Homoclinic orbits of twist maps and billiards

Amadeu Delshams and Rafael Ramírez-Ros
Departament de Matemàtica Aplicada I
Universitat Politècnica de Catalunya
Diagonal 647, 08028 Barcelona, Spain

Abstract

The splitting of separatrices for hyperbolic fixed points of twist maps with d degrees of freedom is studied through a real-valued function, called the Melnikov potential. Its non-degenerate critical points are associated to transverse homoclinic orbits and an asymptotic expression for the symplectic area between homoclinic orbits is given. Moreover, Morse theory can be applied to give lower bounds on the number of transverse homoclinic orbits.

This theory is applied first to elliptic billiards, where non-integrability holds for any non-trivial entire symmetric perturbation. Next, symmetrically perturbed prolate billiards with $d > 1$ degrees of freedom are considered. Several topics are studied about these billiards: existence of splitting, explicit computations of Melnikov potentials, existence of 8 or $8d$ transverse homoclinic orbits, exponentially small splitting, etc.

1 Introduction and plan of the paper

The phenomenon of the splitting of separatrices associated to a hyperbolic fixed point of a map has received a considerable attention, due to its direct relationship with the existence of chaotic motion nearby, as a consequence of the Smale-Birkhoff homoclinic theorem [Sma63, Wig90].

The existence of a recently developed *computable* Melnikov theory [DR96] for maps makes easier the computation of the Melnikov function. In the particular case of area preserving maps in the plane, the Melnikov function M is a periodic function with zero mean, and thus it is in fact the derivative of another periodic function, called the *Melnikov potential* L. The non-degenerate critical points of the Melnikov potential give rise to transverse homoclinic orbits. For analytic maps, the Melnikov function is doubly periodic, and complex variable theory can be used to compute the Melnikov potential.

A particular, but very important, example is provided by the billiard on an analytical convex table. A direct application of the Melnikov theory to perturbed elliptic tables provides explicit formulas for the lobes between separatrices, and also non-integrability for non-trivial perturbations.

The aim of this paper is to generalize such results to higher dimensional billiards. Since the motion inside a billiard can be modeled with the help of a *twist map*, we first develop a theory for twist maps on cotangent bundles with $d \geq 1$ degrees of freedom. Twist maps can be considered as the typical example of *exact symplectic maps*, for which there are several results due to the authors [DR97a] that will be applied along the present paper. Related ideas can be found in [Tre94, Bol95, Tab95b, Lom96b, Lom97].

When there is only one degree of freedom (on the plane), every branch of a coincident separatrix of the unperturbed twist map gives rise, in general, to two homoclinic orbits of the perturbed twist map.

When the number d of degrees of freedom is bigger than one, the (partial) coincidence of the invariant manifolds associated to a fixed hyperbolic point of the unperturbed twist map can take place in different ways. Thus, we will deal with *doubled*, *partially doubled* and *completely doubled* invariant manifolds for the unperturbed case. Different kinds of coincidences between invariant manifolds give rise to different kinds of *separatrices* and *bifurcation sets* and, consequently, to different results about the number of homoclinic orbits of the perturbed case. To avoid any kind of misunderstandings, the introduction of these concepts is carefully performed in section 2.

The main tool of this paper is the *Melnikov potential* L, a scalar function defined on the unperturbed separatrix, which is the natural splitting function for detecting primary homoclinic orbits in twist maps. Its non-degenerate critical points are associated to transverse homoclinic orbits. So, once located its non-degenerate critical points, everything is done. Several analytical results are developed in section 2. Moreover, the Melnikov potential is invariant under the action of the unperturbed twist map, and Morse theory is applied to the Melnikov function defined on a *reduced separatrix*, which turns out to be compact in the completely doubled case. The lower bound provided by the Morse theory on the number of homoclinic orbits is, in general, increased when there exist extra symmetries or reversors.

There is another interpretation of these results, based on variational methods, which allows us to introduce the concepts of *homoclinic action* and *homoclinic area*, as a generalization of the planar case.

These results about twist maps are readily applied to planar billiards in Section 3, providing non-integrability for the elliptic billiard under non-trivial perturbations, and a computable Melnikov potential. Several examples are reviewed.

For more degrees of freedom, in the present paper we do not consider an arbitrary ellipsoidal billiard as the unperturbed case, since the explicit expression of asymptotic motions is still not well-known. Instead, in Section 4 we consider only *prolate billiards*, that is, ellipsoids with all their axis of equal length except one, which is larger. For general non-degenerate perturbations, the Melnikov potential L is defined on $\mathbb{S}^1 \times \mathbb{S}^{d-1}$ for a billiard with d degrees of freedom, and the existence of at least 8 homoclinic orbits is provided by Morse theory. For *reversible* perturbations, the Melnikov potential L is defined on $\mathbb{S}^1 \times \mathbb{P}^{d-1}$, and this lower bound changes to $8d$. The Melnikov potential is explicitly computed for polynomial and quartic perturbations, showing that the lower bound about the number of the homoclinic orbits provided by the Morse theory is effectively attained.

2 General results for twist maps

For the sake of simplicity, we will assume that the objects here considered are smooth. For a general background on symplectic geometry we refer to [Arn76, GS77, AM78]. The basic properties of immersed submanifolds can be found in [GG73, pages 6–11]. More details about twist maps can be found in [Gol94a, Gol94b, BG96].

2.1 Introduction to twist maps

A *twist map* F is a map from a connected subset $\mathcal{P}$ of the cotangent bundle $T^*\mathcal{M}$ of a manifold $\mathcal{M}$ (not necessarily compact) into $\mathcal{P}$, which comes equipped with a twist generating function $\mathcal{L} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ that satisfies

$$F^*(w\,dz) - w\,dz = w'\,dz' - w\,dz = d\mathcal{L}(z, z'), \qquad (z', w') = F(z, w), \quad (2.1)$$

where (z, w) are any cotangent coordinates on $T^*\mathcal{M}$, that is, z are coordinates on $\mathcal{M}$, extended to coordinates (z, w) in the obvious way. The dimension d of the manifold $\mathcal{M}$ will be called the *number of degrees of freedom* of the twist map (2.1).

Condition (2.1) can also be written in a coordinate free manner. Given $\mathcal{L}$, one can retrieve the map (at least implicitly) from

$$w = -\partial_1\mathcal{L}(z, z'), \qquad w' = \partial_2\mathcal{L}(z, z').$$

This can be done globally (i.e., $\mathcal{P} = T^*\mathcal{M}$) only when $\mathcal{M}$ is diffeomorphic to a fiber of $T^*\mathcal{M}$, for example when $\mathcal{M}$ is the covering space of $\mathbb{T}^n$ or a manifold of constant negative curvature.

At this point, it is worth mentioning that an open connected subset $\mathcal{P}$ of a cotangent bundle $T^*\mathcal{M}$ is the typical example of an *exact symplectic manifold*, i.e., a $2d$-dimensional manifold $\mathcal{P}$ endowed with a *symplectic form* ω which is exact: $\omega = -\,d\phi$. Actually, the canonical symplectic forms ω, ϕ, on $T^*\mathcal{M}$ read in cotangent coordinates (z, w) as $\omega = dz \wedge dw$, $\phi = w\,dz$.

A twist map is the typical example of an *exact symplectic map*, i.e., a map $F : \mathcal{P} \to \mathcal{P}$ defined on an exact symplectic manifold $\mathcal{P}$, characterized by the equation $F^*\phi - \phi = dS$ for some function $S : \mathcal{P} \to \mathbb{R}$, called *generating function* of F. For $\mathcal{P} \subset T^*\mathcal{M}$, the fact that the generating function S can be written in terms of old and new coordinates: $S(z, w) = \mathcal{L}(z, z')$, is the *twist condition* that gives the name to the twist maps. Introducing the *canonical projection* $\pi : T^*\mathcal{M} \to \mathcal{M}$, the relation above between generating functions reads as $S(p) = \mathcal{L}(\pi(p), \pi(F(p)))$, for $p \in \mathcal{P}$. Since the example we have in mind (the billiard) can be modeled by a twist map, we will not consider exact symplectic maps anymore, and we refer the reader interested in such theory to [DR97a].

2.2 The unperturbed system

We are given a smooth twist diffeomorphism $F_0 : \mathcal{P} \to \mathcal{P}$, where $\mathcal{P}$ is an open connected subset of a cotangent bundle $T^*\mathcal{M}$. Let $\mathcal{L}_0$ be its twist generating function. We will assume that:

a) There exists a *hyperbolic fixed point* p^0_∞ of F_0.

b) The d-dimensional *unstable* and *stable* invariant manifolds associated to the hyperbolic fixed point p^0_∞,

$$\begin{aligned} \mathcal{W}^u_0 &:= \left\{ p \in \mathcal{P} : \lim_{k\to-\infty} {F_0}^k(p) = p^0_\infty \right\}, \\ \mathcal{W}^s_0 &:= \left\{ p \in \mathcal{P} : \lim_{k\to+\infty} {F_0}^k(p) = p^0_\infty \right\}, \end{aligned}$$

are *doubled*, that is, they coincide:

$$\mathcal{W} := \mathcal{W}^u_0 = \mathcal{W}^s_0.$$

This coincidence of invariant manifolds can take place in many different ways and has several implications upon the topology of the problem. We slow down here to introduce with full details the bifurcation set.

We recall that the invariant manifolds $\mathcal{W}^{u,s}_0$ need not to be submanifolds of $\mathcal{P} \subset T^*\mathcal{M}$, but just connected *immersed submanifolds*. More precisely, $\mathcal{W}^{u,s}_0 = g^{u,s}(\mathbb{R}^d)$ for some one-to-one immersions $g^{u,s} : \mathbb{R}^d \to \mathcal{P}$, such that $g^{u,s}(0) = p_\infty$ and $dg^{u,s}(0)[\mathbb{R}^d]$ is the tangent space to $\mathcal{W}^{u,s}_0$ at p^0_∞ [PM82, II §6].

By a *one-to-one immersion* $g^{u,s} : \mathbb{R}^d \to \mathcal{P}$, we mean that $dg(z)$ has maximal rank d at any point $z \in \mathbb{R}^d$, and that $g^{u,s}$ is one-to-one onto its image $\mathcal{W}^{u,s}_0 = g^{u,s}(\mathbb{R}^d)$. There is a natural way to make $\mathcal{W}^{u,s}_0$ a smooth manifold: the topology on $\mathcal{W}^{u,s}_0$ is the one which makes $g^{u,s}$ a homeomorphism and the charts on $\mathcal{W}^{u,s}_0 = g^{u,s}(\mathbb{R}^d)$ are the pull-backs via $(g^{u,s})^{-1}$ of the charts on $\mathbb{R}^d$.

Figure 1 shows an example of the *double loop* $\mathcal{W}^{u,s}_0 = g^{u,s}(\mathbb{R}^d)$ that takes place when both invariant manifolds are doubled (i.e., they coincide) on the plane ($d = 1$). At p^0_∞, the induced topology on the invariant curves $\mathcal{W}^{u,s}_0$ via the inclusion $\mathcal{W}^{u,s}_0 \subset \mathbb{R}^2$ is not the same as the induced one via $g^{u,s}$. Hence, $\mathcal{W}^{u,s}_0 \setminus \{p^0_\infty\}$ are submanifolds, but not $\mathcal{W}^{u,s}_0$. This situation is a particular case of the following elementary result [GG73, page 11].

Lemma 2.1 *Let $g : \mathbb{R}^d \to \mathcal{P}$ be a one-to-one immersion and set $\mathcal{W} = g(\mathbb{R}^d)$. Let $\Sigma \subset \mathcal{W}$ be the set of points where the two topologies on $\mathcal{W}$ (the one induced by the inclusion $\mathcal{W} \subset \mathcal{P}$ and the one that makes g a homeomorphism) differ. Then, $\Lambda = \mathcal{W} \setminus \Sigma$ is a submanifold of $\mathcal{P}$. Indeed, $\mathcal{W}$ is not a submanifold of $\mathcal{P}$* just *at the points of Σ.*

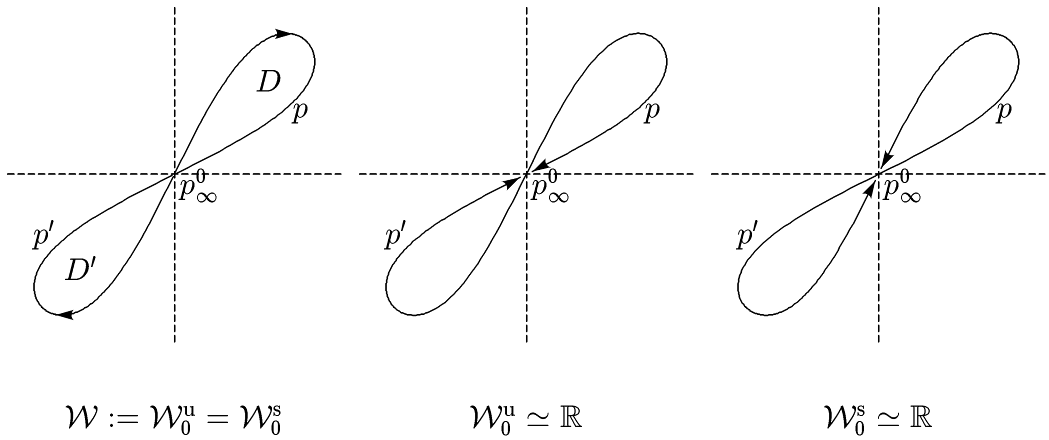

Figure 1: The invariant manifolds $\mathcal{W}_0^u$ and $\mathcal{W}_0^s$ are different as smooth manifolds, and are not submanifolds of $\mathbb{R}^2$. There exist no paths $\gamma^{u,s}$ in $\mathcal{W}_0^{u,s}$ from p to p' such that $\gamma^u = \gamma^s$.

We now recall that we are assuming that the invariant manifolds $\mathcal{W}_0^{u,s}$ are *doubled*, that is, $\mathcal{W} := \mathcal{W}_0^u = \mathcal{W}_0^s$.

Then, we can consider three topologies on $\mathcal{W}$: the one induced by the inclusion $\mathcal{W} \subset \mathcal{P}$, and the two ones induced by the inclusions $\mathcal{W} \subset \mathcal{W}_0^{u,s}$.

We define the *bifurcation set* Σ of this problem as the subset of $\mathcal{W}$ of points where the three topologies do not coincide, and we define the *separatrix* Λ as its complementary in $\mathcal{W}$, that is, $\Lambda := \mathcal{W} \setminus \Sigma$.

By Lemma 2.1, Λ is a submanifold of $\mathcal{P}$. Moreover, the fixed point p_∞^0 is not included in the separatrix Λ. Indeed, this property follows from the fact that $\mathcal{W}_0^u$ and $\mathcal{W}_0^s$ intersect transversely at p_∞^0, so their topology at p_∞^0 as immersed submanifolds can not coincide and $p_\infty^0 \in \Sigma$. Finally, let us note that both the separatrix and the bifurcation set are F_0-invariant, due to the fact that F_0 is a diffeomorphism. We summarize now these properties:

Lemma 2.2 *The bifurcation set Σ and the separatrix Λ have the following properties:*

(i) Λ *is a submanifold of* $\mathcal{P}$ *and* $p_\infty^0 \in \Sigma$.

(ii) Λ *and* Σ *are* F_0*-invariant.*

For simplicity, and due to the application to billiards, we have restricted ourselves to the case that both invariant manifolds of F_0 are doubled, and we have then defined the notion of separatrix. When $\mathcal{W}_0^{u,s}$ are *partially doubled*: $\mathcal{W}_0^u \neq \mathcal{W}_0^s$, but there exists a subset $\Lambda \subset \mathcal{W}_0^u \bigcap \mathcal{W}_0^s$ such that Λ is a d-dimensional submanifold of $\mathcal{P}$, invariant by F_0, and the three topologies on Λ coincide (the ones induced by the inclusions $\Lambda \subset \mathcal{P}$, $\Lambda \subset \mathcal{W}_0^u$ and $\Lambda \subset \mathcal{W}_0^s$), Λ can be taken as a separatrix of the problem, and the Melnikov potential is well defined on it.

2.3 Analytical results for the perturbed system

Consider a perturbed twist map F_ε, and let $\mathcal{L}_\varepsilon = \mathcal{L}_0 + \varepsilon \mathcal{L}_1 + O(\varepsilon^2)$ be the twist generating function of F_ε:

$$(z', w') = F_\varepsilon(z, w) \Longleftrightarrow w = -\partial_1 \mathcal{L}_\varepsilon(z, z'), \quad w' = \partial_2 \mathcal{L}_\varepsilon(z, z'). \tag{2.2}$$

For $0 < |\varepsilon| \ll 1$, there exists a hyperbolic fixed point p_∞^ε of F_ε, close to p_∞^0, with associated invariant manifolds $\mathcal{W}_\varepsilon^{u,s}$. It is not restrictive to normalize the twist generating function by imposing $\mathcal{L}_\varepsilon(z_\infty^\varepsilon, z_\infty^\varepsilon) = 0$, where $z_\infty^\varepsilon = \pi(p_\infty^\varepsilon)$. In particular, $\mathcal{L}_1(z_\infty^0, z_\infty^0) = 0$, where $z_\infty^0 = \pi(p_\infty^0)$.

We now introduce the *Melnikov potential* $L : \Lambda \longrightarrow \mathbb{R}$ by

$$L(p) = \sum_{k \in \mathbb{Z}} \mathcal{L}_1(z_k, z_{k+1}), \quad z_k = \pi(p_k), \ p_k = {F_0}^k(p), \ p \in \Lambda. \tag{2.3}$$

The series above is absolutely convergent since any F_0-orbit $\left({F_0}^k(p)\right)_{k \in \mathbb{Z}}$ in the manifold Λ tends to $p_\infty^0 = (z_\infty^0, w_\infty^0)$ at an exponential rate as $|k| \to \infty$ and $\mathcal{L}_1(z_\infty^0, z_\infty^0) = 0$. We list now some of the main properties of the Melnikov potential.

Theorem 2.1 *Under the above notations and hypotheses:*

a) $L : \Lambda \to \mathbb{R}$ *is well-defined, smooth and invariant under the action of the unperturbed map:* $LF_0 = L$.

b) The differential of the Melnikov potential $M = dL$ *(called the* Melnikov function*), measures, in first order in* ε*, the distance between the perturbed invariant manifolds* $\mathcal{W}_\varepsilon^{u,s}$.

c) If L is not locally constant, the manifolds $\mathcal{W}_\varepsilon^{u,s}$ split for $0 < |\varepsilon| \ll 1$, i.e., they do not coincide.

d) If $p \in \Lambda$ is a non-degenerate critical point of L, the manifolds $\mathcal{W}_\varepsilon^{u,s}$ are transverse along a primary homoclinic orbit $\mathcal{O}_\varepsilon$ of F_ε for $0 < |\varepsilon| \ll 1$, with $\mathcal{O}_0 = \left({F_0}^k(p)\right)_{k\in\mathbb{Z}}$. Moreover, when all the critical points of L are non-degenerate, all the primary homoclinic orbits arising from Λ are found in this way.

The proof of this theorem can be found in [DR97a]. We will restrict ourselves to point out some comments about it.

An essential (and hidden along the present paper) ingredient for the proof of Theorem 2.1 is the fact that the invariant manifolds $\mathcal{W}_\varepsilon^{u,s}$ are exact Lagrangian immersed submanifolds of $\mathcal{P}$. Actually, for *any* cotangent coordinates (x, y) adapted to $\mathcal{W}_0^{u,s}$—that is, in these coordinates the unperturbed invariant manifold $\mathcal{W}_0^{u,s}$ is given locally by $\{y = 0\}$ and the symplectic form ω reads as $x\,dy$—, the perturbed invariant manifold $\mathcal{W}_\varepsilon^{u,s}$ can be expressed locally in the form $y = \varepsilon\partial L_1^{u,s}(x)/\partial x + O(\varepsilon^2)$, for some well-defined smooth function $L_1^{u,s} : \mathcal{W}_0^{u,s} \to \mathbb{R}$ called *infinitesimal generating function* of the perturbed family $\{\mathcal{W}_\varepsilon^{u,s}\}$. Restricting the base points of the unperturbed invariant manifolds to the separatrix Λ where their smooth structures coincide, we can define a smooth function $L = L_1^u - L_1^s : \Lambda \to \mathbb{R}$, whose expression is given in (2.3). From the above discussion, it is obvious that the Melnikov potential is a geometrical object associated to the perturbation, whose differential $M = dL$ gives the first order distance, along the coordinate y in *any* cotangent coordinates (x, y) adapted to the separatrix Λ, between the perturbed invariant manifolds. The F_0-invariance of the Melnikov potential L is a trivial result from its expression, since a shift in the index of the sum does not change its value.

The rest of the properties of Theorem 2.1 follow readily from the properties stated above. By a *primary homoclinic orbit* of our perturbed problem we mean a perturbed homoclinic orbit $\mathcal{O}_\varepsilon \subset (\mathcal{W}_\varepsilon^u \bigcap \mathcal{W}_\varepsilon^s) \setminus \{p_\infty^\varepsilon\}$ of F_ε, defined for $|\varepsilon|$ small enough and depending in a smooth way on ε. These are the kind of orbits that a perturbative theory based on the Melnikov potential can detect.

2.4 Topological results for the perturbed system

Since the transverse homoclinic orbits detected by the Melnikov method are in fact associated to non-degenerate critical points of the Melnikov potential $L : \Lambda \to \mathbb{R}$, Morse theory can be applied to L to provide *lower bounds* on the number of transverse primary homoclinic orbits.

We recall again that we are assuming that the invariant manifolds $\mathcal{W}_0^{u,s}$ are *doubled*, that is,

$$\mathcal{W} := \mathcal{W}_0^u = \mathcal{W}_0^s,$$

and that the separatrix is defined by $\Lambda := \mathcal{W} \backslash \Sigma$, where the bifurcation set Σ is the subset of $\mathcal{W}$ of points where there is a coincidence of the topology induced by the inclusion $\mathcal{W} \subset T^*\mathcal{M}$ and the two ones induced by the inclusions $\mathcal{W} \subset \mathcal{W}_0^{u,s}$.

The separatrix Λ is not a compact submanifold of $\mathcal{P}$. However, by the invariance of L under the action of the unperturbed map F_0, it turns out that the Melnikov potential can de defined on the quotient manifold $\Lambda^* := \Lambda/F_0$, consisting of unperturbed homoclinic orbits of Λ. The quotient manifold $\Lambda^* := \Lambda/F_0$ will be called the *reduced separatrix* (of the unperturbed map).

In general, Λ^* needs not to be compact. One way to ensure that Λ^* is a compact manifold, is by assuming that the bifurcation set is minimal, i.e., $\Sigma = \{p_\infty\}$. (Remember that the hyperbolic fixed point p_∞ is always contained in the bifurcation set Σ, see Lemma 2.2.) This hypothesis is equivalent to require that the separatrix is $\Lambda = \mathcal{W}_0^{u,s} \setminus \{p_\infty\}$. We will say that the invariant manifolds are *completely doubled* in this case.

It is worth remarking that in the planar case with a double loop (∞), the bifurcation set is just the hyperbolic fixed point, i.e., if the invariant manifolds are doubled, then they are completely doubled. But, in general, for more dimensions the situation is not so simple.

For example, let $F_0 : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$, $d > 1$, be the product of d planar maps $f_j : \mathbb{R}^2 \to \mathbb{R}^2$, each one with a double loop $\Gamma_j = \{p_\infty^j\} \cup \Lambda_j$ where $p_\infty^j \in \mathbb{R}^2$ stands for the fixed point of f_j and Λ_j are the two components of $\Gamma_j \setminus \{p_\infty^j\}$, for $j = 1, \ldots, d$. Then, $\Lambda = \Lambda_1 \times \cdots \times \Lambda_d$ has 2^d connected components and $\Sigma = (\Gamma_1 \times \cdots \times \Gamma_d) \setminus \Lambda$ contains *strictly* the hyperbolic fixed point $p_\infty^0 = (p_\infty^1, \ldots, p_\infty^d) \in \mathbb{R}^{2d}$. In particular, $\Lambda \neq \mathcal{W}_0^{u,s} \setminus \{p_\infty\}$, and each connected component of the reduced separatrix Λ^* is homeomorphic to $\mathbb{S}^1 \times \mathbb{R}^{d-1}$. In this example, the invariant manifolds are doubled but not completely doubled, and the reduced separatrix is not compact.

For completely doubled invariant manifolds, we now state a result about the number of primary homoclinic orbits that persist under a general perturbation.

Theorem 2.2 *If the invariant manifolds are* completely doubled*,* Λ^* *is a compact manifold without boundary. Assume that* $L : \Lambda^* \to \mathbb{R}$ *is a Morse function. Then, the number of primary homoclinic orbits is at least* 4*.*

We recall that a real-valued smooth function over a compact manifold without boundary is called a *Morse function* when all its critical points are non-degenerate. It is very well-known that the set of Morse functions is open and dense in the set of real-valued smooth functions [Hir76, page 147]. Thus, to be a Morse function is a condition of generic position.

The proof of the theorem above can be found, again, in [DR97a], along with the exact topological characterization of the manifold Λ^*.

It is very important to notice that additional *symmetries* $I : \mathcal{P} \to \mathcal{P}$ of the map F_ε—that is, involutions I such that $F_\varepsilon I = I F_\varepsilon$ and $I^*\phi = \phi$—can give rise to new invariances for the Melnikov potential: $LI = L$. In particular, L can be considered as a function over the quotient manifold $\Lambda^*_I := \Lambda/\{F_0, I\}$, and the number of homoclinic orbits provided by the Morse theory may be increased. Instead of describing here a general theory, we will apply directly this idea to our examples.

2.5 Variational results for the perturbed system

We finish this account of general results by introducing very briefly some variational results.

There exists a variational principle, due to MacKay, Meiss and Percival [MMP84, Eas91], which establishes that the homoclinic orbits of the perturbed twist map (2.2) are the extremals of the *homoclinic action*

$$W[\mathcal{O}] := \sum_{k\in\mathbb{Z}} \mathcal{L}_\varepsilon(z_k, z_{k+1}), \qquad \mathcal{O} = (z_k)_{k\in\mathbb{Z}},$$

and a *homoclinic area* can be defined for every pair of homoclinic orbits $\mathcal{O}$, $\mathcal{O}'$, and is given by the difference of homoclinic actions

$$\Delta W[\mathcal{O}, \mathcal{O}'] = W[\mathcal{O}] - W[\mathcal{O}'].$$

For a motivation of these names, we again refer the reader to [DR97a]. There, one can see that given $p \in \mathcal{O}$, $p' \in \mathcal{O}'$, and paths $\gamma^{u,s}$ from p to p' *in* $\mathcal{W}_\varepsilon^{u,s}$, such that there exists an oriented 2-chain D with $\partial D = \gamma := \gamma^u - \gamma^s$, then

$$\Delta W[\mathcal{O}, \mathcal{O}'] = \oint_\gamma w\, dz = -\iint_D dz \wedge dw. \tag{2.4}$$

In this formula, by a path $\gamma^{u,s}$ *in* the immersed submanifold $\mathcal{W}_\varepsilon^{u,s}$ we mean that $\gamma^{u,s}$ is contained in $\mathcal{W}_\varepsilon^{u,s}$ and it is *continuous in the topology of* $\mathcal{W}_\varepsilon^{u,s}$. For example, if γ is any of the loops of figure 1, it is a closed path in $\mathbb{R}^2$ but not in $\mathcal{W}$.

The formula above shows clearly that the homoclinic area is a *symplectic invariant*, i.e., it neither depends on the symplectic coordinates used, nor on the choice of the symplectic potential $w\, dz$. The homoclinic action can be considered as the homoclinic area between the homoclinic orbit at hand and the "orbit" of the fixed point p_∞^ε. Thus, it is a symplectic invariant, too.

In particular, if $\mathcal{P} = \mathbb{R}^2 = T^*\mathbb{R}$ with the standard area as the symplectic structure, and $p \in \mathcal{O}$, $p' \in \mathcal{O}'$ are consecutive intersections of the invariant manifolds, then the homoclinic area $\Delta W[\mathcal{O}, \mathcal{O}']$ is simply the (algebraic) area of the associated lobe.

We end this section noting that in terms of the Melnikov potential, there is also a nice expression for the homoclinic action and the homoclinic area. The proof, again, can be found in [DR97a].

Theorem 2.3 *Let $\mathcal{O}_\varepsilon$ be a primary homoclinic orbit with $\mathcal{O}_0 = \left({F_0}^k(p)\right)_{k\in\mathbb{Z}}$ for some $p \in \Lambda$. Then, the homoclinic action admits the asymptotic expression $W[\mathcal{O}_\varepsilon] = W[\mathcal{O}_0] + \varepsilon L(p) + O(\varepsilon^2)$. Given another orbit $\mathcal{O}'_\varepsilon$ such that $\mathcal{O}'_0 = \left({F_0}^k(p')\right)_{k\in\mathbb{Z}}$ for some p' in the same connected component of Λ as p, the homoclinic area is given by*

$$\Delta W[\mathcal{O}_\varepsilon, \mathcal{O}'_\varepsilon] = \varepsilon[L(p) - L(p')] + O(\varepsilon^2).$$

3 Planar billiards

3.1 Convex billiards

Let us consider the problem of the "convex billiard table" [Bir27a, Bir27b]: let C be a smooth closed convex curve of the plane $\mathbb{R}^2$, parameterized by

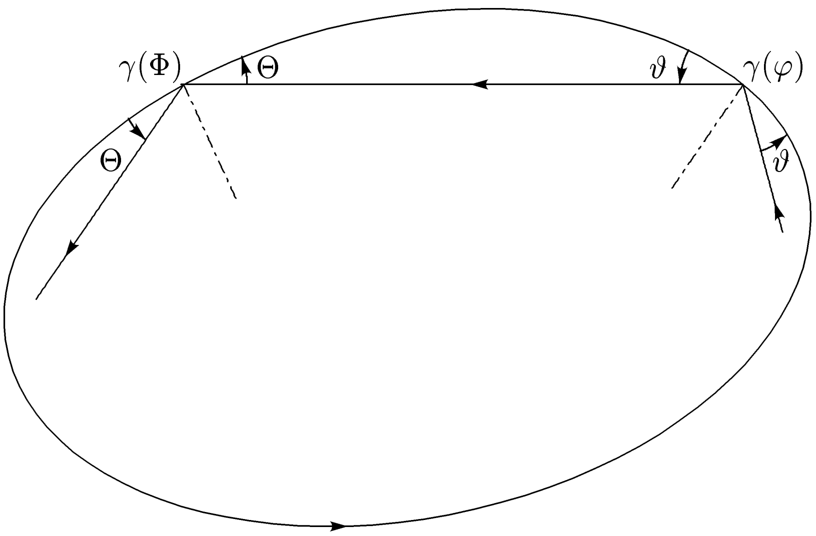

Figure 2: $T(\varphi, v) = (\Phi, V)$, where $v = |\dot\gamma(\varphi)|\cos\vartheta$ and $V = |\dot\gamma(\Phi)|\cos\Theta$.

$\gamma : \mathbb{T} \longrightarrow C$, where $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$, in such a way that C is traveled counterclockwise. Suppose that a material point moves inside C and collides elastically with C according to the law "the angle of incidence is equal to the angle of reflection". Such discrete dynamical systems can be modeled by a smooth twist map (called *billiard map*) T defined on a subset $\mathcal{P}$ of the cotangent bundle of $\mathbb{T}$, that is, the annulus $\mathbb{A} := T^*\mathbb{T} = \mathbb{T} \times \mathbb{R}$.. This subset is defined as

$$\mathcal{P} := \{(\varphi, v) \in \mathbb{A} : \, |v| < |\dot\gamma(\varphi)|\},$$

where the coordinate φ is the parameter on C, and $v = |\dot\gamma(\varphi)|\cos\vartheta$, where $\vartheta \in (0, \pi)$ is the angle of incidence-reflection of the material point. In this way, we obtain the map $T : \mathcal{P} \longrightarrow \mathcal{P}$ given by $(\varphi, v) \longmapsto (\Phi, V)$ that models the billiard (see Figure 2).

This map T is a twist map with one degree of freedom, with

$$G : \{(\varphi, \Phi) \in \mathbb{T} \times \mathbb{T} : \, \varphi \neq \Phi\} \longrightarrow \mathbb{R}, \qquad G(\varphi, \Phi) := |\gamma(\varphi) - \gamma(\Phi)|$$

as its *twist generating function*, since

$$T(\varphi, v) = (\Phi, V) \Longleftrightarrow v = -\partial_1 G(\varphi, \Phi), \qquad V = \partial_2 G(\varphi, \Phi).$$

Indeed, the left-to-right implication is simply a computation:

$$\begin{aligned}
\partial_1 G(\varphi,\Phi) &= \frac{\langle \gamma(\varphi)-\gamma(\Phi),\dot\gamma(\varphi)\rangle}{|\gamma(\varphi)-\gamma(\Phi)|} = -|\dot\gamma(\varphi)|\cos\vartheta = -v, \\
\partial_2 G(\varphi,\Phi) &= \frac{\langle \gamma(\Phi)-\gamma(\varphi),\dot\gamma(\Phi)\rangle}{|\gamma(\varphi)-\gamma(\Phi)|} = |\dot\gamma(\Phi)|\cos\Theta = V,
\end{aligned}$$

whereas the right-to-left one follows from the convexity hypothesis on C.

It is geometrically clear that if C' is another closed convex curve obtained from C by a translation plus a homothety plus an orthogonal linear map (that is, a similarity) then its associated billiard map T' is conjugated to T, and so they are equivalent from a dynamical point of view. We will take advantage of this property, working in the space of smooth closed convex curves modulo similarities.

The billiard map T has no fixed points but it has two-periodic orbits, corresponding to opposite points with the "maximum" and "minimum" distance between them. In these orbits the angle of incidence-reflection is $\pi/2$ and thus $v=0$.

Instead of studying them as fixed points of T^2, we introduce the following simplification, as is usual in the literature [LT93, Tab94, DR96, Lom96a]. We will assume that C is *symmetric* with regard to a point. Modulo a similarity, we can assume that this point is the origin:

$$C = -C.$$

Consequently, we can choose a parameterization γ of C such that satisfies $\gamma(\varphi+\pi)=-\gamma(\varphi)$, in such a way that the two-periodic orbits are of the form $\{(\varphi_0,0),(\varphi_0+\pi,0)\}$, that is, two opposite points over C. Then, the billiard map T and the involution

$$S:\mathcal{P}\to\mathcal{P}, \qquad S(\varphi,v):=(\varphi+\pi,v),$$

commute. This allows us to introduce the *symmetric billiard map*

$$F:\mathcal{P}\to\mathcal{P}, \qquad F:=ST,$$

so that those two-periodic points for T are fixed points for F. Moreover, the dynamics of F and T are equivalent, since $F^2=T^2$.

The map F is also a twist map, with

$$\mathcal{L} : \{(\varphi, \Phi) \in \mathbb{T} \times \mathbb{T} : \varphi \neq \Phi + \pi\} \to \mathbb{R}, \qquad \mathcal{L}(\varphi, \Phi) := |\gamma(\varphi) + \gamma(\Phi)| \quad (3.1)$$

as its twist generating function, since $\gamma(\Phi + \pi) = -\gamma(\Phi)$.

Thus, given a sequence $(p_n)_{n\in\mathbb{Z}}$ such that $p_n = (\varphi_n, v_n) \in \mathcal{P}$, we have that $(p_n)_{n\in\mathbb{Z}}$ is an orbit of F if and only if

$$v_n = -\partial_1\mathcal{L}(\varphi_n, \varphi_{n+1}) = \partial_2\mathcal{L}(\varphi_{n-1}, \varphi_n), \quad \forall n \in \mathbb{Z}. \quad (3.2)$$

This leads us to the following *variational principle*: the orbits of the symmetric billiard map F are in one-to-one correspondence with the critical configurations of the functional (called the *action*)

$$W : \mathbb{T}^{\mathbb{Z}} \to \mathbb{R}, \qquad W[(\varphi_n)_{n\in\mathbb{Z}}] := \sum_{n\in\mathbb{Z}} \mathcal{L}(\varphi_n, \varphi_{n+1}),$$

that is, with the configurations $(\varphi_n)_{n\in\mathbb{Z}} \subset \mathbb{T}$ such that

$$\partial_k W[(\varphi_n)_{n\in\mathbb{Z}}] = \partial_1\mathcal{L}(\varphi_k, \varphi_{k+1}) + \partial_2\mathcal{L}(\varphi_{k-1}, \varphi_k) = 0, \qquad \forall k \in \mathbb{Z}.$$

(Note that although the series for W is in general not convergent, $\partial_k W$ involves only two terms of the series, and therefore ∇W is well defined.) The orbit $(p_n)_{n\in\mathbb{Z}}$ of F can be found from the critical configuration $(\varphi_n)_{n\in\mathbb{Z}}$ of W by using relation (3.2).

Thus, having a twist generating function allows us to work with only half of the coordinates (the *base* coordinates, i.e., the φ's). The *fiber* coordinates (i.e., the v's) are superfluous. We can also work with the coordinate $q = \gamma(\varphi)$ of the impact points on the curve C. We will use indistinctly the p-notation ($p = (\varphi, v) \in \mathcal{P}$), the φ-notation ($\varphi \in \mathbb{T}$), or the q-notation ($q \in C$).

To end the discussion about convex billiards, let us introduce the involution

$$R : \mathcal{P} \to \mathcal{P}, \qquad R(\varphi, v) := (\varphi, -v),$$

which is a *reversor* of the symmetric billiard map F, that is, $F^{-1} = RFR$.

The reversor R and the symmetry S can be interpreted as follows: given an orbit $(q_n)_{n\in\mathbb{Z}}$ of the symmetric billiard map F (respectively, the billiard map T), $(q_{-n})_{n\in\mathbb{Z}}$, $(-q_n)_{n\in\mathbb{Z}}$, and $(-q_{-n})_{n\in\mathbb{Z}}$ are also orbits of F (respectively, T), see Table 3.1. These four orbits are all different, except in the trivial cases of fixed points or two-periodic orbits. Besides, the image of a homoclinic orbit by R, S or RS is another homoclinic orbit. As is usual, we will use this property to save work in looking for the set of primary homoclinic orbits.

	Id	R	S	RS
$\mathcal{P}$	$(p_n)_{n\in\mathbb{Z}}$	$(Rp_n)_{n\in\mathbb{Z}}$	$(Sp_n)_{n\in\mathbb{Z}}$	$(RSp_n)_{n\in\mathbb{Z}}$
$\mathbb{T}$	$(\varphi_n)_{n\in\mathbb{Z}}$	$(\varphi_{-n})_{n\in\mathbb{Z}}$	$(\varphi_n+\pi)_{n\in\mathbb{Z}}$	$(\varphi_{-n}+\pi)_{n\in\mathbb{Z}}$
C	$(q_n)_{n\in\mathbb{Z}}$	$(q_{-n})_{n\in\mathbb{Z}}$	$(-q_n)_{n\in\mathbb{Z}}$	$(-q_{-n})_{n\in\mathbb{Z}}$

Table 1: The effect of the reversor R and the symmetry S on an orbit $(p_n)_{n\in\mathbb{Z}}$ of F. Here $p_n = (\varphi_n, v_n) \in \mathcal{P} \subset \mathbb{T}\times\mathbb{R}$ and $q_n = \gamma(\varphi_n) \in C$.

3.2 Elliptic billiards

The simplest example of closed convex curves are the ellipses. Among them, the circumferences are very degenerate for a billiard, since they have a one-parametric family of two-periodic orbits. So, let us consider now a *non-circular ellipse*:

$$C_0 := \left\{ \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 \right\} = \{\gamma_0(\varphi) = (\alpha\cos\varphi, \beta\sin\varphi) : \varphi \in \mathbb{T}\},$$

with $\alpha^2 \neq \beta^2$. Modulo a similarity, we can assume that $\alpha^2 - \beta^2 = 1$. Thus $\alpha > 1$, $\beta > 0$, the foci of the ellipse are $(\pm 1, 0)$, and the eccentricity is $e = 1/\alpha$. Let us denote $T_0 : \mathcal{P} \longrightarrow \mathcal{P}$ the twist map associated to the ellipse C_0, and $F_0 = S \circ T_0$. The billiard map T_0 is called *elliptic billiard.*

The points $p^r_\infty = (0,0)$ and $p^l_\infty = (\pi, 0)$ form a two-periodic orbit for T_0 that corresponds to the (right and left) vertexes $(\pm\alpha, 0)$ of the ellipse, and hence they are fixed points for F_0. We will check that these two fixed points are hyperbolic with four separatrices connecting them. Thus, we are really dealing with heteroclinic connections. Nevertheless, all the results about the homoclinic case in the previous section can be applied to the symmetric billiard problem. This is due to the fact that we can consider the variable φ defined modulo π in the symmetric case, using the symmetry $S(\varphi, v) = (\varphi + \pi, v)$. Then, the fixed points p^r_∞ and p^l_∞ become the same fixed point, so that the previous connections can be considered homoclinic ones.

Let us recall that a *caustic* is a smooth curve with the following property: if at least one of the segments (or its prolongation) of the polygonal trajectory of the point is tangent to the curve, then all the other segments (or their

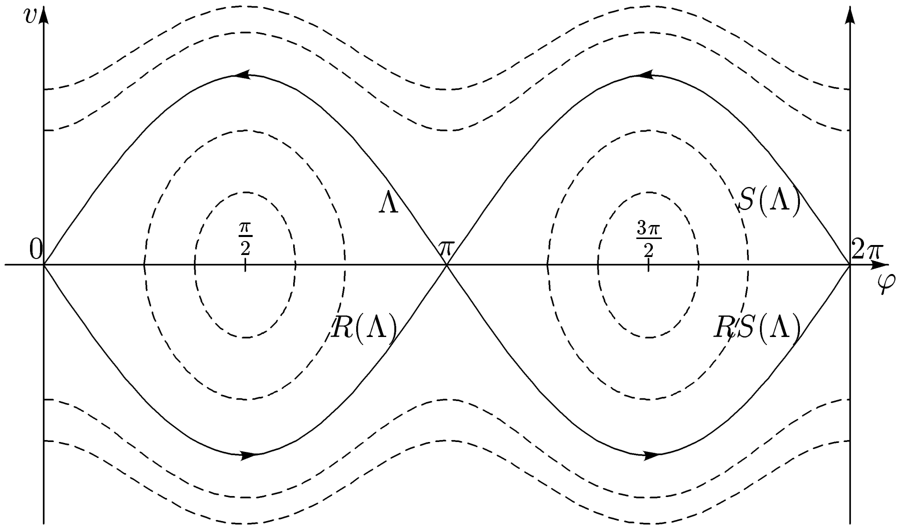

Figure 3: Phase portrait of F_0. Λ, $R(\Lambda)$, $S(\Lambda)$ and $RS(\Lambda)$ are the four separatrices.

prolongations) are tangent to the curve. It is a very well-known fact that all the orbits of an elliptic billiard have a caustic, and actually the caustics are just the family of confocal conics to C_0 (little Poncelet's theorem [KT91, Tab95a]).

This property indicates the integrability of elliptic billiards since the existence of caustics reflects some stability in the system. In fact, it is not difficult to obtain an explicit expression for a first integral of the elliptic billiard in (φ, v) coordinates [Lom96a]. Under the assumption $\alpha^2 - \beta^2 = 1$, a first integral is $I(\varphi, v) = v^2 - \sin^2 \varphi$. As a consequence, the level sets $\{I = \kappa\}_{-1 \leq \kappa < \beta^2}$ are invariant for T_0 and F_0. Thus, the phase portrait of the symmetric billiard map F_0 can be easily obtained, see figure 3.

The main properties of F_0 are listed in the following lemma.

Lemma 3.1 *Let $h > 0$ be determined by the equations*

$$\alpha = \coth(h/2), \quad \beta = cosech(h/2), \quad e = \tanh(h/2). \tag{3.3}$$

a) The points $p^r_\infty = (0,0)$ and $p^l_\infty = (\pi, 0)$ are hyperbolic fixed points of *the symmetric billiard map* F_0*, with* h *as their* characteristic exponent, *that is,* $Spec[\,dF_0(p^{l,r}_\infty)] = \left\{e^h, e^{-h}\right\}$.

b) Let $\mathcal{W}^{u,s}_0(p^{l,r}_\infty)$ *be the unperturbed unstable and stable invariant curves of* F_0 *at* $p^{l,r}_\infty$*. Then,* $\mathcal{W}^s_0(p^{l,r}_\infty) = \mathcal{W}^u_0(p^{r,l}_\infty)$*. Thus,* F_0 *has exactly* four *separatrices (heteroclinic connections):*

$$\begin{array}{rcl} \Lambda & = & \{(\varphi, \sin\varphi) : \varphi \in (0,\pi)\}, \\ R(\Lambda) & = & \{(\varphi, -\sin\varphi) : \varphi \in (0,\pi)\}, \\ S(\Lambda) & = & \{(\varphi, -\sin\varphi) : \varphi \in (\pi,2\pi)\}, \\ RS(\Lambda) & = & \{(\varphi, \sin\varphi) : \varphi \in (\pi,2\pi)\}. \end{array}$$

c) Let $p_0 = (\varphi_0, v_0) : \mathbb{R} \to \Lambda$ *be the diffeomorphism defined by*

$$\varphi_0(t) = \arccos(\tanh t), \qquad v_0(t) = \sin\varphi_0(t) = sech\, t.$$

Then, $p_0(t)$ *is a* natural parameterization *of* Λ*:* $F_0(p_0(t)) = p_0(t+h)$*. Moreover, the natural parameterizations of* $R(\Lambda)$*,* $S(\Lambda)$*, and* $RS(\Lambda)$ *are* $R(p_0(-t))$*,* $S(p_0(t))$*, and* $RS(p_0(-t))$*.*

d) Let $\Phi_0(t) = \varphi_0(t+h)$*. Then,*

$$\beta \frac{\sin\varphi_0(t) + \sin\Phi_0(t)}{|\gamma_0(\varphi_0(t)) + \gamma_0(\Phi_0(t))|} = sech(t + h/2). \tag{3.4}$$

Proof. It is only sketched here. More details can be found in [DR96].

a) We know that p^r_∞ and p^l_∞ are fixed points for F_0. Let

$$\begin{array}{rcl} \mathcal{L}_0(\varphi, \Phi) & = & |\gamma_0(\varphi) + \gamma_0(\Phi)| \\ & = & 2\alpha + \dfrac{(\alpha^2-1)\varphi\Phi - (\alpha^2+1)(\varphi^2+\Phi^2)/2}{2\alpha} + O_3(\varphi,\Phi) \end{array}$$

be the twist generating function of $F_0(\varphi, v) = (\Phi, V)$, where we have used that $\alpha^2 - \beta^2 = 1$. From the implicit equations of F_0 generated by $\mathcal{L}_0$ we get

$$\begin{array}{rcl} trace[\,dF_0(p^r_\infty)] & = & \partial_1\Phi(0,0) + \partial_2 V(0,0) \\ & = & -[\partial_{11}\mathcal{L}_0(0,0) + \partial_{22}\mathcal{L}_0(0,0)]/\partial_{12}\mathcal{L}_0(0,0), \end{array}$$

and a straightforward calculus yields $trace[\, dF_0(p^r_\infty)] = 2(\alpha^2+1)/(\alpha^2-1)$. Moreover, $\det[\, dF_0] \equiv 1$. Thus $\lambda = (\alpha+1)/(\alpha-1) > 1$ is an eigenvalue of $dF_0(p^r_\infty)$, and (3.3) implies $\lambda = e^h$. The proof for p^l_∞ is analogous.

b) This is a direct consequence of the conservation of the first integral I.

c) A tedious (but elementary) computation shows that

$$\partial_1 \mathcal{L}_0(\varphi_0(t), \varphi_0(t+h)) + \partial_2 \mathcal{L}_0(\varphi_0(t-h), \varphi_0(t)) \equiv 0.$$

Thus, the configurations $(\varphi_n)_{n\in\mathbb{Z}} \subset \mathbb{T}$, $\varphi_n = \varphi_0(t+hn)$, are critical points of the action $W_0[(\varphi_n)] = \sum_{n\in\mathbb{Z}} \mathcal{L}_0(\varphi_n, \varphi_{n+1})$, and therefore, by the above-mentioned variational principle, the sequences $(p_n)_{n\in\mathbb{Z}}$, $p_n = p(t+hn)$, are orbits of F_0. This proves that $p(t)$ is a natural parameterization of Λ.

The final part of (c) follows from the equalities $FR = RF^{-1}$ and $FS = SF$.

d) It is another cumbersome computation. □

3.3 Non-integrability of billiards close to ellipses

Birkhoff conjectured that the elliptic billiard is the only integrable smooth convex billiard. Our goal is to see that this is *locally* true for symmetric entire perturbations. Concretely, we shall prove that any non-trivial symmetric entire perturbation of an ellipse is non-integrable. (Roughly speaking, a perturbation of an ellipse will be called trivial when it is again an ellipse.)

To begin with, let us consider an arbitrary symmetric smooth perturbation $C_\varepsilon = -C_\varepsilon$ of the ellipse C_0. Modulo $O(\varepsilon^2)$ terms (which do not play any rôle in our first order analysis) and a similarity, C_ε can be put in the following *parameterized (normal) form*

$$C_\varepsilon = \{\gamma_\varepsilon(\varphi) = (\alpha\cos\varphi, [1+\varepsilon\eta(\varphi)]\beta\sin\varphi) : \varphi\in\mathbb{T}\}, \tag{3.5}$$

for some smooth π-periodic function $\eta(\varphi)$, or in the following *implicit form*

$$C_\varepsilon = \left\{(x,y)\in\mathbb{R}^2 : \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 + 2\varepsilon P(\alpha^{-1}x, \beta^{-1}y)\right\}, \tag{3.6}$$

for some even smooth function $P(u,w)$ such that $P(1,0) = \partial_2 P(1,0) = 0$.

The parameterized form can be considered a normal form for C_ε, whereas the implicit form cannot, since $\eta(\varphi)$ is completely determined by the perturbation, whereas different functions $P(u,w)$ can give rise to the same perturbation C_ε. Because of it, we have preferred to deal with the parameterized form, instead of the implicit one. This does not imply loss of generality, since it is easy to check that the connection between them is simply

$$P(\cos\varphi, \sin\varphi) = \eta(\varphi)\sin^2\varphi. \tag{3.7}$$

The parameterized normal form (3.5) shows that C_ε is an ellipse (up to $O(\varepsilon^2)$ terms, of course) if and only if the function $\eta(\varphi)$ is constant. As a consequence, we will say that C_ε is a *non-trivial (at order 1) symmetric entire perturbation* of the ellipse C_0 if and only if $\eta(\varphi)$ is a non-constant π-periodic entire function.

Let $T_\varepsilon : \mathcal{P} \to \mathcal{P}$ be the billiard map associated to the curve C_ε, and $F_\varepsilon = ST_\varepsilon$ the symmetric billiard map. We note that the hyperbolic fixed points $p_\infty^{l,r}$ are preserved by the perturbations (3.5) and (3.6): $F_\varepsilon(p_\infty^{l,r}) = p_\infty^{l,r}$, since $(\pm\alpha, 0)$ are still the more distant points on the perturbed ellipse C_ε.

For $|\varepsilon| \ll 1$, C_ε is a convex closed curve, and thus F_ε is a twist map, with

$$\mathcal{L}_\varepsilon(\varphi, \Phi) = |\gamma_\varepsilon(\varphi) + \gamma_\varepsilon(\Phi)| = \mathcal{L}_0(\varphi, \Phi) + \varepsilon\mathcal{L}_1(\varphi, \Phi) + O(\varepsilon^2)$$

as its twist generating function, where

$$\begin{aligned} \mathcal{L}_0(\varphi, \Phi) &= |\gamma_0(\varphi) + \gamma_0(\Phi)|, \\ \mathcal{L}_1(\varphi, \Phi) &= \beta^2 \frac{\sin\varphi + \sin\Phi}{|\gamma_0(\varphi) + \gamma_0(\Phi)|}[\sin\varphi\,\eta(\varphi) + \sin\Phi\,\eta(\Phi)]. \end{aligned} \tag{3.8}$$

Using the natural parameterization provided by Lemma 3.1, the formula of $\mathcal{L}_1(\varphi, \Phi)$ given in equation (3.8), and the formula (3.4), the Melnikov potential of our perturbed billiard problem (on the separatrix Λ) is $L(t) = \sum_{n\in\mathbb{Z}} g(t + hn)$, where

$$\begin{aligned} g(t) &= \mathcal{L}_1(\varphi_0(t), \varphi_0(t+h)) \\ &= \beta\, sech(t + h/2)[sech(t)\eta(\varphi_0(t)) + sech(t+h)\eta(\varphi_0(t+h))]. \end{aligned}$$

We have taken t as the coordinate over the separatrix Λ.

Before proceeding to study the Melnikov potential, it is very convenient to arrange the sum $\sum_{n\in\mathbb{Z}} g(t + hn)$, and express the Melnikov potential in the following way:

$$\begin{aligned} L(t) &= \sum_{n\in\mathbb{Z}} f(t + hn), &(3.9)\\ f(t) &= \beta[v_+(t) + v_-(t)]\delta(t) = 2\alpha v_+(t)v_-(t)\eta(\varphi_0(t)), \end{aligned}$$

where

$$v_0(t) := \sin\varphi_0(t) = sech\, t, \quad v_\pm(t) := v_0(t \mp h/2), \quad \delta(t) := v_0(t)\eta(\varphi_0(t)).$$

Now, assume we are given a non-trivial (at order 1) symmetric entire perturbation C_ε of the ellipse. Our aim is to prove the non-integrability of the billiard map T_ε, which is analytic since C_ε so is. For this purpose we only have to prove that the Melnikov potential (3.9) is non-constant. The argument is heavily based in the fact that $\eta(\varphi)$ is a non-constant π-periodic entire function.

Under this hypothesis, $\tau_0 = \pi i/2$ is a singularity of $\eta(\varphi_0(t))$. It suffices to note that $\sin\varphi_0(t) = sech(t)$ and $\cos\varphi_0(t) = \tanh(t)$ have simple poles at τ_0 and no more singularities on $\Im t = \pi/2$. Then, τ_0 is also a singularity of the function $f(t)$ defined in (3.9), since $v_+(t)v_-(t)$ is analytic and non-zero on τ_0. Finally, using that $L(t) - f(t)$ is clearly analytic on τ_0, τ_0 is a singularity of $L(t)$. In particular, the Melnikov potential $L(t)$ is non-constant and we have proved the following result.

Theorem 3.1 *Let C_ε be a non-trivial (at order 1) symmetric entire perturbation of a non-circular ellipse. Then the billiard in C_ε is non-integrable for $0 < |\varepsilon| \ll 1$.*

Given an integer $\ell \geq 1$, a perturbation C_ε of an ellipse C_0 is called *trivial up to order* ℓ if there exists a family of ellipses E_ε such that $C_\varepsilon = E_\varepsilon + O(\varepsilon^{\ell+1})$. The discussion above fails for perturbations trivial up to order 1, but the result of non-integrability can be generalized to *non-trivial perturbations*, that is, except for perturbations that are trivial up to any order $\ell \geq 1$.

Theorem 3.2 *Let C_ε be a non-trivial symmetric entire perturbation of a non-circular ellipse. Then the billiard in C_ε is non-integrable for $0 < |\varepsilon| \ll 1$.*

We describe briefly how this theorem can be proved and the interested reader should fill in the gaps without difficulty.

First, the curve C_ε can be written, modulo a similarity, as (compare with (3.5)):

$$C_\varepsilon = \left\{ \left(\bar\alpha(\varepsilon)\cos\varphi, \bar\eta(\varphi,\varepsilon)\bar\beta(\varepsilon)\sin\varphi \right) : \varphi \in \mathbb{T} \right\},$$

where $\bar{\eta}(\varphi,\varepsilon) = 1+\varepsilon^\ell \eta(\varphi)+O(\varepsilon^{\ell+1})$, $\eta(\varphi)$ is a non-constant π-periodic entire function, $\bar{\alpha}(\varepsilon)$, $\bar{\beta}(\varepsilon)$ are smooth functions such that $\bar{\alpha}(\varepsilon)^2 - \bar{\beta}(\varepsilon)^2 = 1$, and $\ell \geq 1$ is the smallest integer such that C_ε is non-trivial at order ℓ.

Next, we consider the family of ellipses

$$E_\varepsilon = \{(\bar{\alpha}(\varepsilon)\cos\varphi, \bar{\beta}(\varepsilon)\sin\varphi) : \varphi \in \mathbb{T}\},$$

and the biparametric family of curves

$$G_{\varepsilon,\delta} = \{(\bar{\alpha}(\varepsilon)\cos\varphi, \tilde{\eta}(\varphi,\varepsilon,\delta)\bar{\beta}(\varepsilon)\sin\varphi) : \varphi \in \mathbb{T}\},$$

where $\tilde{\eta}(\varphi,\varepsilon,\delta) = 1+\delta\eta(\varphi)+O(\varepsilon\delta)$ is defined in such a way that $C_\varepsilon = G_{\varepsilon,\delta} = E_\varepsilon + O(\delta)$, for $\delta = \varepsilon^\ell$.

Finally, since elliptic billiards are integrable systems with separatrices, we can take E_ε as the unperturbed curve and $G_{\varepsilon,\delta}$ as the perturbation, being $\delta = \varepsilon^\ell$ the perturbation strength. In this setting, Theorem 3.2 follows just along the same lines as Theorem 3.1. The crux of the argument is again that $\eta(\varphi)$ is a non-constant π-periodic entire function.

3.4 Symmetric reversible perturbations

Along this subsection we shall study several topics concerning a special kind of symmetric perturbations, called *reversible*. By definition, these are perturbations preserving the original axial symmetries of the ellipse, that is, perturbations (3.6) such that $P(u,w) = P(-u,w) = P(u,-w)$, or equivalently, $P(u,w) = Q(u^2,w^2)$ for some smooth function $Q : \mathbb{R}^2 \to \mathbb{R}$ such that $Q(1,0) = 0$. Let $\kappa(s)$ be the smooth function defined as $\kappa(s) := Q(1-s^2,s^2)s^{-2}$. Then, relation (3.7) implies that $\eta(\varphi) = \kappa(\sin^2\varphi)$.

The lobe area

Our goal now is to introduce the *lobe area* as a quantity measuring the splitting size.

To such end, we first look for the reversors of the system. We will find two of them, a property that will allow us to state the existence of at least a couple of symmetric heteroclinic orbits $\mathcal{O}^\pm_\varepsilon$. The area of the region enclosed by these orbits will be then defined as the lobe area.

The involution

$$R_0^+ : \mathcal{P} \to \mathcal{P}, \qquad R_0^+(\varphi, v) := (\pi - \varphi, v),$$

is a reversor for the elliptic billiard T_0, and also for $F_0 = ST_0$. The separatrix Λ is R_0^+-symmetric, i.e., $R_0^+\Lambda = \Lambda$, and intersects transversely the fixed set of R_0^+

$$C_0^+ := \{p \in \mathcal{P} : R^+p = p\} = \{(\varphi, v) \in \mathcal{P} : \varphi = \pi/2\}$$

in one point $p_0^+ = (\pi/2, 1)$. The natural parameterization $p_0(t)$ of Λ given in Lemma 3.1 has been chosen to satisfy $p_0(0) = p_0^+$.

Moreover, the involution $R_0^- = F_0R^+$ is another reversor of F_0. The separatrix Λ is also R_0^--symmetric and intersects transversely the fixed set C_0^- of R_0^- in one point p_0^-, and it turns out that $p_0(h/2) = p_0^-$. The associated unperturbed heteroclinic orbits

$$\mathcal{O}_0^+ := \{p_0(hn) : n \in \mathbb{Z}\}, \qquad \mathcal{O}_0^- := \{p_0(h/2 + hn) : n \in \mathbb{Z}\} \tag{3.10}$$

are called *symmetric* heteroclinic orbits, since $R_0^\pm\mathcal{O}_0^\pm = \mathcal{O}_0^\pm$.

For $\varepsilon \neq 0$, since we have restricted the study to reversible perturbations, $R^+ := R_0^+$ is also a reversor of F_ε, as well as the involution $R^- := F_\varepsilon R^+$. Their fixed sets $C^\pm = \{p \in \mathcal{P} : R^\pm p = p\}$ are important because $R^\pm\mathcal{W}_\varepsilon^u(p_\infty^l) = \mathcal{W}_\varepsilon^s(p_\infty^r)$, where $\mathcal{W}_\varepsilon^{u,s}(p_\infty^r)$ and $\mathcal{W}_\varepsilon^{u,s}(p_\infty^l)$ stand for the perturbed invariant curves at the hyperbolic fixed points p_∞^r and p_∞^l. Consequently, any point in the intersection $C^\pm \cap \mathcal{W}_\varepsilon^s(p_\infty^r)$ is a heteroclinic one, and gives rise to a symmetric heteroclinic orbit.

Since the separatrix Λ intersects transversely the unperturbed curve $C_0^\pm$ at the point $p_0^\pm$, there exists a point $p_\varepsilon^\pm = p_0^\pm + O(\varepsilon) \in C^\pm \cap \mathcal{W}_\varepsilon^s(p_\infty^r)$ and, therefore, there exist at least two symmetric heteroclinic orbits, denoted $\mathcal{O}_\varepsilon^\pm$, on the region

$$\{(\varphi, v) \in \mathcal{P} : 0 < \varphi < \pi,\ 0 < v\},$$

for $|\varepsilon|$ small enough. They are called *primary* since they exist for arbitrary small $|\varepsilon|$.

Of course, using the reversor R and the symmetry S, we get that there exist at least *eight* symmetric primary heteroclinic orbits: $\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$.

On the other hand, from $v_0(t) = \sin\varphi_0(t) = \mathit{sech}\, t$ and $\eta(\varphi) = \kappa(\sin^2\varphi)$, it follows that $f(t) = 2\alpha v_+(t)v_-(t)\eta(\varphi_0(t))$ is even, so that the Melnikov potential $L(t) = \sum_{n\in\mathbb{Z}} f(t+hn)$ is even and h-periodic. Its derivative, $M(t) := L'(t)$ is odd and h-periodic; hence $M(nh/2) = 0$, $n \in \mathbb{Z}$. Therefore, $h\mathbb{Z}/2$ is *a* set of critical points for $L(t)$, that generically, are non-degenerate.

We shall prove in Proposition 3.2 that for any given non-zero polynomial perturbation and h small enough, the critical points of $L(t)$ are *just* $h\mathbb{Z}/2$, all of them being non-degenerate. As a consequence, the perturbed billiard map has *just eight* primary heteroclinic orbits: the symmetric ones $\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$. Moreover, the pieces of the perturbed invariant curves between the points $p_\varepsilon^\pm \in \mathcal{O}_\varepsilon^\pm$ enclose a region called *lobe*. Our measure of the splitting size for the planar billiard problem will be the area $A = A(\varepsilon, h)$ of this lobe, which is nothing else but the homoclinic area between $\mathcal{O}_\varepsilon^+$ and $\mathcal{O}_\varepsilon^-$. By Theorem 2.3, it is given by

$$A = \Delta W[\mathcal{O}_\varepsilon^+, \mathcal{O}_\varepsilon^-] = \varepsilon\Omega(h) + O(\varepsilon^2), \qquad \Omega(h) = L(0) - L(h/2).$$

Polynomial perturbations

In order to perform an explicit computation of the Melnikov potential (3.9), we restrict ourselves to symmetric reversible *polynomial* perturbations, that is, perturbations such that the function $P(u,w)$ in the implicit form (3.6) is a polynomial in the variables u^2 and w^2: $P(u,w) = \sum' p_{ij} u^{2i} w^{2j}$, with $\sum'_* p_{ij} = 0$. Here $\sum'$ stands for a finite sum over a range of non-negative integers i and j, whereas $\sum'_*$ denotes the same sum without the terms with $j \neq 0$. (The additional condition is due to the normalization modulo a similarity, which allows us to assume that $P(1,0) = \partial_2 P(1,0) = 0$.)

In the parameterized normal form (3.5), by relation (3.7), these symmetric reversible polynomial perturbations are equivalent to suppose that

$$\eta(\varphi) = \sum_{n=0}^{N} \eta_n \sin^{2n}\varphi, \qquad \eta_N \neq 0,$$

for some integer $N \geq 0$ (called *the order of the perturbation*).

We now address the explicit computation of the Melnikov potential (3.9). Since $v_0(t) = \sin\varphi_0(t) = \mathit{sech}\, t$, then $\eta(\varphi_0(t)) = \sum_{n=0}^{N} \eta_n\, \mathit{sech}^{2n}\, t$, and the function $f(t) = 2\alpha v_+(t) v_-(t) \eta(\varphi_0(t))$ is πi-periodic and meromorphic, so that the Melnikov potential $L(t) = \sum_{n\in\mathbb{Z}} f(t+hn)$ is an *elliptic function* with periods h and πi. This crucial observation goes back to [LT93, DR96, Lev97].

We now review some properties of the elliptic functions (for a general background, we refer to [AS72, WW27]).

Let us recall that a *cell* of an elliptic function of periods ω_1 and ω_2 is any parallelogram P_τ of vertexes τ, $\tau+\omega_1$, $\tau+\omega_2$, and $\tau+\omega_1+\omega_2$, such that its boundary does not contain poles. Then, the set of poles in any given cell is called an *irreducible set of poles.* A direct consequence of the Liouville's Theorem is that two elliptic functions with the same periods, poles, and principal parts, must be the same modulo an additive constant. (By periodicity, in practice it suffices to consider an irreducible set of poles.) This additive constant is not relevant for our purposes, since the intrinsic geometrical object associated to the problem is $L'(t)$ rather than $L(t)$ itself.

Therefore, we are naturally led to the location of an irreducible set of poles for the Melnikov potential $L(t)$, and next to the computation of the associated principal parts.

First, consider $\tau_0 = \pi i/2$ and $\tau_0^\pm = \tau_0 \pm h/2$. By the comments before Theorem 3.1, the poles of $f(t)$ are $\tau_0 + \pi i\mathbb{Z}$, which are of order $2N$, and $\tau_0^\pm + \pi i\mathbb{Z}$, which are simple ones.

Now we focus on their principal parts. We denote by $a_\ell(f,\tau)$ the coefficient of the term $(t-\tau)^\ell$ in the Laurent expansion of $f(t)$ around $t=\tau$.

From the relations

$$a_{2\ell}(v_+,\tau_0) = a_{2\ell}(v_-,\tau_0),$$

the formula $f(t) = \beta[v_+(t)+v_-(t)]\delta(t)$, and the symmetry of $f(t)$ with regard to its "central" pole $\tau_0 = (\tau_0^+ + \tau_0^-)/2$, we get

$$\begin{aligned} a_{-1}(f,\tau_0^+) + a_{-1}(f,\tau_0^-) &= 0, \\ a_{-(2\ell+2)}(f,\tau_0) &= 2\beta a_{-(2\ell+2)}(v_+\delta,\tau_0), \\ a_{-(2\ell+1)}(f,\tau_0) &= 0. \end{aligned}$$

This shows that $\{\pi i/2\}$ is an irreducible set of poles of the Melnikov potential $L(t) = \sum_{n\in\mathbb{Z}} f(t+hn)$. The pole $\pi i/2$ has order $2N$ and

$$a_{-(2\ell+1)}(L,\pi i/2) = 0, \qquad a_{-(2\ell+2)}(L,\pi i/2) = 2\beta a_{-(2\ell+2)}(v_+\delta,\pi i/2),$$

for all $\ell = 0,\ldots,N-1$. Therefore, modulo an additive constant, we can express the Melnikov potential $L(t)$ as a linear combination of even derivatives of the Weierstrass $\wp$-function associated to the periods h and πi evaluated at the point $t - \pi i/2$:

$$L(t) = constant + 2\beta \sum_{\ell=0}^{N-1} \frac{a_{-(2\ell+2)}(v_+\delta, \pi i/2)}{(2\ell+1)!} \wp^{(2\ell)}(t - \pi i/2). \tag{3.11}$$

It suffices to check that both sides of the equality have the same periods, poles, and principal parts. To see this, let us remember that the Weierstrass $\wp$-function associated to the periods $\omega_1 = h$ and $\omega_2 = \pi i$ is defined by the series

$$\wp(t) := t^{-2} + \sum_{n \in \mathbb{Z}_*^2} \{(t-\omega_n)^{-2} - {\omega_n}^{-2}\},$$

where $\omega_{n_1,n_2} = n_1\omega_1 + n_2\omega_2$ and $\mathbb{Z}_*^2 = \mathbb{Z}^2 \setminus \{(0,0)\}$. From its definition, it is obvious that $\wp(t)$ is elliptic with periods h and πi, and $\{0\}$ is an irreducible set of poles for $\wp(t)$, with t^{-2} as the principal part of $\wp(t)$ around $t = 0$. Then, formula (3.11) follows.

For purposes of numerical computations the function $\wp(t)$ is useless on account of the slowness of its convergence. (The general term in the series above is only of order $|n|^{-3}$.) Accordingly, we will introduce another function $\psi(t)$, best suited for pencil-and-paper and/or numerical computations, based in the use of *Jacobian elliptic functions*, such that

$$\psi(t) = constant - \wp(t - \pi i/2). \tag{3.12}$$

Then, we will rewrite formula (3.11) as

$$L(t) = constant - 2\beta \sum_{\ell=0}^{N-1} \frac{a_{-(2\ell+2)}(v_+\delta, \pi i/2)}{(2\ell+1)!} \psi^{(2\ell)}(t). \tag{3.13}$$

This simple formula allows us to compute the Melnikov function in a finite number of steps, for any symmetric reversible polynomial perturbation, that is, for any $\eta(\varphi) = \sum_{n=0}^{N} \eta_n \sin^{2n} \varphi$. We need only to compute the numbers $a_{-(2\ell+2)}(v_+\delta, \pi i/2)$, $\ell = 0, \ldots, N-1$, in each concrete case, where

$$v_+(t) = sech(t - h/2), \qquad \delta(t) = \sum_{n=0}^{N} \eta_n \, sech^{2n+1} t. \tag{3.14}$$

For instance, it is easy to compute $a_{-2N}(v_+\delta, \pi i/2) = (-1)^N \eta_N \alpha\beta$.

The definition of $\psi(t)$ requires the introduction of some additional notations, which we borrow again from [AS72, WW27]. Given the *parameter* $m \in [0,1]$, $K = K(m) := \int_0^{\pi/2}(1 - m\sin\vartheta)^{-1/2}\,d\vartheta$ is the *complete elliptic integral of the first kind*, $K' = K'(m) := K(1-m)$ and $q = q(m) := \exp(-\pi K'/K)$ is the *nome*. If any of the numbers m, K, K', K'/K or q is given, all the rest are determined. From our purposes, it is convenient to determine the value of the quotient K'/K by imposing $K'/K = \pi/h$. From now on, we can consider the quantities m, q and K as functions of h. For instance, the nome is exponentially small in h,

$$q = q(h) = e^{-\pi^2/h}.$$

Under these notations and assumptions it turns out that the elliptic function

$$\psi(t) := \left(\frac{2K}{h}\right)^2 dn^2\left(\frac{2Kt}{h}\middle| m\right), \tag{3.15}$$

where $dn(u) = dn(u|m)$ is one of Jacobian elliptic functions, verifies (3.12). Indeed, it suffices to observe that $dn^2(u|m)$ is an elliptic function of periods $2K$ and $2K'i$, which has $\{K'i\}$ as an irreducible set of poles, $-(u-K'i)^{-2}$ being the principal part of $dn^2(u|m)$ around $u = K'i$. Then, the change of scale $u = 2Kt/h$ makes $\psi(t)$ elliptic with periods h and πi (this is the reason for the choice $K'/K = \pi/h$), and the pre-factor $(2K/h)^2$ prevents a change of its principal part.

In order to convince the reader on the adequacy of $\psi(t)$ for numerical work, we note that its Fourier expansion, valid for $|\Im t| < \pi/2$, is given by

$$\psi(t) = constant + \sum_{k>1} \psi_k \cos(2\pi kt/h), \qquad \psi_k = \left(\frac{2\pi}{h}\right)^2 \frac{2kq^k}{1-q^{2k}}, \quad \forall k \geq 1. \tag{3.16}$$

(The value of $\psi_0 = \int_0^h \psi(t)\,dt\,/h$ is not needed, since we are working modulo additive constants.) Clearly, this series is rapidly convergent for real t (the values we are interested in). This Fourier expansion can be obtained from the relation $dn^2(u|m) = 1 - m\,sn^2(u|m)$ and the Fourier expansion of $sn^2(u|m)$ given in [WW27, page 520].

Quartic perturbations

Let us assume now that C_ε is a symmetric reversible *quartic* perturbation, that is, the function $P(u,w)$ in the implicit normal form (3.6) is a polynomial of degree four:

$$P(u,w) = p_{00}+p_{10}u^2+p_{01}w^2+p_{20}u^4+p_{11}u^2w^2+p_{02}w^4, \qquad p_{00} = -(p_{10}+p_{20}).$$

From relation (3.7), these symmetric reversible quartic perturbations are equivalent to suppose that

$$\eta(\varphi) = \eta_0 + \eta_1 \sin^2 \varphi$$

in the parameterized normal form (3.5), where

$$\eta_1 = p_{20} - p_{11} + p_{02}.$$

(The value of η_0 makes no importance in the following discussion.)

Taking $P(u,w) = w^4$ we get an example of this kind of perturbations, namely

$$C_\varepsilon = \left\{ (x,y) \in \mathbb{R}^2 : \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 + 2\varepsilon \frac{y^4}{\beta^4} \right\},$$

which gives $\eta_0 = 0$ and $\eta_1 = 1$, that is,

$$C_\varepsilon = \left\{ (\alpha \cos\varphi, [1+\varepsilon \sin^2\varphi]\beta \sin\varphi) : \varphi \in \mathbb{T} \right\}. \tag{3.17}$$

Quartic perturbations are interesting because *everything* (Melnikov potential, homoclinic orbits, and lobe areas) can be easily computed. For instance, formula (3.13) takes the simple form

$$\begin{aligned} L(t) &= \textit{constant} + 2\eta_1 \alpha\beta^2 \psi(t) \\ &= \textit{constant} + 2\eta_1 \alpha\beta^2 \left(\frac{2K}{h}\right)^2 dn^2 \left(\frac{2Kt}{h} \middle| m\right). \end{aligned} \tag{3.18}$$

From the properties of the function $dn(u|m)$, the set of real critical points of $L(t)$ is $h\mathbb{Z}/2$, all of them being non-degenerate. According to Theorem 2.1, this gives two homoclinic orbits $\mathcal{O}_\varepsilon^\pm$ close to the unperturbed ones $\mathcal{O}_0^\pm$ given in (3.10).

Taking into account the symmetries and reversors, the perturbed symmetric billiard map has just eight (transverse) symmetric primary heteroclinic orbits: $\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$.

Moreover, since $dn(0|m) = 1$ and $dn(K|m) = \sqrt{1-m}$, the area $A = A(\varepsilon, h)$ of the lobe enclosed by the heteroclinic orbits $\mathcal{O}_\varepsilon^\pm$ is given by

$$A = \varepsilon\Omega(h) + O(\varepsilon^2), \qquad \Omega(h) = L(0) - L(h/2) = 2\eta_1\alpha\beta^2 m (2K/h)^2.$$

We summarize all these results in the following proposition.

Proposition 3.1 *For $0 < |\varepsilon| \ll 1$, the symmetric billiard map associated to the convex curve (3.17) has exactly 8 primary homoclinic orbits $\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$, and all of them are transverse.*

The Melnikov potential has the expression (3.18), and the area of the lobe enclosed by the homoclinic orbits $\mathcal{O}_\varepsilon^\pm$ is given by

$$A = 2\varepsilon\eta_1\alpha\beta^2 m(2K/h)^2 + O(\varepsilon^2).$$

From the formula $\sqrt{2Km^{1/2}/\pi} = 2\sum_{k\geq 0} q^{(k+1/2)^2}$ [WW27, page 479] and the expression of the nome $q = e^{-\pi^2/h}$, we get another expression for $\Omega(h)$:

$$\Omega(h) = 32\pi^2\eta_1\alpha\beta^2 h^{-2} e^{-\pi^2/h}\left\{\sum_{k\geq 0}\exp[-\pi^2 k(k+1)/h]\right\}.$$

This series can be numerically computed in a very fast way, due to the speed of its convergence, even for relatively big values of h.

Clearly, $\Omega(h)$ is exponentially small in h, and we are led naturally to the following duality. For *regular* perturbations ($h > 0$ remains fixed whereas $\varepsilon \to 0$), the Melnikov term $\varepsilon\Omega(h)$ is the dominant term for the formula of the lobe area A. On the contrary, in *singular* perturbations ($h \to 0^+$ and $\varepsilon \to 0$), one is confronted with the difficult problem of justifying the following exponentially small asymptotic expression provided by the Melnikov method:

$$A = A(\varepsilon, h) \sim \varepsilon\Omega(h) \sim 256\pi^2\eta_1\varepsilon h^{-5} e^{-\pi^2/h} \qquad (\varepsilon \to 0, h \to 0^+), \tag{3.19}$$

where we have used that $\alpha = \coth(h/2) \sim 2/h$, and $\beta = cosech(h/2) \sim 2/h$.

We recall that h is the characteristic exponent of the hyperbolic fixed points $p_\infty^{l,r}$ for the symmetric billiard map F_0, see Lemma 3.1. Therefore, singular perturbations correspond to weakly hyperbolic cases. For a justification of an exponentially small asymptotic expression like (3.19), but for other kind of twist maps, we refer the reader to [DR97b, DR97c].

Singular polynomial perturbations

Coming back to a general N, we give a generalization of the exponentially small Melnikov prediction (3.19). Along the following discussion we will assume that h is small enough and that the coefficients η_n, $n = 0, \ldots, N$, of the perturbation

$$\eta(\varphi) = \sum_{n=0}^{N} \eta_n \sin^{2n} \varphi, \qquad \eta_N \neq 0,$$

are fixed. The convex curve is now

$$C_\varepsilon = \left\{ \left(\alpha \cos \varphi, \left[1 + \varepsilon \sum_{n=0}^{N} \eta_n \sin^{2n} \varphi \right] \beta \sin \varphi \right) \, : \, \varphi \in \mathbb{T} \right\}. \tag{3.20}$$

By the definition of $\delta(t)$ given in (3.14), the Melnikov potential (3.13) reads as

$$L(t) = constant\ - 2\beta \sum_{n=1}^{N} \sum_{\ell=0}^{n-1} \frac{B_{n,\ell}}{(2\ell+1)!} \eta_n \psi^{(2\ell)}(t), \tag{3.21}$$

where

$$B_{n,\ell} = a_{-(2\ell+2)}(v_+ \cdot sech^{2n+1}, \pi i/2), \qquad \forall \ell < n. \tag{3.22}$$

(We note that $B_{n,\ell} = 0$, for $\ell \geq n$.) To get the dominant terms of (3.21), we must study the order in h of the functions $\psi^{(2\ell)}(t)$ and the coefficients $B_{n,\ell}$, for $0 \leq \ell \leq n-1$, $1 \leq n \leq N$.

Let us begin with the derivatives of $\psi(t)$. From the Fourier expansion of $\psi(t)$ given in (3.16), we obtain the exponentially small asymptotic expressions

$$\psi^{(2\ell)}(t) = constant\ + (-1)^\ell 2 (2\pi/h)^{2\ell+2} e^{-\pi^2/h} \cos(2\pi t/h) \left[1 + O(e^{-\pi^2/h}) \right], \tag{3.23}$$

for integers $\ell \geq 0$, real t, and small enough $h > 0$.

Next, we focus on the coefficients $B_{n,\ell}$. We split the function v_+ defined in (3.14) in its principal v_+^p and regular $v_+^r (= v_+ - v_+^p)$ part around its singularity $\tau_0^+ = (\pi i + h)/2$. A simple computation gives

$$v_+^p(t) = \frac{-i}{(t - \tau_0^+)}.$$

From the Cauchy inequalities, the coefficients in the Taylor expansion of v_+^r around $\pi i/2$ are $O(1)$, since v_+^r is uniformly bounded, for h small, in a ball of fixed radius centered at $\pi i/2$. Thus,

$$a_\ell(v_+, \pi i/2) = a_\ell(v_+^p, \pi i/2) + a_\ell(v_+^r, \pi i/2) = (2/h)^{\ell+1} i + O(1), \quad \forall \ell \geq 1.$$

Besides, the principal part of $sech^{2n+1}$ around its pole $\pi i/2$ is $O(1)$ and, in particular, $a_{-(2n+1)}(sech^{2n+1}, \pi i/2) = (-1)^{n+1} i$. The discussion above shows that

$$\begin{aligned} B_{n,\ell} &= \sum_{j=0}^{n-\ell} a_{2j+1}(v_+, \pi i/2) \cdot a_{-(2\ell+2j+1)}(sech^{2n+1}, \pi i/2) \\ &= (-1)^n 2^{2n-2\ell} h^{2\ell-2n} \left[1 + O(h^2)\right]. \end{aligned} \tag{3.24}$$

From (3.14) and (3.24) we get

$$\begin{aligned} B_{n,\ell}\psi^{(2\ell)}(t) &= constant + \\ &\quad (-1)^{n+\ell} 2^{2n+4} \pi^{2\ell+2} h^{-(2n+2)} e^{-\pi^2/h} \cos(2\pi t/h) \left[1 + O(h^2)\right], \end{aligned}$$

so that the dominant terms of (3.21) are attained at $n = N$.

Finally, using the relation $\beta = cosech(h/2) = 2/h + O(h)$, we get the following exponentially small asymptotic expression for the Melnikov potential (3.21):

$$L(t) = constant + 2^{-1} \Omega_N \eta_N h^{-(2N+3)} e^{-\pi^2/h} \cos(2\pi t/h) \left[1 + O(h^2)\right], \tag{3.25}$$

where Ω_N is a constant which depends only on the order of the perturbation N, namely

$$\Omega_N = (-1)^N 2^{2N+6} \sum_{\ell=1}^{N} \frac{(-1)^\ell \pi^{2\ell}}{(2\ell-1)!}. \tag{3.26}$$

As π is a transcendental number, $\Omega_N \neq 0$ for all $N \geq 1$ (but $\Omega_N \to 0$ for $N \to \infty$). Thus, the set of real critical points of the Melnikov potential $L(t)$ is $h\mathbb{Z}/2$, all of them being non-degenerate, provided that h is small enough.

As in the quartic perturbation, it follows that for $0 < h \ll 1$, the billiard has just eight (transversal) symmetric primary homoclinic orbits: $\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$. Moreover, the area $A = A(\varepsilon, h)$ of the lobe enclosed by $\mathcal{O}_\varepsilon^\pm$ is given by

$$A = \varepsilon \Omega(h) + O(\varepsilon^2), \qquad \Omega(h) = \Omega_N \eta_N h^{-(2N+3)} e^{-\pi^2/h} \left[1 + O(h^2)\right].$$

We summarize now these results.

Proposition 3.2 *For $h > 0$ small enough, there exists $\varepsilon_0 = \varepsilon_0(h) > 0$ such that for $0 < |\varepsilon| < \varepsilon_0$, the symmetric billiard map associated to the convex curve (3.20) has exactly 8 primary homoclinic orbits $\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$, and all of them are transverse.*

The Melnikov potential has the expression (3.25), and the area of the lobe enclosed by the homoclinic orbits $\mathcal{O}_\varepsilon^\pm$ is given by

$$A = \varepsilon \Omega_N \eta_N h^{-(2N+3)} e^{-\pi^2/h} \left[1 + O(h^2)\right] + O(\varepsilon^2),$$

with $\Omega_N \neq 0$ given in (3.26).

For regular perturbations the Melnikov term $\varepsilon\Omega(h)$ dominates, but for singular perturbations there is a lack of results about the validity of the exponentially small Melnikov prediction

$$A = A(\varepsilon, h) \sim \Omega_N \eta_N \varepsilon h^{-(2N+3)} e^{-\pi^2/h} \qquad (\varepsilon \to 0, h \to 0^+),$$

as in the case before of quartic perturbations.

A geometric interpretation

All the previous results could be expressed in terms of the eccentricity of the unperturbed ellipse $e = \tanh(h/2)$, which is a *natural* parameter for the billiard due to its clear geometric meaning. We have preferred the characteristic exponent h, since it can be considered as the *intrinsic* parameter for the problem.

In that setting, singular perturbations $(h, \varepsilon \to 0)$, can be thought as perturbations of the billiard in a circumference, since the eccentricity of a circumference is $e = 0$, which corresponds to the value $h = 0$.

4 High-dimensional billiards

4.1 Convex billiards

We consider the problem of the "convex billiard motion" in more dimensions. Let $\mathcal{Q}$ be a smooth closed convex hypersurface of $\mathbb{R}^{d+1}$, for $d \geq 2$, parameterized by $\gamma : \mathbb{S}^d \to \mathcal{Q}$, where $\mathbb{S}^d$ is the d-dimensional unit sphere. Suppose that a material point moves inside $\mathcal{Q}$ and collides elastically with $\mathcal{Q}$. Such discrete dynamical systems can be modeled by a smooth twist map (called *billiard map*) T with d degrees of freedom, defined on a suitable open $\mathcal{P}$ of the cotangent bundle of $\mathbb{S}^d$.

In order to describe this twist map, let us introduce the discrete version of the Legendre transformation B of $\mathbb{S}^d \times \mathbb{S}^d$ onto the cotangent bundle of $\mathbb{S}^d$ defined by

$$B(z, z') := (z, w), \qquad w\, dz = -\partial_1 G(z, z')\, dz \tag{4.1}$$

where w is the fiber coordinate, $w\, dz$ is the standard 1-form on the cotangent bundle of $\mathbb{S}^d$, and the function

$$G : \{(z, z') \in \mathbb{S}^d \times \mathbb{S}^d : z \neq z'\} \to \mathbb{R}, \qquad G(z, z') := |\gamma(z) - \gamma(z')|$$

is the Lagrangian of the billiard [Ves91, MV91].

Although generically the Legendre transformation B has only a local inverse, using the convexity condition on $\mathcal{Q}$, it can be easily checked that the billiard Legendre transformation (4.1) is a diffeomorphism from the open set $\mathcal{V} = \{(z, z') \in \mathbb{S}^d \times \mathbb{S}^d : z \neq z'\}$ onto its image $\mathcal{P} = B(\mathcal{V})$. (This is a consequence of the fact that for convex billiards the orbits can be determined either by giving two consecutive different impact points determined by their (base) coordinates z and z', or by giving the (base) coordinate z of an impact point together with the direction of incidence, which is determined by the fiber coordinate w.)

Then, the billiard map is defined by

$$T : \mathcal{P} \to \mathcal{P}, \qquad (z', w') = T(z, w) = B\widehat{T}B^{-1}(z, w)$$

where the diffeomorphism $\widehat{T} : \mathcal{V} \to \mathcal{V}$ maps a couple of consecutive impact points $(z, z') \in \mathcal{V}$ to another couple of consecutive impact points (z', z''), z'' being the impact point following z and z'. The Lagrangian $G(z, z')$ is a twist generating function for the billiard map T, that is,

$$T(z, w) = (z', w') \Longleftrightarrow w'\, dz' - w\, dz = T^*(w\, dz) - w\, dz = dG(z, z').$$

As in the planar case, we shall work in the space of convex hypersurfaces modulo similarities, since billiard maps associated to hypersurfaces related by a similarity are conjugated, and so equal from a dynamical point of view.

The billiard map T has no fixed points, but it has two-periodic orbits. For instance, the two more distant points (on the Euclidean metric in $\mathbb{R}^{d+1}$)

give rise to a two-periodic orbit, which is generically unstable in the linear approximation. In these orbits the fiber coordinate w vanishes.

To study the dynamics of these two-periodic orbits for T, it is better to consider them as fixed points of the square map T^2, and study T^2. But since it is not easy to find the twist generating function for T^2, we instead introduce the same simplification as in the planar case. We will assume that $\mathcal{Q}$ is *symmetric* with regard to the origin:

$$\mathcal{Q} = -\mathcal{Q}.$$

Consequently, it is possible to choose an *odd* parameterization $\gamma : \mathbb{S}^d \to \mathcal{Q}$ in such a way that the two-periodic orbits are of the form $\{(z_0, 0), (-z_0, 0)\}$, that is, two opposite points over $\mathcal{Q}$. Then, the billiard map T and the involution

$$S : \mathcal{P} \to \mathcal{P}, \qquad S(z, w) := (-z, -w),$$

commute.

This allows us to introduce the *symmetric billiard map*

$$F : \mathcal{P} \to \mathcal{P}, \qquad F := ST,$$

so that the two-periodic orbits for T are fixed points for F. Since $F^2 = T^2$, the dynamics of F and T are equivalent. The map F is also a twist map, with

$$\mathcal{L} : \{(z, z') \in \mathbb{S}^d \times \mathbb{S}^d : z + z' \neq 0\} \to \mathbb{R}, \qquad \mathcal{L}(z, z') = |\gamma(z) + \gamma(z')|$$

as its twist generating function, since $\gamma(-z') = -\gamma(z')$.

Finally, let us consider the involution

$$R : \mathcal{P} \to \mathcal{P}, \qquad R(z, w) := (z, -w),$$

which is a reversor for F. We will use the symmetry S and the reversor R to save work in the computation of homoclinic orbits (like in the planar case).

4.2 Prolate ellipsoidal billiards

The simplest examples of smooth convex hypersurfaces are the ellipsoids. Among them, the *spheres* are too degenerate for a billiard system, since there are plenty of (parabolic) two-periodic orbits, formed by all the pairs of opposed points. However, the study of a *generic ellipsoid* (that is, an ellipsoid without axis of equal length) is much more complicated than the study of the non-circular elliptic billiard before, because the explicit expression of the biasymptotic motions in the first case requires the use of analytical tools much more sophisticated than in the second one [Fed97]. Therefore, in order to gain insight into the problem, it is interesting to consider a setting to which the arguments of the planar case can be easily adapted.

This setting is provided by *prolate ellipsoids* (that is, ellipsoids with all its axis of equal length except one, which is larger).

In order to put the involved objects in a compact form, let us introduce the following notation. Given a point $q = (q_0, \ldots, q_d) \in \mathbb{R}^{d+1}$, we denote

$$\bar{q} = q_0 \in \mathbb{R}, \qquad \hat{q} = (q_1, \ldots, q_d) \in \mathbb{R}^d.$$

The same notation is used for points $z = (z_0, \ldots, z_d) \in \mathbb{S}^d$. Now, we can write a prolate ellipsoid as

$$\begin{aligned} \mathcal{Q}_0 &= \left\{ q = (\bar{q}, \hat{q}) \in \mathbb{R}^{d+1} : \frac{\bar{q}^2}{\alpha^2} + \frac{|\hat{q}|^2}{\beta^2} = 1 \right\} \\ &= \left\{ \gamma_0(z) = (\alpha \bar{z}, \beta \hat{z}) : z = (\bar{z}, \hat{z}) \in \mathbb{S}^d \right\}, \end{aligned} \tag{4.2}$$

with $\alpha > \beta > 0$. Modulo a similarity, we can assume that $\alpha^2 - \beta^2 = 1$.

Let us denote $T_0 : \mathcal{P} \to \mathcal{P}$ the twist map associated to the prolate ellipsoid $\mathcal{Q}_0$, and $F_0 = ST_0$. The billiard map T_0 is called (prolate) ellipsoidal billiard.

The points

$$p^{l,r}_\infty = (z^{l,r}_\infty, w^{l,r}_\infty), \qquad z^l_\infty = (-1, 0), \qquad z^r_\infty = (1, 0), \qquad w^{l,r}_\infty = 0,$$

form a two-periodic orbit for T_0 that correspond to the (left and right) vertexes $(\mp\alpha, 0)$ of the prolate ellipsoid on its "horizontal" axis $\{\hat{q} = 0\}$, and hence, they are fixed points for F_0. It turns out that these fixed points are hyperbolic ones, and their invariant manifolds are completely doubled giving rise to two separatrices, in the sense explained in section 2.

Using the symmetry S, we could identify the points $p = (z, w)$ and $S(p) = (-z, -w)$. Then, the fixed points $p^{l,r}_\infty$ become the same point, so that the previous connections could be considered homoclinic ones.

Now, we are confronted to the computation of the heteroclinic orbits for F_0. The rotational symmetry of the prolate ellipsoid $\mathcal{Q}_0$ with regard to its "horizontal" axis $\{\hat{q} = 0\}$ is the essential point to accomplish it. Given a direction $a \in \mathbb{S}^{d-1}$, let Π_a be the plane in $\mathbb{R}^{d+1}$ generated by the directions $(1, 0)$ and $(0, a)$, and let $C_0(a)$ be the section of the prolate ellipsoid $\mathcal{Q}_0$ by the plane Π_a, that is,

$$\Pi_a := [(1,0),(0,a)] = \{q = (\bar{q}, \hat{q}) = (x, ya) : x, y \in \mathbb{R}\},$$

and

$$\begin{aligned} C_0(a) := \mathcal{Q}_0 \cap \Pi_a &= \left\{ q = (\bar{q}, \hat{q}) = (x, ya) \in \mathbb{R}^{d+1} : \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 \right\} \\ &= \{\gamma_0(z) = (\alpha \bar{z}, \beta \hat{z}) : \bar{z} = \cos\varphi, \hat{z} = [\sin\varphi]a, \varphi \in \mathbb{T}\}. \end{aligned}$$

All the sections $C_0(a)$ are ellipses with the same foci: $(\pm 1, 0)$, and the same eccentricity: $e = 1/\alpha$. The key observation is that *if two consecutive impact points are on the same section, the same happens to all the other impact points.* From Lemma 3.1 and this geometric property—which does not hold for a generic ellipsoid—, we get the heteroclinic orbits for F_0. The result is summarized in the following lemma.

Lemma 4.1 *Let $h > 0$ be determined by the equations*

$$\alpha = \coth(h/2), \quad \beta = \mathit{cosech}(h/2), \quad e = \tanh(h/2).$$

Let $q_0 = (\bar{q}_0, \hat{q}_0) = (\alpha \bar{z}_0, \beta \hat{z}_0) : \mathbb{R} \times \mathbb{S}^{d-1} \to \mathcal{Q}_0 \setminus \{(\pm\alpha, 0)\}$ be the diffeomorphism defined by

$$\bar{z}_0(t, a) = \cos\varphi_0(t) = \tanh t, \quad \hat{z}_0(t, a) = [\sin\varphi_0(t)]a = [\mathit{sech}\, t]a,$$

where $\varphi_0 : \mathbb{R} \to (0, \pi)$ stands for the map $\varphi_0(t) = \arccos(\tanh t)$.

a) Given any $(t, a) \in \mathbb{R} \times \mathbb{S}^{d-1}$ the sequences of impact points

$$\mathcal{O}_0^\tau = (q_n^\tau)_{n \in \mathbb{Z}} \subset \mathcal{Q}_0, \qquad \tau \in \{\leftarrow, \rightarrow\}$$

where $q_n^{\rightarrow} = q_0(t + hn, a)$, and $q_n^{\leftarrow} = q_{-n}^{\rightarrow}$, are heteroclinic orbits for F_0. The superscript τ indicates the direction of the orbit: the orbit goes to the left (that is, from the right vertex of the ellipsoid to the left one) for $\tau = \leftarrow$, whereas it goes to the right for $\tau = \rightarrow$. Finally, there are not more heteroclinic orbits for F_0 than the ones obtained in this way.

b) Let $z_0'(t,a) = z_0(t+h,a)$*. Then,*

$$\beta \frac{\hat{z}_0(t,a) + \hat{z}_0'(t,a)}{|\gamma_0(z_0(t,a)) + \gamma_0(z_0'(t,a))|} = [sech(t+h/2)]a. \tag{4.3}$$

The main properties of F_0 are listed in the following lemma, which is a straightforward consequence of the previous one.

Lemma 4.2 *a) The points* $p_\infty^{l,r}$ *are hyperbolic fixed points of the symmetric billiard map* F_0*. Actually,* $Spec[\, dF_0(p_\infty^{r,l})] = \{e^h, e^{-h}\}$*.*

b) Let $\mathcal{W}_0^{u,s}(p_\infty^{r,l})$ *be the unperturbed unstable and stable invariant curves of* F_0 *at* $p_\infty^{r,l}$*. Then,* $\mathcal{W}_0^s(p_\infty^{r,l}) = \mathcal{W}_0^u(p_\infty^{l,r})$*, and* F_0 *has* two *separatrices:*

$$\Lambda^{\rightarrow} := \mathcal{W}_0^u(p_\infty^l) \cap \mathcal{W}_0^s(p_\infty^r) = \left\{ p_0^{\rightarrow}(t,a) : (t,a) \in \mathbb{R} \times \mathbb{S}^{d-1} \right\},$$
$$\Lambda^{\leftarrow} := \mathcal{W}_0^s(p_\infty^l) \cap \mathcal{W}_0^u(p_\infty^r) = \left\{ p_0^{\leftarrow}(t,a) : (t,a) \in \mathbb{R} \times \mathbb{S}^{d-1} \right\},$$

where

$$p_0^{\rightarrow}(t,a) = B(z_0(t,a), z_0'(t,a)), \quad p_0^{\leftarrow}(t,a) = B(z_0'(t,a), z_0(t,a)),$$

are natural parameterizations, that is, $F_0(p_0^{\leftarrow}(t,a)) = p_0^{\leftarrow}(t+h,a)$*, and* $F_0(p_0^{\rightarrow}(t,a)) = p_0^{\rightarrow}(t+h,a)$*.*

The separatrices $\Lambda^{\rightarrow}$ and $\Lambda^{\leftarrow}$ are invariant by the symmetry S, whereas they are interchanged by the reversor R, since R changes the sense of the (discrete) time. In the planar case ($d=1$) we had four separatrices: Λ, $R(\Lambda)$, $S(\Lambda)$, and $RS(\Lambda)$. In the high-dimensional case we have just two: $\Lambda^{\rightarrow}$ and $\Lambda^{\leftarrow}$. A natural questions arises: Why? The answer is easy. If one tries to rewrite the above lemma in the planar case, the variable a moves on $\mathbb{S}^{d-1} = \mathbb{S}^0 = \{\pm 1\}$, which has *two* different connected components. Then, for $d=1$, the set $\Lambda^{\rightarrow} \cup \Lambda^{\leftarrow}$ is formed by four different connected components, each one being a separatrix.

4.3 Splitting in billiards close to prolate ellipsoids

Any ellipsoidal billiard, including the non-prolate ones, is completely integrable [Ves91, MV91]. Thus, it is natural to conjecture that ellipsoidal billiards are the only completely integrable smooth convex billiards, as a generalization of Birkhoff's conjecture in the plane. Nevertheless, we are not ready to tackle this conjecture, not even a local version of it around prolate ellipsoids. The tools at our disposal only allow us to establish the splitting of separatrices under *very general* perturbations of a prolate ellipsoid.

To begin with, let us consider an arbitrary symmetric smooth perturbation $\mathcal{Q}_\varepsilon = -\mathcal{Q}_\varepsilon$ of the prolate ellipsoid $\mathcal{Q}_0$. Up to second order terms in the perturbative parameter ε (which do not play any rôle in our first order perturbative analysis) and a similarity, $\mathcal{Q}_\varepsilon$ can be put in the following *parameterized (normal) form*

$$\mathcal{Q}_\varepsilon = \left\{\gamma_\varepsilon(z) = (\alpha\bar{z}, [1+\varepsilon\nu(z)]\beta\hat{z}) : z = (\bar{z},\hat{z}) \in \mathbb{S}^d\right\}, \tag{4.4}$$

for some even smooth function $\nu : \mathbb{S}^d \to \mathbb{R}$, or in the following *implicit form*

$$\mathcal{Q}_\varepsilon = \left\{ q = (\bar{q},\hat{q}) \in \mathbb{R}^{d+1} : \frac{\bar{q}^2}{\alpha^2} + \frac{|\hat{q}|^2}{\beta^2} = 1 + 2\varepsilon P(\alpha^{-1}\bar{q}, \beta^{-1}\hat{q}) \right\}, \tag{4.5}$$

for some even smooth function $P(\bar{z},\hat{z})$ such that $P(1,0) = d_2P(1,0) = 0$. The connection between the two formulations is very simple, namely

$$P(\bar{z},\hat{z}) = |\hat{z}|^2\, \nu(z), \qquad \forall z = (\bar{z},\hat{z}) \in \mathbb{S}^d. \tag{4.6}$$

In order to make easier the translation of results directly from the planar setting, it is convenient to consider the smooth function $\eta : \mathbb{T}\times\mathbb{S}^{d-1} \to \mathbb{R}$ defined by

$$\eta(\varphi, a) = \nu(z), \qquad z = (\bar{z},\hat{z}), \qquad \bar{z} = \cos\varphi, \qquad \hat{z} = [\sin\varphi]a. \tag{4.7}$$

Now, our aim is to translate neatly the results for ellipses to results for (prolate) ellipsoids. As in the planar case the key point is to elucidate for which "degenerate" perturbations the Melnikov potential is identically constant. The results in the planar case were optimal, since the only "degenerate" perturbations were the trivial ones giving rise to ellipses, which are integrable.

Unfortunately, this is no longer the case in more dimensions. We shall prove that $\mathcal{Q}_\varepsilon$ is a "degenerate" perturbation for the prolate ellipsoid $\mathcal{Q}_0$ when all its sections

$$C_\varepsilon(a) := \mathcal{Q}_\varepsilon \cap \Pi_a = \{(\alpha\cos\varphi, [1+\varepsilon\eta(\varphi,a)[\beta\sin\varphi]a) : \varphi \in \mathbb{T}\}$$

are ellipses (up to second order terms, of course). Obviously, the ellipsoids are a particular case of such perturbations, but there are other ones, as the following lemma shows. We skip its proof, which is a mere computation.

Lemma 4.3 *Under the above notations and assumptions:*

a) $\mathcal{Q}_\varepsilon$ *is an ellipsoid (up to order 1) if and only if* $\eta(\varphi, a) = \langle a, Da\rangle$, *for some symmetric* $d \times d$ *matrix* D.

b) $C_\varepsilon(a)$ *is an ellipse (up to order 1) if and only if* $\eta(\cdot, a)$ *is constant.*

Let $T_\varepsilon : \mathcal{P} \to \mathcal{P}$ be the billiard map associated to hypersurface $\mathcal{Q}_\varepsilon$, and $F_\varepsilon = ST_\varepsilon$. For $|\varepsilon| \ll 1$, $\mathcal{Q}_\varepsilon$ is a convex closed hypersurface, and thus F_ε is a twist map, with

$$\mathcal{L}_\varepsilon(z, z') = |\gamma_\varepsilon(z) + \gamma_\varepsilon(z')| = \mathcal{L}_0(z, z') + \varepsilon \mathcal{L}_1(z, z') + O(\varepsilon^2)$$

as its twist generating function, where

$$\begin{aligned} \mathcal{L}_0(z, z') &= |\gamma_0(z) + \gamma_0(z')| \\ \mathcal{L}_1(z, z') &= \beta^2 \frac{\langle \hat{z} + \hat{z}', \nu(z)\hat{z} + \nu(z')\hat{z}'\rangle}{|\gamma_0(z) + \gamma_0(z')|}. \end{aligned} \tag{4.8}$$

Using the natural parameterization provided by Lemma 4.1, the formula of $\mathcal{L}_1(z, z')$ given in equation (4.8), and the formula (4.3), the Melnikov potential of our perturbed billiard problem (on the separatrix $\Lambda^{\to}$) is

$$L : \mathbb{R} \times \mathbb{S}^{d-1} \to \mathbb{R}, \qquad L(t, a) = \sum_{n\in\mathbb{Z}} g(t + hn, a) = \sum_{n\in\mathbb{Z}} f(t + hn, a), \tag{4.9}$$

where

$$g(t, a) = \beta v_-(t)[\delta(t, a) + \delta(t + h, a)], \qquad f(t, a) = 2\alpha v_+(t) v_-(t) \eta(\varphi_0(t), a),$$

with

$$v_0(t) = sech\, t, \quad v_\pm(t) = v_0(t \mp h/2), \qquad \delta(t, a) = v_0(t)\eta(\varphi_0(t), a).$$

(We have taken (t, a) as the coordinates over the separatrix $\Lambda^{\to}$. Compare with the results in the planar case.)

Now, assume we are given a symmetric perturbation $\mathcal{Q}_\varepsilon$ of the prolate ellipsoid $\mathcal{Q}_0$ such that its section $C_\varepsilon(a_0)$ is a non-trivial (up to order 1) symmetric entire perturbation of the ellipse $C_0(a_0)$, for some $a_0 \in \mathbb{S}^{d-1}$. By definition, $\varphi \mapsto \eta(\varphi, a_0)$ is a non-constant π-periodic entire function. Then, $t \mapsto L(t, a_0)$ is a non-constant function (it suffices to copy the proof for the planar case) and we have proved the following result.

Theorem 4.1 *Let $\mathcal{Q}_\varepsilon$ be symmetric perturbation of the prolate ellipsoid $\mathcal{Q}_0$, such that some of its sections $C_\varepsilon(a)$ is a non-trivial (up to order 1) symmetric entire perturbation of the ellipse $C_0(a)$. Then, the separatrices $\Lambda^\rightarrow$ and $\Lambda^\leftarrow$ split, for $0 < |\varepsilon| \ll 1$.*

In fact, we have proved only that $\Lambda^\rightarrow$ splits, but it is clear that $\Lambda^\leftarrow$ also splits. Indeed, it is enough to observe that the heteroclinic orbits that go to the right are in a one-to-one correspondence with the heteroclinic orbits that go to the left, by means of the reversor R. Therefore, the destruction of a separatrix automatically implies the destruction of the other one.

4.4 Lower bounds

Let us recall that in the planar case $d = 1$, there were at least 8 symmetric primary heteroclinic orbits ($\mathcal{O}_\varepsilon^\pm$, $R\mathcal{O}_\varepsilon^\pm$, $S\mathcal{O}_\varepsilon^\pm$, and $RS\mathcal{O}_\varepsilon^\pm$), for reversible symmetric perturbations of an ellipse. Our goal now is to present similar results for perturbations of a prolate ellipsoid. Obviously, the first step is to define the term reversible for $d \geq 2$.

Following the planar case, a perturbation of the prolate ellipsoid (4.2) will be called *reversible* when it preserves the original symmetries of the ellipsoid with regard to the hyperplane $\{\bar{q} = 0\}$ and the axis $\{\hat{q} = 0\}$, that is, perturbations (4.5) such that $P(\bar{z}, \hat{z})$ is even in $\bar{z}$ and $\hat{z}$. Then, equations (4.6) and (4.7) imply that

$$\eta(\varphi, a) = \eta(\varphi, -a). \tag{4.10}$$

On the one hand, Morse theory provides lower bounds on the number of critical points for functions defined on compact manifolds. On the other hand, the critical points of the Melnikov potential $L : \mathbb{R} \times \mathbb{S}^{d-1} \to \mathbb{R}$ are strongly related to primary heteroclinic orbits. Therefore, it is rather natural to apply Morse theory in order to gain information on the number of primary heteroclinic orbits that persist after perturbation. At a first glance, there exists a technical problem (L is defined on a non-compact manifold), but there is an obvious way to overcome this difficulty: L can be considered as a function defined over the *reduced separatrix* $\mathbb{S}^1 \times \mathbb{S}^{d-1}$, using that $L(t,a)$ is h-periodic in t, and the identification $\mathbb{S}^1 = \mathbb{R}/\{t = t+h\}$.

Under the condition that the Melnikov potential is a Morse function (a condition of generic position), we now state a result about the number of primary heteroclinic orbits that persist under a general perturbation. We will verify the optimality of this result for specific examples.

Theorem 4.2 *Let $\mathcal{Q}_\varepsilon$ be a symmetric smooth perturbation of a prolate ellipsoid $\mathcal{Q}_0$ of dimension d, such that its Melnikov potential $L : \mathbb{S}^1 \times \mathbb{S}^{d-1} \to \mathbb{R}$ is a Morse function. Then, the number of primary heteroclinic orbits after perturbation is at least* 8. *If, in addition, the perturbation $\mathcal{Q}_\varepsilon$ is reversible, there exist at least* $8d$ *primary heteroclinic orbits after perturbation.*

Proof. Since the Melnikov potential $L : \mathbb{S}^1 \times \mathbb{S}^{d-1} \to \mathbb{R}$ is a Morse function, its critical points are in one-to-one correspondence with the primary heteroclinic orbits that emanate from $\Lambda^{\rightarrow}$, which in their turn are in one-to-one correspondence with the primary heteroclinic orbits that emanate from $\Lambda^{\leftarrow}$.

For reversible perturbations, the Melnikov potential can be considered as a (Morse) function defined over $\mathbb{S}^1 \times \mathbb{P}^{d-1}$, where $\mathbb{P}^{d-1} = \mathbb{S}^{d-1}/\{a = -a\}$ is the projective space, since equations (4.9) and (4.10) imply that $L(t,a)$ is even in a. Moreover, each critical point $(t, \pm a) \in \mathbb{S}^1 \times \mathbb{P}^{d-1}$ of $L : \mathbb{S}^1 \times \mathbb{P}^{d-1} \to \mathbb{R}$ corresponds to *two* different critical points of $L : \mathbb{S}^1 \times \mathbb{S}^{d-1} \to \mathbb{R}$.

From the celebrated Morse's inequalities [Hir76, pages 160-164], a Morse function over a d-dimensional compact manifold without boundary X has at least $SB(X;F) := \sum_{q=0}^{d} \beta_q(X;F)$ critical points, where $\beta_q(X;F)$ are the *F-Betti numbers* of X and F is any field, that is, $\beta_q(X;F)$ is the dimension of the q-th singular homology F-vector space of X, noted $H_q(X,F)$.

Consequently, it suffices to check that

$$SB(\mathbb{S}^1 \times \mathbb{S}^{d-1}; \mathbb{Z}_2) = 4, \qquad SB(\mathbb{S}^1 \times \mathbb{P}^{d-1}; \mathbb{Z}_2) = 2d, \tag{4.11}$$

for all $d \geq 2$.

From the well-known $\mathbb{Z}_2$-homologies

$$H_q(\mathbb{S}^m; \mathbb{Z}_2) \cong \begin{cases} \mathbb{Z}_2 & \text{if } q = 0, m \\ 0 & \text{otherwise} \end{cases} \qquad H_q(\mathbb{P}^m; \mathbb{Z}_2) \cong \begin{cases} \mathbb{Z}_2 & \text{if } 0 \leq q \leq m \\ 0 & \text{otherwise} \end{cases},$$

and the Künneth's Formula $H_q(X\times Y;\mathbb{Z}_2)\cong\bigoplus_{p=0}^{q}H_p(X;\mathbb{Z}_2)\otimes H_{q-p}(Y;\mathbb{Z}_2)$, we get

$$
\begin{aligned}
H_q(\mathbb{S}^1\times\mathbb{S}^1;\mathbb{Z}_2) &\cong \begin{cases}\mathbb{Z}_2 & \text{if } q=0,2\\ \mathbb{Z}_2\oplus\mathbb{Z}_2 & \text{if } q=1\\ 0 & \text{otherwise}\end{cases},\\
H_q(\mathbb{S}^1\times\mathbb{S}^{d-1};\mathbb{Z}_2) &\cong \begin{cases}\mathbb{Z}_2 & \text{if } q=0,1,d-1,d\\ 0 & \text{otherwise}\end{cases}, \quad \forall d>2,\\
H_q(\mathbb{S}^1\times\mathbb{P}^{d-1};\mathbb{Z}_2) &\cong \begin{cases}\mathbb{Z}_2 & \text{if } q=0,d\\ \mathbb{Z}_2\oplus\mathbb{Z}_2 & \text{if } q=1,\ldots,d-1\\ 0 & \text{otherwise}\end{cases}, \quad \forall d\geq 2.
\end{aligned}
$$

And (4.11) follows adding dimensions. □

4.5 Polynomial perturbations

We shall study now *polynomial* perturbations of the prolate ellipsoid, that is, perturbations such that the function $P(\bar{z},\hat{z})$ in the implicit form (4.5) is a polynomial. Our goals are the following:

- To compute explicitly the Melnikov potential (and its critical points) for some concrete perturbations.
- To check that the lower bounds given in Theorem (4.2) are optimal.
- To prove that the Melnikov method gives exponentially small (in h) predictions of the splitting size for singular perturbations, as in the planar case.

We shall omit many details in the computations below, since they are a transcription of the same ones in the planar case. The only difference is the additional variable $a\in\mathbb{S}^{d-1}$, which can be considered as a parameter.

Polynomial reversible perturbations

Suppose we are given a polynomial reversible perturbation $\mathcal{Q}_\varepsilon$. Thus, the function $P(\bar{z},\hat{z})$ in the implicit form (4.5) is an even polynomial in the variables $\bar{z}$ and $\hat{z}$. For the sake of simplicity we will assume that

$$P(\bar{z},\hat{z})=|\hat{z}|^2\sum_{n=0}^{N}P_{2n}(\hat{z}),$$

where $P_{2n} : \mathbb{R}^d \to \mathbb{R}$ denotes a homogeneous polynomial of degree $2n$; in particular, this implies that $P(\bar{z}, \hat{z})$ does not depend on the variable $\bar{z}$.

Then, using (4.6) and (4.7), we get that

$$\eta(\varphi, a) = \sum_{n=0}^{N} \eta_n(a) \sin^{2n} \varphi, \qquad \eta_n(a) = P_{2n}(a),$$

and the Melnikov potential is

$$L(t, a) = \textit{constant} - 2\beta \sum_{n=1}^{N} \sum_{\ell=0}^{n-1} \frac{B_{n,\ell}}{(2\ell+1)!} \eta_n(a) \psi^{(2\ell)}(t), \tag{4.12}$$

where the coefficients $B_{n,\ell}$ are given in (3.22) and the elliptic function $\psi(t)$ is defined in (3.15). Furthermore, the dominant terms in h in the singular limit $h \to 0^+$ turns out to be

$$L(t, a) = \textit{constant} + 2^{-1} \Omega_N h^{-(2N+3)} e^{-\pi^2/h} \cos(2\pi t/h) \left[\eta_N(a) + O(h^2) \right],$$

where Ω_N is the non-zero constant defined in (3.26).

This shows how to compute explicitly the Melnikov potential $L(t, a)$ for any polynomial perturbation, and makes evident its exponentially small dependence on h.

A quartic reversible perturbation

Let us consider now the simplest non-trivial case of the previous polynomial perturbations, that is, the case of quartic perturbations: $N = 1$. Concretely, given any symmetric $d \times d$ matrix M, we introduce the perturbation

$$\mathcal{Q}_\varepsilon = \left\{ q = (\bar{q}, \hat{q}) \in \mathbb{R}^{d+1} : \frac{\bar{q}^2}{\alpha^2} + \frac{|\hat{q}|^2}{\beta^2} = 1 + 2\varepsilon \beta^{-4} |\hat{q}|^2 \langle \hat{q}, M \hat{q} \rangle \right\}, \tag{4.13}$$

which gives $\eta_0(a) \equiv 0$ and $\eta_1(a) = \langle a, Ma \rangle$. As in the planar case, *everything* can be computed for quartic perturbations. Using that $B_{1,0} = -\alpha\beta$, we get the Melnikov potential:

$$\begin{aligned} L(t, a) &= \textit{constant} - 2\beta B_{1,0} \eta_1(a) \psi(t) \\ &= \textit{constant} + 2\alpha\beta^2 \left(\frac{2K}{h} \right)^2 \langle a, Ma \rangle \, dn^2 \left(\frac{2Kt}{h} \middle| m \right). \end{aligned}$$

Proposition 4.1 *Let λ_j be the eigenvalues of M and u_j their respective (normalized) eigenvectors: $Mu_j = \lambda_j u_j$ and $u_j \in \mathbb{S}^{d-1}$, for $1 \leq j \leq d$. Suppose that*

$$\lambda_j \neq 0 \quad \forall j, \qquad \lambda_j \neq \lambda_s \quad \forall j \neq s. \tag{4.14}$$

Then, the symmetric billiard map associated to the hypersurface (4.13) has exactly $8d$ primary heteroclinic orbits:

$$\mathcal{O}_\varepsilon^{\tau,\sigma,\pm j} = \left(q_n^{\tau,\sigma,\pm j}(\varepsilon)\right)_{n\in\mathbb{Z}} \subset \mathcal{Q}_\varepsilon, \qquad \tau \in \{\leftarrow,\rightarrow\},\ \sigma \in \{0,1\},\ j \in \{1,\ldots,d\},$$

all of them being transverse, for $0 < |\varepsilon| \ll 1$. The superscript τ indicates the direction of the orbit, as in Lemma 4.1. The functions $q_n^{\tau,\sigma,\pm j}(\varepsilon)$ are smooth in $\varepsilon = 0$, and

$$q_n^{\rightarrow,\sigma,\pm j}(0) = q_0(\sigma h/2 + nh, \pm u_j), \qquad q_n^{\leftarrow,\sigma,\pm j}(\varepsilon) = q_{-n}^{\rightarrow,\sigma,\pm j}(\varepsilon),$$

where $q_0 = (\bar{q}_0, \hat{q}_0) : \mathbb{R} \times \mathbb{S}^{d-1} \to \mathcal{Q}_0 \setminus \{(\pm\alpha, 0)\}$ is the diffeomorphism

$$\bar{q}_0(t,a) = \alpha \tanh t, \quad \hat{q}_0(t,a) = [\beta\, \mathit{sech}\, t]a.$$

Proof. Let Q be the orthogonal matrix whose columns are the eigenvalues u_j. Then, $\mathbb{S}^{d-1} \ni a \mapsto Qa \in \mathbb{S}^{d-1}$ is a diffeomorphism, such that $\eta_1(Qa) = \sum_{j=1}^d \lambda_j {a_j}^2$. Thus, hypotheses (4.14) imply that $\eta_1 : \mathbb{S}^{d-1} \to \mathbb{R}$ has exactly $2d$ critical points: $\{\pm u_j : 1 \leq i \leq d\}$, all of them being non-degenerate. Moreover, we recall that $h\mathbb{Z}/2$ is the set of real critical points of $\psi(t)$, and that these critical points are non-degenerate.

Consequently, $L(t,a)$ is a Morse function over $\mathbb{R}/h\mathbb{Z} \times \mathbb{S}^{d-1}$, which has exactly $4d$ critical points: $(\sigma h/2, \pm u_j)$, for $\sigma \in \{0,1\}$, $1 \leq j \leq d$. They are non-degenerate, too.

Finally, the proposition follows from Theorem 2.1, and Lemma 4.1. □

A quartic non-reversible perturbation

We shall describe similar results obtained for the simplest non-trivial non-reversible perturbation, which is also a quartic one. We shall omit the details since they do not involve any new idea, but only some tedious computations with elliptic functions.

Given a non-zero vector $u \in \mathbb{R}^d$, we consider the perturbation

$$\mathcal{Q}_\varepsilon = \left\{ q = (\bar{q}, \hat{q}) \in \mathbb{R}^{d+1} : \frac{\bar{q}^2}{\alpha^2} + \frac{|\hat{q}|^2}{\beta^2} = 1 + 2\varepsilon\beta^{-4}|\hat{q}|^2 \bar{q}\langle \hat{q}, u\rangle \right\}, \tag{4.15}$$

Then, using the same arguments than in the proof of the preceding proposition, we get the following result.

Proposition 4.2 *The symmetric billiard map associated to the hypersurface (4.15) has exactly* 8 *primary heteroclinic orbits:*

$$\mathcal{O}_\varepsilon^{\tau,\sigma,\pm} = \left(q_n^{\tau,\sigma,\pm}(\varepsilon)\right)_{n\in\mathbb{Z}} \subset \mathcal{Q}_\varepsilon, \qquad \tau \in \{\leftarrow,\rightarrow\},\ \sigma\in\{0,1\},$$

all of them being transverse, for $0<|\varepsilon|\ll 1$. *Moreover,*

$$q_n^{\rightarrow,\sigma,\pm j}(0) = q_0(t_\sigma + nh, \pm u), \qquad q_n^{\leftarrow,\sigma,\pm}(\varepsilon) = q_{-n}^{\rightarrow,\sigma,\pm}(\varepsilon),$$

where $t_0 \in (0,h/2)$ *and* $t_1 \in (3h/4,h)$ *are the only critical points in the interval* $[0,h]$ *of the elliptic function (with periods* h *and* $2\pi i$*):*

$$t \mapsto \sum_{n\in\mathbb{Z}} \frac{\sinh(t+hn)}{\cosh^2(t+hn)}.$$

Some last comments

The previous examples show that the lower bounds on the number of heteroclinic orbits provided by Theorem 4.2 are optimal. The conditions (4.14), for the reversible perturbations, and $u\neq 0$, for the non-reversible ones, are the conditions of generic position for $L(t,a)$ to be a Morse function. The condition $\lambda_j \neq \lambda_s$ for $j\neq s$, is equivalent to the complete breakdown of the symmetry of revolution with regard to the axis $\{\hat{q}=0\}$ of the prolate ellipsoid.

Following the planar case, singular perturbations correspond to perturbations of a spheric billiard, and the Melnikov prediction of the heteroclinic area between some distinguished pairs of heteroclinic orbits is again exponentially small in h. As an example we simply note that, for the reversible quartic perturbation (4.13),

$$\Delta W[\mathcal{O}_\varepsilon^{\tau,1,\pm j}, \mathcal{O}_\varepsilon^{\tau,0,\pm j}] = \varepsilon\Delta_j(h) + O(\varepsilon^2), \qquad \tau\in\{\leftarrow,\rightarrow\},\ j\in\{1,\ldots,d\},$$

where

$$\begin{aligned} \Delta_j(h) &= L(h/2,u_j) - L(0,u_j) \\ &= 32\pi^2\lambda_j\alpha\beta^2 h^{-2} e^{-\pi^2/h}\left\{\sum_{k\geq 0}\exp[-\pi^2 k(k+1)/h]\right\} \\ &\sim 256\pi^2\lambda_j h^{-5} e^{-\pi^2/h} \qquad (h\to 0^+). \end{aligned}$$

Acknowledgments

This work has been partially supported by the EC grant ERBCHRXCT-940460. Research by A. Delshams is also supported by the Spanish grant DGICYT PB94-0215 and the Catalan grant CIRIT 1996SGR–00105. Research by R. Ramírez-Ros is also supported by the U.P.C. grant PR9409. Both authors wish to express their appreciation to Y. Fedorov for several private communications. This work was finished while one of the authors (A.D.), was a visitor at the Institute for Mathematics and Its Applications in Minneapolis, for whose hospitality he is very grateful.

References

[AM78] R.H. Abraham and J.E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Reading, Mass., 1978.

[Arn76] V.I. Arnold. *Les méthodes mathématiques de la mécanique classique*. Mir, Moscou, 1976.

[AS72] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, 1972.

[BG96] P. Boyland and C. Golé. Lagrangian systems on hyperbolic manifolds. Preprint 1996/1, Institute for Mathematical Sciences, Stony Brook, January 1996.

[Bir27a] G.D. Birkhoff. *Dynamical systems*. Amer. Math. Soc. Publ., 1927.

[Bir27b] G.D. Birkhoff. On the periodic motions of dynamical systems. *Acta Math.*, 50:359–379, 1927.

[Bol95] S.V. Bolotin. Homoclinic orbits in invariant tori of Hamiltonian systems. In V.V. Kozlov, editor, *Dynamical systems in classical mechanics*, volume 168 of *Amer. Math. Soc. Transl. Ser. 2*, pages 21–90. Amer. Math. Soc., Providence, RI, 1995. Adv. Math. Sci., 25.

[DR96] A. Delshams and R. Ramírez-Ros. Poincaré-Melnikov-Arnold method for analytic planar maps. *Nonlinearity*, 9(1):1–26, 1996.

[DR97a] A. Delshams and R. Ramírez-Ros. Melnikov potential for exact symplectic maps. Preprint 97-150, `mp_arc@math.utexas.edu`, March 1997. To appear in *Comm. Math. Phys.*

[DR97b] A. Delshams and R. Ramírez-Ros. Exponentially small splitting of separatrices for standard-like maps. Preprint 97-160, `mp_arc@math.utexas.edu`, April 1997. To appear in *J. Nonlinear Sci.*

[DR97c] A. Delshams and R. Ramírez-Ros. Melnikov methods and exponentially small splitting of separatrices for planar maps: An experimental study. Preprint, November 1997.

[Eas91] R.W. Easton. Transport through chaos. *Nonlinearity*, 4:583–590, 1991.

[Fed97] Yu. Fedorov, 1997. Private communication.

[GG73] M. Golubitsky and V. Guillemin. *Stable Mappings and Their Singularities*, volume 14 of *Graduate Texts in Math.* Springer, New York, 1973.

[Gol94a] C. Golé. Optical Hamiltonians and symplectic twist maps. *Phys. D*, 71:185–195, 1994.

[Gol94b] C. Golé. Periodic orbits for Hamiltonian in cotangent bundles. *Trans. Amer. Math. Soc.*, 343(1):327–347, 1994.

[GS77] V. Guillemin and S. Sternberg. *Geometric Asymptotics.* Amer. Math. Soc., Providence, R.J., 1977.

[Hir76] M.W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Math.* Springer-Verlag, New York, Heidelberg, Berlin, 1976.

[KT91] V.V. Kozlov and D.V. Treshchev. *Billiards: A genetic introduction to the dynamics of systems with impacts*, volume 89 of *Transl. Math. Monographs.* Amer. Math. Soc., Providence, RI, 1991.

[Lev97] P. Levallois. Calcul d'une fonction de Melnikov et de ses zeros pour une perturbation algébrique du billiard elliptique. *Ergodic Theory Dynam. Systems*, 17:435–444, 1997.

[Lom96a] H.E. Lomelí. Perturbations of elliptic billiards. *Phys. D*, 99:59–80, 1996.

[Lom96b] H.E. Lomelí. Saddle connections and heteroclinic orbits for standard maps. *Nonlinearity*, 9:649–668, 1996.

[Lom97] H.E. Lomelí. Applications of the Melnikov method to twist maps in higher dimensions using the variational approach. *Ergodic Theory Dynam. Systems*, 17:445–462, 1997.

[LT93] P. Levallois and M.B. Tabanov. Séparation des séparatrices du billard elliptique pour une perturbation algébrique et symétrique de l'ellipse. *C. R. Acad. Sci. Paris Sér. I Math.*, 316:589–592, 1993.

[MMP84] R.S. MacKay, J.D. Meiss, and I.C. Percival. Transport in Hamiltonian systems. *Phys. D*, 13:55–81, 1984.

[MV91] J. Moser and A.P. Veselov. Discrete versions of some classical integrable systems and factorization of matrix polynomials. *Comm. Math. Phys.*, 139:217–243, 1991.

[PM82] J. Palis, Jr. and W. de Melo. *Geometric Theory of Dynamical Systems.* Springer-Verlag, New York, 1982.

[Sma63] S. Smale. Diffeomorphisms with many periodic points. In S.S. Cairns, editor, *Differential and combinatorial topology*, pages 63–80. Princeton Univ. Press, Princeton, New Jersey, 1963.

[Tab94] M.B. Tabanov. Separatrices splitting for Birkhoff's billiard in a symmetric convex domain, close to an ellipse. *Chaos*, 4(4):595–606, 1994.

[Tab95a] S. Tabachnikov. *Billiards*, volume 1 of *Panor. Synth.* Soc. Math. de France, Paris, 1995.

[Tab95b] E. Tabacman. Variational computation of homoclinic orbits for twist maps. *Phys. D*, 85:548–562, 1995.

[Tre94] D.V. Treschev. Hyperbolic tori and asymptotic surfaces in Hamiltonian systems. *Russ. J. Math. Phys.*, 2(1):93–110, 1994.

[Ves91] A.P. Veselov. Integrable maps. *Russian Math. Surveys*, 46(5):1–51, 1991.

[Wig90] S. Wiggins. *Global bifurcations and chaos: Analytical methods*, volume 73 of *Appl. Math. Sci.* Springer, New York, 1990.

[WW27] E.T. Whittaker and G.N. Watson. *A course of modern analysis.* Cambridge Univ. Press, Cambridge, 1927.

Internet access:

All the authors' quoted preprints are available at `http://www-ma1.upc.es` in the `Preprints` pages, or at `ftp://ftp-ma1.upc.es`, in the `pub/preprints` directory.

E-mail addresses: `amadeu@ma1.upc.es`, `rafael@tere.upc.es`.