

# On nonconvex caustics of convex billiards

Oliver Knill

March 12, 1997, Elemente der Mathematics, in Press 1997

## Abstract

We illustrate the dynamics of billiards in tables of constant width. Caustics of such tables can be nowhere differentiable and are related to caustics appearing in differential geometry.

## 1 Introduction

The game of billiards in a general convex table is named after G.D. Birkhoff, who proposed the problem in 1927. Birkhoff used it as an illuminating example to illustrate results obtained in connection with the Newtonian three body problem. Birkhoff billiards are examples of dynamical systems, where results in analysis, calculus of variations or ergodic theory find applications. The problem is closely related to the Dirichlet problem on the domain enclosed by the table. Birkhoff viewed it also as a limiting case for the geodesic flows on two-dimensional surfaces.

In contrast from the real billiard game, where relatively short orbits are considered and other factors like the spin of the ball are relevant, the more academic Birkhoff billiard deals with trajectories of arbitrary length, hitting the boundary according to the precise law of reflection. An orbit is determined by the starting point  $T(s)$  of the table and the angle  $\theta$  between the velocity direction and the tangent at the point  $T(s)$ .

If we shoot a billiard ball and follow the individual angles of impact, it is natural to ask what is the minimal or maximal angle which occurs along an orbit. Is it possible to shoot in such a way that a ball hits the boundary in the future with angles arbitrarily close to 0 and at other times with angles arbitrarily close to  $\pi$ ?

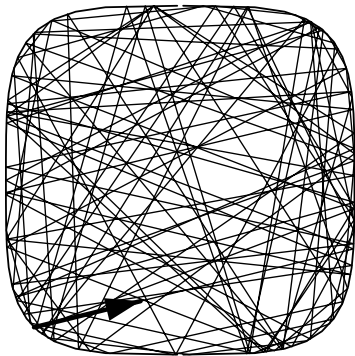


Figure 1. Playing Birkhoff billiards. A trajectory of a ball in the table  $x^4 + y^4 = 1$ . What is the maximal or minimal impact angle, which can occur along orbits in this table? According to G. Birkhoff, the answer to this question is related to the existence of caustics (which, by the way, do not exist for this table).

Birkhoff realized that such an orbit does not exist if and only if there is a continuous invariant curve  $\Gamma = \{(s, \theta(s))\}$  in the phase space of all pairs  $(s, \theta)$ . Such an invariant curve  $\Gamma$  in the phase space is associated to a caustic  $\gamma$  in the interior of the table. The curve  $\gamma$  has the property that if the path of a ball has a tangency at  $\gamma$ , then also the reflected path has a tangency at  $\gamma$ .

An aim of our article is to look closer at the geometry of caustics and to illustrate the story with some pictures. By relating the caustics of a class of billiards with differential geometrical caustics, we will see that this subject is of more general interest. Differential geometric caustics are relevant in applied sciences including observational astronomy and earthquake science: assume that a light source is switched on in

a nonhomogeneous or nonflat medium or assume that an earthquake starts at its epicenter. The set of points which are reached by the light or the sound wave after a given time is called a wave front. For small times, such a wave front is diffeomorphic to a sphere. If the medium is not homogeneous or not flat, these spheres become more and more deformed as time increases and there will be in general points where the front ceases to be an immersed surface, for example points where different parts of the front selfintersect. The set of all points, where the immersion fails, is called the caustic of the light or acoustic source. As we will illustrate, there can be relations between differential geometric caustics and caustics in billiards.

Our story, which begins in the next section can be read first in the fast lane by following the text to the figures.

## 2 Convex billiard tables

Good starting points in the literature of the billiard problem are [23] [4] [24] or [21]. In the case, when the domain is a strictly convex planar region, the return map to the boundary  $T$  defines a map  $\phi$  of the annulus  $A = \mathbf{T} \times [0, 1]$ . The circle  $\mathbf{T}$  parametrizes the table with impact point  $s$  at the boundary  $T$ . The angle of impact  $\theta$  and  $s$  determine the next reflection point  $s_1$  with angle  $\theta_1$ . The reader can see some orbits of the map  $\phi : (s, \cos(\theta)) \mapsto (s_1, \cos(\theta_1))$  in Figures 3 and 4.

Noncontractible invariant curves  $\Gamma$  in  $A$  can define caustics  $\gamma$  in the interior of the table  $T$  (see Figure 2). Caustics are curves with the property that a ball, once tangent to  $\gamma$  stays tangent after every reflection. We will discuss the precise relation between the invariant curve  $\Gamma$  and the caustic  $\gamma$  in the next section. Examples of caustics are drawn in Figures 6,7 and Figure 11a.

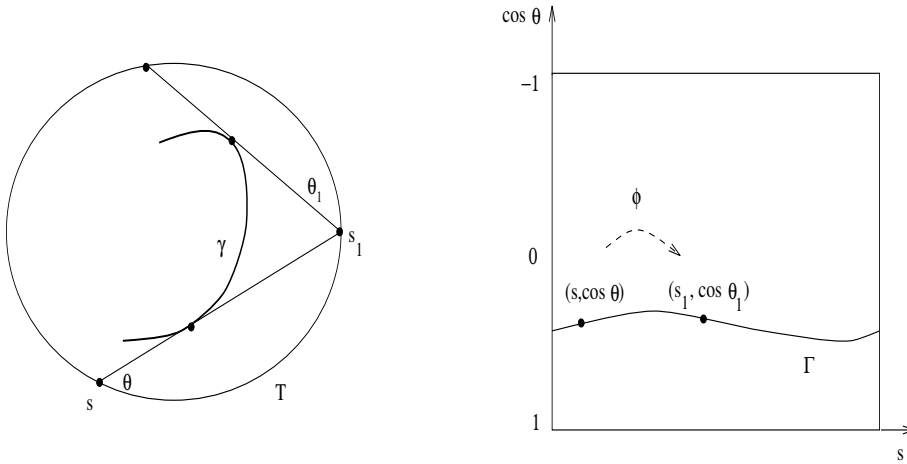


Figure 2. If there is an invariant curve  $\Gamma$  of the billiard map  $\phi$ , one can find a caustic  $\gamma$  in the region bounded by the table  $T$ . The curve  $\gamma$  has the property that a trajectory, once tangent at some point to it, stays tangent to it after the reflection.

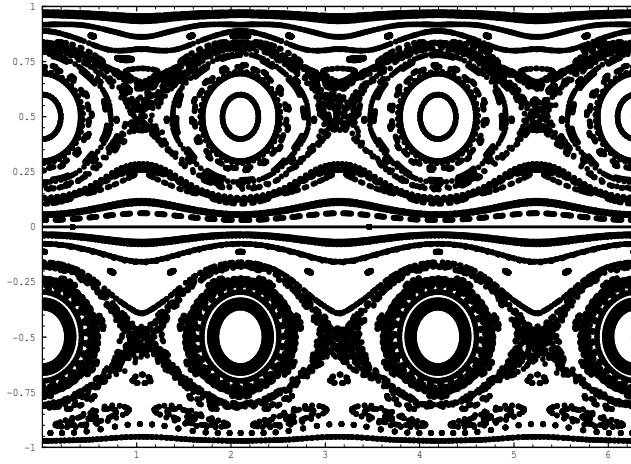


Figure 3. Some numerically computed orbits of a billiard map  $\phi : A \rightarrow A$ . The table is a curve of constant width. Together with the invariant curve  $\{\cos(\theta) = 0\}$  containing periodic orbits of period 2, one can see other invariant curves.

Invariant curves  $\Gamma$  of  $\phi$  are interesting for the dynamics because they divide the phase space  $A$  into invariant regions and therefore prevent ergodicity. They often occur near the boundary of a smooth, strictly convex table and define 'Lazutkin whisper galleries' [8]. Most convex billiards have no invariant curves [11]. Also, if the curvature of the table is unbounded at some point, then there are no invariant curves in a neighborhood of the boundary of  $A$  [15]. Strict convexity is necessary for the existence of caustics [17]. The billiard in Figure 1 for example has no caustics.

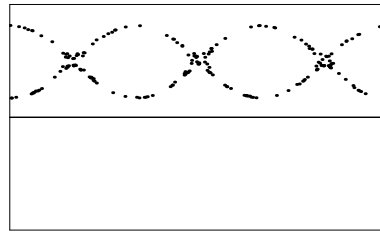
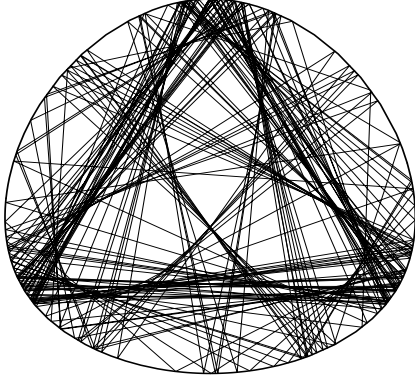


Figure 4a. A finite piece of a chaotic billiard trajectory. The drawing on the right shows the corresponding orbit in the phase space  $A$ .

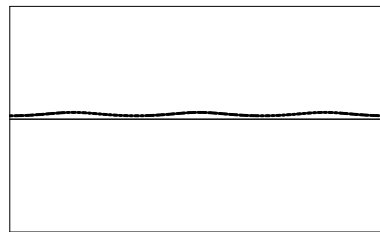
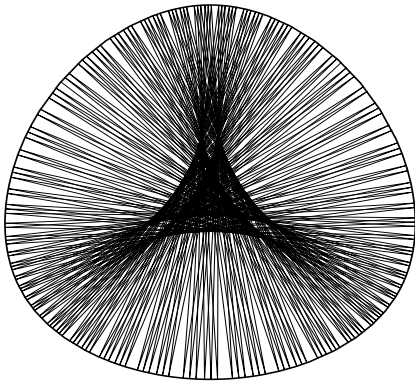


Figure 4b. A finite piece of a billiard trajectory with steep initial impact angle. In the phase space  $A$  the orbit is near the invariant curve  $\{\theta = \pi/2\}$ .

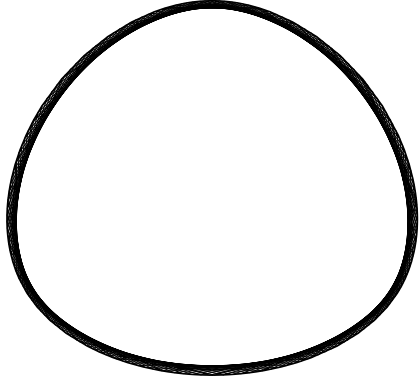


Figure 4c. A finite piece of a "whisper gallery" trajectory with angles of impact near  $\pi$ . The orbit of  $\phi$  in phase space  $A$  is located near the invariant boundary.

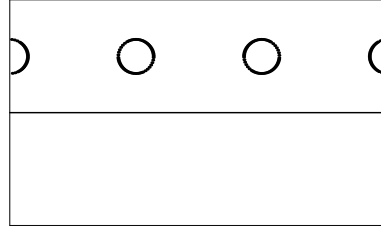
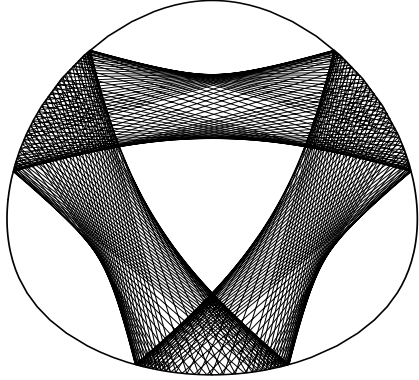


Figure 4d. A trajectory near a linearly stable periodic orbit of period 3.  $\phi^3$  leaves contractible curves invariant which enclose the fixed points of  $\phi^3$ . KAM theory can prove the existence of such curves [18].

We will focus in this article especially on billiard tables of constant width. They all leave the curve  $\Gamma = \{\theta = \pi/2\}$  invariant. Figures 3 and 4 show typical trajectories in such tables. The flatness of the invariant curve  $\Gamma$  will allow us to make statements about the regularity of the corresponding caustics  $\gamma$  depending on the regularity of the table. First however, in the next section, we will clarify the relation between the invariant curve  $\Gamma$  in the phase space and the caustic  $\gamma$  in the interior of the table.

### 3 Caustics of convex billiards

A bounded measurable positive function  $\rho \in L^2(\mathbf{T})$  satisfying  $\int_0^{2\pi} \rho(\beta)e^{i\beta} d\beta = 0$  defines a strictly convex table  $T$  by

$$T(\alpha) = T(0) + \int_0^\alpha \rho(\beta)e^{i\beta} d\beta. \quad (1)$$

$\rho(s)$  is the radius of curvature at the point  $T(s)$ .

Douady remarked that a  $k$ -times differentiable invariant curve  $\Gamma$  of a  $(k+1)$ -times differentiable table defines a  $(k-1)$ -times differentiable caustic [8]. In order to reformulate this result also for intermediate degrees of smoothness, it is convenient to pick  $\rho$  in one of the Hilbert spaces

$$H^\mu(\mathbf{T}) = \{f \in L^2(\mathbf{T}) \mid f(s) = \sum_{k \in \mathbf{Z}} a_k e^{iks}, \|f\|_\mu = \sum_{k \in \mathbf{Z}} |k|^{2\mu} |a_k|^2 < \infty\}$$

which are all contained in the Hilbert space  $L^2(\mathbf{T})$  of square integrable periodic functions. For  $\mu > 1/2$ , a function in  $H^\mu(\mathbf{T})$  has a uniformly convergent Fourier series  $\rho(s) = \sum_{k \in \mathbf{Z}} a_k e^{iks}$ . Moreover, if  $\mu > 1/2 + r$  with  $r \in \mathbf{N}$ , then  $f \in H^\mu(\mathbf{T})$  is  $r$  times continuously differentiable. We use here  $H^\mu$  mainly for notational reason in order to control with a single parameter  $\mu$  both nowhere-differentiable and smooth situations. Not much is lost for a reader who decides to ignore the spaces  $H^\mu$ .

**Lemma 3.1 (R. Douady)** *If both the invariant curve  $\Gamma$  and  $\rho > 0$  are in  $H^\mu(\mathbf{T})$  with  $\mu > 1$ , then  $\gamma$  is in  $H^{\mu-1}(\mathbf{T}, \mathbf{C})$  and given by the formula*

$$\gamma(s) = T(s) + \frac{\sin(\theta(s))}{\kappa(s) + \Gamma'(s)}u(s)$$

Here  $u(s)$  is a unit vector in the direction of the trajectory (see Figure 5),  $\theta(s)$  is the angle between  $u(s)$  and the tangent at  $T(s)$ , and  $\kappa(s) = \rho(s)^{-1}$  is the curvature at  $T(s)$ .

*Proof.* Let  $\delta(s)$  be the angle between the  $x$ -axis in  $\mathbf{R}^2$  and the trajectory connecting  $T(s)$  with  $T(s_1)$ . We have  $\gamma(s) = T(s) + b(s)u(s)$ , where  $b(s) = \sin(\theta)(d\delta/ds)^{-1}$  (see Figure 5). If  $\alpha(s)$  is the angle of the tangent at the point  $T(s)$ , then  $\delta(s) = \theta(s) + \alpha(s)$  and  $\delta' = \theta' + \alpha'(s) = \kappa(s) + \Gamma'(s)$ , so that  $b(s) = \sin(\theta)(\kappa(s) + \Gamma'(s))^{-1}$ .

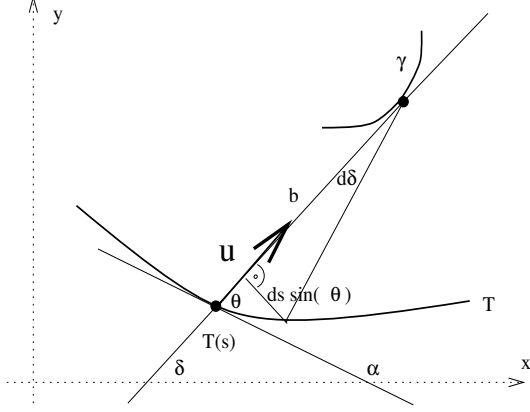


Figure 5. The relation  $b(s) = \sin(\theta)(d\delta/ds)^{-1}$  in the proof of Lemma 3.1. When inserting this relation into the formula  $\gamma = T + bu$ , it gives the formula for the caustic  $\gamma$  in terms of the table  $T$  and the invariant curve  $\Gamma$  in the phase space.

The curvature function  $\kappa$  is in  $H^{\mu-1}$  because the table is in  $H^{\mu+1}$ . Also  $\Gamma'$  is in  $H^{\mu-1}$  by assumption. The strict convexity of  $T$  implies  $\delta' > 0$  and so  $\kappa + \Gamma' > 0$ . Therefore  $\gamma \in H^{\mu-1}$ .  $\square$

In [14] the authors ask how to define, in general, a caustic of a convex two times differentiable table whose invariant curve is only Lipschitz. The equation in Lemma 3.1 does this for a table with invariant circle  $\Gamma$  and bounded measurable curvature function  $\kappa$ . Because the derivative  $\Gamma'(s)$  is also a bounded measurable function by Birkhoff's theorem (for a proof see [16] p.430), the image of the function  $\gamma : \mathbf{T} \rightarrow \mathbf{C}$  is in general a measurable set in the complex plane.

## 4 Caustics of tables of constant width

If  $\Gamma = \{\cos(\theta) = 0\}$  is an invariant curve of the billiard map, then each trajectory belonging to an orbit on  $\Gamma$  both leaves and hits the table orthogonally. The period of the orbit is then 2. Because each of these orbits is a critical point of the length functional, the distance between the two impact points is constant and the table  $T$  is consequently a curve of constant width. It is more convenient to parametrize such a table by the angle  $\alpha$  of the tangent to the boundary instead of using the arc length  $s$  of the table. The table is of constant width if  $\rho$  satisfies  $\rho(\alpha) + \rho(\alpha + \pi) = \text{const}$ . If  $T \in H^{\mu+1}$ , then  $\rho \in H^{\mu}$ . In the case, when the invariant curve  $\Gamma$  satisfies  $\Gamma'(s) = 0$ , the formula in Lemma 3.1 simplifies to

$$\gamma(\alpha) = T(\alpha) + \rho(\alpha)ie^{i\alpha}. \quad (2)$$

which is the formula for an evolute of the table  $T$  (see [5]). If we plug in the Fourier series  $\rho(\alpha) = \sum_{k \in \mathbf{Z}} a_k e^{ik\alpha}$ , we get

$$\gamma(\alpha) = T(0) - i \left( \sum_{k \in \mathbf{Z}} \frac{a_k}{(k+1)} (e^{i(k+1)\alpha} - 1) \right) + i \sum_{k \in \mathbf{Z}} a_k e^{i(k+1)\alpha} = \gamma(0) + i \sum_{k \in \mathbf{Z}} \frac{k}{k+1} a_k (e^{i(k+1)\alpha} - 1).$$

We need  $a_1 = 0$  to assure that the table is closed. The condition  $\rho(\alpha) + \rho(\alpha + \pi) = 2a_0$  holds if  $a_{2k} = 0$  for  $k \neq 0$ . For  $\rho$  to be real, we have to require also  $a_{-k} = \bar{a}_k$ . We summarize:

**Proposition 4.1** Given a positive function  $\rho(\alpha) = \sum_{k \in \mathbf{Z}} a_k e^{ik\alpha} \in H^\mu(\mathbf{T})$  with  $a_{-k} = \overline{a_k}$  satisfying  $a_1 = 0$  and  $a_{2k} = 0$  for  $k \neq 0$ . It defines the  $H^{\mu+1}$  table of constant width

$$T(\alpha) = T(0) - i \sum_{k \in \mathbf{Z}} \frac{a_k}{k+1} (e^{i(k+1)\alpha} - 1).$$

The corresponding billiard map  $\phi$  has the invariant curve  $\Gamma = \{\cos(\theta) = 0\}$  which defines the  $H^\mu$  caustic

$$\gamma(\alpha) = \gamma(0) + i \sum_{k \in \mathbf{Z}} \frac{k}{k+1} a_k (e^{i(k+1)\alpha} - 1)$$

where  $\gamma(0) = T(0) + ia_0$ .

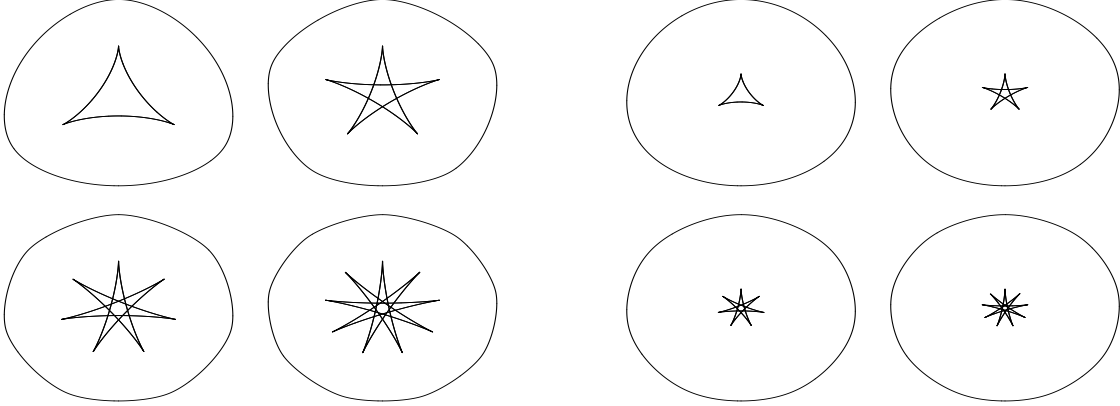


Figure 6a. Caustics of tables of constant width given by the curve determined by  $\rho(\alpha) = a_0 + \cos((1 + 2k)\alpha)$  for  $k = 1, 2, 3, 4$  with  $a_0 = 4$ .

Figure 6b. Since  $a_0$  is arbitrary, as long as  $\rho > 0$ , the tables are not unique. In this figure, we see the same caustics as in Figure 6a) but with tables using  $a_0 = 10$ .

From formulas (1) and (2), we get  $\gamma'(\alpha) = \rho'(\alpha)ie^{i\alpha}$  which shows that  $\rho$  has the same critical points as  $\gamma$ . In other words, the vertices of  $T$  correspond to cusps of  $\gamma$ .

The caustics in Figures 6,7 were drawn using trigonometric polynomials  $\rho(\alpha)$ . While the functions  $\gamma$  were real analytic, in general, the image of  $\gamma$  has singularities. Indeed, the function  $\rho'$  has at least  $N$  zeros if it is a trigonometric polynomial of degree  $N$  [22]. If the zeros are simple, the curve  $\gamma$  has at least  $N$  cusps counted with multiplicity. Note that for the caustics belonging to the invariant curve  $\{\theta = \pi/2\}$ , the map  $\gamma$  is 2 to 1.

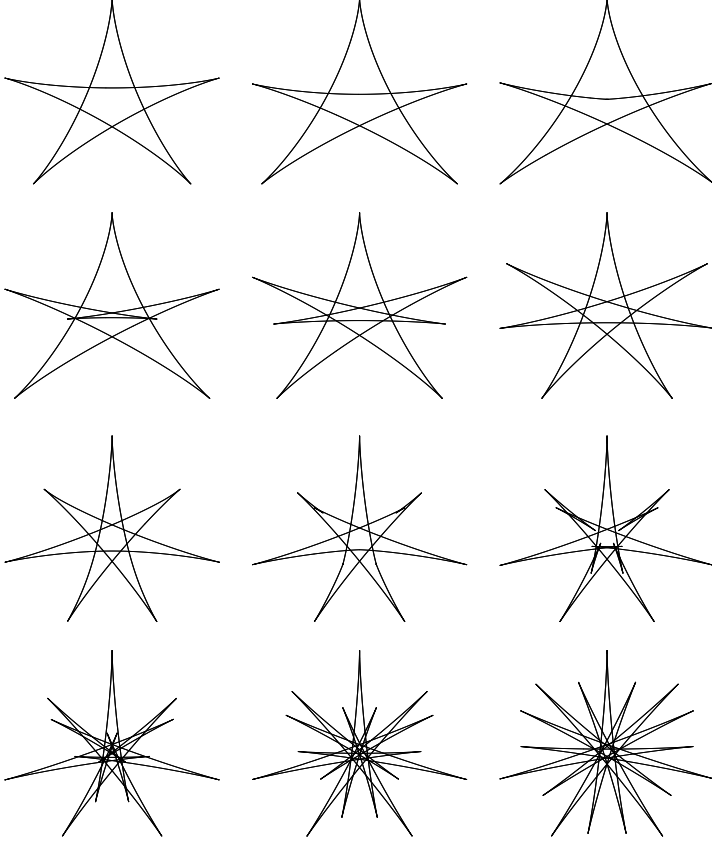


Figure 7a. Caustics of tables with  $\rho(\alpha) = a_0 + g \cos(5k\alpha) + (1 - g) \cos(7k\alpha)$  for  $g = k/7$ ,  $k = 1, \dots, 6$ . We see an interpolation between the caustics with 5-fold and 7-fold symmetry in Figure 6a.

Figure 7b. Examples of caustics obtained from tables determined by  $\rho(\alpha) = a_0 + g \cos(7k\alpha) + (1 - g) \cos(17k\alpha)$  for  $g = k/7$ ,  $k = 1, \dots, 6$ . We see an interpolation between caustics with 7-fold and 17-fold symmetry.

## 5 Nowhere differentiable caustics

**Corollary 5.1** *There exist continuous caustics  $\gamma$  of tables of constant width which have both coordinate functions of  $\gamma$  nowhere differentiable except perhaps at the points  $\gamma(\alpha)$  with  $\alpha = k\pi/2$ ,  $k = 0, \dots, 3$ .*

*Proof.* Let  $1/2 < \mu < 1$ . Take  $a_k = k^{-\mu}$  for  $k = 2^n$  and  $a_k = 0$  otherwise. This Weierstrass function  $\rho(\alpha) = \sum_k a_k e^{ik\alpha}$  is in  $H^{\mu'}$  for every  $\mu' < \mu$  and so continuous. By Lemma 3.1, the corresponding caustic is also continuous. Up to a translation and a rotation in the plane, we have  $\gamma(\alpha) = \sum_{k \in \mathbf{Z}} \frac{k}{k+1} a_k \exp(i(k+1)\alpha)$  so that

$$\begin{aligned}
 \operatorname{Re}(\gamma)(\alpha) &= \sum_{k=2}^{\infty} \frac{k}{k+1} a_k \cos((k+1)\alpha) + \sum_{k=2}^{\infty} \frac{k}{k-1} a_k \cos((k-1)\alpha) \\
 &= \sum_{k=2}^{\infty} a_k \left( \frac{k}{k+1} \cos(2\alpha) + \frac{k}{k-1} \right) \cos((k-1)\alpha) - \frac{k}{k+1} \sin(2\alpha) \sin((k-1)\alpha) \\
 &=: \sum_{k=2}^{\infty} a_k h_k(\alpha).
 \end{aligned}$$

Using  $a_{2^l} h_{2^l}(\alpha_{m,\pm}) = \mp 2^{-l\mu} \frac{2^l}{2^l+1} \sin(2\alpha_{m,\pm})$  for  $\alpha_{m,\pm} = \pm \frac{\pi}{2} 2^{-l} + m2\pi 2^{-l}$ ,  $m \in \mathbf{Z}$ , it is not difficult to see that  $\operatorname{Re}(\gamma)$  is nowhere differentiable except maybe at the points  $\alpha = 0, \pi$ . There is an analogous

statement for  $\text{Im}(\gamma)$ , where we have to exclude the points  $\alpha = \pm\pi/2$ .  $\square$

Note that the symmetry of the curves is responsible for possible differentiability at some finite set of points. The symmetry is broken, if the  $a_k$  are no longer real. We expect then examples with nowhere differentiability without exceptions. Because by a theorem of Banach, nowhere differentiable functions are Baire generic among continuous functions [19], we expect also that a "general" curve of constant width produces nowhere differentiable caustics. Finally, we remark that because every convex curve which is not too long can be extended to a curve of constant width, there are many tables for which the caustic is nowhere differentiable in some interval.

## 6 Nowhere differentiable lines of striction of ruled surfaces

Let us place the billiard table in the plane  $\{z = 0\}$  in the three-dimensional Euclidean space  $\mathbf{R}^3$ . We now also let the billiard ball move in the  $z$  direction by giving it a  $z$ -velocity  $1/l$  if  $l$  is the length of the trajectory between successive impacts  $\{(T(s_k), z_k)\}$  on the boundary of the two-dimensional cylinder table  $T \times \mathbf{R}$ . The special choice of the  $z$  velocity has the consequence that a ball starting at a point  $(s, \theta, z)$  of the cylinder hits the cylinder again in the point  $(\phi(s, \theta), z + 1)$ . The union of all billiard ball trajectories which were tangent to a caustic  $\gamma$  now form a ruled surface, a one-parameter family of straight lines.

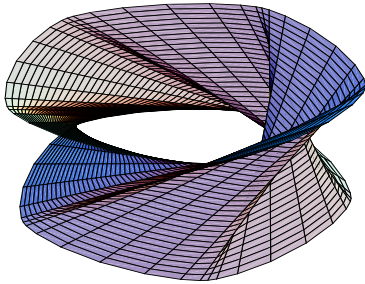


Figure 8a. A ruled surface constructed from a billiard table with  $\rho(\alpha) = p + \cos(5\alpha)$ .  $T$  has 5 flat invariant curves  $\Gamma = \{\theta = \theta_i\}$ , where  $\theta_i$  solves  $\tan(5\theta) = 5 \tan(\theta)$  [13]. The displayed ruled surface belongs to one of these invariant curves different from  $\{\theta = \pi/2\}$ .

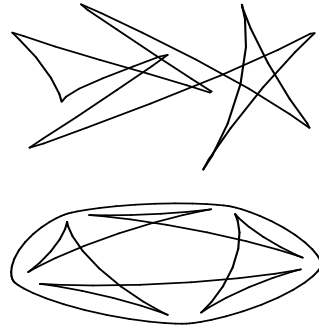


Figure 8b. The line of striction of the surface in Figure 8a) projects to the caustic of the table. For better visibility, the line of striction has been lifted up a bit in this drawing.

The line of striction of the ruled surface constructed from the table with a caustic projects along the  $z$  direction (see Figure 8a) to the caustic. However, unlike the caustic itself, the line of striction does not need to have self intersections. The example in Figure 8a illustrates this.

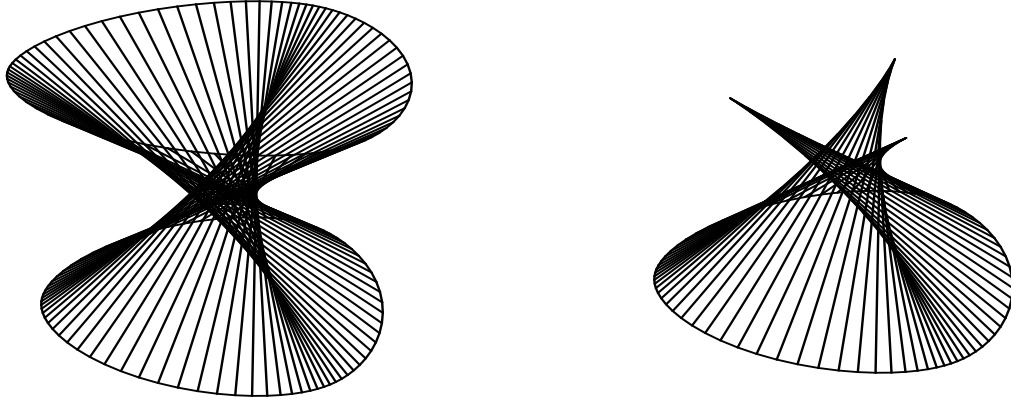


Figure 9. The ruled surface obtained from a table given by  $\rho(\alpha) = a_0 + \cos(3\alpha)$ . The invariant curve is the equator. The projection of the line of striction is the caustic seen in Figure 6a. In the figure to the right, the surface has been drawn only until the line of striction.

Consider a nowhere differentiable caustic as constructed above. The two first coordinates of the line of striction are the two coordinates of the caustics and are therefore nowhere differentiable functions. Because the third coordinate of the line of striction is  $\rho(\alpha)$ , which is a nowhere differentiable Weierstrass function, all coordinate functions are nowhere differentiable.

**Corollary 6.1** *There exist differentiable ruled surfaces in  $\mathbf{R}^3$ , whose line of striction is a curve which has all coordinate functions nowhere differentiable (except perhaps at finitely many points).*

## 7 Nowhere differentiable caustics on spheres

The caustic of a point on a Riemannian manifold is defined as the set of intersection points of infinitesimally closed geodesics at this point. A motivating picture for studying billiards is that the billiard flow can be considered as a limiting case of the geodesic flow. One can now ask, whether caustics of billiards can be related to caustics of a family of geodesics starting at a point in a two-dimensional Riemannian manifold  $M$ . Here, in the case of billiards on tables of constant width, an orbit tangent to the caustic can actually be extended to an orbit of a geodesic flow on a sphere.

**Corollary 7.1** *There exists a metric  $g$  on the two sphere  $S^2$  and a point  $P \in S^2$  such that the caustic of this point has nowhere differentiable coordinate functions (except maybe at four points).*

*Proof.* Take the Northern hemi-sphere with the usual metric and let  $P$  be the North pole (see Figure 10.) A geodesic trajectory starting at  $P$  hits the equator at a right angle. Deform the south-turning-circle below the equator into a curve  $T$  of constant width and fill the South cap with a flat metric. The remaining annulus  $A$  is parametrized by smooth coordinates  $(\phi, u) \in \mathbf{T} \times [0, 1]$  such that the boundary  $\{u = 0\}$  is the equator and  $\{u = 1\} = T$ . We chose a metric in  $A$  such that a geodesic entering  $A$  orthogonally at the equator leaves  $A$  orthogonally to  $T$  and returns to  $P$  after having been tangent to the caustics  $\gamma$ .  $\square$

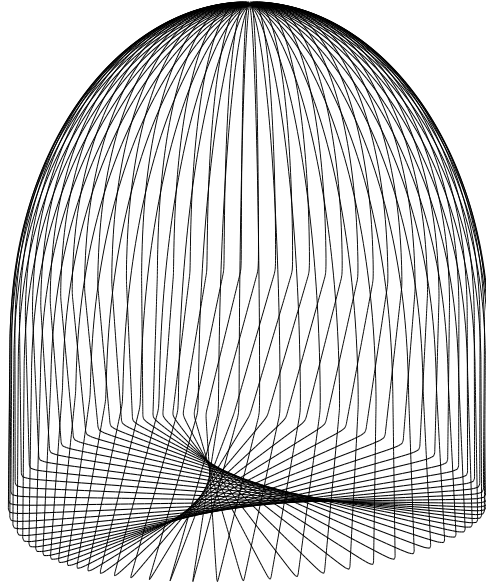


Figure 10. Geodesics starting on the North pole  $P$  of a deformed sphere  $M$  with a Riemannian metric  $g$ . The caustic of  $P$  on  $M$  is a caustic of a table  $T$  of constant width. Every geodesic starting at  $P$  is closed. They have the same length illustrating the more general fact that all closed geodesics through a point on a general two-dimensional surface have the same length [10]).

## 8 Are there examples of fractal caustics?

We have seen that the coordinate functions of caustics  $\gamma$  can be "rough". Can the image of  $\gamma$  become "fractal"? More precisely one can ask: do there exist convex tables  $T$  such that the image of  $\gamma$  has Hausdorff dimension  $\mu$  for some  $\mu > 1$ ? Corresponding questions for graphs of real functions can be difficult and often one is only able to estimate the box-counting dimension (see the case of the Weierstrass function in [9]).

We could try to find the distortion of the map from the graph  $G = \{(\alpha, \rho(\alpha))\} \subset \mathbf{R}^2$  of  $\rho$  to  $\gamma$ . Let  $T(\alpha) = (f(\alpha), g(\alpha))$  be the table defined by a function  $\rho \in H^\mu$ . If  $\rho$  is a Weierstrass function then the graph  $G$  of  $\rho$  has box-counting dimension  $> 1$  [9]. The transformation  $\phi : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ :  $\phi(\alpha, \beta) = (f(\alpha) - \sin(\alpha)\beta, g(\alpha) + \cos(\alpha)\beta)$  maps  $G$  onto the caustic  $\gamma$ . If it were invertible near  $\gamma(\mathbf{T})$ , the image of  $\gamma$  would inherit the fractal properties of  $G$ . Let's see: the Jacobian of  $\phi$  is

$$D\phi(\alpha, \beta) = \begin{pmatrix} f'(\alpha) - \beta \cos(\alpha) & -\sin(\alpha) \\ g'(\alpha) - \beta \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

Because  $T'(\alpha) = \rho(\alpha)e^{i\alpha}$ , one has  $f'(\alpha) = \cos(\alpha)\rho(\alpha)$ ,  $g'(\alpha) = \sin(\alpha)\rho(\alpha)$  and unfortunately,  $\det D\phi$  vanishes exactly on  $G$ . While this approach failed to give caustics with fractional box counting dimension, the calculation shows why: the caustic of a curve of constant width is the place, where the wave front formed by the ray of light starting orthogonally to the table of constant width fails to be an immersed submanifold.

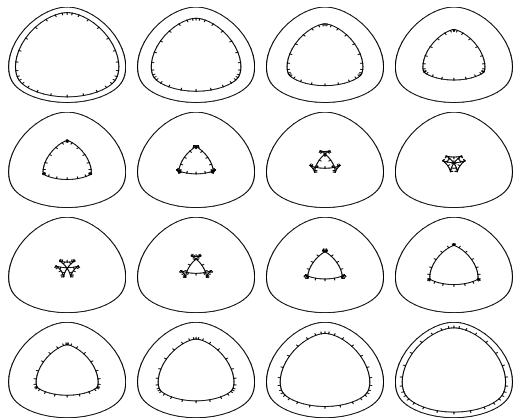


Figure 11a. Legendre collapse: wave fronts from the table with  $\rho(\alpha) = a_0 + \cos(3\alpha)$  radiate from the table and reverse inside out. The wave fronts are the contour lines of the ruled surface of the caustic  $\gamma$ .

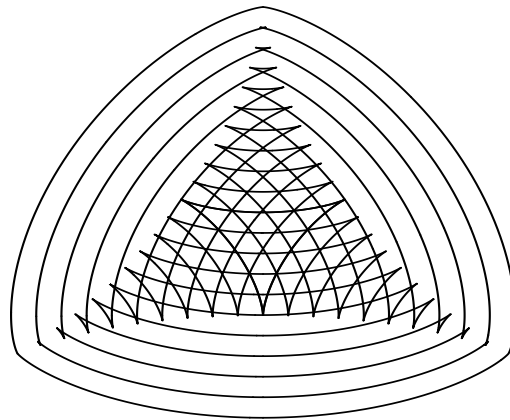


Figure 11b. A simultaneous view of the wave fronts in Figure 11a indicates the location of the caustic. It is a general fact that some of these wave fronts should have at least four cusps [1].

When plotting caustics using larger and larger frequencies in  $\rho$ , the curves usually get more and more cusps (see Figure 12a). From the pictures we can expect that the box-counting dimension or even the Hausdorff dimension can be bigger than 1 for some caustics but we do not know whether this is the case.

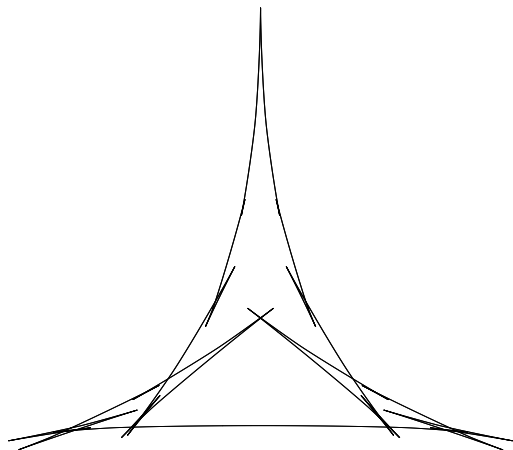


Figure 12a. A caustic  $\alpha \mapsto \gamma(\alpha)$  of a table approximating the table constructed in the proof of Corollary 5.1 using the Weierstrass function  $\rho$ .

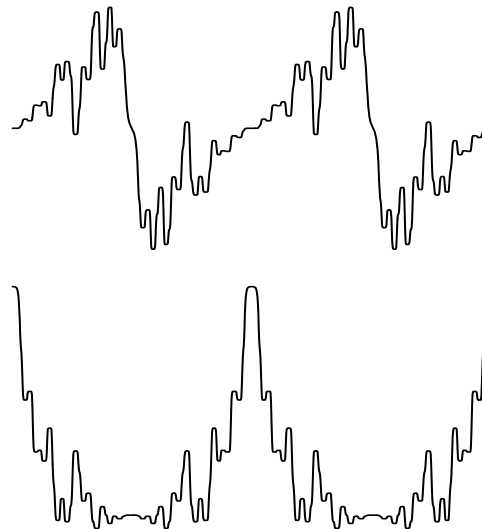


Figure 12b. The coordinate functions of the caustic  $\gamma$  drawn in Figure 12a. The upper graph is the function  $\alpha \mapsto \text{Re}(\gamma)(\alpha)$ , the lower graph is the function  $\alpha \mapsto \text{Im}(\gamma)(\alpha)$ .

## 9 Caustics and billiards in three dimensions

A three-dimensional Birkhoff billiard is defined by a differentiable convex surface  $R$  in  $\mathbf{R}^3$ . The return map to the boundary defines a map  $\phi$  on a four-dimensional manifold  $R \times R$ : two successive impact points  $(r_{n-1}, r_n)$  define the next pair  $(r_n, r_{n+1})$ . Caustics for such billiards do not exist [3] [12]. One can consider however differential geometric caustics of three-dimensional billiards which are defined through the wave front (see Figure 16). We now illustrate this in the case of a surface of equal thickness. Take a curve  $T$  of constant width in the  $x - z$  plane in  $\mathbf{R}^3$ . Assume that  $T$  is reflection symmetric with respect to the  $z$ -axis  $L$ . If we rotate  $T$  around  $L$ , we obtain a surface  $R$  of revolution which is at the same time a surface of constant width (see [6]). A billiard trajectory orthogonal to a point  $P$  of  $R$  intersects the surface  $R$  a second time orthogonally. It corresponds to a 2-periodic orbit of the return map  $\phi$ . By symmetry, a two-dimensional differential geometrical caustic of a three-dimensional billiard is obtained by rotating the caustic  $\gamma$  of  $T$  around  $L$ .

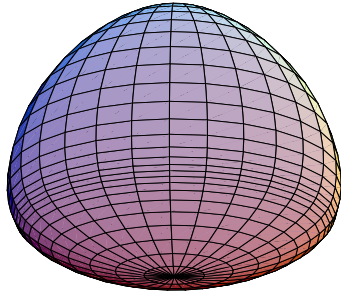


Figure 13a. A surface of constant width in the three dimensional space can be obtained by rotating a curve  $T$  of constant width around a symmetry axes  $L$ .

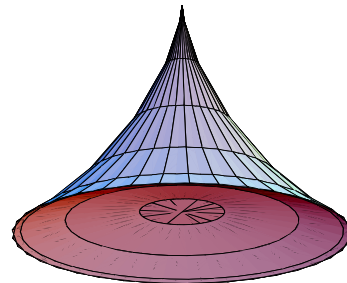


Figure 13b. The differential geometric caustic to the surface of constant width is obtained by rotating the caustic  $\gamma$  of  $T$  around the symmetry axes  $L$ .

The symmetry of the surface of revolution leads to a billiard map  $\phi$  which has an angular momentum integral  $I$ . One has to understand the dynamics of  $\phi$  on three-dimensional leaves  $I = const$ . When the angular momentum is zero, an orbit moves in a plane through the rotation axis  $L$ . The three-dimensional zero angular momentum leaf is foliated by two-dimensional invariant manifolds on which the dynamics is the two-dimensional billiard map (see Figure 14).

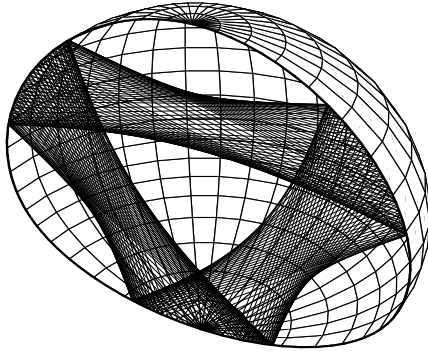


Figure 14. Trajectories with zero angular momentum in a three-dimensional billiard of constant width move in two-dimensional planes, where they are obtained by iterating a two-dimensional billiard map.

Having a two-dimensional Birkhoff billiard as a subsystem of the three-dimensional billiard implies the existence of many periodic orbits, more than are known to exist in general [2]. Also, by the variational theorem in ergodic theory, the topological entropy of the three-dimensional billiard map is bounded below by the topological entropy of the two-dimensional billiard. It would be interesting to know more about the dynamics of  $\phi$  on a nonzero angular momentum leaf. Figure 15 shows some typical numerically computed trajectories.

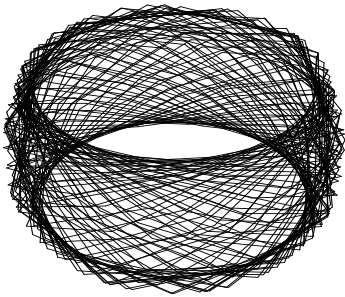


Figure 15a. Playing billiards in a surface of constant width. The picture shows a finite piece of a trajectory in the surface of revolution determined by  $\rho(\alpha) = 4 + \cos(3\alpha)$ .

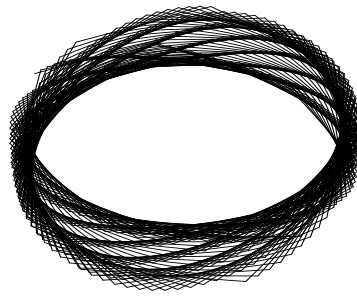


Figure 15b. A finite piece of another orbit with slightly larger angular momentum in the same surface as in Figure 15a.

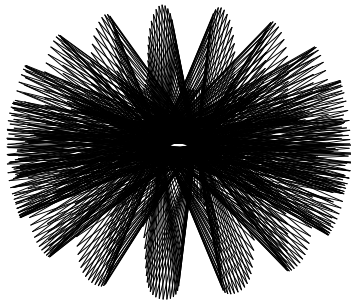


Figure 15c. This piece of a trajectory seems to lie on a two-dimensional torus in the three-dimensional leaf of constant angular momentum.

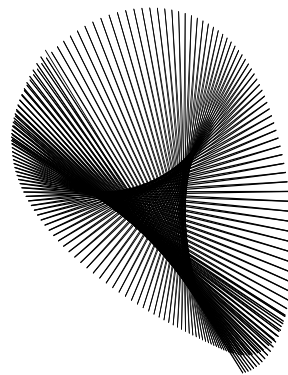


Figure 15d. A piece of an orbit with relatively small angular momentum.

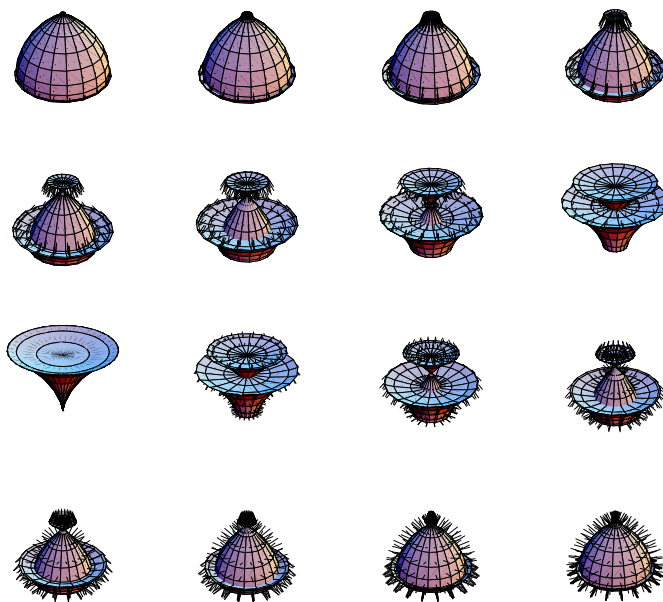


Figure 16. The standard Legendre collapse of a wave front moving from a surface of revolution of constant width. Two-dimensional sections of this collapse are shown in Figure 11a. The surface turns inside out during this collapse as can be seen by the arrows which point inside the surface at the beginning and outside at the end. Differing from the famous movie, this "turning the sphere inside out" is not smooth: the smoothness fails on points of the caustic.

## Appendix: Is there a nonintegrable billiard with an algebraic return map?

For the illustrations in Figure 3 and Figure 4, we constructed the map  $\phi : (\alpha, \theta) \mapsto (\alpha', \alpha' - \alpha - \theta)$  by finding  $\alpha' \neq \alpha$  so that  $(T(\alpha) - T(\alpha'))e^{-i(\alpha+\theta)}$  is real. This problem is equivalent to solving

$$\sum_{k \in \mathbf{Z}} \frac{a_k}{k+1} (\cos(k\alpha + \theta) - \cos((k+1)\alpha' - \alpha - \theta)) = 0.$$

Numerically, the second root of this equation can be found efficiently with a Newton method. (For three-dimensional billiards, the situation is similar but slightly more complicated). In the simplest case  $\rho(\alpha) = a_0 + \cos(3\alpha)$ , one has to solve

$$\cos(\alpha - 4x + \theta) - \cos(3\alpha - \theta) + 4a_0 \cos(\alpha - x + \theta) - 4a_0 \cos(\theta) + 2 \cos(3\alpha + \theta) - 2 \cos(\alpha + 2x + \theta) = 0$$

for  $x = \alpha'$ . We do not know, whether the second root  $x \neq \alpha$  is an algebraic expression in  $\alpha$  and  $\theta$ . One has the problem to find roots of a specific class of polynomials of degree eight. More generally, one can ask, whether there exists a table of constant width different from the circle, for which the return map is algebraic in some coordinates. We do not know of any smooth convex billiard table different from an ellipse, where the return map  $\phi$  is an algebraic expression in some coordinates. This question is related to the open Birkhoff-Poritsky conjecture [20] which claims that the ellipse is the only integrable smooth convex billiard. In [7] it was suggested that curves of equal thickness might provide counter examples to the Birkhoff-Poritsky conjecture.

Acknowledgments. I want to thank E. Gutkin for discussions on billiards and caustics, P. Gruber for literature references, M. Wojtkowski for useful remarks on an earlier version of the manuscript and E. Amiran for helpful suggestions on the final draft. A first version of this paper was written at Caltech.

## References

- [1] V. Arnold. *Topological Invariants of Plane Curves and Caustics*, volume 5 of *University Lecture Series*. American Mathematical Society, 1994.
- [2] I.K. Babenko. Periodic trajectories in a 3-dimensional birkhoff billiard. *Math. USSR-Sbornik*, 71:1–13, 1992.
- [3] M. Berger. Sur les caustiques des surfaces en dimension 3. *C. R. Acad. Sc. Paris*, 311:76–85, 1990.
- [4] M. Berger. La mathématique du billard. *Pour la Science*, 163:76–85, 1991.
- [5] M. Berger and B. Gostiaux. *Differential Geometry: Manifolds, Curves and Surfaces*. Springer-Verlag, New York, Berlin, Heidelberg, 1988.
- [6] G.D. Chakerian and H. Groemer. Convex bodies of constant width. In *Handbook of convex geometry*. Elsevier Science Publishers, 1993.
- [7] Y. Collin de Verdière. Sur les longueurs des trajectoires périodiques d'un billiard. In *Géométrie symplectique et de contact autour du théorème de Poincaré-Birkhoff*, Travaux en cours. Hermann, éditeurs des sciences et des arts, Paris, 1984.
- [8] R. Douady. Application du théorème des tores invariantes. These 3 ème cycle, Université Paris VII, 1982.
- [9] K. Falconer. *Fractal Geometry, Mathematical Foundations and Applications*. John Wiley and Sons, Chichester, 1990.
- [10] P. Funk. Über Flächen mit lauter geschlossenen geodätischen Linien. *Math. Ann.*, 74:278–300, 1913.
- [11] P.M. Gruber. Baire categories in convexity. In *Handbook of convex geometry*. Elsevier Science Publishers, 1993.

- [12] P.M. Gruber. Only ellipsoids have caustics. *Math. Ann.*, 303:185–194, 1995.
- [13] E. Gutkin. Billiard tables of constant width and dynamical characterizations of the circle. Proceedings of the Penn. State Workshop, 1993.
- [14] E. Gutkin and A. Katok. Caustics for inner and outer billiards. *Commun. Math. Phys.*, 173:101–133, 1995.
- [15] A. Hubacher. Instability of the boundary in the billiard ball problem. *Commun. Math. Phys.*, 108:483–488, 1987.
- [16] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its applications*. Cambridge University Press, 1995.
- [17] J. Mather. Glancing billiards. *Ergod. Th. Dyn. Sys.*, 2:397–403, 1982.
- [18] J. Moser. *Stable and random Motion in dynamical systems*. Princeton University Press, Princeton, 1973.
- [19] J.C. Oxtoby. *Measure and Category*. Springer Verlag, New York, 1971.
- [20] H. Poritsky. The billiard ball problem on a table with a convex boundary-an illustrative dynamical problem. *Annals of Mathematics*, 51:456–470, 1950.
- [21] Ya.G. Sinai. *Introduction to ergodic theory*. Princeton University press, Princeton, 1976.
- [22] S. Tabachnikov. Around four vortices. *Russian Math. Surveys*, 45:229–230, 1990.
- [23] S. Tabachnikov. *Billiards*. Panoramas et synthèses. Société Mathématique de France, 1995.
- [24] D.V. Treshchev V.V. Kozlov. *Billiards*, volume 89 of *Translations of mathematical monographs*. AMS, 1991.

Department of Mathematics  
 University of Arizona  
 Tucson, AZ, 85721, USA