Periodic orbits in smooth, strictly convex billiards

C.J. Oelen

Supervisor: prof. dr. B.W. Rink

Department of Mathematics
Vrije Universiteit

15-12-2017

Abstract

It is well known that in smooth, strictly convex billiards there exists a periodic orbit corresponding to the longest possible configuration. With additional assumptions, one can prove existence of another type of orbit and furthermore, both are well-ordered. For certain types of symmetrical billiard tables we will prove existence of other types of orbits than the Birkhoff periodic orbits, namely non-Birkhoff orbits. Also, we will present some aspects of the well-known billiard theory in the context of Aubry-Mather theory. Finally, we will apply the perturbation invariance of the Conley index to conjecture the existence non-symmetrical billiards in a neighborhood elliptical billiards, in which at least one non-Birkhoff orbit of type $(4, 2)$ persists.

Contents

1 Introduction

Proving existence of periodic orbits in dynamical systems is an important theme in dynamical systems theory. In this thesis we will investigate criteria for existence of periodic orbits in what are known as mathematical billiards. More specifically, we will turn our attention to *strictly convex* billiard systems. This subject, praised for its simple and intuitive nature while exhibiting the complexities of general Hamiltonian systems, is a relatively old subject with still many open questions today. We will mainly focus on the existence of periodic orbits, which is only one of the many aspects of mathematical billiards.

The billiard system is a dynamical system which consists of a closed curve (the billiard table) together with a tiny ball (the billiard ball), which is moving inside the region bounded the curve. We assume the collisions at the boundary of the curve are perfectly elastic and that the angle of incidence equals the angle of reflection. This latter property is what we call the "billiard law". We remark that the notion of billiard systems can also be extended to other geometries, but we will purely focus on the Euclidean setting.
The behavior of the billiard ball is of our interest and more specifically, we are interested in the existence of periodic orbits and in classifying them according to certain topological properties. In section 1 we will introduce the formal setting of the billiard system and prove some important properties of the associated billiard map. We introduce the notion of (periodic) orbits and see how different kinds of periodic orbits arise.

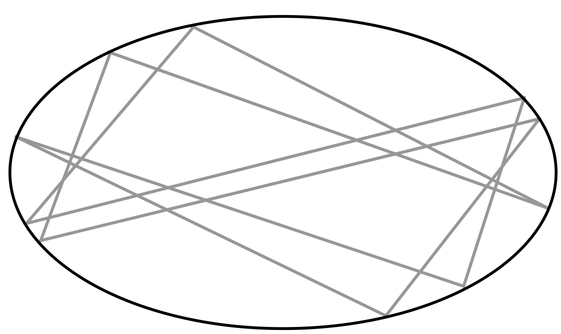

Figure 1: A periodic orbit in a strictly convex billiard table.

It turns out that the billiard system has a special kind of structure: a variational structure. In section 2 more properties are being explored and the variational structure of the billiard problem is explained. After that, we introduce some notions from Aubry-Mather theory, which are applied in finding periodic orbits. One of the aims of the thesis is to investigate how Aubry-Mather theory can be applied to billiard systems and how the theory changes when specialized to billiards. When Aubry-Mather theory is restricted to the context of billiard systems, some features of the theory become less involved than in the general case. We will investigate these alterations and provide simpler proofs than the ones that appear in the general setting. In this way, we obtain results about certain orderings properties of a special type of periodic orbits. In chapter 6 the notion of the gradient flow on an infinite dimensional Banach space is discussed and we show how some results obtained in chapter 5, can be found with this tool. The main purpose of introducing the gradient flow will become apparent in chapter 8 where it is used, in a finite dimensional setting, to find periodic orbits which do not satisfy the order property. This approach is also used to prove existence of periodic orbits that possesses a certain type of symmetry.

2 Formal setting

In this chapter we will give a formal description of the problem of finding and classifying periodic orbits. We will introduce basic notions of billiard systems and prove some useful identities.
Let $\mathcal{C} \subseteq \mathbb{R}^2$ be a closed, strictly convex, and at least C^3 curve. So in particular $\mathcal{C} \cong S^1$, where S^1 is a topological circle. We assume that $\gamma : S^1 \to \mathcal{C}$ is parametrized by arc length so that $\gamma'(t)$ has unit speed. Recall that that strictly convex means that the curvature $\kappa(s) > 0$ for each $s \in \mathcal{C}$, where

$$\kappa(t) = \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}}.$$

The region bounded by this set [1] represents the billiard table and $\mathcal{C}$ is the edge of the table. Since the dynamics only depends on $\mathcal{C}$, we shall refer to $\mathcal{C}$ as being the billiard table.
We consider a point moving inside $\mathcal{C}$ with constant speed and moving in straight lines until it hits the boundary with a certain angle [2]. There it behaves according to the law of reflection: angle of incidence equals angle of reflection. The resulting billiard paths are subject of our investigation in this thesis, and in particular the paths which are periodic. The game of billiards can be considered as a dynamical system, given by a sequence of pairs of positions s_i on $\mathcal{C}$ and angles $\theta_i \in (0, \pi)$ together with a map $T : S^1 \times (0, \pi) \to S^1 \times (0, \pi)$ known as the *billiard map*. The map T is given by

$$T(s_i, \theta_i) = (s_{i+1}, \theta_{i+1}),$$

where s_{i+1} and θ_{i+1} are obtained as follows. Draw a line segment from s_i with angle θ_i w.r.t. the tangent to $\mathcal{C}$ at s_i, to the first point of intersection with $\mathcal{C}$. Denote this intersection point by s_{i+1}, and the angle the line segments makes at s_{i+1} by θ_{i+1}.

[1] There is such a region by the Jordan curve theorem

[2] We shall use this language to mean the angle of the line segment w.r.t. the tangent at the intersection point

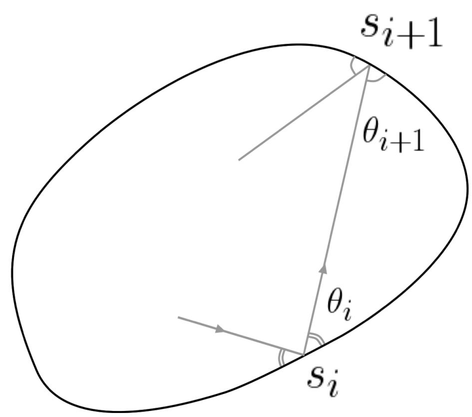

Figure 2: A path that satisfies the billiard law.

Occasionally we shall interpret the map T acting on pairs (x, v), where (x, v) is unit vector at x which makes an angle θ w.r.t. the tangent. So in this case we shall use the notation $T : (x, v) \mapsto (x', v')$ to denote the billiard map.

We impose the natural condition that $s_{i+1} - s_i \in (0, 1)$, which allows us in a later stage to define a notion of "winding around $\mathcal{C}$". It is clear that the map T is continuous if and only if $\mathcal{C}$ is convex. Note that the behavior at $\mathcal{C}$ completely determines the dynamics of the billiard ball. In Proposition 1 it is shown that the billiard map T is area preserving w.r.t. a certain area form. Hence we can interpret the billiard system as described above, as a two dimensional area preserving discrete dynamical system.

Remark 1. *From now on we will call a closed, strictly convex C^k, $k \geq 3$ curve simply a "billiard system" or "billiard domain". Since this class of curves only will be subject of this thesis, no confusion with other kinds of billiard tables should arise. Furthermore, we use the term "reflection points" and "vertices" interchangeably.*

We now introduce the notion of a generating function of the Billiard map. The terminology "generating" is chosen for reasons to become clear later. Denote as usual the norm of a vector in $\mathbb{R}^2$ as $\| \cdot \|$.

Definition 2.1. *Let $t, s \in \mathbb{R}$. The map $S : \{(t, s) \in S^1 \times S^1 : t \neq s\} \to \mathbb{R}$, given by*

$$S(t, s) = -\|\gamma(t) - \gamma(s)\|,$$

is called the generating function *of the billiard map T.*

Note that S is at least C^3, since the billiard domain $\mathcal{C}$ is assumed to be at least C^3. The derivatives of the generating function S play an important role in establishing properties of the billiard map T. In fact, one could easily call them the cornerstone of this chapter in which we explore fundamental properties of the billiard map T. For example, as we will see later on, they help to decide whether a sequence of points is actually part of path a billiard ball will travel. Furthermore, the derivatives of S will play a significant role in the proof of the smoothness of T and the existence of an area-preserving form.

The following lemma describes a relation between the partial derivatives of S and the angle coordinates θ, θ'.

Lemma 2.1. *Let S be the generating function of the billiard map. Write $\gamma(t) = (x(t), y(t))^T$, so that $t \in \mathbb{R}$ is the coordinate along $\mathcal{C}$ and denote the image of (t, θ) under T by (t', θ'). Then the following identities hold:*

$$\frac{\partial S}{\partial t} = \cos(\theta) \quad \text{\textit{and}} \quad \frac{\partial S}{\partial t'} = -\cos(\theta').$$

Proof. Recall that $\mathcal{C}$ is parametrized such that $\|\gamma'(t)\| = 1$. So

$$\begin{aligned}\frac{\partial S}{\partial t} &= -\frac{1}{S(t,t')}[(x(t) - x(t'))x'(t)) + (y(t) - y(t'))y'(t)] \\ &= \frac{1}{S(t,t')}[\gamma'(t) \cdot (\gamma(t') - \gamma(t))] \\ &= \cos(\theta),\end{aligned}$$

where in the last step we used that $u \cdot v = \|u\| \cdot \|v\| \cos(\theta)$, for $u, v \in \mathbb{R}^2$ and where θ is the angle between them. See Figure 3.

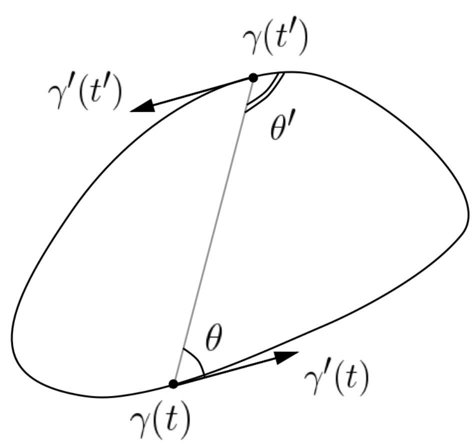

Figure 3: Deriving the relations from Lemma 2.1.

Similarly we obtain the other identity:

$$\begin{aligned}\frac{\partial S}{\partial t'} &= -\frac{1}{S(t,t')}[\gamma'(t')\cdot(\gamma(t')-\gamma(t))] \\ &= -\cos(\theta').\end{aligned}$$

□

A fundamental property of the billiard map is the existence of an area preserving form. This is an easy consequence of the previous lemma.

Proposition 1. *T leaves the area form $\omega = \sin(\theta)dt \wedge d\theta$ invariant.*

Proof. By definition of the differential and Lemma 2.1, we obtain

$$dS = \frac{\partial S}{\partial t}dt + \frac{\partial S}{\partial t'}dt' = \cos(\theta)dt - \cos(\theta')dt'.$$

Hence

$$0 = d^2S = -\sin(\theta)d\theta \wedge dt + \sin(\theta')d\theta' \wedge dt'.$$

So $\sin(\theta)d\theta \wedge dt = \sin(\theta')d\theta' \wedge dt$ and hence $T^*\omega = \omega$, that is, ω is T-invariant. □

One particularly important property is that the map T has a structure what is known as the *twist property*. It is because of this property that there is a variational approach for studying the dynamics of billiard systems. This latter fact will turn out to be useful in our search for periodic orbits.
The twist property in the billiard system has the following geometric interpretation. Let $T : (t,\theta) \mapsto (t',\theta')$ be the billiard map. Increasing the angle θ implies an increment of the position t' on $\mathcal{C}$, assuming the orientation of $\mathcal{C}$ is counter clockwise. This is shown in Figure 4. Formally, the twist property is proved by differentiating the relations found in Lemma 2.1.

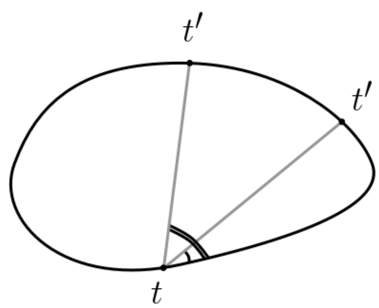

Figure 4: Increasing the angle causes t' to increase along $\mathcal{C}$ as well (keeping t fixed).

Proposition 2. *For the billiard map $T : (t,\theta) \mapsto (t',\theta')$ it holds that $\frac{\partial t'}{\partial \theta} > 0$*

Proof. We differentiate the relations in Lemma 2.1 to obtain

$$\begin{aligned}\frac{\partial^2 S}{\partial t \partial t'} &= \frac{S[-C'(t)\cdot C'(t')] - [C'(t)\cdot(C'(t) - C(t'))]\cos(\theta')}{S^2} \\ &= \frac{-\cos(\theta'+\theta)}{S} - \frac{\cos(\theta)\cos(\theta')}{S} \\ &= \frac{\sin(\theta)\sin(\theta')}{S}\end{aligned}$$

This implies that $\frac{\partial^2 S}{\partial t \partial t'} < 0$ since $0 < \theta, \theta' < \pi$.
Since $\frac{\partial S}{\partial t} = \cos(\theta)$, we obtain $\frac{\partial^2 S}{\partial t \partial t'} = \frac{\partial \cos(\theta)}{\partial t'} = -\sin(\theta)\frac{\partial \theta}{\partial t'} < 0$. Since $\theta \in (0, \pi)$, we have that $\sin(\theta) < 0$. It follows that $(\frac{\partial \theta}{\partial t'})^{-1} = \frac{\partial t'}{\partial \theta} > 0$. This proves the claim. $\square$

The smoothness of the map T is related to the smoothness of the billiard table. More precisely we have the following proposition.

Proposition 3. *T is a C^{k-1} diffeomorphism if $\mathcal{C}$ is C^k.*

Proof. We follow [3]. Let $r := -\cos(\theta)$, $r' := -\cos(\theta')$ and $x, x' \in S^1$. Define $G : S^1 \times (-1, 1) \times S^1 \times (-1, 1) \to \mathbb{R}^2$ by

$$G(x, r, x', r') = \begin{bmatrix} \frac{\partial}{\partial x'} S(x, x') - r \\ \frac{\partial}{\partial x} S(x, x') - r' \end{bmatrix} \tag{1}$$

Observe that G is C^{k-1} since $S = S(t, t')$ is C^k away from $\Delta = \{(t, t') : t = t'\}$ (because $\mathcal{C}$ is C^k). We will apply the implicit function theorem to the function G. Note that $G(x, r, x', r') = 0$ iff

$$\frac{\partial}{\partial x'} S(x, x') = r \text{ and } \frac{\partial}{\partial x} S(x, x') = r'.$$

That is, if (x', r') is the image of (x, r) under T by Lemma 2.1. The Jacobian of G w.r.t. x' and r' is given by

$$JG_{(x',r')} = \begin{bmatrix} \frac{\partial^2}{\partial x'^2} S(x, x') & 0 \\ \frac{\partial^2}{\partial x' \partial x} S(x, x') & -1 \end{bmatrix} \tag{2}$$

Then $\det JG_{(x',r')} = \frac{\partial^2}{\partial x'^2} S(x, x') < 0$ by the twist property, see Proposition 2. The implicit function theorem now says that we can solve, locally, for $x' = x'(x, r)$ and $r' = r'(x, r)$ which both are of class C^{k-1}. These functions are exactly the components of $T(x, r)$, proving that indeed T is a C^{k-1} map. Since the inverse T^{-1} is just T but with the orientation reversed, it follows that T is in fact a C^{k-1} diffeomorphism. $\square$

An *orbit* of T is defined to be a sequence $\{(s_i, \theta_i)\}_{i\in\mathbb{Z}}$ such that $(s_k, \theta_k) = T(s_{k-1}, \theta_{k-1})$. A *periodic orbit* is an orbit $\{(s_i, \theta_i)\}_{i\in\mathbb{Z}}$ for which for some $N > 1$, $T^N(s_i, \theta_i) = (s_i, \theta_i)$ holds for all $i \in \mathbb{Z}$. In Figure 5 two period 5 orbits

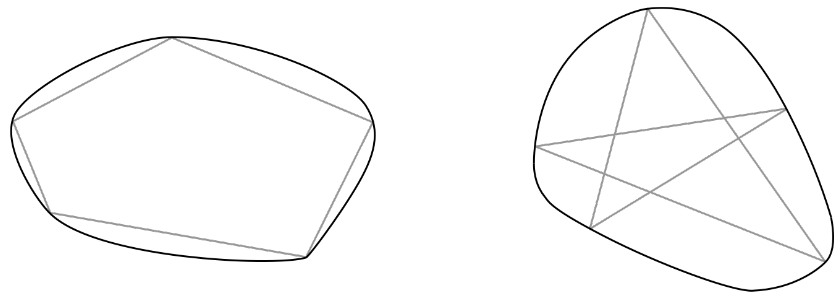

Figure 5: Two topologically different period 5 orbits

are shown.

Let us examine the orbits in Figure 5 a bit more. Clearly, these orbits are qualitatively different. What distinguishes them is the so called *rotation number* of the orbit. Recall that for any configuration it holds that $s_{i+1} = s_i + t_i$, where $t_i \in (0,1)$. Since the orbits are closed, the sum of the t_i's is an integer. This follows because, if p is the period of the orbit, then

$$s_p = s_{p-1} + t_{p-1} = s_{p-2} + t_{p-2} + t_{p-1} = ... = s_0 + \sum_{k=0}^{p-1} t_i.$$

But by periodicity we have that $s_0 = s_p \mod 1$. Hence $\sum_{k=0}^{p-1} t_i \in \mathbb{Z}$ and it is called the rotation number. Observe that this number ρ takes values in $1, ..., p-1$. However, we can assume $\rho \in \{1, ..., [\frac{p}{2}]\}$. This just follows since if we change the orientation of the billiard table $\mathcal{C}$, the rotation number ρ changes into $p - \rho$. This implies we can consider those values of ρ for which $1 \leq \rho \leq [\frac{p}{2}]$. One can also distinguish periodic orbits of the same period and rotation number by a certain type of ordering. In section 3 we will give precise definitions about the types of periodic there are in the billiard system. Our goal will be to prove existence results of certain types of periodic orbits of the map T for certain billiard tables.

A well know result concerning the existence periodic orbits is the following *Birkhoff Theorem*: For every $n \geq 2$ and $\rho \leq [(n-1)/2]$, coprime with n, there exist two geometrically distinct n-periodic billiard orbits with rotation number ρ. For a proof we refer to [4]. We will prove a similar statement but stated in the context of Aubry Mather theory, see section 5.

We remark that the map T has no fixed points, but T^n for $n > 1$ has. There are several ways of proving existence of periodic orbits. For example, a famous result from symplectic topology, known as Poincare's last theorem, can be used to prove existence of periodic points by means of proving there is a certain amount of fixed points of the map T^n. Instead of proving existence of fixed points, we will invoke Aubry-Mather theory to find periodic orbits. This approach will be developed in section 5.

Example 1. (Explicit formula for the billiard map) For a general billiard table it seems not so straightforward to obtain a formula for the map T. In case of a circular billiard $\mathcal{C}$ however, the billiard map is easy to determine. Assume the radius of $\mathcal{C}$ is equal to 1. Let α be the initial angle at some $t \in \mathcal{C}$ in which direction the ball is hit w.r.t. the billiard table. By an elementary geometric argument it is clear that the circular arc between two consecutive reflection points has length 2α, hence t gets mapped to $t + 2\alpha$. By symmetry the angular coordinate remains unchanged. Hence the billiard map for this circular billard is given by $T(t, \alpha) = (t + 2\alpha, \alpha)$.

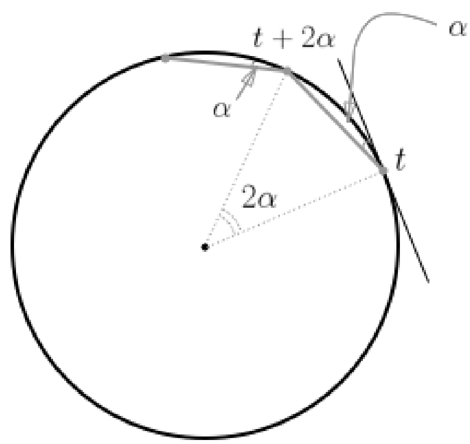

Figure 6: Finding the billiard map T in case of a circular billiard.

$\triangle$

Remark 2. *Although a circular billiards can be considered as (degenerate) elliptical billiards, we want stress that we mean by an elliptical billiard table, a non-degenerate ellipse. This convention will turn out to be convenient in formulating results at a later stage.*

3 Variational structure

A particularly convenient property that is useful for analyzing the billiard problem, is the presence of a variational structure. This is closely related to the fact that the billiard map T is a *twist map* of the cylinder. We will develop the theory of twist maps of the cylinder below and show how the variational structure appears in this context. We will first state some definitions.
An *annulus* is defined as

$$\mathbb{A} = S^1 \times (a, b),$$

where we consider $S^1 = \mathbb{R}/\mathbb{Z}$. The covering map $\pi : \mathbb{R} \times (a, b) \to S^1 \times (a, b)$ is defined as $\pi(x, y) = (x \bmod 1, y)$. We define a lift of some map f of the annulus $\mathbb{A}$ to be a map $\tilde{f} : \mathbb{R} \times (a, b) \to \mathbb{R} \times (a, b)$ for which

$$\pi \circ \tilde{f} = f \circ \pi.$$

This implies that $\tilde{f}(x+1, y) = \tilde{f}(x, y) + (n, 0)$, where n is the degree of the lift. Note that $\tilde{f}$ is unique up to addition of $(n, 0)$, $n \in \mathbb{Z}$.

In the case of the billiard system, we can interpret the billiard map T as a map of the annulus, $T : \mathbb{A} \to \mathbb{A}$, where in this setting $\mathbb{A} = \mathbb{S}^1 \times (0, \pi)$. Note that the annulus $\mathbb{A}$ is open, which due to the fact that the billiard domain is strictly convex. See the discussion in section 1. The cylinder in this context is the phase space of the billiard system. Recall that the coordinates of the phase space as defined in the section 2 are (s, θ). We can lift the map T which amounts in unfolding the annulus (considered as a finite cylinder). In the lift we switch to coordinates (x, y), where the x-coordinate corresponds to the s-coordinate, only now on $\mathbb{R}$, and $y = -\cos(\theta)$. The lift of T is then denoted as $\tilde{T} : \mathbb{R} \times (0, \pi) \to \mathbb{R} \times (0, \pi)$ and write $\tilde{T}(x, y) = (X(x, y), Y(x, y))$. It is natural in the case of the billiard map to impose that $X(x+1, y) = X(x, y) + 1$, so that number of times the billiard ball goes around can get tracked via this definition. This is also reflected in the definition we have made in chapter 1 that $x_{i+1} - x_i \in (0, 1)$. Since the $Y = Y(x, y)$ is related to the angle which does not change when x makes a full turn around $\mathcal{C}$, we will define $Y(x+1, y) = Y(x, y)$. We remark that in practice we shall use the billiard map T and its lift $\tilde{T}$ interchangeably. Basically the only difference between these maps is that in the lift the rotation number is taken into account, while T is just a map on some billiard table $\mathcal{C}$ where we consider the positions on $\mathcal{C}$ modulo 1.

Definition 3.1. *A cylinder map $F : \mathbb{A} \to \mathbb{A}$ is an* exact symplectic positive twist map *if*

1. ***Degree one*** : *$\tilde{F}$ is of degree one: $\tilde{F}(x+1, y) = \tilde{F}(x, y) + (1, 0)$.*
2. ***Exact symplectic*** : *The one-form $F^*(ydx) - ydx$ is exact.*
3. ***Positive Twist****: The map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\phi(x, y) = (x, X(x, y))$ is a diffeomorphism.*

The the positive twist condition means that every vertical fiber $f_x := \{x\} \times \mathbb{R}$ gets twisted around the annulus $\mathbb{A}$. Alternatively this implies that under T the fiber f_x is the graph of a function. In the billiard system the twist property means that increasing the angle yields an increasing in the position as well. This was shown in Proposition 2.

We will now verify that the billiard map T is an exact symplectic map, which is to say that the area is preserved and the *flux* is zero. This last property means that the volume enclosed by cycles $\gamma(t) := (t \bmod 1, 0)$ and its homotopic image

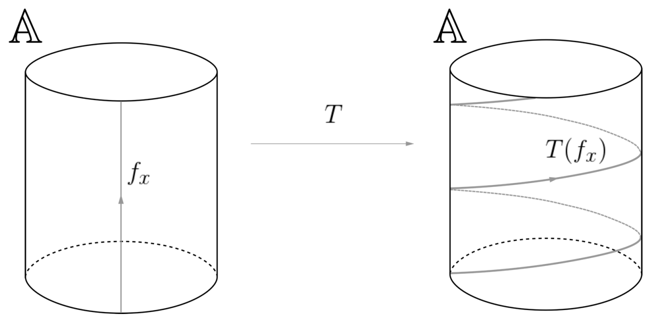

Figure 7: A fiber gets coiled around $\mathbb{A}$ under T

$T \circ \gamma$ is equal to zero.

Condition 2 says there exists a so-called generating function $s : \mathbb{A} \to \mathbb{R}$ such that $ds = T^*(ydx) - ydx$ is exact. This function s lifts to $\tilde{s} : \mathbb{R}^2 \to \mathbb{R}$ which satisfies $\tilde{s}(x, y) = s(x \text{ mod } 1, y)$. Consider the map ϕ from the third condition and write the inverse of ϕ as $\phi^{-1} : (x, X) \mapsto (x, y(x, X))$. The third condition allows us to define the function $S : \mathbb{R}^2 \to \mathbb{R}$ by $S(x, X) := \tilde{s}(x, y(x, X))$. This S will be called the *generating function*. We claim that the function $S(x, X) = -\|\gamma(x) - \gamma(X)\|$ which we defined as the generating function in chapter 1, but now on $\mathbb{R}^2$ and away from the diagonal Δ (see section 1), satisfies this property.

$$dS = \partial_x S dx + \partial_X S dX = -ydx + Y dX = -ydx + T^*(ydx),$$

where we applied Lemma 2.1 in the second step in the new coordinates x and y. Hence T is exact symplectic. Also note that, by periodicity of the parametrization γ, $S(x+1, X+1) = S(x, X)$. The reason that S is called generating, is that the lift $\tilde{T}$ can be recovered from S. To see this, first observe ϕ^{-1} is implicitly given by S: $\phi^{-1}(x, X) = (x, -\frac{\partial S}{\partial x})$, since $y = -\frac{\partial S}{\partial x}(x, X)$. We can write

$$\tilde{T}(x, y) = (X \circ \phi(x, y), \frac{\partial S}{\partial X}(\phi(x, y)).$$

To show that the third condition holds, we will make the following observations. It is clear that ϕ is smooth when $\mathcal{C}$ is smooth enough. The Jacobian $D\phi(x, y)$ is given $D\phi(x, y) = \frac{\partial X(x,y)}{\partial y}$. For billiards we already proved in Proposition 2 the twist condition. Translated in the lift coordinates this means that $\frac{\partial X(x,y)}{\partial y} > 0$. Hence $D\phi \neq 0$. This proves property 3.

Exact symplectic positive twist maps have the property that orbits of the map can be found by solving a variational monotone recurrence relation. In particular this means that one can forget about the second coordinate and consider only the first coordinates to determine the orbit. Intuitively this is clear: two points on the billiard table determine an angle and henceforth the third reflection and so on. Formally, this is expressed in the following lemma.

Lemma 3.1. *Let F denote the lift of a monotone twist map f. Let $S(x,X)$ denotes its generating function. Then there is one-one correspondence between the orbits $\{(x_k,y_k)=F^k(x_{k-1},y_{k-1})\}_{k\in\mathbb{Z}}$ of F and $\{x_k\}_{k\in\mathbb{Z}}$ satisfying*

$$\partial_1 S(x_k,x_{k+1})+\partial_2 S(x_{k-1},x_k)=0 \tag{3}$$

The correspondence is given by $y_k=-\partial_1 S(x_k,x_{k+1})$.

Proof. Assume that $\{(x_k,y_k)\}_k$ is an orbit of F. Then by Lemma 2.1 we have that $y_k=\partial_1 S(x_k,x_{k+1})$ and $y_{k+1}=\partial_2 S(x_k,x_{k+1})$. for all $k\in\mathbb{Z}$. The second equation then gives $y_k=\partial_2 S(x_{k-1},x_k)$. Combining the first and the last we obtain

$$\partial_1 S(x_k,x_{k+1})+\partial_2 S(x_{k-1},x_k)=0.$$

This proves one direction. For the other way around, assume $\{x_k\}_k$ satisfies (3) and that $y_k=-\partial_1 S(x_k,x_{k+1})$. Note that $\phi^{-1}(x_{k-1},x_k)=(x_{k-1},y_{k-1})$. Then

$$\begin{aligned} F(x_{k-1},y_{k-1}) &= F(\phi^{-1}(x_{k-1},x_k))=(x_k,\partial_2 S(x_{k-1},x_k)) \\ &= (x_k,-\partial_1 S(x_k,x_{k+1})=(x_k,y_k). \end{aligned} \tag{4}$$

□

Equation (3) can be seen as a discrete Euler-Lagrange equation in the following sense. Let $N<M$ and fix x_{N-1} and x_{M+1}. Define

$$W_{N,M}=\sum_{k=N-1}^{M} S(x_k,x_{k+1}).$$

Then the orbit segment $\{x_N,...,x_M\}$ is a path of the action functional $W_{N,M}$ iff $dW_{N,M}(\{x_N,...,x_M\})=0$, i.e. when $\{x_N,...,x_M\}$ is a critical point of $W_{N,M}$. We defined a billiard orbit as path satisfying angle of incidence is angle of reflection. We shall show that $dW=0$ indeed implies that we have the equal angle property.

Consider three points on a billiard table as shown in the Figure 8, where x_1,x_3 are fixed. The route from x_1 to x_3 that makes the action functional stationary, is via that point x_2 which makes equal angles with $\mathcal{C}$ (note that this is not necessarily the shortest route). Applying the differential d and invoking Lemma 2.1, we obtain

$$\begin{aligned} dW(x_1,x_2,x_3) &= -d(S(x_1,x_2)+S(x_2,x_3)) \\ &= (\partial_2 S(x_1,x_2)+\partial_1 S(x_2,x_3))dx_2 \\ &= (\cos(\varphi)-\cos(\varphi'))dx_2. \end{aligned}$$

Hence $dW=0$ iff $\cos(\varphi)=\cos(\varphi')$. That is, iff the $\varphi=\varphi'$ since $\varphi,\varphi'\in(0,\pi)$. The same holds for arbitrarily orbits segments by linearity of d and independence of the dx_k's.

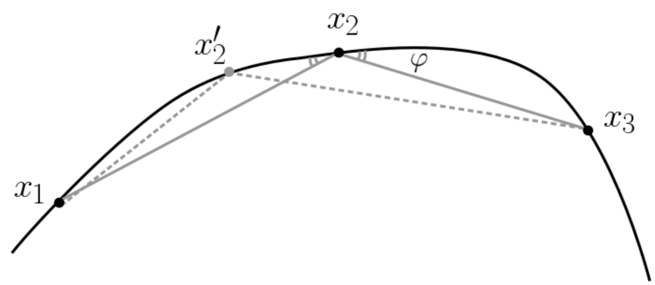

Figure 8: Angle of incidence equals angle of reflection

3.1 Periodic orbits

Periodic orbits in billiard systems were introduced in section 1. We will now discuss lifts of periodic orbits. Because of the presence of a variational structure, this a natural thing to do. Let $F : \mathbb{R} \times (a,b) \to \mathbb{R} \times (a,b)$ be the lift of an exact symplectic positive twist map $f : \mathbb{A} \to \mathbb{A}$. Let $\{(x_k, y_k)\}_{k\in\mathbb{Z}}$ be an orbit of F which satisfies for some $p, q \in \mathbb{Z}$

$$x_{k+p} = x_k + q. \tag{5}$$

Then $f^p(\pi(x_k, y_k)) = \pi(F^p(x_k, y_k)) = \pi(x_k + q, y_k) = \pi(x_k, y_k)$. Hence the F-orbit $\{(x_k, y_k)\}_{k\in\mathbb{Z}}$ satisfying (5) is the lift a periodic orbit of f with period p. We remark again that the variational structure simplifies when dealing with periodic orbits, in the sense that it is enough to consider only half of the coordinates.

Definition 3.2. *A sequence $\{x_k\}_{k\in\mathbb{Z}}$ satisfying (5) is called an (p,q)-sequence. The space of all (p,q)-sequences is denoted by $\mathbb{X}_{p,q}$. So*

$$\mathbb{X}_{p,q} = \{x \in \mathbb{R}^{\mathbb{Z}} : \forall k,\ x_{k+p} = x_k + q\}.$$

In the billiard system, periodic orbits of different pairs of p, q may give rise to qualitatively different orbits. This was illustrated in section 1 where a $(5,1)$ and $(5,2)$ orbit were drawn. The p in this case represents the number of reflections, while q is the winding number of the orbit. In section 1, we used the symbol ρ for the winding number. For our purposes it is clear that only those p, q are relevant for which $p > q \geq 1$.
In some cases there are pairs of p, q that can be excluded because they will lead to the same orbit. When one considers periodic orbits corresponding to minimizers, those pairs of (p,q) orbits are considered where p and q are relatively prime. The reason is that any $W_{p,q}$ minimizer is automatically a $W_{kp,kq}$ minimizer and vice versa. This will be proved in section 5. Furthermore, those pairs are being considered which have $q \leq \frac{p}{2}$ for reasons explained earlier, see section 1.
We record as a lemma the following obvious statement.

Lemma 3.2. *For all $n > 0$: $\mathbb{X}_{p,q} \subseteq \mathbb{X}_{np,nq}$.*

Proof. Let $x \in \mathbb{X}_{p,q}$. Then $x_{k+np} = x_{k+(n-1)p+p} = x_{k+(n-1)p}+q = ... = x_k+nq$. Hence $x \in \mathbb{X}_{np,nq}$. $\square$

4 Spaces of configurations

An important aspect of Aubry-Mather theory is that it deals with comparison principles amongst certain types of orbits. This comparison is made by a partial order on the space of configurations. We shall make precise in this section how this comparison in this context is being made.
We equip the space of all sequences $\mathbb{R}^{\mathbb{Z}}$ with a partial order. These come in various degrees of strictness.

Definition 4.1. *On the space of sequences $\mathbb{R}^{\mathbb{Z}}$ we define $\leq$, $<$, $\ll$ as*

- $x \leq y$ *if* $x_i \leq y_i$ *for all* $i \in \mathbb{Z}$
- $x < y$ *if* $x_i \leq y_i$ *for all* $i \in \mathbb{Z}$ *and* $x \neq y$
- $x \ll y$ *if* $x_i < y_i$ *for all* $i \in \mathbb{Z}$

The topology we put on the space $\mathbb{R}^{\mathbb{Z}}$ is what is known as the *topology of pointwise convergence*: let $x^{(j)} \in \mathbb{R}^{\mathbb{Z}}$. Then $\lim_{j\to\infty} x^{(j)} = x$ if $\lim_{j\to\infty} x_k^j = x_k$ for all $k \in \mathbb{Z}$.

In the theory of circle homeomorphisms there is a notion of ordered orbits. More precisely, if $F : \mathbb{R} \to \mathbb{R}$ is a lift of an orientation preserving homeomorphism $f : S^1 \to S^1$, then it is well known that $F(x+1) = F(x)$ holds and that this implies that if $\{x_k\}_{k\in\mathbb{Z}}$ is an orbit of F, then $x_k \leq x_j+p \Rightarrow x_{k+1} \leq x_{j+1}+p$, for all $k, j, p \in \mathbb{Z}$. It is this definition that is used to define a well-ordering on $\mathbb{R}^{\mathbb{Z}}$. We have then the following definition:

Definition 4.2. *A sequence $\{x_k\}_{k\in\mathbb{Z}}$ is called* well-ordered *or* Birkhoff *if*

$$x_k \leq x_j + p \Rightarrow x_{k+1} \leq x_{j+1} + p,$$

for all $k, j, p \in \mathbb{Z}$. A sequence $\{x_k\}_{k\in\mathbb{Z}}$ is called unordered *or* non-Birkhoff *if it is not Birkhoff. A sequence $\{(s_i, \theta_i)\}_{i\in\mathbb{Z}}$ is (non)- Birkhoff if $\{s_i\}_{i\in\mathbb{Z}}$ is a (non)-Birkhoff sequence.*

We will see later that for periodic orbits the Birkhoff property implies the orbit has the "same structure" as a regular polygon. We shall now discuss some properties of the above introduced notions of Birkhoff and non-Birkhoff orbits and what there structure is in the phase space of the billiard map. We begin with the following definition.

Definition 4.3. *Let F denote the lift of a monotone twist map $f : \mathbb{A} \to \mathbb{A}$. A set $M \subseteq \mathbb{R}^2$ is called $F-$*ordered *if, for $z, z' \in M$,*

$$\pi(z) < \pi(z') \implies \pi(F(z)) < \pi(F(z')),$$

where π is the x-projection.

Definition 4.3 can be used to obtain an alternative characterization of the Birkhoff property of sequences. It is easy to see, cf. [1], that an orbit $\mathcal{O} = \{(x_k, y_k)\}_k$ is Birkhoff if and only if its points form a F-ordered set. On the annulus or in the phase space the points of a Birkhoff sequence $\mathcal{O}$ have a special structure: the y coordinates form a Lipschitz graph over its projection. This is the content of the following proposition.

Proposition 4. *Let $M \subseteq \mathbb{R}^2$ be F-ordered. Then there exists a real number $K = K(F)$ such that if (x, y) and (x', y') are in M, then*

$$|y - y'| \leq K|x - x'|.$$

For a proof we refer to [1].

Definition 4.4. *Let T denote the billiard map of some $\mathcal{C}$. A* rotational invariant *curve $\mathcal{Y}$ (RIC) is a (smooth) homotopically non-trivial curve in the phase space $\mathbb{S}^1 \times (0, \pi)$ such that $T(\mathcal{Y}) = \mathcal{Y}$.*

As a direct consequence of Proposition 4, RIC's are (Lipschitz) graphs of the annulus. In particular, if there exists an invariant closed curve that is not a graph of the annulus, i.e. a homotopically trivial curve, orbits on this curve are necessarily non-Birkhoff. We already remark that this explains why the NB (periodic) orbits in the phase diagram of the elliptical billiard map are to be found on the trivial loops in Figure 60, see section 8. For a proof we refer to [1].

Example 2. In this example we will show that the circular billiard system has no non-Birkhoff orbits.

Let $\mathcal{C}$ denote a circular billiard table and $T : (s, \theta) \mapsto (s + 2\theta, \theta)$ the corresponding billiard map, cf. example 1. It follows from the form of T that the curves $\{\theta = c\}$ for $c \in (0, \pi)$ are invariant under T, so that the whole phase space is foliated by curves $\{\theta = c\},\ c \in (0, \pi)$. See Figure 9.

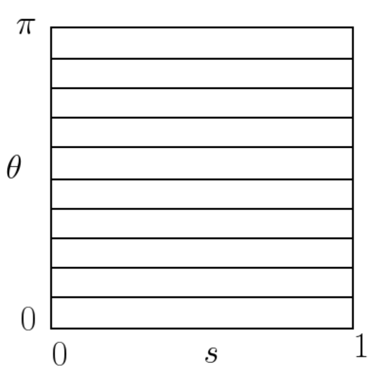

Figure 9: The phase space of a circular billiard system.

Furthermore, in $\mathbb{A} = \mathbb{R}/\mathbb{Z} \times (0, \pi)$ these curves are homotopically non-trivial. So in fact $\{\theta = c\}$ is a RIC. It is clear that each periodic orbit of T is T-ordered in the sense of definition 4.3. In particular, each orbit is Birkhoff. Hence there are no NB orbits in a circular billiard. △

For billiard systems, the existence of RIC's in the phase space give rise to interesting geometric structures inside the billiard table. Namely, one can show that when there is a RIC $\mathcal{Y}$ and (s, θ) a point on $\mathcal{Y}$, there is a curve γ inside or outside the billiard table such that all orbits segments remain tangent to γ. This curve is known as a *caustic*.

Definition 4.5. *A* caustic *is curve γ such that if one segment of an orbit $\mathcal{O}$ is tangent to it, then so are all other segments tangent to it.*

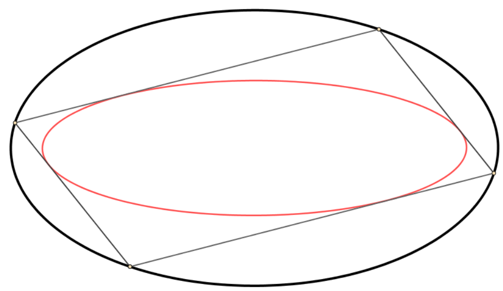

Figure 10: The smaller red ellipse is a caustic for the periodic orbit shown.

The study of caustics is an important subject in the theory of mathematical billiards. This is due to the close relationship between caustics and invariant circles, e.g. see [12]. We shall not dwell too much on this subject, but occasionally it proves its value when considering elliptical billiards.

In analogy to circle homeomorphisms, Birkhoff sequences, as well as non-Birkhoff sequences, can be assigned a so called *rotation number*.

Definition 4.6. *Let $x \in \mathbb{R}^{\mathbb{Z}}$. If $\rho(x) := \lim_{k\to\infty} \frac{x_k}{k}$ exists, it is called the* rotation number *of x.*

Lemma 4.1. Let $x \in \mathcal{B}$. Then $\rho(x) := \lim_{k\to\infty} \frac{x_k}{k}$ exists and

$$|x_k - x_0 - k\rho(x)| \leq 1.$$

The interpretation is that the number $\rho(x)$ for $x \in \mathcal{B}$ is an asymptotic measure for the average displacement $x_{k+1} - x_k$. Recall that the set $\mathbb{X}_{p,q}$ is the collection of all (p, q) periodic configurations. By definition of $\mathbb{X}_{p,q}$, for x it holds that $x_{k+p} = x_k + q$ for each $k \in \mathbb{Z}$. Consider now periodic configurations and recall that the set of such configurations is denoted by $\mathbb{X}_{p,q}$. We will compute for $x \in \mathbb{X}_{p,q}$ the rotation number $\rho(x)$. For any $k \geq 0$ we can write $k = m(k)p + r(k)$

where $r(k) < q$. The function $m = m(k)$ is non-decreasing and $m(k) \to \infty$ if $k \to \infty$. Then

$$\begin{aligned}
\rho(x) &= \lim_{k\to\infty} \frac{x_k}{k} \\
&= \lim_{k\to\infty} \frac{x_{m(k)p+r(k)}}{m(k)p + r(k)} \\
&= \lim_{k\to\infty} \left[\frac{x_{r(k)}}{m(k)p + r(k)} + \frac{m(k)q}{m(k)p + r(k)} \right] \\
&= \frac{q}{p}.
\end{aligned}$$

So in this case the interpretation of $\rho(x)$ just means that the average displacement per iterate equals $\frac{q}{p}$. The space of $x \in \mathcal{B}$ with rotation number $\rho(x) = \frac{p}{q}$ is denoted by $\mathcal{B}_{p/q}$ and we define $\mathcal{B}_{p,q} := \mathbb{X}_{p,q} \cap \mathcal{B}_{p/q}$.

Definition 4.7. *The* translation operator $\tau_{p,q} : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$, *is defined by* $(\tau_{p,q}x)_k := x_{k+p} + q$ *with* $p, q \in \mathbb{Z}$.

The translation operator will turn out to be useful when one deals with Birkhoff sequences. The subset of all Birkhoff sequences in $\mathbb{R}^{\mathbb{Z}}$ is denoted by $\mathcal{B}$. On $\mathbb{R}^{\mathbb{Z}}$ we use the action $\tau_{0,1} : x \mapsto x + 1$ to identify sequences that differ by an integer, i.e. each entry of the sequence differs by an integer. The resulting quotient space is denoted by $\mathbb{R}^{\mathbb{Z}}/\mathbb{Z}$. In terms of the billiard system, this just means all reflection points are shifted a whole number of times around the curve to its initial positions.

Lemma 4.2. *The space $\mathcal{B}_{p,q}/\mathbb{Z}$ is compact for the topology of pointwise convergence.*

Proof. First we observe that $\mathcal{B}_{p,q}$ is closed in the topology for pointwise convergence. This just follows since both "$\leq$" and "$\geq$" are being preserved while taking limits.
By Lemma 4.1, we have that $x_k - x_0 - k\rho(x) \in [-1, 1]$. This implies that $\mathcal{B}_{p,q}/\mathbb{Z}$ is a closed subset of

$$\{[x] \in \mathbb{R}^{\mathbb{Z}} / \mathbb{Z} : x_k = x_0 + k\rho(x) + y_k,\ (x_0, \rho(x), y) \in [0, 1] \times \{\rho(x)\} \times [-1, 1]^{\mathbb{Z}}\}.$$

By Tychonov's theorem this last set is compact for the topology of pointwise convergence, since it is a countable product of compact spaces. As $\mathcal{B}_{p,q}/\mathbb{Z}$ is a closed subset of a compact set, we conclude it is indeed compact. $\square$

We will now show that x is Birkhoff if and only if

$$\tau_{p,q}x \leq x \text{ or } \tau_{p,q}x \geq x,$$

for all $p, q \in \mathbb{Z}$. In practice, this alternative formulation will be employed in the forthcoming proofs.

Lemma 4.3. *A sequence $x \in \mathbb{R}^{\mathbb{Z}}$ is Birkhoff iff $\tau_{p,q}x \leq x$ or $\tau_{p,q}x \geq x$ for all $p, q \in \mathbb{Z}$.*

Proof. Assume x is Birkhoff and let $p, q \in \mathbb{Z}$. Assume w.l.o.g. that for some $k \in \mathbb{Z}$, $\tau_{p,q}x \leq x$, that is $x_{k+p} \leq x_k - q$. Then by definition 4.2

$$x_{k+p+1} \leq x_{k+1} - q.$$

Observe that in the formulation with the translation operators, this can be written as $(\tau_{p,q}x)_{k+1} \leq x_{k+1}$. Inductively, $(\tau_{p,q}x)_l \leq x_l$ holds for all $l \geq k$. To deal with $l < k$, assume for a contradiction that $(\tau_{p,q}x)_{l'} > x_{l'}$ for some $l' < k$. Then by the original definition of the Birkhoff property and a similar argument as before, $(\tau_{p,q}x)_m > x_m$ for all $m \geq l'$. So the translated x will remain above x. However, we assumed that for some $k > l'$, x is below the translate $\tau_{p,q}x$. This is a contradiction. Hence we showed that $\tau_{p,q}x \leq x$ for any choice of $p, q \in \mathbb{Z}$. $\square$

Well ordered sequences can be visualized via what is known as an *Aubry diagram*. Sequences in $\mathbb{R}^{\mathbb{Z}}$ can be interpreted as maps $\mathbb{Z} \to \mathbb{R}$, given by $i \mapsto x_i$, and in this way the dots in an Aubry diagram are plotted. So a dot is placed at (i, x_i) for each $i \in \mathbb{Z}$. Furthermore, consecutive dots are connected by straight line segments. Below an example is given of an $x \in \mathcal{B}$ and the translate $\tau_{1,-1}x$. Note furthermore that by our convention for billiards, that for consecutive reflection points x_i, x_{i+1}, it holds that $x_{i+1} - x_i \in (0, 1)$, so that the Aubry graph is the graph of a strictly increasing function.

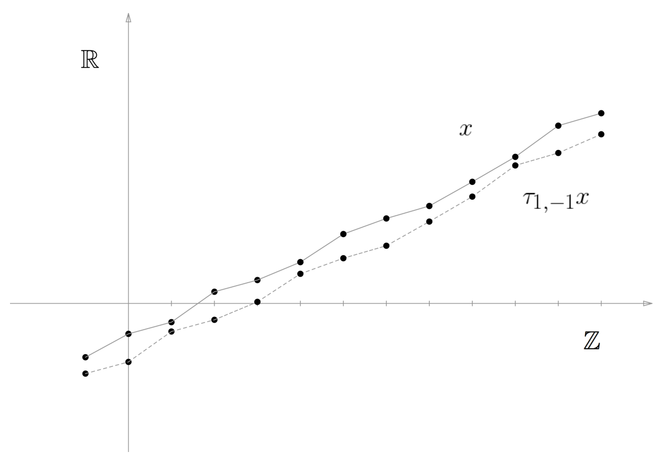

Figure 11: A Birkhoff sequence x and the translate $\tau_{1,-1}x$ (dashed).

For the analysis of billiard orbits, Aubry diagrams can be useful. We will now discuss some aspects of these types of graphs.

Two sequences x and y may *cross* each other and there are two ways this can happen. We say that there is a *crossing* between two sequences x and y at an integer i, if $x_i - y_i$ and $x_{i+1} - y_{i+1}$ have the same sign.

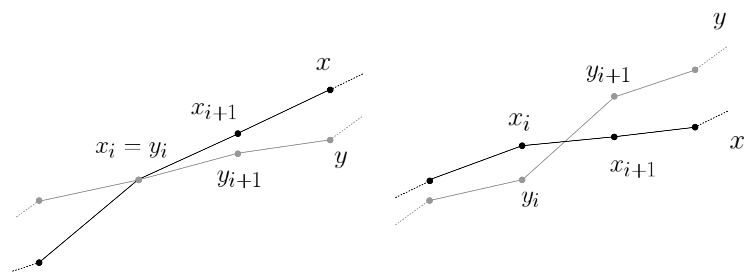

Figure 12: The two types of crossings

The other way this can happen is that x and y may cross at a non-integer $t \in (i, i+1)$ in which case $(x_i - y_i)(x_{i+1} - y_{i+1}) < 0$.

We will now make some observations about Aubry diagrams. A crossing in the Aubry diagram is not directly related to a crossing in the billiard table itself. An easy example is given by two $(2,1)$ orbits in an elliptical billiard system $\mathcal{E}$ which lie on the major and minor axis of the ellipse. In the billiard table they cross, while in the Aubry diagrams of these orbits they do not. Just note that each increment from i to $i+1$ yields an increment of $\frac{1}{2}$ for both graphs. Since they start at different positions, their Aubry graphs remain parallel.Visually, the Birkhoff property of sequence x entails that the graph of x does not intersect with any of its translates $\tau_{k,l}x$ with $k, l \in \mathbb{Z}$. In the billiard system this has the following interpretation. Consider an orbit $x \in \mathbb{R}^{\mathbb{Z}}$. Then $\tau_{k,0}x$ is the orbit x shifted over the vector $(-k, 0)$. For the billiard system it helpful to add a time factor. If we interpret the indices i in x_i as discrete time, meaning that to go from x_i to x_{i+1} takes one unit of time, the shifted orbit starts k time units earlier with going round: at time $i = 0$, x has value x_k. So the horizontal shift results in a cyclic permutation of the vertices. The other type of translation, the vertical shifts, have the following interpretation: a vertically shifted x over l units, already made l complete turns over $\mathcal{C}$ at $i = 0$. The two translations combined results in starting traversing the configuration at $i = 0$ from some (possibly different) vertex. In the case of NB orbits, an appropriate combination of k and l will yield an intersection with $\tau_{k,l}x$ and x itself.

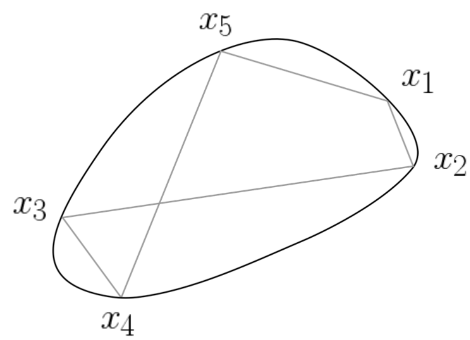

Figure 13: A $(5,2)$ non-Birkhoff orbit.

We now demonstrate with an example how the Birkhoff/non-Birkhoff property for billiard orbits in the corresponding Aubry diagram appear. Consider the $(5,2)$ orbit in Figure 13 and assume it is orientated clockwise. Intuitively we see that in order to go from x_3 to x_4 clockwise, a relatively large distance is needed to bridge. This amounts in a steep slope, compared to the other slopes, from $i = 3$ to $i = 4$ in the corresponding Aubry diagram. This causes its translate $\tau_{1,0}x$ to intersect with x itself, so that x is non-Birkhoff.

In Figure 14, the Aubry diagram of a non-Birkhoff configuration is shown in the right figure. Observe that some translate $\tau_{k,l}x$ (dashed) crosses x. Birkhoff orbits on the other hand, shown on the left in Figure 14, have the property the distances are all more or less close to each other. This implies there is no crossing of the orbit itself with a translate in the Aubry diagram.

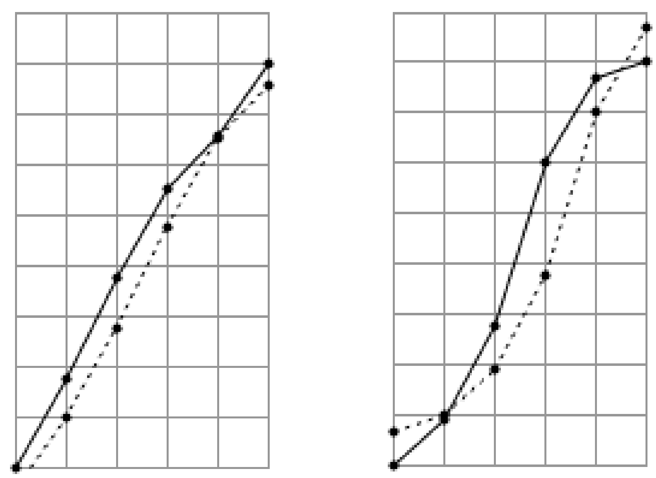

Figure 14: Aubry diagrams of a Birkhoff sequence and a translate (left) and a non-Birkhoff sequence and its translate (right). Note the crossing of the non-Birkhoff sequence with some translate of itself.

There is an alternative description of the Birkhoff and non-Birkhoff property for billiard orbits. Before stating this description we remark that instead of working in a lift of the billiard map, the discussion below applies to the billiard table itself directly. This has as a consequence that we only see the relative position of the reflection points x_i on $\mathcal{C}$, so we do not see the strict increasing nature of the graphs as they are in its Aubry graph.

Definition 4.8. *Fix an orientation of the billiard curve $\mathcal{C}$, say counterclockwise. Define two reflection points x and y on $\mathcal{C}$ to be* consecutive *if, when moving counterclockwise from one to another, there is no vertex in between x and y. Notation $x \prec y$.*

We remark that "$\prec$" is not transitive.

Proposition 5. *Let $\mathcal{C}$ be a billiard table. An orbit $\mathcal{O} = \{(s_i, \theta_i)\}_{i\in\mathbb{Z}}$ with $s_k \in \mathcal{C}$ is Birkhoff if and only for each i, j, it holds that $s_i \prec s_j \implies s_{i+1} \prec s_{j+1}$. Equivalently, writing $z = (s, \theta) \in \mathcal{O}$ and $z' = (s', \theta') \in \mathcal{O}$,*

$$\pi(z) \prec \pi(z') \implies \pi(T(z)) \prec \pi(T(z')),$$

where π is the projection on the first coordinate.

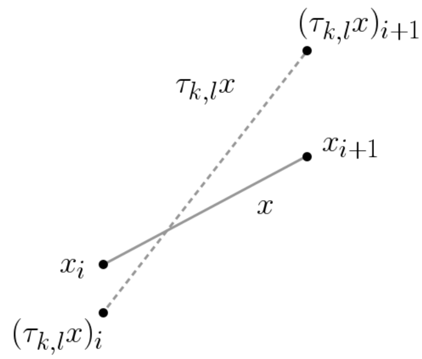

Figure 15: x and $\tau_{k,l}x$ intersect in the interval $[i, i+1]$.

Proof. In Figure 15 this proposition is illustrated for two billiards. Assume x is non-Birkhoff. Then there are some $k, l \in \mathbb{Z}$ such that $\tau_{k,l}x$ intersects with x by definition of the non-Birkhoff property. Suppose the Aubry graphs intersect between i and $i+1$, see figure 15. We may assume that x_i and $(\tau_{k,l}x)_i$ are consecutive reflection points on $\mathcal{C}$ and similarly for x_{i+1} and $(\tau_{k,l}x)_{i+1}$. But this means that while $x_i \prec (\tau_{k,l}x)_i$, it does not hold that $x_{i+1} \prec (\tau_{k,l}x)_{i+1}$. Since $(\tau_{k,l}x)_i$ is just some x_k and $(\tau_{k,l}x)_{i+1}$ is x_{k+1}, the first direction follows.

To prove the second direction, we argue as follows. Assume that x is Birkhoff. Then for any k, l, $\tau_{k,l}x \le x$ or $\tau_{k,l}x \ge x$. Assume the first case. Since there is no crossing, it follows that the ordering is automatically preserved. If x_i and $(\tau_{k,l}x)_i$ are consecutive in the sense of definition 4.8, then there cannot lie a x_k in between x_{i+1} and $(\tau_{k,l}x)_{i+1}$ since otherwise the line segment from x_{k-1} to x_k will actually intersect one of the other line segments. Hence the same orderings property holds for x_{i+1} and $(\tau_{k,l}x)_{i+1}$. This proves that the Birkhoff property implies the order relation stated in Proposition 5. $\square$

Example 3. In Figure 16 a $(4,2)$ periodic orbit $\mathcal{O}$ in an ellipse $\mathcal{E}$ is visualized. Assume the orientation of $\mathcal{E}$ is counterclockwise. Observe that $s_4 \prec s_1$ but not $s_{4+1} = s_1 \prec s_{1+1} = s_2$. Hence $\mathcal{O}$ is a $(4,2)$ non-Birkhoff orbit. For the second periodic orbit $\mathcal{O}'$ we observe that for each i,j it does hold that $s_i \prec s_j \implies s_{i+1} \prec s_{j+1}$. Hence $\mathcal{O}'$ is a Birkhoff orbit. Observe its type is $(7,2)$.

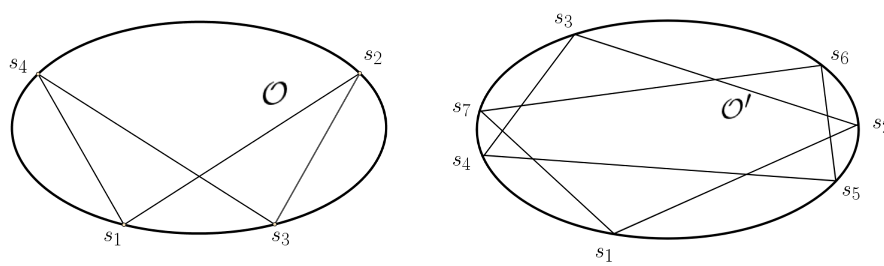

Figure 16: A non-Birkhoff orbit (left) and a Birkhoff orbit (right).

$\triangle$

The description given here will help us to prove an interesting characterization of periodic Birkhoff orbits in billiard systems. As already remarked, non-Birkhoff orbits share the property of possessing some kind of "irregularity". The following proposition makes this precise. I did not find this clear distinction as such in the literature. However, it is well known that orbits obtained by minimizing are (topological) *regular polygons*. Some additional remarks are in place now. First, by a regular polygon we mean the following: for integers p and q, it can be considered as being constructed by connecting every q-th point out of p points regularly spaced in a circular placement. By topological polygon in this context we obviously mean a configuration of straight line segments, ordered like a polygon. So a homeomorphic copy of a regular polygon, in which the "straightness" of the line segments is being preserved. Furthermore, the theory of regular polygons distinguish between convex polygons and star polygons. These polygons can be classified using the *Schläfli symbol*, $\{p\}$ or $\{p/q\}$. The convex class of polygons corresponds to orbits of type $(p,1)$, so with winding number 1 and their Schläfli symbol is $\{p\}$. A winding number $q \geq 2$ corresponds to star polygons with Schläfli symbol $\{p/q\}$, where q stands for that every q-th vertex is joined by a line segment. These types of polygons correspond to periodic orbits of type (p,q). Note the similarity between the

Schläfli symbol and the notation we are using. In Figure 17 examples are shown of the three possible 7-gons.

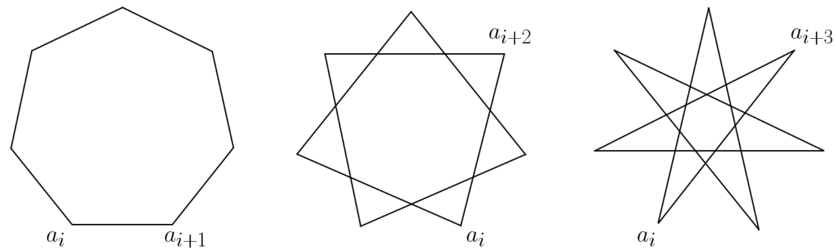

Figure 17: The three regular 7-gons with Schläfli symbols {7}, {7/2} and {7/3} respectively.

Proposition 6. *For billiard systems, periodic Birkhoff orbits of period $p \geq 3$ correspond to topological regular polygons.*

Proof. Suppose $\mathcal{O} = \{(s_i, \theta_i)\}_i$ is a (p, q) Birkhoff periodic orbit. This implies that when s_i joins s_{i+1} the same number of vertices are skipped as in the case where s_k and s_{k+1} are joined, since both s_i and s_k are consecutive as well as s_{i+1} and s_{k+1}. But this is precisely the definition of a regular polygon.
Regular polygons are indeed Birkhoff for the following reason. Consider a $\{p/q\}$ polygon a that is a periodic orbit inside some billiard domain $\mathcal{C}$. The orbit has the form $A = \{(a_{1+iq}, \theta_{1+iq})\}_i$. Fix some vertex a_1. Assume that the vertices are ordered like $a_1, a_2, ..., a_p$ counterclockwise. Then the line segment starting at a_1 goes to vertex a_{1+q} and similarly a_2 goes to a_{q+2}, by definition of a regular polygon. By our convention it holds that $a_{1+q} \prec a_{2+q}$. This just means that $a_1 \prec a_2 \implies \pi(T(a_1, \theta_1)) \prec \pi(T(a_2, \theta_2))$. Since this holds for all a_i with $i = 1, ..., p$, a is indeed Birkhoff, so that regular polygons satisfy the Birkhoff property. $\square$

The task of finding non-Birkhoff orbits in billiards has to do with finding irregular polygons inside billiards according to Proposition 6. Compared to proving existence of Birkhoff orbits, it seems more involved making claims about the existence of non-Birkhoff orbits. The last section of this thesis deals with a particular class of billiard tables in which NB orbits are present, namely those with non-trivial symmetry groups. Furthermore, we make the important observation that it is not necessary for a billiard table to have NB-orbits of a certain type (p, q). This is a consequence of the following recent theorem found by S. Pinto-de-Carvalho and R. Ramírez-Ros. For a proof, we refer to [6].

Theorem 1. *For any p, q relatively prime, there exists a billiard table for which there exist exactly two (p, q) periodic orbits.*

Now, by the results presented in section 5, there are always at least two orbits of different type (to be made precise at a later stage) for any given billiard table, under the assumption that $W_{p,q}$ is Morse. Furthermore, these will be shown to be Birkhoff. Hence in a table constructed as in the above theorem, the two orbits found should be Birkhoff and therefore no NB orbit can exists. We will see later, that for choices of non-relatively prime pairs (p, q), one can find NB-orbits. An important observation is that for all $n \in \mathbb{Z}_{\geq 1}$ the space $\mathcal{B}_{p,q}$ coincides with $\mathcal{B}_{np,nq}$. This will be fundamental in proving the existence of NB-orbits in certain types of billiard tables.

Proposition 7. *For any p, q and $n \in \mathbb{Z}_{\geq 1}$ it holds that $\mathcal{B}_{p,q} = \mathcal{B}_{np,nq}$.*

Proof. It suffices to show that any $x \in \mathcal{B}_{np,nq}$ has in fact period (p, q). Since $\mathcal{B}_{p,q} \subseteq \mathcal{B}_{np,nq}$ trivially, the claim follows.
Assume $x \in \mathcal{B}_{np,nq}$. By definition of $\mathcal{B}_{np,nq} \subseteq \mathbb{X}_{np,nq}$ it holds that $\tau_{np,-nq}x = x$. Clearly, $\tau_{np,-nq} = \tau^n_{p,-q}$. We argue by contradiction: if $x \notin \mathcal{B}_{p,q}$, then $\tau_{p,-q}x \neq x$. But since $x \in \mathcal{B}_{np,-nq}$, it is Birkhoff. In particular, it should hold that $\tau_{p,-q}x < x$ or $\tau_{p,-q}x > x$. Assume w.l.o.g. that the first case holds. Then this implies $\tau_{np,-nq}x = \tau^n_{p,-q}x < x = \tau_{np,-nq}x$. This is a contradiction. Hence any (np, nq) periodic sequence is (p, q) periodic as well. This proves the claim. □

Remark 3. *Proposition 6 might also be used to distinguish between Birkhoff and non-Birkhoff orbits in some cases. For example, consider the orbit x below of type $(6, 2)$ in a smooth triangular shaped billiard system. Since $\mathcal{B}_{3,1} = \mathcal{B}_{6,2}$ by Proposition 7, we only need to observe that x is not 3-periodic, i.e. is not in $\mathcal{B}_{3,1}$. But that is trivially true. Hence x is NB. Of course, one sees immediately that it is not a regular polygon. But later on, this remark proves itself useful for theoretical considerations.*

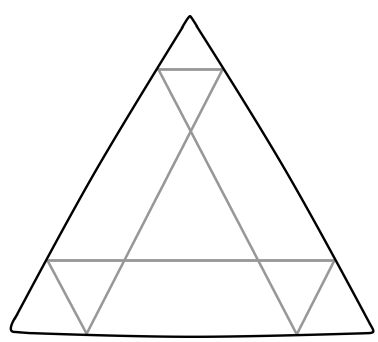

Figure 18: A $(6, 2)$ NB orbit

In section 5 we will see that Birkhoff periodic orbits are "abundant" in convex billiard tables: each convex billiard table has for each p, q at least one Birkhoff periodic orbit. In this sense the existence of these kind of orbits does not reveal much information about the nature of different convex billiard tables. Non-Birkhoff orbits however, are in this sense more interesting. It will turn out that the existence question of (p, q) periodic NB orbits, with p, q relatively prime, is related to the non-existence of *invariant curves*. It is well known that this latter property implies that the map T has positive topological entropy, which means there, in some sense, chaos present. We shall not dwell on this any further.

5 Aubry-Mather Theory

In the previous sections we formulated conditions that need to be satisfied for a sequence $\{x_i\}_{i\in\mathbb{Z}}$ to be a billiard orbit for a general billiard table $\mathcal{C}$. In this section we develop the basics of Aubry-Mather theory to find periodic orbits of any given rational rotation number $\frac{q}{p}$, with $p > q \geq 1$ The approach is to define a *periodic action functional* $W_{p,q}$ and show that on the space $\mathbb{X}_{p,q}$ it has a stationary point. This stationary point corresponds to a periodic orbit of the billiard map. In fact, Aubry-Mather theory deals only with minimizing solutions. In the billiard system, minimizers corresponds to orbits of maximum length with a prescribed number of reflections and winding number. The reason we choose to call orbits of maximal length minimizers, is that it is customary language in this subject. If the action $W_{p,q}$ has the Morse property, the minimizers found via Aubry-Mather theory will be used to construct other types of stationary points. This gives us another method of finding periodic orbits.

By a famous lemma of Aubry, these minimizers can be compared via an orderings principle which was developed in section 4. This principle basically says that minimizers can be strictly ordered w.r.t. to the $\ll$ relation. Furthermore, as we will see later, this has a consequence that the minimizers are Birkhoff. Note we can parametrize $\mathbb{X}_{p,q}$ by the variables $(x_1, ..., x_p)$, so that this finite segment represents the corresponding sequence. Unless where it leads to confusion, we shall not distinguish between the infinite sequence $x \in \mathbb{X}_{p,q}$ and its (finite) parametrization. The main results in this sections are adaptations of a more general setting described in [2].

Remark 4. *We will make the convention that $x_{p+1} = x_1$ in the expressions $S(x_i, x_{i+1})$. This has the advantage that we can consider $\sum_{i=1}^{p} S(x_i, x_{i+1})$ instead of $\sum_{i=1}^{p} S(x_i, x_{i+1}) + S(x_p, x_1 + q)$. This is allowed since it does not change the formula because of translation invariance of S: $S(x_p, x_1 + q) = S(x_p, x_1)$, see the discussion in section 3.*

Fundamental to theory of periodic orbits in billiard systems is the following proposition which relates periodic orbits to critical points of the action functional W.

Proposition 8. A configuration $x \in \mathbb{X}_{p,q}$ is the x-projection of a (p, q) periodic orbit $\{(x_k, y_k)\}_{k\in\mathbb{Z}}$ of the billiard map T if and only if it is a critical point of

$$W_{p,q}(x_1, ..., x_p) := \sum_{j=1}^{p} S(x_j, x_{j+1}). \tag{6}$$

Proof. Let d be the differential operator. Then taking the d of both sides of (6) yields

$$dW_{p,q}(x_1,...,x_p) = \sum_{j=1}^{p} (\partial_1 S(x_j, x_{j+1})dx_j + \partial_2 S(x_j, x_{j+1}))dx_{j+1}$$

Hence $dW_{p,q} = 0$ iff $\partial_2 S(x_k, x_{k+1}) + \partial_1 S(x_{k+1}, x_{k+2}) = 0$, for all $k = 1, ..., p$. These are exactly the Euler-Lagrange equations from Lemma 3.1, proving that indeed a critical point of the periodic action functional corresponds to periodic orbits of the monotone twist map T. □

By Proposition 8, finding periodic orbits in $\mathbb{X}_{p,q}$ corresponds to finding critical points of the action $W_{p,q}$. This fact will be of considerable importance for the rest of this thesis.

The functional $W_{p,q}$ is shift invariant: $W_{p,q}(\tau_{k,l}x) = W_{p,q}(x)$. This is an easy calculation by just using the definition $W_{p,q}(x_1, ..., x_p) = \sum_{j=1}^{p} S(x_j, x_{j+1})$. Let $x \in \mathbb{X}_{p,q}$. Then

$$\begin{aligned} W_{p,q}(\tau_{k,l}x) &= \sum_{j=1}^{p} S((\tau_{k,l}x)_j, (\tau_{k,l}x)_{j+1}) \\ &= \sum_{j=1}^{p} S(x_{k+j} + q, x_{k+j+1} + q) \\ &= \sum_{j=1}^{p} S(x_j, x_{j+1}) \\ &= W_{p,q}(x_1, ...x_p), \end{aligned} \tag{7}$$

where we made use of the translation invariance of S and in the final last step that by definition it holds that $x_{k+p} = x_k + q$ for all $k \in \mathbb{Z}$. Of course this is clear on an intuitive level: relabeling the vertices does not change the length of the configuration.

Now we will find configurations in $X_{p,q}$ that minimize $W_{p,q}$. In the classical theory the *coerciveness* [3] is used to prove existence of minimizers. However, since for billiards it holds that $S(x, X) = -\|\gamma(x) - \gamma(X)\|$ is bounded from below, it cannot go to infinity. So in the context of billiards this property does not hold for the generating function S. Fortunately, with some minor modifications, we can still prove $W_{p,q}$ attains it minimum on $\mathbb{X}_{p,q}$.

Theorem 2. *The action $W_{p,q}$ attains its minimum on $\mathbb{X}_{p,q}$.*

[3] In this context coerciveness means $\lim_{|X-x|\to\infty} S(x, X) = \infty$

Proof. Note that $\mathbb{X}_{p,q}$ is not compact. However, since it holds that $x_{i+1} - x_i \in (0,1)$, we have the following situation. The space of configurations with a given rotation number q is topologically the product of S^1 and a $(p-1)$-dimensional ball. $W_{p,q}$ assumes its minimum on the closure of

$$M_q = \{(x_0, x_1, ..., x_{p-1}) : x_0 \in S^1,\ x_{i+1} = x_i + t_i,\ t_i \in (0,1),\ \sum_{i=0}^{p-1} t_i = q\}.$$

If we can show its minimum does not occur on the boundary of M_q, then we are done since the minimum is then attained in the interior of M_q, and therefore it is a critical point. But this is straightforward. If a minimum did (would) occur at the boundary, then the resulting periodic orbit is in fact a k-gon with $k < p$. This holds since on the boundary of M_q, some points may coincide. But the length of the periodic orbit increases when points that did coincide, were moved apart because of the triangle inequality. Hence the negative of the length decreases so that indeed the minimum is in the interior. This implies the minimum is attained in a critical point.

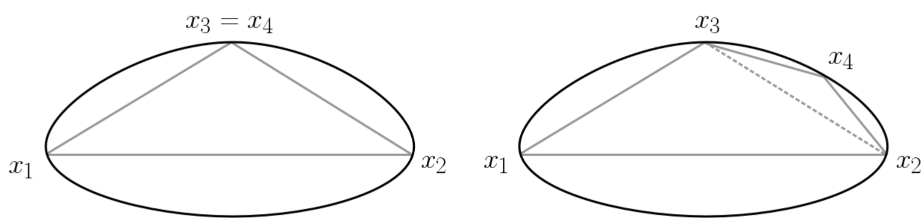

Figure 19: The length of the configuration on the left increases if x_3 x_4 are moved apart.

□

Configurations in $\mathbb{X}_{p,q}$ that minimize $W_{p,q}$ are called (p,q)-minimizers. Later we shall prove existence of other types of stationary configurations in $\mathbb{X}_{p,q}$ which are not-minimizing, but correspond to a "saddle point" of the action functional.

We shall need the following lemma about minimizers in the proof of the famous Aubry's Lemma, which says that minimizers on some space $X_{p,q}$ are ordered in some way.

Lemma 5.1. Let $x, y \in \mathbb{X}_{p,q}$ be minimizers of $W_{p,q}$ and let $m := \min\{x,y\}$ and $M := \max\{x,y\}$. Then m and M are again minimizers of $W_{p,q}$.

Proof. First observe that $m, M \in \mathbb{X}_{p,q}$. Indeed: $m_{k+p} = \min\{x,y\}_{k+p} = \min\{x_{k+p}, y_{k+p}\} = \min\{x_k + q, y_k + q\} = \min\{x,y\}_k + q = m_k + q$. The same holds for M. Write $\alpha := M - x$ and $\beta := m - x$ and observe that $\alpha > 0$ and $\beta < 0$. Then it holds that $\alpha_i \beta_i = 0$. To see this, assume for a contradiction that $\alpha_i > 0$ and $\beta_i < 0$. Then the first inequality gives $x_i < \max\{x_i, y_i\}$ while the second one gives $x_i > \min\{x_i, y_i\}$. This is not possible, so indeed $\alpha_i \beta_i = 0$. We can write $y = M + m - x = \alpha + m = \alpha + \beta + x$. If we show that

$$W_{p,q}(x) + W_{p,q}(y) \geq W_{p,q}(M) + W_{p,q}(m), \tag{8}$$

then we are done. This holds since x and y are both global minima, so the same must then hold for m and M (otherwise x or y cannot minimize the action $W_{p,q}$).
For convenience write $S_j(x) = S(x_j, x_{j+1})$. Then the left hand side can be written as

$$\begin{aligned}
W_{p,q}(x) + W_{p,q}(y) &= \int_0^1 \int_0^1 \frac{\partial^2}{\partial t \partial s} W_{p,q}(x + \alpha t + \beta s) ds dt \\
&= \int_0^1 \int_0^1 \frac{\partial^2}{\partial t \partial s} \Big(\sum_{j=1}^{p} S_j(x + \alpha t + \beta s\Big) ds dt \\
&= \sum_{k,i=1}^{p} \sum_{j=1}^{p} \left(\int_0^1 \int_0^1 \partial_{i,k} S_j(x + \alpha t + \beta s ds dt \right) \alpha_i \beta_k.
\end{aligned}$$

Now, for all $j \neq k$ it holds by the twist condition that $\partial_{i,k} S_j < 0$. At the same time, $\alpha_i \beta_k < 0$ so that all for all $i \neq k$ the above expression is positive. For $i = k$ we have shown that $\alpha_i \beta_i = 0$, hence in for $i = k$ the above expression is zero (since $\partial_{i,i} S$ is bounded). This shows that indeed (8) is satisfied, hence m and M are minimizers as well. $\square$

The next important proposition, known as Aubry's Lemma, yields the well ordering property of minimizers. One of the consequences of this lemma is that it can serve as a tool for distinguishing between the (global) minimum and local minima and also between minima and other types of critical points. We will later see an example of this.

Lemma 5.2. *(Aubry's Lemma) Let $x, y \in \mathbb{X}_{p,q}$ be minimizers of $W_{p,q}$. Then its holds that either $x \ll y$ or $y \ll x$.*

Proof. Let again $m = \min\{x, y\}$. Assume w.l.o.g. that $x < m$ but not $m \ll x$. Then there exist indices i, k with $|i - k| = 1$ for which $x_i = m_i$ and $x_k \geq m_k$ (otherwise $m \ll x$ which we excluded). We compute

$$\begin{aligned}
\sum_{j=1}^{p} (\partial_i S_j(x) - \partial_i S(m) &= \sum_{j=1}^{p} (\partial_i S_j(x) - \partial_i S_j(m)) \\
&= \int_0^1 \frac{d}{dt} \partial_i S_j(tx + (1-t)m) dt \qquad (9) \\
&= \sum_{j,l=1}^{p} \left(\int_0^1 \partial_{i,l} S_j(tx + (1-t)m) dt \right) (x_l - m_l).
\end{aligned}$$

For every $l \neq i$, $\partial_{i,l} S_j(tx + (1-t)m)dt < 0$, by the twist property (Proposition 2) in the coordinates introduced in section 3, while $(x_l - m_l) \geq 0$. Recall that $x_i = m_i$, so that in the above expression, the term with $l = i$ vanishes. So we do not need to worry about the term $\partial_{i,i} S_j$. Hence the above expression is strictly negative. However, since x and m are stationary points, $\nabla W_{p,q}(x) = \sum_{j=1}^{p} \partial_i S_j(x) = 0$ and similarly for m. Thus we get $0 = \sum_{j=1}^{p} \partial_i S_j(x) \neq \sum_{j=1}^{p} \partial_i S_j(m) = 0$. Hence we obtain a contradiction. Therefore the strict order in the statement of the lemma holds. $\square$

Corollary 5.1. *Periodic minimizers are Birkhoff configurations.*

Proof. Let $x \in \mathbb{X}_{p,q}$ be a minimizer. Then $\tau_{k,l}x$ is a minimizer as well since $W_{p,q}$ is $\tau_{k,l}$ invariant. Now, either $\tau_{k,l}x = x$ or Aubry's Lemma implies that either $\tau_{k,l}x \ll x$ or $\tau_{k,l}x \gg x$ holds. This means that $\tau_{k,l}x \leq x$ or $\tau_{k,l}x \geq x$. Hence x is Birkhoff. $\square$

In general it does not hold that all Birkhoff periodic orbits in a billiard system necessarily are minimizers of the action $W_{p,q}$. This follows from the mountain pass lemma, see Lemma 6.3. However, as we will see in Proposition 16, for ellipses it holds true that all minimizing orbits are necessarily Birkhoff. Below a $(8,3)$ periodic minimizer in an ellipse is shown. Observe that this orbit is a topological regular polygon with Schläfli symbol $\{p/q\}$, cf. section 4.

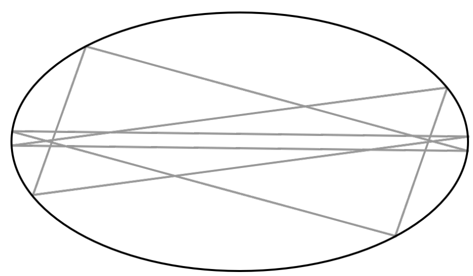

Figure 20: A minimizing $(8,3)$ Birkhoff orbit inside an ellipse

As a consequence of Corollary 5.1, non-Birkhoff orbits cannot be minimizers of $W_{p,q}$. Earlier we remarked that we assumed p and q in the context of (p,q)-minimizers are relatively prime. The reason for this is that for each n, (np,nq)-minimizers are actually $(p,q)-$minimizers as well. Hence a (np,nq) periodic orbit corresponds to (p,q) periodic orbit traversed n times.

Proposition 9. *Let $n \in \mathbb{N}$. Every p,q minimizer is a np,nq minimizer and vice versa.*

Proof. Suppose $x \in \mathbb{X}_{np,nq}$ is $W_{np,nq}$ minimizing, but not p,q periodic. Then $\tau_{p,q}x \neq x$. By Aubry's Lemma we have that $\tau_{p,q}x \ll x$ or $\tau_{p,q}x \gg x$ since both are $W_{np,nq}$ minimizing. Since $\tau_{k,l}$ preserves the ordering, we obtain $\tau_{p,q}^n x =$

$\tau_{np,nq}x \ll x$ or $\tau_{np,nq}x \gg x$, which is a contradiction. Hence any $W_{np,nq}$ minimizer is also a $W_{p,q}$ minimizer.
For the other way around assume that $x \in \mathbb{X}_{p,q}$ minimizes $W_{p,q}$. Observe that also $x \in \mathbb{X}_{np,nq}$:

$$x_{k+np} = x_{k+(n-1)p+p} = x_{k+(n-1)p} + q = ... = x_k + nq.$$

Then it holds that $W_{np,nq}(x) = nW_{p,q}$. Hence if x minimizes $W_{p,q}$, then it also minimizes $W_{np,nq}$. $\square$

6 The Gradient Flow

One of the tools to find periodic orbits that do not have the well-ordered property, is that of the *gradient flow* of the periodic action $W_{p,q}$. In fact is it not only useful in this particular case, but also when one is interested in well-ordered periodic orbits: we will give an alternative proof of the result in section 5 to obtain periodic orbits of any type (p, q).
The interpretation is that when some periodic configuration in $\mathbb{X}_{p,q}$ is allowed to flow, then it might flow to some periodic orbit in the same space. This is a consequence of the fact that periodic orbits correspond to critical points of the action functional $W_{p,q}$. In Figure 21 this is illustrated.

In this section we will introduce the notion of a formal gradient flow of W on some appropriate Banach space of sequences. Firstly, we will give an alternative proof of Theorem 2 about the existence of a minimizer in $\mathbb{X}_{p,q}$. Furthermore, we will derive properties such as the monotonicity of the flow, and introduce a comparison principle. Ultimately, what is developed here comes in play in the last section in proving existence of non-Birkhoff orbits. All results in this section concern infinite dimensional systems. Since we are mainly interested in periodic orbits we could have sticked to the finite dimensional theory. However, we choose not to do so for the following reason. Not only periodic orbits are interesting for research, but also interesting questions can be posed concerning not necessarily periodic orbits. Especially those orbits which have an irrational rotation number are worth studying: we refer to section 8.1 where we described a relation between the existence of such orbits with irrational rotation number and the non-existence of non-Birkhoff orbits with rotation number "close to" that particular irrational. It is therefore natural in the context of billiards not just to stick with the finite dimensional theory. Furthermore, the theory does not become particularly longer or more difficult by presenting it in this way. The results found in this section clearly hold for the periodic action $W_{p,q}$ as well.

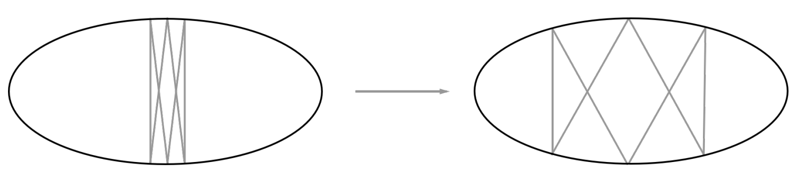

Figure 21: The configuration on the left converges to the periodic orbit on the right.

We consider the following infinite dimensional ODE system

$$\dot{x}_k = -\nabla W(x)_k \qquad (k \in \mathbb{Z}), \tag{10}$$

where $\nabla W : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ is defined as $\nabla W(x)_k = \partial_1 S(x_{k-1}, x_k) + \partial_2 S(x_k, x_{k+1})$. Note that we need the generating function S to be at least C^2 for the system to be well defined. For the billiard system this obtained by assuming the boundary curve is at least C^2 as well. We remark that that x is a *globally* stationary solution if and only if $\nabla W(x) = 0$.

Equip $\mathbb{R}^{\mathbb{Z}}$ with the norm

$$\|x\|_{\mathbb{X}} = \sum_{k=-\infty}^{\infty} \frac{|x_k|}{2^{|k|}},$$

and define $\mathbb{X}$ to be subspace of $\mathbb{R}^{\mathbb{Z}}$ containing elements of finite norm. The space $\mathbb{X}$ is a Banach space: let $\mathbb{X} \ni x^j \to x$. Then $\lim_{j\to\infty} \|x^j - x\|_{\mathbb{X}} = 0$ and hence we get for large enough j that

$$\varepsilon > \sum_{i=-\infty}^{\infty} \frac{|x_i^j - x_i|}{2^{|i|}} \geq \sum_{i=-\infty}^{\infty} \frac{|x_i| - |x_i^j|}{2^{|i|}} = \sum_{i=-\infty}^{\infty} \frac{|x_i|}{2^{|i|}} - ||x^j||_{\mathbb{X}},$$

so that $\|x\|_{\mathbb{X}} < \infty$.
On bounded subsets of $\mathbb{X}$, the topology induced by $\|.\|_{\mathbb{X}}$ is equivalent to the product topology which in turn is equivalent to the topology of pointwise convergence. See for example [1].

Our first lemma states that Birkhoff configurations are contained in $\mathbb{X}$.

Lemma 6.1. *$\mathcal{B} \subseteq \mathbb{X}$.*

Proof. Let $x \in \mathcal{B}$. By Lemma 4.1 we have that $|x_0 - x_k - k\rho(x)| \leq 1$ for all k. This implies that $|x_k| \leq 1 + |x_0| + |k|\rho(x)$. Now we can estimate the norm of x by

$$\|x\|_{\mathbb{X}} = \sum_{k=-\infty}^{\infty} \frac{|x_k|}{2^{|k|}} \leq \sum_{k=-\infty}^{\infty} \frac{|k|\rho(x) + |x_0| + 1}{2^{|k|}} < \infty.$$

So indeed $x \in \mathbb{X}$. □

We now prove a lemma which is used in Proposition 9 about existence of the continuous gradient flow on $\mathbb{X}$.

Lemma 6.2. *Let $x \in \mathbb{X}$. Then the shifted sequence $\tau_{k,l}x \in \mathbb{X}$ and when $y \in \mathbb{X}$ we have that $\|\tau_{k,l}x - \tau_{k,l}y\|_{\mathbb{X}} \leq 2^{|k|}\|x - y\|_{\mathbb{X}}$.*

Proof. For $\|\tau_{k,0}x\|_{\mathbb{X}}$ we find

$$\|\tau_{k,0}x\|_{\mathbb{X}} = \sum_{i=-\infty}^{\infty} \frac{|x_{i+k}|}{2^{|i|}} \leq 2^{|k|} \sum_{i=-\infty}^{\infty} \frac{|x_{i+k}|}{2^{|i+k|}} = 2^{|k|}\|x\|_{\mathbb{X}} < \infty.$$

Hence

$$\|\tau_{k,l}x\|_{\mathbb{X}} = \|\tau_{k,0}x + l\|_{\mathbb{X}} \leq \|\tau_{k,0}x\|_{\mathbb{X}} + \|l\|_{\mathbb{X}} < \infty,$$

so $\tau_{k,l}x \in \mathbb{X}$ and so $\|\tau_{k,l}x - \tau_{k,l}y\|_{\mathbb{X}} = \|\tau_{k,0}(x - y)\|_{\mathbb{X}} \leq 2^{|k|}\|x - y\|_{\mathbb{X}}$. □

We now show that system (10) defines a Lipschitz continuous flow on $\mathbb{X}$ for the topology of pointwise convergence.

Proposition 10. *$\nabla W : \mathbb{X} \to \mathbb{X}$ is globally Lipschitz. Furthermore, the induced flow φ^t is Lipschitz as well.*

Proof. Let $x, y \in \mathbb{X}$. Note that $|\partial_{i,j} S| \leq C$ for some $C > 0$, since S is C^2 and bounded sets in $\mathbb{X}$ are compact. We estimate

$$\begin{aligned} |-\nabla W(y)_i + \nabla W(x)_i| &= |-[\partial_i S(y_i, y_{i+1}) + \partial_i S(y_{i-1}, y_i)] + [\partial_i S(x_i, x_{i+1}) + \partial_i S(x_{i-1}, x_i)]| \\ &= |\int_0^1 \frac{d}{d\tau}(\sum_{k=i-1}^{i} \partial_i S_k(\tau x + (1-\tau)y) d\tau| \\ &\leq \sum_{k=i-1}^{i} |\int_0^1 \sum_{j=k}^{k+1} \partial_{i,j} S_k(\tau x + (1-\tau)y) d\tau| \cdot |x_j - y_j| \\ &\leq C \sum_{k=i-1}^{i} \sum_{j=k}^{k+1} |x_j - y_j|. \end{aligned} \tag{11}$$

This implies that

$$\begin{aligned} \|-\nabla W(y) + \nabla W(x)\|_{\mathbb{X}} &\leq C \sum_{i=-\infty}^{\infty} \frac{1}{2^{|i|}} \sum_{k=i-1}^{i} \sum_{j=k}^{k+1} |x_j - y_j| \\ &= C \sum_{i=-\infty}^{\infty} \frac{1}{2^{|i|}} (|x_{i-1} - y_{i-1}| + |x_i - y_i| + |x_{i+1} - y_{i+1}|) \\ &= 3C \|x - y\|_{\mathbb{X}}. \end{aligned}$$

Hence $\|-\nabla W(y) + \nabla W(x)\| \leq L\|x - y\|$, so that $-\nabla W$ is Lipschitz continuous. Note that this also implies that $-\nabla W$ maps $\mathbb{X}$ into $\mathbb{X}$: Choosing $y = 0$, the above inequality becomes $\|-\nabla W(x)\|_{\mathbb{X}} \leq L\|x\|_{\mathbb{X}} + \|\nabla W(0)\|_{\mathbb{X}}$.
The fact that $-\nabla W$ is Lipschitz implies the global-in-time existence and uniqueness of the initial value problem $\dot{x} = -\nabla W(x)$, $x_0 = x(0)$. To prove that the induced flow is Lipschitz continuous as well, we argue as follows.

$$\begin{aligned} \|x(t) - y(t)\|_{\mathbb{X}} &\leq \|x(0) - y(0)\|_{\mathbb{X}} + \int_0^{|t|} \|\dot{x}(s) - \dot{y}(s)\|_{\mathbb{X}} ds \\ &= \|x - y\|_{\mathbb{X}} + \int_0^{|t|} \|\nabla W(x(\tau)) - \nabla W(y(\tau))\|_{\mathbb{X}} d\tau \\ &\leq \|x - y\|_{\mathbb{X}} + L \int_0^{|t|} \|x(\tau) - y(\tau)\|_{\mathbb{X}} d\tau. \end{aligned}$$

Now Gronwall's inequality yields

$$\|\varphi^t(x) - \varphi^t(y)\|_{\mathbb{X}} \leq \|x - y\|_{\mathbb{X}} \cdot \exp\big(\int_0^{|t|} 1 ds \big) = e^{Lt} \|x - y\|_{\mathbb{X}}.$$

Hence $\|\varphi^t(x) - \varphi^t(y)\|_{\mathbb{X}} \leq M\|x - y\|_{\mathbb{X}}$ with $M = e^{Lt} > 0$. So φ^t is Lipschitz continuous. □

The following result we need, states an intuitively clear fact, namely the equivariance of shift map $\tau_{k,l}$ w.r.t. the flow φ^t. What this means down in the billiard table, is that the gradient flow remains the same after relabeling the reflection points.

Proposition 11. $\varphi^t \circ \tau_{k,l} = \tau_{k,l} \circ \varphi^t$

Proof.

$$\begin{aligned}
\frac{d}{dt}(\tau_{k,l}x)_i = \frac{d}{dt}x_{k+i} &= -\nabla W(x)_{k+i} \\
&= -[\partial_1 S(x_{k+i}, x_{k+i+1}) + \partial_2 S(x_{k+i-1}, x_{k+i})] \\
&= -[\partial_1 S((\tau_{k,l}x)_i, (\tau_{k,l}x)_{i+1}) + \partial_2 S((\tau_{k,l}x)_{i-1}, (\tau_{k,l}x)_i) \\
&= -\nabla W(\tau_{k,l}x)_i,
\end{aligned} \tag{12}$$

where we made use of $\tau_{0,l}$ invariance of S. □

So if $t \mapsto x(t)$ is a solution of $\dot{x} = -\nabla W(x)$, then also $t \mapsto (\tau_{k,l}x)(t)$ is a solution of the same equation. In particular this implies that $\mathbb{X}_{p,q}$ is invariant under the flow: if $x \in \mathbb{X}_{p,q}$, i.e. if $\tau_{p,q}x = x$, then $\tau_{p,q}(\varphi^t(x)) = \varphi^t(\tau_{p,q}x) = \varphi^t(x)$. So $\varphi^t(\mathbb{X}_{p,q}) = \mathbb{X}_{p,q}$.

The ODE $\dot{x} = -\nabla W_{p,q}(x)$ has a property which is similar to the heat flow of parabolic PDE's. Namely, for the heat flow we have that if u, v are solutions to $\dot{x} = -\Delta(x)$, for which $v \leq u$ and $v \neq u$, as soon as the time is turned on, the graphs of $u(t,x)$ and $v(t,x)$ move apart while not touching anymore. The following proposition shows that a similar property holds for the gradient flow on $\mathbb{X}$.

Theorem 3. *Let $x, y \in \mathbb{X}$ such that $x < y$. Then $\varphi^t(x) \ll \varphi^t(y)$ for all $t > 0$*

Proof. Write $x(t) = \varphi^t(x)$ and $y(t) = \varphi^t(y)$. Let $u(t) = y(t) - x(t)$ and note that $u(0) > 0$ since otherwise $x(t) = y(t)$ for all $t > 0$. Note that u satisfies the following ODE:

$$
\begin{aligned}
\dot{u}_i(t) &= -\nabla W(y(t))_i + \nabla W(x(t))_i \\
&= -[\partial_i S(y_i(t), y_{i+1}(t)) + \partial_i S(y_{i-1}(t), y_i(t))] \\
&\quad + [\partial_i S(x_i(t), x_{i+1}(t)) + \partial_i S(x_{i-1}(t), x_i(t))] \\
&= \int_0^1 \frac{d}{d\tau}\Big(\sum_{k=i-1}^{i} \partial_i S_k(\tau x(t) + (1-\tau)y(t)) \Big) d\tau \qquad (13)\\
&= -\sum_{k=i-1}^{i} \sum_{j=k}^{k+1} \Big(\int_0^1 \partial_{i,j} S_k(\tau x(t) + (1-\tau)y(t)) d\tau \Big)(y_j(t) - x_j(t)) \\
&= (H(t)u(t))_i,
\end{aligned}
$$

where we used our previously made convention that $S_k(x) := S(x_k, x_{k+1})$. The operator $H(t) : \mathbb{X} \to \mathbb{X}$ is Lipschitz for all t. The proof of this is similar to the proof of Proposition 10. We aim to show that the solution of (13) is positive, meaning that $u_i(t) > 0$ for all $i \in \mathbb{Z}$ and all $t > 0$. This then implies that $x(t) \ll y(t)$ and hence $\varphi^t(x) \ll \varphi^t(y)$ for all $t > 0$.

Recall that for the generating function S, we have the following bounds: $\partial_{12} S := \partial_{1,2} S < 0$ by the proof of Proposition 2 and $\partial_{ii} S := \partial_{i,i} S < C$ for $i = 1, 2$. Then for some $M > 0$, and all t, the operator $\tilde{H}(t) := H(t) + M\mathrm{Id}$ is positive: $u \geq 0$ implies $\tilde{H}(t)u \geq 0$. Note that $\tilde{H}(t)$ is uniformly bounded since $H(t)$ is Lipschitz. This guarantees that $\dot{u} = H(t)u$ and $\dot{v} = H(t)v$ define well posed initial value problems. If $u(t)$ solves $\dot{u} = H(t)u$, then $v(t) := e^{Mt}u(t)$ solves $\dot{v}(t) = \tilde{H}(t)v(t)$:

$$
\dot{v}(t) = Me^{Mt}u(t) + e^{Mt}\dot{u}(t) = (M\mathrm{Id} + H(t))e^{Mt}u(t) = (M\mathrm{Id} + H(t))v(t).
$$

We will show that if v is a solution to $\dot{v} = H(t)v$, then $v_i(t) > 0$ for all i and $t > 0$. It is clear that then also $u_i(t) > 0$ for all i and $t > 0$.
We will solve $\dot{v}(t) = \tilde{H}(t)v(t)$ by Picard iteration, which is possible since $H(t)$ is Lipschitz. Write for a solution v:

$$
v(t) = \Big(\sum_{i=0}^{\infty} \tilde{H}^{(n)}(t) \Big) v(0).
$$

where $\tilde{H}^{(n)}$ are defined inductively by

$$
\tilde{H}^{(0)}(t) := \mathrm{Id} \text{ and } \tilde{H}^{(n)}(t) := \int_0^t \tilde{H}(s) \circ \tilde{H}^{(n-1)}(s) ds
$$

for $n \geq 1$. Observe that the positivity of $\tilde{H}(t)$ clearly implies that the $B^{(n)}(t)$ are positive as well: $\tilde{H}^{(1)}(t) = \int_0^t \tilde{H}(s)ds$ is positive, since $\tilde{H}(t)$ is. Assume it holds for $n = k$, then $\tilde{H}^{(k)}(t) = \int_0^t \tilde{H}(s) \circ \tilde{H}^{(k-1)}(s)ds \geq 0$. For $\tilde{H}^{(k+1)}(t)$ we then have that

$$
\tilde{H}^{(k+1)}(t) = \int_0^t \tilde{H}(s) \circ \tilde{H}^{(k)}(s) ds \geq 0.
$$

Now, since $v(0) > 0$ we estimate with $|i-k| = 1$, using positivity,

$$\begin{aligned} v_i(t) &= \Big(\sum_{n=0}^{\infty} \tilde{H}^{(n)}(t) \Big) v(0))_i \\ &\geq (\tilde{H}^{(1)}(t) v(0))_i \\ &\geq \Big(\int_0^t \int_0^1 -\partial_{i,k} S_i(\tau x(\tilde{t}) + (1-\tau) y(\tilde{t}) d\tau d\tilde{\tau} \Big) v_k(0). \end{aligned}$$

Since $v(0) > 0$, we can choose a k such that $v_k(0) > 0$. Then this estimate implies that for $i = k-1, k+1$, $v_i(t) > 0$. To show that $v_i(t) > 0$ holds for all i and $t > 0$, we argue as follows. If this inequality holds for some $i = k-1$, then we can argue as before to conclude that the same inequality holds also for $i = (k-1)-1 = k-2$. Similarly, if it holds for $i = k+1$, then also for $i = k+2$. By induction, one obtains $v_i(t) > 0$ for all i and $t > 0$.
What we have shown now is that for the problem $\dot{v}(t) = \tilde{H}(t)v(t)$, the solution satisfies $v_i(t) > 0$ for all i. So $v(t) \gg 0$. But by our earlier remarks, this immediately implies that $u(t) \gg 0$ as well. Hence $y(t) \gg x(t)$, or $\varphi^t(y) \gg \varphi^t(x)$. $\square$

Theorem 3 implies that $\mathcal{B}_{p,q}$ is invariant under the positive gradient flow:

Proposition 12. *Let $p > q \geq 1$. Then $\varphi^t(\mathcal{B}_{p,q}) \subseteq \mathcal{B}_{p,q}$ for all $t > 0$.*

Proof. By definition of Birkhoff sequences we have that for $x \in \mathcal{B}_{p,q}$ either $\tau_{k,l}x \leq x$ or $\tau_{k,l}x \geq x$. Assume w.l.o.g. $\tau_{k,l}x \leq x$. By the above theorem this yields $\varphi^t(\tau_{p,q}x) \ll \varphi^t(x)$ for all $t > 0$, and therefore by equivariance $\tau_{p,q}(\varphi^t(x)) \ll \varphi^t(x)$. Thus $\varphi^t(x) \in \mathcal{B}_{p,q}$ for all $t > 0$. $\square$

Definition 6.1. *Two sequences $x, y \in \mathbb{X}$ are said to be* transverse *if they have no* tangencies*, meaning that if $x_k = y_k$ for some $k \in \mathbb{Z}$, then $x_{k-1} - y_{k-1}$ and $x_{k+1} - y_{k+1}$ have opposite signs. We denote transversality of x and y by $x \pitchfork y$.*

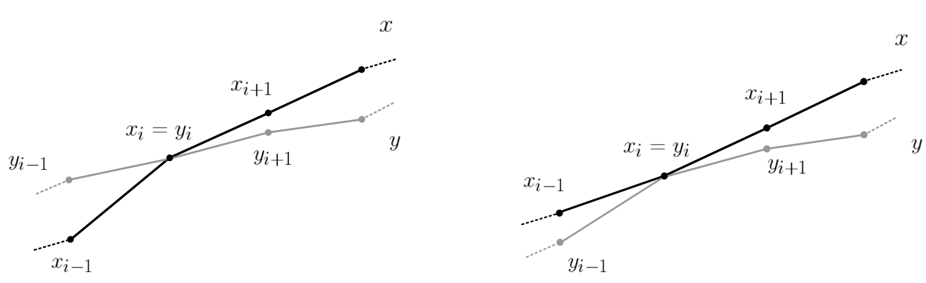

Figure 22: Left: x and y are transverse. Right: x and y are not transverse

When two sequences have no crossings, they are trivially transverse. The number of crossings of transverse sequences x and y is denoted by $I(x,y)$ which is called the *intersection index*. We state the definition for periodic sequences.

Definition 6.2. *Let $x, y \in \mathbb{X}_{p,q}$. If $x \pitchfork y$, then the intersection index of x with y, $I(x,y)$, is defined to be largest integer k for which there are*

$$i_0 < i_1 < ...i_k = i_0 + p,$$

such that

$$(x_{i_j} - y_{i_j})(x_{i_{j+1}} - y_{i_{j+1}}) < 0$$

holds for $j = 0, 1, 2, .., k-1$.

Observe that $k \leq p$ and $I(x,y)$ is an even number when x and y are periodic: if there is a crossing, then by periodicity this crossing is repeated one period later. Hence the sequence crosses back in order to make this crossing again. The following example will be useful for Conjecture 2.

Example 4. Consider the elliptical billiard table shown in Figure 58, section 8.2, and denote by $y = (y_1, y_2, y_3, y_4)$ the $(2,1)$ orbit along the minor axis and by $x = (x_1, x_2, x_3, x_4)$ the V-shaped orbit. Both orbits are of type $(p,q) = (4,2)$ and hence they have an intersection index $I(x,y)$. It is easy to see that $I(x,y) = 2$. Just note there are precisely two intersection points at $i = 1$ and $i = 3$. Since the orbits have $p = 4$, we conclude that the intersection index must indeed be equal to 2. △

We now state an important *Sturmian lemma* which states that the intersection index $I(x,y)$ of two sequences x, y does not increase in time. More formally the lemma reads:

Proposition 13. *Given $x, y \in \mathbb{X}_{p,q}$, the set of $t \in \mathbb{R}$ for which $\varphi^t(x)$ and $\varphi^t(y)$ do not intersect transversally is discrete and $I(\varphi^t(x), \varphi^t(y))$ is a non increasing function of t which has a jump discontinuity exactly at those t for which $\varphi^t(x)$ and $\varphi^t(y)$ do not intersect transversally.*

Proof. See [5]. □

This theorem has the important consequence that any two different orbits (solutions) in $\mathbb{X}_{p,q}$ must be transversal.

Corollary 6.1. *Let $x \neq y \in \mathbb{X}_{p,q}$ be periodic orbits, that is, assume that $\nabla W_{p,q}(x) = \nabla W_{p,q}(y) = 0$. Then x and y are transverse.*

Proof. Define $g(t) = I(\varphi^t(x), \varphi^t(y))$ and note that it is constant since $\varphi^t(x) = x$ for all $t \in \mathbb{R}$ and similarly for y. So g is continuous and therefore the Sturmian Lemma implies that x and y are transverse. □

Another corollary is that if there exists a 1-parameter family of periodic orbits in $\mathbb{X}_{p,q}$, then all its elements have the same intersection index with some fixed orbit also in $\mathbb{X}_{p,q}$.

Corollary 6.2. *Let $\mathcal{F}$ be a 1-parameter family of (p,q) orbits. Let $x_0 \in \mathbb{X}_{p,q}$ be some periodic orbit and assume that $I(x_0, y_0) = k$, for some $y_0 \in \mathcal{F}$ and $k \in \mathbb{Z}_{\geq 0}$. Then $I(x_0, y) = k$ for all $y \in \mathcal{F}$.*

Proof. This trivially follows since the only way the intersection index I can change is when its arguments do not intersect transversally. By assumption however, only periodic orbits are being fed to I as arguments and hence only transverse orbits are considered by Corollary 6.1. $\square$

In Theorem 2, we have shown existence of periodic orbits with any given (rational) rotation number. This was done by finding critical points of (minus) the length functional $W_{p,q}$. Below we shall give an alternative proof of this fact by means of finding fixed points of the gradient flow of the system $\dot{x} = \nabla W_{p,q}(x)$.

Theorem 4. *Let $p > q \geq 1$. Then there is an $x \in \mathcal{B}_{p,q}$ such that $\nabla W_{p,q}(x) = 0$.*

Proof. By Lemma 4.2, $\mathcal{B}_{p,q}$ is closed in $\mathbb{X}_{p,q}$ and in fact compact when we consider sequences modulo $\mathbb{Z}$. The boundary of $\mathcal{B}_{p,q}$ consists of sequences x that touch some translate $\tau_{k,l}x$, but not cross it. This means that all configurations in the boundary are not transverse. The Sturmian Lemma implies that immediately when the flow is turned on, x and its translate become transverse. In particular this means that a boundary configurations instantly become strictly ordered w.r.t. its translates and thus the flow is directed inwards into $\mathcal{B}_{p,q}$.

On solutions of $\dot{x} = -\nabla W_{p,q}(x)$, it holds that

$$\frac{d}{dt} W_{p,q}(x(t)) = -\|\nabla W_{p,q}(x(t))\|^2 < 0.$$

Hence the action strictly decreases along the flow. This implies that the minimum of $W_{p,q}$ cannot lie on the boundary and therefore must be attained in the interior of $\mathcal{B}_{p,q}$. Hence there is $x \in \mathcal{B}_{p,q}$ such that $\nabla W_{p,q}(x) = 0$ and hence we have found a p,q-minimizer.

$\square$

Our search for periodic orbits led us thus far to those orbits which are minimizers for the action functional $W_{p,q}$. Another way of saying this is that we have found index-0 stationary configurations. As remarked before, there also exists other type of configurations. The ones we will consider next are so called saddle points configurations for reasons to become obvious. These periodic orbits have index-1 and their existence is guaranteed by the famous mountain pass lemma. Before we can formulate the precise theorem about existence of index-1 periodic configurations, we will need some basic concepts from Morse theory.

$W_{p,q} : \mathbb{X}_{p,q} \to \mathbb{R}$ is called a *Morse function* if $D^2 W_{p,q}(x)$ is invertible at all critical points x of $W_{p,q}$. Stated differently, the critical points of $W_{p,q}$ are all non-degenerate. In particular this means that for Morse functions the critical points are isolated. When a function is Morse, one defines the *Morse index* $i(x)$ of a non-degenerate critical point x to be the dimension of its unstable manifold of x, considered as a stationary point of $\dot{x} = -\nabla W_{p,q}(x)$. So for example, if $W_{p,q}$ has a (local) minimum at z then the Morse index of z equals 0.

In the billiard system we immediately see two ways for $W_{p,q} : \mathbb{X}_{p,q} \to \mathbb{R}$ to be non-Morse. First, the billiard table itself might cause the degeneracy of the critical points. This is the case for the circular billiard. Just note that any periodic orbit yields an entire 1-parameter family of periodic orbits: just rotate the given periodic orbits to obtain another one. In this case the critical points of $W_{p,q}$ are not isolated. Hence no $W_{p,q}$ can be a Morse function.
In the other case, some choices of p, q might lead to non-Morse functions $W_{p,q}$. Consider an elliptic billiard table $\mathcal{C}$ and any $x_0 \in \mathcal{C}$. Then any orbit with number of reflections $p \geq 3$ comes in a 1-parameter family, see section 8.2, and hence the corresponding action functional $W_{p,q} : \mathbb{X}_{p,q} \to \mathbb{R}$, $p \geq 3$ cannot be Morse. However, it is well known that for $p = 2$ and $q = 1$, the action functional $W_{p,q}$ is in fact Morse.

Although it might seem an inconvenient fact that the our first example of a billiard system is non-Morse, it is in fact quite uncommon. The reason is that Morse functions are dense in the C^2 topology, see for example [1]. Hence with "probability 1" any chosen billiard does satisfy the Morse property. In particular this means that if a function is degenerate, a small perturbation will it in fact non-degenerate. For a proof of this result, we refer to [1].

Definition 6.3. *We call $x \ll y$ in $\mathbb{X}_{p,q}$ consecutive minimizers if there does not exists an index -0 $z \in \mathbb{X}_{p,q}$ such that $x \ll z \ll y$*

The following theorem ensures under some mild assumptions the existence of other types of periodic orbits besides the minimizing orbits.

Lemma 6.3. *(Mountain pass lemma) Assume $W_{p,q}$ is Morse. Let $x \ll y$ of the $W_{p,q}$ be consecutive index-0 configurations. Then there exists an index-1 orbit $z \in \mathbb{X}_{p,q}$ for which $x \ll z \ll y$ holds.*

Proof. The idea is as follows: if there exist at least two isolated index-0 configurations, i.e. local minimums, then to go from one to another, you must go trough a "mountain pass". This mountain pass corresponds to to a critical point of index 1.
To start with the proof, define K to be the closed order interval $K := [x, y]$. Let $\mathcal{C} = \{\gamma : [0,1] \to K : \ \gamma(0) = x, \ \gamma(1) = y, \ \gamma \text{ continuous}\}$. For $\delta \in \mathbb{R}$, define sub-level sets

$$K^\delta := \{x \in K : W_{p,q}(x) \leq \delta\}.$$

The K^δ are forward invariant under the gradient flow because by definition $W_{p,q}$ decreases along solutions.
The claim is that there is a critical point $z \in int(K)$ with the property that $W_{p,q}(z) = c$ where

$$c = \inf_{\gamma \in \mathcal{C}} \max_{t \in [0,1]} W_{p,q}(\gamma(t)).$$

We will prove existence of a critical point z such $x \ll z \ll y$ by means of a contradiction. So suppose there is no such z with $x \ll z \ll y$ and $W_{p,q}(z) = c$. We have $c > \max\{W_{p,q}(x), W_{p,q}(y)\}$ because x and y are index-0 points and

in a small enough neighborhood of x and y $W_{p,q}$ strictly increases. Since $W_{p,q}$ is Morse and hence has only finitely many critical points in K, this implies we can find a 'sub-level strip' around c that does not contain critical points: there are no critical points in $K^{c+\varepsilon} \setminus K^{c-\varepsilon}$ for sufficiently small $\varepsilon > 0$. By shrinking the set a bit, we can obtain by compactness that there is a $\sigma > 0$ such that $||\nabla W_{p,q}||^2 > \sigma$ on $K^{c+\varepsilon} \setminus K^{c-\frac{\varepsilon}{2}}$. This means that if $t \mapsto x(t)$ is a solution curve, as long as $x(t) \in K^{c+\varepsilon} \setminus K^{c-\frac{\varepsilon}{2}}$, it holds that $\frac{d}{dt}W_{p,q}(x(t)) = -||\nabla W_{p,q}||^2 < -\sigma$. This means that any point in $K^{c+\varepsilon}$ will flow in finite time $T > 0$ into $K^{c-\frac{\varepsilon}{2}}$. That is, there is a $T > 0$ such that $\varphi^T(K^{c+\varepsilon}) \subseteq K^{c-\frac{\varepsilon}{2}}$.
However, by definition of c there exists a $\gamma \in \mathcal{C}$ such that $\gamma([0,1]) \subseteq K^{c+\varepsilon}$. By what is shown above, this means that $\varphi^t \circ \gamma \in \mathcal{C}$ lies completely in $K^{c-\frac{\varepsilon}{2}}$. This contradicts the definition of c since we found another element of $\mathcal{C}$ which has a maximum strictly less than c whereas in the definition of c the infimum was taken over all $\gamma \in \mathcal{C}$. Therefore we conclude there exist critical points $x \ll z_1, ..., z_m \ll y$ with $W_{p,q}(z_i) = c$.
The next step is to show that at least one these z_i has index one. Intuitively one argues that if all z_i have index 2, then it is possible to obtain a smaller action. This follows since there are enough directions to travel as to avoid the mountain passes. In a 2-d graph this is trivial, since an index-2 point corresponds to a summit, which can be easily avoided by just walking around the summit and therefore decreasing the action.

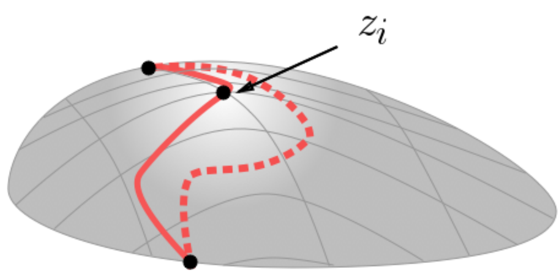

Figure 23: Obtaining a lower action.

More formally, suppose that all of the z_i's have index $k \geq 2$. For all z_i one can find local Morse coordinates near z_i for which, locally, $W_{p,q}$ can be expressed as $W_{p,q}(x_1, ..., x_p) = z_i - x_1^2 - x_2^2 - ... - x_k^2 + x_{k+1}^2 + .. + x_p^2$. Observe that at z_i, $W_{p,q}$ decreases in the $x_1, x_2, ..., x_k$ directions. This follows from the local form of $W_{p,q}$ at z_i: there is a minus sign in front of the x_i^2, for $i = 1, ..., k$. This implies that there are r_i such that if $\gamma \in \mathcal{C}$ intersects $B_{r_i}(z_i)$, it can homotoped into the $x_1, x_2, ..., x_k$ directions to a $\tilde{\gamma}$ which avoids $B_{r_i}(z_i)$ and $W_{p,q}$ has lesser action along $\tilde{\gamma}(t)$, $t \in [0,1]$. This is opposed to the case of index-1 points, in which there is only one route over the saddle point, namely via the saddle point itself.
Now, let γ_n be a sequence of curves in $\mathcal{C}$ for which $c \leq \max_{t\in[0,1]} W_{p,q}(\gamma_n(t)) \leq$

$c + \frac{1}{n}$. We may assume that none of the γ_n intersect the index $k \geq 2$ points z_i. But after letting γ_n flow some suitable chosen time T, we obtain that $\varphi^T(\gamma_n) \subseteq K^{c-\delta}$ for some $\delta > 0$. This is contradictory to the construction of c, since we found another curve, $\varphi^T \circ \gamma_n$ which again lies in $\mathcal{C}$ and has a strictly smaller action on $[0,1]$ Hence there should be at least index 1 point z such that $x \ll z \ll y$. This proves the claim.

□

Definition 6.4. *We will refer to billiard tables for which the action $W_{p,q}$ is Morse as Morse billiard system, or more concisely as Morse billiards*

We now present a result which can be regarded as a converse of the previous theorem, albeit more general. More specifically, the assumption of $W_{p,q}$ being Morse, can be dropped. We do need however an analogue of the Morse index for non-Morse functions. Write $\mathrm{Crit}(W_{p,q})$ for the set of critical points of $W_{p,q}$.

Definition 6.5. *If $W_{p,q}$ is not Morse, and $x \in Crit\ (W_{p,q})$, then we define the generalized Morse Index, $i_g(x)$ of x to be the number of positive eigenvalues of $-\nabla^2 W_{p,q}$.*

Instead of only considering index-1 critical point x as we assume in Theorem 5, any index $i(x) \geq 1$ will work.

Theorem 5. *Let $x \in \mathbb{X}_{p,q}$ with generalized Morse index $i(x) > 0$. Then*

$$a = \sup\{c < x : \nabla W(c) = 0\}, \qquad b = \inf\{d > x : \nabla W(d) = 0\},$$

are critical points of $W_{p,q}$ of index 0.

Proof. We start with noting that the diagonal elements of L are bounded. This follows because S is C^2 and defined on a compact set. Hence there is an $M > 0$ such that $H := -L + M\mathrm{Id}$ has a strictly positive diagonal and non-negative subdiagonals. By the Perron-Frobenius theorem we have that there exists a unique positive eigenvalue $\lambda_0 + M$ of H and the corresponding eigenvector e_0 can be chosen positive: $(e_0)_j > 0$, $j = 1, ..., p$. Since the generalized Morse index of x is greater than 0, λ_0 is a strictly positive eigenvalue of $-L$ as well. Hence, by the strong unstable manifold theorem and positivity of the flow, there exist unique orbits $\alpha_{+,-}(x;t)$ of the flow φ^t for which

$$\alpha_-(x;t) < x < \alpha_+(x;t), \ \ \alpha_{+,-}(x;t) \to x \text{ as } t \to -\infty \tag{14}$$

and $\alpha'(x;t) < 0 < \alpha'(x;t)$. Note that the orbit $\alpha_-(x;.)$ is decreasing and $\alpha_+(x;.)$ is increasing.
Now, since $W_{p,q}$ is $\tau_{k,l}$ invariant, it follows that $\tau_{0,1}x = x+1$ and $\tau_{0,-1}x = x-1$ are critical points as well. Because and $x - 1 < x < x + 1$, we have for large negative t that

$$x - 1 < \alpha_-(x;t) < x < \alpha_+(x;t) < x + 1.$$

Since the flow φ^t is order preserving these inequalities hold in fact for all t: for all s it holds that $x = \varphi^s(x) < \varphi^s(\alpha_+(x;t)) < \varphi^s(x+1) = x+1$ and similarly for α_-.
The limits

$$\lim_{t\to\infty} \alpha_{+,-}(x;t) =: \omega_{+,-}(x)$$

therefore exists by boundedness and the fact $\alpha_{+,-}$ are both monotone. Also they are critical points of $W_{p,q}$.) Define

$$c = \inf_{-\infty<t<\infty} \{\alpha_-(x;t)\}$$

$$d = \sup_{-\infty<t<\infty} \{\alpha_+(x;t)\}.$$

Then by definition $\alpha_-(x;t) \downarrow c$ and $\alpha_+(x;t) \uparrow d$ as $t \to \infty$. Then c and d are critical points of $W_{p,q}$. This follows by continuity of $W_{p,q}$.
By definition of a and b we have that $c \leq a$ and $b \leq d$. We will show that $a = c$ and $b = d$. Assume that $e < x$ is a critical point of W. Then by Theorem 3 $e = \varphi^t(e) \ll \varphi^t(x) = x$ for all $t > 0$. So in fact $e \ll x$. This implies that for large negative t we have $e < \alpha_-(x;t) < x$. Again, by monotonicity of the flow, this hold for all $t \in \mathbb{R}$. So we see that for any critical point it holds that $e < x$ $e \leq c$. In particular this holds for a since it is supremum of all critical points $f < x$. So $a \leq c$. This implies that $a = c$. The proof of $b = d$ is obtained similarly.
We now show that the generalized Morse index of a and b is zero. Suppose that $i(a) > 0$. Then there is a positive eigenvalue of $-L$. This implies there is an orbit $z(t)$ of the gradient flow for which $z(t) \downarrow a$ as $t \to -\infty$ and $z'(t) > 0$. Let $z^* = \sup_{t\in\mathbb{R}} z(t)$. Then z^* is a critical point of $W_{p,q}$ since it is an ω-limit point. Clearly it holds that $a < z^*$. For large negative t, $a < z_+(t) < x$ so that by monotonicity it holds for all $t \in \mathbb{R}$. This implies that also $a < \sup z(t) = z^* \leq x$. For large negative t, $z(t)$ is close to a while for large negative t, $\alpha_-(x;t)$ is close to x. Since $a < x$, we also have then that $z(t) < \alpha_-(x;t)$ for large negative t. In the limit for t to plus infinity, $z(t)$ goes to z^* while $\alpha_-(x;t)$ goes to a. This implies that $z^* \leq a$. This is a contradiction since we earlier derived that $a < z^*$. Hence $i(a) = 0$. The same argument works for proving $i(b) = 0$. $\square$

Lemma 6.3 immediately implies the existence of (p,q) periodic orbits other than minimizers in Morse billiards (which always exist by Theorem 2). This follows by invariance of $W_{p,q}$ by $\tau_{k,l}$. Or more intuitively: if there exists one orbit $x \in \mathbb{X}_{p,q}$, then any of its translates $\tau_{k,l}x$ is again a periodic orbit and they can be chosen consecutive. Of course this corresponds only to relabeling the points of reflection in the orbit down in the billiard system, but in the space $\mathbb{X}_{p,q}$ they are different points. Hence the lemma applies to this situation. In particular this means that given any smooth enough, convex Morse billiard table, there are always at least two different periodic orbits present. In fact, there are always at least two periodic orbits in each space $\mathbb{X}_{p,q}$ with $p > q \geq 1$. For certain types of periodic orbits the minimizer obtained via Aubry-Mather theory, together with the index -1 orbit the Mountain pass lemma yields, are the only periodic orbits. This holds in particular for the space $\mathbb{X}_{2,1}$.

Remark 5. *If the condition that $W_{p,q}$ is Morse is dropped, then there might be no critical points of index-1. This is for example the case for circular billiards. It is well known (and easy to see) that any (p,q) orbit in a circle comes in a full 1-parameter family (by symmetry of the circle). Since no other (p,q) orbits can exists, because all orbits are Birkhoff (by example 2), and the fact minimizers in fact always exists, we conclude that indeed no index-1 orbits are to be found.*

The index -1 configurations who's existence is guaranteed by the mountain pass lemma satisfy an important property: just like the index-0 configurations in the billiard system, they are Birkhoff. In order to prove this, we introduce the notion of a so called *maximal* $index - 0$ *skeleton*. The existence of this structure will immediately imply the fact that the index one configurations are indeed Birkhoff.

Definition 6.6. *A nonempty and strictly ordered collection of configurations*

$$C_0 = \{... \ll x_1 \ll x_0 \ll x_1 \ll ...\} \subseteq \mathbb{X}_{p,q}$$

is called a maximal $index - 0$ *skeleton for the Morse function $W_{p,q}$ if:*

1. *it consists of index-0 critical points of $W_{p,q}$*
2. *it is shift invariant: for all $x \in C_0$ and $k, l \in \mathbb{Z}$ it holds that $\tau_{k,l}x \in C_0$.*
3. *it is maximal: if $y \notin C_0$ is an index-0 point, then there is no $i \in \mathbb{Z}$ with $x_i \ll y \ll x_{i+1}$.*

We will show existence by constructing such a skeleton.
Define $C_0^0 = \{..., x_1, x_0, x_1, ...\}$ to be the collection of global minimizers of $W_{p,q}$. By Aubry's Lemma (Lemma 5.2) these are strictly ordered and if x is minimizer, then $\tau_{k,l}x$ is a minimizer for all $k, l \in \mathbb{Z}$. Hence the collection C_0^0 is indeed nonempty, strictly ordered and shift invariant. Also note that since $W_{p,q}$ is Morse, the collection is discrete.
The maximality is obtained as follows. Suppose now that there exists an $x \notin C_0^0$ of index-0 and for which it holds that $x_i \ll x \ll x_{i+1}$. Adjoin this element x together with all its translates $\tau_{k,l}x$ with $k, l \in \mathbb{Z}$ to the set C_0^0. Denote this set by C_0^1. The extended set C_0^1 clearly satisfies the properties of strictly orderedness and shift invariance. This process of adjoining index-0 points and their translates is repeated. Note that since $[x_0, x_0 + 1]$ is a compact order interval, the Morse property of $W_{p,q}$ guarantees there are only finitely many critical points in $[x_0, x_0 + 1]$. Hence this process will terminate in finitely many steps because any index-0 point in $\mathbb{X}_{p,q}$ has a unique translated copy inside $[x_0, x_0+1]$.

Now by the mountain pass lemma, there exists for each i a index-1 point z_i

between x_i and x_{i+1} in our collection C_0. To show that the index-1 configurations are Birkhoff, let z_k be such a configuration. Then we can fit it into the strictly ordered collection $C_0 = \{..., x_1, x_0, x_1, ...\}$ to obtain

$$C_1 := \{..., x_1 \ll z_1 \ll x_0 \ll ... \ll x_k \ll z_k \ll x_{k+1} \ll ...\}.$$

By maximimality of C_0, this implies that $\tau_{m,n} z_k \ll z_k$ or $\tau_{m,n} z_k \ll z_k \gg z_k$ or $\tau_{m,n} z_k = z_k$. Hence z_k is indeed Birkhoff.

We recapitulate: we proved existence of index-0 configurations, i.e. billiard orbits of any type (p, q) which correspond to minimizers of $W_{p,q}$. These periodic orbits are Birkhoff (cf. Corollary 5.1). The mountain pass lemma guarantees the existence of other types of periodic orbits, namely those of index one, which again are Birkhoff. The next section is devoted to finding periodic orbits that are not Birkhoff and more orbits which are different from the minimizing and saddle point orbits.

7 Conley Index Theory

In this section we will introduce some notions from Conley index theory which we will use in section 10 to prove existence of non-Birkhoff orbits. Our discussion follows [7] to a large extent. Conley index theory is a topological tool which is used to classify certain invariant sets of some flow on a topological space X. In some cases one is able to prove existence of fixed points in certain types of neighborhoods, given that the Conley index of that neighborhood has a particular form. There are several approaches to this theory, and we will outline what is known as the "classical" theory. As remarked, the Conley index is an index of invariant sets. More precisely, if S is an invariant set of a dynamical system, we can assign a topological invariant $h(S)$ to it which characterizes some important features of this set w.r.t. the dynamics. The invariant sets that are considered are isolated, guaranteeing some sort of stability of the index. The powerful feature of the Conley index is that it remains constant under perturbation and even under certain homotopies. The intuitive idea is that some invariant is assigned to an invariant set by applying certain topological constructions: after a neighborhood around the invariant set is found, one identifies at which pieces of this set the flow "moves out". Then one identifies this latter set to point and the homotopy type of the resulting space is the invariant one is interested in. Before we can give the precise definitions, we will recall some basic topological and dynamical systems notions. The following discussion is based on [7].

A *pointed topological* space is a pair (X, x_0) where X is a topological space and $x_0 \in X$ some designated point. We define a *pair* (N, L) to be a collection of subsets such that $L \subseteq N \subseteq X$. Given such a pair, we define

$$(N/L, [L]) := (N \setminus L) \cup [L],$$

where $[L]$ is the equivalence class of the relation $x \sim y \iff x, y \in L$ and "$\setminus$" is the set theoretical complement. A set $U \subseteq N/L$ is open if either U is open in N and $U \cap L = \emptyset$ or the set $(U \cap (N \setminus L)) \cup L$ is open in N. If $L = \emptyset$, then

$$(N/L, [L]) := (N \cup \{*\}, \{*\}),$$

where $*$ denotes the equivalence class consisting of the empty set.

Assume now in addition that X is a locally compact metric space. Let $\varphi : \mathbb{R} \times X \to X$ denote a flow on X and write $\varphi^t(x) := \varphi(t, x)$ for $t \in \mathbb{R}, x \in X$. Note that the spaces of our interest, namely $\mathbb{X}_{p,q}$ with $p > q \geq 1$, trivially satisfy the locally compactness condition, because they are homeomorphic to $\mathbb{R}^p$.

Definition 7.1. *A compact set $N \subseteq X$ is called an* isolating neighborhood *if* $\mathrm{Inv}(N, \varphi) = \{x \in N : \varphi(\mathbb{R}, x) \subseteq N\} \subseteq int(N)$.

Example 5. Let x_0 be attracting fixed point of some dynamical system. Then the set N in the following picture is an isolated invariant set.

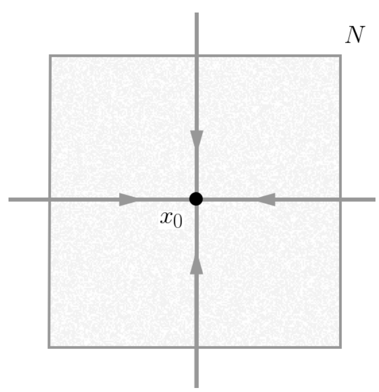

Figure 24: An attracting fixed point x_0 and isolating neighborhood N.

To see this is the case, observe that if x is such that $\varphi(\mathbb{R}, x)$ entirely lies in N, then any point x will exit N in forward or backward time except for $x = x_0$. Clearly $x_0 \in \mathrm{int}(N, \varphi)$. Hence indeed N is an isolating neighborhood of x_0.

$\triangle$

A set S is called an *invariant set* for the flow φ if $\varphi(\mathbb{R}, S) := \bigcup_{t \in \mathbb{R}} \varphi^t(S) = S$. An invariant set S is called an *isolated invariant set* if there exists an N such that $S = \mathrm{Inv}(N, \varphi)$.

Definition 7.2. *(Exit set for a flow)* $L \subseteq N$ is called an exit set for N if, given $x \in N$ and $t_0 > 0$ such that $\varphi^{t_0}(x) \notin N$, then there exists a $t \in [0, t_0)$ such that $\varphi^t(x) \in L$.

This latter definition just means a point $x \in N$ leaves N under the flow φ^t via L. We shall use the shorthand notation $\text{Inv}(N) := \text{Inv}(N, \varphi)$ if φ^t is understood from the context.

We now come to a notion central to Conley index theory, namely the notion of an *index pair* for an isolated invariant set S.

Definition 7.3. *(Index pair) Let S be an isolated invariant set. Then (N, L) for which $L \subseteq N \subseteq X$, with L, N both compact, is called an index pair for S if the following three conditions are being satisfied:*

1. *$N \setminus L$ is a neighborhood of S and $S = cl(Inv(N \setminus L))$.*
2. *L is positively invariant in N, i.e. if $x \in L$ is such that $\varphi([0,t], x) \subseteq N$, then $\varphi([0,t], x) \subseteq L$.*
3. *L is an exit for N.*

Now we have the tools to give the definition of the *homotopy Conley index*.

Definition 7.4. *(Homotopy Conley index) The homotopy Conley index $h(S)$ of an isolated invariant set S is the homotopy type of the pointed topological space obtained by collapsing L to a point:*

$$h(S) \sim (N/L, [L]).$$

Remark 7.1. *This definition is well defined, i.e. if (N, L) and (N', L') are two different index pairs for a invariant isolated set S, then*

$$(N/L, [L]) \sim (N'/L', [L']).$$

For a proof, we refer to [7]

We now present some examples where we find the Conley index of certain isolated invariant sets.

Example 6. We will calculate the Conley index for two systems and state the Conley for the third third example.

1. Consider the system in example 5. There an index pair is given by N and $L = \emptyset$. Hence $h(\{x_0\}) \sim (S^0, *)$
2. Let x_0 be fixed point of index-1. Let $S = \{x_0\}$. Then the set N in the following picture is an isolated invariant set. Let L denote the shaded bars. Then (N, L) is an index pair.

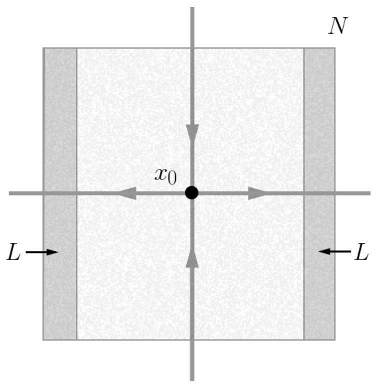

Figure 25: An index pair (N, L) for an index-1 point x_0.

To compute $h(S)$, deform L to a point and paste it together.

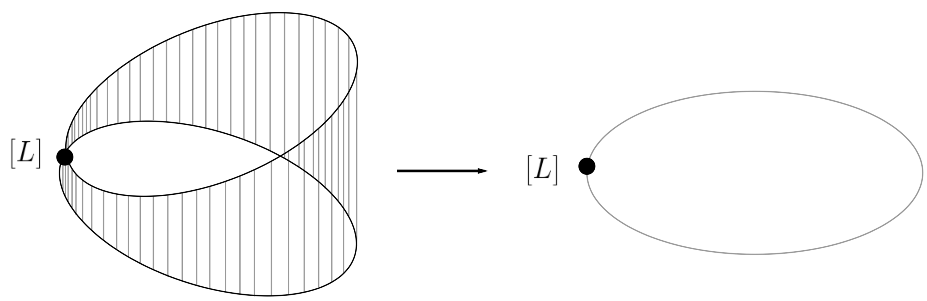

Figure 26: Computing the homotopy type $h(\{x_0\})$.

Now contract it via the shaded are to obtain a space homotopic to $(S^1, *)$. Hence $h(S) \sim (S^1, *)$.

3. More generally, the homotopy Conley index of a fixed point x_0 of index k is given by $h(\{x_0\}) \sim (S^{2k-1}, *)$, see [11].

$\triangle$

We will now list some basic properties of the Conley index.

- If N and N' are isolating neighborhoods such that $\mathrm{Inv}(N) = \mathrm{Inv}(N')$, then
$$h(N) \sim h(N').$$
- If $h(N)$ is non-trivial, then $\mathrm{Inv}(N) \neq \emptyset$

One ideally wants to gain information about the existence of fixed points if $h(N)$ is non-trivial. This is however not possible in general. Fortunately for gradient flows it turns out to be true: we may conclude from $\mathrm{Inv}(N) \neq \emptyset$ that $\mathrm{Inv}(N)$ contains a fixed point. We will formulate this in the following simple lemma.

Lemma 7.1. *Let φ denote the flow of $\dot{x} = \nabla f(x)$. Then it holds that*

$$\mathrm{Inv}(N) \neq \emptyset \implies \mathrm{Crit}(f|_N) \neq \emptyset.$$

Proof. If $\mathrm{Inv}(N) \neq \emptyset$, then $\mathrm{Inv}(N)$ must contain the closure of the orbit of some point in $y \in N$. Since $\omega_{+,-}(x)$ are critical points of f and N is compact, the claim follows. □

The main point for us of using the Conley Index, and in general one of the most important properties of the Conley index, is that it is "invariant under continuation". Intuitively, this means that the Conley index remains invariant under certain homotopies of the flow. This has the convenient implication that we can compute the Conley index for some relatively simple system, and draw conclusions about generic systems if we can show the two systems are related by continuation. We will make this formal by giving the precise definitions and state the invariance result. For $\lambda \in \Lambda \subseteq \mathbb{R}$, let $\varphi^\lambda : \mathbb{R} \times X \to X$ be a continuously parametrized family of flows, where X is compact, locally contractible, connected metric space.

Definition 7.5. *(parametrized flow) The parametrized flow corresponding to the family φ^λ is the continuous flow $\Phi : \mathbb{R} \times X \times \Lambda \to X \times \Lambda$, given by*

$$\Phi(t, x, \lambda) = (\varphi^\lambda(t, x), \lambda).$$

A useful property of the Conley index is that it remains invariant under perturbation. It is this property that we will use in section 9 to get a local existence result about non-Birkhoff orbits in certain billiard systems. We first state a proposition that guarantees that if N is an isolating neighborhood for some flow φ_{λ_0}, then N remains an isolating neighborhood for all nearby flows.

Proposition 14. *There exists $\varepsilon > 0$ such that if $d(\lambda, \lambda_0) < \varepsilon$, then N is an isolating neighborhood for φ_λ.*

The perturbation result about the Conley index now reads as follows.

Theorem 6. *Let N be an isolating neighborhood for φ_{λ_0}. Choose an $\varepsilon > 0$ such that if $d(\lambda, \lambda_0) < \varepsilon$, then N is an isolating neighborhood for φ^λ. Then*

$$h(N, \varphi_\lambda) \sim h(N, \varphi_{\lambda_0}).$$

We are now in a position the give the precise definition of *related by continuation*. Let $N^\lambda = N \cap (X \times \{\lambda\})$, where $N \subseteq X \times \Lambda$.

Definition 7.6. *Let $\lambda_i \in \Lambda$ for $i = 0, 1$, and let S_i be an isolated invariant set for φ^{λ_i}. The S_0 and S_1 are related by continuation if there exists an isolating neighborhood $N \subseteq X \times \Lambda$ of the parametrized flow Φ such that Inv$(N^{\lambda_0}, \varphi^{\lambda_0}) = S_0$ and Inv$(N^{\lambda_1}, \varphi^{\lambda_1}) = S_1$.*

The fundamental theorem is as follows.

Theorem 7. *If S_0 and S_1 are related by continuation, then*

$$h(S_0) \sim h(S_1).$$

8 Non-Birkhoff orbits

In the previous sections we made the following advances towards our goal of finding periodic orbits in the billiard system. First of all, we proved existence of (p, q)-periodic orbits which are minimizers of the action functional $W_{p,q}$. These were found to have the Birkhoff property, as a consequence of Aubry's lemma. In the section about the gradient flow, it was proven that between consecutive minimizing orbits, there is always another type of orbit when $W_{p,q}$ is Morse: saddle-type orbits, or index-1 configurations. These orbits were shown to be ordered w.r.t. the set of minimizing orbits and furthermore, they also have the Birkhoff property. In this section we will find criteria for the existence of non-Birkhoff periodic orbits. It turns out that if the billiard curve has certain symmetry properties, the existence of NB orbits is guaranteed. An example of billiard systems for which non-Birkhoff orbits exists are "sufficiently elongated" ellipses. We will make this statement precise and prove in this section that NB orbits in ellipses are necessarily symmetric . Also other interesting facts about elliptical billiards will be discussed. Another goal of this section is to prove that $(4, 2)$ non-Birkhoff orbits in ellipses persist under small perturbations, in the sense that a perturbed ellipse still possess a $(4, 2)$ NB orbit. The proof exploits properties of the gradient flow as discussed in section 6 and the Conley index theory we developed in section 7. A number of results in sections 8.2 and 8.3 are applications of general results about monotone twist maps found in S. Angenents paper, [5]. These results yield interesting proofs of certain billiard theoretical properties.
We will first discuss NB orbits in general symmetric domains. After that, we will specialize the theorems found in section 8.1 to ellipses and explore further properties in ellipses.

Let $x \in \mathbb{X}_{p,q}$ be a critical point and define

$$\alpha_k(x) := -\partial_{22}S(x_{k-1}, x_k) - \partial_{11}S(x_k, x_{k+1}), \quad \beta_k(x) := -\partial_{12}S(x_{k-1}, x_k).$$

The Hessian matrix of $W_{p,q}$ at $x \in \mathbb{X}_{p,q}$ is given by the tridiagonal matrix with corner elements

$$-\nabla^2 W_{p,q}(x) = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \cdots & \beta_p \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & 0 \\ 0 & \beta_2 & \alpha_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \beta_{p-1} \\ \beta_p & 0 & \cdots & \beta_{p-1} & \alpha_p \end{bmatrix}, \tag{15}$$

where

$$\alpha_i(x) = \sin^2(\theta_i)\left(\frac{1}{S_{i-1,i}} + \frac{1}{S_{i,i+1}}\right) - 2\kappa(x_i)\sin(\theta_i),$$

and

$$\beta_i(x) = -\frac{\sin(\theta_i)\sin(\theta_{i+1})}{S_{i,i+1}},$$

where $S_{i,i+1} := S(x_i, x_{i+1})$, $\kappa(x_i)$ is the curvature at x_i and θ_i is the angle at x_i. The β_i's were already found in Proposition 2. The formula of $\alpha_i(x)$, we found in [10].

Lemma 8.1. *The spectrum of* $-\nabla^2 W_{p,q}(x)$ *is given by*

$$\sigma(-\nabla^2 W_{p,q}(x)) = \{\lambda_0 > \lambda_1 \geq \lambda_2 > ..., \lambda_{p-1}\}.$$

For a proof of Lemma 8.1, we refer to [5].

Another important lemma relates the eigenvalues of the Hessian of twist maps to the number of *signchanges* of corresponding eigenvectors.

Lemma 8.2. *Let* ξ^i *be an eigenvector corresponding to* $\lambda_i \in \sigma(-\nabla^2 W_{p,q}(x))$. *Then* ξ^i *has* $2[\frac{i+1}{2}]$ *signchanges.*

Again we refer to [5] for a proof.

8.1 Billiards with symmetry

It is a well known fact that in any strictly convex billiard there exist Birkhoff periodic orbits of any type (p, q). Non-Birkhoff orbits are less abundant. For example, we have seen in example 2 that the circular billiard has no non-Birkhoff orbits at all, while Theorem 1 has as a consequence that there are billiards which have precisely two orbits of any given type (p, q), where p and q are relatively prime. These two orbits are necessarily Birkhoff by the Birkhoff Theorem, see section 1, and hence for relatively prime pairs p and q, there are no non-Birkhoff orbits. In fact, there is a theorem by Mather [9] that gives a relationship between the non-existence of non-Birkhoff orbits of type (p, q) and the existence of invariant circles with irrational rotation number ω such that $\frac{q}{p}$ is a "sufficiently close" rational of ω. More precisely it reads:

Theorem 8. *An area preserving monotone twist map* f *of an annulus* $\mathbb{A}$ *has no invariant circle of irrational rotation number* ω *if and only if* f *has a non-Birkhoff periodic orbit with rotation number equal to a convergent of* ω.

We shall not consider (p, q) NB orbits with p, q relatively prime, but we will investigate classes of (p, q) orbits where $q|p$.

When the domain has certain symmetry properties, one can prove existence of NB orbits of a special class of (p, q)-orbits. In case that the billiard domain has reflection symmetry, we will prove existence of NB-orbits of type $(2p, p)$. When a billiard domain has rotational symmetry, a similar class of orbits is shown to exists. This is the content of Theorems 9 and 10. I did not find the results in this subsection in the literature.

An important property that is used in the proof is that $\mathcal{B}_{np,nq} = \mathcal{B}_{p,q}$ for all integers $n > 0$. If we can show that there exists an orbit in $\mathcal{B}_{np,nq}$ that is not at the same time in $\mathcal{B}_{p,q}$, it must be a NB orbit. We shall exploit this fact for $p = 2$ and $q = 1$.
We consider billiard domains with two kinds of symmetry. The first one is reflection symmetry and the second one is rotational symmetry. In the latter case, our interest will be about billiard tables with only rotational symmetry about π radians.

We will now give a precise description of the symmetric billiard tables.

Definition 8.1. *A billiard table $\mathcal{C}$ that is reflection symmetric w.r.t. some axis Σ is called* Σ-symmetric*. Configurations that are symmetric w.r.t. Σ are also called Σ-symmetric. A rotational symmetric billiard table that is symmetric w.r.t. a rotation over π radians is called* R_π-symmetric*. Configurations that are symmetric w.r.t. a rotation over π are also called R_π-symmetric.*

Denote by Σ the symmetry axis in which $\mathcal{C}$ is reflection symmetric and $\mathcal{O}_\Sigma$ the corresponding orbit in some $\mathbb{X}_{2p,p}$ along Σ. Recall that $\mathcal{C} \cong \mathbb{R}/\mathbb{Z}$ has length 1. Let the action $S : \mathcal{C} \to \mathcal{C}$ denote the reflection in Σ, given by $\mathcal{S}(x_k) = -x_k \bmod 1$, where the origin is chosen at one of the two points where Σ and $\mathcal{C}$ meet. For an orbit of period p, we can naturally extend the action to tuples:

$$\mathcal{S}(x_1, ..., x_p) := (\mathcal{S}(x_1), ..., \mathcal{S}(x_p)) = (-x_1, ..., -x_p) \bmod 1.$$

In Figure 27 the setting is visualized.

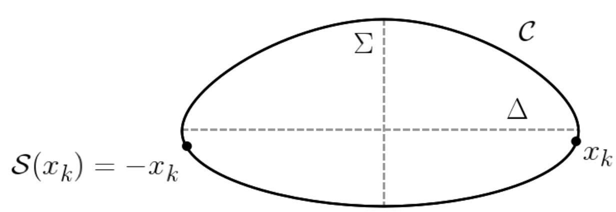

Figure 27: Illustration of the map $\mathcal{S}$ and the $(2, 1)$ orbits in a Σ-symmetric domain.

Let σ denote the shift map $\tau_{1,0}$ and denote by $|\mathcal{O}_\Sigma|$ be the length of the orbit

$\mathcal{O}_\Sigma$. Finally, recall that the curvature at some point t of a plane curve $\gamma(t) = (x(t), y(t))$ is defined as

$$\kappa(t) = \frac{|x'y'' - x''y'|}{(x'^2 + y'^2)^{\frac{3}{2}}},$$

where $'$ stands for differentiating w.r.t. t.

The basic idea of Theorem 9 and Conjecture 1 is that as soon there exists a configuration inside a billiard domain $\mathcal{C}$ with sufficiently small curvature at the intersection points of the shorter symmetry axis and $\mathcal{C}$ itself, then the existence of a periodic orbit of the same topological type is guaranteed.

An important ingredient in both proofs is the existence of an invariant subset L and M of the gradient flow of $W_{p,q}$ which contains Σ-symmetric and R_π-symmetric configurations in $\mathbb{X}_{2n,n}$. L is defined as

$$L = \{x \in \mathbb{X}_{2n,n} : \mathcal{S}(x) = \sigma^n(x)\},$$

and M as

$$M = \{x \in \mathbb{X}_{2n,n} : \sigma^n(x) = R_{\frac{n}{2}}(x)\}.$$

To obtain some intuition we have drawn a configuration in L in Figure 28 where the interpretation of the condition $\mathcal{S}(x) = \sigma^n(x)$ should become clear.

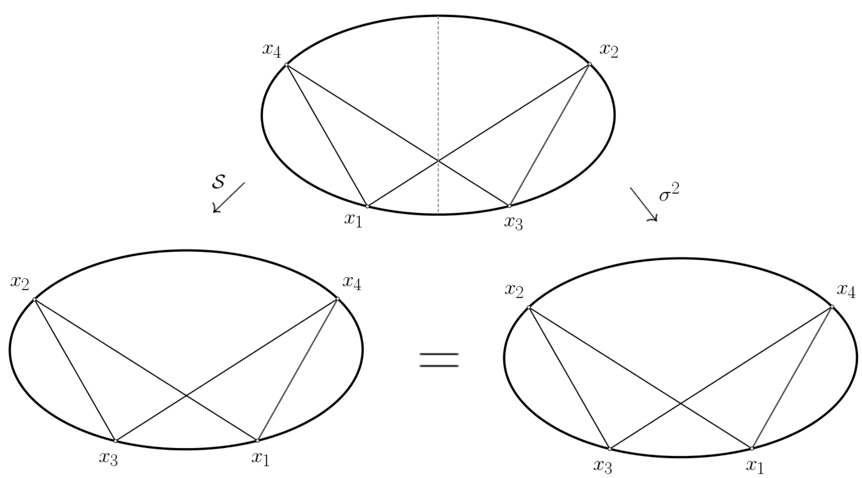

Figure 28: Example of an $x \in L$.

We will now argue that the sets L and M are non-empty, that is to say, there exists $2n$ periodic L-symmetric and M-symmetric configurations that have rotation number n. Consider first L-symmetrical configurations where n is even.

We obtain configurations in L in the following way. Draw $n = 2k$ points ordered like two rows of k points. Connect the points as is shown in Figure 29 and label them as a billiard orbit. Call this configuration x. Since working algebraically is getting involved, we will keep in mind the specific example where $n = 4$, see Figure 29. It is clear that one obtains in general a similar configuration as in

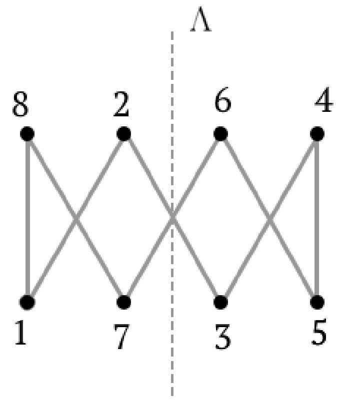

Figure 29: Illustration of the construction for $n = 4$.

the case of $n = 4$. Observe that with this labeling, it holds that $\mathcal{S}(x) = \sigma^n(x)$, where $\mathcal{S}$ is the reflection in the line Λ. Next we need to show that $x \in \mathbb{X}_{2n,n}$, that is, we need to show that the rotation number of x equals n. Considering the fact that mirroring x in the symmetry axis Λ does not change the orientation of x (since $\mathcal{S}(x) = \sigma^n(x)$), one concludes after some thought that the rotation number ρ is equal for both orientation of x. Since $\rho \in \{1, ..., 2n-1\}$, this implies the rotation number must be $\rho = \frac{2n}{2} = n$ so that indeed $x \in \mathbb{X}_{2n,n}$.

The next case is n is odd. Keep in mind the case of $n = 3$ or $n = 5$ in Figure 30. Similarly as in the case n is even, we consider again two rows of $n = 2k + 1$ points which lie symmetry w.r.t. some line Λ. Pick some row and label the points with odd integers k starting from 1. Label the other row with even integers l such that opposite points it holds that $l = k+n \bmod 2n$. Connect all consecutive points and call the resulting configuration x.

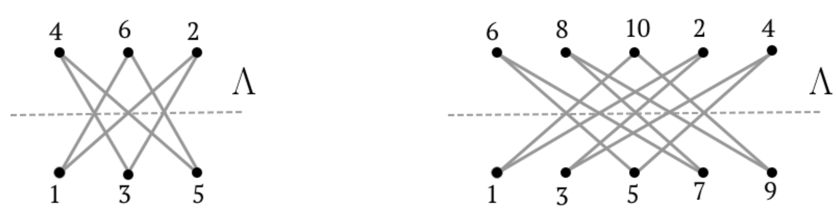

Figure 30: Configuration in L for $n = 3$ and $n = 5$ resp.

Then x satisfies the relation $\mathcal{S}(x) = \sigma^n(x)$. Since the orientation of x does not change under $\mathcal{S}$, one can show that in fact x has rotation number n. That is, $x \in \mathbb{X}_{2n,n}$. Hence for both n even and odd, the set L is non-empty.

We will now discuss the rotational symmetry case. We will first show that if $R_{\frac{n}{2}}(x) = \sigma^n(x)$, then $x \in \mathbb{X}_{2n,n}$. So let x satisfy the symmetry condition. Then for each i, $x_{i+n} = x + \frac{n}{2}$. Hence $x_{i+2n} = x_i + \frac{n}{2} + \frac{n}{2} = x_i + n$. Hence $x \in \mathbb{X}_{2n,n}$. The next thing we need to show is that there exists configurations satisfying the symmetry condition prescribed by M. We only mention that the underlying idea is the same as for the Σ-symmetric case since the symmetry operation R_π also leaves the orientation of a configuration in M invariant. In Figure 31, a $(6,3)$ configuration in M is shown.

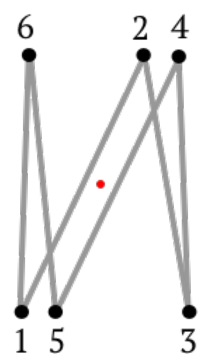

Figure 31: A $(6,3)$ configuration in M. The rotation symmetry over π is w.r.t. the red dot.

Definition 8.2. *The non-Birkhoff constant (NBC) Θ is defined as*

$$\Theta := \sup_{b \in \mathbb{R}} \{b : \exists x \neq 0 \text{ such that } \sqrt{4b^2(1-x^2)} + \sqrt{4b^2(1-x^2)+x^2} > 4b\}.$$

Numerically this yields $\Theta \approx 0.385$. The interpretation of Θ will become clear in the proof of Lemma 8.3. In Lemma 8.3, we characterize elliptical billiard domain for which configurations XYZ exists with the property that the length $|XYZ|$ is greater than $2 \cdot AB$, see Figure 32. This lemma will play a crucial role in Theorem 9 and Conjecture 1 about the existence of NB orbits in certain symmetric billiard tables. Whether there exist such configurations as shown in Figure 32, depends on the geometry of the symmetric table. More precisely, the curvature of the table at the intersection point of the shorter symmetry axis with the billiard table determines if there exists such a state.

For Lemma 8.3, we need Figure 32.

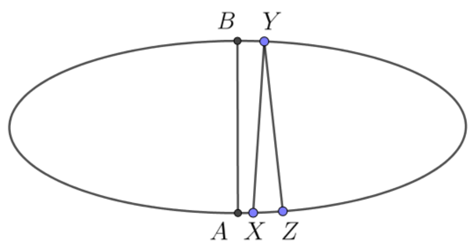

Figure 32: For X, Y and Z close to AB, $|XY| + |YZ| > 2 \cdot AB$ for sufficiently elongated ellipses.

Lemma 8.3. *Consider the elliptical billiard table $\mathcal{E}$ given by $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ with minor axis AB and "wedge" shaped figure XYZ, as in Figure 32. $\mathcal{E}$ is Σ-symmetric w.r.t. $\Sigma := AB$. If $\kappa(A) = \kappa(B) < \frac{2\Theta^2}{|\Sigma|}$, then one can obtain $XY + YZ > 2 \cdot AB$ by moving the wedge sufficiently close to AB. Furthermore, when the latter inequality holds, there exist configurations $z \in L \subseteq \mathbb{X}_{2n,n}$ for all n, such that $|z| > |\mathcal{O}_\Sigma|$. Also, there exist configurations in M for which the claim holds.*

Proof. We prove this by considering the configuration in Figure 33, where CD is parallel with AB. If we can show that $AC + CD > 2 \cdot AB$, then by continuity the same holds for the configurations in Figure 32.

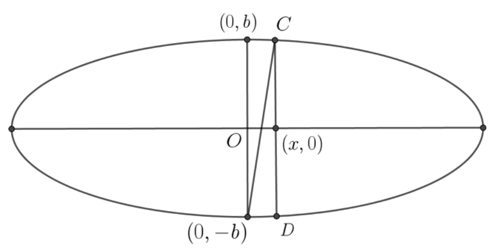

Figure 33: Calculating the length of ABC.

We may assume that $a = 1$. $\mathcal{E}$ is situated in the plane as is shown in Figure 33. Then $A = (0, b)$ and $B = (0, -b)$. Let x be the first coordinate of point C. The length of $AC + CD$ is given by:

$$AC + CD = \sqrt{4b^2(1 - x^2)} + \sqrt{4b^2(1 - x^2) + x^2}.$$

We want to find the curvature at A and B such that it holds that $AC + CD > 2 \cdot AB$, that is, when there exists a non-zero solution of

$$\sqrt{4b^2(1 - x^2)} + \sqrt{4b^2(1 - x^2) + x^2} > 4b.$$

Numerically this yields that if $0 < b < 0.385... =: \Theta$, then the above inequality is being satisfied for small $x > 0$. Hence if $\frac{a}{b} = \frac{1}{b} > \frac{1}{\Theta}$, then such a configuration is possible. We now calculate what this means for the curvature. The curvature for an ellipse at some point t is given by

$$\kappa(t) = \frac{ab}{(a^2 \sin^2(2\pi(t + \frac{1}{4})) + b^2 \cos^2(2\pi(t + \frac{1}{4})))^{\frac{3}{2}}},$$

where $t = 0$ in this formula is at A. Hence we can calculate the following upper bound for the curvature in A and B. Using $\frac{a}{b} > \frac{1}{\Theta}$ and $2b = |\Sigma|$, this yields

$$\kappa(A) = \frac{b}{a^2} = \frac{1}{b} \cdot \frac{b^2}{a^2} < \frac{2\Theta^2}{|\Sigma|}.$$

To prove the second part of the lemma, we argue as follows. Any orbit in $\mathcal{E}$ consists of three type of orbit segments shown in Figures 32, 34 and 35 (or orbits segments which have one vertex on the minor axis, but the arguments are the same in this case). Consider again the wedge XYZ in Figure 32. For any configuration in L, we have that for any two consecutive orbit segments either this case occurs or the case where X is on the left of AB so that XY intersects AB, see Figure 34 or lastly, that the wedge shaped configuration intersects AB twice, see Figure 35.

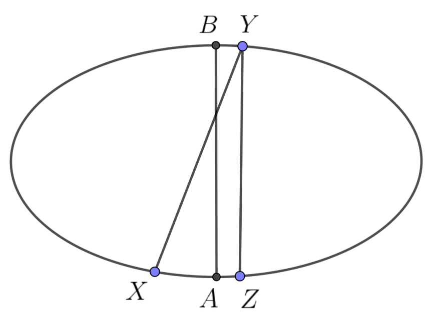

Figure 34: One of three possibilities for two consecutive orbit segments.

In the second case it should be clear the inequality $XY + YZ > 2 \cdot AB$ still holds. Finally, in the last case, in which X and Z lie on the same side of AB and Y on the other side, it should be clear as well the inequality holds.

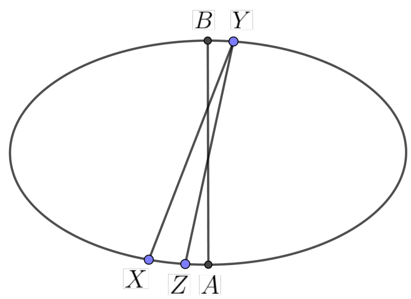

Figure 35: One of three possibilities for two consecutive orbit segments.

In the discussion before definition 8.2, we presented a way to construct elements in L and M. We now argue that these configuration can be made to have greater length than $\mathcal{O}_\Sigma$. This can be obtained by moving all points that are ordered as a line in Figures 29 and 30, closely towards each other in a symmetrical way while not crossing the symmetry axis Λ. Since consecutive points lie "far apart" it is that clear that these configurations shown in Figure 30 and 31 can made longer by applying the ideas in the first part of this lemma. $\square$

As an explicit example, one can squeeze in a configurations z as shown in Figure 36 such that it lies along the minor axis and has, by repeatedly applying the first part of Lemma 8.2, a strictly greater length of z than $\mathcal{O}_\Sigma$.

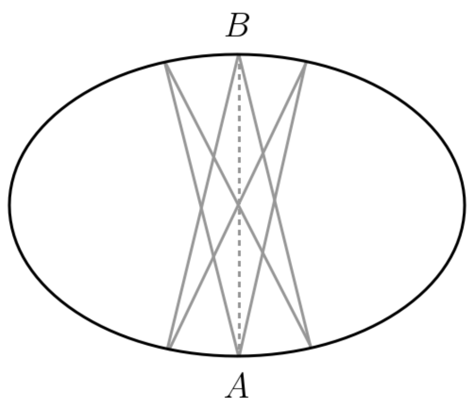

Figure 36: A $(6, 3)$ configuration the corresponding L that lies around AB.

The statements in Lemma 8.3 do not depend on the fact that the ellipse has two symmetry axis. Hence we can generalize the lemma to any billiard table which has one symmetry axis such that the curvature at its intersection point is less than $\kappa(A),\ \kappa(B) < \frac{2\Theta^2}{|\Sigma|}$. Note that it is not necessary to assume

$\kappa(A) = \kappa(B)$ either. This follows since by flattening the billiard curves even more at A or B, this will only make configuration XYZ longer than without this additional deformation. We shall state this consequence as a corollary of Lemma 8.3.

Corollary 8.1. *Assume $\mathcal{C}$ is a Σ-symmetric billiard table for which*

$$\kappa(A),\ \kappa(B) < \frac{2\Theta^2}{|\Sigma|}.$$

Then one can squeeze in a configuration $z \in L$ inside $\mathcal{C}$ with greater length than $\mathcal{O}_\Sigma$.

We so far considered Σ-symmetric billiards only. Since the assumptions of the lemma concern only locally the behavior of the billiard domain around $\mathcal{O}_\Sigma$ and by continuity we can even disturb the reflection symmetry into rotational symmetry, the consequence of the lemma still hold more generally. We shall put this all together in a proposition.

Proposition 15. *Assume $\mathcal{C}$ is a Σ-symmetric or a R_π-symmetric billiard table for which*

$$\kappa(A),\ \kappa(B) < \frac{2\Theta^2}{|\Sigma|}.$$

If $\mathcal{C}$ is Σ-symmetric one can squeeze in a configuration $z \in L$ with greater length than $\mathcal{O}_\Sigma$. If $\mathcal{C}$ is R_π-symmetric a similar statement holds for $z \in M$.

As will be explained after the partial proof of Conjecture 1, if the curvature bound $\kappa(A),\ \kappa(B) < \frac{2\Theta^2}{|\Sigma|}$ is not satisfied, then it might be that some orbits may cease to exist while other orbits remain to exist.

Lemma 8.4. *Let $\mathcal{C}$ be a Σ-symmetric or a R_π-symmetric billiard table. Then $\mathcal{C}$ has two $(2,1)$ orbits. In the case $\mathcal{C}$ has Σ symmetry there are precisely two orbits that are perpendicular to each other, while if $\mathcal{C}$ has R_π symmetry the two $(2,1)$ orbits go through the center of rotation.*

Proof. Assume first $\mathcal{C}$ is Σ-symmetric. Draw a perpendicular line from Σ to $\mathcal{C}$, see Figure 37. Note that the angle $\alpha(x)$ for some $x \in AB$, w.r.t. the tangent vector at some increases from 0 to π in a continuous way. Hence there is some point $z \in \Sigma$ such that $\alpha(z) = \frac{\pi}{2}$. Hence the line segments drawn from x to $\mathcal{C}$ in both directions will constitute a $(2,1)$ orbit.

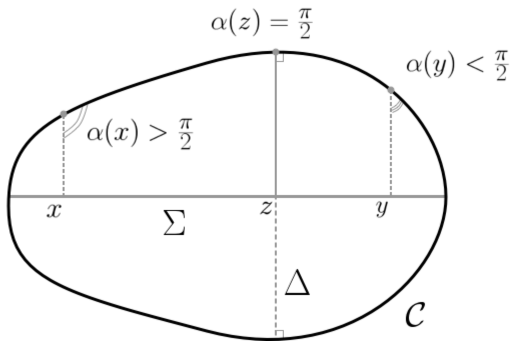

Figure 37: Proof that there exists precisely one diameter perpendicular to Σ.

To deal with the R_π case, we argue as follows. There can be drawn two perpendicular line segments from the center of rotation O to $\mathcal{C}$. Rotate $\mathcal{C}$ now over π radians. By symmetry the extended lines from the center to $\mathcal{C}$ also intersect $\mathcal{C}$ at a right angle. In Figure 38 this is illustrated. By strict convexity there are precisely two such orbits. This proves the lemma.

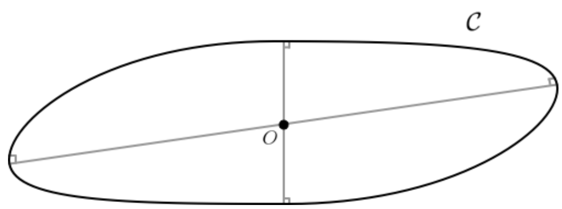

Figure 38: Arguing there exist two diameters in a rotational symmetric billiard table.

□

Recall that given a symmetry axis Σ, we denote by Δ the unique other diameter of $\mathcal{C}$.

Definition 8.3. *Let Σ be a symmetry axis of some $\mathcal{C}$. Denote by Π_Σ the set of all Σ-symmetric billiard tables such that for $A \in \mathcal{C} \cap \Sigma$ it holds that $\kappa(A) < \frac{2\Theta^2}{|\Sigma|}$. Let Σ' be a symmetry axis of some $\mathcal{C}$. Denote by Π_Δ the set of all Σ'-symmetric billiard tables such that for $A \in \mathcal{C} \cap \Delta$, $\kappa(A) < \frac{2\Theta^2}{|\Delta|}$.*

Theorem 9. *Assume $\mathcal{C}$ is Σ-symmetric. Let $A \in \Sigma \cap \mathcal{C}$. If $\kappa(A) < \frac{2\Theta^2}{|\Sigma|}$, then for each $n = 2k$ with $k \geq 1$ there exists a NB orbit in $L \subseteq \mathbb{X}_{2n,n}$. If for $B \in \Delta \cap \mathcal{C}$ it holds that $\kappa(B) < \frac{2\Theta^2}{|\Delta|}$, then there exist $(2n, n)$ NB orbits for each $n = 2k + 1$ with $k \geq 1$.*

Proof. We will prove existence of some invariant set L for the gradient flow of $W_{p,q}$. The set L will contain configurations that possess a kind of symmetry which contains NB orbits and is constructed such that the only Birkhoff orbit in L, up to orientation, is given by $\mathcal{O}_\Sigma$. This forces a configuration in L with greater length than $\mathcal{O}_\Sigma$ to flow to a NB orbit in forward time under the gradient flow.

Recall that $\sigma = \tau_{1,0}$ and that $\mathcal{S}$ is the reflection in the symmetry axis Σ of $\mathcal{C}$. Define the subspace $L \subseteq \mathbb{X}_{2n,n}$ as follows.

$$L = \{x \in \mathbb{X}_{2n,n} : \mathcal{S}(x) = \sigma^n(x)\}.$$

Observe that L is closed in $\mathbb{X}_{2n,n}$ and that $\mathbb{X}_{2n,n}$ is closed in $\mathbb{R}^{\mathbb{Z}}$. Note also that L does not contain Birkhoff configuration besides the configuration $\mathcal{O}_\Sigma \in \mathbb{X}_{2,1} \subseteq \mathbb{X}_{2n,n}$, see the discussion below. We are looking for stationary points of $W_{p,q}$ on L. Note that $\nabla W_{p,q}$ is not defined on the set

$$\Lambda := \bigcup_{i=1}^{2n} \{x \in \mathbb{X}_{2n,n} : x_i = x_{i+1}\}.$$

We claim that $\nabla W_{2n,n}$ leaves L invariant, i.e. that $\nabla W_{2n,n}(L) \subseteq L$. To show this, let $x \in L$. Since reflecting an orbit in a symmetry axis does not change the length of it, we have that $W_{2n,n}(\mathcal{S}(x)) = W_{2n,n}(x)$. Taking the total derivative of both sides at x yields

$$D(W_{2n,n} \circ \mathcal{S})(x) = DW_{2n,n}(x) \implies DW_{2n,n}(\mathcal{S}(x))\mathcal{S} = DW_{2n,n}(x).$$

Taking the transpose of both sides yields $\mathcal{S}(\nabla W_{2n,n}(\mathcal{S}(x))) = \nabla W_{2n,n}(x)$ or, since $\mathcal{S}^2 = \mathrm{Id}$, $\nabla W_{2n,n}(\mathcal{S}(x)) = \mathcal{S}(\nabla W_{2n,n}(x))$. Hence

$$\mathcal{S}(\nabla W_{2n,n}(x)) = \nabla W_{2n,n}(\mathcal{S}(x)) = \nabla W_{2n,n}(\sigma^n(x)) = \sigma^n(\nabla W_{2n,n}(x)),$$

where the last equality follows from Proposition 11. This shows that if $x \in L$, then $\nabla W_{2n,n}(x) \in L$.

This invariance property implies that we can consider the sub-dynamical system on the subspace L. So we can consider

$$\dot{x} = \nabla W_{2n,n}|_L(x),$$

where $\nabla W_{2n,n}|_L : L \to L$. The idea is that we minimize $W_{2n,n}$ on the subspace L to find NB orbits in $\mathbb{X}_{2n,n}$. We need to exclude that the minimum on L is attained at a Birkhoff orbit or at a degenerate configuration in L.

We consider the two cases where $\mathcal{C} \in \Pi_\Sigma$ or $\mathcal{C} \in \Pi_\Delta$. First assume $\mathcal{C} \in \Pi_\Sigma$ and let $n = 2k$ for some $k \geq 1$. Let $|\cdot|$ denote the length of a configuration. By assumption $\kappa(A) < \frac{2\Theta^2}{|\Sigma|}$ for $A \in \Sigma \cap \mathcal{C}$. Hence by Proposition 15, there exist configurations $z \in L$ such that $|z| > |\mathcal{O}_\Sigma|$. We will minimize $W_{2n,n}$ in L, that is, we will *maximize* the length. First we need to exclude that minimizing $W_{2n,n}$ yields the $(2,1)$ orbit along Δ. The reason for this is that the only critical points in $\mathbb{X}_{2n,n}$ besides the NB orbits ones, are $\mathbb{X}_{2,1}$. But this is straightforward to check because $\mathcal{O}_\Delta \notin L$. This follows since $\mathcal{O}_\Delta$ does not satisfy the symmetry criterion prescribed by L, because a reflection of $\mathcal{O}_\Delta$ in Σ corresponds to a shift σ^p, where p is odd. But we assumed $p = 2k$. The next thing to do is excluding that we obtain a degenerate configuration in Λ. We claim that any degenerate configuration $z \in L \cap \Lambda$ can be made longer to obtain a $z' \in L \setminus \Lambda$. The idea is based on the triangle inequality. Moving z_i and z_{i+1} apart, yields by the triangle inequality a strictly larger z' which can be made to remain to lie in L. Note that if one repeats this action of moving points at the other side of Σ , that is move the reflected points $z_{i+n} = \mathcal{S}(z_i)$ and $z_{i+n+1} = \mathcal{S}(z_{i+1})$ in the same way as z_i and z_{i+1}, then z' is forced to lie in L as well. In Figure 39 this is illustrated for $n = 4$.

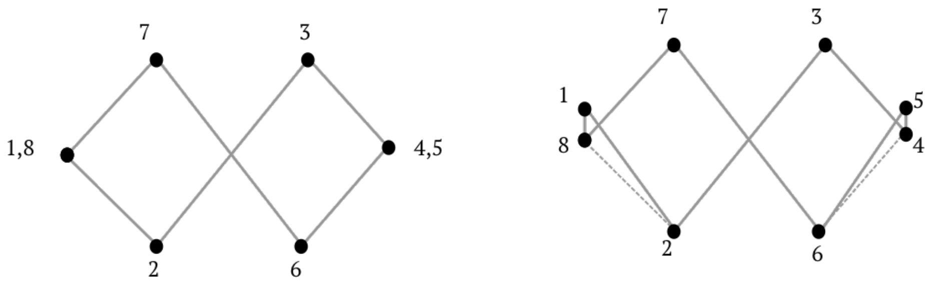

Figure 39: By the triangle inequality, the $(8,4)$ configuration on the right is longer than the degenerate $(8,4)$ configuration on the left. Observe that both configurations are in L.

Now we may conclude that minimizing $W_{2n,n}$ on L, that is maximizing the length, indeed yields a NB orbit of type $(2n, n)$ for any $n = 2k$, $k \geq 1$.

Next we consider $\mathcal{C} \in \Pi_\Delta$. Assume $n = 2k + 1$, with $k \geq 1$. Consider a configuration $x \in \mathbb{X}_{2n,n}$ that is symmetrical w.r.t. Σ, lies around Δ and is such that $|x| > |\mathcal{O}_\Delta|$. Since at $B \in \Delta \cap \mathcal{C}$ we have that $\kappa(B) < \frac{2\Theta^2}{|\Delta|}$, this is possible to arrange. We will again minimize $W_{2n,n}$ on L. Since $|x| > |\mathcal{O}_\Delta|$, we only need to exclude that Σ cannot be the minimum of $W_{2n,n}|_L$. This is clear since $\Sigma \notin L$ because n is odd now. More precisely, reflecting Σ in itself cannot shift the vertices in $\Sigma \cap \mathcal{C}$ an odd number of times away since then they would be at the opposite side from where they started. For the same reason as before, the minimum cannot be obtained in a degenerate configuration in Λ. This proves the minimum of $W_{2n,n}$ on L is indeed a non-Birkhoff orbit of type $(2n, n)$. □

Example 7. In this example we will use a slightly different approach to find a NB orbit. Consider the case of a $(6,3)$ configuration in some $\mathcal{C} \in \Pi_\Delta$, see Figure 40. Observe that since 3 is odd, x_{2k} and x_{2k+1} lie on different sides of Σ for each k. This has the implication that x cannot converge for $t \to \pm\infty$ to $\mathcal{O}_\Sigma$, since otherwise more than half of the vertices will group together at some point of $\Sigma \cap \mathcal{C}$. By assumption $|x| > |\mathcal{O}_\Delta|$ so that $\lim_{t\to\infty} \varphi^t(x) =: x_\infty$ cannot be Birkhoff, where φ^t denotes the gradient flow of $\dot{x} = -\nabla W_{6,3}(x)$. We conclude that x_∞ is non-Birkhoff.

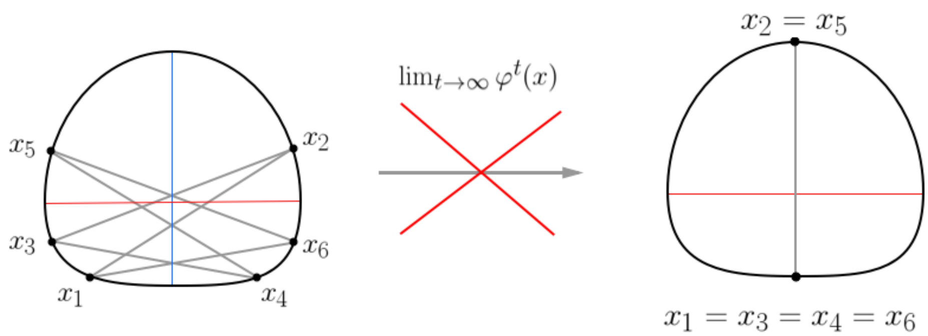

Figure 40: The configuration on the left cannot converge to the configuration on the right since this configuration is degenerate. The red line segment corresponds to Δ.

$\triangle$

We remark that for n is even it may hold that a $(2n, n)$ periodic orbit in fact has period $(n, \frac{n}{2})$. So for example if we find a $(8, 4)$ NB orbit via Theorem 9, a priori it might be in fact the $(4, 2)$ NB orbit already found by applying Theorem 9 for the case $n = 2$.

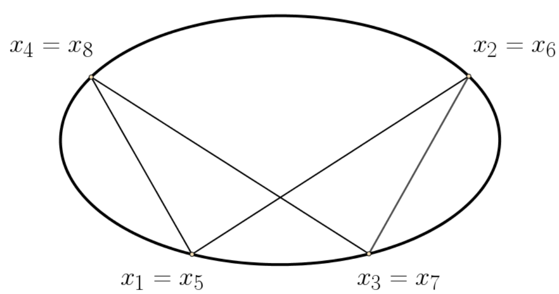

Figure 41: If $\mathcal{S}(x) = \sigma^2(x)$ then $\mathcal{S}(x) \neq \sigma^4(x)$, where $\mathcal{S}$ is the reflection in the minor axis of this ellipse.

However, the symmetry condition prescribed in the definition of L in Theorem 9 automatically excludes this possibility: the symmetry of a $(8, 4)$ NB orbit does not match the symmetry condition for $(4, 2)$ orbits. This easy to see since the shift σ is applied 4 times instead of 2. But the $(4, 2)$ configurations are defined via σ^2. This implies that for each choice of n in Theorem 9, we do obtain genuine NB orbit of period $(2n, n)$.

Another corollary states that in ellipses, mirror symmetric NB orbits $\mathcal{O} \in \mathbb{X}_{2n,n}$ are positioned around the x or y-axis according to the parity of n.

Corollary 8.2. *Let $\mathcal{E}$ be an ellipse given by $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ and parametrize it with period 1. Let $\mathcal{E} \in \mathbb{E}_{\frac{1}{\Theta}}$ denote the set of ellipses for which $\frac{a}{b} > \frac{1}{\Theta}$. Consider an ellipse $\mathcal{E} \in \mathbb{E}_{\frac{1}{\Theta}}$. Then $\mathcal{E} \in \mathbb{E}_{\frac{1}{\Theta}}$ contains a $(2n, n)$ NB orbit $\mathcal{O} \in L$ orbit for each $n \geq 2$. If n is even, then $\mathcal{O}$ is symmetric in the y-axis while if n is odd, then $\mathcal{O}$ is symmetric in the x-axis.*

Proof. We claim that for a suitable choice of Σ and Σ' it holds that $\mathcal{E} \in \Pi_\Sigma \cap \Pi_\Delta$. Note that the curvature at the ends of the major axis can never be smaller than $\frac{\Theta^2}{2b}$ since it equals at those points $\frac{b^2}{a^2} \cdot \frac{1}{b} = \frac{1}{2b\Theta^2}$ and $\frac{1}{b\Theta^2} > \frac{\Theta^2}{b}$.

Take Σ to be the minor axis of the ellipse and let Σ' be the major axis, so that Δ is again the minor axis. For $\mathcal{E} \in \mathbb{E}_{\frac{1}{\Theta}}$ it holds that $\kappa(B) = \kappa(A) < \frac{2\Theta^2}{|\Sigma|}$ for $A \in \Sigma \cap \mathcal{E}$, $B \in \Delta \cap \mathcal{E} = \Sigma \cap \mathcal{E}$ see the proof of Lemma 8.1. The claim follows by applying Theorem 9. $\square$

If the curvature at points in $\Sigma \cap \mathcal{C}$ is too large, some particular type of periodic orbits may not exist. It is possible however that other types exist. This is a general phenomenon that has to do with the kind of orbit segments a configuration is made of. Consider for example the $(6, 3)$ periodic orbit in Figure 36. The curvature in this case can be made slightly larger while its length is still greater $|\mathcal{O}_\Sigma|$, the orbit along the minor axis. The configuration in Figure 21 on the contrary will in fact sooner cease to be longer than $\mathcal{O}_\Sigma$, because some line segments are completely on the left and right of AB. The positions of these segments are the worst case scenario: the curvature needs to be made very small in order to compensate their location. We will later see for ellipses that the curvature bound in Theorem 9 can be weakened. Some periodic orbits might become non-existent, but still there can be found at least one NB orbit of type $(2n, n)$ for each $n \geq 2$. In fact, as soon $\frac{a}{b} > \sqrt{2}$, then there will Σ-symmetric non-Birkhoff orbits present. This will be explained in section 8.2.

Let us continue considering the ellipse and get some insight on how Theorem 9 fails if $\kappa(A)$ is too large in the case of a $(4, 2)$ NB orbit. Consider the V-shaped orbit shown in Figure 42. If $\frac{a}{b} \leq \sqrt{2}$, any segment AX with X varying in the region $y \geq 0$ has length strictly less than the length of the minor axis. Let $F(x)$ denote the square of the length of segment AX. Then $F(x)$ is given by

$$F(x) = (b + b\sqrt{1 - \frac{x^2}{a^2}})^2 + x^2.$$

The derivative of $F(x)$ on $x \in [0, a)$ is non-positive:

$$F'(x) = -2\frac{b^2}{a^2}\frac{1 + \sqrt{1 - \frac{x^2}{a^2}}}{\sqrt{1 - \frac{x^2}{a^2}}}x + 2x \leq -\frac{1 + \sqrt{1 - \frac{x^2}{a^2}}}{\sqrt{1 - \frac{x^2}{a^2}}}x + 2x \leq -2x + 2x = 0,$$

where in the first step we made use of the assumption that $\frac{b^2}{a^2} \geq \frac{1}{2}$. This shows that there does not exists a V-shaped configuration z_0 with greater length. Consider the flow φ^t of $\dot{x} = -\nabla W_{4,2}(x)$. It holds that $\omega_+(z_0) = \mathcal{O}_\Sigma$ and $\omega_-(z_0)$ is the trivial configuration (A, A, A, A) (the ω-limit set notation was introduced in Theorem 6). Hence no V-shaped orbit can exists and therefore no $(4, 2)$ NB orbit exists since either they come in a 1-parameter family or they do not exist, see Theorem 10.

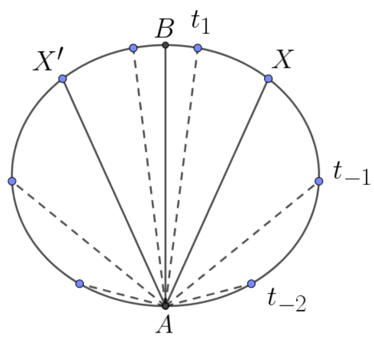

Figure 42: Some V-shaped configuration at times $t_{-2} < t_{-1} < 0 < t_1$.

Example 8. Consider an ellipse $\mathcal{C} \in \mathbb{E}_{\frac{1}{\Theta}}$. According to Theorem 9 there exist orbits of the NB type $(2n, n)$. In Figure 43 a $(10, 5)$ NB orbit in the corresponding subspace L is shown. Note that the orbit is symmetric w.r.t. the major axis, since n is odd in this case, cf. Corollary 8.2. As another example a $(8, 4)$ NB orbit in the corresponding space L is shown in Figure 44. Since n is odd, the orbit is mirror symmetric w.r.t. the minor axis, compare again with Corollary 8.2. $\triangle$

We will now focus on the case where the domain has rotational symmetry over π and prove that a similar statement holds as described in Theorem 9. First we state a lemma about the existence of configurations in M for $\mathcal{E} \in \mathbb{E}_{\frac{1}{\Theta}}$

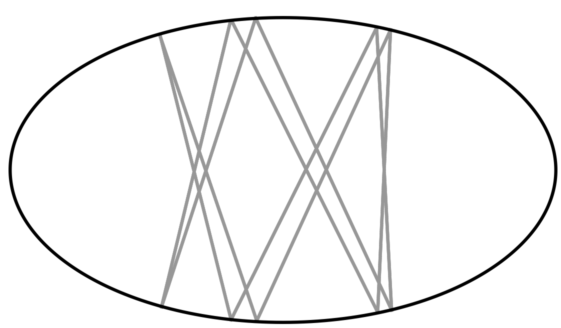

Figure 43: $(10, 5)$ Σ-symmetric periodic orbit

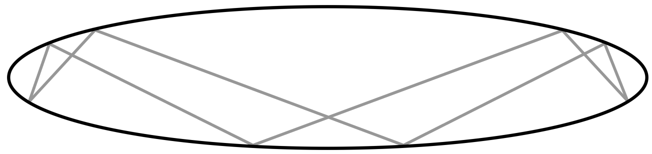

Figure 44: $(8,4)$ Σ-symmetric periodic orbit.

that have non-zero intersection index with $\mathcal{O}_\Sigma$, the $(2,1)$ orbit along the minor axis.

Lemma 8.5. *Let $\mathcal{O}_\Sigma$ be the minor axis an ellipse $\mathcal{E} \in \mathbb{E}_{\frac{1}{\Theta}}$. Then there exists an $x \in M$ such that $I(x, \mathcal{O}_\Sigma) > 0$.*

Proof. Consider the Aubry diagram of $\mathcal{O}_\Sigma$ and fit in close to it a configuration $x \in M$. Since x is NB, it "wiggles" around. Since we have a lot of freedom in where to place x_1, only close to Σ and not on Σ, we can obtain x to have intersections with $\mathcal{O}_\Sigma$. That is, we can obtain $I(x, \mathcal{O}_\Sigma) > 0$. $\square$

I was not able to give a complete proof of Conjecture 1 yet; only a particular case is proved. Numerical experimentation in ellipses provides some evidence that it is true.

Conjecture 1. *Assume $\mathcal{C}$ has R_π-symmetry, let m denote a diameter of $\mathcal{C}$ and let $\mathcal{O}_m$ be the $(2,1)$ orbit along m. Let $A, B \in \mathcal{O}_m \cap \mathcal{C}$. If $\kappa(A), \kappa(B) < \frac{2\Theta^2}{|\Sigma|}$. Then there exists a NB orbit in $\mathbb{X}_{2n,n}$ for each $n \geq 3$ odd.*

Partial proof. We shall only prove Conjecture 1 if the other diameter is smaller than m and the general case only for $n = 3$. The proof follows basically the same idea as in the Σ-symmetry case. We will show again there is a particular subset $M \subseteq \mathbb{X}_{2n,n}$ which is invariant under $\nabla W_{2n,n}$ and thus obtaining again a sub-dynamical system on M. Define for $a \in [0,1]$, $R_a : (S^1)^{2n} \to (S^1)^{2n}$, by $R_a(x) = x + a$ componentwise. We define M to be the subspace of certain types of rotationally symmetric configurations:

$$M = \{x \in \mathbb{X}_{2n,n} : \sigma^n(x) = R_{\frac{n}{2}}(x)\}.$$

We already remarked on p.54 that M is non-empty. Observe that M is closed in $\mathbb{X}_{2n,n}$. In rotational symmetric billiard tables over π, there are always precisely two period $(2,1)$ orbits (up to orientation). This follows from Lemma 8.4.

Let R be any rotation. Observe that $W(R(x)) = W(x)$ since a rotation over $\frac{n}{2}$ of the whole system does nothing to the length of a configuration. Taking the total differential of both sides of this equation, we obtain

$$DW_{2n,n}(R(x))R = DW_{2n,n}(x) \implies \nabla W_{2n,n}(R(x)) = R(\nabla W_{2n,n}(x)).$$

This implies that $\nabla W_{2n,n}(M) \subseteq M$. In particular, the flow of $\dot{x} = -\nabla W_{2n,n}(x)$ leaves M invariant.

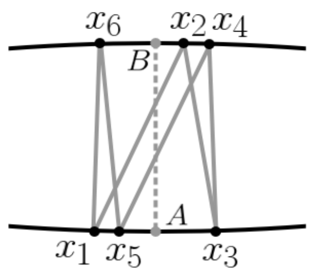

Figure 45: A $(6, 3)$ configuration in M. $\mathcal{O}_m$ is the $(2, 1)$ orbit along AB.

Next we apply similar ideas as in the proof of theorem about fitting M symmetric states with greater length than some $\mathcal{O}_m \in \mathbb{X}_{2,1}$ inside $\mathcal{C}$. One major difficulty in the rotational symmetry case however, is that both $(2, 1)$ orbits lie in the subspace M. This is in contrast with the reflection symmetry case where the lack of this property played a key ingredient in our proof.

If $\mathcal{O}_M$ is not longer than $\mathcal{O}_m$, we can minimize $W_{2n,n}$ on M. Since M is closed, we have that some $x_{\max}$ lies in M. In order to conclude that this is a critical point, we need to exclude it is degenerate, that is we need to show that $x_{\max} \notin \Lambda = \bigcup_{i=1}^{2n} \{x \in \mathbb{X}_{2n,n} : x_i = x_{i+1}\}$. To show this we argue similarly as in the proof of Theorem 9. Any degenerate configuration in M can be made longer by moving coinciding vertices apart, by the triangle inequality. This can be done in such way the resulting configuration is still in M. Hence the minimum (so the maximal length) is not attained in a degenerate configuration. Therefore, it must be a critical point of $W_{2n,n}$. In Figure 46 this is illustrated for a $(6, 3)$ orbit.

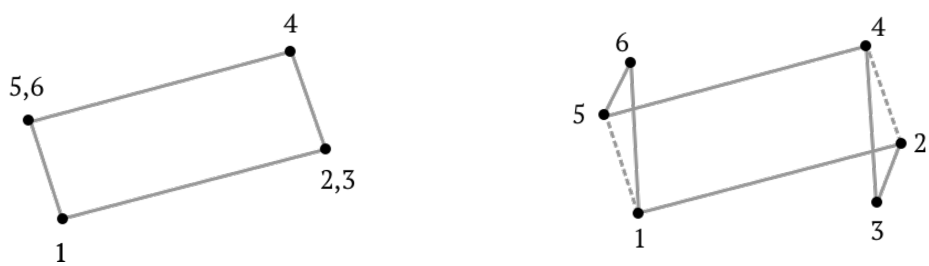

Figure 46: By the triangle inequality the $(6, 3)$ configuration on the right is longer than the $(6, 3)$ configuration on the left. Both are in M.

If the other $(2, 1)$ orbit, call it $\mathcal{O}_M$, is longer than $\mathcal{O}_m$ there exist states in M that might converge under the gradient flow towards it. We argue for $n = 3$ that this cannot happen via M. First of all it holds that NB states $y \in \mathbb{X}_{2n,n}$ sufficiently close to $\mathcal{O}_m$ have non zero intersection index with $\mathcal{O}_m$ since this holds for an elliptical billiard table, see Lemma 8.5. So $I(x, \mathcal{O}_m) \geq 2$.

Assume now that the vertices are ordered such that not all x_i with i even (or odd) are on one side of AB, see Figure 45. Then in order for x to converge to $\mathcal{O}_M$, the intersection index I should decrease to zero because $I(\mathcal{O}_m, \mathcal{O}_M) = 0$. By the Sturmian Lemma, Proposition 13, this happens at some point where y and $\mathcal{O}_{AB}$ are not transverse. In Figure 45 we see that all even vertices must group together and all odd vertices must group together together to from an orbit in $\mathbb{X}_{2,1}$. Now, $I(\varphi^t(x), \mathcal{O}_m)$ can only decrease at times t where x and $\mathcal{O}_m$ are non-transverse. This can happen only can only if $\varphi^t(x)_k$ and $\varphi^t(x)_{k+3}$ move trough the points A and B. It seems however not possible to achieve this because of the symmetry condition posed by M. The explanation is that there are limited ways of x and $\mathcal{O}_m$ getting untangled via the space M. For example, a priori it might happen that they become non-transverse via a whole sequence of touchings in between $\varphi^t(x)_k$ and $\varphi^t(x)_{k+3}$. However, it appears that than too many vertices lie on $\mathcal{O}_m$ so that we cannot make sure that $|x| > |\mathcal{O}_m|$. In fact, this seems to be the only case to check since at the points where x and $\mathcal{O}_m$ become non-transverse, they must go through A and B in order to converge to the major axis, see Figure 45. We explicitly make the remark that for other types of $(6,3)$ configurations the argument immediately fails, for example, if they are more than two intersections.

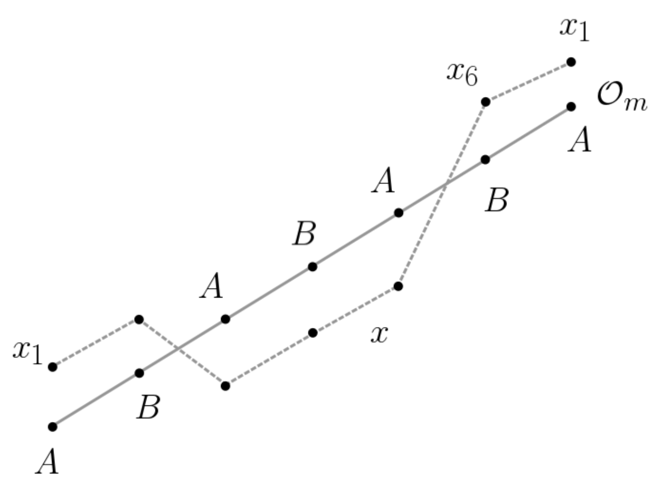

Figure 47: Schematic representation of x and $\mathcal{O}_m$, where x is the configuration from Figure 45. We consider the orientation to be *clockwise*.

Hence $\mathcal{O}_M \neq \lim_{t\to\infty} \varphi^t(x) =: x_\infty$ if $t \to \infty$. This means that x_∞ cannot be a Birkhoff orbit and hence that it must be a NB orbit in M, which is therefore rotational symmetric. $\square$

Example 9. Consider some elliptical billiard table such that the $\kappa(A) < \frac{2\Theta^2}{|\Sigma|}$. Then we can squeeze in the $(6,3)$ configuration z shown in Figure 48 such that its length is greater than $6 \cdot AB$. This follows from Lemma 8.2 applied 3 times.

Conjecture 1 now claims that z converges to a $(6,3)$ NB orbit z_∞. This orbit is also shown in Figure 48.

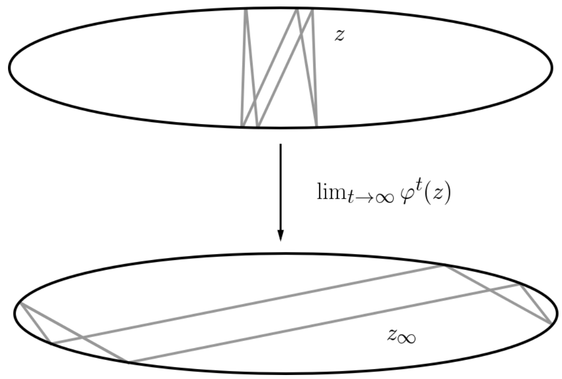

Figure 48: Illustration of example 10.

$\triangle$

In example 8 a reflection symmetric $(10,5)$ NB orbit was shown. Conjecture 1 yields the existence of $(10,5)$ NB orbits that are rotation symmetric, see Figure 49.

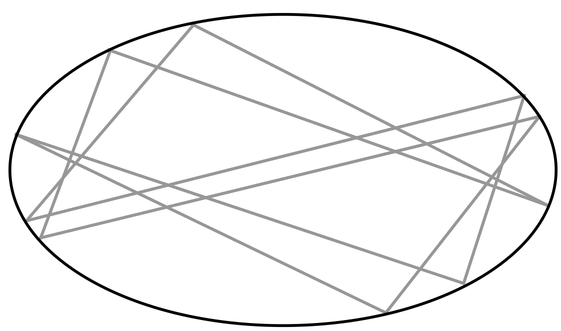

Figure 49: A $(10,5)$ rotation symmetric NB orbit.

Corollary 8.3. *Let $\mathcal{C}$ be a reflection symmetric billiard table with two symmetry axis Σ_a and Σ_b and assume the curvature at Σ_a or Σ_b is less than $\frac{2\Theta^2}{|\Sigma|}$. Denote by O their common point. Then there exist $(2n,n)$ NB orbits, for n odd, which are rotationally symmetric w.r.t. O.*

Proof. Since a domain with two perpendicular symmetry axis is also rotationally symmetric over π, the claim can be proven by an application of Conjecture 1. $\square$

8.2 Elliptical billiards

We will now focus on some properties of elliptical billiard systems and present a complete classification of Birkhoff and non-Birkhoff periodic orbits in these systems. It may appear that this is a well understood theme in the subject of convex billiard systems. After all, this is one of the first non-trivial billiard domain after the circular billiard table. However, after more than 80 years since G.Birkhoff posed his famous conjecture about elliptical billiard, it remains unsettled. We will not dwell on this particular conjecture only for the remark that is has to do with the integrability of the elliptical billiard. More precisely, that the ellipse is the only billiard system that is integrable (regarding the circle as a degenerate ellipse). Besides this famous conjecture by G. Birkhoff, there remain also other conjectures about ellipses. We shall not discuss this here, but we refer to [13].
An important theorem for this section and section 9 is the so-called *Poncelet Theorem.*

Theorem 10. *Let $\mathcal{C}$ and $\mathcal{D}$ be two conics in the plane. Suppose that there is a polygon inscribed in $\mathcal{C}$ and circumscribed about $\mathcal{D}$. Then there are infinitely many such polygons and all of them have the same number of sides. Moreover, each point of $\mathcal{C}$ is a vertex of such a polygonal line.*

Proof. See [8]. □

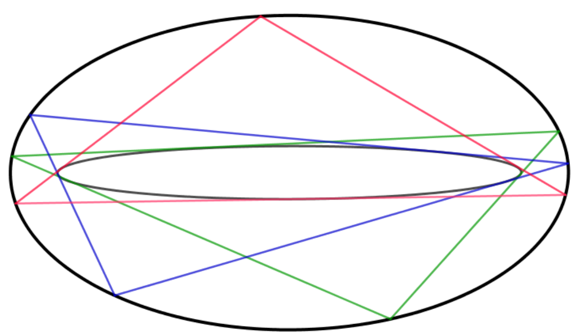

Figure 50: Poncelet's theorem for two confocal ellipses.

Lemma 8.1 and 8.2 can be used to give an interesting proof of the fact that for p, q relatively prime, the only (p, q)-orbits in ellipses are in fact minimizers.

Proposition 16. *Assume $x \in \mathbb{X}_{p,q}$ with $gcd(p, q) = 1$ is an orbit. Then x is a minimizer of $W_{p,q}$.*

Proof. By Poncelet's theorem all such (p, q) orbits as in the statement come in a 1-parameter family $\mathcal{F} = \{x(s) \in \mathbb{X}_{p,q} : s \in S^1\}$ These orbits are in fact regular (star) polygons with Schläfli symbol $\{p/q\}$ by Proposition 6, and for each orbits starting anywhere on $\mathcal{E}$, there is such an orbit. Consider some vertex $x_i(s)$ of the orbit $x(s)$. When $x_i(s)$ is moved, say, counterclockwise, all other vertices $x_j(s)$ of $x(s)$ will move counterclockwise as well. This can be seen as follows. Since all segments $|x_i(s)x_{i+1}(s)|$ of $x(s)$ touch the inner ellipse $\mathcal{E}'$, moving $x_i(s)$ counterclockwise, will force $x_{i+1}(s)$ to move counterclockwise as well, see Figure 51.

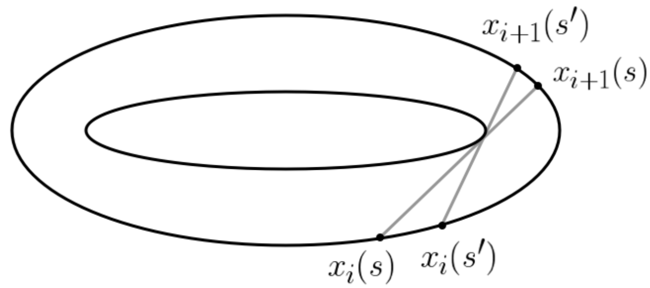

Figure 51: Proving that $x'(s) > 0$.

In particular, $x_j'(s)$ has the same sign as $x_i'(s)$. Note that the eigenvectors corresponding to this family $\mathcal{F}$, $x'(s)$, all have an eigenvalue $\lambda = 0$. This follows simply from that fact that for each $s \in [0, 1]$ it holds that $-\nabla W_{p,q}(x(s)) = 0$ and therefore $-\nabla^2 W_{p,q}(x(s)) \cdot x'(s) = 0$ for each $s \in [0, 1]$. In fact, there are no other eigenvectors with $\lambda = 0$ than these $x'(s)$. This is because of the structure of the spectrum of $-\nabla^2 W_{p,q}(x(s))$ and the fact that $x'(s)$ has zero sign changes for all $s \in \mathcal{E}$. This last statement can be seen as follows.

By Lemma 8.2, since the eigenvector $x'(s)$ has zero signchanges, it has the largest eigenvalue λ_0 and $\lambda_0 = 0$ by the discussion above. But the spectrum of $-\nabla^2 W_{p,q}(x)$ equals $\{\lambda_0 > \lambda_1 \geq \lambda_2 > ... > \lambda_{p-1}\}$. Hence $\lambda_0 = 0$ is in fact simple. This shows there is precisely one 1-parameter family of (p, q) orbits which therefore are necessarily minimizers of $W_{p,q}$.

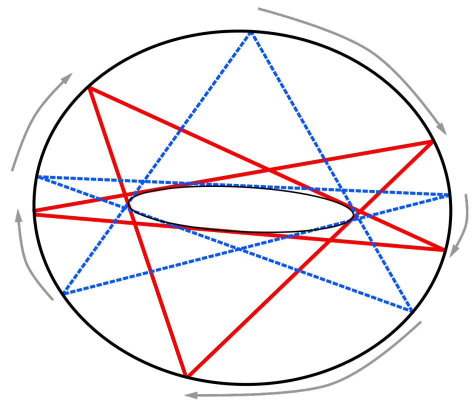

Figure 52: Moving any vertex of a $(5,2)$-orbit clockwise, causes any other vertex to move in the same direction

$\square$

Remark 6. *If the assumption* $\gcd(p,q)=1$ *is dropped, Proposition 16 does not necessarily hold. This follows from Theorem 9 in which we proved the existence of Non-Birkhoff orbits of period* $(2n,n)$ *inside elliptical billiards. Since non-Birkhoff orbits are not minimizers, the claim follows.*

We will now discuss Poncelet's theorem (Theorem 10) for the case of an ellipse and a confocal hyperbola. We start with a proposition.

Proposition 17. *Let* $\mathcal{E}$ *be an ellipse and let* $\mathcal{H}$ *be a confocal hyperbola. Let two lines satisfy the billiard law. If one the lines is tangent to* $\mathcal{H}$*, then so is the other.*

We refer to [8] for a proof. Consider Figure 53. Proposition 17 says that if one segment from a billiard orbit touches the hyperbola ellipse $\mathcal{H}$, then all segments do. The only way to draw a segment tangent to $\mathcal{H}$ is to start at a point on $\mathcal{E}$ in between the two branches of $\mathcal{H}$. Poncelet's theorem then implies that if there is one closed orbit in $\mathcal{E}$ that touches $\mathcal{H}$, then starting from any point on $\mathcal{E}$ in between the two branches of $\mathcal{H}$, there is another closed orbits of the same period tangent to $\mathcal{H}$. In section 8.2 we will see examples of this kind of orbits. This result implies, as in the case of the ellipse, that if a segment from a periodic orbit inside $\mathcal{E}$ touches $\mathcal{H}$, this is holds for any segment of the orbit. We note that the segments might touch $\mathcal{H}$ outside $\mathcal{E}$.

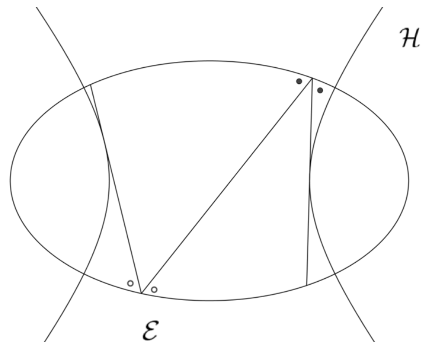

Figure 53: If one line segment touches $\mathcal{H}$, then all line segments do if the billiard law is being satisfied at the points of intersection of the line segments on $\mathcal{E}$.

Proposition 15 yields, together with Theorem 11 a neccesary conditions in which direction one should hit the billiard ball to obtain Birkhoff and non-Birkhoff orbits. The next theorem is taken from [4].

Theorem 11. *Let $\mathcal{E}$ be an ellipse with foci F_1 and F_2. If a segment of a biliard trajectory does not intersect the line segment F_1F_2, then all segments of this trajectory do not intersect F_1F_2; if a segment of a trajectory intersects F_1F_2, then all the segments of this trajectory intersect F_1F_2 and all are tangent to the same hyperbola to the same hyperbola with foci F_1 and F_2.*

Proof. See Theorem 4.4 in [4]. □

For example, if one wants to obtain a Birkhoff periodic orbit, then it is necessary to hit the ball in the direction that will cause the ball not to go through F_1F_1. Obviously this not sufficient since the resulting trajectory might not close up.
Before we state the next proposition, we make an important remark about NB orbits in ellipses. Namely, NB orbits in ellipses must always have an even number of reflection points. This follows from Theorem 11 for the following reason. Since NB orbits have hyperbolae as caustics, each orbit segment crosses F_1F_2 by Proposition 12. Hence, for the orbit to close up in finitely many steps, there must be an even number of crossings of the orbit segments with F_1F_2.

We will now discuss a relation between caustics and the structure of billiard trajectories. Recall that a caustic is a curve c such that if a segment of a trajectory is tangent to it, then all segments of the trajectory are tangent to c, see definition 4.5. The characterization of Birkhoff orbits as regular polygons, together with Poncelet's theorem, enables us to give another distinction between Birkhoff and non-Birkhoff orbits in elliptical billiards in terms of caustics. This formulated in Proposition 18. Before we give of proof of this proposition, we need a lemma.

Lemma 8.6. *Let $\mathcal{P}$ be a regular polygon. Then a line L avoiding all vertices of $\mathcal{P}$ does not intersect all line segments of $\mathcal{P}$.*

Proof. The statement is obvious for 3-gons and 4-gons. Observe that any regular polygon $\mathcal{P}$ with Schläfli symbol $\{p/q\}$ can be circumscribed by a convex p-gon $\mathcal{Q}$ by connecting consecutive vertices by line segments. It is clear that a line L intersecting a convex p-gon $\mathcal{Q}$ not through its vertices, has precisely two intersections in total. In particular, the line L avoids at least one pair of consecutive segments AB and BC from $\mathcal{Q}$ as soon as $p \geq 5$, see Figure 54.

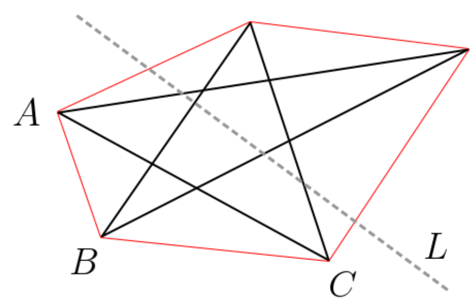

Figure 54: The convex p-gon is shown in red. Since L avoids AB and BC, it will also avoid AC.

Consequently L avoids also the segment AC from $\mathcal{P}$. This proves the claim. □

Proposition 18. *Let $\mathcal{E}$ be an ellipse. Assume $p > 2$. Then $x \in \mathcal{B}_{p,q}$ is a periodic orbit iff x has an ellipse $\mathcal{E}'$ as its caustic. In particular, x is non-Birkhoff iff x has an hyperbola $\mathcal{H}$ as its caustic.*

Proof. We use that in ellipses, there are precisely three possible kind of orbits. The first possibility are orbits that are tangent to an elliptic caustic. The second one are those orbits that are tangent to a hyperbolic caustic. The last kind of orbits are those that pass trough a focal point. It is well known that these third kind of orbits converge to the major axis of $\mathcal{E}$ and therefore cannot be periodic.

Suppose $x \in \mathcal{B}_{p,q}$ is an orbit. By Proposition 6, x is a (topological) regular polygon with Schläfli symbol $\{p/q\}$. Assume that x has instead of an ellipse a hyperbola as its caustic. By a standard result, all segments of the orbit x intersect the segment $|F_1F_2|$ see Theorem 11. We claim that x cannot be a regular polygon and hence cannot be a Birkhoff orbit. This simply follows from the fact that a line intersecting a regular polygon not through a vertex, cannot intersect *all* segments from the polygon. See lemma 8.6. This is a contradiction. If the extended line $|F_1F_2|$ actually intersects a vertex of x, we can can move x by the 1-parameter property to a general position (Theorem 10). Apply now the same argument. Hence x has an ellipse as its caustic. This proves the first direction.

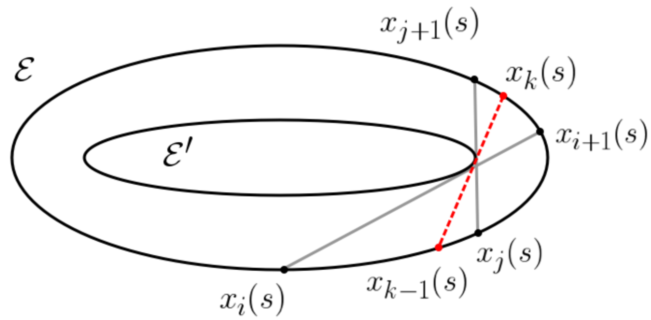

Figure 55: Argument that only regular polygonal orbits can have an elliptic caustic

For the other direction suppose that x touches an elliptic caustic $\mathcal{E}'$ but that it is not a regular polygon. Poposition 5 implies that there exist i and j for which $x_i \prec x_j$ but not $x_{i+1} \prec x_{j+1}$. The order of x_{i+1} and x_{j+1} does remain preserved, by continuity. So there must be some x_k in between x_{j+1} and x_{i+1}, see Figure 55. However, this is not possible because since in this case x_{k-1} should be in between x_i and x_j as well. This is in contradiction with the assumption that $x_i \prec x_j$. Hence x should be a regular polygon, that is, x is Birkhoff. □

Let us recap what we have found thus far about periodic orbits in ellipses. By the existence of minimizers, we have proved existence of (p, q) periodic orbits for p, q relatively prime. Theorem 9 guaranteed the existence of non-Birkhoff orbits of type $(2n, n)$ as well. Poncelet's theorem stated that these orbits, except for the $(2, 1)$ orbits, come in a 1-parameter family, so that there is a whole continuum of these orbits. A natural question is whether there are more NB periodic orbits or if this is everything there is. In fact, it is not so difficult to show that the $(2n, n)$ NB orbits are the only NB orbits in ellipses. This has to do with the fact that only the $(2, 1)$ orbits in ellipses are isolated, while the rest of (p, q) orbits with p, q relatively prime come in a 1-parameter family by Poncelet's theorem. More precisely, we have the following proposition.

Proposition 19. *Non-Birkhoff orbits of type $(2n, n)$, $n \geq 2$, are the only non-Birkhoff orbits in ellipses. In particular, no non-Birkhoff orbits of type $(2n, m)$ where $m \neq n$ exist.*

Proof. Let $x \in \mathbb{X}_{2n,m}$ where we assume that $\gcd(2n, m) \neq 1$, so that NB orbits are a priori possible. Assume x is non-Birkhoff. Let $\alpha = \frac{2n}{\gcd(2n,m)}$, $m \neq n$ and $\beta = \frac{m}{\gcd(2n,m)}$. Consider the space $\mathbb{X}_{\alpha,\beta} \subseteq \mathbb{X}_{2n,m}$. By Theorem 10, there exists a full 1-parameter family $\mathcal{F}$ of (α, β) Birkhoff orbits. Full means in this context that for each position s on the billiard table $\mathcal{E}$, there is a (α, β) Birkhoff orbit.

There are Birkhoff orbits c such that $c < x$, where we interpret $x, c \in \mathbb{X}_{2n,m}$, by just translating c enough downwards. But by the full 1-parameter property we can in fact obtain that such a c touches x from below by choosing a suitable $c_0 = \pi_0(c) \in \mathbb{R}$, where π_0 denotes the projection on the first position of some element in $\mathbb{R}^{\mathbb{Z}}$. Since orbits cannot touch by Corollary 6.1, this implies that x was a Birkhoff itself in the 1-parameter family $\mathcal{F}$. In particular, x cannot be non-Birkhoff. $\square$

Before we continue, recall again the notation $\mathbb{E}_{\sqrt{2}}$ which denote the space of all elliptical billiards for which the ratio of the major axis with the minor axis is greater than $\sqrt{2}$.

We now come to a result, Proposition 20, about NB orbits in ellipses in the space $\mathbb{E}_{\sqrt{2}}$ that will be useful at a later stage in the context of perturbed billiard systems, but which is in fact interesting in its own right. Furthermore, it has an appealing proof which exploits only elementary geometric properties. So in a sense it can be considered as a theorem in Euclidean geometry. I have not found this in the literature. We start with a (known) lemma.

Lemma 8.7. *A circle containing the foci and a point p on the curve will intersect the minor axis at the points of intersection of the tangent and the normal to the curve from point p.*

Proof. We choose to give an elementary geometric proof of this fact.

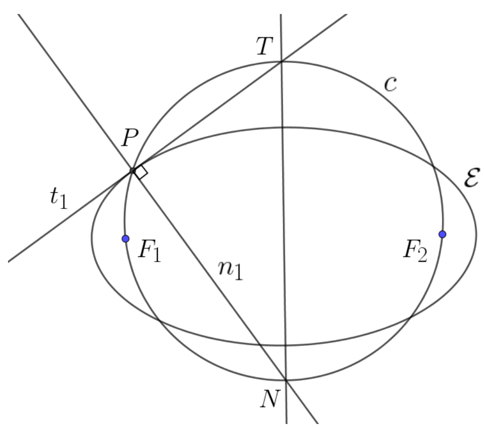

Figure 56: Illustration of Lemma 8.7.

Conider an ellipse $\mathcal{E}$ and draw a circle c through the two foci F_1 and F_2 of $\mathcal{E}$. Note that this circle has a diameter on the minor axis. Denote by P an intersection point of c and $\mathcal{E}$, see Figure 56. Now draw the tangent line t_1 through P as well as the normal n_1 through P. Denote by T and N the intersection point of respectively the tangent and the normal with the circle c. By construction,

ΔTNP is right-angled. Hence there is a unique circle through P, T and N with its center on TN. By uniqueness, the circle must be c. Hence the intersection points T and N indeed lie on the minor axis, since we already remarked that a diameter of c lies on the minor axis. This concludes the proof. $\square$

Proposition 20. *Let $\mathcal{E} \in \mathbb{E}_{\sqrt{2}}$. Then there exists a 1-parameter family of $(4,2)$ non-Birkhoff orbits. Furthermore all these orbits are symmetric w.r.t. the minor axis of $\mathcal{E}$.*

Proof. Let c be a circle that goes through the foci of $\mathcal{E}$. Since $\frac{a^2}{b^2} > 2$, c can be chosen to have three or four intersection points with $\mathcal{E}$. It is easy to show that a diameter of c lies on the extension of the minor axis of $\mathcal{E}$. There are two cases to consider. Case 1: c intersects $\mathcal{E}$ in four points. Case 2: c intersects $\mathcal{E}$ in three points. Assume first we are in case 1. Consider some intersection point P_1 and connect all intersection points with straight line segments in the fashion of Figure 57.

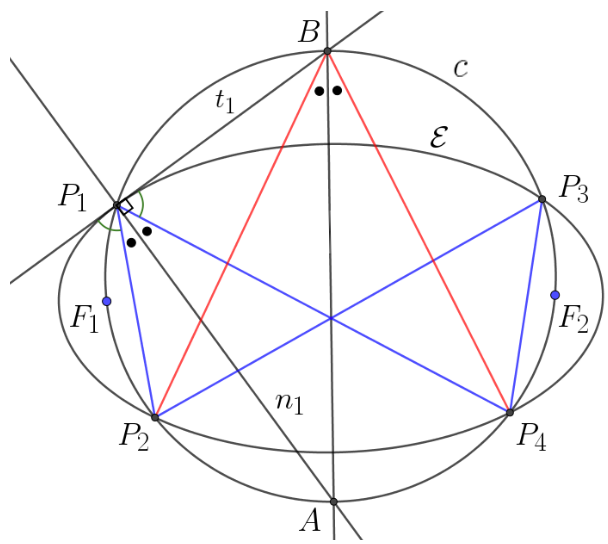

Figure 57: Proving that $P_1P_2P_3P_4$ is a $(4,2)$ billiard orbit.

By Lemma 8.7, the tangent line to P_1 goes through the intersection point B of the extension of the minor axis and c. By the same lemma we also have that the normal n_1 from P_1 goes through the intersection point A. Label now the rest of the points of the configuration so that the configuration is given by $P_1P_2P_3P_4$. Since these four point all lie on the circle c, the angles $\angle P_2P_1P_4$ and $\angle P_2P_3P_4$ are equal and the same holds for the angles at P_2 and P_4. This follows from the Quadrilateral Theorem. In order to show the polygon $P_1P_2P_3P_4$ is in fact an orbit, we need to show that the billiard law holds at the impact points P_i w.r.t. the ellipse $\mathcal{E}$. Since AB is a diameter of c and by symmetry of the ellipse $\mathcal{E}$, the points P_4 and P_2 lie symmetric w.r.t. AB. This implies that $\angle P_2BP_4$ is split into to equal parts by AB. By another application of the quadrilateral theorem, $\angle P_2P_1A = \angle P_2BA$ and also $\angle AP_1P_4 = \angle ABP_4$. Since n_1 is perpendicular to t_1 at P_1, this shows the billiard law holds at P_1. By symmetry the same holds for P_3 and a similar argument yields the equal angle law at P_2 and P_4 as well. This shows that $P_1P_2P_3P_4$ is an fact an orbit and that it is symmetric. In fact, the same holds now for any orbit that has four distinct reflection points on $\mathcal{E}$, since we can continuously change the position of the circle while maintaining four intersection points. Since any family of NB orbit in an ellipse is tangent to a unique confocal hyperbola $\mathcal{H}$, Theorem 10, this whole family has $\mathcal{H}$ as its caustic.

Case 2 deals with the extreme cases that c intersects only three times with $\mathcal{E}$. We claim that the resulting V-shaped configuration is an orbit. Note that A now lies on both c and $\mathcal{E}$. Observe that n_1 is normal to the tangent at P_1 and intersect the extended minor axis at A. Hence if we reflect segment AP_1 in the minor axis, we obtain indeed an orbit. In Figure 58 the orbit is shown in blue.

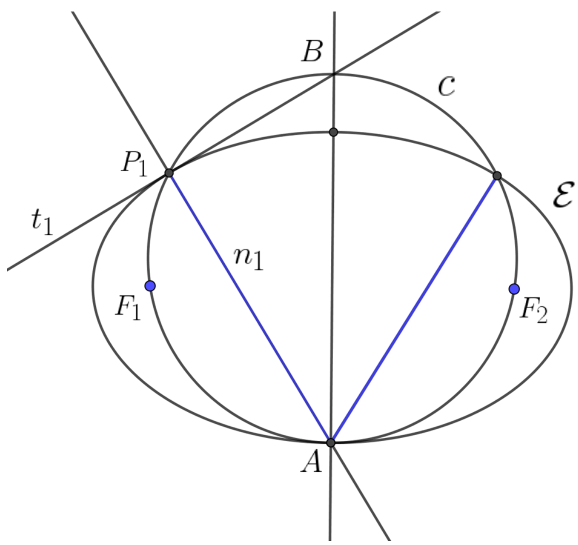

Figure 58: Proof for degenerate $(4, 2)$ orbits.

Fix now some ellipse $\mathcal{E} \in \mathbb{E}_{\sqrt{2}}$. Combining case 1 and case 2, we see that for each choice of a circle c, one obtains a $(4, 2)$ orbit by the constructions discussed above. This completes the proof of the existence of 1-parameter of symmetrical $(4, 2)$ orbits. $\square$

Remark 7. *In the phase diagram of an elliptical billiard map (Fig. 60) the $(2n, n)$ NB orbits are visible as the homotopically trivial circles around the two "eyes" (the (2,1) orbits along the minor axis) that correspond to the minor* *axis of the ellipse. The explanation is that the trivial circles are not graphs of functions and therefore necessarily correspond to non-Birkhoff orbits, see the discussion below Proposition 4. Hence the collection contractible circles must contain NB orbits.*

Lemma 8.8. *Assume $\mathcal{O}$ is a (p,q)-periodic orbit with $q \leq \frac{p}{2}$. Denote by $s : M \to M$, $s = s(x,v)$ the reflection in the inward pointing normal at x. Then it holds that*

$$T \circ s = s \circ T^{p-1}.$$

Also it holds that if p is even, then

$$T^{\frac{p}{2}} \circ s = s \circ T^{\frac{p}{2}}. \tag{16}$$

Proof. The first assertion follows easily from Figure 59.
For the second one we argue as follows. Let $(x,v) \in M$, where M is the phase

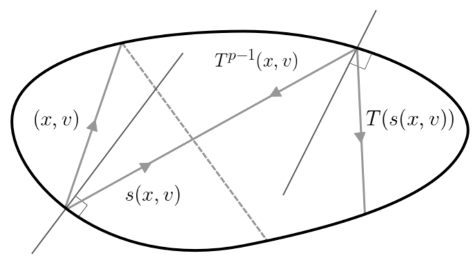

Figure 59: Part of a (p,q) periodic orbit. This figure shows that $T^{\frac{p}{2}} \circ s = s \circ T^{\frac{p}{2}}$.

space. Repeatedly using the first identity, we get

$$\begin{aligned} T^{\frac{p}{2}}(s(x,v)) &= T^{\frac{p}{2}-1}(T(s(x,v))) \\ &= T^{\frac{p}{2}-2}TsT^{p-1}(x,v) \\ &= T^{\frac{p}{2}-2}sT^{2p-2}(x,v) \\ &= \dots \\ &= T^{\frac{p}{2}-\frac{p}{2}}sT^{\frac{p}{2}p-\frac{p}{2}}(x,v) \\ &= sT^{\frac{p^2-p}{2}}(x,v) \\ &= sT^{\frac{p}{2}}(x,v), \end{aligned} \tag{17}$$

where we used that for p even we have that $\frac{p^2-p}{2} = \frac{p}{2} \mod p$. This concludes the proof of the lemma. $\square$

It is well known that the elliptical billiard map T is *integrable* in the sense of Liouville. This means that there exists a smooth function on the phase space that is invariant under map T. The phase space of the billiard in an ellipse is foliated by invariant curves. These T-invariant curves correspond to caustics inside the ellipse. The curves that make a turn around the phase space correspond to elliptical caustics and the contractible curves inside the ∞-shaped curve correspond to the hyperbolic caustics. The area preserving property of the map T, see Proposition 1, implies that one can choose coordinates on the invariant curves such that T becomes a translation $x \mapsto x + c$ on the leaves. The c in this representation depends on the invariant curve. We state this last fact as a lemma and refer to [4] for a proof.

Lemma 8.9. *Let $\mathcal{E}$ be an ellipse and let T be the induced billiard map. Then there are coordinates such that on an invariant curve $\Gamma \subseteq \mathbb{S}^1 \times (0, \pi)$, T restricted to Γ becomes*

$$T|_\Gamma : x \mapsto x + c.$$

Proposition 21. *Non-Birkhoff orbits in ellipses are symmetric w.r.t. reflection in the major axis or minor axis or w.r.t. rotation.*

Proof. There are two main steps in the proof. Step 1 concerns proving the symmetry property of NB orbits that have a vertex P at an intersection of the ellipse $\mathcal{E}$ and the hyperbola $\mathcal{H}$ that touches a given $(2n, n)$ NB orbit. Since the ellipse and hyperbola are confocal, they intersect perpendicular. Hence the direction v_0 at P makes an angle of $\frac{\pi}{2}$ with $\mathcal{E}$. Consider the phase portrait of the billiard map for some chosen a and b. The curves shown in the diagram are level curves of the function $\lambda(s, r) = (a-b)\sin^2(2\pi s)+b-r^2$, where $s \in \mathbb{R}/\mathbb{Z}$ and $r = -\cos(\theta)$. The curves away from the homotopically trivial circles, are graphs of functions on $\mathbb{R}/\mathbb{Z}$. Birkhoff orbits may lie on these rotationally invariant curves. We remark that by symmetry of the ellipse, all level curves are symmetric w.r.t. the line $s = \frac{1}{2}$, while the homotopically trivial curves are additionally symmetric w.r.t. the line $r = 0$. The initial point under consideration is located on some trivial loop at its left most point, denote it by $A \in \mathbb{R}/\mathbb{Z} \times (-1, 1)$. Let S denote the s from Lemma 8.8. That is, the reflection w.r.t. the line $r = 0$, where we now interpret S as map of position and angle rather than position and vector. Then $S(A) = A$ and therefore $T^{\frac{p}{2}}(S(A)) = T^{\frac{p}{2}}(A) = S(T^{\frac{p}{2}}(A))$ by Lemma 8.8. This implies that, considering now $A \in \Gamma$, where $\Gamma \cong S^1$, $T^{\frac{p}{2}}|_\Gamma(A) = A + \pi$, possibly after projecting each trivial loop to the leftmost one. Note that in the $(4, 2)$ case $A + \pi = C$ is already on the leftmost loop.

Now we invoke Lemma 8.9 that says that, on an invariant curve, the billiard map inside an ellipse has the form $T : x \mapsto x+c$ for a suitably chosen coordinate x. If we can show that the c in this expression equals $\frac{2\pi}{p}$ we are done, since then all impact points of the NB orbit lie symmetrically around the corresponding invariant curve. This then implies that also in the billiard table the impact points possess at least one type of symmetry. The argument for determining the c in the expression of T is in fact quite easy. Since $T(x) = x + c$, we obtain that $T^{\frac{p}{2}}(x) = x + \frac{p}{2}c$. But we already showed that $T^{\frac{p}{2}}(x) = x + \pi$. Hence $c = \frac{2\pi}{p}$. From this fact, we may conclude that in the phase space the reflection are symmetrically ordered as a rotation on some trivial loop. But this implies than that the orbit is also symmetric in the billiard table itself. This completes the argument.

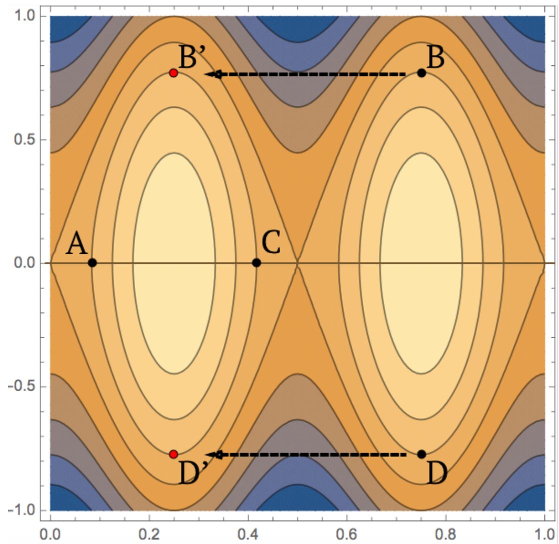

Figure 60: Phase portrait of an elliptical billiard map T with coordinates (s, r). The orbit $ABCD$ represents a $(4, 2)$ NB orbit. Projecting B and D to the leftmost loop Γ on which A and C lie, shows that on Γ all points can be reached by A by a rotation over $\frac{2\pi}{4} = \frac{\pi}{2}$ radians. This implies the symmetry of this particular $(4, 2)$ orbit.

$\square$

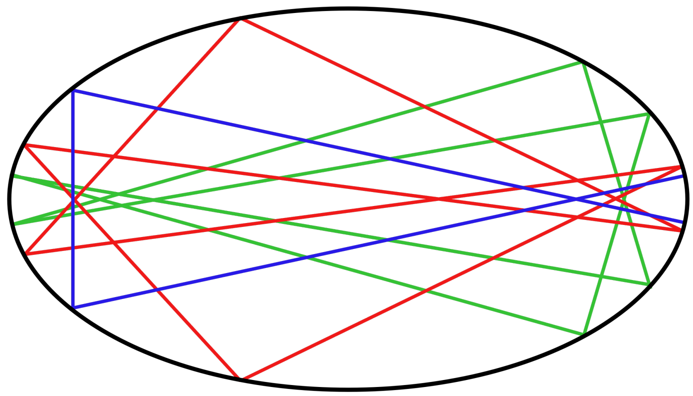

Figure 61: Symmetric $(6,3)$ orbits from a 1 parameter family $\mathcal{F}$. The blue orbit corresponds to a boundary element of $\mathcal{F}$.

We note that Poncelet's theorem implies that Birkhoff periodic orbit are not necessarily symmetric. Proposition 21 yields the following characterization of periodic orbits in billiards: either a periodic orbit is a regular polygon or a (possibly unordered) symmetric orbit. In this sense there is limited geometric shape a periodic orbit inside an ellipse can possess.

9 Non-Birkhoff orbits in non-symmetrical billiards

In this section we will apply the Conley theory developed in section 6.1 to prove existence of NB-orbits in billiards "close to" elliptical billiards. This will be made precise in the discussion below.

Recall that the Morse index of a 1-parameter family of periodic orbits in a billiard system, is defined to be the number of positive eigenvalues of $-\nabla^2 W_{p,q}(x(s))$ for some $s \in I$.

The next theorem is a perturbation result about persistence of NB orbits in a not necessarily symmetric neighborhood of an ellipse. It is well known consequence of KAM theory that resonant invariant curves generally do not survive perturbations, see [14]. However, Theorem 9 implies that although not all $(2p,p)$ NB orbits on a resonant invariant curve persists, there will remain some $(2p,p)$ NB orbits when the elliptical billiard map T is perturbed.
Let $B^3_{\mathcal{C}}$ denote the family of all C^3 billiard curves and recall $\mathbb{E}_{\sqrt{2}}$ denotes the

class of ellipses such that the ratio of the major and minor axis is greater than $\sqrt{2}$. We will only present ideas how to prove this theorem, it not yet completely rigorous.

Conjecture 2. *Let $\mathcal{E} \in \mathbb{E}_{\sqrt{2}}$. Then all billiard tables in a sufficiently small neighborhood of $\mathcal{E}$ in $B^3_{\mathcal{C}}$, posses a NB orbit of type $(4,2)$. In particular, all non-symmetrical billiard tables in a sufficiently small neighborhood of $\mathcal{E}$ have a NB orbit of type $(4,2)$.*

Idea of proof. We will apply Conley Theory to calculate the Conley index of an appropriate isolating invariant set which contains $(4,2)$ NB orbits. Recall that the Conley index is an index of isolating invariant sets. So in order to apply Conley theory to this context, we need to come up with an isolating invariant set and some isolating neighborhood of this set. Let A_1, A_2, A_3 and A_4 denote the intersection points of $\mathcal{H}$ and $\mathcal{E}$ and t_1, t_2, t_3 and t_4 the times such that $\mathcal{E}(t_i) = A_i$. Denote by $[a_1, a_2]$ the (shortest) arc of $\mathcal{E}$ from a_1 to a_2, so that $[a_1, a_2] = \mathcal{E}([t_1, t_2])$. Parametrize $x \in \mathbb{X}_{4,2}$ by (x_0, x_1, x_2, x_3). Define

$$O = \{x \in \mathbb{X}_{4,2} : x_0 \in [a_1, a_2],\ x \pitchfork \mathcal{O}_m,\ I(x, \mathcal{O}_m) = 2\},$$

where $\mathcal{O}_m \in \mathbb{X}_{2,1}$ is the $(4,2)$-orbit along the minor axis of the ellipse $\mathcal{E}$.

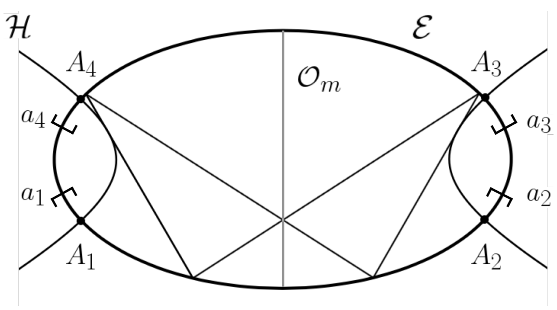

Figure 62: Illustration of the setting and a $(4,2)$ NB orbit in O.

Observe that if there are critical points in O, they are indeed of non-Birkhoff type. This is because the only Birkhoff orbits $x \in \mathbb{X}_{2p,p}$ are in $\mathbb{X}_{2,1}$ and they are along the major or minor axis. So the intersection index $I(x, \mathcal{O}_m) = 0$ in this case, which implies that these orbits cannot lie in O. Denote by φ^t the gradient flow of $\dot{x} = -\nabla W_{4,2}(x)$. Let

$$M = \{x \in O : \overline{\gamma(x)} \subseteq O\},$$

where $\gamma(x) = \{\varphi^t(x) : -\infty < t < \infty\}$. Recall that $\mathrm{Crit}(W_{4,2})$ denotes the set of critical points of $W_{p,q}$ and let $I = [t_1, t_2]$. We will show that the set M equals

$$M = \{x(s) :\ x(s) \in \mathrm{Crit}(W_{4,2}) \cap O,\ s \in I\},$$

and that it is an isolated invariant set for this system. In fact, M is precisely the 1-parameter family $\mathcal{F}$ of all $(4, 2)$ NB orbits in $\mathcal{E}$, which exists by Poncelet's theorem. Note that O cannot be a isolating neighborhood of M since it is not closed. We will first state a lemma and refer to [5] for a proof.

Lemma 9.1. *The following statements are equivalent.*

1. *M is an isolated invariant set.*

2. *M is compact.*

3. *For each $y \in \partial O$, there is an open neighborhood U_y such that $M \cap U = \emptyset$.*

Observe that $\mathcal{O}_m \in \partial O$. By Lemma 9.1, we have to show that for each $y \in \partial O$ there exists a neighborhood U_y of y such that $U \cap M = \emptyset$. We distinguish two cases, whether $y = \mathcal{O}_m$ or $y \neq \mathcal{O}_m$. Recall that $|\cdot|$ denotes the length of configuration. Assume first $y = \mathcal{O}_m$. Consider some $z \in \mathbb{X}_{4,2}$ close to $\mathcal{O}_m$. Close to $\mathcal{O}_m$ there cannot be NB orbits because all NB orbits are tangent to $\mathcal{H}$, hence they are far apart from $\mathcal{O}_m$. Now the ω-limit sets $\omega_{+,-}(z)$ cannot both be NB orbits and consequently, one of them is Birkhoff (no degenerate configuration possible, cf. the proof of Theorem 9). This follows because all NB orbits have the same length, and the gradient flow increases the length of a configuration in forward time and decreases the length in backward time, hence $|\omega_-(z)| < |\omega_+(z)|$. This implies that the intersection index of $\omega_+(z)$ or $\omega_-(z)$ with $\mathcal{O}_m$ is zero. In particular, $\overline{\gamma(z)}$ is *not* contained in O. Hence $z \notin M$ and therefore there exists a $U \ni \mathcal{O}_m$ such that $U \cap M = \emptyset$.
For the case $y \neq \mathcal{O}_m$, we argue as follows. Since $y \in \partial O$, $y \notin \mathcal{F}$, that is, y is not a NB orbit. In particular, at most one of the critical points $\omega_{+,-}(y)$ is non-Birkhoff. This follows because all NB orbits have the same length, and $|\omega_-(y)| < |\omega_+(y)|$. By Lemma 9.1 we have that M is an isolated invariant set.

We now show that M is a hyperbolic set, after which we can compute the Conley index of M. Let $x(s)$ for $s \in [t_1, t_2] =: I$ denote an element of the 1-parameter family $\mathcal{F}$ of $(4, 2)$-orbits. Observe that $-\nabla W(x(s)) = 0$ implies that

$$\nabla^2 W(x(s)) \cdot x'(s) = 0,$$

for all $s \in I$. We need to show that the null space of $-\nabla^2 W(x(s))$ is 1-dimensional, i.e. that $\nabla^2 W(x(s))$ has precisely 1 eigenvalue equal to zero. For this, we will make use of the symmetry of the periodic orbits in M. In the case that $x(s) \in \mathbb{X}_{4,2}$, it holds that $S(x_1, x_2) = S(x_3, x_4)$ and $S(x_2, x_3) = S(x_4, x_1)$. Hence $\alpha_1 = \alpha_3$, $\alpha_2 = \alpha_4$ and $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$. Thus $-\nabla^2 W(x(s))$ assumes the following form:

$$-\nabla^2 W_{p,q}(x) = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \beta_2 \\ \beta_1 & \alpha_2 & \beta_2 & 0 \\ 0 & \beta_2 & \alpha_1 & \beta_1 \\ \beta_2 & 0 & \beta_1 & \alpha_2 \end{bmatrix}. \qquad (18)$$

To prove that $\lambda = 0$ is simple, we look at $-\nabla^2 W(x(s))v = 0$, where $v \in \mathbb{R}^4$.

Recall that $L = \{x \in \mathbb{X}_{4,2} : \mathcal{S}(x) = \sigma^2(x)\}$. The tangent space to L is two dimensional and it consists of vectors $v = (a, b, -a, -b)$. We make a change of basis and write $v = a\mathbf{f}_1 + b\mathbf{f}_2$, where $\mathbf{f}_1 = (1, 0, -1, 0)^T$ and $\mathbf{f}_2 = (0, 1, 0, -1)^T$. Applying $-\nabla^2 W_{4,2}(x(s))$ to the basisvectors $\mathbf{f}_1$ and $\mathbf{f}_2$, gives

$$-\nabla^2 W_{4,2}(x(s))\mathbf{f}_1 = \begin{bmatrix} \alpha_1 \\ \beta_1 - \beta_2 \\ -\alpha_1 \\ \beta_2 - \beta_1 \end{bmatrix}, \quad -\nabla^2 W_{4,2}(x(s))\mathbf{f}_2 = \begin{bmatrix} \beta_1 - \beta_2 \\ \alpha_2 \\ \beta_2 - \beta_1 \\ -\alpha_2 \end{bmatrix}.$$

Rewriting this in terms of the basisvectors yields the following form of the matrix w.r.t. $\mathcal{F} := \{\mathbf{f}_1, \mathbf{f}_2\}$:

$$\big[-\nabla^2 W_{4,2}(x(s))\big]_{\mathcal{F}} = \begin{bmatrix} \alpha_1 & \beta_1 - \beta_2 \\ \beta_1 - \beta_2 & \alpha_2 \end{bmatrix}. \tag{19}$$

Now, we know that $v = a\mathbf{f}_1 + b\mathbf{f}_2 \in \ker\big(\big[-\nabla^2 W_{p,q}(x(s))\big]_{\mathcal{F}}\big)$. So there is an eigenvalue zero. To show that the matrix (20) has precisely one eigenvalue zero, we need to prove that at least one of its entries is non-zero. This follows since if there were two eigenvalues zero, then the matrix itself would be the zero matrix because $-\nabla^2 W_{4,2}(x(s))$ is symmetric and hence the claim follows by diagonalizing. We will show that $\beta_1 - \beta_2 \neq 0$. The argument is split according to whether the orbit is V-shaped (the boundary case) or not.

Case 1 (*generic orbits*)

Proposition 2 tells that

$$\beta_i = -\partial_{1,2} S(x_{i-1}, x_i) = -\frac{\sin(\theta_i)\sin(\theta_{i+1})}{S},$$

where S is the generating function of the billiard map. It is clear by looking at Figure 63 that $\sin(\theta_1)\sin(\theta_2) = \sin(\theta_2)\sin(\theta_3)$ since $\theta_1 = \theta_3$, while if $x_1 \neq x_3$, $S(x_1, x_2) \neq S(x_2, x_3)$. Hence $\beta_1 \neq \beta_2$. This shows that $-\nabla^2 W_{4,2}(x(s))$ is not the zero matrix and consequently that the multiplicity of the zero eigenvalue is one.

Case 2 (*boundary orbits*)

In the boundary case, so when the orbit is V-shaped, the β_i in fact are equal. Hence $\beta_1 - \beta_2 = 0$ and therefore the off-diagonal elements of $-\nabla^2 W_{4,2}(x(s))$ are zero. We see in Figure 64 that indeed $S(x_1, x_2) = S(x_2, x_3) = S(x_3, x_4) = S(x_1, x_4)$ and also $\theta_1 = \theta_3$ and $\theta_2 = \theta_4 = \frac{\pi}{2}$.

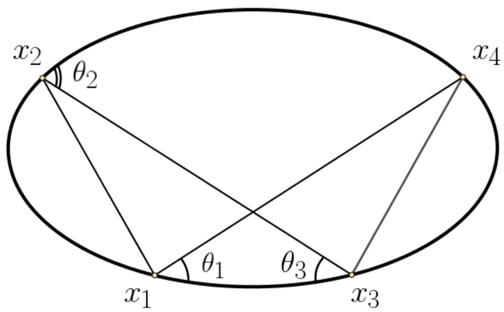

Figure 63: Proving that β_i is non-zero for a generic $(4,2)$ orbit.

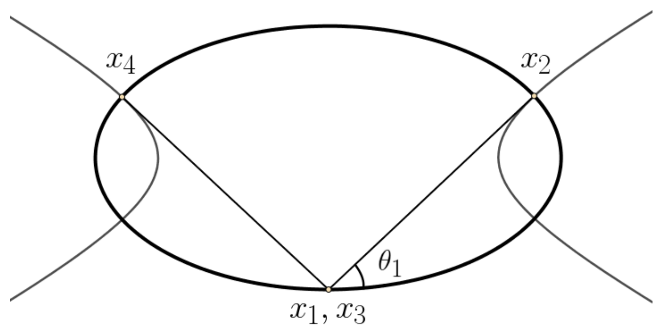

Figure 64: Proof that $\alpha_{1,2}(x)$ are not both zero.

So we need to show that some $\alpha_i \neq 0$ in this case. In the beginning of section 8 we stated that

$$\alpha_2(x) = \sin^2(\theta_2)\left(\frac{1}{S(x_1,x_2)} + \frac{1}{S(x_2,x_3)}\right) - 2\kappa(x_2)\sin(\theta_2).$$

In this case we have that $\sin(\theta_2) = 1$ since $\theta_2 = \frac{\pi}{2}$ and $l := S(x_1,x_2) = S(x_2,x_3)$. Hence

$$\alpha_2(x) = \frac{2}{l} - 2\kappa(x_2).$$

By considering the geometry of the orbit, one easily obtains for α_1 the formula

$$\alpha_1(x) = \sin(\phi_1)\frac{2}{l} - 2\kappa(x_1),$$

see Figure 64. At x_1, the curvature equals $\kappa(x_1) = \frac{b}{a^2}$, since x_1 lies on the minor axis. We will calculate l first. For x and y we have that

$$x = a\sqrt{1 - \frac{b^4}{(b^2 - a^2)^2}},\ y = \frac{b^3}{a^2 - b^2},$$

see the Appendix. Pythagoras gives $l^2 = x^2 + (b + y)^2$, and we get after some work that

$$l^2 = \frac{a^4}{a^2 - b^2}.$$

Then sine of θ_1 is given by

$$\sin(\theta_1) = \frac{ba^2}{l(a^2 - b^2)} = \frac{b}{\sqrt{a^2 - b^2}}.$$

For α_1 at x we have

$$\alpha_1(x) = \frac{2a^2}{l^2(a^2 - b^2)} - \frac{2b}{a^2} = \frac{2 - 2b}{a^2}$$

This is zero if and only if $b = 1$. We will now calculate $\alpha_2(x)$ under the assumption $b = 1$. If $\alpha_2(x) \neq 0$ for each $\mathcal{E} \in \mathbb{E}_{\sqrt{2}}$ with $b = 1$, then we may conclude that the matrix in (17) has at least one non-zero entry.

$$\alpha_2(x) = \frac{2\sqrt{a^2 - 1}}{a^2} - \frac{2(a^2 - 1)^{\frac{3}{2}}}{a^2}.$$

We observe that $\alpha_2(x) \neq 0$ if and only if $a = \sqrt{2}$ (or $a = -\sqrt{2}$). This implies that the only ellipse for which the V-shaped orbit x has $\alpha_1(x) = \alpha_2(x) = 0$, is the ellipse with $\frac{a}{b} = \sqrt{2}$ which is excluded from the set $\mathbb{E}_{\sqrt{2}}$. This now implies that the matrix $-\nabla^2 W_{4,2}(x(s))$ has precisely one eigenvalue $\lambda = 0$, for each $s \in I$. We therefore obtain that M is an hyperbolic invariant set of dimension M one. The space M is diffeomorphic to an embedded line segment. Just note that there are no crossings, since assuming that $x(s) = x(s')$ for some $s, s' \in I$ implies that $s = s'$. This follows simply because each $x(s)$ is distinguished $(4, 2)$ orbit (where we consider orientation as a distinguishing property). We are now in a position to calculate the Conley index of M. We suggest a way to do this based on [5].

The product of pointed topological spaces (X, x_0) and (Y, y_0), denoted by $(X, x_0) \wedge (Y, y_0)$, is given by

$$X \times Y / X \times \{y_0\} \cup \{x_0\} \times Y.$$

The Conley Index of M is the Conley index of an appropiate product of isolated invariant sets, which we describe below. The Conley index of a product is simply the product of Conley indices of the separate isolated invariant sets. So we first need to calculate the Conley index of both separate systems. By the unstable manifold theorem, the flow near M is the product of that near a hyperbolic fixed point of index k and the trivial flow on a line segment, cf. [5]. We will first look at the trivial flow along a line segment.

Consider the space $X = [a, b]$ and consider the trivial flow φ_0^t on it. Since X is compact, one easily sees that

$$\operatorname{Inv}(X, \varphi_0^t) := \{x \in X : \varphi_0(\mathbb{R}, x) \subseteq X\} = X \subseteq int(X).$$

The exit set is given by $L = \emptyset$. Let $S = \operatorname{Inv}(X, \varphi_0^t)$. Then

$$h(S) \sim (X, [\emptyset]) \sim (\{pt\}, [\emptyset]),$$

where $pt \in X$. The Conley index for a hyperbolic fixed point x_0 of index k was already computed in example 6. We found that

$$h(\{x_0\}) = (S^{2k-1}, *).$$

Hence by the product rule for the Conley index, see section 7, we obtain

$$\begin{aligned} h(M) &\sim h(X \times \{x_0\}) \\ &\sim (\{pt\}, [\emptyset]) \wedge (S^{2k-1}, \{*\}) \\ &\sim (\{pt\} \times S^{2k-1} / \{pt\} \times \{*\} \cup [\emptyset] \times \{pt\}) \\ &\sim (S^{2k-1}, *). \end{aligned}$$

In particular, the Conley index is non-zero, implying that $\operatorname{Inv}(N)$ is non-empty, see section 7. Since we are considering gradient flows, we conclude there is a fixed point in M by Lemma 7.1. By the perturbation invariance of the Conley Index, Theorem 7, this implies that for a small perturbation of the flow such that we still obtain a flow in a convex billiard table, N remains to be an isolating neighborhood. Therefore the Conley index remains constant and in particular non-zero. Hence there are perturbed systems in which we can still find a $(4, 2)$-NB orbit. This proves the claim that there exist (not necessarily symmetric) perturbed elliptical billiard tables in which NB orbits of type $(4, 2)$ are present. $\square$

In the appendix we collect elementary calculations for some results in the proof of Conjecture 2.

A 1

Given an ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ with $\frac{a}{b} > \sqrt{2}$, we determine the coordinates of the vertices of the (unique) V-shaped orbit. Since an orbit is a critical point of the length function, we only need to solve $F'(x) = 0$, where $F(x)$ is the square of the length AX:

$$F(x) = (b + b\sqrt{1 - \frac{x^2}{a^2}})^2 + x^2.$$

Performing the differentiation and solving $F(x) = 0$, we obtain $x = 0$ or

$$\frac{1 + \sqrt{1 - \frac{x^2}{a^2}}}{\sqrt{1 - \frac{x^2}{a^2}}} = \frac{a^2}{b^2}.$$

The solution $x = 0$ corresponds to the $(2, 1)$ orbit along the minor axis, so we can discard it. Solving the second equation for yields

$$\sqrt{1 - \frac{x^2}{a^2}} = \frac{b^2}{a^2 - b^2}.$$

Letting x_0 be a solution (there exists one if $\frac{a}{b} > \sqrt{2}$), we find that

$$y_0 = b\sqrt{1 - \frac{x_0^2}{a^2}} = \frac{b^3}{a^2 - b^2}.$$

For x_0 we find that

$$x_0 = a\sqrt{1 - \frac{y_0^2}{b^2}} = a\sqrt{1 - \frac{b^4}{(a^2 - b^2)^2}}.$$

B 2

The proof of the formula for l^2 is given below.

$l^2 = x_0^2 + (b + y_0)^2$. Plugging in the solutions found in A1, we get

$$l^2 = a^2(1 - \frac{b^4}{(b^2 - a^2)^2}) + b^2 + \frac{2b^4}{a^2 - b^2} + \frac{b^6}{(a^2 - b^2)^2}.$$

Rewriting this expression gives

$$\begin{aligned}
l^2 &= \frac{a^2(a^4 - 2a^2b^2 + b^4) - a^2b^4}{(a^2-b^2)^2} + \frac{b^2(a^2-b^2)^2 + 2b^4(a^2-b^2) + b^6}{(a^2-b^2)^2} \\
&= \frac{a^6 - 2a^4b^2 + a^2b^4 - a^2b^4 + a^4b^2 - 2a^2b^4 + b^6 + 2a^2b^4 - 2b^6 + b^6}{(a^2-b^2)^2} \\
&= \frac{a^6 - a^4b^2}{(a^2-b^2)^2} \\
&= \frac{a^4(a^2-b^2)}{(a^2-b^2)^2} \\
&= \frac{a^4}{a^2-b^2}.
\end{aligned}$$

C 3

The computation of the curvature at x_2 goes as follows. Parametrize the ellipse as $(x(t), y(t)) = (a\cos(t), b\sin(t))$, $t \in [0, 2\pi]$. Recall that the formula for the curvature at some $t \in [0, 2\pi]$ is given by

$$\kappa(t) = \frac{ab}{(a^2\sin^2(t) + b^2\cos^2(t))^{\frac{3}{2}}}. \tag{20}$$

Evaluating (21) at $(x_0, y_0) = (a\sqrt{1 - \frac{b^4}{(a^2-b^2)^2}}, \frac{b^3}{a^2-b^2})$ results in

$$\begin{aligned}
\kappa(x_2) &= \frac{ab}{\left(\frac{b^2}{a^2}\cdot a^2(1 - \frac{b^4}{(a^2-b^2)^2}) + \frac{a^2}{b^2}\cdot\frac{b^6}{(a^2-b^2)^2}\right)^{\frac{3}{2}}} \\
&= \frac{ab}{b^3\left(\frac{a^4-2a^2b^2+b^4-b^4+a^2b^2}{(a^2-b^2)^2}\right)^{\frac{3}{2}}} \\
&= \frac{a}{b^2\left(\frac{a^4-a^2b^2}{(a^2-b^2)^2}\right)^{\frac{3}{2}}} \\
&= \frac{a}{b^2\left(\frac{a^2}{a^2-b^2}\right)^{\frac{3}{2}}} \\
&= \frac{(a^2-b^2)^{\frac{3}{2}}}{a^2b^2}
\end{aligned}$$

References

[1] Christophe Gole, *Symplectic Twist Maps*, World Scientific 1st edition, 2001.

[2] Blaz Mramor and Bob Rink, *Ghost circles in lattice Aubry-Mather theory*, 2011.

[3] Boris Hasselblat and Anatole Katok, *A first course in dynamics*, Cambridge University Press 1st edition, 2003.

[4] Serge Tabachnikov, *Geometry and Billiards*, American Mathematical Society 1st edition, 2001.

[5] Sigurd Angenent, *The periodic orbits of an area preserving twist map*, Communications in Mathematical Physics, 1988.

[6] Sonia Pinto-de-Carvalho and Rafael Ramirez-Ros, *Billiards with a given number of (k,n)-orbits*, American Institute of Physics, 2012.

[7] Ludwig Arnold, Christopher K.R.T Jones, Konstantin Mischaikow, Genevieve Raugel, *Dynamical systems*, Springer, 1994.

[8] Vladimir Dragovic, Milena Radnovic, *Poncelet Porisms and Beyond*, Birkhäuser, 2010.

[9] Philip L.Boyland, Glen Richard Hall, *Invariant circles and the order structure of periodic orbits in monotone twist maps*, Elsevier, Volume 26, Issue 1, 1987.

[10] Marek Ryszard Rychlik *Periodic points of the billiard ball map in a convex domain*, J. Differential Geom. Volume 30, Number 1 (1989), 191-205

[11] Charles Conley, *Isolated invariant sets and the Morse index*, CBMS Regional Conference Series in Mathematics, 38. American Mathematical Society, Providence, R.I., 1978

[12] Oliver Knill *On Nonconvex Caustics of Convex Billiards* Elemente der Mathematik, Birkhäuser Verlag, Basel, 1998

[13] Pablo S. Casas and Rafael Ramírez-Ros *The frequency map for billiards inside ellipsoids*, $https://www.ma.utexas.edu/mp_arc/c/10/10-68$.*pdf*

[14] Sonia Pinto-de-Carvalho and Rafael Ramírez-Ros *Nonpersistence of resonant caustics in perturbed elliptic billiards*, $https://www.ma.utexas.edu/mp_arc/c/11/11-118$.*pdf*