

# The oval billiard table and periodic orbits

## The mathematics behind the "Simulator" computer program

Created 1.10.2025

*This document analyses mathematical billiards on convex billiard tables. Interesting statements can be derived elementarily, especially about periodic orbits. The mathematical investigations are supported by experiments with the computer program "Simulator".*

*The document is intended as a stimulus for extended maths lessons at intermediate level, whether for courses outside the compulsory curriculum or for individual work by interested pupils.*

*The entire series of topics related to the "Simulator" includes:*

- *The chaotic properties of logistic growth*
- ***The oval billiard table and periodic orbits***
- *Newton iteration and the complex roots of unity*
- *Iteration of quadratic functions in the complex plane*
- *Numerical methods and coupled pendulums*
- *Planetary motion and the three-body problem*
- *Strange attractors and the weather forecast of Edward Lorenz*
- *Fractal sets and Lindenmayer systems*
- *The history of chaos theory*
- *Programming your own dynamic systems in the "simulator"*

*Each topic is dealt with in a separate document.*

*The computer program "Simulator" enables the simulation of simple dynamic systems and experimentation with them. The code is publicly available on GitHub, as is a Microsoft Installer version. The corresponding link is: <https://github.com/HermannBiner/Simulator>. The following documentation is integrated into the "Simulator" in German and English:*

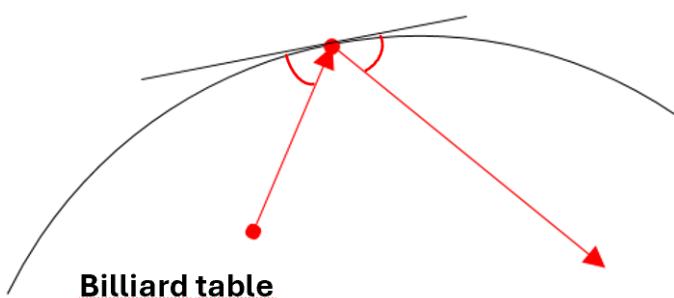
- *Mathematical documentation with examples and exercises*
- *Technical documentation with a detailed description of the functionality*
- *User manual with examples*
- *Version history*

## Contents

1.	The mathematical billiard table .....	2
2.	Billiards in a circle .....	7
3.	Elliptical billiards .....	10
4.	Caustics in elliptical billiards .....	17
5.	Periodic points in elliptical billiards .....	20
6.	The phase portrait .....	23
7.	Oval billiard .....	27
8.	The C-diagram .....	29
9.	Billiards in a stadium .....	33
10.	The convex billiard table .....	35
11.	Exercise examples .....	38
	Further reading .....	40

## 1. The mathematical billiard table

In this paper, we consider a convex region of the plane to be a billiard table. We will define what we mean by this later. On the billiard table, an idealised (massless and point-shaped) billiard ball moves frictionlessly and in a straight line until it hits the edge of the billiard table. Here it is reflected according to the law of reflection: angle of projection = angle of incidence. The angle of incidence is defined as the angle between the ball path before the collision and the tangent to the edge curve at the collision point. Similarly, the angle of projection is the angle between this tangent at the point of collision and the trajectory of the ball after the collision.



Reflection of the billiard ball (red) at the edge of the billiard table

We will analyse the elliptical, oval and stadium-shaped billiard table. These table shapes are implemented in the simulator, and any number of balls can be launched simultaneously.

Now we define the billiard table more precisely, namely as a finite area of the plane, which is bounded by a simply closed, regular and continuously differentiable curve.

### Definition 1.1

By a *plane curve* we mean a continuous mapping:

$$\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$$

□

Note:  $\gamma$  denotes the mapping and  $\vec{\gamma}(t) \in \mathbb{R}^2$  the image of  $t$  under the mapping  $\gamma$ . The vector notation for the image is usually omitted. However, because we are building on an elementary mathematical level, we prefer the vector notation because we also need the scalar product, which is defined for vectors.

The cushion of the billiard table is a plane curve as defined above. In mathematical billiards, a billiard ball must be reflected at the cushion of the billiard table. This means that the tangent should be defined for each edge point. The tangent to a curve at a point  $t$  is defined by the derivative  $\dot{\gamma}(t)$ , if it exists (we write a point for the derivative in this paper). The derivative of a curve at a point  $t \in I$  is defined analogously to the one-dimensional case of real functions:

### Definition 1.2

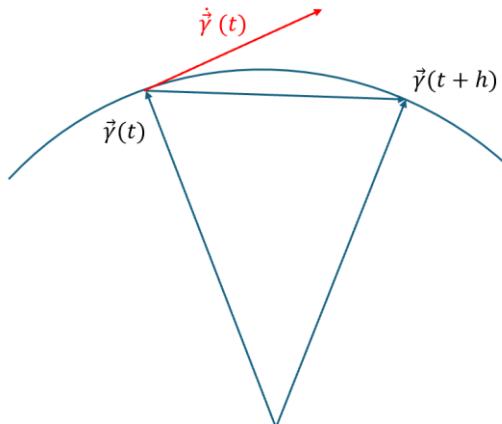
The derivative of a curve  $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$  at the point  $t$  is defined by:

$$\dot{\gamma}(t) := \lim_{h \rightarrow 0} \frac{\vec{\gamma}(t+h) - \vec{\gamma}(t)}{h}$$

If this limit exists.

□

The derivative is the *tangent* to the curve  $\gamma$  at the point  $t$ .



The derivative of a curve at the point  $t$

For the billiard table, we will require that  $\dot{\gamma}(t)$  is *continuous*. Next, we want the plane curve bounding the billiard table to be closed, so that the billiard table represents a finite region of the plane. This leads to the next definition:

### Definition 1.3

A curve  $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$  is *closed* if  $I = [a, b]$  is a finite real interval with  $\vec{\gamma}(a) = \vec{\gamma}(b)$ . If the curve is differentiable, the following should also apply:  $\dot{\gamma}(a) = \dot{\gamma}(b)$ .

The curve is *simply closed* if it contains no double points. This means that the mapping  $\gamma$  on  $[a, b[$  is *injective*.

□

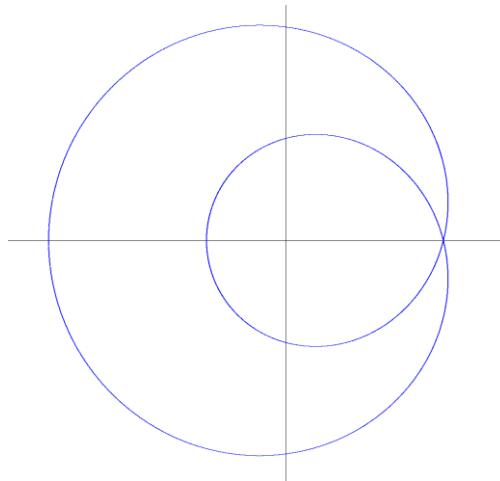
### Examples

The parabola  $\gamma: t \in \mathbb{R} \rightarrow \vec{\gamma}(t) = (t, t^2)$  is a plane curve but not closed.

The circle  $\gamma: t \in \mathbb{R} \rightarrow \vec{\gamma}(t) = (\cos t, \sin t)$  is a closed plane curve, but not simply closed because  $(\cos(t + 2\pi k), \sin(t + 2\pi k)) = (\cos t, \sin t)$ . Therefore,  $\gamma$  is not injective.

On the other hand, the circle parameterised by  $\gamma: t \in [0, 2\pi[ \rightarrow \vec{\gamma}(t) = (\cos t, \sin t)$  is simply closed.

The curve  $\gamma: t \in [0, 2\pi[ \rightarrow \vec{\gamma}(t) = (2 + \sin t)(\cos 2t, \sin 2t)$  has a double point:  $\vec{\gamma}(0) = \vec{\gamma}(\pi)$  and is not simply closed.



The graph of the above curve

Every carpenter who builds a "real" billiard table will have difficulties with double points.

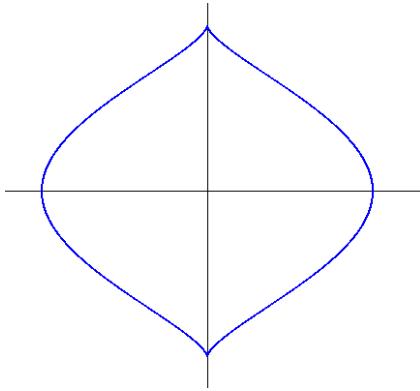
Have we characterised a billiard table according to our imagination? As we can see in the next example, something is still missing.

### Example

The plane curve

$$\gamma: t \in [0, 2\pi[ \rightarrow \vec{\gamma}(t) = \begin{pmatrix} \cos^3 t \\ \sin t \end{pmatrix}$$

Is simply closed. It is continuously differentiable but has "peaks" at  $t = \frac{\pi}{2}, \frac{3\pi}{2}$ . We want to rule this out for a billiard table.



Graph of the curve above

The first derivative of the curve is:

$$\dot{\vec{\gamma}}(t) = \begin{pmatrix} -3\cos^2 t \cdot \sin t \\ \cos t \end{pmatrix}$$

At the points  $t = \frac{\pi}{2}, \frac{3\pi}{2}$  this derivative disappears, i.e.  $|\dot{\vec{\gamma}}(t)| = 0$  for  $t = \frac{\pi}{2}, \frac{3\pi}{2}$ ,

□

If we require that  $|\dot{\vec{\gamma}}(t)| \neq 0$ , this leads to the next definition:

#### **Definition 1.4**

A curve  $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$  is called *regular* if the magnitude of the derivative does not disappear anywhere, i.e.

$$|\dot{\vec{\gamma}}(t)| \neq 0, \forall t \in I$$

□

Does this definition exclude "peaks"? For a continuously differentiable curve, at every point  $t_0 \in I$  applies the approximation:

$$\vec{\gamma}(t_0 + h) \approx \vec{\gamma}(t_0) + h \dot{\vec{\gamma}}(t_0) \text{ if } h \approx 0$$

But for  $|\dot{\vec{\gamma}}(t_0)| \neq 0$  this is the equation of a straight line, and this has no peaks.

Now we can define what we mean by a mathematical billiard table.

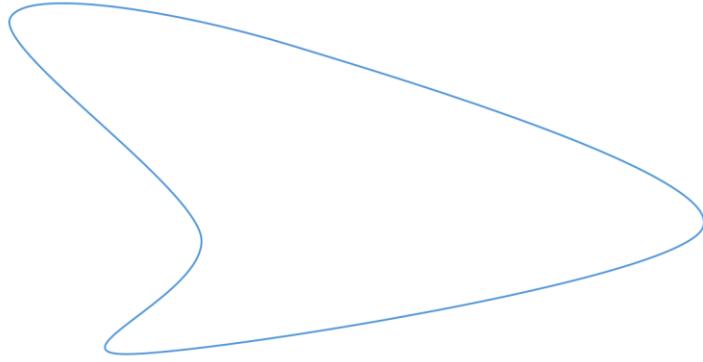
#### **Definition 1.5**

A *mathematical billiard table* is a (finite) region of the plane which is bounded by a simple closed, continuously differentiable and regular plane curve.

□

With this definition, we have excluded rectangular billiard tables, among others. Investigations of billiards in a rectangle are interesting but are not the subject of this paper. You can find relevant literature on the Internet.

For us, billiard tables would still be possible where the edge "jumps in" at some points.



This billiard table has an "inward-curved" part.

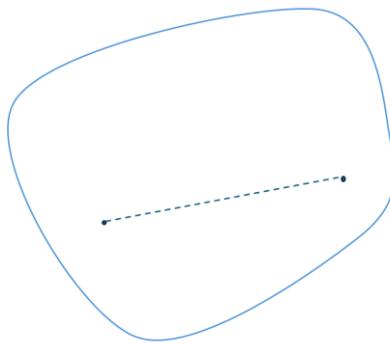
However, we want to restrict ourselves to the simple case that the billiard table does not have any such inward-curved parts. This leads to the last definition in this section:

#### **Definition 1.6**

A mathematical billiard table is called *convex* if, for any two points inside the billiard table, their connecting line also lies entirely inside the billiard table. A convex billiard table can still contain straight segments. If it does not, it is called *strictly convex*.

□

In the following, we will always consider convex billiard tables.



Convex billiard table

We will need the derivation of  $|\vec{\gamma}(t)|$  in several places. In preparation, we note the following:

#### **Lemma 1.7**

Let  $\gamma$  be a continuously differentiable curve  $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$  with  $|\vec{\gamma}(t)| \neq 0, \forall t$

Assertion:

$$\frac{d}{dt} |\vec{\gamma}(t)| = \frac{\vec{\gamma}(t) \cdot \dot{\vec{\gamma}}(t)}{|\vec{\gamma}(t)|}$$

□

Proof:

$$\frac{d}{dt} |\vec{\gamma}(t)| = \frac{d}{dt} \left| \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \right| = \frac{d}{dt} \sqrt{x_1^2(t) + x_2^2(t)} = \frac{2x_1(t) \cdot \dot{x}_1(t) + 2x_2(t) \cdot \dot{x}_2(t)}{2\sqrt{x_1^2(t) + x_2^2(t)}} = \frac{\vec{\gamma}(t) \cdot \dot{\vec{\gamma}}(t)}{|\vec{\gamma}(t)|}$$

□

Remark:

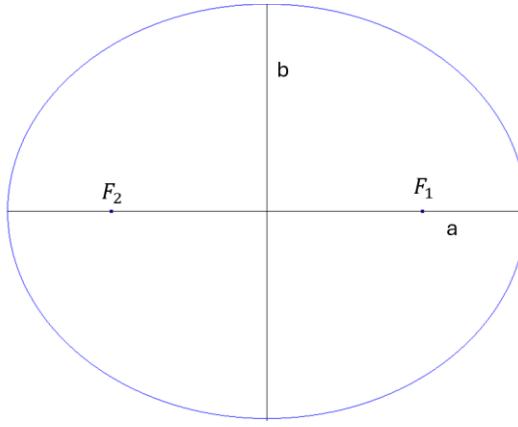
If  $\vec{\gamma}(t) \perp \dot{\vec{\gamma}}(t)$ , then  $\frac{d}{dt} |\vec{\gamma}(t)| = 0$ . If  $\vec{\gamma}(t) \parallel \dot{\vec{\gamma}}(t)$ , then  $\frac{d}{dt} |\vec{\gamma}(t)| = |\dot{\vec{\gamma}}(t)|$ .

*Example*

The elliptical billiard table is an example of a strictly convex billiard table. We will analyse it in more detail in a later section. It is defined by the mapping:

$$\gamma: t \in [0, 2\pi[ \rightarrow \vec{\gamma}(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}, a, b \in \mathbb{R}^+$$

$a$  defines the major axis and  $b$  the minor axis of the ellipse.  $F_{1,2}$  are the so-called focal points.



Elliptical billiard table

For  $a = b$  you get a circle with radius  $a$  and the focal points coincide with the zero point.

## 2. Billiards in a circle

The first simple example of mathematical billiards is a circular billiard table. It is strictly convex.

The standard parameterisation of the unit circle is:

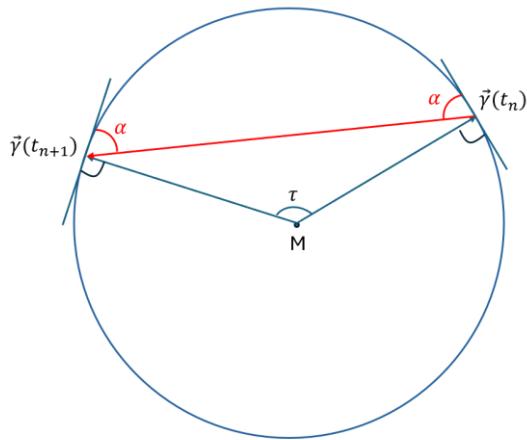
$$\gamma: t \in [0, 2\pi[ \rightarrow \vec{\gamma}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

and this describes a billiard table according to our definition. The parameter  $t$  is also the angle between the position vector  $\vec{\gamma}(t)$  and the positive x-axis.

Starting from a starting point  $\vec{\gamma}(t_1)$  and a starting angle  $\alpha_1$  between the tangent in  $\vec{\gamma}(t_1)$  and the ball trajectory, the further collisions and thus the ball trajectory are defined by a mapping:

$$(t_n, \alpha_n) \rightarrow (t_{n+1}, \alpha_{n+1})$$

You can see that the collision angle  $\alpha$  is constant:

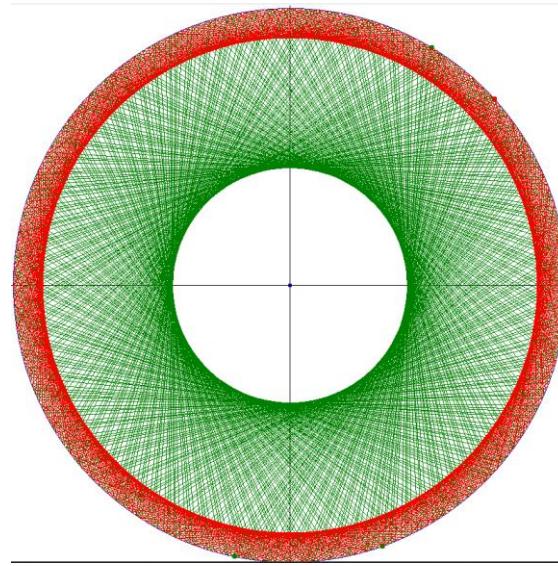


Billiards in a circle

The triangle with the two collision points and the centre point M is equilateral, as the distance M from each collision point is equal to the radius. The base angles of this triangle are therefore equal. The base angles  $\alpha$  are each the complement of a base angle to  $\pi/2$ , i.e. also equal. This means that the collision angle  $\alpha$  is constant during the entire movement of the ball. The parameter  $t$  increases by a constant value  $\tau = 2\alpha$  for the next collision point. The mapping is therefore explicitly given by:

$$\begin{cases} \alpha_{n+1} = \alpha \text{ (constant)} \\ t_{n+1} = t_1 + 2n\alpha \end{cases}$$

Here is an image generated by the "simulator" for two different balls: red with a flat reflection angle and green with a steep one:



Billiard in a circle

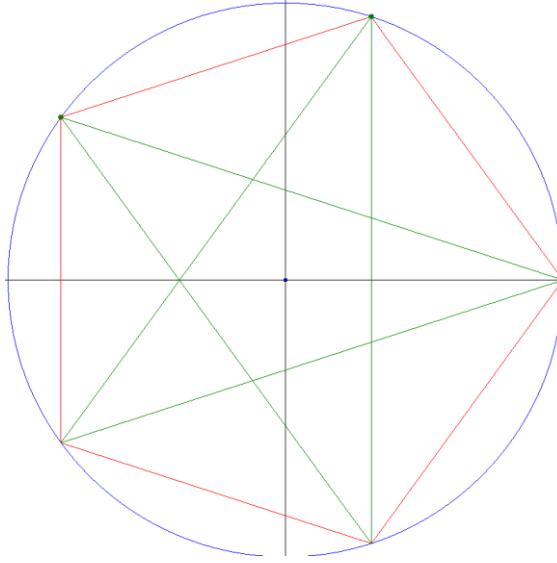
You can also see that the orbits are tangential to an "inner" circle. This is known as the *caustic*. It is easy to check that this inner circle has the radius  $\cos\alpha$  if  $\alpha$  is the constant angle of collision.

Obviously, there are also periodic orbits. Let us examine these in more detail.

Assume that the ball starts at the point  $t_1 = 0$  and the starting angle  $\alpha$  is a rational multiple of  $\pi$ . Then  $t_n = 2\alpha(n - 1)$  is a rational multiple of  $2\pi$ . Let us assume that  $\frac{p}{q} \in \mathbb{Q}$  is a truncated fraction

and  $\alpha = \pi \cdot \frac{p}{q}$ . With each collision, the angle of rotation  $t_n$  increases by  $2\alpha = 2\pi \cdot \frac{p}{q}$ . We only consider the case  $p < q$ , because otherwise the integer part of the fraction only results in a rotation by a multiple of  $2\pi$ .

The path is periodic if  $t_n$  is a multiple of  $2\pi$  after a certain number of collisions. However, this is the case after  $q$  collisions, so  $q$  is the period of the trajectory. Then  $t_n$  has grown to the value  $2\pi p$ . If  $p < \frac{q}{2}$ , then  $p$  is the number of windings of the path around the centre of the circle in a negative clockwise direction. If  $p > \frac{q}{2}$ , then  $q - p$  is the number of windings of the path around the centre of the circle in a positive clockwise direction.



Trajectories with period  $q = 5$  and number of revolutions  $p = 1$  (red) or  $p = 2$  (green)

If the start angle  $\alpha$  of the orbit is an irrational multiple of  $\pi$ , then the centre angle  $\tau$  is an irrational multiple of  $2\pi$ . In this case, the orbit will never close, and the orbit is aperiodic. The points of such an orbit lie close to the edge of the circle. The reason for this is provided by the following consideration:

Let  $\vartheta_1 = 2\alpha$  be the (constant) angle of rotation of the path between two successive collision points and  $\vartheta_1$  not be a rational multiple of  $2\pi$ . Starting from a starting point  $\vec{\gamma}_1$  on the edge of the circle, we then perform as many collisions as possible, until the path comes so close to the point  $\vec{\gamma}_1$  for the first time that it passes  $\vec{\gamma}_1$  on the next collision. This is the case after  $n$  collisions. Starting from  $\vec{\gamma}_1$ , the path will never exactly hit the point  $\vec{\gamma}_1$  itself, because  $\vartheta_1$  would otherwise be a rational multiple of  $2\pi$ .

We denote  $n$  rotations around the angle  $\vartheta_1$  with  $D_{\vartheta_1}^n$ . Then  $\vec{\gamma}_1$  lies between the points:

$$D_{\vartheta_1}^n(\vec{\gamma}_1) < \vec{\gamma}_1 < D_{\vartheta_1}^{n+1}(\vec{\gamma}_1)$$

Whereby the arc length is used as the measure. Either  $\vec{\gamma}_1$  is then closer to the point  $D_{\vartheta_1}^n(\vec{\gamma}_1)$  or otherwise closer to the point  $D_{\vartheta_1}^{n+1}(\vec{\gamma}_1)$ . W.l.o.g. we assume that  $\vec{\gamma}_1$  is closer to  $D_{\vartheta_1}^{n+1}(\vec{\gamma}_1)$ . This distance is then smaller than  $\frac{\vartheta_1}{2}$ . We then set:

$$D_{\vartheta_1}^{n+1} =: D_{\vartheta_2}$$

and replace  $D_{\vartheta_1}^{n+1}$  by a rotation  $D_{\vartheta_2}$  with  $\vartheta_2 < \frac{\vartheta_1}{2}$ . It is then valid for a certain multiple of the rotation  $D_{\vartheta_2}$ :

$$D_{\vartheta_2}^m(\vec{\gamma}_1) < \vec{\gamma}_1 < D_{\vartheta_2}^{m+1}(\vec{\gamma}_1)$$

With the same argument we replace w.l.o.g.:

$$D_{\vartheta_2}^{m+1} =: D_{\vartheta_3}$$

There is therefore a rotation  $D_{\vartheta_3}$  with  $\vartheta_3 < \frac{\vartheta_2}{2}$ , which is a multiple of the rotation  $D_{\vartheta_2}$ . We repeat the same argument enough times until we obtain an arbitrarily small rotation angle  $\vartheta_k < \frac{\vartheta_1}{2^{k-1}}$ , whereby the rotation by this angle is a multiple of  $D_{\vartheta_1}$ :  $D_{\vartheta_k} = D_{\vartheta_1}^r$  for a sufficiently large  $r \in \mathbb{N}$ . So, if we rotate enough times, the path will hit every arbitrarily small interval.

As a result of our considerations, we have

### Theorem 2.1

Let the billiard table be a circle.

Assertion:

- 1) Let  $\frac{p}{q} \in \mathbb{Q}$  be a truncated fraction and  $p < \frac{q}{2}$ . Then the path with the start angle  $\alpha = \pi \cdot \frac{p}{q}$  has the period  $q$  and the winding number  $p$ . For every pair  $(p, q)$  of natural numbers with  $p < \frac{q}{2}$  that are not related to each other, there is an orbit with period  $q$  and winding number  $p$ .
- 2) Let the start angle  $\alpha$  not be a rational multiple of  $\pi$ . Then the path will never close, and it comes arbitrarily close to every point on the edge of the circle.  $\square$

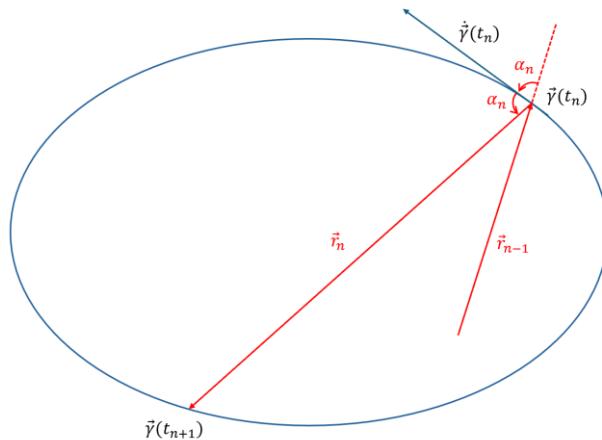
## 3. Elliptical billiards

In this chapter we consider an elliptical billiard table.

The usual parameter representation of an ellipse with principal axes  $a$  and  $b$  is given by:

$$\gamma: t \in [0, 2\pi[ \rightarrow \vec{\gamma}(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}, a, b \in \mathbb{R}^+$$

A ball rolls frictionless on this table and is reflected from the edge according to the law of reflection.



Before the nth collision, the ball travels in the direction of a vector  $\vec{r}_{n-1}$ . It then hits the next collision point  $\vec{\gamma}(t_n)$ . The angle of collision occurring there  $\alpha_n$  is defined by

$$\alpha_n = \arccos \frac{\vec{r}_{n-1} \cdot \dot{\vec{\gamma}}(t_n)}{|\vec{r}_{n-1}| |\dot{\vec{\gamma}}(t_n)|}$$

As this angle lies in the interval  $]0, \pi[$ , this definition is well-defined. The angle is always measured in an anticlockwise direction. In the sketch above, for example, the angle of reflection at the next collision point is  $\alpha_{n+1} > \frac{\pi}{2}$ .

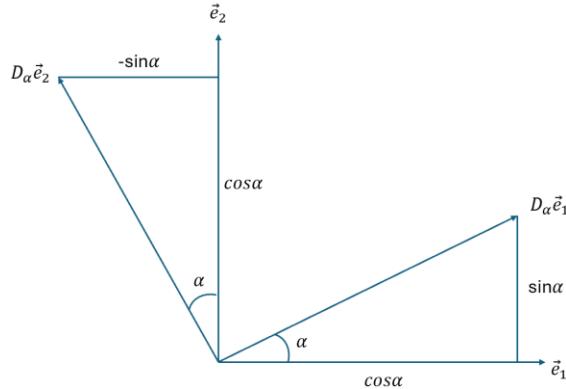
Due to the law of reflection, the direction of the next trajectory is obtained by rotating  $\dot{\vec{\gamma}}(t_n)$  by the angle  $\alpha_n$ . If  $D_{\alpha_n}$  is this rotation around the angle  $\alpha_n$  (anticlockwise), then the following applies:

$$\vec{r}_n = D_{\alpha_n} \dot{\vec{\gamma}}(t_n)$$

We first consider what this rotation looks like.

If the rotation around the zero point and the angle  $\alpha$  is anticlockwise, the basis vectors are mapped as follows:

$$D_\alpha: \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix}$$



Rotation of the basis vectors

Any vector is a linear combination of the basis vectors. In the case of a rotation, it does not matter whether you first combine linearly and then rotate or whether you combine the rotated basis vectors linearly. Any vector is therefore rotated as follows:

$$D_\alpha: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow x_1 \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} + x_2 \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix}$$

Or in matrix notation:

$$D_\alpha \vec{x} = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \cos \alpha - x_2 \sin \alpha \\ x_1 \sin \alpha + x_2 \cos \alpha \end{pmatrix}$$

The image is calculated by forming the scalar product of a matrix row with the vector  $\vec{x}$ . This then provides the corresponding component of the image vector.

So if  $\alpha_n$  is known and the rotation is  $D_{\alpha_n} = \begin{bmatrix} \cos \alpha_n & -\sin \alpha_n \\ \sin \alpha_n & \cos \alpha_n \end{bmatrix}$ , then the ball runs on the straight line after the nth collision:

$$\vec{x}(u) = \vec{\gamma}(t_n) + u \cdot D_{\alpha_n} \dot{\vec{\gamma}}(t_n) = \begin{pmatrix} a \cos t_n \\ b \sin t_n \end{pmatrix} + u \cdot \begin{bmatrix} \cos \alpha_n & -\sin \alpha_n \\ \sin \alpha_n & \cos \alpha_n \end{bmatrix} \begin{pmatrix} -a \sin t_n \\ b \cos t_n \end{pmatrix}, u \in \mathbb{R}$$

This straight line has two intersection points with the ellipse: one for  $u = 0$  (the current collision point) and the other as the new collision point. To determine this, we insert the components of the straight line into the ellipse equation

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

A somewhat complex calculation then provides the result:

$$u^2 \cos^2 \alpha_n + u^2 \sin^2 \alpha_n \cdot \left[ \frac{b^2}{a^2} \cos^2 t_n + \frac{a^2}{b^2} \sin^2 t_n \right] + 2u^2 \sin t_n \cos t_n \sin \alpha_n \cos \alpha_n \cdot \left[ \frac{b}{a} - \frac{a}{b} \right] - 2u \sin \alpha_n \left[ \frac{b}{a} \cos^2 t_n + \frac{a}{b} \sin^2 t_n \right] = 0$$

Since we exclude  $u = 0$ , division by  $u$  provides the condition for the second solution:

$$u \left( \cos^2 \alpha_n + \sin^2 \alpha_n \cdot \left[ \frac{b^2}{a^2} \cos^2 t_n + \frac{a^2}{b^2} \sin^2 t_n \right] + 2 \sin t_n \cos t_n \sin \alpha_n \cos \alpha_n \cdot \left[ \frac{b}{a} - \frac{a}{b} \right] \right) = 2 \sin \alpha_n \left[ \frac{b}{a} \cos^2 t_n + \frac{a}{b} \sin^2 t_n \right]$$

This solution provides the next point of collision on the ellipse:  $\vec{\gamma}(t_{n+1}) = \begin{pmatrix} a \cos t_{n+1} \\ b \sin t_{n+1} \end{pmatrix}$  and this can be used to calculate  $t_{n+1}$ . Then you have  $\dot{\vec{\gamma}}(t_{n+1})$  and therefore also  $\alpha_{n+1}$ .

You therefore have a mapping for the elliptical billiard:

$$(t_n, \alpha_n) \rightarrow (t_{n+1}, \alpha_{n+1}), [0, 2\pi[ x ]0, \pi[ \rightarrow [0, 2\pi[ x ]0, \pi[$$

This mapping is iterated in elliptical billiards.

The simulator chooses another variant for the explicit calculation of the collision points and collision angles. This is described in the mathematical documentation for the simulator.

### Theorem 3.1

For the elliptical billiard, the corresponding mapping  $(t_n, \alpha_n) \rightarrow (t_{n+1}, \alpha_{n+1}), [0, 2\pi[ x ]0, \pi[ \rightarrow [0, 2\pi[ x ]0, \pi[$  can be calculated explicitly according to the procedure described above.

The next collision point results from the intersection between the straight line

$$\vec{x}(u) = \vec{\gamma}(t_n) + u \cdot D_{\alpha_n} \dot{\vec{\gamma}}(t_n)$$

and the ellipse. The next collision angle is

$$\alpha_{n+1} = \arccos \frac{\vec{r}_n \cdot \dot{\vec{\gamma}}(t_{n+1})}{|\vec{r}_n| |\dot{\vec{\gamma}}(t_{n+1})|}$$

□

### Example

We want to check whether the above formulae are correct for the billiard in a circle. The following applies to the unit circle:  $a = b = 1$ . The condition for the second solution in the above formula is then reduced to

$$u = 2\sin\alpha_n$$

If you insert this into the equation, you get the following for the next point of collision:

$$\begin{aligned} \begin{pmatrix} \cos t_{n+1} \\ \sin t_{n+1} \end{pmatrix} &= \begin{pmatrix} \cos t_n \\ \sin t_n \end{pmatrix} + 2\sin\alpha_n \begin{bmatrix} \cos\alpha_n & -\sin\alpha_n \\ \sin\alpha_n & \cos\alpha_n \end{bmatrix} \begin{pmatrix} -\sin t_n \\ \cos t_n \end{pmatrix} \\ &= \begin{pmatrix} \cos t_n - 2\sin\alpha_n \cos\alpha_n \sin t_n - 2\sin^2\alpha_n \cos t_n \\ \sin t_n - 2\sin^2\alpha_n \sin t_n + 2\sin\alpha_n \cos\alpha_n \cos t_n \end{pmatrix} \\ &= \begin{pmatrix} \cos t_n (\cos^2\alpha_n - \sin^2\alpha_n) - 2\sin\alpha_n \cos\alpha_n \sin t_n \\ \sin t_n (\cos^2\alpha_n - \sin^2\alpha_n) + 2\sin\alpha_n \cos\alpha_n \cos t_n \end{pmatrix} \\ &= \begin{pmatrix} \cos t_n \cos 2\alpha_n - \sin t_n \sin 2\alpha_n \\ \cos t_n \sin 2\alpha_n + \sin t_n \cos 2\alpha_n \end{pmatrix} = D_{2\alpha_n} \begin{pmatrix} \cos t_n \\ \sin t_n \end{pmatrix} \end{aligned}$$

The new collision point is therefore created when the previous collision point is rotated by the angle  $2\alpha_n$ . This means:  $t_{n+1} = t_n + 2\alpha_n$ .

Furthermore

$$\cos\alpha_{n+1} = \frac{\vec{r}_n \cdot \dot{\vec{\gamma}}(t_{n+1})}{|\vec{r}_n| |\dot{\vec{\gamma}}(t_{n+1})|}$$

Now is:  $\vec{\gamma}(t_{n+1}) = D_{2\alpha_n} \vec{\gamma}(t_n)$ , i.e.  $\dot{\vec{\gamma}}(t_{n+1}) = D_{2\alpha_n} \dot{\vec{\gamma}}(t_n)$ . On the other hand,  $\vec{r}_n = D_{\alpha_n} \dot{\vec{\gamma}}(t_n)$ . The expression

$$\frac{\vec{r}_n \cdot \dot{\vec{\gamma}}(t_{n+1})}{|\vec{r}_n| |\dot{\vec{\gamma}}(t_{n+1})|}$$

is the cosine of the angle between  $\vec{r}_n$  and  $\dot{\vec{\gamma}}(t_{n+1})$ , i.e. between  $D_{2\alpha_n} \dot{\vec{\gamma}}(t_n)$  and  $D_{\alpha_n} \dot{\vec{\gamma}}(t_n)$ . But this is just  $\alpha_n$ . Thus applies:

$$\cos\alpha_{n+1} = \cos\alpha_n$$

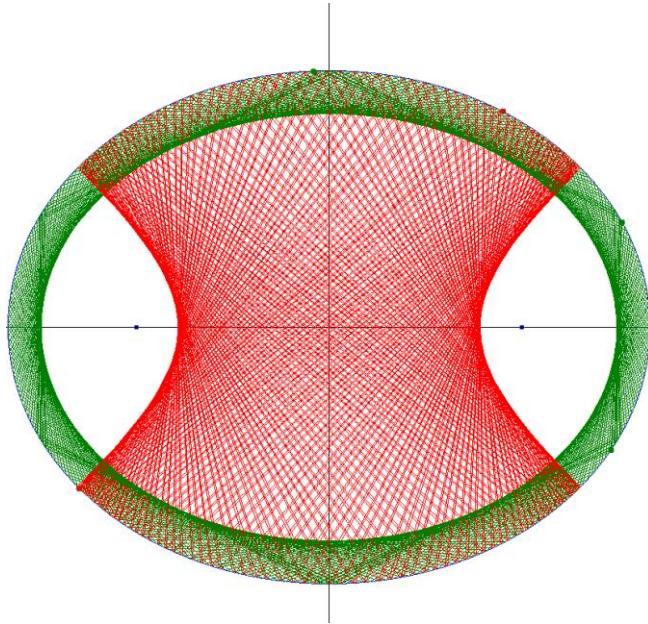
And because of  $\alpha_n \in ]0, \pi[$  applies:  $\alpha_{n+1} = \alpha_n = \text{constant}$ .

This result is the mapping  $(t_n, \alpha_n) \rightarrow (t_{n+1}, \alpha_{n+1})$  for circular billiards.

That was a bit complicated, but it served to check our previous considerations about elliptical billiards.

□

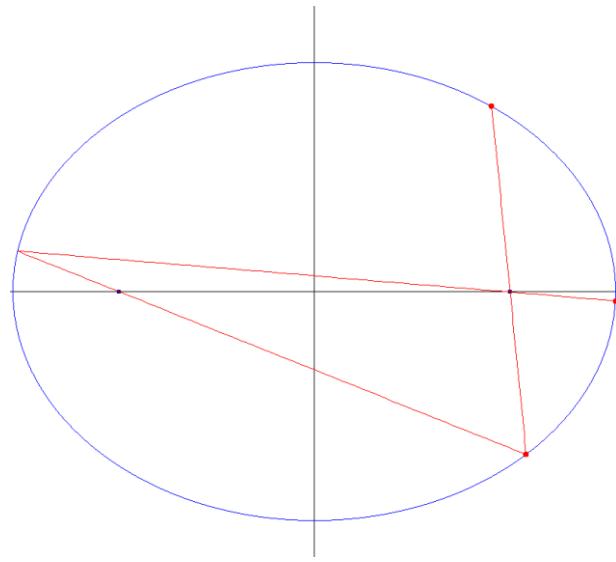
We will now carry out an experiment with the simulator:



Trajectory of two balls on the elliptical billiard table

You can see in the picture above: If the starting angle is chosen to be flat (green above), then the ball path runs around the focal points of the ellipse. If the starting angle is steep (red at the top), then the ball path runs up and down between the focal points.

We know from optics that rays emanating from one focal point are bundled at the other focal point. A test with the simulator shows this behaviour, whereby a billiard table with the axis ratio  $c := \frac{b}{a} = 0.76$  is selected in the following image  $t_1 = 0.941304146267719$  and  $\alpha_1 = 2.17540450580406$ .



A ball starts at the top right, and the path runs through the focal point on the right at the start

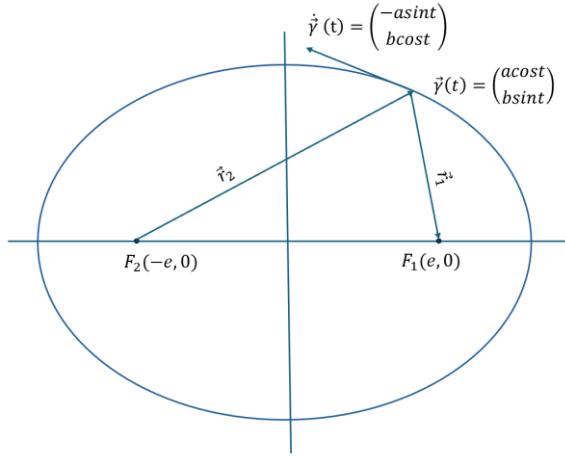
There are elegant elementary geometric proofs for the assertion that a ball trajectory that passes through one focal point passes through the other focal point after the collision. Here we will give a proof using vector geometry.

The focal points of an ellipse with major axis  $a$  and minor axis  $b$  have the coordinates  $(\pm e, 0)$  where  $e^2 := a^2 - b^2$ . Let  $t$  again be the parameter of the parameter representation of the ellipse defined in the beginning.

As preparation, we state

$$e \cdot \cos t \leq e < \sqrt{e^2 + b^2} = a$$

Thus  $a - e \cdot \cos t > 0$



Focal points of the ellipse

The following applies to the magnitude of the vectors  $\vec{r}_{1,2} = (\pm acost \mp e, \pm bsint)$ :

$$\begin{aligned} |\vec{r}_{1,2}|^2 &= a^2 \cos^2 t + b^2 \sin^2 t + e^2 \mp 2eacost \\ &= a^2(1 + \cos^2 t) + b^2(\sin^2 t - 1) \pm 2eacost \\ &= a^2 + a^2 \cos^2 t - b^2 \cos^2 t \pm 2eacost \\ &= a^2 + e^2 \pm 2eacost = (a \pm ecost)^2 \end{aligned}$$

Because  $a - e \cdot \cos t > 0$  therefore applies:  $|\vec{r}_{1,2}| = a \pm ecost$

Now we calculate:

$$\vec{r}_{1,2} \cdot \dot{\gamma} = -easint \mp a^2 costsint \pm b^2 sintcost = -easint \mp e^2 sintcost = -esint(a \pm ecost)$$

Thus applies:

$$\cos \alpha_{1,2} = \frac{\vec{r}_{1,2} \cdot \dot{\gamma}}{|\vec{r}_{1,2}| |\dot{\gamma}|} = \frac{-esint}{|\dot{\gamma}|}$$

If  $\alpha_{1,2}$  is the angle between  $\vec{r}_{1,2}$  and the tangent  $\dot{x}$ . However, the right-hand side is independent of whether you consider  $\vec{r}_1$  or  $\vec{r}_2$  and therefore  $\alpha_1 = \alpha_2$ , as these angles lie in the interval  $]0, \pi[$ .

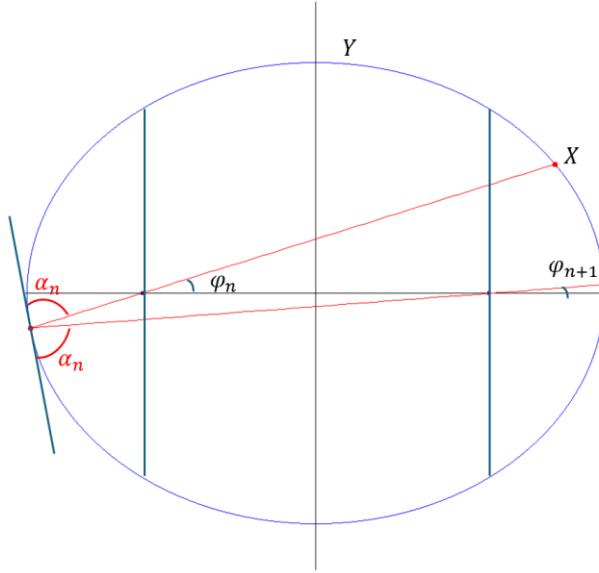
Result:

### Theorem 3.2

If a section of the billiard ball's path passes through a focal point of the ellipse, then each section of the ball's path passes through one of the focal points in turn.

□

Let's look again at the case of an orbit that passes through a focal point.



Orbit through the focal points

If the ball starts at a point  $X$  and then passes through the distant focal point, the ball's trajectory forms an angle  $\varphi_n$  with the x-axis. If, as outlined above,  $X$  lies above the x-axis, then the angle of reflection is  $\alpha_n < \frac{\pi}{2}$ . After the collision, the following applies to the next angle between the ball's trajectory and the x-axis:  $\varphi_{n+1} = \varphi_n - (\pi - 2\alpha_n)$ . For reasons of symmetry, the same applies to the next collision. Thus, the angle  $\varphi_n$  decreases with each collision. The lower limit is reached when  $\pi - 2\alpha_n \approx 0$ , or when the ball travels back and forth between the vertices of the main axis.

If the ball had started at a point  $Y$ , then the above case would be reached after the first collision. If the ball starts at  $X$  but passes through the nearby focal point, it will land at a point in the range  $Y$  after the first collision, and after the second collision, the above case will occur again. Thus, we have the following result:

### Theorem 3.3

A trajectory that passes through the focal points of the ellipse comes arbitrarily close to the major axis of the ellipse during the iteration.

□

A tantalising question is *whether elliptical billiards could be understood as an affine image of circular billiards*. Similar to the school project *The chaotic properties of logistic growth*, where we were able to visualise logistic growth with the parameter  $a = 4$  as a conjugate of the tent map. This gave us a lot of clarity about logistic growth.

So, if  $g: (s, \beta) \mapsto (s', \beta')$  describes the circular billiard with certain initial parameters, then the elliptical billiard  $f: (t, \alpha) \mapsto (t', \alpha')$  should be representable as a conjugate in the form:

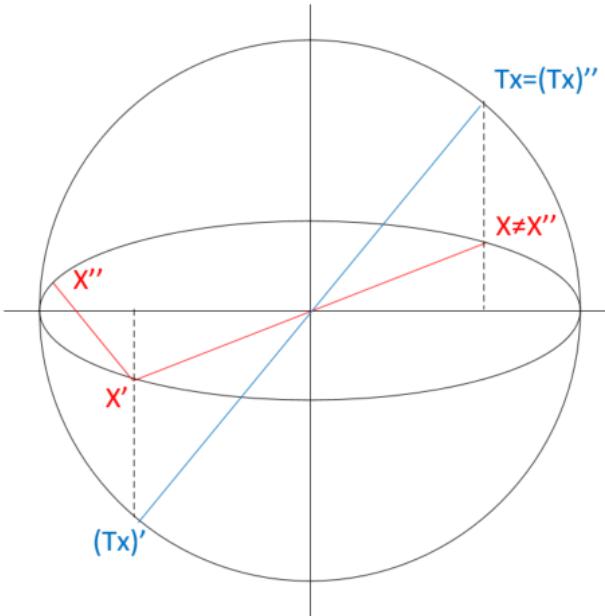
$$f(t, \alpha) = T^{-1} \circ g \circ T(t, \alpha) = T^{-1} \circ g(s, \beta) = T^{-1}(s', \beta') = (t', \alpha')$$

where  $T$  is the stretching of the plane in the y-direction by the factor  $a/b$ . In this case, an ellipse with axes  $a$  and  $b$  is transformed into a circle with radius  $a$ . The following then applies to a point  $\vec{x}(t)$  on the ellipse:

$$T: \vec{x}(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix} \mapsto \vec{s}(t) = \begin{pmatrix} a \cos t \\ a \sin t \end{pmatrix}$$

We would therefore first map the starting point and the starting angle of the ball through  $T$  to the corresponding starting point of a ball in the circle. Then a collision is performed in the circle, which leads to the next collision point, whereby the start angle of the circle remains constant during the collision. We then transform everything back to the ellipse and hope that we get the same result as if we had performed the collision directly in the ellipse according to the law of reflection.

Can this work? For example, the focal points of the ellipse remain fixed in the image  $T$  but lose their significance in the circle. We are therefore looking for a counterexample that shows that the proposed conjugation does not work in this way. Let's look at the following figure:



Affine mapping between ellipse and circle

We consider the point  $X$ . In billiards, it will end up at  $X''$  after two collisions. The image of  $X$  under the affine figure  $T$  is  $TX$ . After two collisions, this point lands back at itself and not at  $TX''$ .

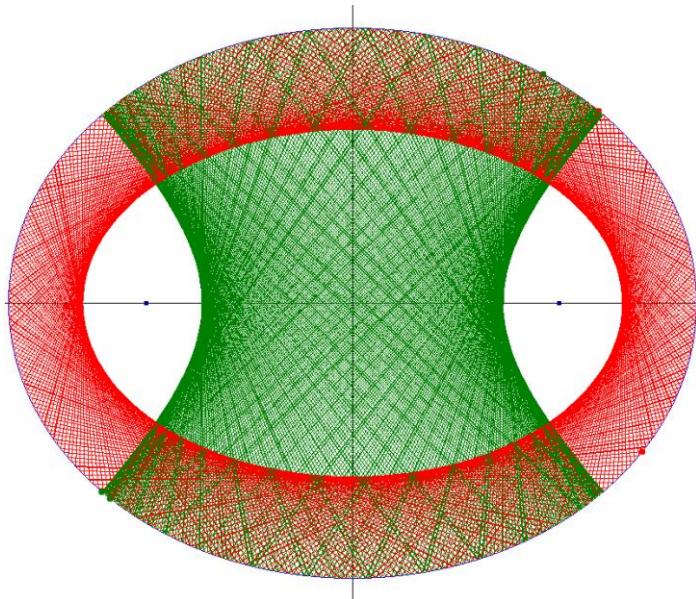
#### Theorem 3.4

*It is not possible to represent the elliptic billiard as an affine conjugate of the circular billiard.*

□

## 4. Caustics in elliptical billiards

A *caustic* is a term from optics and describes an arc to which light rays of an optical system are tangents. In the case of billiards, the ball paths are considered instead of light rays. We have already seen pictures of elliptical billiards:

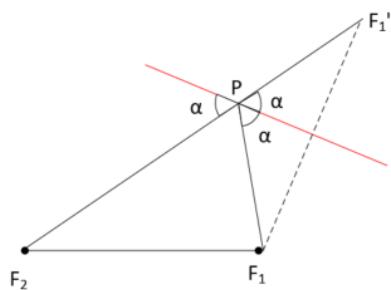


Caustics in elliptical billiards

The picture above shows the trajectories of two different balls, one in green with a starting direction between the focal points, and one in red with a starting direction outside the focal points of the ellipse. You can see that the caustic in both cases is obviously a conic section. An ellipse in the red case and a hyperbola in the green case. These have the same focal points as the ellipse of the billiard table.

We want to prove this in the case of the red curve. With the help of vector geometry, this can be time-consuming. We resort to an elementary geometric proof.

We consider the elliptical billiard table and a fixed, given section of the path of a ball. Firstly, we must consider where this section of the path touches the caustic. To do this, we consider a (different) ball that starts at one focal point, is mirrored at the given section of the path according to the law of reflection and travels to the other focal point. (There is always such a path.)



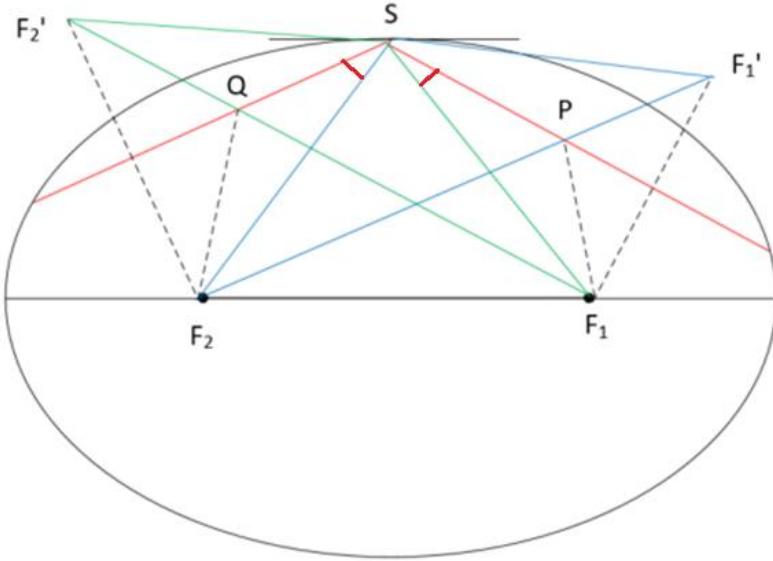
Construction of the point  $P$  on the path section (red)

The path section is shown in red above. We must select the desired meeting point  $P$  in such a way that the law of reflection applies to a focal beam. To do this, we mirror  $F_1$  at the path section and then connect the mirrored point  $F_1'$  with the other focal point  $F_2$ . At the intersection point  $P$  with the path section, all marked angles are equal and therefore the law of reflection applies. This also applies exactly when the caustic touches the path section at the point  $P$ . Perhaps the caustic is not yet an ellipse, but there is an ellipse  $E_1$  with focal points  $F_1$  and  $F_2$ , so that the path segment is a tangent to this ellipse.

You can easily check that we get the same point  $P$  if we had mirrored the other focal point  $F_2$  instead of  $F_1$ .

We then construct the next meeting point  $Q$  for the next section of the path using the same procedure. This point also has an ellipse  $E_2$  with focal points  $F_1$  and  $F_2$ , so that the path section is a tangent to this ellipse. It must now be shown that  $E_1 = E_2$  and this is the caustic.

We consider the following figure with the original ellipse, the path segments (red), the collision point  $S$  and the points constructed as before  $P, Q$ :



Two consecutive path segments (red) with the collision point  $S$  on the ellipse

We first consider the angles at the centre of collision. A ball that starts at the focal point  $F_2$  in the direction of the blue line and is reflected at the collision point ends up at the other focal point  $F_1$  according to theorem 3.2. The law of reflection applies:

$$\text{Angle (blue line through } F_2 \text{, tangent)} = \text{angle (green line through } F_1 \text{, tangent)}$$

As the red lines are trajectories of the ball, the law of reflection also applies here and therefore

$$\text{Angle (red line after the collision, tangent)} = \text{angle (red line before the collision, tangent)}$$

It follows, however, that the difference angles (marked red in the figure) are also the same on both sides. Since  $F_2$  and  $F_2'$  are mirror-symmetrical, the angles of the mirrored difference angles are also the same. This gives us a first intermediate result:

At the point  $S$ , the interior angles of the blue and green triangles are equal.

Because of the mirroring, the following now applies:  $|SF_2'| = |SF_2|$  and  $|SF_1'| = |SF_1|$ . This means that the blue and green triangles are *congruent*. This means that the third side opposite the point  $S$  is also the same length for both triangles and the following applies:

$$|F_2'F_1| = |F_1'F_2| \Rightarrow |F_2Q| + |QF_1| = |F_2P| + |PF_1|$$

Because an ellipse is the location of all points that have a constant sum of distances from two (focal) points,  $P$  and  $Q$  lie on the same ellipse and  $E_1 = E_2$  applies.

We will carry out an analogue consideration for the case that the path sections run between the focal points in the form of an exercise and see that in this case a hyperbola with the focal points  $F_1$  and  $F_2$  appears as a caustic.

In the specific case, this ellipse (or hyperbola) can be determined explicitly by proceeding as in the construction above and calculating the respective major axis length  $|F_2P| + |PF_1|$ .

Result:

#### Theorem 4.1

In elliptical billiards, the caustic is

- a) An ellipse if the ball trajectory runs outside the focal points
- b) A hyperbola if the ball trajectory runs inside the focal points

Both conic sections have the same focal points as the original ellipse.

□

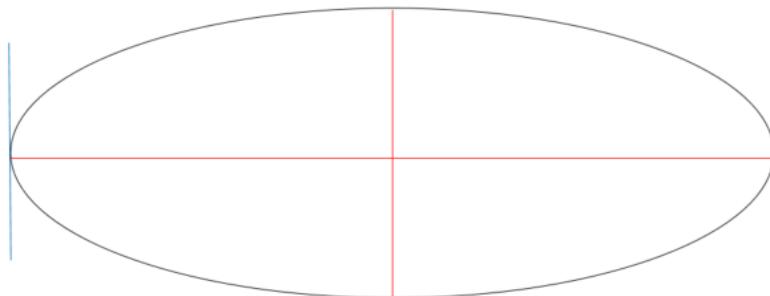
Note: In the case of the hyperbola, the point of contact of the (extended) ball path may also lie outside the billiard table.

As soon as the starting point and the first angle of reflection of a ball on an elliptical billiard table are determined, the associated caustic is also clearly defined. This means that the further path of the ball is also defined by the caustic. A tangent is laid from a butt point to the caustic, intersected with the ellipse and the next butt point is obtained.

## 5. Periodic points in elliptical billiards

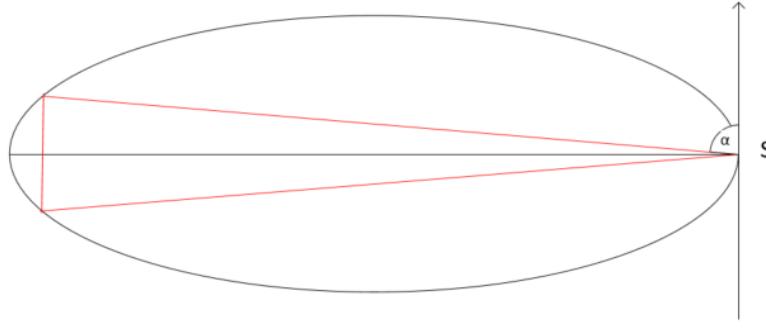
In this section, we examine periodic paths in elliptical billiards.

In the case of a 2-periodic point, the direction of the incident ball path must be rotated by the angle  $\pi$  so that the outgoing beam hits the original starting point again. This means that the direction of incidence is perpendicular to the tangent at the point of collision. The same also applies when the ball hits the starting point again. The only possible trajectories are therefore



2-period trajectories in elliptical billiards

Trajectories of period three exist. If you look at the following sketch:



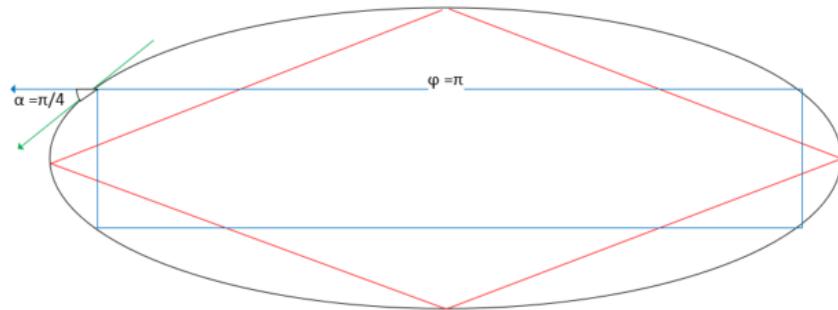
A three-periodic path

Starting from the point  $S$ , select  $\alpha$  so that the ball travels vertically downwards after the first collision and returns to the point  $S$  after the second collision. If you experiment with the simulator, you will see that this is possible. If  $\alpha$  is selected slightly too small, you will end up slightly below  $S$ . If  $\alpha$  is set too large, you will end up above  $S$ . If  $\alpha$  is constantly changed in between, then there must be a  $\alpha$  such that the point  $S$  is hit again.

It seems difficult to find an analytical or elementary geometric solution to the problem. A vector-geometric approach leads to a complicated equation for the required  $\alpha$ .

However, the simulator can be used to find an approximate solution. In an experiment, we start with a slightly too small  $\alpha_1$  and then select the next  $\alpha_2$  slightly too large. The  $\alpha$  you are looking for must therefore lie somewhere in between for reasons of continuity. Using an interval nesting method, we can then narrow down the  $\alpha$  we are looking for more precisely.

Orbits of period 4 are again easy to find:



Orbits of period four

A path that runs from one vertex to the next is shown in red. It can be travelled clockwise or anticlockwise.

Another 4-period orbit is shown in blue. The point of collision at the angle  $\alpha$  is easy to determine: The new direction angle of the path must point vertically downwards after the collision. The direction angle of the ball before the collision is equal to  $\pi$ . The direction of the tangent relative to the positive x-axis must then be  $5\pi/4$ . This provides the coordinates for the desired collision point

$$\vec{r} \approx \begin{pmatrix} -a \cdot 0.707 \\ b \cdot 3.927 \end{pmatrix}$$

Experiments with the simulator show that apparently non-symmetrical paths of period 4 also exist, e.g. with starting point  $t = 1$  a corresponding starting angle  $\alpha$  lies in the interval  $[0.721, 0.722]$ .

This method can obviously be used to find starting angles that lead to orbits of any given period. In fact, elliptical billiards have trajectories of any period. The following applies:

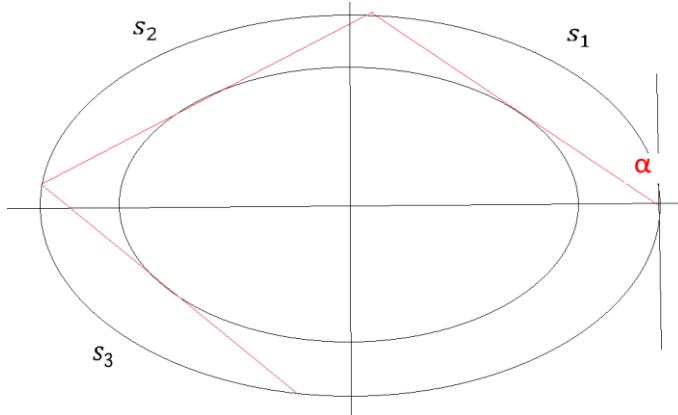
### Theorem 5.1

In elliptical billiards, for every pair of  $(p, q)$  natural numbers with  $q \geq 2$  and  $p < [q/2]$  that have no common divisor, there is a  $q$ -periodic path with a winding number  $p$ .

$[q/2]$  here denotes the next smaller natural number less than  $q/2$ .

□

*Proof*



A ball starts at the right vertex and a starting angle  $\alpha$

If a ball starts at the right vertex and the angle  $\alpha$ , then the caustic is defined by the first path section. The other sections of the path are well-defined tangents to the caustic. The direction of rotation remains the same. The arc lengths of the path sections  $s_1, s_2, s_3, \dots$  on the edge of the ellipse lie on the outside of the path sections relative to the caustic and are also well-defined.

Now we consider the functions

$$L_q(\alpha) := \sum_{i=1}^q s_i$$

Where  $q$  is fixed. This is the total arc length on the ellipse, which corresponds to the path of the ball after  $q$  collisions.  $L_q(\alpha)$  is continuous and strictly monotonically increasing in  $\alpha$ , since the starting point is fixed at the right vertex of the ellipse. Now we let  $\alpha$  move towards  $\pi/2$ . Then the spherical path approaches the two-periodic path between the main vertices. In the case of  $q$  joints, the arc length  $q \cdot \frac{U}{2}$  then corresponds to the spherical path if  $U$  is the circumference of the ellipse. Thus  $L_q(\alpha)$  is a continuous mapping:

$$L_q(\alpha): ]0, \frac{\pi}{2}] \rightarrow ]0, q \cdot \frac{U}{2}]$$

Which assumes the maximum in  $\frac{\pi}{2}: L_q\left(\frac{\pi}{2}\right) = \frac{q}{2} \cdot U$ . The ball then has made exactly  $p = [\frac{q}{2}]$  revolutions.

According to the intermediate value theorem,  $L_q(\alpha)$  assumes every value in the interval  $]0, q \cdot \frac{U}{2}]$ , i.e. in particular also all values  $p \cdot U$  for  $p < [\frac{q}{2}]$ . For each fixed  $p < [\frac{q}{2}]$  there is therefore a  $q$ -

periodic path, so that the ball returns to the starting point after  $q$  collisions and after  $p$  revolutions at the latest. The question now is whether  $p$  is the correct winding number of the path, or whether the ball returns to the starting point earlier, i.e. after  $q' < q$  collisions for the first time. Let's assume that the corresponding winding number is  $p' < p$ . Since the ball is in any case at the starting point after  $q$  collisions and  $p$  revolutions, there must be natural numbers such that  $s \cdot q' = q$  and  $s \cdot p' = p$ , i.e.  $p$  and  $q$  would have a common divisor. The assertion follows from this in reverse.

□

## 6. The phase portrait

Let a billiard table be defined by the edge curve  $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$ . Each collision is uniquely described by a pair of  $(t_n, \alpha_n)$  with  $t_n \in I$  and  $\alpha \in ]0, \pi[$ .

### **Definition 6.1**

Each collision can be represented by a point in the space  $I \times ]0, \pi[$ . This space of all possible collisions is called *phase space*.

□

The trajectory of a ball is determined if a starting point on the edge of the billiard table and a starting angle are given. Each subsequent collision is defined by a billiard mapping

$$(t_n, \alpha_n) \rightarrow (t_{n+1}, \alpha_{n+1})$$

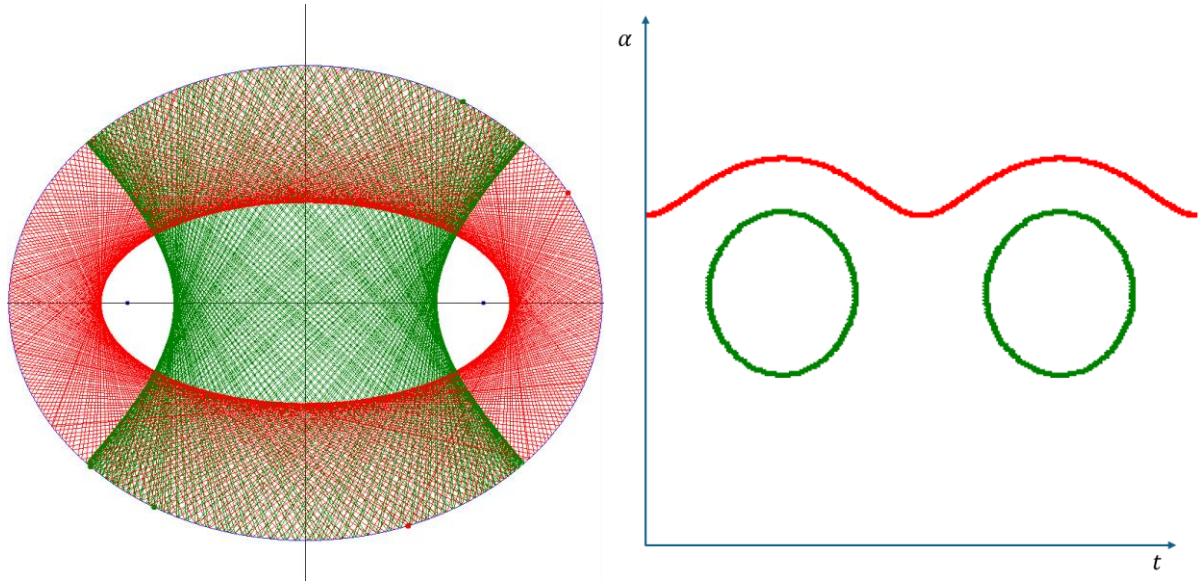
Where  $t_n \in I$  and  $\alpha \in ]0, \pi[$ . Therefore, if  $t_1$  and  $\alpha_1$  are given, the orbit consists of all subsequent pairs  $(t_n, \alpha_n)$ .

### **Definition 6.2**

The set of all pairs  $(t_n, \alpha_n)$  that occur in an orbit can be represented in phase space. This set is called the *phase portrait* of the orbit.

□

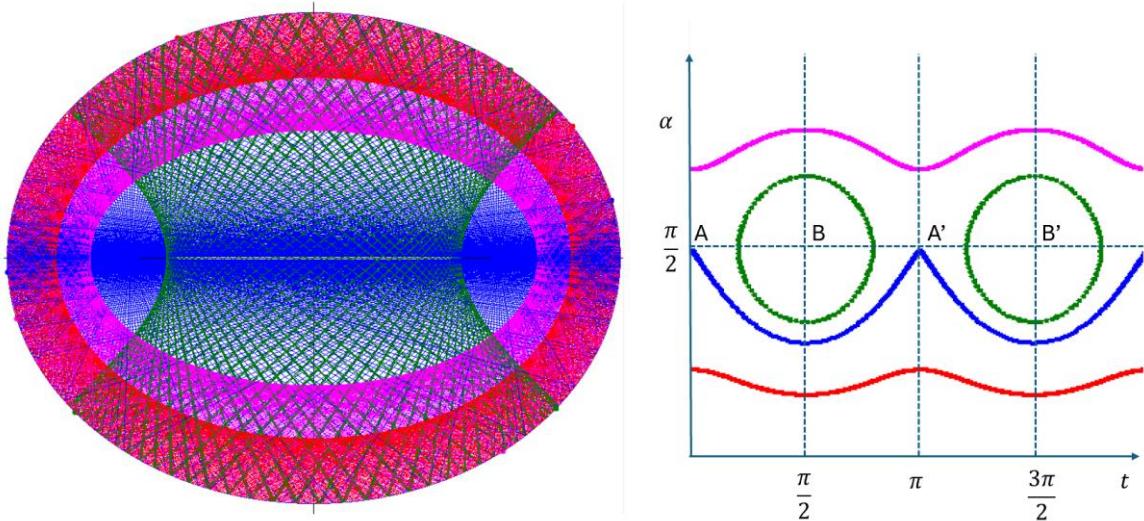
Let us look at an example in the case of elliptical billiards and one red and one green orbit. The simulator also shows the corresponding phase portrait for each orbit:



Two ball tracks on the left and the corresponding phase portrait on the right

In the phase portrait, the parameter  $t$  is plotted on the x-axis and the collision angle  $\alpha$  on the y-axis. You can see that the red orbit comes close to all edge points on the ellipse over time. The collision angle  $\alpha$  oscillates within a limited range. This leads to the red wavy line on the right, which however extends in the x-direction over the entire interval  $I$ . In the green path, the collision points are either at the upper edge or the lower edge. The phase portrait therefore breaks down into two separate areas in the interval  $I$ , which are hit alternately because the ball swings back and forth between the upper and lower edge of the billiard table. Here too, the collision angle  $\alpha$  lies within a limited range.

Now let's examine the phase portrait in more detail. Firstly, due to the periodicity of the sine and cosine functions, we realise that for  $t = 0$  and  $t = 2\pi$  we have identical impact points with identical impact angles for each path. The phase space is therefore a cylinder. The phase space is "glued together" at the straight lines  $t = 0$  and  $t = 2\pi$ .



Phase portrait of special paths

In the diagram above, the points  $A, A'$  belong to a trajectory that starts at the right-hand vertex of the large semi-axis, i.e. at  $t = 0$ , at an angle of  $\alpha = \frac{\pi}{2}$ . It then hits the edge of the billiard at the left-hand vertex, i.e. at  $t = \pi$ , and is reflected there again at an angle of  $\alpha = \frac{\pi}{2}$ . This means that the phase portrait of this path consists of the points  $A, A'$  and swings back and forth between them. Similarly, the pair of points  $B, B'$  belongs to a path that oscillates back and forth between the vertices of the minor semi-axis. In the green path,  $t, \alpha$  remains within a limited area and the nudge points oscillate back and forth between the left and right green curves. The blue curve shows a path that passes through the focal points of the ellipse. For the red path, the first impact angle was  $\alpha < \frac{\pi}{2}$ . It is clear from the theorems on caustics that this path rotates anti-clockwise and that all impact angles remain  $< \frac{\pi}{2}$ . Similarly, the first impact angle for the path in magenta is  $\alpha > \frac{\pi}{2}$  and this path rotates clockwise, whereby the impact angles always remain  $> \frac{\pi}{2}$ .

We consider a ball that starts at a specific starting position on the elliptical billiard table and rolls outside the focal points. Then, the associated caustic is uniquely defined. These caustics each define a unique mapping

$$\tau_l: t \in [0, 2\pi[ \rightarrow \alpha \in ]0, \frac{\pi}{2}]$$

And

$$\tau_r: t \in [0, 2\pi[ \rightarrow \alpha \in ]\frac{\pi}{2}, \pi[$$

Starting from a starting point on the edge, there is a left-rotating path (this is the first case above) in which each section of the path is the tangent from the starting point to the caustic. There is also a right-turning path (this is the second case above).

If the ball runs within the focal points, an analogue consideration applies.

### **Definition 6.3**

Given an elliptical billiard table and a caustic. Then we call the set of pairs of points  $(t, \tau_l(t))$  or  $(t, \tau_r(t))$  a *trajectory* in phase space according to the figures above.

□

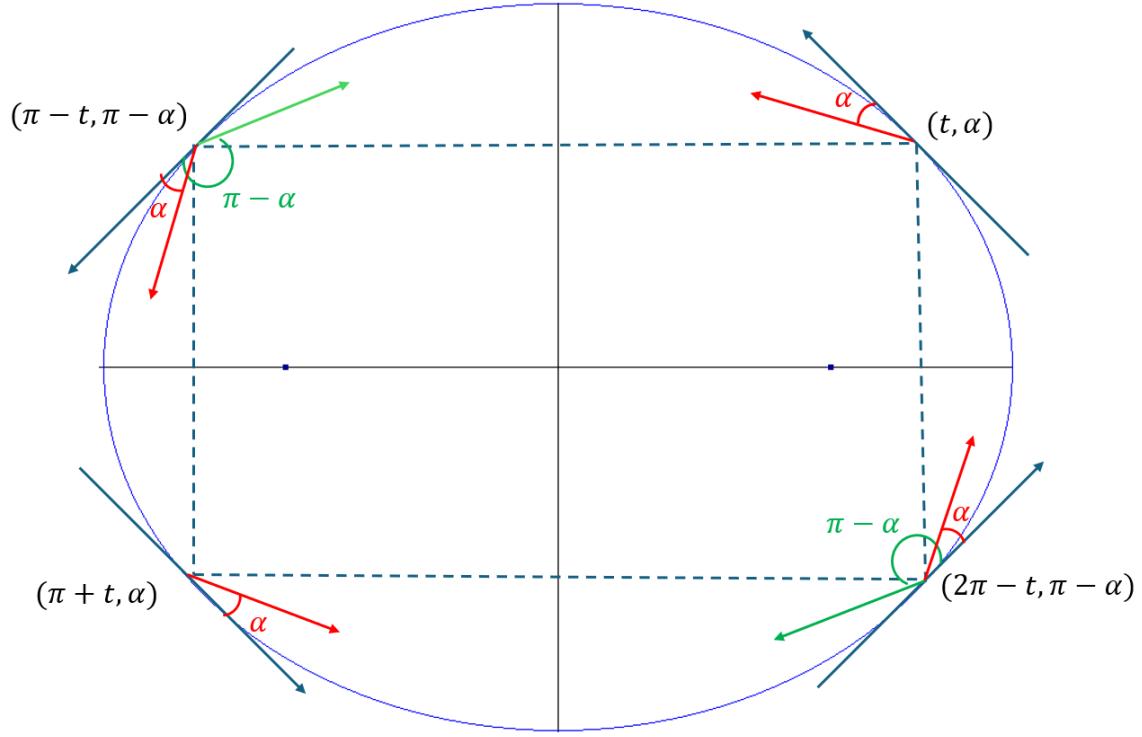
This immediately results in

### **Corollary 6.4**

If two trajectories have the same caustic and the same direction of rotation, then their phase portrait lies on the same trajectory in phase space.

□

We can now investigate the apparent symmetries in phase space in more detail. Consider the following sketch:



Trajectories with symmetrical initial parameters

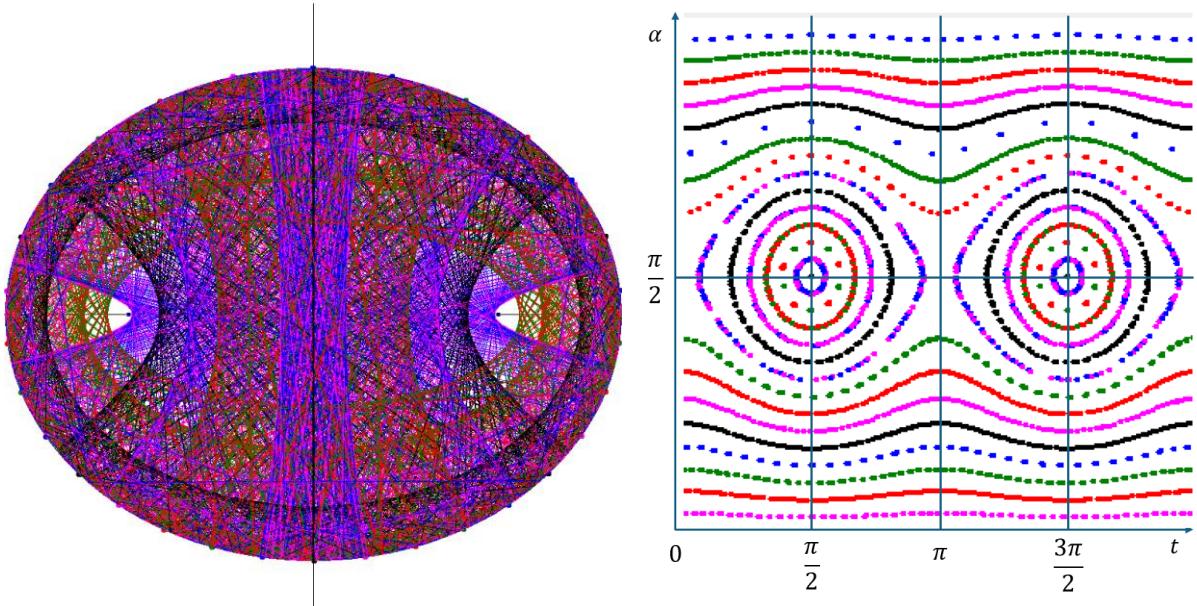
Due to the axial symmetry of the ellipse, the left-rotating red orbit with the starting point  $(t, \alpha)$  and the right-rotating green orbit with the starting point  $(\pi - t, \pi - \alpha)$  have the same caustics. Because of the law of reflection, the right-turning red path also has the same caustic. If you extend the argument to the other starting points shown above, you can see that all red trajectory sections lie on the same trajectory in phase space, as do all green trajectory sections. In the red case, this trajectory has the same value for the values of  $t$  and  $\pi - t$ . It is therefore mirror-symmetrical to the perpendicular line  $t = \frac{\pi}{2}$ . The same applies to the symmetry axis  $t = \frac{3\pi}{2}$  if you look at the lower red points in the sketch above. Since the same value of  $\alpha$  also occurs for  $t$  and  $2\pi - t$ , we also have the axis of symmetry  $t = \pi$ . Furthermore, the left-turning, red trajectory and the associated right-turning, green trajectory are symmetrical, whereby for a fixed  $t$  the associated green angle  $\pi - \alpha$  belongs to a red angle  $\alpha$ . This means that the horizontal straight line  $\alpha = \frac{\pi}{2}$  is also an axis of symmetry of the phase space. If the ball runs within the focal points, we leave an analogue consideration as exercise. As a result, we have

### Theorem 6.5

In elliptical billiards, the phase space as a cylinder has four vertical axes of symmetry, namely  $t = 0, t = \frac{\pi}{2}, t = \pi, t = \frac{3\pi}{2}$  and the horizontal axis of symmetry  $\alpha = \frac{\pi}{2}$ .

□

The phase portrait and the phase space are useful because they provide a good overview of the set of all trajectories. In the simulator, you can distribute many balls with different starting angles across the billiard table, let them run and then view the phase space.



Phase portrait of many ball trajectories with symmetry axes

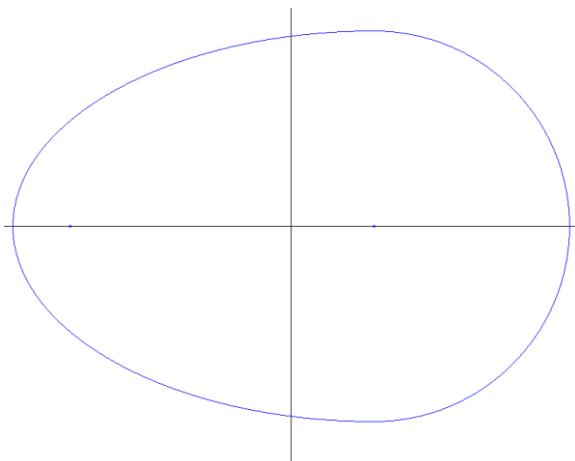
## 7. Oval billiard

Chickens would probably not enjoy certain eggs, which we will use as a billiard table below. The billiard table is made up of a semicircle and half an ellipse. Where  $b$  is always the radius of the circle and if  $a \geq b$   $a$  is the main axis of the ellipse and  $b$  its minor axis. If  $a < b$  it is the other way around. In the simulator, you can freely select the ratio  $c := \frac{b}{a}$  and the shape of the billiard table adapts depending on your choice.

We select the parameter representation of the oval:

$$\vec{\gamma}(t) = \begin{cases} \left( \begin{array}{c} m + b \cos t \\ b \sin t \end{array} \right), & t \in [-\pi/2, \pi/2] \\ \left( \begin{array}{c} m + a \cos t \\ b \sin t \end{array} \right), & t \in ]\pi/2, 3\pi/2[ \end{cases}$$

Where  $m := (a - b)/2$  is the x-coordinate of the centre of the semicircle. Its y-coordinate is 0.



Oval billiard table: Half an ellipse on the left, a semicircle on the right

The curve parameterised in this way is continuous at the connection points. The following applies, for example:

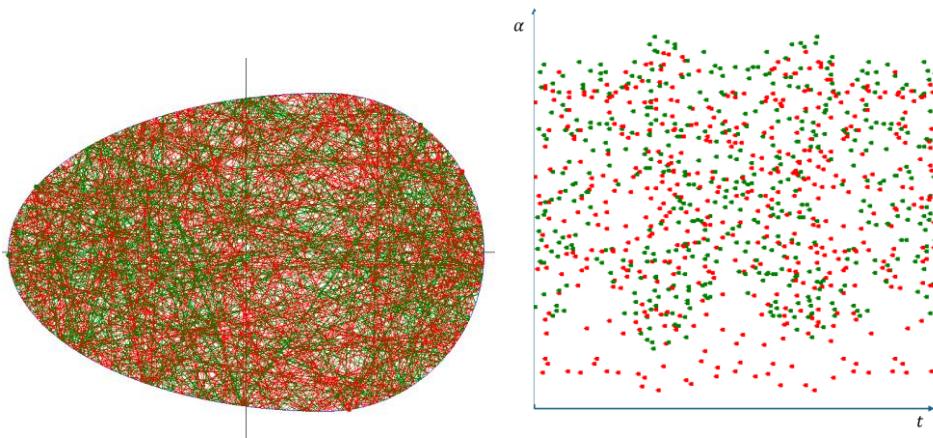
$$\lim_{t \rightarrow \pi/2^+} \vec{\gamma}(t) = \begin{pmatrix} m \\ b \end{pmatrix} = \lim_{t \rightarrow \pi/2^-} \vec{\gamma}(t) = \vec{\gamma}(\pi/2)$$

However, it is not differentiable at the transition points if  $a \neq b$ . One has e.g.:

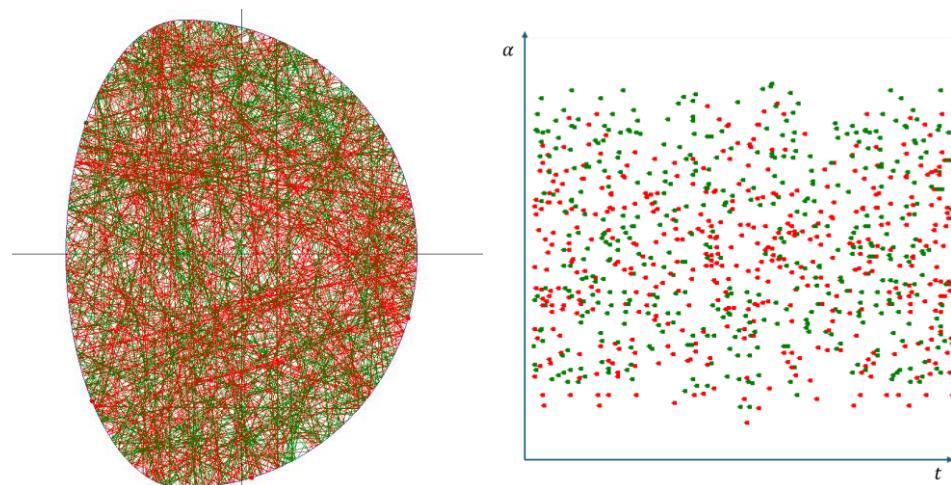
$$\lim_{t \rightarrow \pi/2^+} \dot{\vec{\gamma}}(t) = \begin{pmatrix} -a \\ 0 \end{pmatrix} \neq \begin{pmatrix} -b \\ 0 \end{pmatrix} = \lim_{t \rightarrow \pi/2^-} \dot{\vec{\gamma}}(t)$$

For the reflection, however, we only need the direction of the tangent at this point, and this is the same at the transition points 0.

The calculation of the next collision points and collision angle is carried out in the same way as for the elliptical billiard and is described in the mathematical documentation for the simulator. Let's look at some images generated with the simulator.



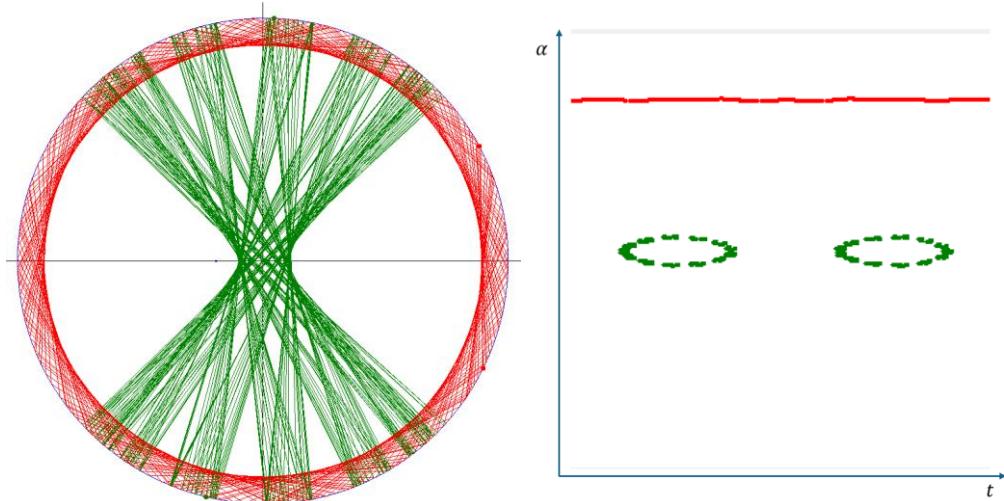
The trajectory of two balls with  $c = 0.5$  and the corresponding phase portraits



The same for two balls with  $c = 2$

If you let the balls run for longer, the points in the phase portrait spread out more over the entire space. This suggests chaotic behaviour, without us being able to justify this here. For the exact definition of chaotic behaviour, see the school project *The chaotic properties of logistic growth*.

However, if you choose  $c$  close to 1, then the trajectories are more like those in an elliptical billiard table:



Two balls with  $c = 0.98$

If  $c$  is close to 1, then the trajectories and their phase portrait resemble an elliptical billiard table.

The periodic trajectories can also be discussed for oval billiards. We will leave this as an exercise. We will see later that the statement in Theorem 4.1 also applies to oval billiards.

## 8. The C-diagram

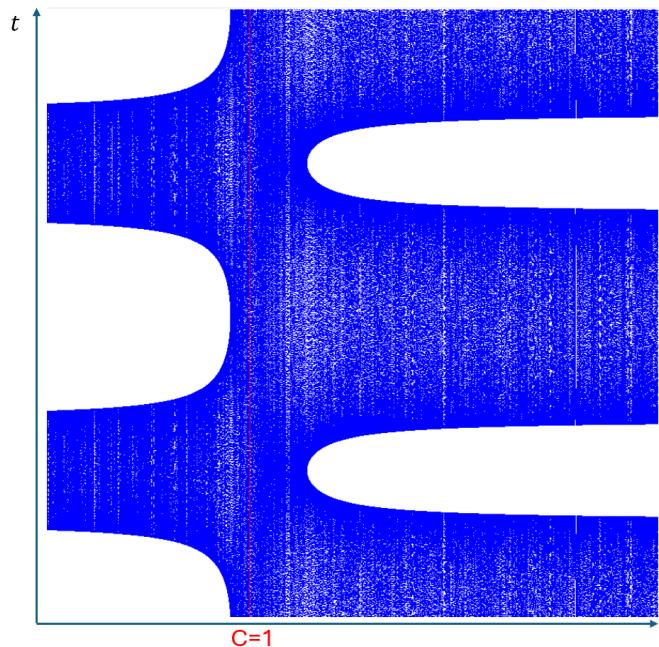
The parameter  $C$  determined the shape of the billiard table in elliptical and oval billiards. In the first case,  $C$  was the ratio of the minor axis to the major axis of the ellipse, in the second case the ratio of the circle radius to the major axis of the ellipse.

What effect does the continuous change of  $C$  have on ball tracks? This can be analysed in the simulator. A ball is always started from a definable fixed position. The simulator then generates the following diagram:

$C$  is varied along the x-axis. The y-axis is then used to plot either the reflection angle or collision parameters. This shows the range in which these values move for each value of  $C$ .

*Example*

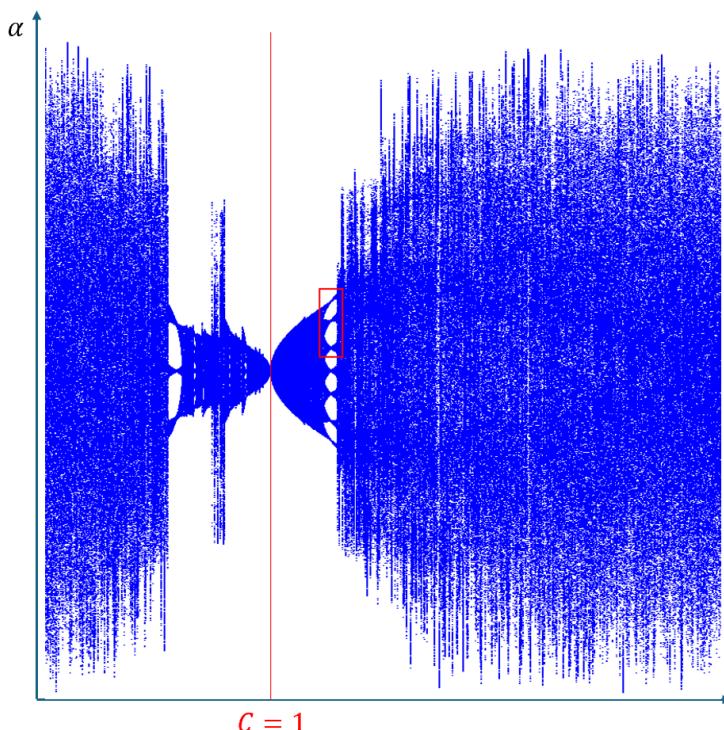
The following diagram is obtained for the elliptical billiard and the plotted collision parameter:



$C$  diagram for the elliptical billiard and the parameter  $t$

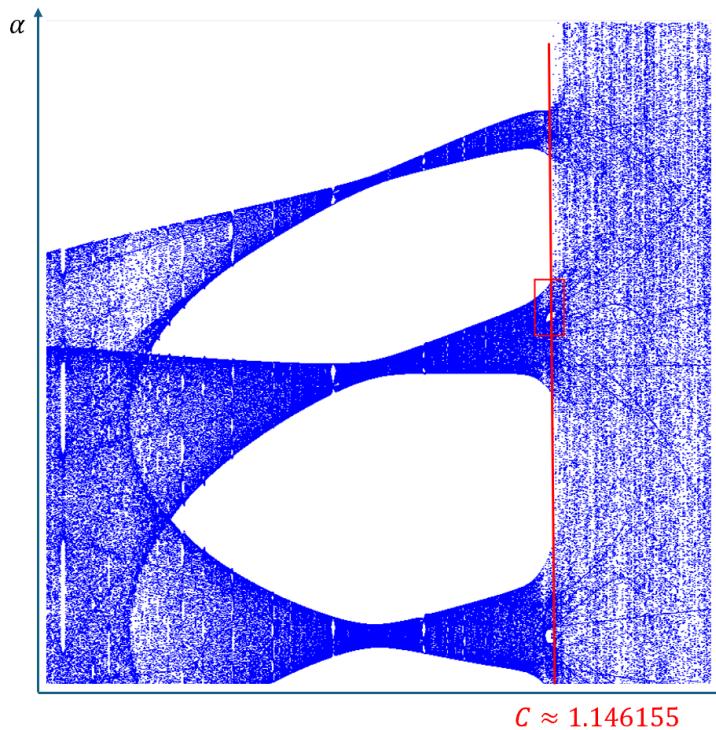
$C$  varies between 0.5 and 2. For  $C = 1$  you have a circle and the parameter  $t$  is distributed over the entire interval  $]0, 2\pi[$ . Due to its periodicity, the image above is a cylinder, with the lower and upper edges glued together in the direction of the  $x$ -axis. In the neighbourhood of  $C = 1$  the ball runs around the outside of the focal points and  $t$  is distributed throughout  $]0, 2\pi[$ . To the left and right of this,  $t$  runs between the focal points and is only distributed in two sub-intervals of  $]0, 2\pi[$ .

It is interesting to analyse the angle of reflection  $\alpha$  for the oval billiard.



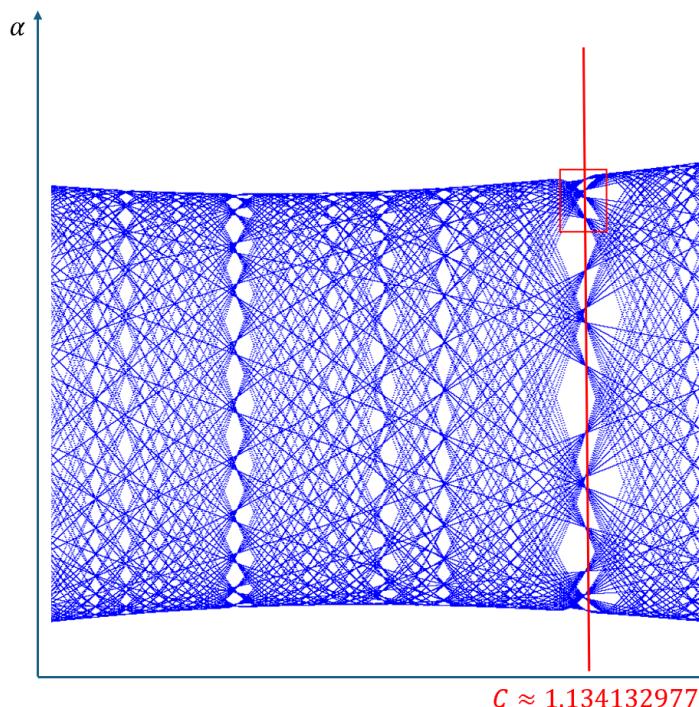
$C$  diagram for the oval billiard and the angle of reflection  $\alpha$

For  $C = 1$  you have the circle and  $\alpha$  is constant. On the far left and far right,  $\alpha$  is "smeared" in the interval  $]0, \pi[$ . However, there also appear to be periodic windows in the neighbourhood of  $C = 1$ . You can use the simulator to mark certain areas and then zoom in. A periodic area is marked at the top. The centre of the rectangle is at  $C \approx 1.1333$ . The zoomed image then looks like this:



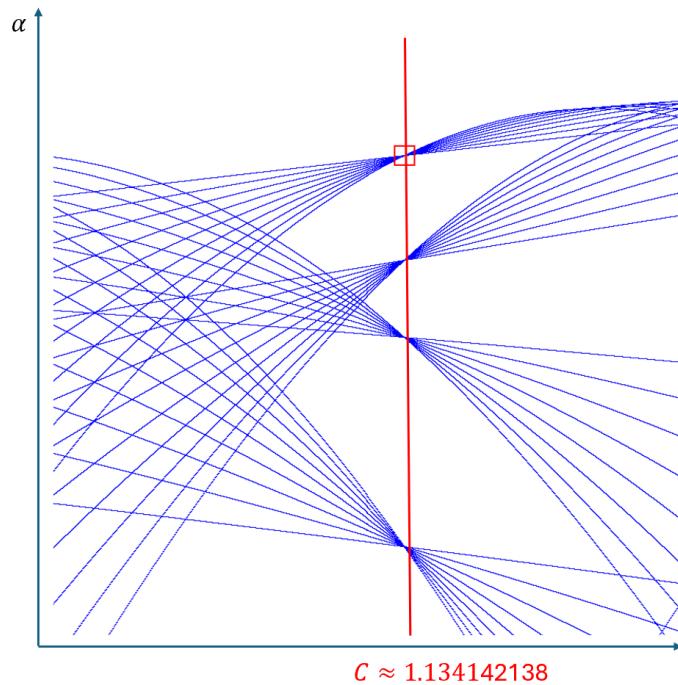
Zoom into the first section

Here we have again marked a section into which we are zooming.



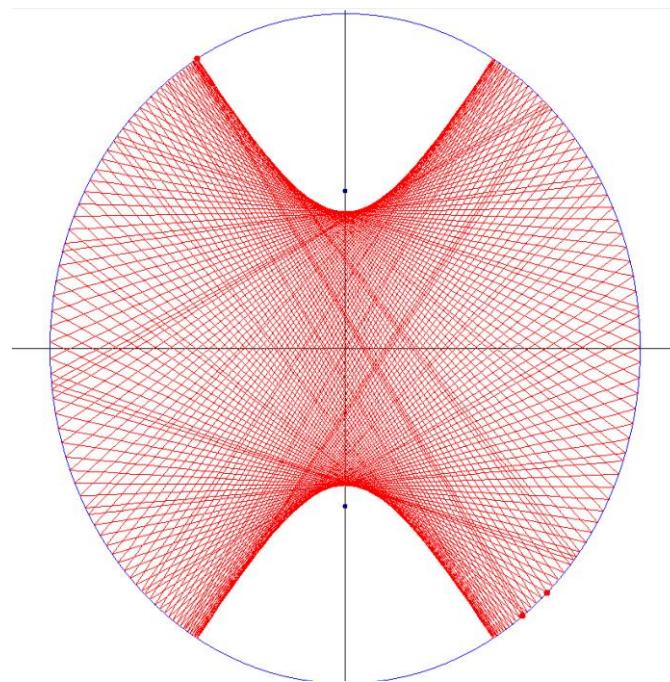
Zoom into the next section

A rather complex-looking period is already visible here. We zoom in again:



Zoom into the last section

In the simulator, we can estimate the values of  $C$ . We now go back to the oval billiard with the value  $C = 1.134142138$  and the start parameters that were selected for the C-diagram, namely  $t = 2.0944, \alpha = 1.5708$ .

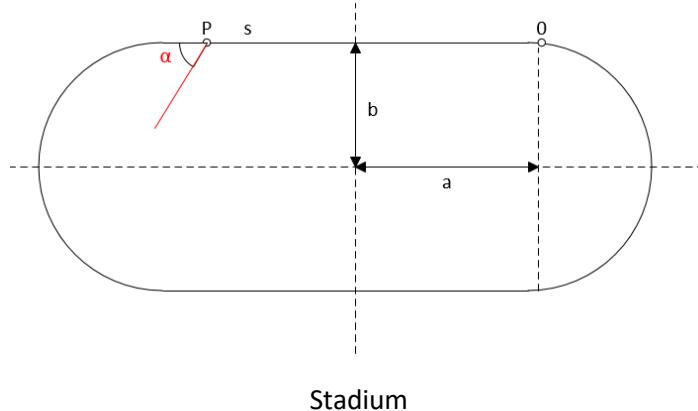


The ball orbit with the above parameters

We leave the further examination of the elliptical billiard to the reader, in particular diagrams for the reflection angle  $\alpha$ .

## 9. Billiards in a stadium

Here we analyse a stadium-shaped billiard table. This is only convex, but no longer strictly convex.

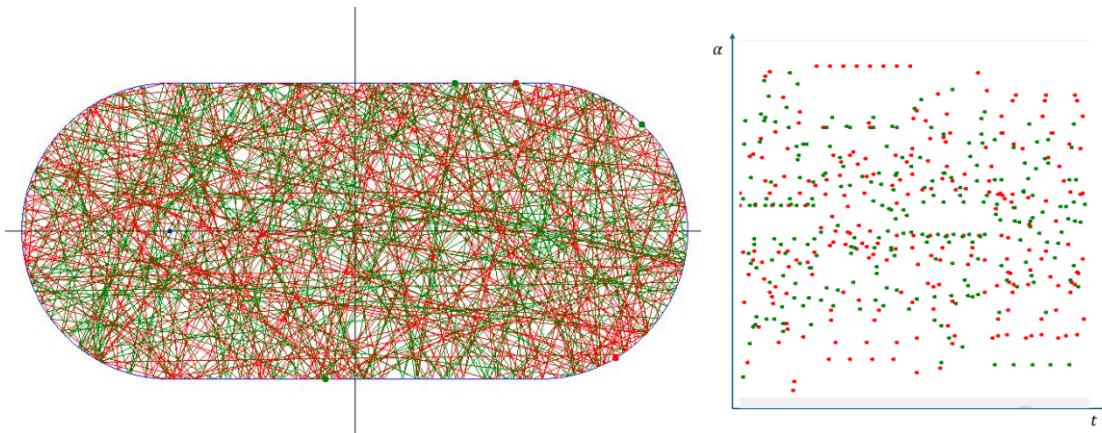


The stadium is made up of a rectangle with a width of  $2a$  and a height of  $2b$ . A semicircle with a radius  $b$  is added to each of the vertical sides. We use a coordinate system with the zero point at the centre of the stadium and the symmetry lines as axes. Furthermore:  $c := \frac{b}{a}$ .

A billiard ball is reflected at the edge of the stadium according to the law of reflection.

For the parameterisation, we fix a zero point on the edge, namely the point  $(a, b)$ . To describe a point  $P$  on the edge, we use the arc length  $s$  between this point and the fixed zero point on the edge. We calculate this modulo the circumference of the stadium, i.e. modulo  $L = 4a + 2\pi b$ . This means that each collision point of the billiard ball is uniquely defined by the parameter  $s$ . As the second parameter, we again use the angle  $\alpha$  between the ball path and the curve tangent at the point of collision. If the ball hits a rectangular side, this tangent is just identical to the rectangular side. In the other case, the curve tangent is the circle tangent at the point of collision.

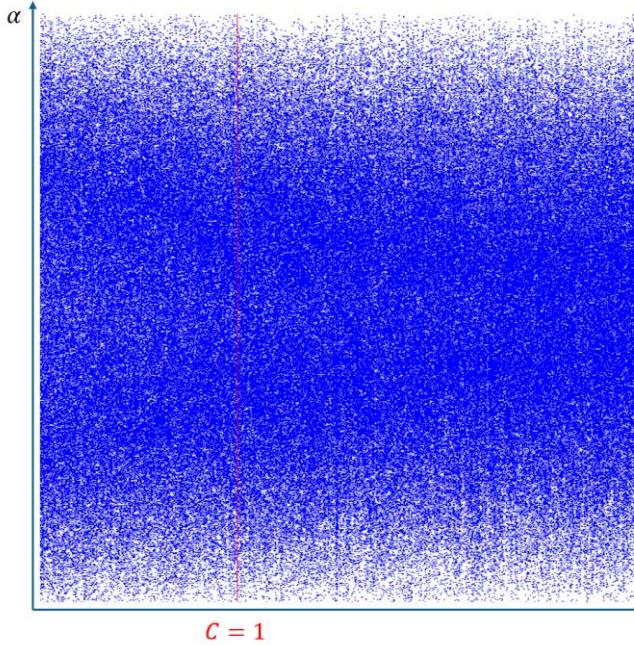
The procedure for finding the next collision point is analogous to the procedure for elliptical billiards, whereby a multiple case differentiation is necessary here, depending on the area of the billiard table in which the ball collisions. The method used by the simulator is described in the mathematical documentation for the simulator. We conduct an experiment with the simulator. A red and a green ball are launched.



Billiards in the stadium

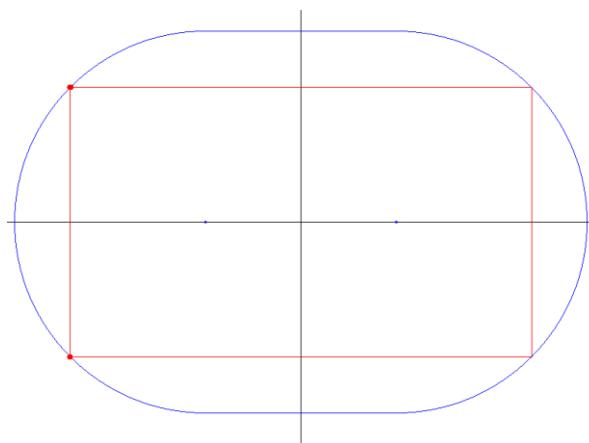
As you can see, the points corresponding to the individual collisions are distributed over the entire phase space. In fact, billiards in the stadium is chaotic. The Russian mathematician Leonid Bunimovich (1947 - ) proved this.

The C-diagram for billiards in the stadium is also an indication of this:



C-diagram for billiards in the stadium and the angle of reflection  $\alpha$

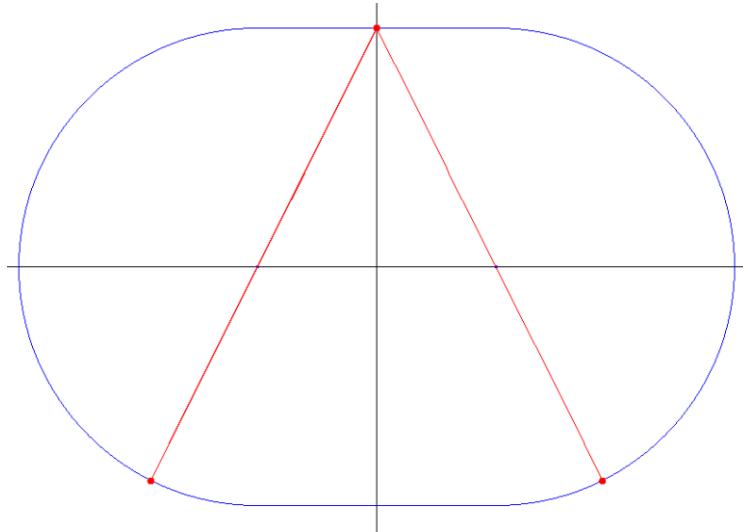
It is also easy to find periodic points for billiards in the stadium. For example, starting from a vertex  $(0, a + b)$  there are paths of every period  $2k, k \in \mathbb{N}$ , which can be found using corresponding paths in the circle.



A path of period 4 on the table  $c = 2$

As you can easily calculate, the following applies to the above path  $s = 2a + b \cdot \frac{\pi}{4}$  and  $\alpha = \frac{\pi}{4}$

If a path passes through the centre of a circle and is then mirrored at it, it passes through the centre again on the way back. This allows you to find further periodic paths.

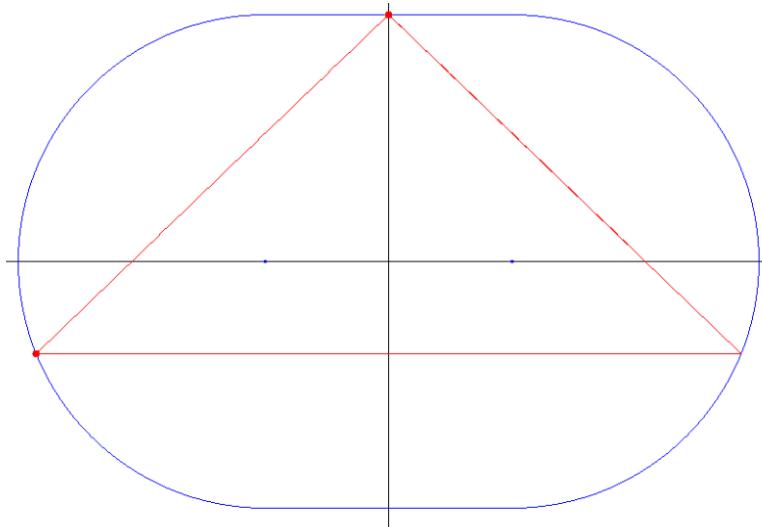


Another path of period four on the table  $c = 2$

In the above case,  $s = 0.31\bar{6}$  is used again, i.e. the starting point is  $(0, b)$ . The angle  $\alpha$  is:  $\alpha = \arctan \frac{b}{a} = 1.10715 \dots$

For paths of odd period, we can again determine the corresponding start parameters at least approximately by means of interval nesting. As an example, we are looking for a three-period orbit with a starting point in  $(0, b)$  for the factor  $c = 2$ . The start parameter for this factor is  $s = a = 0.31\bar{6}$ . We start with an angle  $\alpha = 0.75$  and adjust it continuously so that the parameter  $s = 0.31\bar{6}$  for the position of the ball is reached as well as possible after three collisions in the protocol.

After a few iteration steps, we obtain the following approximation  $\alpha \approx 0.766239$ :



Approximate start parameters for a three-periodic path on the table  $c = 2$

## 10. The convex billiard table

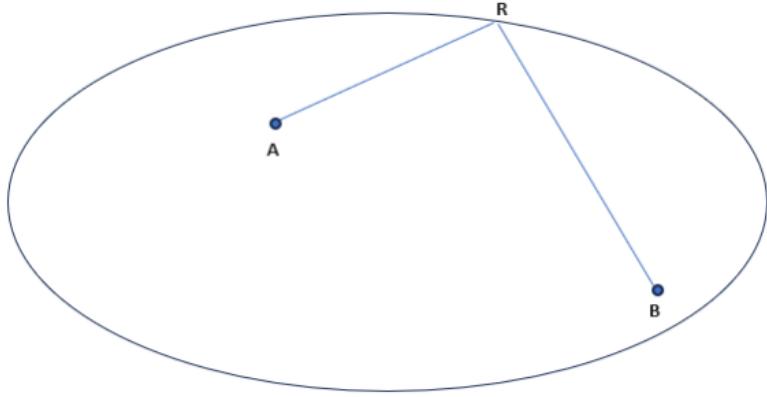
Some elementary statements are valid for arbitrary convex billiard tables and in this section, we will formulate two of them and prove them elementarily.

First, we define the term "reflection point".

### Definition 10.1

Let  $A, B$  be two different points inside a convex billiard table. Then a point  $R$  on the edge of the billiard table is called a *reflection point for  $A, B$*  if a section of the path starting from  $A$  is reflected in  $R$  in such a way that it arrives at  $B$  after reflection.

□



$R$  is a reflection point for  $A, B$

#### Example

In elliptical billiards, every point on the ellipse is a reflection point for the focal points of the ellipse.

A first question is: Is *there always a reflection point  $R$  for any point  $A \neq B$  that lies inside a convex billiard table?*

We consider the function

$$f(t) = |\overrightarrow{AR}| + |\overrightarrow{RB}|, t \in I$$

Where  $I$  is the parameter interval for the boundary curve that limits the billiard table. Let  $\vec{r}(t)$  be the position vector associated with  $R$ . Then:

$$f(t) = |\vec{r}(t) - \vec{a}| + |\vec{r}(t) - \vec{b}|$$

$f$  is a real-valued, continuous function and  $I$  is a closed interval. If  $f$  is constant, then  $A$  and  $B$  are focal points of an ellipse and then each edge point is a reflection point. Otherwise,  $f$  assumes a maximum and a minimum on  $I$ .  $I$  can be continued periodically as the boundary curve is closed. This means that the extremes of  $f$  occur in the interior of  $I$  or in the interior of the periodic continuation. Because the boundary curve is continuously differentiable,  $f$  is also continuously differentiable. The derivative of  $f$  at the extreme points is equal to zero. Let  $t_1, t_2$  be the parameter values for which  $f$  is maximum or minimum. Then the following applies:

$$\frac{d}{dt} f(t) = \frac{(\vec{r}(t) - \vec{a})}{|\vec{r}(t) - \vec{a}|} \cdot \dot{\vec{r}}(t) + \frac{(\vec{r}(t) - \vec{b})}{|\vec{r}(t) - \vec{b}|} \cdot \dot{\vec{r}}(t) = 0 \text{ for } t = t_{1,2}$$

$\frac{(\vec{r}(t) - \vec{a})}{|\vec{r}(t) - \vec{a}|} =: \vec{e}_a$  is a unit vector and likewise  $\frac{(\vec{r}(t) - \vec{b})}{|\vec{r}(t) - \vec{b}|} =: \vec{e}_b$  in each direction of the corresponding path section through the point  $A$  or  $B$ . The sum of the unit vectors  $\vec{e}_a + \vec{e}_b$  is the angle bisector of these two path sections. The following applies:

$$(\vec{e}_a + \vec{e}_b) \cdot \dot{\vec{r}}(t_{1,2}) = 0$$

This angle bisector for  $t = t_{1,2}$  is therefore perpendicular to the tangent at these points and therefore the connecting lines  $AR$  and  $BR$  form the same angle with the tangent.

Since the billiard table is convex, the connections  $AR$  and  $BR$  lie completely inside the table and represent sections of the path of a ball that starts at  $A$ , is reflected at  $R$  and then hits  $B$ .

As a result, we have:

### Theorem 10.2

If  $A$  and  $B$  are any two points inside a convex billiard table. Then there are at least two reflection points  $R_{1,2}$  for  $A$  and  $B$  on the edge of the billiard table. At these points, the sum of the distances  $|\overrightarrow{AR}| + |\overrightarrow{BR}|$  has a local extremum.

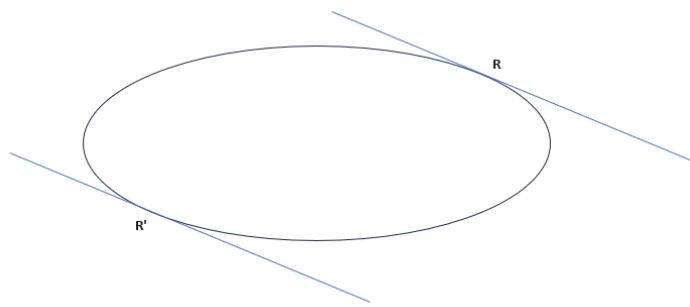
□

You can also analyse the case where  $A$  and  $B$  coincide. The same result also applies in this case: There are at least two reflection points, so that a ball that starts from  $A$  returns to  $A$  after the collision. The proof is an exercise.

In some examples, we were able to approximate start parameters that led to periodic orbits with interval nesting. The question is whether there are always periodic orbits for convex billiard tables.

Let us consider a (possibly non-convex) billiard table and a tangent to its edge at any edge point. Assume that there are points on both sides of the tangent that belong to the interior of the billiard table. Then it is possible to find points on different sides of the tangent whose connection is no longer completely inside the billiard table. The billiard table would then not be convex. Conversely, if the billiard table is convex and we place a tangent in any edge point, then all points that belong to the billiard table lie on the same side of the tangent.

Let's take a convex billiard table and place the tangent at an edge point  $R$ . If we let this edge point run around the edge of the table once, the tangent line will have rotated by  $2\pi$ . Due to the continuity of the derivative, there is therefore a boundary point  $R'$  at which the tangent is parallel to the original one. If we now consider such a pair of tangents, their distance is the "width" of the billiard table with respect to these tangents.



A pair of parallel tangents on a convex billiard table

The width of the billiard table as a function of the edge point  $R$  or thus the associated parameter  $t$  of the ellipse edge is a periodic, restricted real function and thus assumes a maximum and a minimum. This function can be represented as:

$$W(t, t') = |\vec{r}(t) - \vec{r}(t')|$$

At the extreme points, the derivative from  $W$  to  $t$  is zero. The following therefore applies:

$$\frac{d}{dt} W(t) = \frac{\vec{r}(t) - \vec{r}(t')}{|\vec{r}(t) - \vec{r}(t')|} \cdot \dot{\vec{r}}(t) = 0$$

The vector  $\overrightarrow{RR'}$  is therefore perpendicular to the tangent at the point  $R$ . As the tangent in  $R'$  is parallel to the tangent in  $R$ , this vector is also perpendicular to this tangent. This means that a ball emanating from  $R$  is reflected in  $R'$  and returns to  $R$ . This results in a 2-periodic path. As there are at least two extreme points, the following applies:

### Theorem 10.3

A convex billiard table has at least two 2-periodic paths. They correspond to the minimum and maximum width of the table.

□

Based on our experiments with the simulator, we can assume that it is possible to find periodic trajectories of every period for a convex billiard table. This is indeed the case, because

### Theorem 10.4 (George Birkhoff 1884 - 1944)

For a strictly convex billiard table, there are two geometrically different  $p$ -periodic orbits with winding number  $q$  for every pair of natural numbers  $(q, p)$  with  $p \geq 2$  and  $q \leq [(p-1)/2]$ .

□

$[(p-1)/2]$  here denotes the next smallest natural number less than or equal to  $(p-1)/2$ .

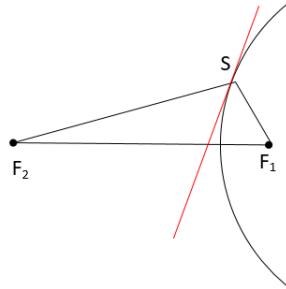
This theorem is a consequence of a much more general fixed-point theorem by Poincaré-Birkhoff, which cannot be explained here. The exact formulation of this fixed-point theorem and a proof can be found in [2]. Also, the application of this theorem to the convex billiard, from which theorem 10.4 follows.

## 11. Exercise examples

1. Examine the billiard in a rectangle. Mirror the rectangle on all sides and continue these reflections on the new rectangles. This technique is called *the unfolding method*. The ball track then becomes a straight line. What periodic paths are there? Do aperiodic paths fill the rectangle tightly? Can the path be interpreted as a path on a torus by "gluing" suitable sides of the rectangular grid together?
2. Examine the graph of the curve  $\gamma: t \in [0, 2\pi] \rightarrow \vec{\gamma}(t) = (2 + \sin t)(\cos pt, \sin pt)$  for  $p \in \mathbb{N}$ . How many double points are there? On which coordinate axis do they lie?

3. Prove this: In a circular billiard, the caustic is a circle with radius  $\cos\alpha$ , if  $\alpha$  is the angle of reflection.
4. In section 3 on elliptical billiards, we outlined the condition for calculating the second collision point with the parameter  $u$ . Perform the corresponding calculation explicitly.
5. An elliptical billiard table is given by  $\gamma: t \rightarrow \left( \begin{smallmatrix} 2\cos t \\ \sin t \end{smallmatrix} \right)$ ,  $[0, 2\pi] \rightarrow \mathbb{R}^2$ . A ball starts at the point  $t_1 = 0$  with the angle  $\alpha_1 = \frac{\pi}{4}$  relative to the first tangent at the point  $t_1 = 0$ . Determine the second point of collision and the second angle of collision using the calculations in section 3.
6. The hyperbola is defined as the location of all points in the plane, so that the amount of the difference between the two distances of two fixed focal points is constant. This constant is usually referred to as  $2a$ , analogous to the ellipse. Let  $P$  be a point on the hyperbola and  $F_1, F_2$  be the two focal points. Then both  $|PF_1| - |PF_2| = 2a$  and  $|PF_2| - |PF_1| = 2a$  are constant. The hyperbola therefore has two branches.

Instead of the "reflection" of a ray emanating from a focal point, this ray is reflected at the hyperbola tangent at the meeting point of the ray. After reflection, this ray then passes through the other focal point.



A focal ray is reflected at the hyperbolic tangent (red)

This means that in the figure above,  $\overline{SF_1}$  and  $\overline{SF_2}$  form the same angle with the hyperbolic tangent. Now consider the elliptical billiard if the ball path passes between the focal points. Show that in this case the caustic of the trajectory is a hyperbola.

Hint: Proceed in the same way as in the case of an elliptical caustic.

7. Given an elliptical billiard table with axis ratio  $\frac{b}{a} = 0.8$  and the usual parameter representation. A ball starts at the point  $t_1 = 0$ . Determine a starting angle  $\alpha$ , which leads to a three-periodic orbit, by experimenting with the simulator and an interval nesting method. Also investigate orbits of other periods.
8. In elliptical billiards, analyse the phase portrait of an orbit that runs as precisely as possible through the focal points of the ellipse.
9. Discuss the periodic orbits in oval billiards and billiards in the stadium. Examine low-period orbits. Use simulator experiments and interval nesting to approximate the associated start

angles.

10. Consider a billiard table that is bounded by a convex and continuously differentiable curve  $\gamma$ . Furthermore, let  $\vec{\gamma}(t)$  be positively orientated, i.e. the billiard table is always to the left of the curve tangent  $\dot{\vec{\gamma}}(t)$ . Let:

$\psi_n$  = the angle between the curve tangent at the nth collision point and the positive x-axis  
 $\alpha_n$  = the angle of reflection at the nth collision  
 $\varphi_n$  = the angle between the nth path section and the positive x-axis

Show that applies:

$$\begin{cases} \alpha_{n+1} = \psi_{n+1} - \varphi_n \\ \varphi_{n+1} = \psi_{n+1} + \alpha_{n+1} \\ \varphi_{n+1} = 2\psi_{n+1} - \varphi_n \end{cases}$$

11. Analyse the C-diagram for the oval billiard and examine any structures that appear to be periodic.
12. Let  $\gamma: I \subset \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \vec{\gamma}(t)$  be a regular and twice continuously differentiable plane curve. If the curve is traversed as a function of the parameter  $t$ ,  $\dot{\vec{\gamma}}(t)$  can be understood as a (tangential) velocity vector and  $\ddot{\vec{\gamma}}(t)$  as acceleration. This acceleration can be divided into a vector parallel and one perpendicular to the velocity vector:

$$\ddot{\vec{\gamma}}(t) = \vec{a}^{\parallel}(t) + \vec{a}^{\perp}(t)$$

Show: If  $\gamma$  is parameterised according to the arc length  $s$ , then the following applies:  $\vec{a}^{\parallel}(s) = 0, \forall s$ .

13. The usual parameter representation of the circle with radius  $r$  is:

$$\vec{\gamma}(t) = r \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, t \in [0, 2\pi[, r > 0$$

What does a parameter representation with the arc length as a parameter explicitly look like?

14. Let the point  $A$  be inside a convex billiard table. Prove: Then there are at least two reflection points on the edge of the billiard table such that a ball emanating from  $A$  that is reflected in a reflection point returns to  $A$ .

## Further reading

[1] Serge Tabachnikov: Geometry and Billiards, Springer Spektrum 2013

[2] Anna-Maria Vocke: The Poincaré-Birkhoff fixed point theorem and applications to billiards, Bachelor thesis, Mathematical Institute, Westfälische Wilhelms-Universität Münster, 2014