & \frac{\partial y_2}{\partial y_1} &= 1, & \frac{\partial y_2}{\partial I_1} &= 0, \\ \frac{\partial I_2}{\partial x_1} &= -\sqrt{\varepsilon}\frac{c}{a}, & \frac{\partial I_2}{\partial I_1} &= 1, & \frac{\partial I_2}{\partial y_1} &= 0,\end{aligned}$$

$$\frac{\partial^2 y_2}{\partial x_1^2} = \mathcal{O}(y^2 + y\sqrt{\varepsilon}), \quad \frac{\partial^2 y_2}{\partial x_1 \partial y_1} = \mathcal{O}(1), \quad \frac{\partial^2 y_2}{\partial x_1 \partial I_1} = \mathcal{O}(y), \quad \frac{\partial^2 I_2}{\partial x_1 \partial *_1} = \mathcal{O}(\sqrt{\varepsilon})$$

and the other second derivatives are zero.

Proof. Let $X(x, y, I, \tau), Y(x, y, I, \tau), \mathcal{I}(x, y, I, \tau)$ denote the solutions of (E.1) with initial conditions

$$(E.13) \quad (X, Y, \mathcal{I})(x, y, I, 0) = (x, y, I).$$

Then

$$y_2 = Y(x_1, y_1, I_1, \tau) \quad I_2 = \mathcal{I}(x_1, y_1, I_1, \tau)$$

where

$$(E.14) \quad X(x_1, y_1, I_1, \tau) = \delta.$$

Differentiating (E.14) and using (E.13) together with its x -derivative we get

$$\begin{aligned} \frac{\partial X}{\partial x} + a(\delta, Y, \mathcal{I}) \frac{\partial \tau}{\partial x} &= 0, \\ \frac{\partial a}{\partial x} \frac{\partial \tau}{\partial x} + \frac{\partial a}{\partial y} b \left(\frac{\partial \tau}{\partial x} \right)^2 + \sqrt{\varepsilon} \frac{\partial a}{\partial I} c \left(\frac{\partial \tau}{\partial x} \right)^2 + a \frac{\partial^2 \tau}{\partial x^2} &= 0. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial \tau}{\partial x} &= -\frac{1}{a}, \\ \frac{\partial^2 \tau}{\partial x^2} &= \frac{1}{a^2} \frac{\partial a}{\partial x} - \frac{1}{a^3} \left(\frac{\partial a}{\partial y} b + \sqrt{\varepsilon} \frac{\partial a}{\partial I} c \right). \end{aligned}$$

Now

$$\begin{aligned} \frac{\partial Y}{\partial x_1} &= \frac{\partial Y}{\partial x} + b(\delta, Y, \mathcal{I}, \tau) \frac{\partial \tau}{\partial x} = -\frac{b}{a}, \\ \frac{\partial^2 Y}{\partial x_1^2} &= \frac{\partial b}{\partial x} \frac{\partial \tau}{\partial x} + \frac{\partial b}{\partial y} b \left(\frac{\partial \tau}{\partial x} \right)^2 + \sqrt{\varepsilon} \frac{\partial b}{\partial I} c \left(\frac{\partial \tau}{\partial x} \right)^2 + b \frac{\partial^2 \tau}{\partial x^2} \end{aligned}$$

Using (E.2) the last expression is

$$-\frac{1}{a} \frac{\partial b}{\partial x} + \frac{b}{a^2} \left(\frac{\partial b}{\partial y} + \frac{\partial a}{\partial x} \right) + \mathcal{O}(y^2 + y\sqrt{\varepsilon})$$

By (E.3) the second term is $\mathcal{O}(y\sqrt{\varepsilon})$. To estimate the first term observe that

$$\begin{aligned} \frac{\partial b}{\partial x}(x_2, y_2, I_2) &= \frac{\partial b}{\partial x}(x_2, 0, I_2) + \frac{\partial^2 b}{\partial x \partial y}(x_2, 0, I_2) y_1 + \mathcal{O}(y^2) \\ &= \frac{\partial b}{\partial x}(x_2, 0, I_2) - \frac{\partial^2 a}{\partial^2 x}(x_2, 0, I_2) y_1 + \mathcal{O}(y^2). \end{aligned}$$

The first term here vanishes due to (E.2) while the second term is $\mathcal{O}(y_1^2 + \sqrt{\varepsilon} y_1)$ due to (E.4).

The estimates for other derivatives are similar but easier because there is no need to use (E.3) and (E.4). \square

Lemma E.7. *We have*

$$\frac{\partial y_2}{\partial x_0} = \frac{|b(x_0, y_0, I_0, 0)|}{a(x_1, y_1, I_1, 0)} + \mathcal{O}((\sqrt{\varepsilon\tau} + x_0^2)\tau), \quad \frac{\partial y_2}{\partial I_0} = \mathcal{O}((\sqrt{\varepsilon\tau} + x_0^2)\tau),$$

$$\frac{\partial I_2}{\partial x_0} \sim -\frac{\sqrt{\varepsilon}c}{\frac{\partial a}{\partial x_0}}, \quad \frac{\partial I_2}{\partial I_0} = 1 + \mathcal{O}(\sqrt{\varepsilon}\tau^2),$$

$$\frac{\partial^2 y_2}{\partial x_0^2} = \mathcal{O}\left(\frac{1}{x_0}\right), \quad \frac{\partial^2 y_2}{\partial x_0 \partial I_0} = \mathcal{O}\left(\frac{1}{x_0}\right), \quad \frac{\partial^2 y_2}{\partial I_0^2} = \mathcal{O}(\tau^2),$$

$$\frac{\partial^2 I_2}{\partial x_0^2} = \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_0^2}\right), \quad \frac{\partial^2 I_2}{\partial x_0 \partial I_0} = \mathcal{O}\left(\frac{\sqrt{\varepsilon}\tau^2}{x_0}\right), \quad \frac{\partial^2 I_2}{\partial I_0^2} = \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_0}\right).$$

Proof. The inequalities are obtained by direct computation. To get the asymptotics of $\frac{\partial y_2}{\partial x_0}$ observe that

$$\begin{aligned} \frac{\partial y_2}{\partial x_0} &= \frac{\partial y_2}{\partial x_1} \frac{\partial x_1}{\partial x_0} + \frac{\partial y_2}{\partial y_1} \frac{\partial y_1}{\partial x_0} + \frac{\partial y_2}{\partial I_1} \frac{\partial I_1}{\partial x_0} = \frac{\partial y_1}{\partial x_0} - \left(\frac{b}{a}\right)(x_1, y_1, I_1, \varepsilon) \frac{\partial x_1}{\partial x_0} \\ &= -\frac{\Delta_1}{a(x_1, y_1, I_1, \varepsilon)} = \frac{|b(x_0, y_0, I_0, \varepsilon)|}{a(x_1, y_1, I_1, \varepsilon)} + \mathcal{O}((\sqrt{\varepsilon\tau} + x_0^2)\tau) \end{aligned}$$

where the last equality follows from Lemma E.3.

To get the asymptotics of $\frac{\partial I_2}{\partial x_0}$ observe that

$$\frac{\partial I_2}{\partial x_0} = \frac{\partial I_2}{\partial x_1} \frac{\partial x_1}{\partial x_0} + \frac{\partial I_2}{\partial y_1} \frac{\partial y_1}{\partial x_0} + \frac{\partial I_2}{\partial I_1} \frac{\partial I_1}{\partial x_0} = \frac{\partial I_2}{\partial x_1} \frac{\partial x_1}{\partial x_0} + \frac{\partial I_2}{\partial I_1} \frac{\partial I_1}{\partial x_0} = \frac{\partial I_2}{\partial x_1} \frac{\partial x_1}{\partial x_0} + \mathcal{O}\left(\frac{\sqrt{\varepsilon}\delta}{x_0}\right).$$

Now by Lemmas E.4 and E.6

$$\frac{\partial I_2}{\partial x_1} \frac{\partial x_1}{\partial x_0} \sim -\sqrt{\varepsilon} \frac{c}{a} \frac{\delta}{x_0}.$$

Since $a \sim \delta \frac{\partial a}{\partial x}$ the result follows. \square

The above formulas describe the transition between (I, x) and (I, y) coordinates. We now return to (I, E) coordinates. Due to the smoothness of (x, y, I) near the separatrix we have

$$(E.15) \quad H = \sqrt{\varepsilon}\kappa(I) + q(I)xy + HOT$$

Thus on $\{y = \delta\}$ the following bound hold

$$\frac{\partial H}{\partial x} = q(I)y + HOT = \delta q(I) + HOT \quad \text{and} \quad \frac{\partial H}{\partial I} \sim \frac{\partial q}{\partial I} x \delta + HOT.$$

It follows that the passage near the separatrix has the following derivatives in (I, H) coordinates

$$\begin{pmatrix} 1 & 0 \\ \mathcal{O}(\delta x_0) & \delta q(I) \end{pmatrix} \begin{pmatrix} 1 & -\frac{c\sqrt{\varepsilon}}{\frac{\partial a}{\partial x} x_0} \\ \mathcal{O}((\sqrt{\varepsilon\tau} + x_0^2)\tau) & \frac{b}{a} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \mathcal{O}(x_0) & \frac{1}{\delta q(I)} \end{pmatrix} + HOT$$

$$(E.16) \quad = \begin{pmatrix} 1 & -\frac{c\sqrt{\varepsilon}}{\frac{\partial_a}{\partial x}x_0\delta q(I)} \\ \mathcal{O}(x_0\delta) & \frac{b}{a} \end{pmatrix}.$$

Observe that if $H_0 \gg \sqrt{\varepsilon}$ then by (E.15) the term in the upper corner equals to

$$-\sqrt{\varepsilon} \frac{c}{\lambda(I)H_0} (1 + o(1)).$$

In our setting this can be rewritten as

$$\frac{\partial I_2}{\partial H_0} \sim \frac{-\sqrt{\varepsilon}\alpha_1(I, \phi(I), \theta_{cr}(I))}{\sqrt{\left|\frac{\partial^2 U}{\partial \theta^2}(\theta_{cr}(I), I)\right|}}.$$

Concerning the second derivatives Lemma E.7, (E.16) and Lemma C.1 imply that those bounds can have at most $\frac{1}{d^2}$ extra factor comparing with maps in \mathcal{T} .

E.4. Derivative bounds of Proposition 6.2. Here we prove Proposition 6.2(a) and (b). For simplicity we consider orbits which pass only once near the saddle. On figure 3 those orbits pass on the right of the saddle point. There are also orbits passing twice near the saddle. On figure 3 those orbits pass below and then above the saddle point. The analysis of orbits experiencing two passages is similar to one passage case but requires a slightly longer computations. Namely we need to consider the composition of three maps: first passage of the δ -neighborhood of the saddle, motion along the separatrix loop and the second passage of δ -neighborhood of the saddles. Since such compositions are studied in Appendix F we leave the proof of the two passage case to the reader who may refer to section F.2 and in particular Lemma F.3 for details.

The proof of parts (a) and (b) of Proposition 6.2 for is the same as for Proposition 6.1 except that now for orbits passing (once) near the separatrix we split step (4) into three substeps:

- (a) Landing to $y = \delta$;
- (b) Landing to $x = \delta$
- (c) Hit of $r = R$.

Now the jump of I inside the δ -neighborhood of the saddle can be computed using (E.1). Namely if t_{\mp} are the beginning and the end of the passage then

$$\Delta I_{saddle} = \sqrt{\varepsilon} \int_{t_-}^{t_+} c(I(s), \phi(s), \theta(s)) ds.$$

Now near the saddle we have $c(I(s), \phi(s), \theta(s)) = c(I, \phi_j(I), \theta_{jk}(I) + o(1))$, so

$$\Delta I_{saddle} = \sqrt{\varepsilon} c(I, \phi_j(I), \theta_{jk}(I)) (t_+ - t_-) (1 + o(1)).$$

Now the jump outside the δ -neighborhood is at most $\text{Const} \sqrt{\varepsilon \bar{t}}$ where \bar{t} is the largest time spent on step (4) by an orbit avoiding δ neighborhood of the saddles. Now if δ is fixed and $d \rightarrow 0$ we get $\bar{t} \leq \text{Const}(\delta)$ whereas $t_+ - t_- \rightarrow \infty$. Letting $\delta \rightarrow 0$ sufficiently slowly we obtain equation (6.3) Proposition 6.2(a).

To get the bounds for derivatives observe that the maps of steps (4a) and (4c) are in \mathcal{T}_0 (see Lemmas C.1, C.7 and E.6). Therefore if the map of step (4b) were in \mathcal{T} the estimates of the Proposition 6.1 would remain valid. However because of the step (4b) the estimates are actually worse. Namely for the first derivative we loose a factor of $O(\frac{1}{d})$ and for the second derivative we loose a factor of $O(\frac{1}{d^2})$. Now to obtain the first derivative estimates we need to multiply the estimates of steps (1)–(11). Since bounds for all factors stay as before except for extra $(\frac{1}{d})$ factor at step (4) we loose at most $(\frac{1}{d})$. Similarly then computing the second derivative step (4) contributes either the second derivative or the first derivative squared. In the first case we loose at most $\frac{\ln^m d}{d^2}$ for some number m , in the second case we loose at most $\frac{1}{d^2}$. Finally to get (6.4) we argue as in the proof of Proposition B.1 examining each monomial of $\frac{\partial J_7}{\partial H_0}$. Again there is only one monomial which is better than $o(\sqrt{\varepsilon}/d)$. Thus

$$\frac{\partial J_7}{\partial H_0} \sim \frac{\partial J_7}{\partial I_6} \frac{\partial I_6}{\partial I_5} \frac{\partial I_5}{\partial I_{4c}} \frac{\partial I_{4c}}{\partial I_{4b}} \frac{\partial I_{4b}}{\partial H_{4a}} \frac{\partial H_{4a}}{\partial H_3} \frac{\partial H_3}{\partial H_2} \frac{\partial H_2}{\partial H_1} \frac{\partial H_1}{\partial H_0} \sim -\sqrt{\varepsilon} \frac{c}{\lambda d} (1 + o(1)).$$

Combining this with

$$\frac{\partial H_{11}}{\partial H_0} \sim \frac{\partial H_{11}}{\partial J_7} \frac{\partial J_7}{\partial H_0} \sim \frac{1}{\varepsilon} \frac{\partial w}{\partial I} \frac{\partial J_7}{\partial H_0}$$

we obtain (6.4).

E.5. Measure bounds of Proposition 6.2. Here we prove Proposition 6.2(c) and (d).

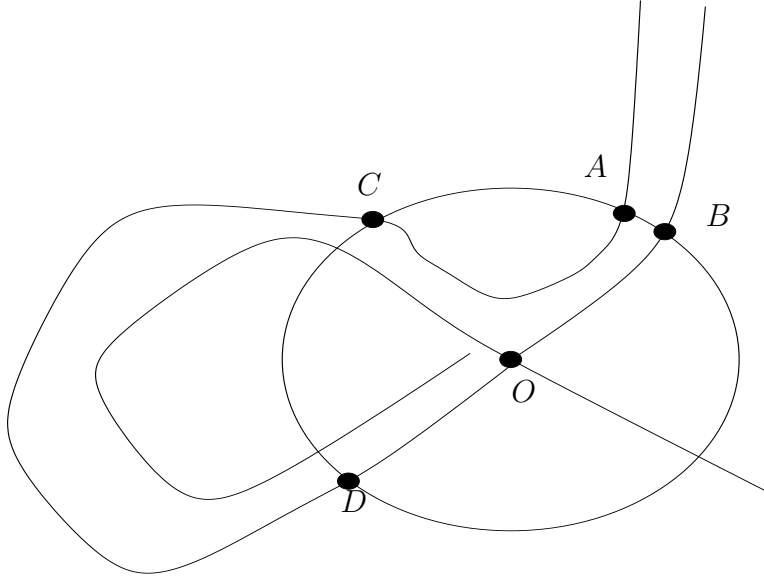
To get (c) observe that the maps of steps (1)–(3) are in \mathcal{T}_0 so their compositions are in \mathcal{T}_0 . Thus using the notation of the previous section we have

$$\frac{\partial H_{4a}}{\partial H_0} = \frac{\partial H_{4a}}{\partial H_3} \frac{\partial H_3}{\partial H_0} + \frac{\partial H_{4a}}{\partial I_3} \frac{\partial I_3}{\partial H_0} = 1 + o(1) + \mathcal{O}(\sqrt{\varepsilon}) o(1/\sqrt{\varepsilon}) = 1 + o(1).$$

Next, since the composition of the maps of steps (1)-(4a) is in \mathcal{T} we have that the image of γ satisfies

$$(E.17) \quad \frac{dI_{4a}}{dH_{4a}} = \frac{\frac{\partial I_{4a}}{\partial H_0} + \frac{\partial I_{4a}}{\partial I_0} g'(H_0)}{\frac{\partial E_{4a}}{\partial H_0} + \frac{\partial H_{4a}}{\partial I_0} g'(H_0)} = \frac{\mathcal{O}(\sqrt{\varepsilon})}{1 + o(1) + \mathcal{O}(\frac{1}{\sqrt{\varepsilon}})\mathcal{O}(\varepsilon^{1/2+\delta})} = \mathcal{O}(\sqrt{\varepsilon}).$$

So this image is transversal to the line $H = \text{Const}$. Thus the set $\{d \leq \xi\}$ has measure comparable to the measure of the set $\{H_{4a} \leq \xi\}$. This proves (c).



Using (2.11) we see that (d) reduces to

$$H_+^{(jk)} - H_-^{(jk)} = \sqrt{\varepsilon} M_{jk}(I) + o(\sqrt{\varepsilon}).$$

Let A and B be the images of $H_-^{(jk)}$ and $H_+^{(jk)}$ respectively under the map of steps (1)-(4a). By the foregoing discussion

$$\frac{H_+^{(jk)} - H_-^{(jk)}}{H(B) - H(A)} = 1 + o(1)$$

so we need to estimate the denominator.

Let CD be the component of the orbit of A outside the δ neighborhood of the saddle. Fig. 7 projects everything into $I = \text{Const}$ plane ignoring the fact that the orbits of B and D hit \mathcal{N}_ε at different points O_B and O_D . Now using the smooth dependence of stable and unstable manifolds on parameters we get

$$(E.18) \quad |H(B) - H(O_B)| \leq \text{Const} \delta \sqrt{\varepsilon},$$

$$(E.19) \quad |I(B) - I(O_B)| \leq \text{Const} \delta \sqrt{\varepsilon},$$

$$(E.20) \quad |H(D) - H(O_D)| \leq \text{Const} \delta \sqrt{\varepsilon},$$

$$(E.21) \quad |I(D) - I(O_D)| \leq \text{Const} \delta \sqrt{\varepsilon},$$

Next

$$H(D) - H(C) = \sqrt{\varepsilon} \int \left[r^2 \bar{\beta}(I, 0, \theta, 0) - (L(I) + g(I, \theta)) \bar{\eta}(I, 0, \theta, 0) + \frac{\partial \hat{H}}{\partial I} \bar{\alpha}(I, 0, \theta, 0) \right] dt$$

where the integral is taken along the orbit from C to D . Thus

$$(E.22) \quad H(D) - H(C) = \sqrt{\varepsilon} [M_{jk}^\delta(I) + \Delta_{\delta, \varepsilon}]$$

where $M_{jk}^\delta(I)$ denotes the integral (2.16) over the part of Γ which lies outside δ -neighborhood of O and $\Delta_{\delta, \varepsilon} \rightarrow 0$ as $\varepsilon \rightarrow 0$. Also since it takes time $\mathcal{O}(|\ln \delta|)$ to go from C to D we have

$$I(D) - I(C) = \mathcal{O}(|\ln \delta| \sqrt{\varepsilon}).$$

Thus by (E.21)

$$(E.23) \quad I(O_D) - I(C) = \mathcal{O}(|\ln \delta| \sqrt{\varepsilon}).$$

Let O_A and O_C be the points on \mathcal{N}_ε having the same I coordinate as A and C respectively. Using that \mathcal{N}_ε and its derivatives depend smoothly on $\sqrt{\varepsilon}$ and remembering that H is zero on \mathcal{N} by our choice of $K_{jk}(I)$ we get

$$(E.24) \quad |H(O_C) - H(O_D)| \leq \text{Const} \sqrt{\varepsilon} |I(O_C) - I(O_D)| \leq \text{Const} \varepsilon |\ln \delta|.$$

Next we claim that

$$(E.25) \quad H(A) - H(O_A) = H(C) - H(O_C) + \mathcal{O}(\delta \sqrt{\varepsilon}).$$

Indeed if both $H(A) - H(O_A)$ and $H(C) - H(O_C)$ are less than $\delta \sqrt{\varepsilon}$ then there is nothing to prove. Otherwise the result follows from Lemma E.2(c) (applied to either system (E.1) or its time reversal).

Combining (E.18), (E.22), (E.24) and (E.25) we get

$$(E.26) \quad H(B) - H(A) - [H(O_B) - H(O_A)] = \sqrt{\varepsilon} [M_{jk}^\delta(I) + \Delta_{\delta, \varepsilon}] + \mathcal{O}(\delta \sqrt{\varepsilon}).$$

On the other hand we have

$$\begin{aligned} |I(O_A) - I(O_B)| &= |I(A) - I(O_B)| \leq |I(A) - I(B)| + |I(B) - I(O_B)| \\ &\leq \text{Const} [\sqrt{\varepsilon} |H(A) - H(B)| + \delta \sqrt{\varepsilon}] \end{aligned}$$

where the last inequality uses (E.17) and (E.19). Thus

$$H(O_A) - H(O_B) \leq \text{Const} [\sqrt{\varepsilon} |H(A) - H(B)| + \delta \sqrt{\varepsilon}].$$

Combining this with (E.26) we get

$$H(B) - H(A) = \sqrt{\varepsilon} [M_{jk}^\delta(I) + \Delta_{\delta,\varepsilon}] + \mathcal{O}(\delta\sqrt{\varepsilon}).$$

Letting ε and δ to 0 at appropriate speed we obtain the statement required.

APPENDIX F. CAPTURED POINTS.

F.1. Dividing the trajectory. Our task is to establish (3.8)–(3.9) rigorously. To this end we divide the captured trajectory into three parts: entrance part, middle part and exit part. The middle part will be defined by the condition that $|H| \geq \varepsilon^{1/4-\delta}$. For the middle part the standard averaging theory of Appendix D can be applied. On the other hand, for the entrance and the exit parts the orbit passes near the saddle point several times and for each passage the results of Appendix E can be used.

Since our goal is to prove Proposition 6.2* we only consider the orbits which do not come closer than $\sqrt{\frac{\varepsilon}{|\ln \varepsilon|}}$ to the saddle point.

Lemma F.1. (a) *The entrance map satisfies the following estimates.*

$$|H_1| = \varepsilon^{1/4-\delta} + \mathcal{O}(\sqrt{\varepsilon}), \quad I_1 - I_0 \sim \frac{\varepsilon^{1/4-\delta} c(0, 0, I_0) (\frac{1}{4} - \delta)}{M_{ij}(I_0) \lambda(I_0)} |\ln \varepsilon|.$$

$$\frac{\partial H_1}{\partial H_0} \sim 1, \quad \frac{\partial I_1}{\partial I_0} \sim 1, \quad \frac{\partial H_1}{\partial I_0} = o(\varepsilon^\delta), \quad \frac{\partial I_1}{\partial H_0} \sim \frac{c(0, 0, I_0)(1/4 + \delta) |\ln \varepsilon|}{M_{ij}(I_0) \lambda(I_0)}.$$

$$\frac{\partial^2(H_1, I_1)}{\partial(H_0, I_0)^2} = \mathcal{O}(\varepsilon^{-(3/4+2\delta)}).$$

(b) *The middle map satisfies the following.*

$$I_2 = \mathbf{s}(I_1) + o(1), \quad |H_2| = \varepsilon^{1/4-\delta} + \mathcal{O}(\sqrt{\varepsilon}).$$

The first derivatives of (H_2, I_2) with respect to (H_1, I_1) are given by (3.8)–(3.9) (with (H_0, I_0) replaced by (H_1, I_1) and (H_f, I_f) by (H_2, I_2)).

The second derivatives are $\mathcal{O}(\varepsilon^{-1/4})$.

(c) *The exit map satisfies the following estimates.*

$$|H_3| = \mathcal{O}(\sqrt{\varepsilon}), \quad I_3 - I_2 \sim \frac{\varepsilon^{1/4-\delta} c(0, 0, I_f) (\frac{1}{4} - \delta)}{M_{ij}(I_f) \lambda(I_f)} |\ln \varepsilon|.$$

$$\frac{\partial H_3}{\partial H_2} \sim 1, \quad \frac{\partial I_3}{\partial I_2} \sim 1, \quad \frac{\partial H_3}{\partial I_2} = o(\varepsilon^\delta), \quad \frac{\partial I_3}{\partial H_2} \sim \frac{c(0, 0, I_f)(1/4 + \delta) |\ln \varepsilon|}{M_{ij}(I_f) \lambda(I_f)}.$$

$$\frac{\partial^2(H_3, I_3)}{\partial(H_2, I_2)^2} = \mathcal{O}(\varepsilon^{-(3/4+2\delta)}).$$

The next result is obtained from Lemma F.1 by a direct computation.

Corollary F.2. *The derivatives of the map $(I_0, H_0) \rightarrow (I_3, H_3)$ satisfy (3.8)-(3.9) with (H_f, I_f) replaced by (H_3, I_3) and $\ln H_0, \ln H_f$ replaced by $\frac{\ln \varepsilon}{2}$.*

The second derivatives are $\mathcal{O}(\varepsilon^{-(3/4+3\delta)})$.

Corollary F.2 implies the following modification of Proposition B.1.

Proposition B.1*. *The inner map for the captured orbits satisfies the following estimates.*

$$\frac{\partial J_7}{\partial H_0} \sim \frac{-C^* \mathbf{c}(J_0) \mathbf{c}(J_f)}{4M(J_f)} \ln^2 \varepsilon,$$

where C^* is defined by (3.7) and $c(J)$ is defined by equation (3.6).

$$\frac{\partial J_7}{\partial J_0} = \mathcal{O}\left(\frac{\ln^2 \varepsilon}{\sqrt{\varepsilon}}\right),$$

$$\frac{\partial H_7}{\partial H_0} = \mathcal{O}\left(\frac{\ln^2 \varepsilon}{\sqrt{\varepsilon}}\right), \quad \frac{\partial H_7}{\partial J_0} = \mathcal{O}\left(\frac{\ln^2 \varepsilon}{\varepsilon}\right)$$

The second derivatives bounds are worse than the bounds of Proposition B.1 by a factor of $\varepsilon^{-(5/4+3\delta)}$.

Proof. To get the information about the first derivatives we directly multiply the bounds

$$(F.1) \quad \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\sqrt{\varepsilon}) \\ \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right) & \mathcal{O}(1) \end{pmatrix} \begin{pmatrix} \mathcal{O}(|\ln \varepsilon|) & \mathcal{O}(\ln^2 \varepsilon) \\ \mathcal{O}(1) & \mathcal{O}(|\ln \varepsilon|) \end{pmatrix} \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\sqrt{\varepsilon}) \\ \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right) & \mathcal{O}(1) \end{pmatrix}$$

(the middle term is given by Corollary F.2 while two other terms come from Appendix C). To get the asymptotics of $\frac{\partial J_7}{\partial H_0}$ we observe that the bound for top right corner of the product (F.1) comes from products of the top left corner of the first matrix, the top right corner of the second matrix and the bottom right corner of the third matrix. Therefore the result follows from (3.8) and the fact that the maps of steps 1–3 and 5–7 of Appendix C as well as the maps of steps (4a), (4c) of Section E.4 are in \mathcal{T}_0 .

To obtain the bounds for the second derivatives we observe that comparing to the proof of Proposition B.1 only the terms of step 4 are different. Now the terms containing $\frac{\partial^2(I_4, H_4)}{\partial^2(I_3, H_3)}$ get worse by a factor $\mathcal{O}(\varepsilon^{-(5/4+3\delta)})$ and the terms containing $\left(\frac{\partial(I_4, H_4)}{\partial(I_3, H_3)}\right)^2$ get worse by a factor of $\mathcal{O}(\varepsilon^{-1} \ln^4 \varepsilon)$. \square

F.2. Entrance phase. We consider the iterates of the first return map to $\{y = \delta\}$. The first return map satisfies the following estimates.

Lemma F.3. *The first return map satisfies the following estimates.*

$$(a) \quad |H_{n+1}| - |H_n| \sim \sqrt{\varepsilon} M_{ij}(I), \quad I_{n+1} - I_n \sim \sqrt{\varepsilon} \alpha_1(I_n, \phi(I_n), \theta_{cr}(I_n)) \frac{|\ln x_n|}{\lambda(I_n)}.$$

$$(b) \quad \frac{\partial x_{n+1}}{\partial x_n} = \frac{b_{n+1}}{b_n} + \mathcal{O}((\sqrt{\varepsilon} + x_n^2) \tau), \quad \frac{\partial x_{n+1}}{\partial I_n} = \mathcal{O}((\sqrt{\varepsilon} + x_n^2) \tau),$$

$$\frac{\partial I_{n+1}}{\partial I_n} = 1 + \mathcal{O}(\sqrt{\varepsilon} \tau^2), \quad \frac{\partial I_{n+1}}{\partial x_n} \sim -\frac{\sqrt{\varepsilon} \alpha_1(I, \phi(I), \theta_{cr}(I))}{\lambda(I) x_n}$$

where b_n denotes $b(x_n, \delta, I_n, 0)$.

$$(c) \quad \frac{\partial^2 x_{n+1}}{\partial x_n^2} = \mathcal{O}\left(\frac{1}{x_n}\right), \quad \frac{\partial^2 x_{n+1}}{\partial x_n \partial I_n} = \mathcal{O}\left(\frac{1}{x_n}\right), \quad \frac{\partial^2 x_{n+1}}{\partial I_n^2} = \mathcal{O}(\tau^2),$$

$$\frac{\partial^2 I_{n+1}}{\partial x_n^2} = \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_n^2}\right), \quad \frac{\partial^2 I_{n+1}}{\partial x_n \partial I_n} = \mathcal{O}\left(\frac{\sqrt{\varepsilon} \tau^2}{x_n}\right), \quad \frac{\partial^2 I_{n+1}}{\partial I_n^2} = \mathcal{O}\left(\frac{\sqrt{\varepsilon}}{x_n}\right).$$

Proof. (a) The formula for the change of H is proven similarly to Section E.5. To establish the formula for the change of I observe that

$$I(t_2) - I(t_1) = \mathcal{O}(\sqrt{\varepsilon} |t_2 - t_1|).$$

The orbit spends most of the time near $(0, 0, I_n)$ where $\dot{I} \sim \sqrt{\varepsilon} c(0, 0, I_n)$. Also by Lemma E.2(d) the passage time satisfies $\tau \sim \frac{|\ln x_n|}{\lambda(I_n)}$.

(b),(c) We represent our map as a composition of two maps: landing to $\{x = \delta\}$ and landing to $\{y = \delta\}$. The first map was analyzed in Section E.1. The second map can be treated using the standard perturbation theory. Thus its derivative is $\mathcal{O}(\sqrt{\varepsilon})$ perturbation of $\varepsilon = 0$ map which is

$$\begin{pmatrix} \frac{a}{b} & 0 \\ 0 & 1 \end{pmatrix}.$$

(The derivation of the first column is similar to but much easier than the results of Section E.1.) Concerning the second derivatives, the derivatives of x are $\mathcal{O}(1)$ and the derivatives of I are $\mathcal{O}(\sqrt{\varepsilon})$. Now the result follows easily. \square

Lemma F.4. *Let (\tilde{x}, \tilde{I}) be the last point in the entrance zone. Then*

$$(a) \quad |\tilde{H}| = \varepsilon^{1/4-\delta} + \mathcal{O}(\sqrt{\varepsilon}), \quad \tilde{I} - I_0 \sim \frac{\varepsilon^{1/4-\delta} \alpha_1(I_0, \phi(I_0), \theta_{cr}(I_0)) \left(\frac{1}{4} - \delta\right)}{M_{ij}(I_0) \lambda(I_0)} |\ln \varepsilon|.$$

$$(b) \quad \frac{\partial \tilde{x}}{\partial x_0} \sim 1, \quad \frac{\partial \tilde{I}}{\partial I_0} \sim 1, \quad \frac{\partial \tilde{x}}{\partial I_0} = o(1),$$

$$\frac{\partial \tilde{I}}{\partial x_0} \sim -\frac{\bar{\alpha}(I_0, \phi(I_0), \theta_{cr}(I_0)) \delta(1/4 + \delta) q(I_0) \ln \varepsilon}{\lambda(I_0) M_{ij}(I_0)}.$$

(c) The second derivatives are $\mathcal{O}(\varepsilon^{-(3/4+2\delta)})$.

Proof. Part (a) immediately follows from Lemma F.3.

To establish part (b) we first show by induction that if K_1, K_2, K_3 and K_4 are sufficiently large then

$$\left| \frac{\partial x_n}{\partial x_0} - \frac{b_n}{b_0} \right| \leq K_1 [\sqrt{\varepsilon} n + \varepsilon n^3] (\ln n + |\ln \varepsilon|)^2, \quad \left| \frac{\partial x_n}{\partial I_0} \right| \leq K_2 [\sqrt{\varepsilon} n + \varepsilon n^3] (\ln n + |\ln \varepsilon|)^2,$$

$$\left| \frac{\partial I_n}{\partial I_0} - 1 \right| \leq K_3 [\sqrt{\varepsilon} n + \varepsilon n^3] (\ln n + |\ln \varepsilon|)^2, \quad \left| \frac{\partial I_n}{\partial x_0} \right| \leq K_4 \ln n.$$

This readily gives the estimates of part (b) except that for the asymptotics of $\frac{\partial \tilde{I}}{\partial x_0}$. However the above bounds imply that for $n > 0$

$$\frac{\partial I_{n+1}}{\partial x_0} - \frac{\partial I_n}{\partial x_0} \sim -\frac{\alpha_1(I_0, \phi(I_0), \theta_{cr}(I_0)) \sqrt{\varepsilon}}{\lambda(I_0) x_n}.$$

By (E.15) and Lemma F.3(a)

$$x_n \sim -\frac{n M_{ij}(I_0)}{\delta q(I_0)} \sqrt{\varepsilon}.$$

Since the number of steps is $\mathcal{O}(\varepsilon^{-(1/4+\delta)})$ we get

$$(F.2) \quad \frac{\partial \tilde{I}}{\partial x_0} \sim \frac{\alpha_1(I_0, \phi(I_0), \theta_{cr}(I_0)) \delta(1/4 + \delta) q(I_0)}{M_{ij}(I_0) \lambda(I_0)} |\ln \varepsilon|.$$

Now part (b) follows easily.

Moreover a similar argument shows that for any n

$$\left\| \frac{\partial(\tilde{x}, \tilde{I})}{\partial(x_n, I_n)} \right\| \leq K |\ln \varepsilon|.$$

Now part (c) follows from Lemma F.3(c). \square

Combining Lemma F.4 with (E.15) proves part (a) of Lemma F.1. Part (c) of Lemma F.1 follows from part (a) by time reversal.

Observe that (F.2) can be rewritten as

$$\frac{\partial \tilde{I}}{\partial H_0} \sim \frac{\mathbf{c}(I_0)(1/4 + \delta)}{M_{ij}(I_0)} |\ln \varepsilon|$$

where \mathbf{c} is defined by (3.6). This matches the asymptotics predicted by (3.8).

F.3. Middle phase. We need to study equations of the form

$$\dot{\theta} = r + \sqrt{\varepsilon}P, \quad \dot{r} = -\frac{\partial U}{\partial \theta} + \sqrt{\varepsilon}Q, \quad \dot{I} = \sqrt{\varepsilon}R.$$

It will be convenient to change the time to ensure that the orbits do not hang near the saddle for a long time. Accordingly introduce a new time variable s by

$$(F.3) \quad ds = \sqrt{r^2 + \left(\frac{\partial U}{\partial \theta}\right)^2} dt.$$

Next we rewrite our system in action-angle coordinates. Namely define ψ by the equation

$$\frac{\partial \psi}{\partial r} = -\frac{\sqrt{r^2 + \left(\frac{\partial U}{\partial \theta}\right)^2}}{T(H) \left(\frac{\partial U}{\partial \theta} + \sqrt{\varepsilon}Q\right)}$$

where $T(H)$ is the normalization factor

$$T(H) = -\oint \frac{\sqrt{r^2 + \left(\frac{\partial U}{\partial \theta}\right)^2}}{\left(\frac{\partial U}{\partial \theta} - \sqrt{\varepsilon}Q\right)} dr$$

and the integration is over the energy level.

This leads to the system

$$\psi' = 1, \quad H' = \sqrt{\varepsilon}\tilde{X}(H, I, \psi), \quad I' = \sqrt{\varepsilon}\tilde{Y}(H, I, \psi).$$

Observe that \tilde{X} and \tilde{Y} are nonsingular away from the set $\tilde{H} = 0$. Indeed the only other possible singularity set is the set \mathbf{S} of elliptic rest points. However by assumption (H) the entrance-exit map is defined for all $I \in \mathbf{G}$. Also by assumptions (J)–(K) the only way the solution of the inner averaged system (3.3) can accumulate on \mathbf{S} is if it approaches a fixed point. But by assumption (H) no orbit starting from \mathbf{G} converges to a fixed point on \mathbf{S} . So all the solutions are uniformly bounded away from \mathbf{S} . Next we get the asymptotics of \tilde{X} and \tilde{Y} near the singularities.

Lemma F.5.

$$\begin{aligned} \tilde{X} &= \mathcal{O}(1), \quad \tilde{Y} = \mathcal{O}\left(\frac{1}{\sqrt{|H|}}\right), \\ \frac{\partial^{p+q+l}\tilde{X}}{\partial H^p \partial I^q \partial \psi^l} &= \mathcal{O}\left(|H|^{-(p+\frac{q+l}{2})} |\ln^q |H||\right), \\ \frac{\partial^{p+q+l}\tilde{Y}}{\partial H^p \partial I^q \partial \psi^l} &= \mathcal{O}\left(|H|^{-(p+\frac{q+l+1}{2})} |\ln^q |H||\right), \end{aligned}$$

$$\oint \left| \frac{\partial^{p+q+l} \tilde{X}}{\partial H^p \partial I^q \partial \psi^l} \right| ds = \mathcal{O} \left(|H|^{-(p+\frac{q+l-1}{2})} |\ln^q |H|| \right),$$

$$\oint \left| \frac{\partial^{p+q+l} \tilde{Y}}{\partial H^p \partial I^q \partial \psi^l} \right| ds = \mathcal{O} \left(|H|^{-(p+\frac{q+l}{2})} |\ln^q |H|| \right).$$

where the integration is over the energy level.

The proof of Lemma F.5 is given in Section F.4.

Lemma F.6. *The first return map satisfies the following.*

$$(a) \quad H_{n+1} - H_n \sim \sqrt{\varepsilon} X, \quad I_{n+1} - I_n \sim \sqrt{\varepsilon} Y.$$

$$(b) \quad \frac{\partial H_{n+1}}{\partial H_n} - 1 \sim \sqrt{\varepsilon} \frac{\partial X}{\partial H}, \quad \frac{\partial H_{n+1}}{\partial I_n} \sim \sqrt{\varepsilon} \frac{\partial X}{\partial I},$$

$$\frac{\partial I_{n+1}}{\partial H_n} - 1 \sim \sqrt{\varepsilon} \frac{\partial Y}{\partial H}, \quad \frac{\partial I_{n+1}}{\partial I_n} \sim \sqrt{\varepsilon} \frac{\partial Y}{\partial I}$$

where X and Y are inner averaged vector fields (see (3.3)).

Proof. Both part (a) and part (b) are proven as in Appendix D. We sketch part (b), part (a) is easier.

We introduce improved variables

$$\xi = \delta H - \sqrt{\varepsilon} Z_1 \delta H - \sqrt{\varepsilon} Z_2 \delta I, \quad \eta = \delta H - \sqrt{\varepsilon} Z_3 \delta H - \sqrt{\varepsilon} Z_4 \delta I$$

where

$$\frac{\partial Z_1}{\partial \psi} = \left(\frac{\partial \tilde{X}}{\partial H} - \frac{\partial X}{\partial H} \right), \quad \frac{\partial Z_2}{\partial \psi} = \left(\frac{\partial \tilde{X}}{\partial I} - \frac{\partial X}{\partial I} \right),$$

$$\frac{\partial Z_3}{\partial \psi} = \left(\frac{\partial \tilde{Y}}{\partial H} - \frac{\partial Y}{\partial H} \right), \quad \frac{\partial Z_4}{\partial \psi} = \left(\frac{\partial \tilde{Y}}{\partial I} - \frac{\partial Y}{\partial I} \right).$$

Proceeding as in the proof of Lemma D.1 we see that the errors of averaging are controlled by the following terms

$$\begin{aligned} \frac{\partial H_{n+1}}{\partial H_n} : & \quad \varepsilon \oint \left| \frac{\partial^2 \tilde{X}}{\partial H^2} \right| ds = \mathcal{O} \left(\varepsilon |H|^{-3/2} \right) = \mathcal{O} \left(\varepsilon^{\frac{5}{8} + \frac{3\delta}{2}} \right), \\ \frac{\partial H_{n+1}}{\partial I_n} : & \quad \varepsilon \oint \left| \frac{\partial^2 \tilde{X}}{\partial I \partial H} \right| ds = \mathcal{O} \left(\varepsilon |H|^{-1} |\ln |H|| \right) = \mathcal{O} \left(\varepsilon^{\frac{3}{4} + \delta} |\ln \varepsilon| \right), \\ \frac{\partial I_{n+1}}{\partial H_n} : & \quad \varepsilon \oint \left| \frac{\partial^2 \tilde{Y}}{\partial H^2} \right| ds = \mathcal{O} \left(\varepsilon H^{-2} \right) = \mathcal{O} \left(\varepsilon^{\frac{1}{2} + 2\delta} \right), \\ \frac{\partial I_{n+1}}{\partial I_n} : & \quad \varepsilon \oint \left| \frac{\partial^2 \tilde{Y}}{\partial I \partial H} \right| ds = \mathcal{O} \left(\varepsilon |H|^{-3/2} |\ln |H|| \right) = \mathcal{O} \left(\varepsilon^{\frac{5}{8} + \frac{3\delta}{2}} |\ln \varepsilon| \right). \end{aligned}$$

(The reason why $\frac{\partial}{\partial I} \frac{\partial}{\partial H}$ appears in the second and fourth lines instead of a more dangerous $\left(\frac{\partial}{\partial H}\right)^2$ is because $\frac{\partial^2 \tilde{X}}{\partial H^2}$ and $\frac{\partial^2 \tilde{Y}}{\partial H^2}$ come with the factor δH and in the second and fourth lines we have $|\delta H| = \mathcal{O}(\sqrt{\varepsilon})$.) \square

Lemma F.6 gives the bounds of Lemma F.1(b) related to the change of I, H and their first derivatives. To obtain the bounds on the second derivative we consider the variational equation for the second derivative

$$\frac{d}{dt}(\delta^2 I, \delta^2 H) = \sqrt{\varepsilon} \frac{\partial(X, Y)}{\partial(I, H)}(\delta^2 I, \delta^2 H) + \sqrt{\varepsilon} \frac{\partial^2(X, Y)}{\partial(I, H)^2}((\delta I, \delta H), (\delta I, \delta H)).$$

This is a linear inhomogeneous equation where the inhomogeneous part is $\mathcal{O}(|H|^{-2} \sqrt{\varepsilon} |\ln^4 \varepsilon|)$ (see Lemma 3.1), the fundamental solution of the corresponding linear system is $\mathcal{O}(\ln^2 \varepsilon)$ and $|H|$ grows as $\sqrt{\varepsilon} t$ near the entrance phase and has a similar decay near the exit phase. Now the estimates on the second derivatives follows easily.

F.4. Estimates of the derivatives. We begin with the following general result.

Lemma F.7. (a) Let

$$\Phi(I, \mathcal{H}) = \int_{\sqrt{\mathcal{H}}}^{s^*} \frac{f(s, I, \mathcal{H})}{\sqrt{s^2 - \mathcal{H}}} ds$$

where f is a smooth bounded function. Then for $p \geq 0$

$$\frac{\partial^p \Phi}{\partial I^p} = \frac{1}{2} \frac{\partial^p f}{\partial I^p}(0, I, 0) |\ln \mathcal{H}| + \mathcal{O}(1)$$

and for $q > 0$

$$\frac{\partial^{p+q} \Phi}{\partial I^p \partial \mathcal{H}^q} = \frac{1}{2} \frac{\partial^p f}{\partial I^p}(0, I, 0) \frac{(-1)^q q!}{\mathcal{H}^q} + \mathcal{O}\left(\frac{1}{\mathcal{H}^{q-1}}\right).$$

(b) If $f(0, I, 0) \equiv 0$ then

$$\frac{\partial^p \Phi}{\partial I^p} = \int_0^{s^*} \frac{\frac{\partial^p f}{\partial I^p}(s, I, 0)}{\sqrt{s^2 - \mathcal{H}}} + \mathcal{O}(\mathcal{H} |\ln \mathcal{H}|),$$

$$\frac{\partial^{p+1} \Phi}{\partial I^p \partial \mathcal{H}} = -\frac{1}{2} \frac{\partial^{p+1} f}{\partial I^p \partial \mathcal{H}}(0, I, 0) \ln \mathcal{H} + \mathcal{O}(1)$$

and for $q \geq 2$

$$\frac{\partial^{p+1} \Phi}{\partial I^p \partial \mathcal{H}^q} = -\frac{1}{2} \frac{(-1)^{q-1} (q-1)!}{\mathcal{H}^{q-1}} \frac{\partial^{p+1} f}{\partial I^p \partial \mathcal{H}}(0, I, 0) + \mathcal{O}\left(\frac{1}{\mathcal{H}^{q-2}}\right).$$

(c) Let

$$\Psi(I, \mathcal{H}) = \int_{\sqrt{\mathcal{H}}}^{s^*} \frac{f(s, I, \mathcal{H})}{\sqrt{s - \mathcal{H}}} ds$$

Then Ψ is a smooth function.

Remark. In a typical application of this lemma we will have $\mathcal{H} = -H$.

Proof. (a) It is enough to estimate Φ and its \mathcal{H} derivatives since the differentiation with respect to I just replaces the integrand by its I derivatives.

Write

$$f(s, I, \mathcal{H}) = f(0, I, \mathcal{H}) + s \frac{\partial f}{\partial s}(0, I, \mathcal{H}) + s^2 f_2(s, I, \mathcal{H})$$

and split $\Phi = \Phi_0 + \Phi_1 + \Phi_2$ where Φ_i denotes the contribution of the corresponding terms in the above formula. Then

$$\begin{aligned} \Phi_0 &= \int_{\sqrt{\mathcal{H}}}^{s^*} \frac{f(0, I, \mathcal{H})}{\sqrt{s^2 - \mathcal{H}}} ds = \int_1^{s^*/\sqrt{\mathcal{H}}} \frac{f(0, I, \mathcal{H})}{\sqrt{u^2 - 1}} du \\ &= \int_1^{s^*/\sqrt{\mathcal{H}}} \frac{f(0, I, \mathcal{H})}{u} du - \int_1^{s^*/\sqrt{\mathcal{H}}} \frac{f(0, I, \mathcal{H})}{u\sqrt{u^2 - 1}(\sqrt{u^2 - 1} + u)} du \\ &= f(0, I, \mathcal{H}) \ln \frac{s^*}{\sqrt{\mathcal{H}}} + \mathcal{O}(1). \end{aligned}$$

Also

$$(F.4) \quad \frac{\partial \Phi_0}{\partial \mathcal{H}} = \int_1^{s^*/\sqrt{\mathcal{H}}} \frac{\frac{\partial f}{\partial \mathcal{H}}(0, I, \mathcal{H})}{\sqrt{u^2 - 1}} du - \frac{1}{2\mathcal{H}} \frac{f(0, I, \mathcal{H})}{\sqrt{1 - \frac{\mathcal{H}}{(s^*)^2}}}.$$

Thus the main contribution comes from the second term which equals to $-\frac{f(0, I, 0)}{2\mathcal{H}} + \mathcal{O}(1)$ while the first term is $\mathcal{O}(|\ln \mathcal{H}|)$.

(F.4) easily implies that

$$\frac{\partial^q \Phi_0}{\partial \mathcal{H}^q} = \frac{f(0, I, 0)}{2} \frac{(-1)^q q!}{\mathcal{H}^q} + \mathcal{O}\left(\frac{1}{\mathcal{H}^{(q-1)}}\right)$$

so it remains to show that the contributions of Φ_1 and Φ_2 are of lower order. We have

$$\Phi_1 = \int_{\sqrt{\mathcal{H}}}^{s^*} \frac{s \frac{\partial f}{\partial s}(0, I, \mathcal{H})}{s^2 - \mathcal{H}} ds = \frac{1}{2} \frac{\partial f}{\partial s}(0, I, \mathcal{H}) \sqrt{(s^*)^2 - \mathcal{H}}.$$

To handle Φ_2 we split

$$\Phi_2 = \int_{\sqrt{\mathcal{H}}}^{2\sqrt{\mathcal{H}}} \frac{s^2 f_2(s, I, \mathcal{H})}{\sqrt{s^2 - \mathcal{H}}} ds + \int_{2\sqrt{\mathcal{H}}}^{s^*} \frac{s^2 f_2(s, I, \mathcal{H})}{\sqrt{s^2 - \mathcal{H}}} ds.$$

The first term here equals

$$\Phi_1(I, \mathcal{H}) = \mathcal{H} \int_1^2 \frac{u^2 f(u\sqrt{\mathcal{H}}, I, \mathcal{H})}{\sqrt{u^2 - 1}} dz$$

so it admits the Taylor expansion in powers of $\sqrt{\mathcal{H}}$. To analyze the second term we expand the denominator into a power series (with unit radius of convergence)

$$\int_{2\sqrt{\mathcal{H}}}^{s^*} \frac{s f_2(s, I, \mathcal{H})}{\sqrt{1 - (\mathcal{H}/s^2)}} ds = \sum_k \omega_k \int_{2\sqrt{\mathcal{H}}}^{s^*} \frac{f(s, I, \mathcal{H}) \mathcal{H}^k}{s^{2k-1}} ds$$

and notice that the last integral is

$$\mathcal{O}\left(\frac{\mathcal{H}}{(2k-2)2^{k-2}}\right)$$

so that $\Phi_2 = \mathcal{O}(\mathcal{H})$. A similar argument shows that

$$\frac{\partial^q \Phi_2}{\mathcal{H}^q} = \mathcal{O}(\mathcal{H}^{1-q}).$$

This completes the proof of part (a).

The proof of part (b) is similar except that the estimate of Φ_0 is different. Namely we have $f(0, I, \mathcal{H}) = \mathcal{H} \tilde{f}(I, \mathcal{H})$ for a smooth function \tilde{f} so we gain an extra factor of \mathcal{H} compared to part (a).

(c) Introducing a new variable $s - \mathcal{H}$ instead of s we obtain

$$\Psi(I, \mathcal{H}) = \int_H^{s^*} \frac{h(s - \mathcal{H}, I, \mathcal{H})}{\sqrt{s - \mathcal{H}}} ds$$

for a smooth function h . Now the change of variables $u = \sqrt{s - \mathcal{H}}$ transforms

$$\Psi(I, \mathcal{H}) = 2 \int_0^{\sqrt{s^* - \mathcal{H}}} h(u^2, I, \mathcal{H}) du.$$

□

Proof of Lemma 3.1. We have

$$(F.5) \quad Y(I, H) = 2 \oint \bar{\alpha}(I, 0, \theta(s), 0) ds = \sqrt{2} \int_{\theta_1(H)}^{\theta_2(H)} \frac{\bar{\alpha}(I, 0, \theta, 0)}{\sqrt{H - U}} d\theta.$$

Choose some θ^* between $\theta_1(0)$ and $\theta_2(0)$ and split $Y = Y_1 + Y_2$ where Y_1 involves the integral from θ_1 to θ^* and Y_2 involves the integral from θ^* to θ_2 . To estimate the first term observe that θ_{jk} is the maximum of U and $U(\theta_{jk}(I), I) = 0$. Therefore $U(\theta, I) = -(\theta - \theta_{jk})^2 V(I, \theta)$ for a positive function V . Accordingly the denominator of (F.5) takes form $\sqrt{(\theta - \theta_{jk})^2 V - |H|}$. Introducing a new variable s by $\theta - \theta_{jk} = \frac{s}{\sqrt{V}}$ reduces Y_1 to the form of Lemma F.7(a). Similarly Y_2 can be estimated using Lemma F.7(c).

The estimates of X are similar except we use Lemma F.7(b) instead of Lemma F.7(a) to handle the first integral. □

Proof of Lemma F.5. The estimates for \tilde{X} and \tilde{Y} follow from the expressions for \dot{H}, \dot{I} and (F.3). (Recall, see (2.17) that the $\sqrt{\varepsilon}$ coefficient of \dot{H} vanishes near the origin).

To estimate the derivatives of \tilde{X} and \tilde{Y} we need to bound the derivatives of the coordinate change $(\theta, r, I) \rightarrow (H, \psi, I)$. We represent it as a composition of two changes.

(1) $(\theta, r, I) \rightarrow (H, r, I)$. Since $H = \frac{r^2}{2} + U$ we have

$$(F.6) \quad \delta\theta = \frac{\delta H - r\delta r - \frac{\partial U}{\partial I}\delta I}{\frac{\partial U}{\partial\theta}}.$$

Observe that

$$(F.7) \quad \frac{\partial U}{\partial\theta} \sim (\theta - \theta_{jk}) \text{ and}$$

$$(F.8) \quad r \leq \text{Const}|\theta - \theta_{jk}|$$

since $r^2 \leq 2|U|$.

(2) $(H, r, I) \rightarrow (H, \psi, I)$. We have

$$\psi = -\frac{1}{T(H, I)} \int_0^r \frac{\sqrt{(\frac{\partial U}{\partial\theta})^2 + z^2}}{\frac{\partial U}{\partial\theta} - \sqrt{\varepsilon}Q}(\theta(z, H, I), I) dz$$

where $\theta(z, H, I)$ is defined by the condition

$$(F.9) \quad U(\theta, I) + \frac{z^2}{2} = H.$$

By the same analysis as in Lemma F.7 we have

$$|T(H)| \leq \text{Const}, \quad \frac{\partial^{p+q}T}{\partial^p I \partial^q H} = \mathcal{O}\left(H^{\frac{1}{2}-q}\right).$$

Next,

$$(F.10) \quad \frac{\partial\psi}{\partial r} = \frac{\sqrt{(\frac{\partial U}{\partial\theta})^2 + r^2}}{T(\frac{\partial U}{\partial\theta} - \sqrt{\varepsilon}Q)} = \mathcal{O}(1)$$

due to (F.7) and (F.8). Further, $\frac{\partial\psi}{\partial H} = I + II + III + IV$ where

$$I = -\frac{\frac{\partial T}{\partial H}}{T^2}\psi = \mathcal{O}(1/\sqrt{|H|}),$$

$$II = -\frac{1}{T} \int_0^r \frac{z^2 \frac{\partial^2 U}{\partial\theta^2} \frac{\partial\theta}{\partial H}}{\sqrt{(\frac{\partial U}{\partial\theta})^2 + z^2} (\frac{\partial U}{\partial\theta} - \sqrt{\varepsilon}Q)^2} dz,$$

$$\begin{aligned} \mathbb{I} &= -\frac{\sqrt{\varepsilon}}{T} \int_0^r \frac{Q \frac{\partial U}{\partial \theta} \frac{\partial \theta}{\partial H}}{\sqrt{\left(\frac{\partial U}{\partial \theta}\right)^2 + z^2 \left(\frac{\partial U}{\partial \theta} - \sqrt{\varepsilon}Q\right)^2}} dz, \\ \mathbb{I}V &= -\frac{\sqrt{\varepsilon}}{T} \int_0^r \frac{\sqrt{\left(\frac{\partial U}{\partial \theta}\right)^2 + z^2 \frac{\partial Q}{\partial \theta} \frac{\partial \theta}{\partial H}}}{\left(\frac{\partial U}{\partial \theta} - \sqrt{\varepsilon}Q\right)^2} dz. \end{aligned}$$

Differentiating (F.9) we get $\frac{\partial \theta}{\partial H} = 1/\left(\frac{\partial U}{\partial \theta}\right)$ so

$$\begin{aligned} \text{(F.11)} \quad |\mathbb{I}| &\leq \text{Const} \int_0^r \frac{z^2}{\left(\frac{\partial U}{\partial \theta}\right)^4} dz \leq \text{Const} \int_0^r \frac{z^2}{U^2} dz \\ &\leq \text{Const} \int_0^r \frac{z^2}{\left(|H| + \frac{z^2}{2}\right)^2} dz \leq \frac{\text{Const}}{\sqrt{|H|}}. \end{aligned}$$

Similarly $|\mathbb{I}V| \leq \frac{\text{Const}\sqrt{\varepsilon}}{H} \leq \frac{\text{Const}}{\sqrt{|H|}}$, $|\mathbb{I}V| \leq \text{Const} \sqrt{\frac{\varepsilon}{|H|}}$.

Likewise computing $\frac{\partial \psi}{\partial I}$ reduces to estimating

$$\int_0^r \frac{z^2}{\left(\frac{\partial U}{\partial \theta}\right)^3} \frac{\partial \theta}{\partial I} dz.$$

Differentiating (F.9) we get

$$\frac{\partial \theta}{\partial I} = \frac{-\frac{\partial U}{\partial I}}{\frac{\partial U}{\partial \theta}} = \mathcal{O}(1)$$

since

$$0 = \frac{\partial U}{\partial \theta}(\theta_{jk}(I), I) \frac{\partial \theta_{jk}}{\partial I} + \frac{\partial U}{\partial I}(\theta_{jk}(I), I) = \frac{\partial U}{\partial I}(\theta_{jk}(I), I)$$

and so

$$\left| \frac{\partial U}{\partial I} \right| \leq \text{Const} |\theta - \theta_{jk}| \leq \text{Const} \left| \frac{\partial U}{\partial I} \right|.$$

Thus

$$\left| \frac{\partial \psi}{\partial I} \right| \leq \text{Const} \int \frac{z^2 dz}{(z^2/2 + |H|)^{3/2}} \leq \text{Const} |\ln |H||.$$

Accordingly

$$\delta \psi = \mathcal{O}(1) \delta r + \mathcal{O}\left(\frac{1}{\sqrt{|H|}}\right) \delta H + \mathcal{O}(|\ln |H||) \delta I.$$

Conversely by (F.10)

$$\text{(F.12)} \quad \delta r = \mathcal{O}(1) \delta \psi + \mathcal{O}\left(\frac{1}{\sqrt{|H|}}\right) \delta H + \mathcal{O}(|\ln |H||) \delta I.$$

Substituting this into (F.6) we get

(F.13)

$$\delta\theta = \frac{\delta H - \frac{\partial U}{\partial I} \delta I}{\frac{\partial U}{\partial \theta}} - \frac{r}{\left(\frac{\partial U}{\partial \theta}\right)} \left[\mathcal{O}(1) \delta\psi + \mathcal{O}\left(\frac{1}{\sqrt{|H|}}\right) \delta H + \mathcal{O}(|\ln |H||) \delta I \right].$$

The derivatives of \tilde{X} and \tilde{Y} with respect to (H, ψ, I) are obtained from the derivatives with respect to (θ, r, I) by substitution (F.12)–(F.13). Taking $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \theta}$ brings an extra factor of $(\theta - \theta_{jk})^{-1}$ (that is an extra factor of $\mathcal{O}\left(\frac{1}{\sqrt{|H|}}\right)$ whereas the substitution (F.12)–(F.13) contributes another factor $\mathcal{O}\left(\frac{1}{\sqrt{|H|}}\right)$ for H -derivatives and $\mathcal{O}(|\ln |H||)$ for I -derivatives. Integrals of $(\theta - \theta_{jk})^{-l}$ are estimated as above (cf e.g. (F.11)). The result follows. \square

F.5. Proof of Proposition 6.2*.

Proof. Part (a) follows from the estimates of Lemma F.1. Parts (b) and (c) are obtained similarly to the proof of Proposition 6.1 except we use Proposition B.1* instead of Proposition B.1. For part (c) we let $\bar{\delta} = 3\delta$ and observe that even though the second derivative bounds of Proposition B.1* are worse by the factor of $\varepsilon^{-(5/4+\bar{\delta})}$ it only results in $\mathcal{O}(\varepsilon^{-(1+\bar{\delta})})$ deterioration of the second derivative bounds in Proposition 6.2* since $\frac{\partial^2(J_\tau, H_\tau)}{\partial(J_0, H_0)^2}$ does not give the leading contribution in Proposition 6.1.

Part (d) follows since the orbits we consider make $\mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)$ rotations and the longest rotation takes time $\mathcal{O}(\sqrt{\varepsilon} |\ln \varepsilon|)$. To obtain part (e) we observe that the orbit can pass near the separatrix either during the entrance into resonance or during the exit from it. The measure of the former orbits is estimated similarly to Section E.5. To estimate the measure of the orbits which come too close to the resonance during the exit we observe that due to Corollary F.2 the image of each captured component consists of $\mathcal{O}(|\ln \varepsilon|)$ components (each component consisting of the points making the same number of rotation during the capture) and for each component the relative measure of the points coming too close to the separatrix is $\mathcal{O}\left(\frac{1}{\sqrt{|\ln \varepsilon|}}\right)$ (this is because in the notation of Corollary F.2 the map $H_0 \rightarrow H_3$ has bounded distortion). \square

G.1. Example 1. Here we compute the parameters of the limiting process for Example 1 of Section 4.

From (4.2) we have

$$L(I) = U'(Z(I)).$$

The critical points are given by the equation

$$\sin \theta_{cr} = U'(Z(I)).$$

We are interested in the saddle point corresponding to the maximum of U so $\theta_{cr} = \sin^{-1} U'(Z(I))$. The inner hamiltonian takes form

$$H = U'(Z(I))\theta + \cos \theta - [(U'(Z(I))\theta_{cr} + \cos \theta_{cr})].$$

Since

$$\frac{\partial \theta_{cr}}{\partial I} = \frac{U''}{U' \cos \theta_{cr}}$$

we have

$$\frac{H'}{\sqrt{\varepsilon}} = \frac{U''}{U'}(Z(I)) [r^2 - (\theta - \theta_{cr})] \sin \theta.$$

Introduce functions

$$\Lambda_{j,k}(N, E) = \oint r^j \sin^k \theta dt, \quad \Upsilon = \oint \theta \sin \theta dt,$$

where the integration is over the energy E curve of the pendulum with constant torque

$$(G.1) \quad x'' = N + \sin \theta.$$

Then the integral over the separatrix loop is computed as

$$\iint_{\Omega} \sin \theta dr d\theta = \Lambda_{2,1}(-U'(Z(I)), 0),$$

whereas

$$M = \frac{U''}{U'}(Z(I)) [-\Lambda_{2,1}(-U'(Z(I)), 0) + (\Upsilon(-U'(Z(I)), 0) - \theta_{cr} \Lambda_{0,1}(-U'(Z(I)), 0))].$$

Finally the inner averaged equation used to compute the entrance-exit function takes form

$$(G.2) \quad \begin{aligned} H' &= \frac{U''}{U'}(Z(I)) [\Lambda_{2,1}(-U'(Z(I)), H) \\ &- (\Upsilon(-U'(Z(I)), H) - \theta_{cr} \Lambda_{0,1}(-U'(Z(I)), H))] \\ I' &= \Lambda_{0,1}(-U'(Z(I)), H). \end{aligned}$$

Thus denoting by $\tau(I)$ the period of the averaged system (4.1) we get the following formulas for the limiting process

- The limiting equation is

$$(G.3) \quad \frac{d\mathcal{I}}{dt} = \frac{1}{\tau(\mathcal{I})} \sum_j \Lambda_{2,1}(-U'(Z_j(\mathcal{I})), 0)$$

where the summation is over all points where $U(Z_j) = \mathcal{I}$.

- The killing intensity is

$$(G.4) \quad \lambda(\mathcal{I}) = \sum_j \lambda_j(\mathcal{I})$$

where $\lambda_j(\mathcal{I}) =$

$$\left(\frac{U''(Z_j(\mathcal{I}))}{U'(Z_j(\mathcal{I}))\tau(\mathcal{I})} [-\Lambda_{2,1}(-U'(Z_j(\mathcal{I})), 0) + (\Upsilon(-U'(Z_j(\mathcal{I})), 0) - \theta_{cr}\Lambda_{0,1}(-U'(Z_j(\mathcal{I})), 0))] \right)_+.$$

- The entrance-exit function is computed using (G.2).

G.2. Example 2. Here we compute the limiting process for Example 2.

From (4.4) the inner hamiltonian is

$$H = \frac{r^2}{2} \mp (N\theta - \cos\theta) \sqrt{\frac{2I}{N}} \pm (N\theta_{cr} - \cos\theta_{cr}) \sqrt{\frac{2I}{N}}$$

where θ_{cr} is the same as in the Example 1. Thus we have

$$L(I) = \sqrt{2NI}.$$

Observe that there are four resonances corresponding to possible choices of signs z and $\cos\psi$ but only the sign of $z\cos\psi$ is important. For example, if $z\cos\psi > 0$ we get

$$\frac{H'}{\sqrt{\varepsilon}} = \frac{1}{\sqrt{2NI}} \left(r^2(N + \sin\theta) + (N\theta - \cos\theta + N\theta_{cr} - \cos\theta_{cr}) \sqrt{\frac{2I}{N}} \sin\theta \right)$$

Using the equality

$$(N\theta - \cos\theta + N\theta_{cr} - \cos\theta_{cr}) \sqrt{\frac{2I}{N}} = \frac{r^2}{2} - H$$

we obtain

$$\frac{H'}{\sqrt{\varepsilon}} = \frac{Nr^2 + \frac{3}{2}r^2 \sin\theta - H \sin\theta}{\sqrt{2NI}}.$$

Likewise in case $z\cos\psi < 0$ we obtain

$$\frac{H'}{\sqrt{\varepsilon}} = -\frac{Nr^2 + \frac{3}{2}r^2 \sin\theta - H \sin\theta}{\sqrt{2NI}}.$$

Next observe the change of variables $\bar{\theta} = \pi - \theta$ transforms equation (G.1) into

$$\bar{\theta}'' = -\theta'' = -N - \sin \theta = -N - \sin \bar{\theta}.$$

Accordingly the contributions of four resonances to the drift term cancel out and so the limiting equation is $\mathcal{I}' = 0$. Next the change of time $\tilde{t} = (2I/N)^{1/4}t$ transform the unperturbed inner system into (G.1) and the velocity becomes rescaled by $r = \tilde{r}(2I/N)^{1/4}$ and the total energy is rescaled by $\tilde{H} = H\sqrt{\frac{N}{2I}}$. Observe that the separatrix integrals corresponding to $z \cos \psi = \pm 1$ have opposite signs so one of them is positive and the other is negative. Thus the total contribution of the separatrix integrals is

$$\sum_{j=1}^4 \frac{(M_j(I))_+}{L_j(I)} = \frac{|3\Lambda_{2,1}(I) + 2N\Lambda_{2,0}(I)|}{(8N^5 I^3)^{1/4}}.$$

The inner averaged system used to compute the entrance-exit function takes form

$$\begin{aligned} \tilde{H}' &= \pm \left(\frac{2I}{N}\right)^{1/4} \frac{1}{2I} \left[N\Lambda_{2,0}(N, \tilde{H}) + \frac{3}{2}\Lambda_{2,1}(N, \tilde{H}) - \tilde{H}\Lambda_{0,1}(N, \tilde{H}) \right], \\ I' &= \mp \left(\frac{2I}{N}\right)^{1/4} \Lambda_{0,1}(N, \tilde{H}). \end{aligned}$$

(Here we have used the variable $\tilde{H} = H\sqrt{\frac{N}{2I}}$ rather than H to compute inner averaged equation since it leads to simpler formulas while the entrance–exit function is the same). After a further change of time the inner averaged system takes form

$$(G.5) \quad \tilde{H}' = \pm \frac{N\Lambda_{2,0}(N, \tilde{H}) + \frac{3}{2}\Lambda_{2,1}(N, \tilde{H}) - \tilde{H}\Lambda_{0,1}(N, \tilde{H})}{2I},$$

$$(G.6) \quad I' = \mp \Lambda_{0,1}(N, \tilde{H}).$$

(Here the sign is chosen so that \tilde{H} is decreasing near $(0, I_0)$.)

Denote by $\bar{\tau}(I)$ is the period of the averaged system (4.3). Then the limiting process is a jump process with jump intensity

$$(G.7) \quad \lambda(\mathcal{I}) = \frac{|3\Lambda_{2,1}(I) + 2N\Lambda_{2,0}(I)|}{\bar{\tau}(I)(8N^5 I^3)^{1/4}}.$$

and the jump function computed using the inner averaged system (G.5)–(G.6).

APPENDIX H. DISTORTION BOUND.

Lemma H.1. *Suppose that $\tilde{\gamma}$ is a subcurve with coordinates $[E_1, E_2]$ such that $|\bar{E}_2 - \bar{E}_1| \leq \Delta$. If*

$$\left| \frac{d^2 \bar{E}}{dE^2} \right| \leq \mathbf{L} \left(\frac{d\bar{E}}{dE} \right)^2 \quad \text{and} \quad \frac{d\bar{E}}{dE} \neq 0$$

then we have the following distortion bound: for all $E_3, E_4 \in [E_1, E_2]$ we have

$$(H.1) \quad e^{-\mathbf{L}\Delta} \leq \frac{|E_4 - E_3| |\bar{E}_2 - \bar{E}_1|}{|\bar{E}_4 - \bar{E}_3| |E_2 - E_1|} \leq e^{\mathbf{L}\Delta}$$

Proof. By the Intermediate Value Theorem

$$\frac{|E_4 - E_3| |\bar{E}_2 - \bar{E}_1|}{|\bar{E}_4 - \bar{E}_3| |E_2 - E_1|} = \left| \frac{d\bar{E}}{dE}(E') / \frac{d\bar{E}}{dE}(E'') \right|$$

for some $E' \in [E_1, E_2]$, $E'' \in [E_3, E_4]$. On the other hand

$$\left| \frac{d}{dE} \ln \left(\frac{d\bar{E}}{dE} \right) \right| \leq \mathbf{L} \left| \frac{d\bar{E}}{dE} \right|$$

Integrating we get

$$e^{-\mathbf{L}|\bar{E}' - \bar{E}''|} \leq \left| \frac{d\bar{E}}{dE}(E') / \frac{d\bar{E}}{dE}(E'') \right| \leq e^{\mathbf{L}|\bar{E}' - \bar{E}''|}.$$

The lemma follows. □

REFERENCES

- [1] Afraimovich V. S. & Shilnikov L. P. *The ring principle in problems of interaction between two self-oscillating systems*, (Russian) J. Appl. Math. Mech. **41** (1977) 618–627.
- [2] Anosov D. V. *Averaging in systems of ordinary differential equations with rapidly oscillating solutions*, (Russian) Izv. Akad. Nauk SSSR Ser. Mat. **24** (1960) 721–742.
- [3] Arnold V. I. *Small denominators and problems of stability of motion in classical and celestial mechanics*, Russian Math. Surv. **18** (1963) 85–191.
- [4] Arnold V. I. *Applicability conditions and an error bound for the averaging method for systems in the process of evolution through a resonance*, Soviet Math. Doklady **6** (1965) 331–334.
- [5] Arnold V. I., Kozlov, V. V., & Neishtadt, A. I. *Mathematical aspects of classical and celestial mechanics*, 3d ed. Encyclopaedia Math. Sci. **3** (2006) Springer, Berlin.
- [6] Bakhtin V. I. *Averaging in a general-position single-frequency system*, Dif. Eq. **27** (1991) 1051–1061.
- [7] Bakhtin V. I. *Cramer asymptotics in the averaging method for systems with fast hyperbolic motions*, (Russian) Tr. Mat. Inst. Steklova **244** (2004) 65–86.
- [8] Brin M. & Freidlin M. *On stochastic behavior of perturbed Hamiltonian systems*, Erg. Th., Dyn. Sys. **20** (2000) 55–76.
- [9] Bogoliubov N. N. & Mitropolsky Y. A. *Asymptotic methods in the theory of non-linear oscillations*, International Monographs on Advanced Math., Phys., Hindustan Publishing Corp., Delhi, Gordon and Breach Science Publishers, New York, 1961.
- [10] Bogolyubov N. N. & Zubarev D. N. *Method of asymptotic approximation for systems with rotating phase and its application to motion of charged particles in a magnetic field*, Ukr. Mat. Zh. **7** 5-17 (1955) (in Russian).
- [11] Cary J. R., Escande D. F. & Tennyson J. L. *Adiabatic invariant change due to separatrix crossing*, Phys. Rev. A **34** (1986) 4256–4275.
- [12] Cary J. R., Skodje R. T. *Phase change between separatrix crossings*, Physica D **36** (1989) 287–316.
- [13] Chernov N. & Dolgopyat D. *Brownian Motion-1*, Memoirs AMS v. **198** no. 927 (2009).
- [14] Chirikov, B. V. *The Passage of a Nonlinear Oscillating System through Resonance*, Soviet Physics Doklady **4** (1959) 390.
- [15] de Simoi J. *Abundance of escaping orbits in a family of antiintegrable limits of the standard map*, PhD thesis, University of Maryland, 2009.
- [16] Dolgopyat D. *Limit theorems for partially hyperbolic systems*, Trans. AMS **356** (2004) 1637–1689.
- [17] Dolgopyat D. *On differentiability of SRB states for partially hyperbolic systems*, Invent. Math. **155** (2004), no. 2, 389–449.
- [18] Dolgopyat D. *Averaging and invariant measures*, Mosc. Math. J. **5** (2005) 537–576.
- [19] Dolgopyat D. *Bouncing balls in non-linear potentials*, Discrete Contin. Dyn. Syst. **22** (2008) 165–182.
- [20] Fatou P. *Sur le mouvement d'un système à des forces de courte période*, Bull. Soc. Math. France **56** (1928) 98–139.

- [21] Fredlin M. I. *The averaging principle and theorems on large deviations*, Russian Math. Surv. **33** (1978) 117176.
- [22] Freidlin M. I. *Random and deterministic perturbations of nonlinear oscillators*, Proc. ICM-98, Doc. Math., 1998, Extra Vol. III, 223–235.
- [23] Fredlin M. I. *Autonomous stochastic perturbations of dynamical systems*, Acta Appl. Math. **78** (2003) 121–128.
- [24] Freidlin M. I. & Wentzell A. D. *Random perturbations of dynamical systems*, 2d ed. Grundlehren der Mathematischen Wissenschaften **260** (1998) Springer-Verlag, New York.
- [25] Gihman I. I. & Skorohod A. V. *The theory of stochastic processes-I*, Die Grundlehren der mathematischen Wissenschaften **210** (1974) Springer-Verlag, New York-Heidelberg.
- [26] Goldreich P. *An explanation of the frequent occurrence of commensurable mean motions in the solar system*, Monthly Notices Royal Astronomical Soc. **130** (1965) 159–181.
- [27] Goldreich P. & Peale, S. *Spin-orbit coupling in the solar system*, Astronomical J. **71** (1966) 425–437.
- [28] Guckenheimer J., Wechselberger M. & Young L.-S. *Chaotic attractors of relaxation oscillators*, Nonlinearity **19** (2006) 701–720.
- [29] Hirsch, M. W., Pugh C. C. & Shub M. *Invariant manifolds*, Lect. Notes Math., **583** (1977) Springer-Verlag, Berlin-New York.
- [30] Henrard J. and Lemaitre A. *A mechanism of formation for the Kirkwood gaps*, Icarus **55** (1983) 482–494.
- [31] Itin A. P., Neishtadt A. I. & Vasiliev, A. A. *Captures into resonance and scattering on resonance in dynamics of a charged relativistic particle in magnetic field and electrostatic wave*, Phys. D **141** (2000) 281–296.
- [32] Kasuga T. *On the adiabatic theorem for the Hamiltonian system of differential equations in the classical mechanics I-III* Proc. Japan Acad. Sci. **37** (1961) 366–382.
- [33] Kifer Yu. *Averaging in dynamical systems and large deviations*, Invent. Math. **110** (1992) 337–370.
- [34] Kifer Yu. *Some recent advances in averaging*, in *Modern dynamical systems and applications*, (Ed. M. Brin, B. Hasselblatt & Ya. Pesin) 385–403, Cambridge Univ. Press, Cambridge, 2004.
- [35] Kolmogorov A. N. *On conservation of conditionally periodic motions for a small change in Hamilton's function*, Dokl. Akad. Nauk SSSR (N.S.) **98**, (1954). 527–530 (Russian).
- [36] Kruskal M. *Asymptotic theory of Hamiltonian and other systems with all solutions nearly periodic*, J. Mathematical Phys. **3** (1962) 806–828.
- [37] Lochak P. & Meunier C. *Multiphase Averaging for Classical Systems*, Springer Appl. Math. Sci. **72** 1988.
- [38] Mandelstam L. I & Papalexi N. D. *Über die Begründung einer Methode für die Näherungslosung von Differentialgleichungen*, J. Exp. Theor. Phys. **4** (1934) 117.
- [39] Moser J. *Stable and random motions in dynamical systems. With special emphasis on celestial mechanics*, Annals of Math. Studies **77** Princeton University Press, Princeton, NJ.

- [40] Neishtadt, A. I. *Passage through a separatrix in a resonance problem with a slowly-varying parameter*, J. Appl. Math. Mech. **39** (1975) 594-605.
- [41] Neishtadt, A. I. *Passage through a resonance in the two-frequency problem*, Sov. Phys. Dokl. **20** (1975) 189-191.
- [42] Neishtadt A. I. *Averaging in multifrequency systems-II*, Sov. Phys., Dokl. **21** (1976) 80-82.
- [43] Neishtadt A. I. *On resonant problems in nonlinear systems*, PhD thesis, (1975) Moscow State University.
- [44] Neishtadt, A. I. *Change of an adiabatic invariant at a separatrix*, Sov. J. Plasma Phys. **12** (1986) 568-573.
- [45] Neishtadt A. I. *On the change in the adiabatic invariant on crossing a separatrix in systems with two degrees of freedom*, J. Appl. Math. Mech. **51** (1987) 586-592.
- [46] Neishtadt A. I. *On perturbation theory of nonlinear resonant systems*, Habilitation (1988) Moscow State University.
- [47] Neishtadt A. I. *Averaging, capture into resonances, and chaos in nonlinear systems*, in *Chaos* (Ed. D. K. Campbell) Amer. Inst. Phys., New York, 1990, 261-273.
- [48] Neishtadt A. I. *Averaging and passage through resonances*, Proc. Int. Congress Math. (Kyoto, 1990), Math. Soc. Japan, Tokyo, 1991, 1271-1283.
- [49] Neishtadt A. I. *On probabilistic phenomena in perturbed systems*, Selecta Math. Soviet. **12** (1993) 195-210.
- [50] Neishtadt A.I. (1993) *On destruction of adiabatic invariants in multi-frequency systems*, In: Perelló C., Simó C., Solà-Morales J. (eds.) *Equadiff 91*, International conference on differential equations, Vol. 1, pp. 195-207.
- [51] Neishtadt A.I. *On adiabatic invariance in two-frequency systems*, In: Simó C. (ed.) *Hamiltonian systems with three or more degrees of freedom*, Proceedings of NATO ASI, Series C, vol. **533** Kluwer, Dordrecht, pp. 193-212.
- [52] Neishtadt A. I, Sidorenko V. V & Treschev D. V. *Stable periodic motions in the problem of passage through a separatrix*, Chaos **7** (1997), 2-11.
- [53] Neishtadt A. & Vasiliev A. *Phase change between separatrix crossings in slow-fast Hamiltonian systems*, Nonlinearity **18** (2005) 1393-1406.
- [54] Peixoto M. M. *Structural stability on two-dimensional manifolds*, Topology **1** (1962) 101-120.
- [55] Operchuk I.M. *Study of statistical properties of multiple resonance passages*, Diploma paper, Moscow State University 2003.
- [56] Ott W. & Wang Q. *Dissipative homoclinic loops and rank one chaos*, preprint, arXiv:0802.4283.
- [57] Piro O. & Feingold M. *Diffusion in three-dimensional Liouvillean maps*, Phys. Rev. Lett. **61**, 1799 - 1802 (1988)
- [58] Sanders J. A., Verhulst F. & Murdock J. *Averaging methods in nonlinear dynamical systems*, 2d ed. Appl. Math. Sci. **59** (2007) Springer, New York.
- [59] Vainchtein D. L., Neishtadt, A. I. & Mezić I. *On passage through resonances in volume-preserving systems*, Chaos **16** (2006), no. 4 paper 043123.
- [60] Vainchtein D. L., Rovinsky E. V., Zelenyi, L. M. & Neishtadt, A. I. *Resonances and particle stochasticization in nonhomogeneous electromagnetic fields*, J. Nonlinear Sci. **14** (2004) 173-205.

- [61] Vainchtein D. L., Vasiliev A. A. & Neishtadt A. I. *Electron dynamics in a parabolic magnetic field in the presence of an electrostatic wave*, Plasma Physics Reports **35** (2009) 1021–1031.
- [62] Vasilev A. A., Zaslavskii G. M., Natenzon M. Ya., Neishtadt A. I., Petrovichev B. A., Sagdeev R. Z. & Chernikov A. A. *Attractors and stochastic attractors of motion in a magnetic field*, Soviet Phys. JETP **67** (1988) 2053–2062.
- [63] Wang Q. & Young L.-S. *From invariant curves to strange attractors*, Comm. Math. Phys. **225** (2002) 275–304.