On the phase change for perturbations of Hamiltonian systems with separatix crossing

Anatoly Neishtadt and Alexey Okunev

Abstract

We consider general perturbations of Hamiltonian systems with one degree of freedom such that the orbits of the perturbed system cross a separatix of the unperturbed system. The point where the separatix crossing occurs is described by a parameter called the pseudo-phase, it is known that this point depends on the initial conditions in a quasi-random way. We prove an asymptotic formula for the dependence of the pseudo-phase on the initial conditions. Such formula is a necessary ingredient for the study of quasi-random phenomena associated with multiple separatix crossings. Our proof is based on a detailed study of the averaged system of order 2. We also estimate how well the solutions of the averaged system of order 2 approximate the solutions of the perturbed system up to the separatix.

1 Introduction

Let us consider a Hamiltonian system with one degree of freedom

\[
\begin{align*}
\dot{q} &= \frac{\partial H}{\partial p}, & \dot{p} &= -\frac{\partial H}{\partial q}, & \dot{z} &= 0
\end{align*}
\]

(1.1)

with the Hamiltonian \( H(p, q, z) \) depending on a parameter \( z \). The parameter \( z \) may be one- or multidimensional; we will assume \( z \) to be a column-vector \((z_1, \ldots, z_k)\), \( k \geq 1 \). Let us also assume that for all values of \( z \) this system has a non-degenerate saddle \( C(z) \) with two separatix loops \( l_1 \) and \( l_2 \). For definiteness, we will assume that the separatix loops form a figure eight in the plane (Fig. 1). However, one may also consider other phase portraits, e.g. the phase portrait of a pendulum on the cylinder, or in the plane one of the separatix loops may be inside the other. Indeed, restricted to a small neighborhood of the union of the separatrices, these phase portraits are the same as the figure eight. The separatrices split the phase space into three domains. Let us denote by \( G_3 \) the “outer” domain adjacent to \( l_1 \cup l_2 \) and by \( G_1 \) and \( G_2 \) the domains inside the loops adjacent only to \( l_1 \) and only to \( l_2 \).

![Figure 1: The unperturbed system.](image)

*The work was supported by the Leverhulme Trust (Grant No. RPG-2018-143).*
Now let us add a small perturbation $\varepsilon f$:

\[
\dot{q} = \frac{\partial H}{\partial p} + \varepsilon f_q(p, q, z, \varepsilon), \\
\dot{p} = \frac{\partial H}{\partial q} + \varepsilon f_p(p, q, z, \varepsilon), \\
\dot{z} = \varepsilon f_z(p, q, z, \varepsilon).
\]  

(1.2)

We assume that $H$ is analytic and $f$ is $C^2$. We use that $H$ is analytic to apply the local normal form [5] in a neighborhood of $C$. It will also be convenient to assume that $H(C) = 0$ for all $z$.

The equation (1.2) is a general form for perturbations of Hamiltonian systems with one degree of freedom. The solutions of the perturbed system may cross the separatrices of the unperturbed system, this is called separatix crossing. The phenomena associated with separatix crossings are much better studied for the case when the perturbed system is also Hamiltonian. It may be a slow-fast Hamiltonian system, or (this could be treated as a particular case of slow-fast Hamiltonian systems) a system with the Hamiltonian slowly depending on the time. For such systems separatix crossing leads to a jump of the improved adiabatic invariant of order $\varepsilon \ln \varepsilon$ (without separatix crossings the improved adiabatic invariant may only change by $O(\varepsilon^2)$ over times of order $\varepsilon^{-1}$ for the one degree of freedom case). There are formulas for the value of this jump ([11], [3], [7], [8]), and this value depends on a parameter called the pseudo-phase (we use the terminology from [10], another name appearing in the literature is “crossing parameter” [4]) which describes the phase at the moment of separatix crossing. For dissipative perturbations, tracking the evolution of slow variables after a separatix crossing with accuracy better than $O(\varepsilon)$ also requires knowing the value of pseudo-phase. In the Hamiltonian case there are also formulas for phase change when approaching the separatix (or, equivalently, for the pseudo-phase). Such formulas were obtained (using the averaging method) in [4] for Hamiltonian systems with one degree of freedom and slow time dependence; in [10] for slow-fast Hamiltonian systems with one degree of freedom corresponding to fast motion. In [2] the authors use the averaging method to compute the phase change for perturbed strongly nonlinear oscillators. Unlike [4] and [10], they do not provide an estimate for the accuracy of using the averaging method, but instead check that the result compares well with numerical experiments. These formulas together with the formulas for the change of the adiabatic invariant allow to study trajectories with multiple separatix crossings. Let us mention the existence of stability islands [9], [12] established using the formulas for phase change.

Separatix crossing also leads to probabilistic effects. As the value of the pseudo-phase may change by $O(1)$ for a change of $O(\varepsilon)$ in the initial condition, the jump of adiabatic invariant mentioned above may be considered as a random value. Moreover, solutions starting in the domain $G_3$ adjacent to both separatrices may cross any of the separatrices and proceed to one of the domains $G_1$ and $G_2$. Which separatix will be crossed is also determined by the pseudo-phase, so different outcomes are mixed in the phase space. There are two natural ways to define the probability of capture in each $G_i$. In one of them, the initial data is fixed together with its small neighborhood, and then the relative measure of the points in this neighborhood captured in $G_i$ is taken for $\varepsilon \to 0$. In the other one, the initial data is fixed and the relative measure of $\varepsilon < \varepsilon_0$ such that the trajectory is captured in $G_i$ is taken for $\varepsilon_0 \to 0$. A formula for the probability of capture in the first sense is proved in [6], and it was suggested that for the other definition the same formula also works.

We show that a formula for the pseudo-phase similar to the formula from [10] for slow-fast Hamiltonian systems also holds for arbitrary perturbations (1.2) of Hamiltonian systems with one degree of freedom. This generalises the formula from [10] (however, our error term is slightly worse). This formula for the pseudo-phase allows to compute the probability of capture in both senses above, thus proving that it is indeed given by the same formula. We also estimate how well the solutions of the perturbed system are approximated by the solutions of the averaged system of order 2 when approaching the separatrices. This is needed to prove the formula for the pseudo-phase, but is also of independent interest.

The general plan of the proof of the formula for the pseudo-phase is close to the one in [10]. However, instead of the improved adiabatic invariant considered in [10] we consider the averaged system of order 2. An important part of our paper is obtaining estimates for the coefficients of this system. It is worth to note that the rate of change of the Hamiltonian $H$ of the unperturbed system along the solutions of the the averaged system of order 2 is $1 + O(\varepsilon)$ times the rate of change of $H$ along the solutions of the averaged system of order 1 even near the separatix. This means that the solutions of the averaged system of order 2 cross the separatrices of the unperturbed Hamiltonian system.

Let us briefly discuss the structure of this paper. In Sections 2 – 6 we give the necessary preliminaries on the averaging method and state the results of this paper. The results are proved in Sections 7 – 13. For definiteness, the proofs are for the figure eight phase portrait and for trajectories starting in $G_3$. This is always assumed in Sections 7 – 13. The proofs generalise for other cases straightforwardly. In Sections 2 – 3 we discuss the action-angle variables of the unperturbed system and the coordinate change provided by the averaging
method. In Section 4 the averaged system of order 2 is defined, and we present the estimate on how well it approximates the solutions of (1.2). This estimate is proved in Section 7. In Section 5 we state the formula (5.1) for the pseudo-phase. It is proved in Sections 8 – 9. In Section 6 we discuss the probabilities of capture into different domains after separatrices crossing. Sections 10 – 13 are devoted to technical details. In Section 10 details related to the coordinate change provided by the averaging method are presented. In Sections 11 – 13 estimates related to the action-angle variables, the coordinate change provided by the averaging method, and the averaged system of order 2 are established.

2 Energy-angle variables

We are interested in trajectories starting in one of the domains \( G_1, G_2, G_3 \) (let us denote this domain \( G_i \)) and approaching the separatrix/separatrices of the unperturbed system. It will be convenient to assume that \( H(z) = 0 \) for all \( z \) and \( H > 0 \) in \( G_i \). Let us denote by \( f_\varepsilon(p, q; z, \varepsilon) = f_{\varepsilon, \text{int}} + f_{\varepsilon, \text{ext}} + f_{\varepsilon, \text{var}} \) the rate of change of \( H \) divided by \( \varepsilon \). For \( i = 1, 2 \) denote \( \Theta_i(z) = -f_\varepsilon(p(t), q(t), z; 0)dt \) (here \( t \) is the time for the unperturbed system). Let \( \Theta_3 = \Theta_1 + \Theta_2 \). We assume that \( \Theta_i > 0 \) for all considered values of \( z \), this means that the trajectories of the perturbed system starting close to the separatrix/separatrices eventually approach the separatrix/separatrices of the unperturbed system.

Let us state a useful relation between the derivatives of the components of \( f_\varepsilon \).

\[
\begin{align*}
\partial h_\varepsilon &= \frac{\partial f_\varepsilon}{\partial p} + \sum_{i=1}^{n} \partial f_{\text{var}} \partial w_i + \frac{1}{T} \frac{\partial T}{\partial h} f_h = \text{div}(f), \\
\text{where } \text{div}(f) &= \frac{\partial f_\varepsilon}{\partial q} + \frac{\partial f_\varepsilon}{\partial p} + \sum_{i=1}^{n} \frac{\partial f_{\text{var}}}{\partial z_i}. 
\end{align*}
\]

Let us state a useful relation between the derivatives of the components of \( f_\varepsilon \).

Lemma 2.1.

Proof. Let us first prove that \( \frac{\partial f_\varepsilon}{\partial q} + \frac{\partial f_\varepsilon}{\partial p} + \sum_{i=1}^{n} \frac{\partial f_{\text{var}}}{\partial z_i} = \frac{\partial f_\varepsilon}{\partial q} + \frac{\partial f_\varepsilon}{\partial p} + \sum_{i=1}^{n} \frac{\partial f_{\text{var}}}{\partial z_i}. \) Here \( I \) is the action of the unperturbed system and \( f_\varepsilon, f_w, f_\varepsilon \) is the vector field \( f \) written in the variables \( I, w, \varphi \).

Recall that the divergence of a vector field \( v \) with respect to a volume form \( \alpha \) is a function \( \text{div}_\alpha(v) \) such that \( \mathcal{L}_v(\alpha) = \text{div}_\alpha(v) \cdot \alpha \) (here \( \mathcal{L} \) denotes the Lie derivative). In the coordinates \( x_1, \ldots, x_n \) for the euclidean volume form the vector field \( \partial x_1 \wedge \cdots \wedge \partial x_n \) we have \( \text{div}_\alpha(dx_1 \wedge \cdots \wedge dx_n) = \frac{\partial}{\partial x_1} + \cdots + \frac{\partial}{\partial x_n}. \)

Hence the equality rewrites as \( \text{div}(dp \wedge dq \wedge dz_1 \wedge \cdots \wedge dz_n) = \frac{\partial}{\partial x_1} + \cdots + \frac{\partial}{\partial x_n}. \)

Finally, using that \( \frac{\partial f_\varepsilon}{\partial q} \) (this follows from the Hamiltonian equations in the action-angle variables) and \( f_3 = \frac{\partial f_\varepsilon}{\partial h_\varepsilon} \), we can compute that \( \frac{\partial f_\varepsilon}{\partial q} = \frac{\partial f_\varepsilon}{\partial h_\varepsilon} + \frac{\partial f_\varepsilon}{\partial h_\varepsilon} f_h. \)

3 Averaging chart

We start with the system (2.1). In line with the general approach of the averaging method, let us find a change of variables

\[
\begin{align*}
h = \bar{h} + \varepsilon u_{h,1}(\bar{h}, \bar{w}, \bar{\varphi}, \varepsilon) + \varepsilon^2 u_{h,2}(\bar{h}, \bar{w}, \bar{\varphi}, \varepsilon), \\
w = \bar{w} + \varepsilon u_{w,1}(\bar{h}, \bar{w}, \bar{\varphi}, \varepsilon) + \varepsilon^2 u_{w,2}(\bar{h}, \bar{w}, \bar{\varphi}, \varepsilon), \\
\varphi = \bar{\varphi} + \varepsilon u_{\varphi,1}(\bar{h}, \bar{w}, \bar{\varphi}, \varepsilon)
\end{align*}
\]

(3.1)
| Expression | Estimates | Obtained in |
|------------|-----------|-------------|
| $T$        | $T, \frac{\partial T}{\partial u}, \frac{\partial^2 T}{\partial u^2} = O(\ln(h)); \frac{\partial T}{\partial w}, \frac{\partial^2 T}{\partial w^2} = O(h^{-1}); \frac{\partial^2 T}{\partial w \partial u} = O(h^{-2})$ | Section 11.3 |
| $\omega$   | $\omega, \frac{\partial \omega}{\partial u}, \frac{\partial^2 \omega}{\partial u^2} = O(\ln^{-1} h); \frac{\partial \omega}{\partial w}, \frac{\partial^2 \omega}{\partial w^2} = O(h^{-1} \ln^{-2} h); \frac{\partial^2 \omega}{\partial w \partial u} = O(h^{-2} \ln^{-2} h)$ | Section 11.3 |
| $f_{w_i}$  | $f_{w_i}, \frac{\partial f_{w_i}}{\partial w_{i}}, \frac{\partial^2 f_{w_i}}{\partial w_{i}^2} = O(1); \frac{\partial f_{w_i}}{\partial w}, \frac{\partial^2 f_{w_i}}{\partial w^2} = O(h_1^{-1} \ln^{-1} h); \frac{\partial^2 f_{w_i}}{\partial w \partial w_i} = O(h_2^{-1} \ln^{-1} h); \frac{\partial^2 f_{w_i}}{\partial w \partial u} = O(h_2^{-1} \ln^{-2} h)$ | Section 11.5 |
| $f_h$      | $f_h, \frac{\partial f_h}{\partial w}, \frac{\partial^2 f_h}{\partial w^2} = O_*(1)$, other estimates as for $f_{w_i}$ | Section 11.5 |
| $\text{div } f$ | As for $f_{w_i}$ | Section 11.5 |
| $f_\varphi$ | $f_\varphi, \frac{\partial f_\varphi}{\partial \varphi} = O_*(h_1^{-1} \ln^{-2} h); f_\varphi(h, w, 0) = O(h_1^{-1/2} \ln^{-1} h); \frac{\partial f_\varphi}{\partial \varphi} = O(h_2^{-1} \ln^{-2} h)$ | Section 11.5 |
| $u_{h,1}$  | $u_{h,1}, \frac{\partial u_{h,1}}{\partial \varphi}, \frac{\partial^2 u_{h,1}}{\partial \varphi^2} = O(1); \frac{\partial u_{h,1}}{\partial w}, \frac{\partial^2 u_{h,1}}{\partial w \partial \varphi} = O(h_1^{-1} \ln^{-1} h); \frac{\partial^2 u_{h,1}}{\partial w \partial w_1} = O(h_2^{-1} \ln^{-1} h)$ | Section 13 |
| $u_{h,1,1}$ | As for $u_{h,1}$ | Section 13 |
| $u_{w,1}$  | $u_{w,1}, \frac{\partial u_{w,1}}{\partial \varphi}, \frac{\partial^2 u_{w,1}}{\partial \varphi^2} = O(h_1^{-1} \ln^{-1} h); \frac{\partial u_{w,1}}{\partial w} = O(h_2^{-1} \ln^{-1} h)$ | Section 13 |
| $f_{w,1}$  | $f_{w,1}, \frac{\partial f_{w,1}}{\partial \varphi} = O(1), \frac{\partial f_{w,1}}{\partial w} = O(h_1^{-1} \ln^{-2} h)$ | Section 13 |
| $h_1$      | $h_1, \frac{\partial h_1}{\partial \varphi}, \frac{\partial^2 h_1}{\partial \varphi^2} = O(h_1^{-1} \ln^{-3} h); \frac{\partial h_1}{\partial w} = O(h_2^{-1} \ln^{-3} h)$ | Section 13 |
| $u_{h,2}$  | $u_{h,2}, \frac{\partial u_{h,2}}{\partial \varphi} = O(h_1^{-1}); \frac{\partial u_{h,2}}{\partial w} = O(h_1^{-1} \ln^{-1} h); \frac{\partial u_{h,2}}{\partial w_2} = O(h_2^{-1})$ | Section 13 |
| $u_{w,2}$  | As for $u_{h,2}$ | Section 13 |
| $f_{h,2}$  | $f_{h,2} = O(\ln^{-1} h), \frac{\partial f_{h,2}}{\partial \varphi} = O(h_1^{-2} \ln^{-1} h), \frac{\partial f_{h,2}}{\partial w} = O(h_2^{-1} \ln^{-1} h)$ | Section 13 |
| $f_{w,2}$  | $f_{w,2} = O(\ln^{-1} h), \frac{\partial f_{w,2}}{\partial \varphi} = O(h_1^{-2} \ln^{-1} h), \frac{\partial f_{w,2}}{\partial w} = O(h_2^{-1} \ln^{-1} h)$ | Section 13 |
| $\varphi_2$ | $\varphi_2, f_{\varphi,2} = O(\h_1^{-2} \ln^{-2} h) + O_*(h_1^{-2} \ln^{-1} h)$ for $h > C_h \varepsilon$; $f_{\varphi,2} = O(h_2^{-1} \ln^{-2} h) + O_*(h_1^{-2} \ln^{-1} h)$ for $h > C_h \varepsilon$. | Section 13 |
| $f_{h,3}$  | $f_{h,3} = O(h_1^{-2} \ln^{-1} h) + O_*(h_2^{-1} \ln^{-2} h)$ for $h > C_h \varepsilon$. | Section 13 |
| $f_{w,3}$  | $f_{w,3} = O(h_1^{-2} \ln^{-1} h) + O_*(h_2^{-1} \ln^{-2} h)$ for $h > C_h \varepsilon$. | Section 13 |
| $t_{*,*}$  | The estimates for $f_{a,i}$ and its derivatives are as for $f_{a,i}$. | Section 13 |

Table 1: Estimates used in this paper. Here $g = O_*(h_1 \ln^{-3} h)$ means $g = O(h_1 \ln^{-3} h) e^{-\omega[|\hat{t}|](|\hat{t}| + 1)^\gamma}$ for some $\gamma$, where $t$ is one of the coordinates $t_i$ introduced in Section 11.1. At each point one or two coordinates $t_i$ are defined. If there are two, they are both $O(1)$, so we may chose any of them as $t$. Roughly speaking, $|\hat{t}|$ is the time needed for the trajectory of the unperturbed system to leave some neighborhood of C, while outside C we have $|\hat{t}| + 1 \sim 1$. The constant $C_h > 0$ depends only on $H$ and $f$, it will be defined in Section 13. Let us also note that some of the expressions above are vectors, for them their norm is estimated, i.e. $\frac{\partial T}{\partial w} = O(\ln^{-1} h)$ means $\| \frac{\partial T}{\partial w} \| = O(\ln^{-1} h)$. |
that transforms (2.1) to the following form:

\[
\begin{align*}
\tilde{T} &= \varepsilon \tilde{T}_{h,1}(h, w, \varepsilon) + \varepsilon^2 \tilde{T}_{h,2}(h, w, \varepsilon) + \varepsilon^3 \tilde{T}_{h,3}(h, w, \varepsilon), \\
\tilde{w} &= \varepsilon \tilde{T}_{h,1}(h, w, \varepsilon) + \varepsilon^2 \tilde{T}_{w,1}(h, w, \varepsilon) + \varepsilon^3 \tilde{T}_{w,2}(h, w, \varepsilon), \\
\tilde{v} &= \omega(h, w) + \varepsilon \tilde{T}_{\varphi,1}(h, w, \varepsilon) + \varepsilon^2 \tilde{T}_{\varphi,2}(h, w, \varepsilon).
\end{align*}
\]

(3.2)

Let us call the new chart \( h, w, \varphi \) the averaging chart. For brevity we will often omit the dependence of the functions \( f, \tilde{T}, \) and \( u_*, \) on \( \varepsilon. \)

It is convenient to denote by \( \nu \) the column vector \((h, w)\) and by \( \upsilon \) the column vector \((\tilde{h}, \tilde{w})\).

Let \( \mathcal{T}_{\nu,1} = \langle \mathcal{T}_{h,1}, \mathcal{T}_{w,1}, \rangle \), \( u_{\nu,1} = (u_{h,1}, u_{w,1}). \)

**Lemma 3.1.** For \( k = h, w, i = 1, 2 \) and for \( k = \varphi, i = 1 \) we have

\[
\mathcal{T}_{k,i}(h, w) = \langle Y_{k,i}(h, w, \varphi) \rangle_{\varphi},
\]

(3.3)

\[
\mathcal{T}_{k,i}(h, w) + \omega(h, w) \frac{\partial u_{k,i}}{\partial \varphi}(h, w, \varphi) = Y_{k,i}(h, w, \varphi)
\]

(3.4)

with

\[
\begin{align*}
Y_{h,1} &= f_h, \\
Y_{w,1} &= f_w, \\
Y_{\varphi,1} &= f_\varphi + \frac{\partial}{\partial \varphi} u_{w,1}. \\
Y_{h,2} &= \frac{\partial f_h}{\partial \upsilon} u_{v,1} + \frac{\partial f_h}{\partial \varphi} u_{\varphi,1} - \frac{\partial u_{h,1}}{\partial \upsilon} \mathcal{T}_{v,1} - \frac{\partial u_{h,1}}{\partial \varphi} \mathcal{T}_{\varphi,1}, \\
Y_{w,2} &= \frac{\partial f_w}{\partial \upsilon} u_{v,1} + \frac{\partial f_w}{\partial \varphi} u_{\varphi,1} - \frac{\partial u_{w,1}}{\partial \upsilon} \mathcal{T}_{v,1} - \frac{\partial u_{w,1}}{\partial \varphi} \mathcal{T}_{\varphi,1}. \\
\end{align*}
\]

(3.5)

The formulas for \( \mathcal{T}_{h,3}, \mathcal{T}_{w,3} \) and \( \mathcal{T}_{\varphi,2} \) are stated in Lemma 10.1 below.

We will prove this lemma in Section 10. The formulas above uniquely define \( \mathcal{T}_{h,i}, \) \( u_{\nu,i} \) under an additional assumption that for \( k = h, w; i = 1, 2 \) and for \( k = \varphi; i = 1 \) we have (in the formula below \( \langle * \rangle_\varphi \) denotes averaging with respect to \( \varphi \))

\[
\langle u_{k,i} \rangle_{\varphi} = 0.
\]

We will always assume this to hold.

For \( h \to 0 \) many expressions introduced above tend to infinity. We will use the estimates given in Table 1, these estimates will be proved below.

**Lemma 3.2.** There exists a constant \( C_{\text{inv}} > 0 \) depending on the perturbed system (1.2) such that for \( h > C_{\text{inv}} \varepsilon \) the coordinate change given by (3.1) is invertible.

**Proof.** Let us denote by \( F \) the map \( \langle \nu, \upsilon \rangle \to \langle v, \varphi \rangle \) given by (3.1). Let us consider the domain \( \tilde{h} > C \varepsilon, \) where the constant \( C > 0 \) is large. From Table 1 (note that as the values of \( u_{\nu,*} \) are taken at \( \langle \nu, \upsilon \rangle \), so we should plug \( h = \tilde{h} \) in the estimates in Table 1) and \( h^{-1} < C^{-1} \) we have \( \lvert h - \tilde{h} \rvert = O(\varepsilon), \) so for large enough \( C \) we have \( 0.5h < \tilde{h} < 2h. \) This means that \( h > C_{\text{inv}} \varepsilon \) implies \( \tilde{h} > C \varepsilon. \) By the inverse function theorem this implies that \( F \) is a local diffeomorphism. Moreover, for \( \tilde{h} > C \varepsilon \) we have \( \|F(x) - x\| = O(h \ln h) \). Indeed, \( \varepsilon u_{v,1} = O(\varepsilon^2 h \ln h) \) since \( \mathcal{T}_{\nu,1} \) is \( O(h \ln h) \). Hence, for some \( C \) this determinant lies in \([0.5, 2] \) for \( \tilde{h} > C \varepsilon. \) By the inverse function theorem this implies that \( F \) is a local diffeomorphism. Moreover, for \( \tilde{h} > C \varepsilon \) we have \( \|F(x) - x\| = O(h \ln h) \). Indeed, \( \varepsilon u_{v,1} = O(\varepsilon^2 h \ln h) \) since \( \mathcal{T}_{\nu,1} \) is \( O(h \ln h) \). Therefore, \( F \) is invertible as a local diffeomorphism that is \( C^0 \)-close to the identity.

Using that \( \langle \frac{\partial u_{k,i}}{\partial \varphi} \rangle_{\varphi} = 0, \langle \frac{\partial u_{k,i}}{\partial \nu} \rangle_{\varphi} = \frac{\partial u_{k,i}}{\partial \nu}(u_{k,i}) \varphi = 0, \) we can simplify (3.3) for \( \mathcal{T}_{h,2}, \mathcal{T}_{w,2} : \)

\[
\begin{align*}
\mathcal{T}_{h,2} &= \frac{\partial f_h}{\partial h} u_{h,1} + \frac{\partial f_h}{\partial w} u_{w,1} + \frac{\partial f_h}{\partial \varphi} u_{\varphi,1}, \\
\mathcal{T}_{w,2} &= \frac{\partial f_w}{\partial h} u_{h,1} + \frac{\partial f_w}{\partial w} u_{w,1} + \frac{\partial f_w}{\partial \varphi} u_{\varphi,1}. \\
\end{align*}
\]

(3.6)

The following formula is similar to Formula 2 from [7].
Lemma 3.3.

\[ u_{a,1}(h, w, t_0) = \frac{1}{T} \int_0^T \left( t - \frac{T}{2} \right) f_a(h, w, t + t_0) dt \quad \text{for } a = h, w_1, \ldots, w_k. \]  

(3.7)

Here the third argument in \( u_{a,1} \) and \( f_a \) is not \( \varphi \), as usual, but the time \( t = \varphi T/(2\pi) \). We use the notation \( f_a(h, w, t) = f_a(h, w, \varphi(h, w, t)) \) and a similar notation for \( u_{a,1} \).

This can also be rewritten as follows:

\[ u_{a,1}(h, w, t_0) = \frac{1}{2\pi} \int_0^T (\varphi(t) - \pi) f_a(h, w, t + t_0) dt \quad \text{for } a = h, w_1, \ldots, w_k. \]  

(3.8)

Proof of Lemma 3.3. The function \( u_{a,1} \) is uniquely determined by two properties. The first one is that \( \frac{\partial u_{a,1}}{\partial t} = f_a(t) - \langle f_a \rangle \) (this follows from (3.4), (3.5)). Denote by \( U \) the expression on the right hand side of (3.7). We have

\[ \frac{\partial U}{\partial t_0} = \frac{1}{T} \int_0^T \left( t - \frac{T}{2} \right) \frac{\partial f_a}{\partial t}(t + t_0) dt. \]

Integrating by parts, this can be rewritten as

\[ \frac{\partial U}{\partial t_0} = \frac{1}{T} \left[ \int_{t_0}^T f_a(t + t_0) \left( t - \frac{T}{2} \right) dt \right] - \frac{1}{T} \int_0^T f_a(t + t_0) dt = f_a(t_0) - \langle f_a \rangle. \]

Hence the first property of \( u_{a,1} \) holds for \( U \).

The second property is that \( \langle u_{a,1} \rangle_t = 0 \). This also holds for \( U \), it is checked by writing \( \int \frac{\partial U}{\partial t_0} dt_0 \) as a double integral and changing the order of integration.

\[ \square \]

4 Averaged system of order 2

The coefficients of the initial system (3.2) in the averaging chart depend on \( \epsilon \). We would like the coefficients of that system in the averaging chart to be independent of \( \epsilon \). To this end, let us introduce some notation. First, let us expand

\[ f(p, q, z, \epsilon) = f^0(p, q, z) + \epsilon f^1(p, q, z) + \epsilon^2 f^2(p, q, z, \epsilon), \]  

(4.1)

where \( f^0(p, q, z) = f(p, q, z, 0) \) and \( f^1(p, q, z) = \frac{\partial f^0}{\partial \epsilon}(p, q, z, 0) \). Clearly, \( f^0, f^0, f^0, f^0, f^1, f^1 \) are smooth functions of \( p, q \) and \( z \). The functions \( f^2, f^2, f^2, f^2 \) are smooth functions of \( p, q \) and \( z \) that depend on \( \epsilon \) and are uniformly bounded by some constant independent of \( \epsilon \) (by Taylor’s theorem with the Lagrange form of remainder). Let us also consider the perturbed system (2.1) with the perturbation \( \epsilon f^0(h, w, \varphi) \) instead of \( f(h, w, \varphi, \epsilon) \). For such system we may also consider a coordinate change of form (3.1) that transforms it to the form (3.2). Let us add an upper index 0 to the coefficients of these equations (e.g. \( \hat{u}^0_{h,1}, \tilde{f}^0_{\varphi,1} \)) to show that we started with the perturbation \( \epsilon f^0 \). The coefficients \( \hat{u}^0_{a,1} \) and \( \tilde{T}^0_{a,1} \) are determined by the same formulas as \( u_{a,1} \) and \( T_{a,1} \), but we should plug \( f^0 \) instead of \( f \) into those formulas.

Now let us rewrite (3.2) in such way that only the coefficients next to the largest powers of \( \epsilon \) depend on \( \epsilon \). This is done simply by expanding the coefficients similarly to (4.1). The resulting system will be

\[ \hat{h} = \hat{f}_{h,1}(\hat{h}, \omega) + \epsilon^2 \hat{f}_{h,2}(\hat{h}, \omega) + \epsilon^3 \hat{f}_{h,3}(\hat{h}, \omega), \]
\[ \hat{v} = \hat{f}_{w,1}(\hat{h}, \omega) + \epsilon^2 \hat{f}_{w,2}(\hat{h}, \omega) + \epsilon^3 \hat{f}_{w,3}(\hat{h}, \omega), \]
\[ \hat{\varphi} = \omega(\hat{h}, \hat{w}) + \epsilon \hat{f}_{\varphi,1}(\hat{h}) + \epsilon^2 \hat{f}_{\varphi,2}(\hat{h}, \omega, \varphi), \]  

(4.2)

where

\[ \hat{f}_{h,1} \] is the \( h \)-component of \( \hat{f} \) written in \( (h, w, \varphi) \) coordinates, \( \hat{f}_{w,1} \) and \( \hat{f}_{w,2} \) are the \( \hat{w} \)-components of \( \hat{f} \) written in \( (h, w, \varphi) \) coordinates, and \( \hat{f}_{\varphi,1} \) and \( \hat{f}_{\varphi,2} \) satisfy the estimates in Table 1. The estimates for \( \hat{f}_{h,1} \) will be proved in Lemma 13.3 below, one can also find formulas for \( \hat{f}_{\varphi,2}, \hat{f}_{h,3}, \hat{f}_{w,3} \) there. Also note that by [6, Corollary 3.1] we have \( \int \hat{f}_{h,2} dt = -\Theta_1(w) + O(h \ln h) \), so we have

\[ \hat{f}_{h,1} = \frac{-\Theta_1(w) + O(h \ln h)}{T}. \]  

(4.3)

The averaged system of order 2 is obtained from the system (4.2) by removing all terms on the right hand side that depend on \( \omega \):

\[ \hat{h} = \epsilon \hat{f}_{h,1}(\hat{h}, \omega) + \epsilon^2 \hat{f}_{h,2}(\hat{h}, \omega), \]
\[ \hat{\omega} = \epsilon \hat{f}_{w,1}(\hat{h}, \omega) + \epsilon^2 \hat{w}_{h,2}(\hat{h}, \omega), \]
\[ \hat{\varphi} = \omega(\hat{h}, \hat{w}) + \epsilon \omega_1(\hat{h}, \hat{w}). \]  

(4.4)
Here we denote $\omega_1(h) = \tilde{f}_{\varphi,1}(\hat{h}, \hat{w})$ in order to match with [10]. We will sometimes call this system simply the averaged system.

Let us introduce the slow time $\tau = \varepsilon t$. Then the first two equations in (4.5) can be written as follows:

$$\frac{d\hat{h}}{d\tau} = \hat{f}_{h,1}(\hat{h}, \hat{w}) + \varepsilon \hat{f}_{h,2}(\hat{h}, \hat{w}),$$

$$\frac{d\hat{w}}{d\tau} = \hat{f}_{w,1}(\hat{h}, \hat{w}) + \varepsilon \hat{f}_{w,2}(\hat{h}, \hat{w}).$$

(4.6)

By (4.4) and the estimate on $\hat{f}_{h,2}$ from Table 1 we get that

$$\frac{d\hat{h}}{d\tau} = -\Theta_1(\hat{w}) + O(h \ln h) + O(\varepsilon).$$

(4.7)

As $\Theta_i > 0$, this means that any solution $\hat{h}(\tau), \hat{w}(\tau), \hat{\phi}(\tau)$ of the averaged system of order 2 starting close to the separatrices crosses the separatrix of the initial unperturbed Hamiltonian equation. Denote by $\tau_*$ the slow time at the moment of crossing, $\hat{h}(\tau_*) = 0$. From (4.7) we also see that for small $\varepsilon$, $h$ and $\tau < \tau_*$ the function $\hat{h}(\tau)$ is decreasing. By (4.7) we also have that along the solution of the averaged system

$$\frac{d\tau}{d\hat{h}} = -\frac{T}{\Theta_1(\hat{w})}(1 + O(h \ln h) + O(\varepsilon)).$$

(4.8)

The lemma below estimates how the solutions of the averaged system of order 2 approximate the solutions of (3.2) while approaching the separatrices. It will be proved in Section 7.

**Lemma 4.1.** There exists $C_2 > 0$ such that the following holds. Consider a solution $\overline{\tau}(t)$ of (3.2), where $\overline{\tau}(t) = (\overline{\tau}(t), \overline{\tau}(t))$, with initial condition $\overline{\tau}(0), \overline{\tau}(0)$. Consider also a solution $\hat{\tau}(\tau) = (\hat{h}(\tau), \hat{w}(\tau))$ of (4.6) with initial condition $v(0)$ such that $\|\overline{\tau}(0) - \hat{\tau}(0)\| \leq C_1 \varepsilon^2$ for some $C_1 > 0$. Then for all small enough $\varepsilon$ for any $t$ such that

$$\hat{h}(\varepsilon t) > C_2 \varepsilon$$

(4.9)

we have the following estimates (in the error terms below we write $h$ for $\hat{h}(\varepsilon t)$, e.g. $O(h)$ instead of $O(h(\varepsilon t))$):

$$\|\overline{\tau}(t) - \hat{\tau}(t)\| = O(\varepsilon^2 h^{-1}),$$

$$\overline{\tau}(t) - \overline{\tau}(0) = \varepsilon^{-1} \int_0^{\varepsilon t} \left(\omega (\hat{\tau}(\tau')) + \varepsilon \omega_1 (\hat{\tau}(\tau'))\right) d\tau' + O(\varepsilon h^{-1} \ln^{-1} h).$$

**5 Formula for the pseudo-phase**

In this section we state the formula (5.1) for the pseudo-phase. This formula is similar to the one from [10], see also Section 1 for more references. This formula holds for any possible phase portrait and we may consider trajectories starting in any domain $G_i$, $i = 1, 2, 3$. Recall that we assume $h > 0$ in $G_i$. We need the condition $\Theta_i(\hat{w}) > 0$, this means that the trajectories starting in $G_i$ approach the separatrix/separatrices.

Consider a solution $v(t), \varphi(t)$ with $v(t) = (h(t), w(t))$ of the perturbed equation (2.1) that approaches the separatrix/separatrices. To define the pseudo-phase, it will be convenient to assume that if $i = 3$, we have $\Theta_1, \Theta_2 > 0$ in addition to $\Theta_3 > 0$. Let $h_{-1}$ be the value of $h(\tau)$ at the last crossing of the transversal $\varphi = 0$ with $h(\tau) > 0$. For $w_i$ defined below in this section (it will be close to the value of $w(t)$ at the moment of the separatrix crossing) denote $\Theta_i = \Theta_i(w_i)$. As $h$ decreases by approximately $\varepsilon \Theta_i$ during one turn, we have $0 \leq h_{-1} < \varepsilon \Theta_i + O(\varepsilon^{3/2})$ (this inequality follows from [6, Proposition 5.1] and (9.5), (9.6) below). We will assume $c_1 \varepsilon^{3/2} < h_{-1} < \varepsilon \Theta_i$ for $c_1 > 0$ chosen in Lemma 9.3 below. This holds for most initial conditions, what happens if this condition does not hold is discussed in Remark 5.2 below. The pseudo-phase is defined as $\frac{d\tau}{d\tilde{h}}$ (10).

Let the initial conditions for our solution be $v(0) = v_0, \varphi(0) = \varphi_0$. Set

$$\dot{v}_0 = v_0 - \varepsilon v_{0,1}(v_0, \varphi_0).$$

By (3.1) and Lemma 13.3 this approximates the value of $\varphi$ in the averaging chart corresponding to $v_0, \varphi_0$ with error $O(\varepsilon^2)$. Let $\hat{\tau}(\tau) = (\hat{h}(\tau), \hat{w}(\tau))$ be the solution of (4.6) with this initial condition. Let $\tau_*$ be the first time such that $\hat{h}(\tau_*) = 0$. In Section 4 we have shown that $\tau_*$ exists. Denote $w_{\tau_*} = \hat{w}(\tau_*)$. Denote $u_i = \frac{1}{2}(\Theta_1(w_{\tau_*}) - \Theta_2(w_{\tau_*}))$ for $i = 3$ (the separatrices should be enumerated as stated in Remark 5.1) and $u_0 = \frac{\varepsilon}{2}$ for $i = 1, 2$. Note that $u_* = \lim_{\tau \to \tau_*} u_{h_{-1}}(\hat{\tau}(\tau), 0)$ (for $i = 3$ this is by Lemma 9.1 and for $i = 1, 2$ this can be proved in the same way). Let us also recall the notation $\omega_1 = \tilde{f}_{\varphi,1} - \tilde{f}_{\varphi,1}$. Then we have

$$\frac{\hat{h}_{-1}}{\varepsilon \Theta_{3*}} = \left\{ \frac{1}{2\pi} \left( \varphi_0 + \frac{1}{2} \int_{\tau_0}^{\tau_*} (\omega (\hat{\tau}(\tau')) + \varepsilon \omega_1 (\hat{\tau}(\tau'))) d\tau' \right) + \frac{u_*}{\Theta_{3*}} + O(\varepsilon^{1/3} \ln^{1/3} \varepsilon) \right\}. \quad (5.1)$$

Here the curly brackets $\{ \}$ denote the fractional part.
Remark 5.1. To simplify the proof, we assume that \( \varphi = 0 \) corresponds to the choice of transversals \( \Gamma(z) \) defined in Section 11.1. However, one may easily check that (5.1) also holds if we take as \( \varphi = 0 \) any family of transversals tangent to the bisector of the angle between the segments of separatrices adjacent to the saddle. For \( i = 1, 2 \) only one such bisector lies in \( G \), and for \( i = 3 \) the choice of the bisector should match with the enumeration of the separatrices: \( \Theta_2 \) should correspond to \( 0 < \varphi < \pi \) and \( \Theta_1 \) to \( \pi < \varphi < 2\pi \).

Remark 5.2. Denote the values of \( \Theta \) on the right-hand side of (5.1) gives \( \{ h_{\Theta_{1,2}} \} \) and if \( h_{\Theta_1} > \varepsilon \Theta_1 \), it gives \( \{ h_{\Theta_{2,3}} \} \). We have also assumed earlier that \( \Theta_1, \Theta_2 > 0 \). If they have different signs with \( \Theta_3 = \Theta_1 + \Theta_2 > 0 \), the last transversal crossing can happen for \( h > \varepsilon \Theta_3 \). In this case we should find the first \( k \in \mathbb{N} \) such that \( h(t) > c_1 \varepsilon^{3/2} \) during all the time before the moment corresponding to \( h_{\Theta_3} \). Then the right-hand side of (5.1) gives \( \{ h_{\Theta_{2,3}} \} \).

6 Probabilities of capture

A trajectory starting in \( G_3 \) may be captured into \( G_1 \) or \( G_2 \) after separatrix crossing with the outcome determined by the pseudo-phase as stated in [6, Proposition 5.1]. Let us state a corollary of this proposition here.

Corollary 6.1. Any solution of the perturbed system with the pseudo-phase in \( [O(\varepsilon^{3/2}), \Theta_{1,3}] \) is captured in \( G_2 \) and with the pseudo-phase in \( [(\Theta_{1,2})_1 + O(\varepsilon^{3/2}), 1 - O(\varepsilon^{3/2})] \) is captured in \( G_1 \).

It turns out that the initial data for different outcomes are mixed, so it makes sense to consider captures in \( G_1 \) and \( G_2 \) as random events with some probabilities. One natural definition of the probability of capture is stated in [1]. This definition is based on the relative measures of the sets of points from a small neighborhood of the initial condition in \( G_3 \) that will be captured into \( G_1 \) and \( G_2 \) for \( \varepsilon \to 0 \). For this definition the probability of capture into \( G_j \) is close to \( \Theta_j \), this is proved in [6].

Let us sketch how formula (5.1) implies this formula for the probability of capture. Given initial data \( v_0, \varphi_0 \), let us fix some \( \varphi \) close to \( \varphi_0 \) and vary \( v \) near \( v_0 \). Denote by \( \Theta(v) \) the pseudo-phase of the solution of the perturbed system with the initial condition \( v, \varphi \). From (5.1) we have \( \frac{\Theta(v)}{\Theta_3} \sim \varepsilon^{-1} \). Using (5.1) and Corollary 6.1, it is possible to show that most points in the neighborhood of \( v_0 \) are covered by disjoint interchanging stripes of width \( O(\varepsilon) \) formed by values of \( v \) such that the trajectory is captured in \( G_1 \) and \( G_2 \), and the widths of the stripes captured into \( G_j \) are proportional to \( \Theta_j \). Then, naturally, the relative measure of the values of \( v \) captured into \( G_j \) is \( \frac{\Theta_j}{\Theta_3} \). Integrating this by \( \varphi \), we get the formula for the probability of capture.

Another way of defining the probability of capture was suggested by D.V. Anosov.

Denote by \( \mu \) the Lebesgue measure on \( \mathbb{R} \). Let us fix the initial data \( v_0, \varphi_0 \). Denote

\[ U_j(\varepsilon_0) = \{ \varepsilon \in (0, \varepsilon_0) : \text{the trajectory of } v_0, \varphi_0 \text{ is captured into } G_j \} \]

Then one can define the probability of capture into \( G_j \) as \( \lim_{\varepsilon_0 \to 0} \mu(U_j(\varepsilon_0))/\varepsilon_0 \). It was suggested in [6] that for this definition the probability of capture into \( G_j \) is also \( \frac{\Theta_j}{\Theta_3} \). Using (5.1), we can prove this statement.

Proposition 6.2. We have

\[ \frac{\mu(U_j(\varepsilon_0))}{\varepsilon_0} = \frac{\Theta_j}{\Theta_3} + O(\varepsilon^{1/3} \ln^{1/3} \varepsilon_0). \]

Proof. As \( \mu(U_j(\varepsilon_0)) + \mu(U_2(\varepsilon_0)) \leq \varepsilon_0 \), it is enough to show that

\[ \frac{\mu(U_j(\varepsilon_0))}{\varepsilon_0} > \frac{\Theta_j}{\Theta_3} + O(\varepsilon^{1/3} \ln^{1/3} \varepsilon_0). \]

Denote by \( \psi(\varepsilon) \) the right-hand side of (5.1) without the fractional part and the error term. Note that the integrals in (5.1) are computed along the solution of the averaged system of order 2 and so they depend on \( \varepsilon \). Denote \( a = (2\pi)^{-1} \int_{\tau=0}^{\tau=1} \omega(\bar{v}_1(\tau)) d\tau \), where \( \bar{v}_1(\tau) \) is the solution of the averaged system of order 1 (i.e. (4.6) with \( \varepsilon = 0 \)) with the initial condition \( \bar{v}_1(0) = \bar{v}_0(0) \). We may check that

\[ \psi = \varepsilon^{-1} a + O(1), \quad \frac{d\psi}{d\varepsilon} = -\varepsilon^{-2} a + O(\varepsilon^{-1}). \]

\(^1\text{This was a comment in a meeting of the Moscow Mathematical Society, this definition was first discussed in literature in [6].}\)
Hence, for small enough $\varepsilon_0$ we have $\psi \to \infty$ monotonically when $\varepsilon$ decreases from $\varepsilon_0$ to 0 and
\[
\frac{d\varepsilon}{d\psi} = -\psi^{-2} a + O(\psi^{-3}).
\] (6.2)

Without loss of generality we can take $j = 1$ in (6.1). By Corollary 6.1 and (5.1) for all $\varepsilon$ such that $\varepsilon \in [n + \frac{\varepsilon_0}{\varepsilon}, O(\varepsilon^{1/3} \ln^{1/3} \varepsilon)], n + 1 - O(\varepsilon^{1/3} \ln^{1/3} \varepsilon)], n \in \mathbb{N}$, the trajectory with the initial condition $v_n, \phi_0$ is captured in $G_1$. Using (6.2), we can estimate the length of the union of the preimages of such segments for the map $\varepsilon \to \psi$ as follows ($n_0 \sim \varepsilon_0^{-1}$ in the formula below):
\[
\mu(U_1(\varepsilon_0)) \geq \frac{a}{n_0 \varepsilon_0} n^{-2} \left( \frac{\Theta_1}{\varepsilon_0} + O(n^{-1/3} \ln^{1/3} n) \right) + O(\varepsilon_0^2).
\] On the other hand, we have
\[
\varepsilon_0 + O(\varepsilon_0^2) = a \sum_{n \geq n_0} (n^{-2} + O(n^{-3})),
\] as the union of the preimages of the segments $[n, n + 1], n \geq n_0$ is $(0, \varepsilon_0 + O(\varepsilon_0^2))$. These two formulas imply (6.1).

7 Proof of the approximation lemma

Starting with this section, we will assume that the phase portrait of the unperturbed system is a figure eight and we will only consider trajectories approaching the separatrices from the outside (i.e. from $G_3$). The proofs easily generalise to the other cases.

In this section we prove Lemma 4.1. As far from the separatrices solutions of the averaged system approximate solutions of the perturbed system in the averaged chart with accuracy $O(\varepsilon^2)$ for time intervals $O(\varepsilon^{-1})$, we may assume that $\Theta_0(0) > 0$ is small enough. Then $\hat{h}(\tau)$ will decrease monotonically. It will be convenient to use the notation $\Theta(\tau), \Theta(\tau), \Theta(\tau), \Theta(\tau) = \Theta(t), \Theta(t), \Theta(t), \Theta(t)$ with $\tau = \varepsilon^{-1} t$.

Let us start with the estimates for $\Theta(t) - \hat{h}(\tau)$ and $\Theta(t) - \hat{h}(\tau)$. We will first only consider what happens up to some moment $\tau_{fin}$ such that for all $\tau < \tau_{fin}$ we have
\[
0.5 \hat{h}(\tau) \leq \hat{h}(\tau) \leq 2 \hat{h}(\tau),
\] (7.1)

In order to receive a better estimate, let us switch from $h$ to the action $I$. Denote $\tilde{T} = I(\tilde{v})$, $\tilde{r} = I(\tilde{w})$. We have $f_{1, i} = \frac{\partial I}{\partial h} f_{1, i} + \frac{\partial I}{\partial w} f_{1, w,i}$ and $f_{r, i} = (f_{r, 1, i}, f_{r, w,i})$. We need the following lemma.

Lemma 7.1. We have
\[
\frac{\partial h}{\partial t} = \omega, \quad \left\| \frac{\partial h}{\partial w} \right\| = O(\ln h), \quad \left\| \frac{\partial h}{\partial r} \right\| = O(\ln^{-1} h),
\]
\[
\left\| \frac{\partial f_{1, i}}{\partial r} \right\| = O(h^{-1} \ln^{-3} h), \quad \left\| \frac{\partial f_{2, i}}{\partial r} \right\| = O(h^{-2} \ln^{-1} h), \quad \left\| f_{r, i} \right\| = O(h^{-2} \ln h) + O(h^{-2}).
\]

Proof of Lemma 7.1. From the Hamiltonian equations we have $\frac{\partial h}{\partial t} = \omega$. By [6, Corollary 3.2] we have
\[
\frac{\partial I}{\partial w} = O(1), \quad \left\| \frac{\partial^2 I}{\partial w \partial h} \right\| = O(\ln h), \quad \left\| \frac{\partial^2 I}{\partial h^2} \right\| = O(1).
\]
As $\frac{\partial I}{\partial w} = \omega^{-1}$, the first estimate implies $\left\| \frac{\delta I}{\delta w} \right\| = O(\ln h)$. We have $\frac{\partial I}{\partial w} (\frac{\partial I}{\partial h})_{t \text{const}} + \frac{\partial I}{\partial w} = 0$, this gives $\left\| \frac{\delta I}{\delta w} \right\| = O(\ln^{-1} h)$ and $\left\| \frac{\delta I}{\delta w} \right\| = O(\ln^{-1} h)$.

We have
\[
\frac{\partial f_{1, i}}{\partial t} = \frac{\partial f_{1, i}}{\partial h} f_{1, i} + \frac{\partial f_{1, i}}{\partial w} f_{1, w,i}.
\] (7.2)

For $i = 1$ this rewrites as
\[
\frac{\partial f_{1, i}}{\partial t} = (2\pi)^{-1} \int_{H=h} f_{1, i}^0 dt + \frac{\partial f_{1, i}}{\partial w} f_{1, w,i}.
\] (7.3)

By [6, Lemma 3.2] we have
\[
\int_{H=h} f_{1, i}^0 dt = O(1), \quad \frac{\partial}{\partial h} \int_{H=h} f_{1, i}^0 dt = O(\ln h), \quad \frac{\partial}{\partial w} \int_{H=h} f_{1, i}^0 dt = O(1).
\]

Plugging the first estimate in (7.3) gives $f_{1, i} = O(1)$. Plugging these estimates and the estimates in Table 1 in the derivatives of (7.3) gives
\[
\left\| \frac{\partial f_{1, i}}{\partial h} \right\| = O(h^{-1} \ln^{-2} h), \quad \left\| \frac{\partial f_{1, i}}{\partial w} \right\| = O(1).
\]
As \( \| \frac{\partial f_{1,3}}{\partial r} \| = O(\ln^{-1} h) \), we have \( \| \frac{\partial f_{1,3} \partial h}{\partial \phi} + \frac{\partial f_{1,3} \partial w}{\partial r} \| = O(h^{-1} \ln^{-3} h) \). We can prove that \( \| \frac{\partial f_{1,2}}{\partial h} \| = O(h^{-1} \ln^{-3} h) \) in the same way, so \( \| \frac{\partial f_{1,2}}{\partial \phi} \| = O(h^{-1} \ln^{-3} h) \).

By (7.2) for \( a = h, u, \) we have

\[
\frac{\partial f_{1,2}}{\partial \phi} = \frac{\partial I}{\partial \phi} + \frac{\partial I}{\partial \phi} \frac{\partial f_{1,2}}{\partial \phi} + \frac{\partial I}{\partial \phi} \frac{\partial w_{1,2}}{\partial \phi} \frac{\partial w_{1,2}}{\partial \phi}
\]

this yields \( \frac{\partial f_{1,2}}{\partial h} = O(h^{-2}), \) \( \| \frac{\partial f_{1,2}}{\partial h} \| = O(h^{-1}) \) by Table 1. As \( \| \frac{\partial f_{2}}{\partial h} \| = O(\ln^{-1} h) \), we have \( \| \frac{\partial f_{2}}{\partial h} \| = O(h^{-2} \ln^{-1} h) \).

Finally, the estimate on \( f_{1,3} \) follows from (7.2) and Table 1, while the estimate on \( f_{w,3} \) is given by Table 1.

As \( \vec{\tau}, \vec{\tau} \) is a solution of (3.2), it is also a solution of (4.2). Rewriting (4.2) and (4.5) using \( r \) instead of \( v \) gives

\[
\vec{\tau} = \vec{f}_{r,1}(\tau) + \epsilon^2 \vec{f}_{r,2}(\tau) + \epsilon^3 \vec{f}_{r,3}(\tau, \vec{\tau}, \epsilon),
\]

\[
\vec{v} = \vec{f}_{r,1}(\tau) + \epsilon^2 \vec{f}_{r,2}(\tau).
\]

(7.4)

Denote \( \Delta(\tau) = |[\vec{\tau}(\tau) - \hat{r}(\tau)]| \). From (7.4) we have the following differential inequality for \( \Delta \):

\[
\frac{d\Delta}{d\tau} \leq a(\tau) \Delta + \epsilon^2 b(\tau),
\]

where \( a(\tau) = \left( \| \frac{\partial f_{1,3}}{\partial r} \| + \epsilon(\| \frac{\partial f_{1,2}}{\partial h} \| \right) \) and \( b(\tau) = \left( \| f_{1,3}(\vec{\tau}(\tau), \vec{\tau}(\tau)) \| \right) \). Here the notation \( \left( \| \frac{\partial f_{1,3}}{\partial r} \| \right) \) means that each row of this matrix is taken at some intermediate point in \( [\vec{\tau}(\tau), \hat{r}(\tau)] \).

By (7.1), (4.9) and Lemma 7.1 we have \( a(\tau) = O(h^{-1} \ln^{-1} h) + \epsilon O(h^{-2} \ln^{-1} h) \). By Lemma 7.1 we have \( b(\tau) = O, (h^{-2} \ln h) + O(h^{-2}). \)

As in [10], we use the following estimate for \( \Delta \) obtained by solving (7.5):

\[
\Delta(\tau) \leq \exp \left( \int_0^\tau a(\tau') d\tau' \right) \left( \Delta(0) + \epsilon^2 \int_0^\tau b(\tau') d\tau' \right).
\]

Using (4.8) and the estimates for \( a \) and \( b \), we can make a change of variable and compute the integrals above as integrals \( dh \). We have

\[
\int_0^\tau a(\tau') d\tau' = O(1) + \epsilon O(h^{-1}) = O(1), \text{ where } h = \hat{h}(\tau).
\]

The integral of \( b \) can be estimated in the same way. As \( \int \hat{O}_s(h^{-2} \ln h) d\tau \) during each wind of the trajectory of the perturbed system around the figure eight is \( O(h^{-2} \ln h) \) and this wind takes time \( O(h^{-1}) \), we can replace this function with its average \( O(h^{-2}) \) if we also add the integral over the last incomplete wind:

\[
\int_0^\tau \hat{O}_s(h^{-2} \ln h) d\tau \sim \int_0^\tau \hat{O}(h^{-2}) d\tau + O(\epsilon h^{-2} \ln h), \text{ where in the last term } h = \hat{h}(\tau).
\]

Hence,

\[
\int_0^\tau b(\tau') d\tau' = \int_0^\tau \hat{O}(h^{-2}) d\tau' + O(\epsilon h^{-2} \ln h) = O(h^{-1} \ln h).
\]

Note that \( O(\epsilon h^{-2} \ln h) \) is \( O(h^{-1} \ln h) \) as \( \epsilon h^{-1} < C_{9}^{-1} \).

As \( \Delta(0) = O(\epsilon^2) \), this gives the estimate \( \Delta(\tau) = O(\epsilon^2 h^{-1} \ln h) \). As \( \| \frac{\partial v}{\partial \phi} \| = O(\ln^{-1} h) \) (here \( h \sim \hat{h}(\tau) \) by (7.1)), we have \( |[\vec{\tau}(\tau) - \hat{r}(\tau)]| = O(\epsilon^2 h^{-1}). \)

(7.6)

From the estimate on \( \vec{\tau}(\tau) - \hat{r}(\tau) \) we have just proved and (4.9) we get that \( \vec{\tau}(\tau) - \hat{r}(\tau) = C_{9}^{-1} O(\epsilon) < C_{9}^{-2} O(\hat{h}(\tau)) < 0.5 \hat{h}(\tau) \) for large enough \( C_{9} \), so the condition (7.1) actually holds for all \( t \) considered in this lemma.

The estimate for the difference in \( w \) in (7.6) can be improved. By (7.6) and Table 1 for \( h = \hat{h}(\tau), \hat{v} = \hat{v}(\tau) \) and \( \vec{\tau} = \vec{\tau}(\tau) \) we have

\[
\| \hat{f}_{w,1}(\vec{\tau}) + \epsilon \hat{f}_{w,2}(\vec{\tau}) + \epsilon^2 \hat{f}_{w,3}(\vec{\tau}, \vec{\tau}(\tau), \epsilon) - \hat{f}_{w,1}(\hat{v}) - \epsilon \hat{f}_{w,2}(\hat{e}) \| = O(\epsilon^2 h^{-2} \ln^{-1} h) + O_s(\epsilon^2 h^{-2}).
\]

Arguing as above, we can estimate the integral of this expression \( d\tau \):

\[
|\hat{\vec{\tau}}(\tau) - \vec{\tau}(\tau)| = O(\epsilon^2 h^{-1}).
\]

Let us now prove the estimate for \( \phi \). Denote \( \omega_0(v) = \omega(v) + \epsilon \omega_1(v) \). Then from (4.2) we have

\[
|\vec{\tau}(\tau) - \vec{\tau}(0)| = \epsilon^{-1} \int_0^\tau \left( \omega_0(\vec{\tau}(\tau')) + \epsilon^2 \hat{f}_{w,3}(\vec{\tau}(\tau'), \vec{\tau}(\tau)) \right) d\tau'.
\]

10
From Table 1 and (4.9) we have \( \frac{\partial \omega_1}{\partial h} = O(h^{-1} \ln^{-2} h) \). We also have \( \left\| \frac{\partial \omega_1}{\partial w} \right\| = O(\ln h) \). Thus from (7.6) we have \( |\omega_1(\tau) - \omega_1(\hat{\tau}(\tau))| = O(\epsilon^2 h^{-1} \ln^{-2} h) \). From Table 1 we have \( f_{\sigma,2} = O(h^{-1} \ln^{-2} h) + O(\epsilon h^{-1} \ln^{-2} h) \). So

\[
\varphi(\tau) - \varphi(0) = \epsilon^{-1} \int_{0}^{\tau} \omega_1(\dot{\hat{\tau}}(\tau'))d\tau' + \epsilon \int_{0}^{\tau} O(h^{-2}(\tau') \ln^{-2} \hat{\hat{\tau}}(\tau')) + C \int_{0}^{\tau} \hat{\hat{\tau}}(\tau') \ln^{-1} \hat{\hat{\tau}}(\tau')d\tau'.
\]

The second integral can be estimated in the same way as \( \int_{0}^{\tau} b(\tau')d\tau' \) above:

\[
\epsilon \int_{0}^{\tau} O(h^{-2}(\tau') \ln^{-2} \hat{\hat{\tau}}(\tau')) + C \int_{0}^{\tau} \hat{\hat{\tau}}(\tau') \ln^{-1} \hat{\hat{\tau}}(\tau')d\tau' = O(\epsilon h^{-1} \ln^{-1} h).
\]

This proves the formula for \( \varphi \).

8 Cancellation lemma

In this section we prove the following lemma. It will be useful when we prove the formula for the pseudo-phase, because due to this lemma two terms will cancel out. Denote \( \omega_1(h, w) = \tilde{f}_{\varphi, 1} = \int_{\varphi} f_1 \) to match the notation in [10].

**Lemma 8.1.** Consider a solution \( \hat{h}(\tau), \hat{\varphi}(\tau) \) of the averaged system (4.5). Take \( \tau_1 < \tau_2 < \tau_* \) such that \( \hat{h}(\tau_1) \) is small enough. Denote \( h_1 = \hat{h}(\tau_1) \), \( h_2 = \hat{h}(\tau_2) \), \( w = \hat{\varphi}(\tau_1), \Theta_3 = \Theta_3(w) \).

Then

\[
\int_{\tau_1}^{\tau_2} \omega_1(\hat{h}(\tau), \hat{\varphi}(\tau))d\tau = -\frac{2\pi}{\Theta_3(w)} \int_{\tau_1}^{\tau_2} \hat{w}_{h,1}(\hat{h}(\tau), \hat{\varphi}(\tau), 0) + O(\epsilon h \ln h). \tag{8.1}
\]

Let us first estimate \( \omega_1 \). Denote \( \mathcal{I}(h, w) = \int_{0}^{2\pi} t(\varphi)f_{w}^{0}(\varphi)d\varphi \).

**Lemma 8.2.**

\[
\omega_1 = \frac{1}{T} \frac{\partial \mathcal{I}}{\partial h} + O(h^{-1/2} \ln^{-1} h). \tag{8.2}
\]

**Proof.** Integrating by parts, we can write

\[
2\pi \omega_1 = \int_{0}^{2\pi} f_{w}^{0} d\varphi = \int_{0}^{2\pi} \varphi f_{w}^{0} - \int_{0}^{2\pi} \varphi \frac{\partial f_{w}^{0}}{\partial \varphi} d\varphi.
\]

Using (2.2) and the equality \( \frac{1}{T} \frac{\partial}{\partial h}(Tf_{w}^{0}) = \frac{\partial f_{w}^{0}}{T \partial h} + \frac{1}{T} \frac{\partial}{\partial \varphi}(Tf_{w}^{0}) \), this rewrites as

\[
2\pi \omega_1 = 2\pi f_{w}^{0}(h, w, 0) - \int_{0}^{2\pi} \varphi \text{div}(f_{w}^{0}) d\varphi + \sum_{i=1}^{k} \int_{0}^{2\pi} \varphi \frac{\partial f_{w_i}^{0}}{\partial w_i} d\varphi + \int_{0}^{2\pi} \varphi \frac{1}{T} \frac{\partial}{\partial h} \mathcal{I}(h, w) d\varphi.
\]

By Table 1 the first term is \( O(h^{-1/2} \ln^{-1} h) \). The second term is \( O(1) \) as \( \text{div}(f_{w}^{0}) \) is bounded. The third term is \( O(1) \) by Table 1. As \( \frac{\partial}{\partial h} \) commutes with integrating by \( \varphi \), we can rewrite the last term as \( \frac{1}{T} \frac{\partial}{\partial h} \int_{0}^{2\pi} \varphi T f_{w_i}^{0} d\varphi = \frac{1}{T} \frac{\partial}{\partial h} \mathcal{I}(h, w) \). We have obtained (8.2).

**Lemma 8.3.**

\[
\frac{\partial}{\partial w_i} \left( \int_{0}^{2\pi} t(\varphi)f_{w_i}^{0} d\varphi \right) = O(1), \quad i = 1, \ldots, k.
\]

**Proof.** As \( t = (2\pi)^{-1} T \varphi \), we have

\[
\frac{\partial}{\partial w_i} \left( \int_{0}^{2\pi} t(\varphi)f_{w_i}^{0} d\varphi \right) = (2\pi)^{-1} \frac{\partial \mathcal{I}}{\partial w_i} \int_{0}^{2\pi} \varphi f_{w_i}^{0} d\varphi + (2\pi)^{-1} T \int_{0}^{2\pi} \varphi \frac{\partial f_{w_i}^{0}}{\partial w_i} d\varphi =
\]

\[
= T^{-1} \frac{\partial \mathcal{I}}{\partial w} \int_{0}^{2\pi} \varphi \Theta_3(1) dt + \int_{0}^{2\pi} \varphi \Theta_3(1) dt = O(1).
\]

**Proof of Lemma 8.1.** For small enough \( h_1 \), the value of \( \hat{h}(\tau) \) decreases, so we may use \( h = \hat{h}(\tau) \) as a coordinate along the solution of the averaged system. We will also take \( \frac{dT}{dh}, \frac{dw}{dh} \) and \( \frac{d\varphi}{dh} \) along this solution. For convenience let us recall (4.8) here:

\[
\frac{dT}{dh} = -\frac{T}{\Theta_3(\varphi)}(1 + O(h \ln h) + O(\epsilon)) = O(\ln h).
\]

By Lemma 8.3 we have \( \frac{\partial \varphi}{\partial h} \| = O(1) \). We can write

\[
\| \frac{\partial \varphi}{\partial h} \| = \left\| \frac{dT}{dh}(f_{w,1} + \epsilon f_{w,2}) \right\| = O(\ln h) + \epsilon O(h^{-1} \ln^{-2} h),
\]

\[
\frac{dT}{dh} = \frac{dT}{dh} + \frac{dT}{dw} \frac{dw}{dh} = O(\ln h) + \epsilon O(h^{-1} \ln^{-2} h).
\]
As \(\omega_1 = O(h^{-1} \ln^{-3} h)\) and so \(\int_{h_1}^{h_2} |T \omega_1| dh = O(h_1)\), we have
\[
\int_{\tau_1}^{\tau_2} \omega_1(\hat{h}(\tau), \hat{w}(\tau)) d\tau = -\int_{h_1}^{h_2} \frac{T \omega_1}{\Theta_3(\hat{w})} dh + O(h_1) + O(\varepsilon \ln^{-1} h_2).
\]
Integrating the estimate for \(\|\frac{dT}{dh}\|\), we get
\[
\|\hat{w} - w_*\| = O(h_1 \ln h_1) + O(\varepsilon \ln^{-1} h_1).
\] (8.3)
As by [6, Lemma 3.2]
\[
\frac{\partial \Theta_3}{\partial w} = O(1),
\] (8.4)
this means
\[
|\Theta_3(\hat{w}) - \Theta_3(w_*)| = O(h_1 \ln h_1) + O(\varepsilon \ln^{-1} h_1)
\] (8.5)
and
\[
\int_{\tau_1}^{\tau_2} \omega_1(\hat{h}(\tau), \hat{w}(\tau)) d\tau = -\frac{1}{\Theta_3} \int_{h_1}^{h_2} T \omega_1 dh + O(h_1) + O(\varepsilon \ln^{-1} h_1).
\]
By Lemma 8.2 this can be rewritten as
\[
\int_{\tau_1}^{\tau_2} \omega_1 d\tau = -\frac{1}{\Theta_3} \int_{h_1}^{h_2} \frac{T \Theta}{dh} dh + O(h_1^{1/2}) + O(\varepsilon \ln^{-1} h_1) =
\]
\[
= -\frac{1}{\Theta_3} \int_{h_1}^{h_2} \frac{dT}{dh} dh + O(h_1^{1/2}) + O(\varepsilon \ln^{-1} h_1) =
\]
\[
= -\frac{1}{\Theta_3} \int_{\tau_1}^{\tau_2} T(\hat{h}(\tau), \hat{w}(\tau)) + O(h_1^{1/2}) + O(\varepsilon \ln^{-1} h_1).
\] (8.6)
As \(dt = \frac{T dh}{\dot{\Theta}_3},\) by (3.7) we have
\[
u_{h,1}^0(h, w, 0) = \frac{1}{T} \int_0^T \left( t - \frac{T}{2} \right) f^0_h(t) dt = \frac{1}{2\pi} \int_0^{2\pi} t f^0_h(t) d\varphi - \frac{1}{2} \int_0^T f^0_h dt.
\]
By [6, Corollary 3.1] \(\int_0^T f^0_h(t) dt = -\Theta_3(w) + O(h \ln h)\). Hence,
\[
\int_{\tau_1}^{\tau_2} \nu_{h,1}^0(\hat{h}(\tau), \hat{w}(\tau), 0) = \frac{1}{2\pi} \int_{\tau_1}^{\tau_2} T(\hat{h}(\tau), \hat{w}(\tau)) + O(h_1 \ln h_1) + O(\varepsilon \ln^{-1} h_1).
\]
Comparing this with (8.6), we get (8.1).

9 Proof of the formula for the pseudo-phase

In this section we prove the formula (5.1) for the pseudo-phase. We use the notation from Section 5. First let us prove some auxiliary statements.

**Lemma 9.1.** We have \(\lim_{\tau_1 \to \tau_2, \varepsilon \to 0} \nu_{1,1}^0(\hat{w}(\tau), 0) = \frac{1}{2} (\Theta_1(w_*) - \Theta_2(w_*)) = u_*\).

**Proof.** Recall that \(\Theta_2\) corresponds to \(0 < \varphi < \pi\) and \(\Theta_3\) to \(\pi < \varphi < 2\pi\). For \(\tau \to \tau_1, \varepsilon = 0\) we have \(\hat{h}(\tau) \to 0\). Let us split the integral expression (3.8) (with \(f\) replaced by \(f^0_h\)) for \(u_{h,1}^0(\hat{w}(\tau), 0)\) into the integrals over the part of the trajectory near \(l_1\) and near \(l_2\). For the first part the value of \(\varphi(t) - \pi\) is close to \(\pi/2\) far away from the saddle \(C\). But close to \(C\) we have \(f^0_h \approx 0\), so the integral near \(l_1\) is close to \(\Theta_1/4\). Similarly, the integral near \(l_2\) is close to \(-\Theta_2/4\).

**Lemma 9.2.** Take \(\tau_1 < \tau_2,\) denote \(h_1 = \hat{h}(\tau_1)\). Then we have
\[
\int_{\tau_1}^{\tau_2} \omega(\hat{w}(\tau)) d\tau = 2\pi \Theta_3 h_1 + O(\varepsilon h_1) + O(h_1^2 \ln h_1).
\] (9.1)

**Proof.** As \(\omega T = 2\pi, (4.8)\) and (8.5) implies that
\[
\int_{\tau_1}^{\tau_2} \omega(\hat{h}(\tau)) d\tau = -2\pi \Theta_3^{-1} \int_{h_1}^{h_2} \left( 1 + O(h \ln h) + O(\varepsilon) \right) dh = 2\pi \Theta_3 \int_{h_1}^{h_2} \left( 1 + O(h \ln h) + O(\varepsilon) \right) dh,
\]
which gives the required estimate.

**Lemma 9.3.** Assume \(\Theta_1, \Theta_2 > 0\). Then there exist \(c_1, c_2 > 0\) such that for all small enough \(\varepsilon\) the following holds. Take a point \((h_0, w_0, 0)\) on the transversal \(\varphi = 0\) with \(\varepsilon \Theta_3(w_0) + 2c_1 \varepsilon^{3/2} \leq h_0 < c_2\). Then the orbit of this point intersects the transversal \(\varphi = 0\) once more with
\[
h = h_0 - \varepsilon \Theta_3(w_0) + O(\varepsilon h_0 \ln h_0) + O(\varepsilon^2 h_0^{-1/2}) > c_1 \varepsilon^{3/2}
\]
and the time passed between these two intersections is \(O(\ln h_0)\).
Proof. By [6, Lemma 3.5, Corollary 3.4] there are $c_2 > 0$ and $c_3 > 0$ such that for $c_2 \varepsilon \leq h_0 \leq c_2 \varepsilon$ the orbit crosses the transversal again after the time $O(\ln h)$ passes and we have $h = h_0 - \varepsilon \int_{h_0}^{h} f_0 dt + O(\varepsilon^2 h_0^{-1/2})$. By (4.1) we have $f_h - f_0 = \varepsilon \psi(p, q, z, \varepsilon)$ for some smooth $\psi$ with $\psi(C) = 0$. By [6, Lemma 3.2] $\int_{h_0}^{h_{\varepsilon}} \psi dt = O(1)$, so $\int_{h_0}^{h} f_0 dt = \int_{h_0}^{h_{\varepsilon}} f_0 dt + O(\varepsilon) = -\Theta_3(u_0) + O(h_{\ln h_0}) + O(\varepsilon)$ by (4.4). As $\varepsilon = O(h)$, this gives the required estimate.

By [6, Proposition 5.1] there is $c_4 > 0$ such that for $\varepsilon \Theta_3 + c_4 \varepsilon^{3/2} \leq h_0 < c_2 \varepsilon$ the orbit of our point intersects the transversal $\varphi = 0$ once more (the condition $\Theta_1, \Theta_2 > 0$ is used here). Moreover, arguing as in the proof of [6, Proposition 5.1], we can get that for this new intersection $h = h_0 - \varepsilon \Theta_3(u_0) + O(\varepsilon^{3/2})$ and the time between these two intersections is $O(\ln \varepsilon)$. As $\Theta_3 \varepsilon \leq h_0 < c_2 \varepsilon$, these estimates are equivalent to the estimates claimed in the lemma. □

Recall that $h_{-2}, h_{-3}, \ldots$ denote the values of $h$ at the consecutive crossings of the transversal $\varphi = 0$ before $h_{-1}$.

**Lemma 9.4.** For $\varepsilon^{9.9} < h_{-n} < \varepsilon^{0.1}$ we have

$$h_{-n} = h_{-1} + \varepsilon(n - 1)\Theta_3(w_{-n}) + O(h_{-n}^2 \ln h_{-n}) + O(h_{-n}^{1/2}).$$

(9.2)

Proof. First, let us note that as the considered $h_{-n}$ are in $[c_2 \varepsilon^{3/2}, \varepsilon^{0.1}]$, we have $\ln h \sim \ln \varepsilon \sim \ln h_{-n}$. So the time passed between two consecutive intersections is $O(\ln h_{-n})$. As $n \sim \varepsilon^{-1} h_{-n}$, the total time between the moments corresponding to $h_{-n}$ and $h_{-1}$ is $O(\varepsilon^{-1} h_{-n} \ln h_{-n})$. As $w = O(\varepsilon)$, we have $\|w - w_{-n}\| = O(h_{-n} \ln h_{-n})$ for all encountered values of $w$ and by (8.4) we have $\Theta_3(w) - \Theta_3(w_{-n}) = O(h_{-n} \ln h_{-n})$. Now the required estimate follows from Lemma 9.3 by summation. □

Let us return to the proof of the formula for pseudo-phase. Denote by $v(\tau), \varphi(\tau) = v(t), \varphi(t)$, where $t = \varepsilon^{-1} \tau$ and $v(\tau) = (h(\tau), w(\tau))$, the solution of the perturbed system written using the slow time $\tau$. We denote by $\overline{\tau}(\tau), \overline{\varphi}(\tau)$, where $\overline{\tau}(\tau) = \left(\overline{h}(\tau), \overline{w}(\tau)\right)$, the solution $v(\tau), \varphi(\tau)$ of the perturbed system, written in the averaged chart (3.1). Denote $\overline{v}_0 = (\overline{h}_0, \overline{w}_0) = (\overline{h}(0), \overline{w}(0))$. We have $\|\overline{v}_0 - \overline{v}_0\| = O(\varepsilon^2)$, so we may use Lemma 4.1.

**Lemma 9.5.** There exist $C_3 > 0$ such that for all $\tau$ with

$$\hat{h}(\tau) > C_3 \varepsilon$$

(9.3)

the solutions $v(\tau), \overline{\tau}(\tau)$ and $\overline{v}(\tau)$ are close:

$$\|\overline{v} - \overline{\tau}\| = O(\varepsilon^2 \hat{h}^{-1}) < \hat{h}/4, \quad \|\tau - v\| = O(\varepsilon) < \hat{h}/4.$$  

(9.4)

Proof. The first two estimates are given by Lemma 4.1. To obtain the last one, we just plug the estimates from Table 1 into the equation $v = \overline{\tau} + \varepsilon v_{+1} + \varepsilon^2 v_{+2}$ from (3.1). □

Now we are ready to prove (5.1). Consider a moment $\tau_1$ such that $\varphi(\tau_1) = 0$ and $\hat{h}(\tau_1)$ is as close as possible to $\varepsilon^{2/3} \ln^{-1/3} \varepsilon$. Note that we have (9.3) for $\tau = \tau_1$. We may check that under the condition (9.3) the difference between $\hat{h}(\tau)$ for consecutive times $\tau$ with $\varphi(\tau) = 0$ is $O(\varepsilon)$. Indeed, the time between consecutive fast times of crossing the transversal $\varphi = 0$ is $O(T)$ and $\hat{h}$ is $O(T^{-1})$. Hence,

$$\hat{h}(\tau_1) = (1 + o(1))\varepsilon^{2/3} \ln^{-1/3} \varepsilon.$$  

(9.5)

Denote by $h_1, \varphi_1, \overline{h}_1, \overline{\varphi}_1, \hat{h}_1, \hat{\varphi}_1$ the values of $h, \varphi, \overline{h}, \overline{\varphi}, \hat{h}, \hat{\varphi}$ at the slow time $\tau_1$. As justified by (9.4), we may write $h_1$ instead of $h$ and $\overline{h}_1$ in the error terms. For brevity let us even denote $h = h_1$ for the error terms and write simply $O(h)$.

**Lemma 9.6.** For any $\tau \geq \tau_1$ until separatrices crossing (i.e. with $h(\tau) > 0$) we have

$$w(\tau) = w_* + O(h_1 \ln h_1), \quad \Theta_3(w(\tau)) = \Theta_3 + O(h_1 \ln h_1).$$  

(9.6)

Proof. By (8.3) we have $w(\tau_1) = w_* + O(h_1 \ln h_1) + O(\varepsilon \ln h_1)$. By (9.4) we have $w(\tau_1) = w(\tau_1) + O(\varepsilon^2 \hat{h}_1^{-1})$. Arguing as in the proof of Lemma 9.4 gives $w(\tau) = w(\tau_1) + O(h_1 \ln h_1)$. Combining these estimates gives the first statement (the term $O(h_1 \ln h_1)$ absorbs other terms for $h_1 \gtrsim \varepsilon$); the second statement follows from the first by (8.4). □

Let us split the integral in (5.1) into integrals from 0 to $\tau_1$ and from $\tau_1$ to $\tau_2$. First, let us check that

$$\varphi_0 + \frac{1}{\varepsilon} \int_{\tau=0}^{\tau_1} \left( \omega(\hat{\varphi}(\tau)) + \varepsilon \omega_1(\hat{\varphi}(\tau)) \right) d\tau = 2\pi m + O(\varepsilon h^{-1} \ln^{-1} h),$$  

(9.7)

where $m \in \mathbb{Z}$. By Lemma 4.1 we have

$$\frac{1}{\varepsilon} \int_{\tau=0}^{\tau_1} \left( \omega(\hat{\varphi}(\tau)) + \varepsilon \omega_1(\hat{\varphi}(\tau)) \right) d\tau = \overline{\varphi}_1 - \overline{\varphi}_0 + O(\varepsilon h^{-1} \ln^{-1} h).$$
We also have \( \varphi = \overline{\varphi} + \varepsilon u_{x,1} \). By Table 1 \( u_{x,1} = O(h^{-1} \ln^{-1} h) \), so \( \overline{\varphi}_1 - \overline{\varphi}_0 = \varphi_1 - \varphi_0 + O(\varepsilon h^{-1} \ln^{-1} h) \). As \( \varphi_1 = 2\pi n \), this gives the required equality (9.7).

Now let us use (9.1) and (8.1) (in (8.1) we pass to the limit for \( \tau_2 \to \tau_1 \to 0 \), by Lemma 9.1 we have \( u_{y,1}(\hat{v}(\tau_2),0) \to u_0 \) to compute the remaining terms in (5.1). We have

\[
\frac{1}{2\pi \varepsilon} \left( \int_{\tau = \tau_1}^{\tau} \left( \omega(\hat{v}(\tau)) + \varepsilon \omega_1(\hat{v}(\tau)) \right) d\tau \right) + \frac{u_0}{\Theta_{\Theta_{3\varepsilon}}} = \frac{1}{\varepsilon \Theta_{3\varepsilon}} \left( \hat{h}_1 + \varepsilon u_{y,1}(\hat{v}_1,0) \right) + O(h^{1/2} + O(\varepsilon^{-1} h \ln h)).
\]

(9.8)

Note that the term \( O(\varepsilon^{-1} h) \) from (8.1) is absorbed into \( O(h^{1/2}) \) by (9.3). As \( h_1 \approx \varepsilon \), by Table 1 we have \( \frac{\partial}{\partial \varepsilon} u_{y,1}(\varepsilon) = O(h^{-1} \ln h) \). Hence, by (9.4) we have

\[
\hat{h}_1 + \varepsilon u_{y,1}(\varepsilon) = \hat{h}_1 + \varepsilon u_{y,1}(\varepsilon) + O(h^{-1} \ln h) + O(\varepsilon^{-1} h \ln h).
\]

The last equality is justified by Lemma 13.3. The error term \( O(\varepsilon^2 h^{-1}) \) appears, but it is absorbed into \( O(\varepsilon^2 h^{-1}) \). As \( \varepsilon_1 = \hat{\varphi}_1 + \varepsilon u_{x,1} \), by Table 1 we have \( \hat{\varphi}_1 = O(h^{-1} \ln^{-1} h) \). Hence, by the estimate \( \frac{\partial}{\partial \varepsilon} u_{y,1}(\varepsilon) = O(\ln h) \) from Table 1 we get

\[
\varepsilon u_{y,1}(\varepsilon) = \varepsilon u_{y,1}(\varepsilon) + O(\varepsilon^{-1} h \ln h)
\]

and

\[
\hat{h}_1 + \varepsilon u_{y,1}(\varepsilon) = \hat{h}_1 + \varepsilon u_{y,1}(\varepsilon) + O(\varepsilon^{-1} h \ln h)
\]

As by (3.1)

\[
\hat{h}_1 = \hat{h}_1 + \varepsilon u_{y,1}(\hat{\varphi}_1,\varepsilon) + \varepsilon^2 u_{x,2}(\hat{\varphi}_1,\varepsilon)
\]

the estimate \( \varepsilon^2 u_{x,2}(\hat{\varphi}_1,\varepsilon) = O(\varepsilon^2 h^{-1}) \) from Table 1 yields

\[
\hat{h}_1 + \varepsilon u_{y,1}(\varepsilon) = h_1 + O(\varepsilon^{-1} h \ln h)
\]

Combining this with (9.8), we get

\[
\frac{1}{2\pi \varepsilon} \left( \int_{\tau = \tau_1}^{\tau} \left( \omega(\hat{v}(\tau)) + \varepsilon \omega_1(\hat{v}(\tau)) \right) d\tau \right) + \frac{u_0}{\Theta_{3\varepsilon}} = \frac{h_1}{\varepsilon \Theta_{3\varepsilon}} - R(h_1)
\]

with the error term

\[
R = O(h^{1/2} + O(\varepsilon h^{-1} \ln h + O(\varepsilon^{-1} h^2 \ln h)).
\]

After taking a sum with (9.7), we get

\[
\frac{h_1}{\Theta_{3\varepsilon}} = \frac{1}{2\pi} \left( \varphi_0 + \frac{1}{\varepsilon} \int_{\tau = \tau_0}^{\tau} \left( \omega(\hat{v}(\tau)) + \varepsilon \omega_1(\hat{v}(\tau)) \right) d\tau \right) + \frac{u_0}{\Theta_{3\varepsilon}} - m + R(h_1).
\]

Note that \( R \) absorbs the error term in (9.7). Let us now apply (9.2) for \( v_{-n} = v_1 \). We have \( n \approx \varepsilon^{-1} h_1 \); by (9.6) we have \( (n-1) \Theta_{3\varepsilon} = (n-1) \Theta_{3\varepsilon} + O(\varepsilon^{-1} h_1^2 \ln h) \), so this yields the required formula (5.1), but with the error term \( R(h_1) \) depending on \( h_1 \). Note that the error term above and the error term in (9.2) divided by \( \varepsilon \) are not greater than \( R \). Then we just plug in the expression (9.5) for \( h_1 \) and obtain \( R = O(\varepsilon^{-1/2} \ln^{-1/3} \varepsilon) \). One may check that (9.5) minimizes the error term. Indeed, first we check that up to some power of \( \ln \varepsilon \) the value of \( R \) is minimal for \( h \approx \varepsilon^{3/3} \). Then \( \ln h \approx (2/3) \ln \varepsilon \), and from this we see that \( R \) is minimal for \( h \) given by (9.5). This completes the proof of formula (5.1).

10 Formulas for the averaging chart

In this section we present formulas for \( \tilde{T}_{x,2} \) and \( \tilde{T}_{x,3} \) from Lemma 3.1 and prove this lemma. We use the notation introduced in Section 3. We will also need the following notation.

- Denote by \( x \) the column vector \((h, w, \varphi)\) and by \( \overline{x} \) the column vector \((\overline{h}, \overline{w}, \overline{\varphi})\). Let \( f_0 = \left(f_0, f_w, f_{x,1} = (f_{x,1}, f_{x,2}, f_{x,3})\right) \), \( u_{x,1} = (u_{x,1}, u_{w,1}, u_{\varphi,1}) \).
- Given \( k = x, v, h, \varphi \), let us denote \( u_{x,2} = u_{x,1} + \varepsilon u_{x,2} \), \( \tilde{T}_{x,1,2} = \tilde{T}_{x,1} + \varepsilon \tilde{T}_{x,2} \), \( \tilde{T}_{x,3} = \tilde{T}_{x,1} + \varepsilon \tilde{T}_{x,2} + \varepsilon^2 \tilde{T}_{x,3} \). For \( k = x \) the terms \( u_{x,2}, \tilde{T}_{x,3} \) appear, we set \( u_{x,2} = \tilde{T}_{x,3} = 0 \).
- Given a vector-function \( g(x) = (g_1, \ldots, g_l) \), denote \( \left( \frac{\partial}{\partial \varphi} \right) \right|_{x} = (\frac{\partial g_1}{\partial \varphi}(\xi_1), \ldots, \frac{\partial g_l}{\partial \varphi}(\xi_l)) \), \( \left( \frac{\partial^2}{\partial \varphi^2} \right) \right|_{x} = (\frac{\partial^2 g_1}{\partial \varphi^2}(\eta_1), \ldots, \frac{\partial^2 g_l}{\partial \varphi^2}(\eta_l)) \), where \( \xi, \eta \) are some intermediate points on the segment \([x, \overline{x}]\).
Lemma 10.1. We have the following system of linear equations determining \( \mathcal{F}_{\varphi,2} \) and \( \mathcal{F}_{v,3} = (\mathcal{F}_{h,3}, \mathcal{F}_{w,3}) \):

\[
(1 + \varepsilon \frac{\partial u_{\varphi,1}}{\partial \varphi}) \mathcal{F}_{\varphi,2} + \varepsilon^2 \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{v,3} = \\
\frac{\partial \omega}{\partial v} u_{v,2} + \frac{1}{2} \int_{x,1,2} \frac{\partial^2 \omega}{\partial x^2} u_{x,1,2} + \left( \frac{\partial f_\varphi}{\partial x} \right) u_{x,1,2} - \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,1},
\]

\[
(1 + \varepsilon \frac{\partial u_{v,1}}{\partial \varphi}) \mathcal{F}_{v,3} + \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{\varphi,2} = \\
\frac{\partial f_v}{\partial v} u_{v,2} + \frac{1}{2} \int_{x,1,2} \frac{\partial^2 f_v}{\partial x^2} u_{x,1,2} - \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{v,2} - \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,1}.
\]  

(10.1)

Proof of lemmas 5.1 and 10.1. We shall differentiate the coordinate change (3.1) with respect to the time and rewrite all emerging terms as functions of \( \varphi \). For brevity the equations on \( h \) and \( w \) will be grouped together as an equation on \( v \). The derivatives of the left hand sides of (3.1) are given by (2.1). They are functions of \( x \), let us write Taylor’s expansions at the point \( \varphi \). We group together the terms of order at least 3 for the coordinate change in \( v \) and 2 for the change in \( \varphi \)

\[
\dot{\varphi} = \omega(x) + \varepsilon f_\varphi(x) = \omega(\varphi) + \varepsilon f_\varphi(\varphi) + \varepsilon^2 (\frac{\partial f_\varphi}{\partial x} u_{x,1,2}) + \varepsilon^3 \left( \frac{\partial^2 f_\varphi}{\partial x^2} u_{x,1,2} \right),
\]

\[
\dot{\omega} = \frac{\partial \omega}{\partial v} u_{v,2} + \frac{1}{2} \int_{x,1,2} \frac{\partial^2 \omega}{\partial x^2} u_{x,1,2} - \varepsilon^3 \left( \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,1} \right) - \varepsilon^2 \left( \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{v,2} \right) - \varepsilon \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,2}.
\]

Now we write the terms containing the derivatives of \( u_{v,i} \):

\[
\varepsilon \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{\varphi,2} + \varepsilon^2 \frac{\partial u_{v,2}}{\partial \varphi} \mathcal{F}_{v,3} = \\
= \varepsilon \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{\varphi,2} + \varepsilon^2 \frac{\partial u_{v,2}}{\partial \varphi} \mathcal{F}_{v,3} + \varepsilon^3 \left( \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{\varphi,1} \right) + \varepsilon^2 \left( \frac{\partial u_{v,2}}{\partial \varphi} \mathcal{F}_{v,2} \right) + \varepsilon \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,2} + \varepsilon^2 \frac{\partial u_{\varphi,2}}{\partial \varphi} \mathcal{F}_{v,3} + \varepsilon \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,3} + \varepsilon^2 \frac{\partial u_{\varphi,2}}{\partial \varphi} \mathcal{F}_{v,3}.
\]

Let us plug these expressions together with (3.2) into the time derivative of (3.1). Equating the terms of the same order in \( \varepsilon \) (grouping together the terms with order at least 3 for the equation on \( v \) and 2 for the equation on \( \varphi \)), we get (3.4) and (3.5), as well as the following equations:

\[
\mathcal{F}_{\varphi,2} = \frac{\partial \omega}{\partial v} u_{v,2} + \frac{1}{2} \int_{x,1,2} \frac{\partial^2 \omega}{\partial x^2} u_{x,1,2} + \left( \frac{\partial f_\varphi}{\partial x} \right) u_{x,1,2} - \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,1},
\]

\[
\mathcal{F}_{v,3} = \frac{\partial f_v}{\partial v} u_{v,2} + \frac{1}{2} \int_{x,1,2} \frac{\partial^2 f_v}{\partial x^2} u_{x,1,2} - \frac{\partial u_{v,1}}{\partial \varphi} \mathcal{F}_{v,2} - \frac{\partial u_{\varphi,1}}{\partial \varphi} \mathcal{F}_{\varphi,2},
\]

which are equivalent to (10.1), we just expand some terms like \( \mathcal{F}_{x,1,2,3} \) in order to move the terms containing \( \mathcal{F}_{\varphi,2} \) and \( \mathcal{F}_{v,3} \) to the left hand side. \( \square \)

11 Estimates related to the energy-angle variables

11.1 The coordinates \( \tilde{h}, \tilde{w}, \tilde{t}_i \)

Our goal in this section is to estimate how \( q, p \) (or, more generally, a smooth function \( \psi(q, p, z) \)) depend on \( h, w, \varphi \) for \( h \to 0 \). To do so, we introduce new coordinates \( \tilde{h}, \tilde{w}, \tilde{t}_i \). The subscript \( i \) is here because there will be different coordinate systems in different parts of the phase space. Then we will estimate how \( q, p \) depend on \( \tilde{h}, \tilde{w}, \tilde{t}_i \) and how \( \tilde{h}, \tilde{w}, \tilde{t}_i \) depend on \( h, w, \varphi \). Combining these estimates, we will get the required estimates of the dependence of \( q, p \) on \( h, w, \varphi \).

For simplicity we will assume that the Hamiltonian \( H \) is analytic. Then by [5]2 one can find a new coordinate system \( x, y \) in the neighborhood of the saddle \( C \) such that this coordinate change is analytic and volume preserving, and the unperturbed system in the new coordinates

\[ \text{[5]} \]

2 The result of [5] is for the case when \( H \) periodically depends on the time, but one may check that when \( H \) does not depend on the time the coordinate change constructed in [5] also does not depend on the time. The dependence on the parameter is also absent in [5], but the proof may be easily adapted for the parametric case.
is determined by a Hamiltonian $H_{x,y} = H_{x,y}(xy, z)$ with $H_{x,y}(C) = 0$ for all $z$ (we may subtract $H_{x,y}(C)$ from $H_{x,y}$ if this does not hold). Let $\tilde{h} = xy, \tilde{w} = w = z$, denote $a(\tilde{h}, \tilde{w}) = \frac{\partial H_{x,y}}{\partial h}$ (we have $a \neq 0$). Then in the new chart the unperturbed system rewrites as

$$\dot{x} = a(\tilde{h}, \tilde{w})x, \quad \dot{y} = -a(\tilde{h}, \tilde{w})y.$$  \hspace{1cm} (11.1)

Note that $\tilde{h}$ is a first integral of this system. Also note that $\tilde{h}$ is a smooth function of $h, z$, as one can find $\tilde{h}$ from the equality $H_{x,y}(h, z) = h$. This also means that $\tilde{h}$ is defined on the whole phase space, even far from $C$. We also have for any fixed value of $z$

$$\lim_{h \to 0} \frac{\tilde{h}}{h(h, z)} = a(0, z).$$  \hspace{1cm} (11.2)

We will assume that the coordinates $x, y$ are as drawn in Figure 2, else we can rotate this coordinate system by $\pi n/2$. Then, as $h > 0$ for $\tilde{h} = xy > 0$, we have $a > 0$. Rescaling $p, q, x$ and $y$ if needed, we may assume that the neighborhood of $C$ where the new coordinates are defined contains the square $S = \{x, y : -1 \leq x, y \leq 1\}$ for all $z$.

![Figure 2: Domains where $\tilde{t}$ are defined.](image)

The diagonals $x = \pm y$ split $S$ into four triangles adjacent to each of its sides. In each such triangle let us introduce the time $\tilde{t}$ (it can be positive or negative) that passes after the trajectory of the unperturbed system intersects the adjacent side of $S$. The time $\tilde{t}$ can also be continued outside the square to the neighborhood of the separatix crossing the transversal $\tilde{t}_i = 0$ (it is a side of $S$). Domains where each $\tilde{t}_i$ is defined are drawn in figure 2. Note that the coordinate systems $\tilde{h}, \tilde{w}, \tilde{t}_i$ cover the whole phase space (we only consider non-negative values of $\tilde{h}$ close to zero here).

We will assume that $\varphi = 0$ corresponds to the transversal $\Gamma$ given by $x = y \geq 0$. Note that here we consider the angle coordinate in the domain $G_3$, for the domains $G_1$ and $G_2$ the transversal $\Gamma$ would be given by $\{x = \pm y\} \cap G_i$.

### 11.2 Estimates on how $q, p$ depend on $\tilde{h}, \tilde{w}, \tilde{t}_i$

Outside of $S$ each point of the phase space is covered by two coordinate systems $\tilde{h}, \tilde{w}, \tilde{t}_i$. For both of them the coordinate change $p, q, z \leftrightarrow \tilde{h}, \tilde{w}, \tilde{t}_i$ is defined and is smooth without singularities. So we only need to consider what happens inside $S$. For definiteness, let us restrict ourselves to the triangle $\{1 \geq x \geq y \geq 0\}$. For brevity we will write just $\tilde{t}$ for the coordinate $\tilde{t}_i$ defined in this triangle. This means that $\tilde{t}$ is the time after the trajectory intersects the line $x = 1$. Note that $\tilde{t} \leq 0$ inside our triangle. We have

$$x = e^{a(\tilde{h}, \tilde{w})\tilde{t}}, \quad y = \tilde{h}e^{-a(\tilde{h}, \tilde{w})\tilde{t}}, \quad z = \tilde{w};$$

$$\tilde{h} = xy, \quad \tilde{w} = z, \quad \tilde{t} = \frac{\ln x}{a(xy, z)}. \hspace{1cm} (11.3)$$

$$\frac{\partial x}{\partial \tilde{h}} = \frac{\partial a}{\partial \tilde{h}}(\tilde{h}, \tilde{w})\tilde{t}x, \quad \frac{\partial x}{\partial \tilde{w}} = \frac{\partial a}{\partial \tilde{w}}(\tilde{h}, \tilde{w})\tilde{t}x, \quad \frac{\partial x}{\partial \tilde{t}} = a(\tilde{h}, \tilde{w})x;$$

$$\frac{\partial y}{\partial \tilde{h}} = -\frac{\partial a}{\partial \tilde{h}}(\tilde{h}, \tilde{w})\tilde{t}y + \frac{1}{x}, \quad \frac{\partial y}{\partial \tilde{w}} = -\frac{\partial a}{\partial \tilde{w}}(\tilde{h}, \tilde{w})\tilde{t}y, \quad \frac{\partial y}{\partial \tilde{t}} = -a(\tilde{h}, \tilde{w})y;$$

$$\frac{\partial z}{\partial \tilde{w}} = 1, \quad \frac{\partial z}{\partial \tilde{h}} = 0, \quad \frac{\partial z}{\partial \tilde{t}} = 0. \hspace{1cm} (11.4)$$
Note that $\dot{t} = \dot{t} e^{it} = O(1)$, as $a t < 0$. We also have $x \geq h_{1/2}$, as $x \geq y$. It follows that
\[
\frac{\partial y}{\partial h} = O(h^{-1/2}), \frac{\partial x}{\partial h} \frac{\partial y}{\partial w} \frac{\partial x}{\partial w} = O(\bar{t}^1 + 1)e^{-a|\bar{t}|}; \frac{\partial y}{\partial \bar{t}} \frac{\partial x}{\partial \bar{t}} = O(1)e^{-a|\bar{t}|}; \frac{\partial z}{\partial \bar{h}} = O(1) = 1; \frac{\partial z}{\partial \bar{t}} = 0, \frac{\partial \bar{z}}{\partial \bar{t}} = 0.
\]
(11.5)

Note that by (11.2) we may write $O(h^k)$ instead of $O(h^k)$. It also follows from (11.4) that
\[
\frac{\partial^2 y}{\partial h^2} = -\frac{2i}{x} \frac{\partial a}{\partial h} + \cdots = O(\bar{t}^1 + 1)h^{-1/2}; \frac{\partial^2 y}{\partial \bar{h} \partial \bar{t}} = O(\bar{t}^1 + 1)h^{-1/2};
\]
\[
\frac{\partial^2 y}{\partial \bar{h}^2} = O(h^{-1/2}); \frac{\partial^2 x}{\partial \bar{h} \partial \bar{t}} = O(\bar{t}^1 + 1)e^{-a|\bar{t}|}; \frac{\partial^2 y}{\partial h^2} = O(e^{-a|\bar{t}|});
\]
\[
\frac{\partial^2 y}{\partial \bar{t}^2} = O((\bar{t}^1 + 1))^2e^{-a|\bar{t}|}; \frac{\partial^2 \bar{t}}{\partial \bar{h}} = 0.
\]
(11.6)

Now let us return from $(x, y)$ to $(q, p)$. Let us consider a smooth function $\psi(x, y, z)$ without singularities, e.g. $\psi = q$ or $\psi = p$. We will use the following formula $(a_i, b_i$ are some coordinate systems and $c$ is some function)
\[
\frac{\partial^2 \psi}{\partial a_i \partial b_j} = \sum_{i} \frac{\partial^2 b_i}{\partial a_i \partial a_j} \frac{\partial c}{\partial b_i} + \sum_{k, l} \frac{\partial b_i}{\partial a_j} \frac{\partial b_k}{\partial a_j} \frac{\partial^2 c}{\partial b_i \partial b_l}.
\]
(11.7)

We can estimate the derivatives of $\psi$, using the chain rule for the first derivatives and (11.7) for the second derivatives, and (11.5), (11.6). This gives us
\[
\frac{\partial \psi}{\partial h} = O(h^{-1/2}); \frac{\partial \psi}{\partial w} = O(1); \frac{\partial \psi}{\partial \bar{t}} = O(e^{-a|\bar{t}|});
\]
\[
\frac{\partial^2 \psi}{\partial h^2} = O(h^{-1}); \frac{\partial^2 \psi}{\partial \bar{h} \partial \bar{t}} = O(\bar{t}^1 + 1)h^{-1/2}; \frac{\partial^2 \psi}{\partial \bar{h}^2} = O(h^{-1/2});
\]
\[
\frac{\partial^2 \psi}{\partial \bar{t}^2} = O(1); \frac{\partial^2 \psi}{\partial h \partial \bar{t}} = O(|\bar{t}|e^{-a|\bar{t}|}); \frac{\partial^2 \psi}{\partial \bar{t}^2} = O(e^{-a|\bar{t}|}).
\]
(11.8)

Outside of $S$ we can take as $\tilde{t}$ any of the two coordinates $\tilde{t}_i$, defined near each separatrix, we have $|\bar{t}| + 1 \sim 1$. These estimates are valid everywhere: we obtained them in a part of $S$, in other parts of $S$ they can be obtained similarly, and outside of $S$ we even have $O(1)$ on all right hand sides as the considered coordinate change is smooth.

Let us also consider a function $\psi_0$ with $\psi_0(C) = 0$ (e.g. $\psi_0 = f_h$). As $C$ corresponds to $x = y = 0$, the functions $\psi_0, \frac{\partial \psi_0}{\partial w}, \frac{\partial^2 \psi_0}{\partial h^2}$ (here the derivatives are taken for fixed $x, y$) all vanish at $C$ and so are $O(e^{-a|\bar{t}|})$. Some of the estimates above turn out to be better for $\psi_0$:
\[
\psi_0 = O(e^{-a|\bar{t}|}); \frac{\partial \psi_0}{\partial h} = O(e^{-a|\bar{t}|})(|\bar{t}| + 1); \frac{\partial^2 \psi_0}{\partial \bar{t}^2} = O(e^{-a|\bar{t}|})(|\bar{t}| + 1)^2.
\]
(11.9)

### 11.3 Estimates on how $\tilde{h}, \tilde{t}_i$ depend on $h, w, \varphi$

First, recall that $\tilde{h}$ is an analytic function of $h, w$. As $\tilde{h}(0, w) = 0$, all summands in the series for $\tilde{h}$ contain $h$. Hence, we can write $\tilde{h} = \tilde{h}_0(h, w)$, where $\tilde{h}_0(h, w)$ is also analytic. From this we have
\[
\frac{\partial \tilde{h}}{\partial w} \frac{\partial^2 \tilde{h}}{\partial w^2} = O(h).
\]
(11.10)

Denote by $S(h, w)$ the time that the solution of the unperturbed system with given $h, w$ takes to get from the diagonal of the square $S$ to its side. Then the total time spent inside $S$ during each period is $4S$. From (11.3) we have $S = \frac{\ln(h, w)}{2\pi(h, w)}$. Hence, by (11.10) we have
\[
\frac{\partial S}{\partial \bar{h}} = O(h^{-2}), \frac{\partial \bar{S}}{\partial \bar{w}} = O(h^{-1}), \frac{\partial^2 \bar{S}}{\partial \bar{w}^2} = O(h^{-1}), \frac{\partial^2 \bar{S}}{\partial \bar{h} \partial \bar{w}} = O(h^{-1}), \frac{\partial^2 S}{\partial w^2} = O(h^{-1}).
\]
(11.11)

Denote by $T_{reg, 1}(h, w)$ and $T_{reg, 2}(h, w)$ the times that the solution of the unperturbed system spends outside $S$ near each of the separatix loops during each period. These are smooth functions of $h, w$. Then
\[
T = 4S + T_{reg, 1} + T_{reg, 2}.
\]

From (11.11) we get the estimates on $T, \omega$ from Table 1.

Let us recall that for the unperturbed system we denote by $t = T^\varphi/(2\pi)$ the time passed after crossing the transversal $\varphi = 0$ given by $x = y > 0$. For each $\tilde{t}_i$ we have $\tilde{t}_i = t - t_{0,i}$, where $t_{0,i}$ is the value of $t$ corresponding to $\tilde{t}_i = 0$. We have (see Figure 2)
\[
t_{0,i} = kS + k_1T_{reg, 1} + k_2T_{reg, 2} \text{ with } k \in \{1, 3\}; k_1, k_2 \in \{0, 1\}.
\]
(11.12)
Hence, we have

$$\tilde{t}_i = \left(4S + T_{reg,1} + T_{reg,2}\right) \frac{q}{2\pi} - kS - k_1T_{reg,1} - k_2T_{reg,2}.$$ 

This may also be rewritten as

$$\tilde{t}_i = S(h, w) \left(\frac{2\sigma}{\pi} - k\right) + T_{reg}(h, w, \varphi),$$

where $T_{reg}$ has no singularities and $\frac{2\sigma}{\pi} - k = O(\ln^{-1} h)(|\tilde{t}_i| + 1)$.

From these formulas, (11.11), smooth dependence of $\tilde{t}$ on $h$, $w$ and (11.10) we get

$$\frac{\partial \tilde{t}_i}{\partial h} = O(h^{-1} \ln^{-1} h)(|\tilde{t}_i| + 1); \quad \frac{\partial \tilde{t}_i}{\partial w} = O(|\tilde{t}_i| + 1); \quad \frac{\partial \tilde{t}_i}{\partial \varphi} = O(\ln h);$$

$$\frac{\partial \tilde{h}}{\partial h} = O(1); \quad \frac{\partial \tilde{h}}{\partial w} = 1; \quad \frac{\partial \tilde{h}}{\partial \varphi} = O(h);$$

$$\frac{\partial^2 \tilde{t}_i}{\partial h^2} = O(h^{-1} \ln^{-1} h)(|\tilde{t}_i| + 1); \quad \frac{\partial^2 \tilde{t}_i}{\partial w^2} = O(\ln h); \quad \frac{\partial^2 \tilde{t}_i}{\partial \varphi^2} = O(h^{-1});$$

$$\frac{\partial^2 \tilde{h}}{\partial h^2} = O(1); \quad \frac{\partial^2 \tilde{h}}{\partial w^2} = O(h); \quad \frac{\partial^2 \tilde{h}}{\partial \varphi^2} = 0; \quad \frac{\partial^2 \tilde{h}}{\partial * \partial \varphi} = 0$$

*(11.13)*

### 11.4 Estimates on how $q$, $p$ depend on $h$, $w$, $\varphi$

As above, let $\psi(x, y, z)$ be a smooth function without singularities, e.g. $\psi = q$ or $\psi = p$. Applying to (11.8) and (11.13) the chain rule for first derivatives and formula (11.7) for second derivatives, we get the following estimates (here $\bar{t}$ is one of the coordinates $t_i$ as in Section 11.2):

$$\frac{\partial \psi}{\partial h} = O(h^{-1} \ln^{-1} h)e^{-a|\bar{t}|}(|\bar{t}| + 1); \quad \frac{\partial \psi}{\partial w} = O(1); \quad \frac{\partial \psi}{\partial \varphi} = O(\ln h)e^{-a|\bar{t}|};$$

$$\frac{\partial^2 \psi}{\partial h^2} = O(h^{-1} \ln^{-1} h)e^{-a|\bar{t}|}(|\bar{t}| + 1); \quad \frac{\partial^2 \psi}{\partial w^2} = O(\ln h)e^{-a|\bar{t}|}(|\bar{t}| + 1)^2; \quad \frac{\partial^2 \psi}{\partial \varphi^2} = O(1);$$

*(11.14)*

Let us also note that for a function $\psi_0$ with $\psi_0(C) = 0$ we can use (11.9) and some of the estimates above turn out to be better:

$$\psi_0 = O(e^{-a|\bar{t}|}); \quad \frac{\partial \psi_0}{\partial w} = O(e^{-a|\bar{t}|})(|\bar{t}| + 1); \quad \frac{\partial^2 \psi_0}{\partial w^2} = O(e^{-a|\bar{t}|})(|\bar{t}| + 1)^2.$$  

*(11.15)*

Finally, as $\frac{\partial \psi}{\partial w} = O(|\bar{t}| + 1)e^{-a|\bar{t}|}$ by (11.15), we have

$$\frac{\partial \psi}{\partial w} = \frac{\partial \psi}{\partial z} + O(|\bar{t}| + 1)e^{-a|\bar{t}|}.$$  

*(11.16)*

### 11.5 Estimates on $f$

Here we obtain the estimates on $f_h$, $f_w$, and $f_\varphi$ from Table 1. The estimates on $f_w$, together with its derivatives follow from (11.14) as $f_w = f_{z_i}$ is smooth without singularities. The estimates on $f_h$ follow from (11.14) and (11.15), as $f_h = f_{\tau} + f_{\varphi} + f_{\omega} + f_{\epsilon}$ is smooth without singularities and $f_h(C) = 0$ (as by [6, Lemma 2.1] we have $\frac{\partial \varphi}{\partial \varphi} = O(\ln h)\frac{\partial \varphi}{\partial \varphi} (C)$, $\frac{\partial \varphi}{\partial \varphi} (C)$).

Let us estimate $f_h(h, w, 0)$. Recall that $t$ is the time passed after the solution of the unperturbed system crosses the transversal $x = y > 0$. For $x, y > 0$ we have $t = \frac{1}{2a(h, w)}(\ln x - \ln y)$, this is obtained by solving (11.11) with initial conditions $x = y = \hat{h}^{1/2}$ for $t = 0$. For $\varphi = 0$ (and therefore $t = 0, x = y = \hat{h}^{1/2}$) we have

$$\frac{\partial \tau}{\partial x} = \frac{\partial}{\partial x}(\omega t) = \frac{\omega}{2a(h, w)} \frac{\partial \tau}{\partial x} = \frac{\omega}{2a(h, w)} x, \quad \frac{\partial \varphi}{\partial y} = \frac{\omega}{2a(h, w)} y, \quad \frac{\partial \varphi}{\partial z} = 0,$$

$$f_h(h, w, 0) = f_\tau \frac{\partial \varphi}{\partial x} + f_\varphi \frac{\partial \varphi}{\partial y} = \frac{\omega}{2a(h, w)} (x f_\tau - y f_\varphi) = \frac{\omega}{2a(h, w)} (f_\varphi - f_\varphi).$$
Here \( f_x, f_y \) are the components of the vector field \( f \) written in the \( x, y \) chart, they are \( O(1) \). Hence, \( f_x(h, w, 0) = O(h^{-1/2} \ln^{-1} h) \). We can apply (11.14) to \( \psi = \frac{f_x - f_y}{2\omega} \), together with (11.10) this gives

\[
\frac{\partial f_x(h, w, 0)}{\partial h} = O(h^{-3/2} \ln^{-1} h), \quad \left\| \frac{\partial f_x(h, w, 0)}{\partial w} \right\| = O(h^{-1/2} \ln^{-1} h).
\]

Denote \( g(h, w) = f_x(h, w, \varphi_0) \), where \( \varphi_0 \approx \pi \) corresponds to \( x = y < 0 \). As \( \varphi_0 \) corresponds to \( t = 2\pi + T_{reg,1} \), we have \( \varphi_0 = \pi + 0.5\omega(T_{reg,1} - T_{reg,2}) \). We can write \( g = g_0 + g_1 \), where \( g_0 \) is computed as if \( x, y < 0 \) corresponds to \( \varphi = 0 \) and \( g_1 = \frac{\partial f_x}{\partial \varphi} f_y + \frac{\partial f_y}{\partial \varphi} f_x \). As for \( x = y \) we have \( e^{-|t|} = O(h^{1/2}) \) and \( f_x = O(h^{1/2}), \frac{\partial f_x}{\partial h} = O(h^{1/2} \ln h), \frac{\partial f_x}{\partial w} = O(h^{-1/2}) \) by (11.14) and (11.15), we have \( g_1 = O(h^{-1/2} \ln^{-1} h) \) and \( g_{1,h} = O(h^{-3/2} \ln^{-2} h), g_{1,w} = O(h^{-1/2} \ln^{-1} h) \). For \( g_0 \) and its derivatives we can use the estimates for \( f_x(h, w, 0) \) proved above. Hence, the estimates for \( f_x(h, w, 0) \) proved above also hold for \( g, \frac{\partial f_x}{\partial h} \) and \( \frac{\partial f_x}{\partial w} \).

**Lemma 11.1.** Suppose that for a function \( f(h, w, \varphi) \) we have the estimate \( \alpha(h, w, \varphi) = O_*(\tilde{\alpha}) \) with \( \tilde{\alpha} = \tilde{\alpha}(h) \). Then for any \( \varphi_0, \varphi_1 \in [0, 2\pi] \) we have \( \int_{\varphi_0}^{\varphi_1} \alpha \omega d\varphi = O(\tilde{\alpha} \ln^{-1} h) \).

**Proof.** It is enough to show that \( \int_0^T |\alpha| dt = O(\tilde{\alpha}) \). This integral \( dt \) can be split into four integrals \( \tilde{t}_i \), and each of them is \( O(1) \) as the estimate on \( \alpha \) contains a term that decays exponentially with the growth of \( \tilde{t} \).

From (2.2) and the estimates on \( f_h \) and \( f_w \) from Table 1 we have \( \frac{\partial f_x}{\partial \varphi} = O_*(h^{-1} \ln^{-1} h) \). Note that the estimates for \( \text{div}(f) \) are given by (11.14), as this function is smooth. For given \( h, w \) denote by \( \varphi_\ast(\varphi) \) the angle corresponding to the “nearest” intersection of the solution with \( h \) and the line \( x = y \). We have \( \varphi_\ast = 0 \) for \( \varphi < \pi/2 \) or \( \varphi > 3\pi/2 \) and \( \varphi_\ast(\varphi) \approx \pi \) for \( \pi/2 \leq \varphi \leq 3\pi/2 \). As \( \alpha \) is in the proof of Lemma 11.1, from \( \frac{\partial f_x}{\partial \varphi} = O_*(h^{-1} \ln^{-1} h) \) we can obtain \( f_x(\varphi) = f_x(\varphi_\ast) = O_*(h^{-1} \ln^{-1} h) \). As for \( x = y \) we have \( e^{-|t|} \sim h^{1/2} \) and \( f_x(\varphi_\ast) = O(h^{-1/2} \ln^{-1} h) = O(h^{-1} \ln^{-2} h) e^{-|t|}(\tilde{t} + 1) \), we have \( f_x = O_*(h^{-1} \ln^{-1} h) \).

Let us apply \( \frac{\partial}{\partial h} \) to (2.2), this gives \( \frac{\partial^2 f_x}{\partial \varphi \partial h} = O_*(h^{-2} \ln^{-1} h) \). Arguing as above, we get \( \frac{\partial f_x}{\partial \varphi} = O_*(h^{-1} \ln^{-1} h) \). The estimate \( \left\| \frac{\partial f_x}{\partial \varphi} \right\| = O_*(h^{-1} \ln^{-2} h) \) is obtained in the same way.

### 12 Estimates for \( \tilde{T}_{h,2} \) and \( \tilde{T}_{w,2} \)

For brevity, in this section we will write \( u_a \) instead of \( u_{a,1}, \tilde{T}_a \) instead of \( \tilde{T}_{a,1} \), and so on.

#### 12.1 Expressions for \( \tilde{T}_{h,2} \) and \( \tilde{T}_{w,2} \)

**Lemma 12.1.**

\[
2\pi \tilde{T}_{h,2} = \int_0^{2\pi} \left( \text{div} f - \sum_{w_i} \frac{\partial f_w}{\partial u_{w_i}} u_h + \sum_{w_i} \frac{\partial f_h}{\partial u_{w_i}} u_w \right) d\varphi - \omega^{-1} \sum_{w_i} \frac{\partial \omega}{\partial u_{w_i}} \int_0^{2\pi} f_h u_w d\varphi.
\]

\[
2\pi \tilde{T}_{w,2} = \frac{\partial}{\partial h} \int_0^{2\pi} f_w u_h d\varphi - \omega^{-1} \sum_{w_i} \frac{\partial \omega}{\partial u_{w_i}} \int_0^{2\pi} f_h u_w d\varphi + \int_0^{2\pi} \text{div}(f) u_a d\varphi + \sum_{w_i} \frac{\partial f_w}{\partial u_{w_i}} u_a \frac{\partial f_h}{\partial u_{w_i}} u_a d\varphi \quad \text{for } a = w_1, \ldots, w_k.
\]

**Proof.** Fix \( a \in \{h, w_1, \ldots, w_k\} \). By (3.6) we have

\[
2\pi \tilde{T}_{w,2} = \int_0^{2\pi} \frac{\partial f_w}{\partial u_a} u_h + \frac{\partial f_h}{\partial u_a} u_w + \sum_{w_i} \frac{\partial f_w}{\partial u_{w_i}} u_a \frac{\partial f_h}{\partial u_{w_i}} u_a d\varphi.
\]

By Lemma 3.1 we have \( \frac{\partial u_i}{\partial \varphi} = \frac{1}{\omega}(f_v - \tilde{T}_v), \quad b = h, w_1, \ldots, w_k; \quad \frac{\partial u_b}{\partial \varphi} = \frac{1}{\omega}(f_v - \tilde{T}_v + \frac{\partial u_b}{\partial \varphi} u_h + \sum_{w_i} \frac{\partial f_w}{\partial u_{w_i}} u_h) \). We also have \( \int_0^{2\pi} u_b d\varphi = 0, \quad b = h, w, w_1, \ldots, w_k \). Hence, \( \int_0^{2\pi} \tilde{T}_w u_a d\varphi = 0 \) for \( b, c = h, \varphi, w_1, \ldots, w_k \). Integrating by parts, we have

\[
\int_0^{2\pi} \frac{\partial f_w}{\partial \varphi} u_a d\varphi = -\int_0^{2\pi} f_w \frac{\partial u_a}{\partial \varphi} d\varphi = -\omega^{-1} \int_0^{2\pi} f_w \frac{\partial u_a}{\partial \varphi} d\varphi + \frac{2\pi}{\omega} \tilde{T}_w f_w d\varphi + \omega^{-1} \sum_{b=h,w_1,\ldots,w_k} \frac{\partial \omega}{\partial b} \int_0^{2\pi} f_w u_b d\varphi.
\]

Similarly, we have

\[
\int_0^{2\pi} \frac{\partial f_h}{\partial \varphi} u_a d\varphi = -\int_0^{2\pi} f_h \frac{\partial u_a}{\partial \varphi} d\varphi = -\omega^{-1} \int_0^{2\pi} f_h \frac{\partial u_a}{\partial \varphi} d\varphi + \frac{2\pi}{\omega} \tilde{T}_w f_h d\varphi.
\]
Hence, we have
\[ 2\pi \tilde{f}_{a,2} = \int_0^{2\pi} \frac{\partial f_a}{\partial h} u_h - \frac{\partial f_x}{\partial \varphi} u_u + \sum_{w_i} \frac{\partial f_a}{\partial w_i} u_{w_i} d \varphi - \omega^{-1} \sum_{b=h,w_1,...,w_k} \frac{\partial \omega}{\partial b} \int_0^{2\pi} f_a u_b d \varphi. \] (12.2)

Note that for \( b = h, w_1, \ldots, w_k \) we have
\[ \int_0^{2\pi} f_b u_b d \varphi = \omega \int_0^{2\pi} u_b \frac{\partial u_b}{\partial \varphi} d \varphi = 0. \] (12.3)

Therefore, for \( a = h \) we have
\[ 2\pi \tilde{f}_{h,2} = \int_0^{2\pi} \left( \frac{\partial f_h}{\partial h} + \frac{\partial f_x}{\partial \varphi} \right) u_h + \sum_{w_i} \frac{\partial f_a}{\partial w_i} u_{w_i} d \varphi - \omega^{-1} \sum_{b=h,w_1,...,w_k} \frac{\partial \omega}{\partial b} \int_0^{2\pi} f_h u_b d \varphi. \]

Rewriting this using (2.2) and (12.3) yields the first formula in (12.1).

Now we assume that \( a = w_1, \ldots, w_k \). We can compute
\[ \omega \frac{\partial}{\partial h} \int_0^T f_a u_b dt = \omega \frac{\partial}{\partial h} \left( \omega^{-1} \int_0^{2\pi} f_a u_b d \varphi \right) = -\omega^{-1} \omega \frac{\partial}{\partial h} \int_0^{2\pi} f_a u_b d \varphi + \frac{\partial}{\partial h} \int_0^{2\pi} f_a u_b d \varphi. \]

As \( \frac{\partial}{\partial h} f_a u_b d \varphi = \int_0^{2\pi} \frac{\partial f_a}{\partial h} u_a + \sum_{w_i} \frac{\partial f_a}{\partial w_i} u_{w_i} d \varphi - \omega^{-1} \sum_{b=h,w_1,...,w_k} \frac{\partial \omega}{\partial b} \int_0^{2\pi} f_h u_b d \varphi \), from (12.2) we have
\[ 2\pi \tilde{f}_{a,2} = \int_0^{2\pi} \frac{\partial f_a}{\partial \varphi} u_a - f_a \frac{\partial u_a}{\partial h} \frac{\partial}{\partial h} \int_0^T f_a u_b dt - \omega^{-1} \sum_{b=h,w_1,...,w_k} \frac{\partial \omega}{\partial b} \int_0^{2\pi} f_h u_b d \varphi. \]

As \( (\frac{\partial u_a}{\partial h})_a = 0 \), we have
\[ \int_0^{2\pi} f_a \frac{\partial u_a}{\partial h} d \varphi = \omega \int_0^{2\pi} \frac{\partial u_a}{\partial h} \frac{\partial}{\partial h} \int_0^{2\pi} f_h u_b d \varphi = \omega \int_0^{2\pi} u_b \frac{\partial u_a}{\partial h} (\omega^{-1} (f_h - T)) d \varphi = \]
\[ = \int_0^{2\pi} u_b \frac{\partial f_h}{\partial h} d \varphi + \omega^{-1} \frac{\partial \omega}{\partial h}, \int_0^{2\pi} f_h u_b d \varphi. \]

Hence,
\[ 2\pi \tilde{f}_{a,2} = \int_0^{2\pi} \left( \frac{\partial f_h}{\partial h} + \frac{\partial f_x}{\partial \varphi} \right) u_a + \sum_{w_i} \frac{\partial f_a}{\partial w_i} u_{w_i} d \varphi - \omega^{-1} \sum_{b=h,w_1,...,w_k} \frac{\partial \omega}{\partial b} \int_0^{2\pi} f_h u_b d \varphi - \omega^{-1} \frac{\partial \omega}{\partial h}, \int_0^{2\pi} f_h u_b d \varphi. \]

Rewriting this using (2.2) and \( T^{-1} \frac{\partial T}{\partial \psi} + \omega^{-1} \frac{\partial \omega}{\partial h} = \frac{\partial}{\partial h} (T + \omega) = 0 \) yields the second formula in (12.1).

### 12.2 Estimate for \( \tilde{f}_{h,2} \)

**Lemma 12.2.** Let \( \psi \) be either a smooth function \( \psi(p, q, z) \) or the function \( \frac{\partial T}{\partial \psi} \), \( b = h, w_1, \ldots, w_k \), \( i = 1, \ldots, k \) and \( u(h, w, \varphi) \) be a function with \( u = O(1) \), \( (u)_{\varphi} = 0 \) (e. g. \( u_{h,i} \) for \( b = h, w_1, \ldots, w_k \)). Then
\[ \int_0^{2\pi} \psi \cdot u \ d \varphi = O(\ln^{-1} h). \]

**Proof.** First, let us prove that \( \int_0^T |\psi - a| dt = O(1) \), where \( a = \psi(C) \) for smooth \( \psi \) and \( a = \frac{\partial f_h}{\partial h}(C) \) for \( \psi = \frac{\partial f_h}{\partial h} \). For smooth \( \psi \) this follows from [6, Lemma 3.2]. For \( \psi = \frac{\partial f_h}{\partial h} \) by (11.16) we have \( \psi - \frac{\partial f_h}{\partial h} = O(1) \) \( + |\varphi_i| \) \( (\text{here } \frac{\partial f_h}{\partial h} \text{ is smooth}) \). From this we have \( \int |\psi - \frac{\partial f_h}{\partial h}| dt = O(1) \), so the required statement follows from the smooth case with \( \psi = \frac{\partial f_h}{\partial h} \).

As \( u = O(1) \), the estimate above implies \( \int_0^T (\psi - a) u dt = O(1) \). As \( (u)_{\varphi} = 0 \), we have \( \int_0^T a \cdot u dt = 0 \) and \( \int_0^T \psi \cdot u dt = O(1) \). Changing the variable, we obtain the required estimate.

**Lemma 12.3.**
\[ \tilde{f}_{h,2} = O(\ln^{-1} h) \]

**Proof.** By Table 1 we have \( \omega^{-1} \frac{\partial \omega}{\partial h} = O(1) \). Plugging this and the estimate of Lemma 12.2 in (12.1) yields the required estimate.
12.3 Estimate for $\tilde{T}_{w,2}$

**Lemma 12.4.** For any smooth function $\psi(p,q,z)$ we have

$$\frac{\partial}{\partial \tilde{h}} \left( \int_{t=0}^{T} \psi u_h dt \right) = O(h^{-1} \ln^{-2} h)$$

**Proof.** We will assume $\psi(C) = 0$, as we can replace $\psi$ by $\psi - \psi(C)$ due to $u_h(\phi) = 0$. We will use the integral expression for $u$ given by (3.8):

$$u_h(t_0) = \int_{t=0}^{T} \left( \frac{t - 1}{T - 2} \right) f_h(t + t_0) dt.$$ 

We have (in the formula below $t_1 = t + t_0 + kT$, where $k \in \mathbb{Z}$ is such that $t_1 \in [0,T]$; $\{x\}$ denotes the fractional part of $x$, i.e. such number $y$ in $[0,1)$ that $x - y \in \mathbb{Z}$)

$$\int_{t_0 = 0}^{T} \psi u_h dt_0 = \int_{t=0}^{T} \left( \frac{t - 1}{T - 2} \right) f_h(t + t_0) \psi(t) dt_0 dt_0 = \int_{t=0}^{T} \left( \frac{t_1 - t_0}{T} \right) f_h(t_1) \psi(t) dt_1 dt_0.$$ 

We will use the following notation from sections 11.1-11.3: $x, y, \tilde{x}, \tilde{h}, S$. The phase space can be split by the lines $x = y, \tilde{t}_1 = -\tilde{t}_2$ and $\tilde{t}_3 = -\tilde{t}_4$ (see Figure 2) into four parts, such that each part is covered by one of the coordinates $\tilde{t}_1, \ldots, \tilde{t}_4$. Let all possible values of $t_1$ in its part span the segment $[a_1, b_1]$. Note that for the second two lines the values of the coordinates $\tilde{t}_i$ defined there (and also the values of the corresponding $a_i$ or $b_i$) are smooth functions of $h$, $w$ without singularities. For example, for $\tilde{t}_1 = -\tilde{t}_2$ we have $\tilde{t}_1 = a_1 = T_{reg,i}/2$ and $\tilde{t}_2 = a_2 = -T_{reg,i}/2$, where $T_{reg,i}(h,w)$ is the time between the points with $t_1 = 0$ and $t_2 = 0$.

Denote $\lambda_{i,j}(t_0, t_1) = \left\{ \frac{t(t_1 + t_1) - (t(t_1 + t_1))}{T} \right\} - \frac{1}{2}$. The integral above can be split into a sum of the following 16 integrals for $i, j = 1, \ldots, 4$:

$$\int_{t_0 = a_1}^{b_1} \int_{t_1 = a_2}^{b_2} \lambda_{i,j}(t_0, t_1) f_h(t_1) \psi(t_1) dt_1 dt_0.$$ 

(12.4)

As $\frac{\partial}{\partial \tilde{h}} = \frac{\partial}{\partial \tilde{h}} + \frac{\partial}{\partial \tilde{h}} = O(1) \frac{\partial}{\partial \tilde{h}}$, we will estimate the $\tilde{h}$-derivative of (12.4) instead of its $h$-derivative. Let us first note that the discontinuity of $\lambda_{i,j}(t_0, t_1)$ corresponds to $\tilde{t}_1 = t_0$ and $\tilde{t}_2 = t_1$, giving the same point, so $i = j$ and $t_0 = t_1$, and this discontinuity does not create additional terms in the $\tilde{h}$-derivative of (12.4). By (11.12) we have $\lambda_{i,j}(t_0, t_1) + 0.5 = \left\{ \frac{t_1 - t_0 + k_{i,j}S + T_{reg,i,j}}{T} \right\}$ (here $k_{i,j} \in \mathbb{Z}$ and $T_{reg,i,j}$ is a smooth function of $h, w$), so

$$\frac{\partial \lambda_{i,j}(t_0, t_1)}{\partial \tilde{h}} = O(h^{-1} \ln^{-2} h)(|t_1 - t_0| + 1).$$ 

As $f_h(C) = \psi(C) = 0$, by (11.15) we have $f_h(t_1), \psi(t_1) = O(e^{-n(t_1)})$ for $s = 1, \ldots, 4$, this means $\int |t_0 - t_1| f_h(t_1) \psi(t_0) dt_0 dt_1 = O(1)$ and so

$$\int \lambda_{i,j} f_h(t_1) \psi(t_0) dt_0 dt_1 = O(h^{-1} \ln^{-2} h).$$ 

(11.8) We have $\int \lambda_{i,j} f_h \psi \frac{dt_0 dt_1}{\partial \tilde{h}} = O(h^{-0.5} \ln^2 h)$ and $\int \lambda_{i,j} f_h \frac{dt_0 dt_1}{\partial \tilde{h}} = O(h^{-0.5} \ln^2 h)$.

The $\tilde{h}$-derivative of (12.4) also has terms associated with the change of the domain of integration. There are four similar terms, let us consider just one of them:

$$\frac{\partial \lambda}{\partial \tilde{h}} \int_{t_1 = a_2}^{b_2} \lambda_{i,j}(t_0, t_1) f_h(t_1) \psi(t_1) dt_1.$$ 

There are two cases. First, $a_1$ may correspond to $x = y = O(\sqrt{T})$, then $\frac{\partial \lambda}{\partial \tilde{h}} = O(h^{-1})$, $\psi = O(\sqrt{T})$, and our term is $O(h^{-0.5} \ln h)$. Otherwise, we have $\frac{\partial \lambda}{\partial \tilde{h}} = O(1)$ and our term is $O(\ln h)$.

Combining these estimates, we see that the $\tilde{h}$-derivative (and so also the $h$-derivative) of (12.4) is $O(h^{-1} \ln^{-2} h)$. This proves the lemma.

**Lemma 12.5.**

$$\tilde{T}_{w,2} = O(h^{-1} \ln^{-3} h), \quad i = 1, \ldots, k.$$ 

**Proof.** By Table 1 we have $\omega^{-1} \frac{\partial \omega}{\partial \tilde{h}} = O(1)$. This and the estimate of Lemma 12.2 gives the estimate $O(\ln^{-1} h)$ for all terms of the expression (12.1) for $\tilde{T}_{w,2}$ except the first one. The first term is estimated by Lemma 12.4.

13 Estimates related to the averaging chart

In this section we prove the estimates from Table 1 for the functions $u_{h,i}$ and $\tilde{T}_{h,i}$. The following lemma allows to mass-produce such estimates. However, these estimates are not always good, so we will estimate some of these functions in a different way.
Lemma 13.1. Given a function $Y(h, w, \varphi)$, let
\[ \mathcal{T} = (Y)_{\varphi} \] (13.1)
and let the function $u$ be determined by the equation $\omega \frac{\partial u}{\partial \varphi} = Y - \mathcal{T}$ and the condition $\langle u \rangle_{\varphi} = 0$. Denote $Y_t = T \cdot Y$. Let $v = h, w, h, hw, wu$ and $w$ denote the corresponding first or second derivative. Then we can estimate the functions $\mathcal{T}$ and $u$ and their derivatives using estimates for $Y$ and $Y_t$ and their derivatives (these estimates are denoted by $\bar{Y}, \bar{Y}_t, \bar{Y}_v, \bar{Y}_{tv}$ below, they depend only on $h$) in the following way:

1. $\mathcal{T} = O(\bar{Y})$ for $Y = O(\bar{Y})$; $\mathcal{T} = O(\bar{Y} \ln^{-1} h)$ for $Y = O_h(\bar{Y})$. 
2. $\frac{\partial^2 u}{\partial \varphi^2} = O(\bar{Y}_v)$ for $\frac{\partial^2 u}{\partial \varphi^2} = O(\bar{Y}_v)$, $\frac{\partial u}{\partial \varphi} = O(\bar{Y}_v \ln^{-1} h)$ for $\frac{\partial u}{\partial \varphi} = O(\bar{Y}_v)$. 
3. $\frac{\partial^2 u}{\partial \varphi \partial \psi} = O(\bar{Y}_v)$ for $\mathcal{T} = O(\bar{Y}_v)$ or $Y_t = O(\bar{Y}_v)$.
4. $\frac{\partial^2 u}{\partial \varphi \partial \psi} = O(\bar{Y}_v)$ for $\frac{\partial^2 u}{\partial \varphi \partial \psi} = O(\bar{Y}_v)$ or $\frac{\partial^2 u}{\partial \varphi \partial \psi} = O(\bar{Y}_v)$. 
5. $u = O(\bar{Y}_v)$ for $Y_t = O(\bar{Y}_v)$; $u = O(\bar{Y}_v \ln^{-1} h)$ for $Y_t = O(\bar{Y}_v)$.
6. $\frac{\partial^2 u}{\partial \varphi^2} = O(\bar{Y}_v)$ for $\frac{\partial^2 u}{\partial \varphi^2} = O(\bar{Y}_v)$ or $\frac{\partial^2 u}{\partial \varphi^2} = O(\bar{Y}_v)$. 

Remark 13.2. As the maps $\mathcal{T} \mapsto u$ and let the function $u$ be determined by the equation $\omega \frac{\partial u}{\partial \varphi} = Y - \mathcal{T}$ and the condition $\langle u \rangle_{\varphi} = 0$. Denote $Y_t = T \cdot Y$. Let $v = h, w, h, hw, wu$ and $w$ denote the right hand sides of (10.1)):

\[ \sum_{\gamma} \frac{\partial u}{\partial \varphi} \partial \psi = O(\bar{Y}_v) \]

The functions $u_{h,1}$ and $\mathcal{T}_{h,1}$ are given by Lemma 3.1 and Lemma 10.1. Lemma 13.1 allows to obtain the estimates for $u_{h,1}$ and their derivatives in the functions $u_{w,1}$. However, for $Y = f_w - f_w(C)$ we get the same value of $u_{w,1}$, but better estimates, as we may use (11.15).

The estimate for the functions $\mathcal{T}_{h,1}$ and their derivatives are also obtained by the lemma above. Note that $\mathcal{T}_{h,1} = \langle f_w \rangle_{\varphi}$, as $\langle u_{h,1} \rangle_{\varphi} = 0$. Using the estimates above, we can estimate $u_{w,1}$, $u_{h,1}$, $u_{w,2}$, $\mathcal{T}_{h,2}$, $\mathcal{T}_{w,2}$ and their derivatives by Lemma 13.1. However, for the functions $\mathcal{T}_{h,2}$ and $\mathcal{T}_{w,2}$ themselves better estimates are obtained in sections 12.2 and 12.3.

To estimate the functions $\mathcal{T}_{h,3}$ and $\mathcal{T}_{w,3}$, we need to assume that

\[ h > C_h \varepsilon. \] (13.2)

The large enough constant $C_h > 0$ will be chosen below. It will be greater than the constant $C_{inw}$ from Lemma 3.2. By (10.1) we have the following system of equations (here $v = (h, w)$ and $A_v$, $A_w$ denote the right hand sides of (10.1)):

\[
(1 + \varepsilon \frac{\partial u_{w,1}}{\partial \varphi}) \mathcal{T}_{\varphi,2} + \varepsilon^2 \frac{\partial u_{w,1}}{\partial \varphi} \mathcal{T}_{w,3} = A_v,
\]

\[
(1 + \varepsilon \frac{\partial u_{w,1}}{\partial \varphi}) \mathcal{T}_{w,3} + \frac{\partial u_{w,1}}{\partial \varphi} \mathcal{T}_{\varphi,2} = A_w.
\]

From (13.2), (3.1) and the estimates on $u_{h,1}$ and $u_{h,2}$ for large enough $C_h$ we have

\[ h \in [0.5h, 2h]. \]

This allows us to estimate the intermediate values from (10.1) as if they were at the point $\bar{h}$. Using Table 1 and (13.2), we have $A_v = O(h^{-3} \ln^{-1} h)$, $A_w = O(h^{-3} \ln^{-1} h) + O(h^{-3} \ln^{-1} h)$. We can substitute the expression for $\mathcal{T}_{\varphi,3}$ from the second equation into the first one. This yields

\[
\mathcal{T}_{\varphi,2} \left( 1 + \varepsilon \frac{\partial u_{w,1}}{\partial \varphi} - \varepsilon^2 \frac{\partial u_{w,1}}{\partial \varphi} (1 + \varepsilon \frac{\partial u_{w,1}}{\partial \varphi}) \right) = A_v - \varepsilon^2 \frac{\partial u_{w,1}}{\partial \varphi} (1 + \varepsilon \frac{\partial u_{w,1}}{\partial \varphi}) A_w.
\]

From (13.2) and Table 1 we see that for large enough $C_h$

\[
\left\| \frac{\partial u_{w,1,2}}{\partial \varphi} \right\| = O(\ln h); \quad \left\| \frac{\partial u_{w,1,2}}{\partial \varphi} \right\| < 0.1, \quad \left\| \varepsilon \frac{\partial u_{w,1,2}}{\partial \varphi} \right\| = O(C_h^{-2} \ln^{-1} h).
\]
For large enough $C_h$ we have $\left| \frac{\partial u_{1,2}}{\partial \phi} - \varepsilon^2 \frac{\partial u_{1,2}}{\partial \psi} (1 + \varepsilon \frac{\partial u_{1,2}}{\partial \psi})^{-1} \frac{\partial u_{1,2}}{\partial \phi} \right| < 0.5$. Hence, we have $\mathcal{F}_{\varepsilon^2} = O(h^{-2} \ln^{-1} h) + O(h^{-2} \ln^{-2} h)$ (let us note that for $h > C_h, \varepsilon |\ln \varepsilon|^{0.5}$ we have $\left\| \varepsilon^2 \frac{\partial \mathcal{F}_{\varepsilon^2}}{\partial \varepsilon} \right\| = O(\ln^{-2} h)$ and this yields slightly better estimate $\mathcal{F}_{\varepsilon^2} = O(h^{-2} \ln^{-2} h)$). Then from the second equation we obtain $\mathcal{F}_{\varepsilon^3} = O(h^{-2}) + O(h^{-2} \ln^{-1} h)$.

Lemma 13.3. The estimates for the functions $\mathcal{F}_{\varepsilon^a}$ (for $a = h, w, i = 1, 2, 3$ and $a, \psi, w, \varepsilon$) and their derivatives stated in Table 1 also hold for the corresponding functions $u_{a,1}$ and their derivatives. Moreover, we have $|u_{a,1}(h, w, \varepsilon) - u_{a,1}(h, w, \psi)| = O(\varepsilon)$ for $a = h, w, i, \ldots, w_k$.

Proof. Recall that the expressions $\mathcal{F}_{\varepsilon^a}$ are computed by the same formulas as $\mathcal{F}_{\varepsilon^a}$, with the perturbation $f$ replaced by $f^0$. This means that the estimates we have for $\mathcal{F}_{\varepsilon^a}$ (they are valid for any smooth perturbation $f$) also hold for $\mathcal{F}_{\varepsilon^a}$. By (4.3) we have $f_{h,1} = \mathcal{F}_{h,1}^0, f_{w,1} = \mathcal{F}_{w,1}^0$ and $f_{w,1} = \mathcal{F}_{w,1}^0$, so for these expressions and their derivatives the lemma holds.

By (4.3) we also have $f_{h,2} = \mathcal{F}_{h,2}^0 + (f_{h,1}^0 (h, w, \varphi))_{\varphi}$. Similarly, to the estimate on $\mathcal{F}_{h,1}$ above ($\alpha$ is computed exactly as $\mathcal{F}_{h,1}$ if we start with $f^i$ instead of $f$), we have $\alpha = O(\ln^{-1} h), \frac{\partial\alpha}{\partial h} = O(h^{-1} \ln^{-2} h)$ and $\frac{\partial\alpha}{\partial w} = O(1)$. Therefore, the estimates for $\mathcal{F}_{h,2}$ and $\mathcal{F}_{w,2}$ from Table 1 also hold for $f_{h,2}$, the estimates for $\mathcal{F}_{w,2}$ are obtained in the same way.

We have $f_{\varepsilon^2} = \mathcal{F}_{\varepsilon^2} + \varepsilon^{-1} (\mathcal{F}_{\varepsilon^2}^0 - \mathcal{F}_{\varepsilon^2})$. Using (4.1), we get $\varepsilon^{-1} (\mathcal{F}_{\varepsilon^2}^0 - \mathcal{F}_{\varepsilon^2}) = (f_{\varepsilon^2}^0 + \varepsilon f_{\varepsilon^2}^1)_{\varphi}$, where $f_{\varepsilon^2}^0$ is the $\varphi$-component of $f^1$ written in the energy-angle coordinates. As the estimate for $\mathcal{F}_{h,1}$ holds for any smooth $f$, we can plug in $f^1 + \varepsilon f^2$ instead of $f$ and get the estimate $(f_{\varepsilon^2}^0 + \varepsilon f_{\varepsilon^2}^1) = O(h^{-1} \ln^{-2} h)$. As $f_{\varepsilon^2}^0$ and $f_{\varepsilon^2}^1$ are uniformly bounded by a constant independent of $\varepsilon$, one may check that this estimate is uniform in $\varepsilon$. Therefore, the estimate for $\mathcal{F}_{\varepsilon^2}$ also holds for $f_{\varepsilon^2}$.

Before estimating $f_{h,3}$, let us prove the second statement of the lemma. For $b = h, w, \varphi$ the map $U_t : f \rightarrow u_{1,1}$ is linear by (3.4) and (3.5). Hence, for $\psi = f^1 + \varepsilon f^2$ and $u_{h,1} = U(\psi)$ we have

$$u_{1,1}(h, w, \varphi, \varepsilon) = u_{0,1}(h, w, \varphi) + \varepsilon u_{1,1}^0(h, w, \varphi, \varepsilon).$$

As $\psi$ is smooth with respect to $p, q, z$ and uniformly bounded with respect to $\varepsilon$, the estimate $u_{1,1}(h, w, \varphi, \varepsilon) = O(1)$ for $a = h, w, \ldots, w_k$ also holds for $u_{1,1}^0$.

We have $f_{h,2} = \mathcal{F}_{h,2}^0 + (f_{h,1}^0 (h, w, \varphi))_{\varphi}$. Clearly, $(f_{h,1}^0)_{\varphi} = O(1)$. As we have $f_{h,2} = \sum_a (\partial f_{h,1}^0 / \partial a)_{u_a}$, $a = h, w, \ldots, w_k$ and the functions $u_{a,1}$ linearly depend on $f$, for $\psi = f^1 + \varepsilon f^2$ we can write

$$\mathcal{F}_{h,2} = \sum_a \left( \frac{\partial f_{h,1}^0}{\partial a} u_a^0 + \frac{\partial f_{h,1}^0}{\partial a} u_a^0 + \frac{\partial f_{h,1}^0}{\partial a} u_a^0 \right)_{\varphi}, \quad a = h, w, \ldots, w_k.$$

Here the upper index $\psi$ means that the function is obtained using $\psi$ instead of $f$. The estimates on $\frac{\partial f_{h,1}^0}{\partial a}, u_a$ from Table 1 are also valid for $\frac{\partial f_{h,1}^0}{\partial a}, u_a^0$. Using these estimates, we obtain $\varepsilon^{-1} (\mathcal{F}_{h,2} - \mathcal{F}_{h,2}^0) = O(h^{-1})$, thus proving the estimate for $f_{h,3}$. In a similar way we obtain $\varepsilon^{-1} (\mathcal{F}_{w,2} - \mathcal{F}_{w,2}^0) = O(h^{-1})$, thus proving the estimate for $f_{w,3}$. □

References

[1] V.I. Arnold. “Small denominators and problems of stability of motion in classical and celestial mechanics”. *Russ. Math. Surv* 18.6 (1963), pp. 85–191.

[2] F. J. Bourland, R. Haberman, and W. L. Kath. “Averaging methods for the phase shift of arbitrarily perturbed strongly nonlinear oscillators with an application to capture”. *SIAM Journal on Applied Mathematics* 51.4 (1991), pp. 1150–1167.

[3] J.R. Cary, D.F. Escande, and J.L. Tennyson. “Adiabatic-invariant change due to separatrix crossing”. *Physical Review A* 34.5 (1986), p. 4256.

[4] J.R. Cary and R.T. Skodje. “Phase change between separatrix crossings”. *Physica D: Nonlinear Phenomena* 36.3 (1989), pp. 287–316.

[5] J. Moser. “The analytic invariants of an area-preserving mapping near a hyperbolic fixed point”. *Communications on Pure and Applied Mathematics* 9.4 (1956), pp. 673–692.

[6] A.L. Neishtadt. “Averaging method for systems with separatrix crossing”. *Nonlinearity* 30.7 (2017), p. 2871.
[7] A.I. Neishtadt. “Change of an adiabatic invariant at a separatrix”. *Fizika plazmy* 12.8 (1986), p. 992. [Engl. transl.: Sov. J. Plasma Phys. 12, 568-573 (1986)].

[8] A.I. Neishtadt. “On the change in the adiabatic invariant on crossing a separatrix in systems with two degrees of freedom”. *Prikl. Mat. Mekh.* 51.5 (1987), pp. 750-757. [Engl. transl.: J. Appl. Math. Mech. 51, No. 5, 586-592 (1987)].

[9] A.I. Neishtadt, V.V. Sidorenko, and D.V. Treschev. “Stable periodic motions in the problem on passage through a separatrix”. *Chaos* 7.1 (1997), pp. 2–11.

[10] A.I. Neishtadt and A.A. Vasiliev. “Phase change between separatrix crossings in slow–fast Hamiltonian systems”. *Nonlinearity* 18.3 (2005), p. 1393.

[11] A.V. Timofeev. “On the constancy of an adiabatic invariant when the nature of the motion changes”. *Sov. Phys. — JETP* 48 (1978), p. 656.

[12] A.A. Vasiliev, A.I. Neishtadt, C. Simó, and D.V. Treschev. “Stability islands in domains of separatrix crossings in slow-fast Hamiltonian systems”. *Proceedings of the Steklov Institute of Mathematics* 259.1 (2007), p. 236.

Anatoly Neishtadt,
Department of Mathematical Sciences,
Loughborough University, Loughborough LE11 3TU, United Kingdom;
Space Research Institute, Moscow 117997, Russia
E-mail : a.neishtadt@lboro.ac.uk

Alexey Okunev,
Department of Mathematical Sciences,
Loughborough University, Loughborough LE11 3TU, United Kingdom
E-mail : a.okunev@lboro.ac.uk